Started on Friday, 4 June 2021, 3:03 PM
State Finished
Completed on Friday, 4 June 2021, 3:41 PM
Time taken 38 mins 14 secs

Question **1**
Complete
Marked out of 1.00

_____ is a online NoSQL developed by Cloudera.

- A. HCatalog
- B. Hbase
- C. Oozie
- D. Imphala

Question **2**
Complete
Marked out of 1.00

Which of the following are the core components of Hadoop?

- A. Map Reduce
- B. HBase
- C. HDFS
- D. Both (a) and (b)

Question **3**
Complete
Marked out of 1.00

Which of the following is not a NoSQL database?

- A. None of the mentioned
- B. SQL Server
- C. Cassandra
- D. MongoDB

Question **4**

Complete

Marked out of 1.00

Which of the following is a wide-column store?

- A. Redis
- B. Cassandra
- C. Riak
- D. MongoDB

Question **5**

Complete

Marked out of 1.00

Most NoSQL databases support automatic _____ meaning that you get high availability and disaster recovery.

- A. all of the mentioned
- B. processing
- C. replication
- D. scalability

Question **6**

Complete

Marked out of 1.00

Which of the following are the simplest NoSQL databases?

- A. Wide-column
- B. All of the mentioned
- C. Key-value
- D. Document

Question **7**

Complete

Marked out of 1.00

_____ stores are used to store information about networks, such as social connections.

- A. Graph
- B. Key-value
- C. Wide-column
- D. Document

Question **8**

Complete

Marked out of 1.00

NoSQL databases is used mainly for handling large volumes of _____ data.

- A. semi-structured
- B. All of the mentioned
- C. unstructured
- D. structured

Question **9**

Complete

Marked out of 1.00

What are the main components of Big Data?

- A. Map Reduce
- B. YARN
- C. HDFS
- D. All of these

Question **10**

Complete

Marked out of 1.00

What are the different features of Big Data Analytics?

- A. Open-Source
- B. Scalability
- C. All the above
- D. Data Recovery

Question **11**

Complete

Marked out of 2.00

Which of the following can be used to control the number of part files in a map reduce program output directory?

- A. Number of Reducers
- B. Counter
- C. Partitioner
- D. Number of Mappers

Question **12**

Complete

Marked out of 2.00

Which of the following is/are true about combiners?

- A. Mappers can be used as a combiner class
- B. Combiners are primarily aimed to improve Map Reduce performance
- C. Combiners can be used for any Map Reduce operation
- D. Combiners can be used for mapper only job

Question **13**

Complete

Marked out of 2.00

Who will initiate the mapper?

- A. Job tracker
- B. Task tracker
- C. Combiner
- D. Reducer

Question **14**

Complete

Marked out of 2.00

How are keys and values presented and passed to the reducers during a standard sort and shuffle phase of MapReduce?

- A. Keys are presented to a reducer in random order; values for a given key are sorted in ascending order.
- B. Keys are presented to reducer in sorted order; values for a given key are not sorted.
- C. Keys are presented to a reducer in random order; values for a given key are not sorted.
- D. Keys are presented to reducer in sorted order; values for a given key are sorted in ascending order.

Question **15**

Complete

Marked out of 2.00

When is the earliest point at which the reduce method of a given Reducer can be called?

- A. It depends on the Input Format used for the job.
- B. As soon as at least one mapper has finished processing its input split.
- C. As soon as a mapper has emitted at least one record.
- D. Not until all mappers have finished processing all records.

Question **16**

Complete

Marked out of 2.00

Why we should use NoSQL?

When it comes to handling humongous data, we are often required to upgrade our hardware which is expensive. To resolve that we store data in multiple hosts. This is where we require NoSQL. The structure of many different forms of data is more easily handled and evolved with a NoSQL database. NoSQL databases are often better suited to storing and modeling structured, semi-structured, and unstructured data in one database.

Question **17**

Complete

Marked out of 2.00

What is key-value and document data models?

In a **key-value data model** store the data is considered to be inherently opaque to the database. key-value databases store all key-value pairs together in a single namespace, which is analogous to a relational schema.

A **document-oriented data model** system relies on internal structure in the document in order to extract metadata that the database engine uses for further optimization. Document databases organize documents into groups called collections, which are analogous to the tables in relational databases.

Question **18**

Complete

Marked out of 2.00

Explain peer-peer replication?

- SQL Server Peer to Peer replication is a replication type where the publisher server replicates data to multiple subscriber servers at the same time.
- Peer to peer SQL Server replication is useful for multiple data center locations across the globe. One centralized data center manages the data on the other datacenter data.
- It distribute the responsibility of the master to each node in the cluster. In this situation, testing is much easier since you can remove any node in the cluster and the other nodes will continue to function.
- The disadvantage of peer-to-peer networks is that there's an increased complexity and communication overhead that must occur for all nodes to be kept up to date with the cluster status.

Question **19**

Complete

Marked out of 2.00

Difference between sharding and replication?

- **Data replication** is the process of making multiple copies of data and storing them at different locations to improve their overall accessibility across a network. Similar to data mirroring, data replication can be applied to both individual computers and servers. The data replicates can be stored within the same system, on-site and off-site hosts, and cloud-based hosts. It refers to a database setup in which several copies of the same dataset are hosted on separate machines. The main reason to have replication is redundancy.
- **Sharding** is the process of breaking up large tables into smaller chunks called shards that are spread across multiple servers. A shard is essentially a horizontal data partition that contains a subset of the total data set, and hence is responsible for serving a portion of the overall workload. Sharding allows for horizontal scaling of data writes by partitioning data across multiple servers using a shard key. It's important to choose a good shard key.

Question **20**

Complete

Marked out of 2.00

Explain map-reduce?

The **MapReduce** algorithm contains two important tasks, namely Map and Reduce. The Map task takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key-value pairs). The Reduce task takes the output from the Map as an input and combines those data tuples (key-value pairs) into a smaller set of tuples. The reduce task is always performed after the map job. MapReduce is a programming model and an associated implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster.

Different phases of Map Reduce

- Input Phase
- Map
- Intermediate Keys
- Combiner
- Shuffle and Sort
- Reducer
- Output Phase

Example -

Count – Generates a token counter per word.

◄ Mid-1

Jump to...