Multiple Choice Q&A

Unit -I:

1. What is the classification of digital data?
a) Structured and unstructured
b) Analog and digital
c) Binary and hexadecimal
d) Textual and numerical
Answer: a) Structured and unstructured

2. What is the definition of big data?
a) Data that is extremely large in size
b) Data that is complex and difficult to process
c) Data that exceeds the processing capacity of conventional databases
d) All of the above
Answer: d) All of the above

3. How does traditional business intelligence differ from big data?
a) Traditional BI focuses on historical data, while big data focuses on real-time data
b) Traditional BI relies on structured data, while big data can handle both structured and unstructured data
c) Traditional BI uses relational databases, while big data uses distributed systems like Hadoop
d) All of the above
Answer: d) All of the above

4. What is the coexistence of big data and data warehouse?
a) Big data replaces data warehouses completely
b) Data warehouses are used exclusively for big data analytics
c) Big data and data warehouses complement each other in an organization's data strategy
d) Big data and data warehouses are mutually exclusive concepts
Answer: c) Big data and data warehouses complement each other in an organization's data strategy

5. What is the importance of big data analytics?
a) Identifying trends and patterns in data

b) Making data-driven decisions
c) Improving operational efficiency
d) All of the above
Answer: d) All of the above

6. What is not considered as big data analytics?
a) Predictive modeling
b) Traditional business reporting
c) Data visualization
d) Natural language processing
Answer: b) Traditional business reporting

7. What contributes to the sudden hype around big data analytics?
a) Increased availability of data sources
b) Advancements in data processing technologies
c) Growing demand for data-driven insights
d) All of the above
Answer: d) All of the above

8. How are analytics classified?
a) Descriptive, diagnostic, predictive, prescriptive
b) Structured, unstructured, semi-structured
c) Batch processing, real-time processing
d) None of the above
Answer: a) Descriptive, diagnostic, predictive, prescriptive

9. What are the greatest challenges that prevent businesses from capitalizing on big data?
a) Lack of skilled personnel
b) Privacy and security concerns
c) Data integration issues
d) All of the above
Answer: d) All of the above

10. Which of the following is not a top challenge facing big data?
a) Scalability
b) Data quality
c) Cost of storage

d) Lack of data sources
Answer: d) Lack of data sources

11. What is the primary goal of big data analytics?
a) To process data faster than traditional methods
b) To make data more accessible to decision-makers
c) To uncover hidden patterns and insights in large datasets
d) To replace traditional business intelligence systems
Answer: c) To uncover hidden patterns and insights in large datasets

12. What is the term used to describe the interdisciplinary field that uses scientific methods, algorithms, and systems to extract knowledge and insights from structured and unstructured data?
a) Big data analytics
b) Data mining
c) Data science
d) Machine learning
Answer: c) Data science

13. Which of the following is not a terminology commonly used in big data environments?
a) MapReduce
b) Hadoop
c) SQL
d) NoSQL
Answer: c) SQL

14. What type of data is characterized by having a defined and predictable structure?
a) Unstructured data
b) Semi-structured data
c) Structured data
d) None of the above
Answer: c) Structured data

15. Which of the following statements about big data is true?
a) Big data refers to data that can be processed using traditional database technologies

b) Big data analytics always requires real-time processing
c) Big data is characterized by the volume, velocity, and variety of data
d) Big data is primarily used for storing structured data
Answer: c) Big data is characterized by the volume, velocity, and variety of data

16. In big data analytics, what does the term "velocity" refer to?
a) The speed at which data is generated and processed
b) The size of the data being analyzed
c) The variety of data sources
d) The accuracy of the data
Answer: a) The speed at which data is generated and processed

17. What is the primary function of traditional business intelligence systems?
a) To analyze unstructured data
b) To provide real-time analytics
c) To support operational decision-making
d) To replace big data analytics
Answer: c) To support operational decision-making

18. Which of the following is an example of unstructured data?
a) Customer transaction records in a database
b) Sensor data from IoT devices
c) Text documents
d) Structured financial data
Answer: c) Text documents

19. What technology is commonly used for processing and analyzing large volumes of unstructured data?
a) Hadoop
b) SQL
c) Relational databases
d) NoSQL
Answer: a) Hadoop

