Multiple Choice Question & Answers

Unit – 3

1. What is the primary programming language for writing MapReduce jobs in Hadoop?

   a) C++

   b) Java

   c) Python

   d) PHP

   Answer: b) Java

2. Which Hadoop component is responsible for data serialization?

   a) YARN

   b) HDFS

   c) Avro

   d) MapReduce

   Answer: c) Avro

3. What is Apache Pig primarily used for in the Hadoop ecosystem?

   a) Data warehousing

   b) Data analysis and transformation scripting

   c) Real-time data processing

d) Data visualization

Answer: b) Data analysis and transformation scripting

4. What is a major feature of Apache Hive in Hadoop?

a) Provides machine learning capabilities

b) Enables writing MapReduce jobs easily

c) Allows SQL-like querying of data

d) Handles real-time data streaming

Answer: c) Allows SQL-like querying of data

5. Which tool is used for transferring bulk data between Hadoop and structured datastores?

a) Flume

b) ZooKeeper

c) Sqoop

d) Oozie

Answer: c) Sqoop

6. What does the 'Reduce' step in MapReduce do?

a) Splits data into chunks

b) Aggregates the results from the Map step

c) Distributes processing jobs to nodes

d) Serializes data into HDFS

Answer: b) Aggregates the results from the Map step

7. In Hadoop, what is ZooKeeper used for?

a) Data storage

b) Job scheduling

c) Cluster coordination and configuration management

d) Data analysis

Answer: c) Cluster coordination and configuration management

8. What is a characteristic of a NameNode in Hadoop?

a) It stores actual data

b) It is used for real-time data processing

c) It manages the file system metadata

d) It performs data serialization

Answer: c) It manages the file system metadata

9. Which of the following is true about Hadoop 3.x compared to 2.x?

a) It removes support for YARN

b) It introduces support for GPU hardware

c) It eliminates the HDFS

d) It replaces MapReduce with a new framework

Answer: b) It introduces support for GPU hardware

10. How does Apache Flume primarily interact with Hadoop?

a) It manages Hadoop's file system namespace

b) It provides a service for log and event data ingestion

c) It offers an SQL-like interface for Hadoop

d) It is used for data serialization

Answer: b) It provides a service for log and event data ingestion

11. What is the purpose of Oozie in the Hadoop ecosystem?

a. Data visualization

b. Workflow scheduling and coordination

c. Data storage management

d. Real-time data processing

Answer: b) Workflow scheduling and coordination

12. What does Hadoop's Map step in MapReduce do?

a) Aggregates the results

b) Manages the cluster resources

c) Processes and filters data in parallel

d) Stores data into HDFS

Answer: c) Processes and filters data in parallel

13. What type of file system is HDFS?

a. Relational file system

b. Distributed file system

c. Local file system

d. Network file system

Answer: b) Distributed file system

14. What does Apache Ambari provide in the context of Hadoop?

a) Data analysis scripting

b) Cluster management and monitoring

c) Real-time data processing

d) Workflow scheduling

Answer: b) Cluster management and monitoring

15. What is the benefit of Hadoop's distributed computing model?

a) Faster real-time processing

b) High availability and fault tolerance

c) Automated data cleaning

d) Integrated data visualization

Answer: b) High availability and fault tolerance

16. What is a common use of Apache Spark in the Hadoop ecosystem?

a) Managing HDFS

b) Real-time data processing and analytics

c) Data serialization and deserialization

d) Workflow scheduling

Answer: b) Real-time data processing and analytics

17. In Hadoop, what is the role of a Secondary NameNode?

a) To store the actual data

b) To take over if the Primary NameNode fails

c) To perform periodic checkpoints of HDFS metadata

d) To handle real-time data processing

Answer: c) To perform periodic checkpoints of HDFS metadata

18. What is a major feature of Apache HBase in the Hadoop ecosystem?

a) It provides SQL-like querying

b) It is used for batch data processing

c) It is a distributed, column-oriented database

d) It is a tool for data ingestion

Answer: c) It is a distributed, column-oriented database

