1. Which of the following is not a characteristic of relational databases?
a) Schema flexibility
b) ACID compliance
c) NoSQL structure
d) Support for complex queries

Answer: c) NoSQL structure
Explanation: Relational databases follow a structured query language (SQL) and adhere to the principles of ACID (Atomicity, Consistency, Isolation, Durability) transactions. NoSQL databases, on the other hand, do not adhere to the relational model and provide more flexibility in schema design.

2. What are GFS and HDFS examples of?
a) Relational databases
b) Cloud file systems
c) Key-Value stores
d) NoSQL databases

Answer: b) Cloud file systems
Explanation: Google File System (GFS) and Hadoop Distributed File System (HDFS) are both examples of cloud file systems designed to store and manage large amounts of data across clusters of computers.

3. Which of the following is a feature of GFS and HDFS?
a) Strong consistency
b) Low fault tolerance
c) POSIX compliance

d) Horizontal scalability

Answer: d) Horizontal scalability
Explanation: Both GFS and HDFS are designed to scale horizontally, meaning they can efficiently handle increasing amounts of data by adding more servers to the cluster.

4. BigTable and HBase are examples of:
a) Relational databases
b) NoSQL databases
c) Cloud file systems
d) Key-Value stores

Answer: b) NoSQL databases
Explanation: BigTable and HBase are both NoSQL databases designed to handle large volumes of structured data across distributed clusters.

5. Which system provides fault tolerance and scalability for large-scale distributed systems?
a) MapReduce
b) BigTable
c) HDFS
d) Dynamo

Answer: d) Dynamo
Explanation: Dynamo is a highly available and scalable distributed data store developed by Amazon. It provides fault tolerance and scalability for large-scale distributed systems.

6. What does MapReduce primarily focus on?
a) Real-time data processing

b) Batch processing
c) Interactive querying
d) Online transaction processing

Answer: b) Batch processing
Explanation: MapReduce is primarily designed for processing large volumes of data in batch mode, rather than real-time or interactive processing.

7. Which of the following is a characteristic of the MapReduce model?
a) Low scalability
b) High latency
c) Distributed computing
d) Single-threaded processing

Answer: c) Distributed computing
Explanation: MapReduce leverages distributed computing to process large datasets across multiple nodes in a cluster, enabling parallel processing and scalability.

8. What is the primary advantage of MapReduce in handling large-scale data processing?
a) Low fault tolerance
b) High complexity
c) Parallel efficiency
d) Sequential processing

Answer: c) Parallel efficiency
Explanation: MapReduce allows for parallel processing of data across multiple nodes in a cluster, leading to improved efficiency in handling large-scale data processing tasks.

9. Which operation is commonly associated with relational databases but can also be implemented in MapReduce?
a) Map
b) Shuffle
c) Reduce
d) Join

Answer: d) Join
Explanation: Join operations, which involve combining data from multiple tables based on a related column, are commonly associated with relational databases but can also be implemented in MapReduce frameworks.

10. What is an example of an enterprise batch processing task suitable for MapReduce?
a) Real-time stock trading
b) Online gaming
c) Log analysis
d) Web page rendering

Answer: c) Log analysis
Explanation: Log analysis, which involves processing and analyzing large volumes of log data, is a typical enterprise batch processing task suitable for MapReduce.

11. Which of the following is a characteristic of GFS and HDFS but not of traditional file systems?
a) Single point of failure
b) High throughput
c) Limited scalability

d) Strong consistency

Answer: b) High throughput
Explanation: GFS and HDFS are designed for high throughput and are optimized for handling large files and streaming data, which is not a characteristic of traditional file systems.

12. Which of the following systems provides automatic data partitioning and replication for fault tolerance?
a) GFS
b) HBase
c) Dynamo
d) BigTable

Answer: c) Dynamo
Explanation: Dynamo provides automatic data partitioning and replication across multiple nodes in a distributed system to ensure fault tolerance and high availability.

13. Which characteristic makes HDFS suitable for storing large files?
a) Low fault tolerance
b) Strong consistency
c) Data replication
d) Single-node architecture

Answer: c) Data replication
Explanation: HDFS replicates data across multiple nodes in the cluster to ensure fault tolerance and high availability, making it suitable for storing large files.

14. What is the primary difference between GFS/HDFS and traditional file systems in terms of

scalability?
a) Traditional file systems scale vertically
b) Traditional file systems use distributed computing
c) GFS/HDFS scale horizontally
d) GFS/HDFS have limited scalability

Answer: c) GFS/HDFS scale horizontally
Explanation: Unlike traditional file systems that typically scale vertically by adding more resources to a single server, GFS and HDFS scale horizontally by adding more nodes to the distributed system.

15. Which of the following databases is known for its column-oriented storage?
a) MongoDB
b) BigTable
c) Cassandra
d) Redis

Answer: b) BigTable
Explanation: BigTable is known for its column-oriented storage, where data is stored in columns rather than rows, allowing for efficient querying and retrieval of specific columns.

16. Which system provides automatic sharding and replication of data for high availability?
a) Dynamo
b) HBase
c) MapReduce
d) BigTable

Answer: b) HBase
Explanation: HBase provides automatic sharding and replication of data across multiple nodes to ensure high availability and fault tolerance in distributed environments.

17. What is the primary purpose of MapReduce?
a) Real-time data processing
b) Interactive querying
c) Batch processing
d) Online transaction processing

Answer: c) Batch processing
Explanation: MapReduce is primarily used for batch processing of large datasets, where data is processed in parallel across distributed nodes in a cluster.

18. In the MapReduce paradigm, what does the “map” phase primarily involve?
a) Data aggregation
b) Data partitioning
c) Data sorting
d) Data transformation

Answer: d) Data transformation
Explanation: In the MapReduce paradigm, the “map” phase involves transforming input data into intermediate key-value pairs, which are then passed to the “reduce” phase for further processing.

19. Which of the following is a key benefit of using MapReduce for parallel computing?
a) Sequential processing

b) Single-threaded execution
c) Scalability
d) High latency

Answer: c) Scalability
Explanation: MapReduce enables parallel computing by distributing data processing tasks across multiple nodes in a cluster, allowing for scalability to handle large datasets efficiently.

20. What type of data processing task is best suited for MapReduce?
a) Real-time streaming analytics
b) Interactive data querying
c) Batch processing of large datasets
d) Online transaction processing

Answer: c) Batch processing of large datasets
Explanation: MapReduce is well-suited for batch processing tasks that involve analyzing large volumes of data in parallel across distributed nodes, making it ideal for processing large datasets 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