20. What is the primary goal of descriptive analytics?
a) To explain why certain events occur
b) To predict future outcomes
c) To summarize historical data

d) To prescribe actions based on data insights
Answer: c) To summarize historical data

21. Which of the following is not a challenge associated with big data analytics?
a) Data privacy and security
b) Lack of data sources
c) Scalability
d) Complexity of data integration
Answer: b) Lack of data sources

22. What term is used to describe the process of extracting meaningful information from text data?
a) Data mining
b) Text analytics
c) Predictive modeling
d) Data visualization
Answer: b) Text analytics

23. What type of analytics focuses on predicting future outcomes based on historical data?
a) Descriptive analytics
b) Diagnostic analytics
c) Predictive analytics
d) Prescriptive analytics
Answer: c) Predictive analytics

24. What is one of the main reasons behind the sudden hype around big data analytics?
a) Decrease in data storage costs
b) Increased demand for real-time analytics
c) Advancements in data processing technologies
d) All of the above
Answer: d) All of the above

25. In big data analytics, what does the term "variety" refer to?
a) The number of data sources being analyzed
b) The accuracy of the data

c) The diversity of data types and formats
d) The speed at which data is generated
Answer: c) The diversity of data types and formats

26. What is the primary difference between big data analytics and traditional business intelligence?
a) Big data analytics focuses on real-time data, while traditional BI focuses on historical data.
b) Big data analytics relies on relational databases, while traditional BI uses distributed systems.
c) Big data analytics can handle both structured and unstructured data, while traditional BI focuses only on structured data.
d) Big data analytics uses machine learning algorithms, while traditional BI uses statistical methods.
Answer: c) Big data analytics can handle both structured and unstructured data, while traditional BI focuses only on structured data.

27. What term describes the process of combining and analyzing data from multiple sources to create unified datasets?
a) Data integration
b) Data visualization
c) Data preprocessing
d) Data transformation
Answer: a) Data integration

28. Which of the following is NOT a common challenge faced by businesses in capitalizing on big data?
a) Lack of skilled personnel
b) High cost of data storage
c) Privacy and security concerns
d) Inability to process real-time data
Answer: d) Inability to process real-time data

29. What role does data science play in big data analytics?
a) Data science focuses on collecting and storing large volumes of data.
b) Data science involves building and maintaining data warehouses.
c) Data science applies scientific methods and algorithms to extract insights from data.

d) Data science focuses on visualizing data for business stakeholders.
Answer: c) Data science applies scientific methods and algorithms to extract insights from data.

30. Which of the following terms is used to describe a technology that allows for the storage and processing of large datasets across distributed computing clusters?
a) SQL
b) Hadoop
c) MongoDB
d) Tableau
Answer: b) Hadoop

31. What is the significance of MapReduce in big data analytics?
a) MapReduce is a programming model used for distributed processing of large datasets.
b) MapReduce is a database management system for storing big data.
c) MapReduce is a data visualization tool for analyzing big data.
d) MapReduce is a machine learning algorithm used in big data analytics.
Answer: a) MapReduce is a programming model used for distributed processing of large datasets.

32. What term refers to the process of analyzing large datasets to uncover hidden patterns, unknown correlations, and other useful information?
a) Predictive modeling
b) Data mining
c) Descriptive analytics
d) Text analytics
Answer: b) Data mining

33. Which of the following is NOT a classification of analytics?
a) Descriptive analytics
b) Diagnostic analytics
c) Predictive analytics
d) Real-time analytics
Answer: d) Real-time analytics

34. What is the primary focus of prescriptive analytics?
a) Explaining historical data patterns

b) Predicting future outcomes
c) Recommending actions based on data insights
d) Summarizing large datasets
Answer: c) Recommending actions based on data insights

35. What term refers to the process of transforming raw data into a more usable format for analysis?
a) Data integration
b) Data preprocessing
c) Data visualization
d) Data transformation
Answer: b) Data preprocessing

36. Which of the following is an example of unstructured data?
a) Excel spreadsheet
b) Relational database
c) Email messages
d) Transactional data
Answer: c) Email messages