19. How does Hadoop achieve high data throughput?

a) By using in-memory computations

b) Through data replication and parallel processing

c) With a centralized processing model

d) By using a single-threaded process

Answer: b) Through data replication and parallel processing

20. In Hadoop, what is a common strategy to optimize MapReduce performance?

a) Increasing the memory of the NameNode

b) Reducing the number of mappers and reducers

c) Fine-tuning the configuration of mappers and reducers

d) Using only in-memory data storage

Answer: c) Fine-tuning the configuration of mappers and reducers

21. What is Apache Kafka used for in the context of Hadoop?

a) Distributed database management

b) Data serialization and deserialization

c) Building data lakes

d) Real-time streaming data pipelines

Answer: d) Real-time streaming data pipelines

22. How does Hadoop handle hardware failures?

a) By relying on external backup systems

b) Through manual intervention by administrators

c) Automatically, through data replication and task reruns

d) Using RAID configurations

Answer: c) Automatically, through data replication and task reruns

23. What is the significance of Apache Storm in the Hadoop ecosystem?

a) It is a tool for batch processing

b) It provides capabilities for real-time data processing

c) It is used for data serialization

d) It offers a graphical user interface for Hadoop

Answer: b) It provides capabilities for real-time data processing

24. What is a Rack in the context of Hadoop?

a) A storage unit in HDFS

b) A collection of DataNodes in a physical location

c) A type of data serialization format

d) A job scheduling tool

Answer: b) A collection of DataNodes in a physical location

25. How is data redundancy achieved in HDFS?

a) By storing multiple copies of data across different nodes

b) Using a centralized backup system

c) Through RAID configurations

d) By storing data in a single master node

Answer: a) By storing multiple copies of data across different nodes

Unit – 4

26. What is the key difference between RDBMS and Hadoop?

a) RDBMS is for large-scale data processing, while Hadoop is not

b) Hadoop is primarily for online transaction processing

c) RDBMS uses structured data, while Hadoop handles both structured and unstructured data

d) Hadoop uses SQL for data querying, RDBMS does not

Answer: c) RDBMS uses structured data, while Hadoop handles both structured and unstructured data

27. What is the primary function of Hadoop?

a) Real-time analytics

b) Large-scale data processing

c) Online transaction processing

d) Data visualization

Answer: b) Large-scale data processing

28. Which company is a major distributor of Hadoop?

a) Oracle

b) Microsoft

c) Cloudera

d) Adobe

Answer: c) Cloudera

29. What is the role of a NameNode in Hadoop?

a. Data processing

b. Resource management

c. Storing data

d. Managing file system metadata

Answer: d) Managing file system metadata

30. What does HDFS stand for?

a) High-Density File System

b) Hadoop Data File System

c) Hadoop Distributed File System

d) High-Definition File System

Answer: c) Hadoop Distributed File System

31. Which daemon is responsible for storing actual data in HDFS?

a) NameNode

b) DataNode

c) Secondary NameNode

d) ResourceManager

Answer: b) DataNode

32. What happens during a file write operation in HDFS?

a) Data is written to the NameNode first

b) Data is directly written to DataNodes

c) Data is stored in a centralized server

d) Data is replicated in the client machine

Answer: b) Data is directly written to DataNodes

33. What is the role of the Secondary NameNode in Hadoop?

a) Backup for the NameNode

b) Storing additional copies of data

c) Processing data requests

d) Performing housekeeping tasks for the NameNode

Answer: d) Performing housekeeping tasks for the NameNode

34. How is data read from HDFS?

a) The client directly interacts with the NameNode

b) The NameNode reads data from the DataNode and passes it to the client

c) The client interacts with the DataNode directly after initial interaction with the NameNode

d) Data is streamed from a centralized storage

Answer: c) The client interacts with the DataNode directly after initial interaction with the NameNode

