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

SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

SIXTH SEMESTER INTEGRATED MCA DEGREE EXAMINATION (R), MAY 2023
(2020 SCHEME)

Course Code: 20IMCAT304

Course Name: **Fundamentals of Data Science**

Max. Marks: **60**   **Duration: 3 Hours**

PART A
(Answer all questions. Each question carries 3 marks)

1. Explain the elements of big data analysis.
2. Differentiate classification and clustering.
3. Elaborate on the CAP theorem.
4. Enumerate the characteristics of NoSQL Databases.
5. Which are the different phases in data analytics life cycle?
6. Give a brief note on quantitative analysis.
7. What is the role of Zookeeper? Explain.
8. Compare Hadoop Yarn with Oozie.
9. Write short notes on any three applications of big data.
10. Discuss the functionalities of security layer.

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

MODULE I

11. Discuss on the knowledge discovery process.   (6)

OR

12. Explain any six data mining techniques.   (6)

MODULE II

13. Elaborate on the concept of sharding and its types.   (6)

OR

14. Explain the different In-memory storage devices.   (6)

MODULE III

15. Explain the different types of statistical analysis techniques.   (6)

OR

16. Illustrate visual analysis techniques.   (6)

MODULE IV

17. Describe the HDFS architecture with a neat diagram.   (6)

OR

18. With a suitable example, explain MapReduce.   (6)

MODULE V

19. Elaborate on the importance of data source layer and ingestion layer in processing big data.   (6)

OR

20. Write short note on
    a. Storage layer   (6)
    b. Physical infrastructure layer

****