

Code No: MB194D1 /19

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY- GURAJADA VIZIANAGARAM

MBA IV Semester Regular/Supplementary Examinations, May-2025

Big Data Analytics (Systems)

Time: 3 Hours

Max. Marks: 75

Answer any FIVE Questions One Question From Each Unit

All Questions Carry Equal Marks

UNIT-I

- | | | | |
|----|---|---|----|
| 1. | a | Define Big Data. Discuss the five V's of Big Data with examples. | 6M |
| | b | Explain the architecture of Hadoop and describe the role of its core components | 6M |

OR

- | | | | |
|----|---|--|----|
| 2. | a | Compare data storage in traditional data warehouses vs Hadoop-based systems. | 6M |
| | b | Describe the evolution of Hadoop. What are the challenges that led to the development of Hadoop? | 6M |

UNIT-II

- | | | | |
|----|---|---|----|
| 3. | a | Describe the HDFS architecture and explain how data is stored and retrieved. | 6M |
| | b | Explain the concept of Hadoop I/O. How are compression and serialization handled in Hadoop? | 6M |

OR

- | | | | |
|----|---|--|----|
| 4. | a | What is the role of distcp in Hadoop? Illustrate with a scenario where distcp is beneficial. | 6M |
| | b | Explain the process of reading and writing files using the Hadoop Java API. Provide a sample code snippet. | 6M |

UNIT-III

- | | | | |
|----|---|---|----|
| 5. | a | Write a basic MapReduce program to count word frequency in a given dataset. Explain each stage clearly. | 6M |
| | b | What are the advantages of using Hadoop Streaming in MapReduce programming? | 6M |

OR

- | | | | |
|----|---|---|----|
| 6. | a | Explain the concept of chaining MapReduce jobs. How is data passed between jobs? | 6M |
| | b | Discuss how data from different sources can be joined using MapReduce. Provide a high-level algorithm or pseudo-code. | 6M |

UNIT-IV

- | | | | |
|----|---|---|----|
| 7. | a | Compare Hive with traditional RDBMS in terms of architecture and data querying. | 6M |
| | b | Explain the function of Hive Metastore and User Defined Functions (UDFs) in Hive with examples. | 6M |

OR

- | | | | |
|----|---|---|----|
| 8. | a | Describe the differences between HBase and RDBMS. What are the key advantages of using HBase for unstructured data? | 6M |
| | b | Write a short note on Big SQL and its relevance in the Big Data ecosystem. | 6M |

UNIT-V

- | | | | |
|----|---|---|----|
| 9. | a | Explain the PageRank algorithm with an example. How is it implemented using MapReduce? | 6M |
| | b | Define Bloom Filters. How do they help in reducing false positives in large data systems? | 6M |

OR

- | | | | |
|-----|---|--|----|
| 10. | a | Discuss the various types of machine learning techniques in R used for Big Data Analytics. | 6M |
| | b | What is BigR? How does it support Big Data Analytics? Give suitable examples. | 6M |

CASE STUDY

- 11 An e-commerce company wants to enhance its recommendation system, detect fraudulent transactions, and analyze customer reviews for sentiment analysis. 15M
- a)** Propose a Big Data solution architecture for implementing the above features.
 - b)** Identify appropriate Hadoop ecosystem tools for each task and justify their use.
 - c)** Suggest how R can be integrated to perform advanced analytics on the data.