35. In Hadoop, what is the purpose of the MapReduce framework?

a) Data storage

b) Data serialization

c) Processing large datasets in a distributed manner

d) Resource management

Answer: c) Processing large datasets in a distributed manner

36. What is HBase used for in Hadoop?

a) Real-time querying of large datasets

b) Data visualization

c) Job scheduling

d) Workflow management

Answer: a) Real-time querying of large datasets

37. What type of data model does HBase use?

a) Relational

b) Graph

c) Columnar

d) Document-oriented

Answer: c) Columnar

38. What is Apache Hive mainly used for in Hadoop?

a) Machine learning

b) Data warehousing and SQL-like querying

c) Real-time data processing

d) Graph processing

Answer: b) Data warehousing and SQL-like querying

39. Which Hadoop tool is known for its data flow scripting capabilities?

a) HBase

b) Hive

c) Pig

d) Sqoop

Answer: c) Pig

40. How does Hive execute queries?

a. By converting queries into Java code

b. By using a specialized Hive query execution engine

c. By converting queries into MapReduce jobs

d. Through direct interaction with the DataNode

Answer: c) By converting queries into MapReduce jobs

41. What is the primary file storage system used by Hadoop?

a) NFS

b) HDFS

c) S3

d) EFS

Answer: b) HDFS

42. Which component in Hadoop is responsible for resource allocation and job scheduling?

a. NameNode

b. DataNode

c. YARN

d. HBase

Answer: c) YARN

43. In Hadoop's MapReduce, what does the Mapper do?

a) Combines the output of the Reducer

b) Splits input data into smaller chunks for processing

c) Manages the distribution of data blocks

d) Aggregates the results into a final output

Answer: b) Splits input data into smaller chunks for processing

44. What is a significant advantage of using Hadoop for data processing?

a) Low latency data access

b) Support for complex transactions

c) Scalability and fault tolerance

d) Built-in real-time processing

Answer: c) Scalability and fault tolerance

45. Which of the following is true about Hadoop's scalability?

a) It is limited to a single server

b) It scales horizontally by adding more nodes

c) It scales only vertically by adding more resources to existing nodes

d) Scalability is not a feature of Hadoop

Answer: b) It scales horizontally by adding more nodes

46. What is the main function of Pig in the Hadoop ecosystem?

a. Data visualization

b. High-level data processing and analysis

c. Managing HDFS

d. Real-time data processing

Answer: b) High-level data processing and analysis

47. How are Hadoop clusters typically deployed?

a) In a single data center only

b) Exclusively in the cloud

c) On-premises, in the cloud, or in hybrid environments

d) On personal computing devices

Answer: c) On-premises, in the cloud, or in hybrid environments

48. What is the primary use case for Apache Sqoop in Hadoop?

    a. Data warehousing

    b. Transferring bulk data between Hadoop and relational databases

    c. Real-time data processing

    d. Data serialization

    Answer: b) Transferring bulk data between Hadoop and relational databases

49. What does the Hadoop YARN component do?

    a) Manages the file system

    b) Handles job scheduling and cluster resource management

    c) Processes and filters data

    d) Stores data

    Answer: b) Handles job scheduling and cluster resource management

50. What type of architecture does Hadoop use?

    a) Client-server

    b) Monolithic

    c) Peer-to-peer

    d) Distributed computing

    Answer: d) Distributed computing

51. What kind of processing is MapReduce best suited for?

a) Real-time processing

b) Transactional processing

c) Batch processing

d) Graph processing

Answer: c) Batch processing

52. In Hadoop, what is a common way to achieve high availability for the NameNode?

a) Using RAID configurations

b) Running multiple NameNodes on a single machine

c) Having a secondary NameNode as a hot standby

d) Frequent backups to an external database

Answer: c) Having a secondary NameNode as a hot standby

53. How does HDFS ensure data integrity?

a) Through regular data backups

b) Using checksums for data blocks

c) By storing all data in a single location

d) Through manual verification

Answer: b) Using checksums for data blocks

54. What is a key feature of Hadoop's DataNode?

a) It only stores metadata

b) It performs real-time data analysis

c) It stores and retrieves data blocks

d) It schedules MapReduce jobs

Answer: c) It stores and retrieves data blocks

55. How does Pig interact with Hadoop?

a) It replaces HDFS for storage

b) It compiles scripts into MapReduce jobs for data processing

c) It is used for real-time data processing

d) It provides a graphical interface for Hadoop configuration

Answer: b) It compiles scripts into MapReduce jobs for data processing

56. What is a characteristic of Hadoop's write-once-read-many model?

a) Data can be updated in place

b) Data is immutable once written

c) HDFS supports transactional updates

d) Data is stored temporarily

Answer: b) Data is immutable once written

57. What is the role of the JobTracker in Hadoop 1.x?

    a. It stores data

    b. It manages the file system namespace

    c. It schedules and manages MapReduce jobs

    d. It processes data

    Answer: c) It schedules and manages MapReduce jobs

58. How does Hadoop provide fault tolerance?

    a) By using a centralized monitoring system

    b) Through data replication across different nodes

    c) By automatically redirecting requests to a backup server