37. What is the primary goal of descriptive analytics?
a) Predicting future outcomes
b) Explaining historical data patterns
c) Recommending actions based on data insights
d) Summarizing large datasets
Answer: d) Summarizing large datasets

38. What technology is commonly used for real-time processing of streaming data in big data analytics?
a) Hadoop
b) Spark
c) MongoDB
d) Tableau
Answer: b) Spark

39. What term refers to the ability of a system to handle increasing amounts of work or data?
a) Scalability
b) Reliability

c) Efficiency
d) Flexibility
Answer: a) Scalability

40. What is the primary objective of data visualization in big data analytics?
a) To summarize historical data patterns
b) To predict future outcomes
c) To communicate insights effectively
d) To preprocess raw data
Answer: c) To communicate insights effectively

41. Which of the following is NOT a common challenge faced by businesses in capitalizing on big data?
a) Lack of skilled personnel
b) High cost of data storage
c) Lack of data sources
d) Inability to process real-time data
Answer: c) Lack of data sources

42. What role does data science play in big data analytics?
a) Data science focuses on collecting and storing large volumes of data.
b) Data science involves building and maintaining data warehouses.
c) Data science applies scientific methods and algorithms to extract insights from data.
d) Data science focuses on visualizing data for business stakeholders.
Answer: c) Data science applies scientific methods and algorithms to extract insights from data.

43. Which of the following terms is used to describe a technology that allows for the storage and processing of large datasets across distributed computing clusters?
a) SQL
b) Hadoop
c) MongoDB
d) Tableau
Answer: b) Hadoop

44. What is the significance of MapReduce in big data analytics?

a) MapReduce is a programming model used for distributed processing of large datasets.
b) MapReduce is a database management system for storing big data.
c) MapReduce is a data visualization tool for analyzing big data.
d) MapReduce is a machine learning algorithm used in big data analytics.
Answer: a) MapReduce is a programming model used for distributed processing of large datasets.

45. What term refers to the process of analyzing large datasets to uncover hidden patterns, unknown correlations, and other useful information?
a) Predictive modeling
b) Data mining
c) Descriptive analytics
d) Text analytics
Answer: b) Data mining

46. Which of the following is NOT a classification of analytics?
a) Descriptive analytics
b) Diagnostic analytics
c) Predictive analytics
d) Real-time analytics
Answer: d) Real-time analytics

47. What is the primary focus of prescriptive analytics?
a) Explaining historical data patterns
b) Predicting future outcomes
c) Recommending actions based on data insights
d) Summarizing large datasets
Answer: c) Recommending actions based on data insights

48. What term refers to the process of transforming raw data into a more usable format for analysis?
a) Data integration
b) Data preprocessing
c) Data visualization
d) Data transformation
Answer: b) Data preprocessing

49. Which of the following is an example of unstructured data?
a) Excel spreadsheet
b) Relational database
c) Email messages
d) Transactional data
Answer: c) Email messages

50. What is the primary goal of descriptive analytics?
a) Predicting future outcomes
b) Explaining historical data patterns
c) Recommending actions based on data insights
d) Summarizing large datasets
Answer: d) Summarizing large datasets

51. What are the key features of Hadoop?
a) Fault tolerance and reliability
b) Scalability
c) Distributed processing
d) All of the above
Answer: d) All of the above

52. What are the key advantages of Hadoop?
a) Cost-effectiveness
b) Ability to handle large volumes of data
c) Flexibility
d) All of the above
Answer: d) All of the above

53. How many major versions of Hadoop have been released so far?
a) 2
b) 3
c) 4
d) 5
Answer: c) 4

54. What is the Hadoop ecosystem?
a) A software platform for distributed computing
b) A collection of open-source projects related to Hadoop
c) A proprietary technology developed by a single company

d) None of the above
Answer: b) A collection of open-source projects related to Hadoop

55. What is the main purpose of Hadoop distributions?
a) To provide different versions of Hadoop software
b) To customize Hadoop for specific use cases
c) To simplify the deployment and management of Hadoop clusters
d) All of the above
Answer: c) To simplify the deployment and management of Hadoop clusters

56. What is the primary need for Hadoop in the context of big data?
a) To process structured data
b) To store data in relational databases
c) To handle large volumes of data that traditional systems cannot manage
d) None of the above
Answer: c) To handle large volumes of data that traditional systems cannot manage

57. What is the primary difference between RDBMS and Hadoop?
a) RDBMS uses SQL for querying, while Hadoop uses MapReduce.
b) RDBMS can handle structured data, while Hadoop can handle both structured and unstructured data.
c) RDBMS is a distributed computing platform, while Hadoop is a relational database management system.
d) None of the above
Answer: b) RDBMS can handle structured data, while Hadoop can handle both structured and unstructured data.

