Long Questions

Unit – 3

1. How does Hadoop integrate with other big data technologies?
2. Explain the importance of Hadoop's scalability in big data processing.
3. What are the functionalities of Apache Ambari in the Hadoop ecosystem?
4. Describe the impact of Hadoop on data warehousing.
5. How does Hadoop contribute to cost-effective big data processing?
6. What is the role of Hadoop in data democratization?
7. How does Hadoop handle data privacy and compliance?
8. Describe the role of Apache Storm in real-time data processing in Hadoop.
9. Explain the significance of Apache Kafka in the Hadoop ecosystem.
10. How does Hadoop support data recovery and fault tolerance?
11. Discuss the role of machine learning in the Hadoop ecosystem.
12. What is the impact of Hadoop on cloud computing and big data?
13. How does Hadoop assist in predictive analytics?
14. Discuss Hadoop’s role in IoT (Internet of Things) data processing.
15. How does Hadoop enable multi-structured data processing?
16. What is the role of Apache Oozie in the Hadoop ecosystem?
17. Describe how Hadoop supports data governance and metadata management.
18. How is Hadoop utilized in business intelligence (BI) and analytics?
19. Discuss the evolution of Hadoop and its future prospects in big data.
20. Explain the role of data lakes in the Hadoop ecosystem.
21. How does Hadoop enable effective ETL (Extract, Transform, Load) processes?
22. Describe the role of Apache Tez in the Hadoop ecosystem.

23. How does Hadoop facilitate data archiving and historical analysis?

24. Discuss the challenges of data quality and consistency in Hadoop.

25. Explain the role of Apache Nifi in the Hadoop ecosystem.

Unit – 4

26. How does Hadoop's architecture differ from traditional RDBMS in handling large datasets?

27. What are the key components of Hadoop's ecosystem, and how do they contribute to its functionality?

28. Can you explain the role of Hadoop distributors and their impact on Hadoop's adoption in the industry?

29. What is HDFS, and how does it enable high-throughput access to application data?

30. How do HDFS daemons function, and what are their responsibilities in the Hadoop ecosystem?

31. Can you detail the process involved in the anatomy of file write and read in HDFS?

32. What is the role of the NameNode in HDFS, and how does it manage the filesystem namespace?

33. How does the Secondary NameNode enhance the reliability and efficiency of the Hadoop cluster?

34. What functions do DataNodes serve in HDFS, and how do they interact with the NameNode?

35. Can you explain the overall architecture of HDFS and its key design principles?

36. How is Hadoop configured for different environments, and what are the key configuration parameters?

37. What is the MapReduce framework, and how does it process large data sets in Hadoop?

38. How does HBase integrate with Hadoop for Big Data processing, and what are its unique features?

39. What role does Hive play in the Hadoop ecosystem, and how does it facilitate data warehousing?

40. How does Pig complement other Hadoop components in processing large datasets?

41. In what ways is Hadoop's architecture more scalable and fault-tolerant compared to traditional RDBMS?

42. What are the common challenges faced when deploying Hadoop in large-scale environments?

43. How does data replication work in HDFS, and what are its advantages?

44. Can you describe the process of a typical MapReduce job execution in Hadoop?

45. What are the security mechanisms in place within the Hadoop ecosystem?

46. How does Hadoop's ecosystem support real-time data processing?

47. What are the significant differences in data modeling between Hadoop and traditional RDBMS?

48. How do Hadoop's data processing capabilities benefit businesses in terms of insights and decision-making?

49. What are the limitations of Hadoop in handling certain types of data or computational tasks?

50. How has the Hadoop ecosystem evolved over the years, and what future enhancements are anticipated?

Unit – 5

51. How can R be used for predictive modeling in supervised machine learning, and what are some common algorithms for this purpose?

52. What are the key differences between supervised and unsupervised learning in the context of data analytics with R?

53. How can unsupervised learning techniques in R be applied to identify patterns and clusters in large datasets?

54. What are some best practices for preprocessing data in R for machine learning applications?

55. Can you explain the concept of overfitting in machine learning and how it can be addressed using R?

56. How does collaborative filtering work in the context of recommendation systems, and how can it be implemented using R?

57. What are the challenges and advantages of using R for social media analytics, particularly for sentiment analysis?

58. How can R be used to analyze mobile analytics data, and what unique considerations are there for this type of data?

59. What is BigR in the context of big data analytics, and how does it integrate with R?

60. How can R be leveraged for real-time analytics in a big data environment?

61. What are some visualization techniques in R that are particularly useful for interpreting machine learning model results?

62. How can time series analysis be conducted in R for forecasting purposes, and what are some common methods?

63. What role does feature selection play in building effective machine learning models in R, and how can it be performed?

64. How can text mining be applied in R for natural language processing, especially in the context of social media data?

65. What are some common challenges in mobile analytics data interpretation, and how can R help to overcome these challenges?

66. How can big data analytics with BigR be integrated with other technologies like Hadoop or Spark?

67. What are some methods for evaluating the performance of a machine learning model in R?

68. How can R be used to implement and evaluate different types of neural networks for deep learning applications?

69. What are some approaches for handling missing or incomplete data in R for machine learning projects?

70. How can R be utilized for geospatial data analysis in the context of mobile analytics?

71. What are the best practices for scaling R-based machine learning models to handle large datasets?

72. How can R be applied for customer segmentation and personalization in e-commerce using collaborative filtering techniques?

73. Can you discuss a case study where R was used effectively for big data analytics in a business context?

74. What are some advanced regression techniques available in R for predictive modeling?

75. How can R be used for anomaly detection in large datasets, and what are some of the key algorithms for this purpose?