

**THIRD SEMESTER B.Sc. DEGREE EXAMINATION, NOVEMBER 2023**  
**COMPUTER SCIENCE AND MATHEMATICS (DOUBLE MAIN)**  
**GDCS3A01T: INTRODUCTION TO DATA SCIENCE**

**Time: 2 ½ Hours**

**Maximum Marks: 80**

**SECTION A: Answer the following questions. Each carries *two* marks.**  
**(Ceiling 25 Marks)**

1. What is data?
2. Define Apache spark.
3. What is streaming data?
4. Define ordinal data with example.
5. List out any four sources of big data.
6. What is meant by batch processing?
7. Comment on data modeling.
8. State the use of logical model.
9. What is meant by secondary data?
10. Mention the use of RDF graphs.
11. What is the need for security in data science?
12. How can consistency and trust be related?
13. What is meant by data ownership?
14. List out some benefits of data security.
15. What do you mean by data masking?

**SECTION B: Answer the following questions. Each carries *five* marks.**  
**(Ceiling 35 Marks)**

16. Explain the advantages of data science.
17. Describe any three facets of data with example.
18. Explain any four advantages of NoSQL?
19. Discuss about spark framework.
20. Explain on Relational modeling.
21. Discuss the source of big data.
22. What is the future scope of data science?
23. “An ethical challenge should be part of the hiring process”. Substantiate.

**SECTION C: Answer any *two* questions. Each carries *ten* marks.**

24. What is data? Explain the types of data.
25. What is big data? Explain key components of big data ecosystem.
26. With neat diagram explain spark frameworks.
27. Explain the five Cs of data in detail? Also explain how to implement five Cs.

**(2 x 10 = 20 Marks)**