1. What is Hadoop primarily designed for?

a) Real-time data processing
b) Handling structured data
c) Storing and processing large volumes of data
d) Running relational database queries

Answer: c) Storing and processing large volumes of data

Explanation: Hadoop is designed to handle large volumes of data that exceed the capabilities of traditional databases. It enables distributed storage and processing of big data across clusters of computers.

2. Which of the following is NOT a core component of Hadoop?

a) HDFS
b) YARN
c) MapReduce
d) Spark

Answer: d) Spark

Explanation: While Spark is often used alongside Hadoop for processing big data, it is not a core component of the Hadoop ecosystem. Core components include HDFS, YARN, and MapReduce.

3. What does HDFS stand for in the context of Hadoop?

a) Hadoop Distributed File Storage
b) High-Performance Data Format System
c) Hadoop Distributed File System
d) Hyper-Dense File Sharing

Answer: c) Hadoop Distributed File System

Explanation: HDFS is the primary storage system used by Hadoop for distributed storage of data across multiple nodes in a cluster.

4. Which component of the Hadoop ecosystem is responsible for resource management and job scheduling?

a) HDFS
b) MapReduce
c) YARN
d) Hive

Answer: c) YARN

Explanation: YARN (Yet Another Resource Negotiator) is the resource management layer in Hadoop that is responsible for managing and scheduling resources across applications in a Hadoop cluster.

5. What is the physical architecture of Hive?

a) Master-slave architecture
b) Peer-to-peer architecture

c) Client-server architecture
d) Distributed architecture

Answer: a) Master-slave architecture

Explanation: Hive typically follows a master-slave architecture where there is a single master node responsible for coordinating operations and multiple worker nodes that execute tasks as directed by the master.

6. What are some limitations of Hadoop?

a) Limited scalability
b) Poor fault tolerance
c) Inefficient for real-time processing
d) Lack of support for unstructured data

Answer: c) Inefficient for real-time processing

Explanation: Hadoop is not well-suited for real-time processing due to its batch-oriented nature and high latency in data processing.

7. How does Hadoop differ from traditional RDBMS?

a) RDBMS offers better scalability
b) Hadoop lacks support for structured data
c) Hadoop is designed for distributed storage and processing
d) RDBMS is more cost-effective

Answer: c) Hadoop is designed for distributed storage and processing

Explanation: Hadoop is designed to handle large volumes of unstructured or semi-structured data across distributed clusters, whereas traditional RDBMS systems are typically centralized and optimized for structured data.

8. Which programming model is used by Hadoop for processing large datasets in parallel across a distributed cluster?

a) MPI
b) SQL
c) MapReduce
d) Spark

Answer: c) MapReduce

Explanation: MapReduce is the programming model used by Hadoop for processing large datasets in parallel across distributed clusters, dividing the workload into map and reduce tasks.

9. What does Hadoop YARN stand for?

a) Yet Another Resource Negotiator
b) Yet Another Resource Node
c) Yet Another Resource Network
d) Yet Another Resource Name

Answer: a) Yet Another Resource Negotiator

Explanation: YARN stands for Yet Another Resource Negotiator, which is the resource management layer in Hadoop responsible for managing resources and scheduling jobs.

10. How does Hadoop handle data processing tasks?

a) Sequentially
b) In parallel
c) Randomly
d) Hierarchically

Answer: b) In parallel

Explanation: Hadoop processes data in parallel across distributed clusters, allowing for efficient and scalable data processing tasks. This parallel processing capability is essential for handling large volumes of data efficiently.

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