58. What is the main challenge of distribution computing?
a) Ensuring data security
b) Handling data redundancy
c) Managing communication and coordination among distributed nodes
d) None of the above
Answer: c) Managing communication and coordination among distributed nodes

59. What is the history of Hadoop's development?
a) It was initially developed by Google.

b) It was inspired by the MapReduce paper published by Google.
c) It was first released as an open-source project by Yahoo.
d) All of the above
Answer: d) All of the above

60. What does HDFS stand for?
a) Hadoop Distributed File System
b) Hadoop Database File System
c) Hadoop Data Formatting System
d) Hadoop Data Filtering System
Answer: a) Hadoop Distributed File System

61. What are the key features of Hadoop?
a) Fault tolerance and reliability
b) Scalability
c) Distributed processing
d) All of the above
Answer: d) All of the above

62. What are the key advantages of Hadoop?
a) Cost-effectiveness
b) Ability to handle large volumes of data
c) Flexibility
d) All of the above
Answer: d) All of the above

63. How many major versions of Hadoop have been released so far?
a) 2
b) 3
c) 4
d) 5
Answer: c) 4

64. What is the Hadoop ecosystem?
a) A software platform for distributed computing
b) A collection of open-source projects related to Hadoop
c) A proprietary technology developed by a single company
d) None of the above

Answer: b) A collection of open-source projects related to Hadoop

65. What is the main purpose of Hadoop distributions?
a) To provide different versions of Hadoop software
b) To customize Hadoop for specific use cases
c) To simplify the deployment and management of Hadoop clusters
d) All of the above
Answer: c) To simplify the deployment and management of Hadoop clusters

66. What is the primary need for Hadoop in the context of big data?
a) To process structured data
b) To store data in relational databases
c) To handle large volumes of data that traditional systems cannot manage
d) None of the above
Answer: c) To handle large volumes of data that traditional systems cannot manage

67. What is the primary difference between RDBMS and Hadoop?
a) RDBMS uses SQL for querying, while Hadoop uses MapReduce.
b) RDBMS can handle structured data, while Hadoop can handle both structured and unstructured data.
c) RDBMS is a distributed computing platform, while Hadoop is a relational database management system.
d) None of the above
Answer: b) RDBMS can handle structured data, while Hadoop can handle both structured and unstructured data.

68. What is the main challenge of distribution computing?
a) Ensuring data security
b) Handling data redundancy
c) Managing communication and coordination among distributed nodes
d) None of the above
Answer: c) Managing communication and coordination among distributed nodes

69. What is the history of Hadoop's development?
a) It was initially developed by Google.
b) It was inspired by the MapReduce paper published by Google.

c) It was first released as an open-source project by Yahoo.
d) All of the above
Answer: d) All of the above

70. What does HDFS stand for?
a) Hadoop Distributed File System
b) Hadoop Database File System
c) Hadoop Data Formatting System
d) Hadoop Data Filtering System
Answer: a) Hadoop Distributed File System

71. What are the key features of Hadoop?
a) Fault tolerance and reliability
b) Scalability
c) Distributed processing
d) All of the above
Answer: d) All of the above

72. What are the key advantages of Hadoop?
a) Cost-effectiveness
b) Ability to handle large volumes of data
c) Flexibility
d) All of the above
Answer: d) All of the above

73. How many major versions of Hadoop have been released so far?
a) 2
b) 3
c) 4
d) 5
Answer: c) 4

74. What is the Hadoop ecosystem?
a) A software platform for distributed computing
b) A collection of open-source projects related to Hadoop
c) A proprietary technology developed by a single company
d) None of the above
Answer: b) A collection of open-source projects related to Hadoop

