1. What is Hadoop MapReduce primarily used for?

a) Real-time data processing
b) Batch processing of large datasets
c) Stream processing
d) In-memory data processing

Answer: b) Batch processing of large datasets

Explanation: Hadoop MapReduce is designed for processing large datasets in batch mode, making it suitable for tasks like data warehousing, ETL (Extract, Transform, Load), and analytics where data can be processed in parallel across a distributed system.

2. Which of the following components are essential for creating Hadoop MapReduce jobs?

a) NameNode and DataNode
b) JobTracker and TaskTracker
c) ResourceManager and NodeManager
d) Mapper and Reducer

Answer: d) Mapper and Reducer

Explanation: The Mapper and Reducer are the key components of a Hadoop MapReduce job. The Mapper processes input data and emits intermediate key-value pairs, while the Reducer aggregates and processes these intermediate results to produce the final output.

3. What is the purpose of distributing data processing across server farms in Hadoop MapReduce?

a) To minimize network latency
b) To reduce the load on individual servers
c) To achieve fault tolerance and scalability
d) To increase power consumption

Answer: c) To achieve fault tolerance and scalability

Explanation: Distributing data processing across server farms in Hadoop MapReduce enables fault tolerance by replicating data and computations across multiple nodes. It also allows for scalability, as additional nodes can be added to handle increased workloads.

4. How can you monitor the progress of job flows in Hadoop MapReduce?

a) Using Hadoop Distributed File System (HDFS)
b) Through the ResourceManager web interface
c) By accessing the NameNode logs
d) Using the Hadoop shell commands

Answer: b) Through the ResourceManager web interface

Explanation: The ResourceManager web interface provides real-time information about the status and progress of Hadoop MapReduce jobs, including details such as job duration, resource usage, and task progress.

5. Which of the following are examples of Hadoop daemons?

a) NameNode and DataNode
b) Mapper and Reducer

c) ResourceManager and NodeManager
d) JobTracker and TaskTracker

Answer: a) NameNode and DataNode

Explanation: NameNode and DataNode are examples of Hadoop daemons responsible for managing the distributed file system (HDFS) and storing data blocks, respectively.

6. In Hadoop MapReduce, which execution mode is suitable for development and testing on a single machine?

a) Fully distributed mode
b) Pseudo-distributed mode
c) Local mode
d) Remote mode

Answer: c) Local mode

Explanation: Local mode in Hadoop MapReduce allows developers to test their code on a single machine without the need for a cluster setup. It simulates the Hadoop environment locally.

7. What is the primary function of the NameNode in Hadoop Distributed File System (HDFS)?

a) Storing data blocks
b) Managing metadata and namespace
c) Executing MapReduce tasks
d) Allocating system resources

Answer: b) Managing metadata and namespace

Explanation: The NameNode in HDFS is responsible for storing metadata and managing the namespace, including information about file locations, permissions, and replication.

8. Which execution mode in Hadoop MapReduce mimics a real distributed cluster but runs on a single machine?

a) Local mode
b) Pseudo-distributed mode
c) Fully distributed mode
d) Hybrid mode

Answer: b) Pseudo-distributed mode

Explanation: Pseudo-distributed mode in Hadoop MapReduce simulates a distributed cluster environment on a single machine, allowing developers to test their code in a setup similar to a real cluster.

9. What is the primary role of the ResourceManager in Hadoop YARN (Yet Another Resource Negotiator)?

a) Managing storage resources
b) Managing compute resources
c) Distributing data across nodes
d) Monitoring job progress

Answer: b) Managing compute resources

Explanation: The ResourceManager in Hadoop YARN is responsible for managing compute resources in a Hadoop cluster, including allocating resources to applications and tracking resource utilization.

10. Which Hadoop component is responsible for launching and monitoring MapReduce tasks on individual nodes in the cluster?

a) ResourceManager
b) NodeManager
c) JobTracker
d) TaskTracker

Answer: b) NodeManager

Explanation: NodeManager is responsible for launching and monitoring MapReduce tasks on individual nodes in a Hadoop cluster, managing resources and reporting status back to the ResourceManager.