    d) Using a RAID storage configuration

    Answer: b) Through data replication across different nodes

59. What is the main benefit of using HBase in Hadoop?

    a) Real-time access to large datasets

    b) SQL-like querying capabilities

    c) Data visualization

    d) Scripting data flow processes

    Answer: a) Real-time access to large datasets

60. Which file format is commonly used in Hadoop for optimizing large data processing?

a) CSV

b) JSON

c) Parquet

d) XML

Answer: c) Parquet

61. How does Hadoop handle large-scale data processing?

a) Through a single centralized database

b) By processing all data in memory

c) Through distributed computing and storage

d) Using only SQL-based queries

Answer: c) Through distributed computing and storage

62. What distinguishes Hadoop from traditional RDBMS in terms of storage?

a) Hadoop is suitable only for small datasets

b) Hadoop stores data in a centralized manner

c) Hadoop is designed for horizontal scaling with large volumes of data

d) Hadoop does not support unstructured data

Answer: c) Hadoop is designed for horizontal scaling with large volumes of data

63. What is the significance of reducers in Hadoop's MapReduce?

a) They split data into smaller chunks

b) They combine the output from mappers to produce the final result

c) They are responsible for data storage

d) They perform real-time data processing

Answer: b) They combine the output from mappers to produce the final result

64. What is a typical use case for Apache Pig in Hadoop?

a) Real-time data streaming

b) Complex data transformations and analysis

c) Data storage and retrieval

d) Cluster management and monitoring

Answer: b) Complex data transformations and analysis

65. How does Hadoop's HDFS deal with node failures?

a) By quickly switching to a standby node

b) Through automatic data replication on other nodes

c) By pausing operations until the node is restored

d) Using manual data recovery methods

Answer: b) Through automatic data replication on other nodes

66. In Hadoop, what is a primary advantage of using YARN?

a) It automates data serialization

b) It enhances the ability to manage and utilize cluster resources efficiently

c) It simplifies data storage

d) It is used for interactive SQL querying

Answer: b) It enhances the ability to manage and utilize cluster resources efficiently

67. What does the Hadoop ecosystem primarily offer?

a) A single tool for all data needs

b) A collection of tools for different big data processing requirements

c) Real-time processing capabilities

d) A graphical user interface for big data analytics

Answer: b) A collection of tools for different big data processing requirements

68. How does Hadoop's HDFS achieve high throughput?

a) By using high-speed SSDs

b) Through parallel processing of data across multiple nodes

c) By limiting the number of read/write operations

d) Using a single high-capacity node

Answer: b) Through parallel processing of data across multiple nodes

69. What is the role of Apache ZooKeeper in a Hadoop cluster?

a) It is used for SQL-like querying

b) It manages HDFS

c) It is a centralized service for maintaining configuration information and providing distributed synchronization

d) It processes data

Answer: c) It is a centralized service for maintaining configuration information and providing distributed synchronization

70. What is a typical characteristic of a Hadoop cluster?

a) Limited to small-scale deployments

b) Composed of a few high-capacity nodes

c) Consists of many commodity hardware nodes

d) Requires specialized hardware and storage

Answer: c) Consists of many commodity hardware nodes

71. In Hadoop, what is a major benefit of using the MapReduce programming model?

a) It is optimized for transaction processing

b) It automatically handles parallel computation and fault tolerance

c) It is primarily used for real-time analytics

d) It simplifies the management of the HDFS

Answer: b) It automatically handles parallel computation and fault tolerance

72. What does Hadoop's ResourceManager primarily manage?

a) Data storage in HDFS

b) Hadoop configuration settings

c) Memory and CPU resources of the cluster

d) Data serialization and deserialization

Answer: c) Memory and CPU resources of the cluster

73. What is the advantage of using Apache Hive for data warehousing in Hadoop?

a) It supports real-time data processing

b) It provides an SQL-like interface for querying and managing large datasets

c) It is used for graph processing

d) It is a real-time monitoring tool

Answer: b) It provides an SQL-like interface for querying and managing large datasets

74. In the Hadoop ecosystem, what is the purpose of Apache Flume?

a) It is used for batch processing

b) It manages the file system namespace

c) It is a service for efficiently collecting, aggregating, and moving large amounts of log data

d) It is a data visualization tool

Answer: c) It is a service for efficiently collecting, aggregating, and moving large amounts of log data

75. What is the typical block size in HDFS?

a) 64 MB

b) 128 MB

c) 1 GB

d) 256 MB

Answer: b) 128 MB

Unit – 5

76. What does the R function read.csv() do?

a) Writes data to a CSV file

b) Reads data from a CSV file

c) Plots data from a CSV file

d) Converts data to CSV format

Answer: B) Reads data from a CSV file

77. Which package in R is used for data manipulation?

a) ggplot2

b) dplyr

c) shiny

d) tidyr

Answer: B) dplyr

78. Machine Learning: Introduction

What is machine learning?