75. What is the main purpose of Hadoop distributions?
a) To provide different versions of Hadoop software
b) To customize Hadoop for specific use cases
c) To simplify the deployment and management of Hadoop clusters
d) All of the above
Answer: c) To simplify the deployment and management of Hadoop clusters

76. Which of the following is NOT a key advantage of Hadoop?
a) Fault tolerance
b) Limited scalability
c) Cost-effectiveness
d) Ability to handle large volumes of data
Answer: b) Limited scalability

77. What is the main goal of the Hadoop ecosystem?
a) To provide a single software solution for all big data needs
b) To create a platform for developing custom big data applications
c) To integrate various open-source projects that work with Hadoop
d) None of the above
Answer: c) To integrate various open-source projects that work with Hadoop

78. Which of the following is NOT a commonly used Hadoop distribution?
a) Cloudera
b) Hortonworks
c) MongoDB
d) MapR
Answer: c) MongoDB

79. What is the primary purpose of Hadoop's distributed computing model?
a) To improve data security
b) To distribute data across multiple servers
c) To enable parallel processing of large datasets
d) None of the above
Answer: c) To enable parallel processing of large datasets

80. Who initially developed Hadoop?
a) Google

b) Facebook
c) Yahoo
d) Microsoft
Answer: c) Yahoo

81. What is the role of Hadoop Distributed File System (HDFS) in Hadoop?
a) To store data in a relational database format
b) To provide fault tolerance and high availability for data storage
c) To process data using MapReduce
d) None of the above
Answer: b) To provide fault tolerance and high availability for data storage

82. What is the primary advantage of Hadoop over traditional relational database management systems (RDBMS)?
a) Hadoop is more cost-effective
b) Hadoop can handle both structured and unstructured data
c) Hadoop is faster at processing queries
d) None of the above
Answer: b) Hadoop can handle both structured and unstructured data

83. What is one of the main challenges of distributed computing?
a) Ensuring data consistency
b) Managing network bandwidth
c) Coordinating tasks across distributed nodes
d) None of the above
Answer: c) Coordinating tasks across distributed nodes

84. Who is credited with the development of the original MapReduce algorithm?
a) Jeff Bezos
b) Larry Page
c) Sergey Brin
d) Doug Cutting
Answer: d) Doug Cutting

85. What does Hadoop use for processing and analyzing large datasets?
a) SQL queries
b) MapReduce programming model

c) MongoDB database
d) None of the above
Answer: b) MapReduce programming model

86. What is the significance of the term "distributed" in Hadoop's architecture?
a) It refers to the distribution of data across multiple servers.
b) It refers to the distribution of tasks across multiple nodes in a cluster.
c) It refers to the distribution of processing power across multiple cores.
d) None of the above
Answer: b) It refers to the distribution of tasks across multiple nodes in a cluster.

87. Which of the following is NOT a common component of the Hadoop ecosystem?
a) Hive
b) Spark
c) Cassandra
d) HBase
Answer: c) Cassandra

88. What is the primary role of Hadoop's MapReduce framework?
a) To store and manage large datasets
b) To process and analyze large datasets in parallel
c) To provide fault tolerance for distributed data storage
d) None of the above
Answer: b) To process and analyze large datasets in parallel

89. What does HDFS stand for in Hadoop?
a) Hadoop Data File System
b) Hadoop Distributed File System
c) High Data File System
d) None of the above
Answer: b) Hadoop Distributed File System

90. What is the primary advantage of Hadoop distributions?
a) They provide different versions of the Hadoop software.
b) They customize Hadoop for specific use cases.

c) They simplify the deployment and management of Hadoop clusters.
d) None of the above
Answer: c) They simplify the deployment and management of Hadoop clusters

91. What is the primary purpose of Hadoop's distributed computing model?
a) To improve data security
b) To distribute data across multiple servers
c) To enable parallel processing of large datasets
d) None of the above
Answer: c) To enable parallel processing of large datasets

