

NOVEMBER 2023

51432/SU45A/SE26B

Time : Three hours

Maximum : 75 marks

PART A — ($10 \times 2 = 20$ marks)

Answer any TEN questions each in 30 words.

1. What is Data Science?
2. What are the characteristics of big data?
3. Define data cleansing.
4. What is an Outliers?
5. What are the components of model building phase?
6. What are the different phases of modeling process?
7. What is an example of semi-supervised learning?
8. What is MapReduce?
9. What is CAP Theorem?
10. What are the benefits of NoSQL?
11. What is data mart?
12. What are the techniques to handle missing data?

PART B — ($5 \times 5 = 25$ marks)

Answer any FIVE questions each in 200 words.

13. Explain the benefits of Data Science.
14. Explain different types of databases used in data science.
15. How to retrieve the data in Data Science?
16. Why we need the machine learning in Data Science?
17. Write short notes on BASE principles of NoSQL.
18. Discuss the limitations of Hadoop.
19. Briefly explain about data preparation phase of data science process.

PART C — ($3 \times 10 = 30$ marks)

Answer any THREE questions each in 500 words.

20. Explain the data science process with a neat diagram.
21. Explain Exploratory data analysis phase of data science process in detail;
22. Explain the Supervised machine learning in detail.

23. Explain Spark components in detail.
 24. Explain the following phases of data science process:
 - (a) Setting the research goal
 - (b) Presentation and automation
-