

**M. Sc. DEGREE EXAMINATION, APRIL 2024**  
**BIOINFORMATICS**  
**FOURTH SEMESTER**

**COURSE : CORE**  
**PAPER : BIG DATA ANALYSIS**  
**SUBJECT CODE : 19BI/PC/BD44**  
**TIME : 3 HOURS** **MAX. MARKS: 100**

**SECTION – A**

**ANSWER ALL QUESTIONS** **(20 X 1=20)**

1. How Does a Data Scientist Work?
2. What is probabilistic Modelling in NLP?
3. Why is image analysis important in the medical field?
4. Define network modelling.
5. List out the basic concepts of big data?
6. Summarize important features that a big data analysis platform needs?
7. What is big data's role in medical decision making?
8. List out the advantages of machine-generated data?
9. What are the 5 characteristics of big data?
10. Classify data in to two types.
11. List out five goals of exploratory data analysis
12. Compare data and big data?
13. What is communication in data analysis?
14. Name the 5 P's of data science
15. List out the steps involved in data preprocessing
16. What are the 4 big data strategies?
17. Summarize the features of MapReduce
18. Name any 4 components of Hadoop?
19. List out the different types of scalability in cloud computing?
20. Which programming language is used for big data?

**SECTION – B**

**ANSWER ANY FOUR QUESTIONS..** **(4 x 10 = 40)**

21. Categorize the Probabilistic Models in Machine Learning.
22. Briefly describe about major challenges in managing big data.
23. Explain the types of Big Data.
24. Summarize on 5-step process to structure your analysis.
25. List out characteristics of big data.
26. Elaborate on the significance of data science process.
27. Discuss the role of cloud computing in big data analytics.

**SECTION – C**

**ANSWER ANY TWO QUESTIONS..**

**(2 x 20 = 40)**

28. Discuss the Impact of Data Science on Healthcare.
29. Elaborate the Types and benefits of Machine Data.
30. Explain the steps in preprocessing and exploring data.
31. Summarize the YARN Features and architecture in detail.

\*\*\*\*\*