92. Who initially developed Hadoop?
a) Google
b) Facebook
c) Yahoo
d) Microsoft
Answer: c) Yahoo

93. What is the role of Hadoop Distributed File System (HDFS) in Hadoop?
a) To store data in a relational database format
b) To provide fault tolerance and high availability for data storage
c) To process data using MapReduce
d) None of the above
Answer: b) To provide fault tolerance and high availability for data storage

94. What is the primary advantage of Hadoop over traditional relational database management systems (RDBMS)?
a) Hadoop is more cost-effective
b) Hadoop can handle both structured and unstructured data
c) Hadoop is faster at processing queries
d) None of the above
Answer: b) Hadoop can handle both structured and unstructured data

95. What is one of the main challenges of distributed computing?
a) Ensuring data consistency
b) Managing network bandwidth
c) Coordinating tasks across distributed nodes
d) None of the above

Answer: c) Coordinating tasks across distributed nodes

96. Who is credited with the development of the original MapReduce algorithm?
a) Jeff Bezos
b) Larry Page
c) Sergey Brin
d) Doug Cutting
Answer: d) Doug Cutting

97. What does Hadoop use for processing and analyzing large datasets?
a) SQL queries
b) MapReduce programming model
c) MongoDB database
d) None of the above
Answer: b) MapReduce programming model

98. What is the significance of the term "distributed" in Hadoop's architecture?
a) It refers to the distribution of data across multiple servers.
b) It refers to the distribution of tasks across multiple nodes in a cluster.
c) It refers to the distribution of processing power across multiple cores.
d) None of the above
Answer: b) It refers to the distribution of tasks across multiple nodes in a cluster.

99. Which of the following is NOT a common component of the Hadoop ecosystem?
a) Hive
b) Spark
c) Cassandra
d) HBase
Answer: c) Cassandra

100. What is the primary role of Hadoop's MapReduce framework?
a) To store and manage large datasets
b) To process and analyze large datasets in parallel
c) To provide fault tolerance for distributed data storage

d) None of the above
Answer: b) To process and analyze large datasets in parallel
101. What is the primary purpose of processing data with Hadoop?
a) To increase data security
b) To distribute data across multiple servers
c) To enable parallel processing of large datasets
d) None of the above
Answer: c) To enable parallel processing of large datasets

102. What is MapReduce programming?
a) A programming language developed by Google
b) A programming model for processing and generating large datasets
c) A data storage format used in Hadoop
d) None of the above
Answer: b) A programming model for processing and generating large datasets

103. What is the role of a mapper in MapReduce programming?
a) To process input data and generate intermediate key-value pairs
b) To combine intermediate key-value pairs into a final output
c) To partition input data across multiple nodes
d) None of the above
Answer: a) To process input data and generate intermediate key-value pairs

104. What does a reducer do in MapReduce programming?
a) It processes intermediate key-value pairs and generates final output
b) It performs data partitioning
c) It formats input data for processing
d) None of the above
Answer: a) It processes intermediate key-value pairs and generates final output

105. What is the purpose of a combiner in MapReduce programming?
a) To combine multiple datasets into a single output
b) To preprocess input data before mapping
c) To reduce the amount of data shuffled between mapper and reducer
d) None of the above
Answer: c) To reduce the amount of data shuffled between mapper and reducer

106. What is a partitioner in MapReduce programming?

a) A component that divides input data into equal-sized chunks
b) A component that assigns keys to reducers based on a partitioning function
c) A component that formats output data into a structured format
d) None of the above
Answer: b) A component that assigns keys to reducers based on a partitioning function

107. What are the different types of NoSQL databases?
a) Relational, Document, Key-Value, Columnar
b) MySQL, PostgreSQL, Oracle, SQL Server
c) MongoDB, Cassandra, Redis, DynamoDB
d) None of the above
Answer: a) Relational, Document, Key-Value, Columnar

108. What is one advantage of using NoSQL databases?
a) They are highly normalized
b) They provide strong ACID transactions
c) They can handle unstructured and semi-structured data
d) None of the above
Answer: c) They can handle unstructured and semi-structured data

109. How is NoSQL used in the industry?
a) For storing and processing structured data only
b) For handling large volumes of unstructured data
c) For running complex SQL queries
d) None of the above
Answer: b) For handling large volumes of unstructured data

