Short Questions

Unit - I:

1. What are the types of digital data according to classification?
2. How has big data evolved over time?
3. Define big data in the context of modern data analysis.
4. How does traditional business intelligence differ from big data analytics?
5. Can big data and data warehouses coexist? Explain.
6. What is the significance of big data analytics in modern business?
7. Explain what big data analytics isn't.
8. Why has there been sudden hype around big data analytics?
9. How can analytics be classified in the context of big data?
10. What are some of the greatest challenges businesses face in capitalizing on big data?
11. List the top challenges that big data faces in implementation.
12. Why is big data analytics important in the contemporary business landscape?
13. Define data science and its relation to big data analytics.
14. What are some terminologies commonly used in big data environments?
15. How do organizations utilize big data analytics for strategic decision-making?
16. Can you differentiate between structured and unstructured data in big data analytics?
17. Explain the concept of data lakes in big data ecosystems.
18. What role does machine learning play in big data analytics?
19. How does real-time analytics contribute to big data processing?
20. Discuss the scalability challenges associated with big data analytics.
21. What are the ethical considerations in big data analytics?
22. How do privacy concerns impact big data analytics practices?
23. Explain the concept of predictive analytics in big data.
24. How does big data analytics contribute to personalized marketing strategies?
25. What are the security implications of handling massive amounts of data in big data analytics?
26. Discuss the role of cloud computing in facilitating big data analytics.
27. How does data visualization aid in understanding big data insights?
28. What are some common tools and platforms used in big data analytics?
29. How does big data analytics empower industries like healthcare and finance?
30. Discuss the challenges of data governance in big data environments.

31. What is the role of data preprocessing in big data analytics?
32. Explain the importance of data quality in big data analytics.
33. How does big data analytics help in risk management?
34. Discuss the impact of big data analytics on customer relationship management.
35. Can you explain the concept of sentiment analysis in big data?
36. What are some emerging trends in big data analytics?
37. How do businesses tackle the issue of data silos in big data analytics?
38. What role does natural language processing play in analyzing unstructured data?
39. Discuss the concept of data mining in the context of big data analytics.
40. How does big data analytics contribute to supply chain optimization?
41. What are some challenges associated with integrating big data analytics into existing IT infrastructure?
42. How do businesses address the issue of data bias in big data analytics?
43. Explain the concept of prescriptive analytics in big data.
44. What are the economic implications of big data analytics?
45. How does big data analytics enhance fraud detection mechanisms?
46. Discuss the role of data fusion in combining heterogeneous data sources in big data analytics.
47. What ethical guidelines should be followed in conducting big data analytics?
48. How does big data analytics contribute to improving operational efficiency?
49. Explain the concept of dark data and its relevance in big data analytics.
50. What are some challenges in ensuring regulatory compliance in big data analytics?

Unit -II:

51. What are the key features of Hadoop?
52. What are the key advantages of using Hadoop?
53. Can you list some versions of Hadoop?
54. Provide an overview of the Hadoop ecosystem.
55. What are Hadoop distributions, and why are they important?
56. Why is there a need for Hadoop in modern data processing?
57. How does Hadoop compare to traditional RDBMS systems?
58. What challenges does distributed computing present, and how does Hadoop address them?
59. Can you outline the history of Hadoop's development?

60. Could you provide an overview of Hadoop's architecture?
61. What is HDFS, and how does it fit into the Hadoop ecosystem?
62. How does Hadoop handle fault tolerance in distributed environments?
63. What role does MapReduce play in Hadoop's processing model?
64. How does Hadoop ensure scalability for handling large datasets?
65. Can you explain the concept of data locality in Hadoop?
66. What are some common components of the Hadoop ecosystem beyond HDFS and MapReduce?
67. How does Hadoop handle parallel processing of data across distributed nodes?
68. What security measures does Hadoop offer to protect data integrity and privacy?
69. How does Hadoop support various data formats and structures?
70. What are the main components of Hadoop's resource management framework?
71. How does Hadoop support high availability and reliability in data processing?
72. What are some popular Hadoop distributions available in the market?
73. How does Hadoop address the challenges of processing unstructured and semi-structured data?
74. What are the limitations of using Hadoop for certain types of data processing tasks?
75. Can you explain the concept of YARN in the Hadoop ecosystem?
76. How does Hadoop handle data compression and optimization for storage efficiency?
77. What role do Hadoop ecosystem projects like Hive and Pig play in data processing?
78. How does Hadoop handle data replication for fault tolerance and data redundancy?
79. What are some considerations when choosing a Hadoop distribution for enterprise deployment?
80. How does Hadoop support batch processing of large datasets?
81. What role does ZooKeeper play in managing distributed systems within the Hadoop ecosystem?
82. How does Hadoop address data processing bottlenecks and performance optimization?
83. Can you compare and contrast Hadoop's approach to distributed computing with other frameworks?
84. How does Hadoop handle data ingestion from various sources into its ecosystem?

85. What are the differences between Hadoop's storage model and traditional file systems?
86. How does Hadoop ensure data consistency and coherence in distributed processing?
87. What are some use cases where Hadoop is particularly well-suited for data processing tasks?
88. How does Hadoop handle data shuffling and sorting in MapReduce jobs?
89. What are some challenges organizations may face when adopting Hadoop for their data processing needs?
90. How does Hadoop support data governance and regulatory compliance requirements?
91. What are some considerations for optimizing resource utilization in a Hadoop cluster?
92. How does Hadoop handle schema evolution and data schema flexibility?
93. Can you explain the concept of speculative execution in Hadoop's processing model?
94. How does Hadoop integrate with existing data management and analytics tools?
95. What are the advantages of using Hadoop for processing large-scale graph data?
96. How does Hadoop support data replication across multiple data centers for disaster recovery?
97. Can you describe the role of Hadoop's ecosystem projects like Spark and HBase in data processing?
98. What are some emerging trends in the evolution of the Hadoop ecosystem?
99. How does Hadoop handle resource scheduling and job prioritization in a multi-tenant environment?
100. What are some best practices for deploying and managing Hadoop clusters in production environments?

Unit - III:

101. What is the process of processing data with Hadoop?
102. Can you introduce the concept of MapReduce programming?
103. What are the main components of a MapReduce program?
104. Explain the role of the Mapper in MapReduce programming.
105. What is a Reducer, and how does it function in MapReduce?
106. Describe the purpose and function of a Combiner in MapReduce.

107. How does a Combiner help optimize MapReduce jobs?
108. What is the significance of a Partitioner in MapReduce?
109. How does a Partitioner contribute to the efficiency of data processing in MapReduce?
110. Can you outline the steps involved in processing data with Hadoop using MapReduce?
111. What are the different types of NoSQL databases?
112. Explain the advantages of using NoSQL databases.
113. How is NoSQL utilized in various industries?
114. What are the key differences between SQL and NoSQL databases?
115. Can you introduce the concept of NewSQL?
116. How does NewSQL differ from traditional SQL databases?
117. Compare and contrast NoSQL, SQL, and NewSQL databases.
118. What are the primary characteristics of NoSQL databases?
119. Discuss the scalability features of NoSQL databases.
120. How do NoSQL databases handle unstructured data?
121. Explain the flexibility of schema design in NoSQL databases.
122. What are the main considerations when choosing between SQL and NoSQL databases?
123. How does the CAP theorem apply to NoSQL databases?
124. What are some popular use cases for NoSQL databases?
125. How does the performance of NoSQL databases compare to SQL databases?