11. What is the primary advantage of using Hadoop MapReduce for data processing?

a) Low latency
b) Real-time processing
c) Fault tolerance
d) Limited scalability

Answer: c) Fault tolerance

Explanation: Hadoop MapReduce provides fault tolerance by replicating data and

computations across multiple nodes in a distributed cluster, ensuring that processing can continue even if individual nodes fail.

12. Which execution mode in Hadoop MapReduce is suitable for production deployments across a cluster of machines?

a) Local mode
b) Pseudo-distributed mode
c) Fully distributed mode
d) Hybrid mode

Answer: c) Fully distributed mode

Explanation: Fully distributed mode in Hadoop MapReduce is suitable for production deployments across a cluster of machines, where data processing tasks are distributed across multiple nodes for parallel execution.

13. What is the role of the DataNode in Hadoop Distributed File System (HDFS)?

a) Managing metadata
b) Storing data blocks
c) Managing compute resources
d) Monitoring job progress

Answer: b) Storing data blocks

Explanation: DataNode in HDFS is responsible for storing actual data blocks on the disk and performing read and write operations as requested by clients or other Hadoop components.

14. Which Hadoop daemon is responsible for maintaining a global picture of the cluster and managing resources?

a) ResourceManager
b) NodeManager
c) NameNode
d) DataNode

Answer: a) ResourceManager

Explanation: ResourceManager in Hadoop YARN is responsible for maintaining a global view of the cluster and managing its resources, including allocating resources to applications and monitoring resource utilization.

15. What does the Reducer component do in a Hadoop MapReduce job?

a) Processes input data and emits intermediate key-value pairs
b) Aggregates and processes intermediate results to produce final output
c) Manages compute resources in a Hadoop cluster
d) Allocates storage resources in Hadoop Distributed File System

Answer: b) Aggregates and processes intermediate results to produce final output

Explanation: The Reducer component in Hadoop MapReduce aggregates and processes intermediate results generated by the Mapper, producing the final output of the job.

16. Which mode of execution in Hadoop MapReduce is recommended for production environments where data processing needs to be distributed across a large cluster?

a) Local mode
b) Pseudo-distributed mode
c) Fully distributed mode
d) Hybrid mode

Answer: c) Fully distributed mode

Explanation: Fully distributed mode in Hadoop MapReduce is recommended for production environments where data processing tasks need to be distributed across a large cluster of machines for parallel execution.

17. What is the primary function of the Job Tracker in Hadoop MapReduce?

a) Managing storage resources
b) Managing compute resources
c) Monitoring job progress
d) Distributing data across nodes

Answer: c) Monitoring job progress

Explanation: The JobTracker in Hadoop MapReduce is responsible for monitoring the progress of MapReduce jobs, scheduling tasks, and handling task failures in a Hadoop cluster.

18. In Hadoop MapReduce, what is the purpose of the Mapper component?

a) Aggregating intermediate results
b) Processing input data and emitting key-value pairs
c) Distributing data across nodes

d) Monitoring job progress

Answer: b) Processing input data and emitting key-value pairs

Explanation: The Mapper component in Hadoop MapReduce processes input data and emits intermediate key-value pairs based on the logic defined by the developer.

19. Which execution mode in Hadoop MapReduce simulates a real distributed cluster but runs on a limited number of machines?

a) Local mode
b) Pseudo-distributed mode
c) Fully distributed mode
d) Hybrid mode

Answer: b) Pseudo-distributed mode

Explanation: Pseudo-distributed mode in Hadoop MapReduce simulates a real distributed cluster environment but runs on a limited number of machines, making it suitable for testing and development purposes.

20. What is the primary benefit of distributing data processing across server farms in Hadoop MapReduce?

a) Minimizing network latency
b) Reducing the load on individual servers
c) Achieving fault tolerance and scalability
d) Increasing power consumption

Answer: c) Achieving fault tolerance and scalability

Explanation: Distributing data processing across server farms in Hadoop MapReduce enables fault tolerance and scalability by replicating data and computations across multiple nodes, allowing for parallel execution of tasks and handling of failures.

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