Short Questions

Unit – 3

1. What is Hadoop Streaming, and how does it enable the use of non-Java programming languages with Hadoop?

2. Explain the role of Hadoop in handling unstructured data in Big Data analytics.

3. What is the Hadoop Distributed File System (HDFS) block size, and why is it important?

4. How does Hadoop's data replication strategy contribute to fault tolerance?

5. What is Hadoop's speculative execution, and why is it used in MapReduce jobs?

6. How does Hadoop address the problem of data locality in distributed computing?

7. Can you explain the purpose of Hadoop's Secondary NameNode?

8. What is the role of Apache Hive in the Hadoop ecosystem, and how does it simplify data querying and analysis?

9. Explain the differences between Hadoop and traditional relational databases in handling Big Data.

10. What are the advantages of using Hadoop's HBase for NoSQL data storage and real-time processing?

11. How does Hadoop support data governance and compliance in enterprise settings?

12. Explain the role of ZooKeeper in managing distributed applications within the Hadoop ecosystem.

13. Can you describe the process of scaling a Hadoop cluster, and what factors should be considered when planning for scalability?

14. How does Hadoop handle data shuffling and sorting during the MapReduce process?

15. Explain the concept of speculative execution in the context of Hadoop's fault tolerance mechanism.

16. How does Hadoop's architecture contribute to horizontal scalability?

17. What is the purpose of the ResourceManager in YARN, and how does it manage cluster resources?

18. How does Hadoop handle data compression, and what are the benefits of compressing data in HDFS?

19. Explain the concept of data skew in Hadoop, and how can it be mitigated during data processing?

20. What is the purpose of Hadoop's ResourceManager and NodeManager components in the YARN architecture?

21. Can you explain the differences between the Hadoop 1.x and Hadoop 2.x architectures and their implications for Big Data processing?

22. What is the role of Apache Pig in the Hadoop ecosystem, and how does it simplify data processing tasks?

23. How does Hadoop ensure data security, and what authentication and authorization mechanisms does it support?

24. What is the purpose of the Hadoop Ecosystem, and how do its components work together to provide a comprehensive Big Data solution?

25. How does Hadoop handle data replication and redundancy, and what are the benefits of this approach in a distributed system?

Unit – 4

26. What is Hadoop?
27. How does Hadoop differ from traditional RDBMS?
28. What is HDFS in Hadoop?
29. Can you name some Hadoop distributors?
30. What are HDFS Daemons?
31. How does file writing work in HDFS?
32. How does file reading work in HDFS?
33. What is the NameNode in HDFS?
34. What is the role of the Secondary NameNode?
35. What is the purpose of DataNode in HDFS?
36. Can you describe the HDFS architecture?
37. How is Hadoop configured?
38. What is the MapReduce framework?
39. How does HBase contribute to big data processing?
40. What is Hive in Hadoop ecosystem?
41. Can you explain what Pig is in Hadoop?

42. What is the difference between HDFS and a regular filesystem?
43. Why is Hadoop considered good for big data processing?
44. What is a Hadoop Cluster?
45. How does Hadoop ensure data reliability and fault tolerance?
46. What is YARN in Hadoop?
47. How is load balancing achieved in HDFS?
48. Can you explain block replication in HDFS?
49. What is a Data Lake and how does Hadoop relate to it?
50. How does Hadoop handle concurrent read and write operations?
51. What is the significance of the Hadoop ecosystem?
52. How does Hadoop contribute to cost savings in data processing
53. What are some common use cases for Hadoop?
54. How does Hadoop support scalability?
55. What is a NameNode federation in HDFS?
56. What is a rack awareness in HDFS?
57. How does Hadoop handle hardware failures?
58. What is speculative execution in Hadoop?

59. What is the role of Zookeeper in Hadoop?

60. What is the primary function of the NameNode in HDFS?

61. How does Secondary NameNode assist the NameNode?

62. Can you explain the role of a DataNode in HDFS?

63. What is the Hadoop MapReduce framework?

64. How does HBase fit into the Hadoop ecosystem?

65. What is HIVE in Hadoop?

66. Explain the role of PIG in Hadoop.

67. What is the purpose of a JobTracker in Hadoop?

68. Describe the process of writing a file in HDFS.

69. How does reading a file work in HDFS?

70. Can you explain the concept of Hadoop Rack Awareness?

71. What is a TaskTracker in Hadoop?

72. What are the benefits of using Hadoop for big data processing?

73. How does Hadoop ensure data replication and reliability?

74. What is the significance of YARN in Hadoop

75. How does Hadoop differ from traditional RDBMS in data processing?

Unit – 5

76. What is R used for in data analytics?

77. How do you import data into R for analysis?

78. Can R handle large datasets?

79. What is machine learning in simple terms?

80. How is machine learning implemented in R?

81. What is supervised learning in machine learning?

82. Can you give an example of a supervised learning algorithm in R?

83. What's the difference between classification and regression in supervised learning?

84. What is unsupervised learning?

85. Can you name a common unsupervised learning method in R?

86. What is collaborative filtering in simple terms?

87. How is collaborative filtering used in machine learning?

88. What is social media analytics?

89. How can R be used for social media analytics?

90. What does mobile analytics focus on?

91. How is data collected for mobile analytics?

92. What is BigR in the context of big data analytics?

93. 93 How does BigR help in big data analytics?

94. Can BigR integrate with Hadoop ecosystems?

95. What is a key advantage of using R for big data analytics?

96. How does BigR handle memory management with large datasets?

97. Is R suitable for real-time big data analytics?

98. Can BigR be used for predictive modeling on big data?

99. How do you visualize big data results in R?

100. What are the challenges of using R with big data?

101. Can BigR handle streaming data?

102. How does BigR ensure data security and privacy?

103. Is R's BigR package user-friendly for beginners in big data analytics?

104. How scalable is R's BigR package for growing data needs?

105. Can BigR be integrated with cloud-based big data solutions?

106. What type of data formats can BigR handle?

107. How does BigR compare to Python's big data tools?

108. Can BigR handle complex machine learning algorithms on big data?

109. What is the role of data visualization in BigR analytics?

110. How does BigR handle data from different sources?

111. Can BigR be used for time-series analysis on big data?

112. What kind of statistical methods can be applied using BigR?

113. How important is data preprocessing in BigR analytics?

114. Can BigR be used for sentiment analysis on big datasets?

115. How does BigR ensure accuracy in data analysis?

116. Is BigR suitable for academic research involving big data?

117. Can BigR be used for geospatial data analysis?

118. How does BigR handle missing data in large datasets?

119. Can you perform network analysis with BigR?

120. How do machine learning and BigR interact in data analysis?

121. What are the limitations of using R and BigR in enterprise settings?

122. How does BigR manage data lineage and auditing?

123. Is BigR a good choice for real-time analytics?

124. How do you ensure data quality when using BigR?

125. Can BigR handle text analytics on large text datasets?