By OnlineInterviewQuestions.com

Cassandra Interview Questions

Practice Best Cassandra Interview Questions

Nowadays, the industry is developing an interest in using Schema-less databases. For that reason, NoSQL is growing in this sector at a great pace. So, to prepare for your interviews, here we present some, **Cassandra Interview Questions**, which is the **NoSQL** database. Also, if you check the salary trend of NoSQL database developers, it is quite high. So, you can go blindly with this field and start preparing from today onwards. Let's have a look:

Q1. What do you understand by Cassandra?

It is a NoSQL based technology which is highly selected by the users and customers. This company is run by Apache. Cassandra is so popular because it is very capable to store and manage huge data without any loss or damages. It is written in Java. The most amazing feature of Cassandra is that it has no chance of failure. Cassandra is the mixture of the key-value store and column-oriented where Key-value represents the external chamber for an application while column represents the keyspace thing.

Q2. What are the main benefits of using Cassandra?

- First of all, the extraordinary points that makes the people so attracted to it that it has no chance of failure.
- It is very efficient as it delivers an exact time execution that is very helpful in analyzing the data. By this, it is quite easy to handle the work for engineers, developers etc.
- It is designed as on an equal footing, not like the master-slave.
- Again, it is very flexible for the users. As anyone can insert a number of nodes to any Cassandra in any of the data centers.
- Users are able to send requests to the servers.
- As correspondence to the new technologies, it leaves no issues of competition. It expedites scalability in which users can easily scale up or scale down as per the needs of the users. And further, it does not need any kind of refreshment in the processing of the operation while scaling.
- The next great point is replication. Users are able to copy the data as much as the copies they want. They can even store their data at different nodes. In the case of failure of any nodes, users can back up their data from another node.
- It is chosen as the most favored NoSQL DB by many companies and the organizations because of its excellent performance.
- Slicing is very easy and simple in Cassandra because it operates on column-oriented. This makes the many more function like accessing data and redemption.
- Last but not least it holds schema-free or schema-optional data pattern.

Q3. What do you mean by Tunable Consistency?

Tunable Consistency is used to keep the fresh and co-exist data rows on all their replicas. It permits the clients a better option in which they can select a consistency level as per their requirement. Tunable consistency is one of a kind features that make the users, developers, and architects having Cassandra their primary choice. Basically, it supports two kinds of consistencies

Eventual consistency- In this consistency, all the data is accessible from the last update, it has no new update. It is just mean to achieve replication of data.

Strong consistency- In this type of consistency, it supports some kind of conditions. These are: -

R+ W > N, where

N stands for the number of replications of data

W stands for the number of nodes that demand to grant for a prosperous write

R stands for the number of nodes that demand to grant for a prosperous read.

Q4. Explain the concept of compaction in Cassandra.

Compaction is very efficient in maintaining the process of arrangement for data update of the data structure on disk. Compaction is beneficial at the time of interaction with Memtable.

Generally, there are two kinds of Compaction

1. **Minor compaction** -It is a type of compaction in which equally sized SS Tables are adjusted into one. It does not need to start, as it starts automatically when a fresh SS Table is formed.
2. **Major compaction**- It can't start automatically, there is a node tool used as a trigger. It is used to condense the SS Table of a column family into the one.

Q5. Explain the concept of Cassandra Data Model?

Cassandra Data Model is composed of four main components:

Cluster: -It is inclusive of a lot of nodes and key spaces.

Keyspace: It consists of a namespace to the group having a lot of column family, particularly, one per division

Column: It is inclusive of a name of the column, timestamp, and value.

Column family: It consists of a number of the columns with row key referral.

Q6. Discuss the precautions that are needed to take care while adding a column?

- The column name is not matched with an already present column name
- The table is not limited to a compressed storage prospect.

Q7. What do you understand by the Super Column in Cassandra?

Cassandra Super Column is used to collect the same kind of data. These are really key-value sets. These values are referred to the column. It is a grouping arrangement of columns. They follow a sequel that is

Key store> column family > super column> column data structure in JSON (JavaScript Object Notation).

Q8. Differentiate between the terms: node, a cluster, and data center in Cassandra?

These all are the basic component of Cassandra. A node is a work as an individual machinery, a cluster is an accumulation of a great number of nodes and these nodes have a similar kind of collected data. While at the time of serving the customers where they are located at different locations Data centers are useful. In combination, we can say that it helps to group various nodes of a cluster into various data centers.