a) A technique for data storage

b) A method for manually coding algorithms

c) A process of teaching computers to learn from data

d) A hardware improvement in computing

Answer: C) A process of teaching computers to learn from data

79. Which of these is a type of machine learning algorithm?

a) Regression

b) Multiplication

c) Subtraction

d) Division

Answer: A) Regression

80. What characterizes supervised learning?

a) The use of labeled data

b) Learning without any data

c) The use of unlabeled data

d) The absence of algorithms

Answer: A) The use of labeled data

81. Which is an example of a supervised learning task?

a) Clustering

b) Dimensionality Reduction

c) Classification

d) Association

Answer: C) Classification

82. 82.What is unsupervised learning?

a) Learning with labeled data

b) Learning with partially labeled data

c) Learning without labeled data

d) Learning with direct supervision

Answer: C) Learning without labeled data

83. Which algorithm is used for unsupervised learning?

a) Linear Regression

b) Decision Trees

c) K-Means Clustering

d) Logistic Regression

Answer: C) K-Means Clustering

84. Collaborative filtering is mainly used in:

a) Image Recognition

b) Recommendation Systems

c) Fraud Detection

d) Weather Forecasting

Answer: B) Recommendation Systems

85. What does collaborative filtering rely on?

a) Similarity in user behavior

b) Past user experiences

c) Product details

d) A and B

Answer: D) A and B

86. Social media analytics primarily helps in:

a) Improving network speed

b) Understanding user behavior

c) Increasing storage capacity

d) Enhancing hardware performance

Answer: B) Understanding user behavior

87. A key tool for social media analytics is:

a) Regression Analysis

b) Text Mining

c) Calculus

d) Database Management

Answer: B) Text Mining

88. Mobile Analytics is used to:

a) Repair mobile devices

b) Analyze mobile app usage

c) Increase mobile battery life

d) Design mobile phones

Answer: B) Analyze mobile app usage

89. Which metric is important in mobile analytics?

a) User engagement

b) Screen size

c) Battery capacity

d) Processor speed

Answer: A) User engagement

90. BigR is a tool for:

a) Small-scale data analysis

b) Large-scale data analysis in R

c) Gaming

d) Social networking

Answer: B) Large-scale data analysis in R

91. In the context of Big Data, 'Velocity' refers to:

a) Speed of data generation

b) Variety of data

c) Volume of data

d) Value of data

Answer: A) Speed of data generation

92. R is primarily used for:

a) Web development

b) Statistical computing and graphics

c) 3D modeling

d) Video editing

Answer: B) Statistical computing and graphics

93. Which of the following is a supervised learning algorithm?

a) K-means clustering

b) Apriori algorithm

c) Neural Networks

d) Principal Component Analysis

Answer: C) Neural Networks

94. What is 'Big Data' characterized by?

a) The Three Vs: Volume, Velocity, Variety

b) Its small size and complexity

c) The speed of internet connectivity

d) Its structured format only

Answer: A) The Three Vs: Volume, Velocity, Variety

95. In R, which function is used to create linear models?

a) lm()

b) plot()

c) sum()

d) read.table()

Answer: A) lm()

96. What is the main goal of data analytics?

a) Data storage

b) Data cleaning

c) Extracting useful insights from data

d) Creating large databases

Answer: C) Extracting useful insights from data

97. What does 'Machine Learning' primarily focus on?

a) Developing self-aware machines

b) Allowing machines to learn from data and make decisions

c) Mechanical engineering

d) Building physical robots

Answer: B) Allowing machines to learn from data and make decisions

98. In the context of machine learning, 'overfitting' means:

a) The model is too simple

b) The model performs well on the training data but poorly on unseen data

c) The data is too complex

d) The model is perfectly fit

Answer: B) The model performs well on the training data but poorly on unseen data

99. A 'Random Forest' is a:

a) Type of decision tree

b) Forest conservation method

c) Machine learning algorithm

d) Data visualization technique

Answer: C) Machine learning algorithm

100. The main purpose of unsupervised learning is to:

a) Predict the outcome for new data

b) Classify data based on labels

c) Find hidden patterns in data

d) Follow explicit instructions from a programmer

Answer: C) Find hidden patterns in data

101. Collaborative filtering is commonly used in:

a) Robotics

b) Weather forecasting

c) E-commerce recommendations

d) Data compression

Answer: C) E-commerce recommendations

102. In social media analytics, sentiment analysis is used to:

a) Measure the speed of internet connections

b) Determine the emotional tone behind a body of text

c) Calculate the total number of likes

d) Analyze the structure of social networks

Answer: B) Determine the emotional tone behind a body of text

103. The main focus of mobile analytics is on:

a) Hardware performance of mobile devices

b) Analyzing data from mobile app usage and user engagement

c) The number of mobile devices sold

d) Mobile network coverage areas

Answer: B) Analyzing data from mobile app usage and user engagement

104. BigR is particularly useful for:

a) Handling small datasets in R

b) Handling large-scale data processing in R

c) Gaming purposes

d) Social media analysis

Answer: B) Handling large-scale data processing in R

105. In Big Data, 'Variety' refers to:

a) The different types of data (structured, unstructured, semi-structured)