110. What is the primary difference between SQL and NoSQL databases?
a) SQL databases use a schema, while NoSQL databases are schema-less
b) SQL databases are horizontally scalable, while NoSQL databases are vertically scalable
c) SQL databases use a key-value store, while NoSQL databases use a relational model
d) None of the above
Answer: a) SQL databases use a schema, while NoSQL databases are schema-less

111. What is the primary purpose of processing data with Hadoop?
a) To increase data security
b) To distribute data across multiple servers
c) To enable parallel processing of large datasets
d) None of the above
Answer: c) To enable parallel processing of large datasets

112. What is MapReduce programming?
a) A programming language developed by Google
b) A programming model for processing and generating large datasets
c) A data storage format used in Hadoop
d) None of the above
Answer: b) A programming model for processing and generating large datasets

113. What is the role of a mapper in MapReduce programming?
a) To process input data and generate intermediate key-value pairs
b) To combine intermediate key-value pairs into a final output
c) To partition input data across multiple nodes
d) None of the above
Answer: a) To process input data and generate intermediate key-value pairs

114. What does a reducer do in MapReduce programming?
a) It processes intermediate key-value pairs and generates final output
b) It performs data partitioning
c) It formats input data for processing
d) None of the above
Answer: a) It processes intermediate key-value pairs and generates final output

115. What is the purpose of a combiner in MapReduce programming?
a) To combine multiple datasets into a single output
b) To preprocess input data before mapping
c) To reduce the amount of data shuffled between mapper and reducer
d) None of the above
Answer: c) To reduce the amount of data shuffled between mapper and reducer

116. What is a partitioner in MapReduce programming?
a) A component that divides input data into equal-sized chunks
b) A component that assigns keys to reducers based on a partitioning function

c) A component that formats output data into a structured format
d) None of the above
Answer: b) A component that assigns keys to reducers based on a partitioning function

117. What are the different types of NoSQL databases?
a) Relational, Document, Key-Value, Columnar
b) MySQL, PostgreSQL, Oracle, SQL Server
c) MongoDB, Cassandra, Redis, DynamoDB
d) None of the above
Answer: a) Relational, Document, Key-Value, Columnar

118. What is one advantage of using NoSQL databases?
a) They are highly normalized
b) They provide strong ACID transactions
c) They can handle unstructured and semi-structured data
d) None of the above
Answer: c) They can handle unstructured and semi-structured data

119. How is NoSQL used in the industry?
a) For storing and processing structured data only
b) For handling large volumes of unstructured data
c) For running complex SQL queries
d) None of the above
Answer: b) For handling large volumes of unstructured data

120. What is the primary difference between SQL and NoSQL databases?
a) SQL databases use a schema, while NoSQL databases are schema-less
b) SQL databases are horizontally scalable, while NoSQL databases are vertically scalable
c) SQL databases use a key-value store, while NoSQL databases use a relational model
d) None of the above
Answer: a) SQL databases use a schema, while NoSQL databases are schema-less

121. What is the primary purpose of processing data with Hadoop?
a) To increase data security

b) To distribute data across multiple servers
c) To enable parallel processing of large datasets
d) None of the above
Answer: c) To enable parallel processing of large datasets

122. What is MapReduce programming?
a) A programming language developed by Google
b) A programming model for processing and generating large datasets
c) A data storage format used in Hadoop
d) None of the above
Answer: b) A programming model for processing and generating large datasets

123. What is the role of a mapper in MapReduce programming?
a) To process input data and generate intermediate key-value pairs
b) To combine intermediate key-value pairs into a final output
c) To partition input data across multiple nodes
d) None of the above
Answer: a) To process input data and generate intermediate key-value pairs

124. What does a reducer do in MapReduce programming?
a) It processes intermediate key-value pairs and generates final output
b) It performs data partitioning
c) It formats input data for processing
d) None of the above
Answer: a) It processes intermediate key-value pairs and generates final output

125. What is the purpose of a combiner in MapReduce programming?
a) To combine multiple datasets into a single output
b) To preprocess input data before mapping
c) To reduce the amount of data shuffled between mapper and reducer
d) None of the above
Answer: c) To reduce the amount of data shuffled between mapper and reducer