1. What is the primary purpose of integrating disparate data stores in big data processing?

a) To increase the complexity of data analysis
b) To simplify data retrieval processes
c) To ensure data consistency and accuracy
d) To minimize the need for data transformation

Answer: c) To ensure data consistency and accuracy

Explanation: Integrating disparate data stores helps in bringing together various sources of data into a unified format, ensuring consistency and accuracy in the data being processed.

2. In big data processing, what does mapping data to the programming framework involve?

a) Assigning geographical coordinates to data points
b) Linking data sets to a specific programming language
c) Defining how data will be processed within the chosen programming model
d) Sorting data into alphabetical order

Answer: c) Defining how data will be processed within the chosen programming model

Explanation: Mapping data to the programming framework involves defining how data will be processed within the selected programming model, ensuring efficient utilization of computational resources.

3. Which stage of big data processing involves connecting and extracting data from storage?

a) Data transformation

b) Data integration
c) Data retrieval
d) Data mapping

Answer: c) Data retrieval

Explanation: Data retrieval involves connecting to data storage systems and extracting the required data for processing.

4. What is the purpose of transforming data in big data processing?

a) To increase data complexity
b) To reduce storage requirements
c) To prepare data for analysis
d) To remove redundant data

Answer: c) To prepare data for analysis

Explanation: Transforming data involves converting raw data into a format suitable for analysis, including cleaning, filtering, and structuring it appropriately.

5. Which technique is commonly used for subdividing data in preparation for Hadoop MapReduce?

a) Data sharding
b) Data serialization
c) Data replication
d) Data compression

Answer: a) Data sharding

Explanation: Data sharding involves dividing large datasets into smaller, manageable chunks for parallel processing in frameworks like Hadoop MapReduce.

6. What does the acronym “HDFS” stand for in the context of big data processing?

a) Hadoop Distributed File System
b) High Data Flow System
c) Hierarchical Data Formatting System
d) Heterogeneous Data Fusion System

Answer: a) Hadoop Distributed File System

Explanation: HDFS (Hadoop Distributed File System) is a distributed file system designed to store large volumes of data across multiple machines in a Hadoop cluster.

7. Which of the following is a key advantage of using Hadoop MapReduce for big data processing?

a) Real-time data processing
b) High throughput for small datasets
c) Fault tolerance and scalability
d) Simple programming model

Answer: c) Fault tolerance and scalability

Explanation: Hadoop MapReduce provides fault tolerance and scalability by distributing

computation across a cluster of commodity hardware, enabling efficient processing of large datasets.

8. What role does Apache Spark play in big data processing?

a) Data storage
b) Data visualization
c) Data processing and analytics
d) Data indexing

Answer: c) Data processing and analytics

Explanation: Apache Spark is a fast and general-purpose distributed computing system used for big data processing, including data processing and analytics tasks.

9. Which component of big data processing frameworks is responsible for job scheduling and resource management?

a) YARN (Yet Another Resource Negotiator)
b) HBase
c) Kafka
d) Spark SQL

Answer: a) YARN (Yet Another Resource Negotiator)

Explanation: YARN (Yet Another Resource Negotiator) is a key component of Hadoop that manages resources and schedules jobs in a distributed environment, facilitating efficient resource utilization.

10. What is the purpose of data serialization in big data processing?

a) To optimize data storage
b) To convert data into a readable format
c) To transfer data over a network
d) To enable efficient data processing

Answer: c) To transfer data over a network

Explanation: Data serialization involves converting complex data structures into a format suitable for transmission over a network, facilitating efficient data transfer between different components of a big data processing system.

Related posts:

1. Introduction to Information Security
2. Introduction to Information Security MCQ
3. Introduction to Information Security MCQ
4. Symmetric Key Cryptography MCQ
5. Asymmetric Key Cryptography MCQ
6. Authentication & Integrity MCQ
7. E-mail, IP and Web Security MCQ