b) The volume of data

c) The speed of data processing

d) The accuracy of data

Answer: A) The different types of data (structured, unstructured, semi-structured)

106. In R, which package is commonly used for creating plots and graphs?

a) stringr

b) ggplot2

c) jsonlite

d) Rcpp

Answer: B) ggplot2

107. What is a 'Neural Network' in the context of machine learning?

a) A computer network

b) A type of deep learning algorithm

c) A database management system

d) A cybersecurity tool

Answer: B) A type of deep learning algorithm

108. The term 'Volume' in Big Data refers to:

a) The amount of data

b) The speed of data processing

c) The variety of data types

d) The sound level in data centers

Answer: A) The amount of data

109. Which of the following is an unsupervised learning algorithm?

a) Decision Trees

b) Support Vector Machines

c) K-means Clustering

d) Linear Regression

Answer: C) K-means Clustering

110. What does the R function ggplot() do?

a) Cleans data

b) Performs statistical tests

c) Creates data visualizations

d) Writes data to a database

Answer: C) Creates data visualizations

111. In machine learning, 'training data' is used for:

a) Repairing machines

b) Teaching a model to recognize patterns

c) Storing data permanently

d) Increasing data processing speed

Answer: B) Teaching a model to recognize patterns

112. What is a key characteristic of 'supervised learning'?

a) Using unlabeled data

b) No need for data

c) Using a predefined set of labels for data

d) Learning without any supervision

Answer: C) Using a predefined set of labels for data

113. An example of unsupervised learning is:

a) Classification

b) Regression

c) Clustering

d) Reinforcement Learning

Answer: C) Clustering

114. Collaborative filtering algorithms focus on:

a) Individual user preferences

b) Predicting weather patterns

c) Calculating financial risks

d) Analyzing hardware performance

Answer: A) Individual user preferences

115. Social media analytics can help businesses:

a) Improve their product design

b) Understand customer sentiments

c) Increase the speed of their internet connection

d) Reduce their tax liabilities

Answer: B) Understand customer sentiments

116. A key aspect of mobile analytics is:

a) Monitoring app downloads

b) Designing mobile interfaces

c) Developing new mobile devices

d) Researching new battery technologies

Answer: A) Monitoring app downloads

117. BigR is designed for working with:

a) Small datasets in Python

b) Large datasets in R

c) Gaming algorithms

d) Social media platforms

Answer: B) Large datasets in R

118. What is a common use case for using R in data analytics?

a) Building computer networks

b) Statistical analysis and data visualization

c) Hardware design

d) Software gaming development

Answer: B) Statistical analysis and data visualization

119. Which of these is a challenge in machine learning?

a) Decreasing the size of data

b) Overfitting and underfitting models

c) Making computers faster

d) Reducing the cost of internet

Answer: B) Overfitting and underfitting models

120. In unsupervised learning, the most common task is:

a) Prediction

b) Classification

c) Clustering

d) Regression

Answer: C) Clustering

121. Collaborative filtering is most effective when:

a) User data is limited

b) There is a large amount of user data

c) The system is offline

d) User preferences are static

Answer: B) There is a large amount of user data

122. An important aspect of social media analytics is:

a) Network cabling

b) Analyzing trends and patterns in user data

c) Selling online advertisements

d) Increasing server capacity

Answer: B) Analyzing trends and patterns in user data

123. Mobile analytics is crucial for:

a) Building mobile towers

b) Understanding user engagement with mobile apps

c) Repairing mobile devices

d) Selling mobile phones

Answer: B) Understanding user engagement with mobile apps

124. BigR helps in handling:

a) Simple, small-scale calculations

b) Complex, large-scale data analytics in R

c) Mobile game development

d) Social media account management

Answer: B) Complex, large-scale data analytics in R

125. In R, the dplyr package is used for:

a) Data manipulation

b) Creating animations

c) Developing web applications

d) Performing linear regression

Answer: A) Data manipulation