Q9. What do you mean by Column Family?

Column family as the name suggests it relates to a structure that has a large number of rows. These are associated with a key-value set. Key represents the title of the column while value suggests the column data. You can relate it to the hash map exist in Java. The Column family is very manageable as it provides one rows having a hundred of columns while the others provide just 2 columns. There are no limitations to list of columns.

Q10. What do you mean by SS Table and explain how it is different from the other original tables?

SS Table stands for Sorted String Table which indicates the presence of an important file in Cassandra and it accepts the repeated number of written memtables. These memtables are stockpiled on disk. It remains for every Cassandra table. A main feature of the SS Table is that it provides stability to the data files as it does not allow any changes once the data is written. Moreover, Cassandra generates three split files. These files are like bloom filter, partition summary and partition index.

Q11. Elucidate the concept of CAP theorem?

CAP is efficiently used at the time of handling and managing the scaling tactics. Whenever a desire of scaling is observed, CAP theorem play its vital role. CAP Theory stands for Consistency Availability and Partition tolerance theory which states that in the system same as Cassandra users cannot use all the three characteristics, they have to choose two of them and one is needed to sacrifice.

These three characteristics are: -

- Consistency: It gives the warranty for returning of recent write for the user.
- Availability-: It is a source of giving a reasonable reply within minimum time.
- Partition: It represents that the system will work also at the time when the network barrier or partitions occur.

Q12. Explain what do you understand by Cassandra- CQL collections?

Cassandra- CQL collections serve the clients to reserve a large number of values just in one variable. There are many ways to use the CQL collection in Cassandra. These are: -

- **List-** In arranging and managing the system of the data, a list is used. Moreover, it is also useful to store the value numerous times.
- **SET-** In order to keep and returned the group of elements in classified orders, SET is used. MAP- MAP is used to keep the key-value set of components.

Q13. Explain the Memtable in Cassandra?

As the name suggests, Memtable is related to memory. The data that is written is in a structure (in-memory) by Cassandra is termed as Memtable. All the content that is stored as key/column takes place in these structures. With the use of the key, it is easy to classify the data. For every Column Family, there is a definite Memtable and it is also useful at the time of regaining the column data from the key.

Q14. What is the procedure of data storage in Cassandra?

The data stored in Cassandra is in bytes. When the user or client is sure about the approver, then these bytes are encoded by the Cassandra according to the need. After the completion, a comparator orders the encoding based on the column.

Composites have a particular coding and are patterned in bytes. For each and every component there is always a storage of two-byte length and it is supported by the byte-encoded element which is further accompanied by a termination bit.

Q15. Describe the different consistency levels for read operation in Cassandra?

- **All**
  It is extremely consistent. It is compulsory to a write needs to be written to memtable and commit log which is on copy nodes in the group
- **EACH_QUORUM**
  It is compulsory for a write needs to be written to memtable and commit log on quorum which exists on copy nodes in all data centers
- **LOCAL_QUORUM**
  It is compulsory for a write needs to be written to memtable and commit log on the quorum of copy nodes but only in the same center.
- **ONE**
  It is compulsory for a write needs to be written to memtable and commit log with one or more replica node.
- **TWO**
  It is compulsory for a write needs to be written to memtable and commit log with one or more replica node.
- **THREE**
  Same as the above but it should be with three replica nodes, sequentially

Q16. Is it possible to add or delete Column Families in a working group?

Yes, it is possible to add or delete Column Families in a working group but before doing it, there has some precaution or procedure that the client has to follow. These precautions are: -

- Very first, users must assure that the commit log is clear and it can be done by 'node tool drain'.
- No data should be left in the commit log. For this Cassandra has to be turned off.
- Lastly, it is vital to delete the SS Table files for the raised CFs.

Q17. Discuss and explain the various types of Partitioners in Cassandra?

Murmur3 Partitioner: It is the default and the most important partitioner as it is better and well performed than the others. Its speed is more than Random Partitioner. With all of this, it is also functional for even distribution. It uses 64-bit hash values with Range: 263 to 263-1

Random Partitioner: Before the arrival of Cassandra 1.2, Random Partitioner was identified as the default. It is worked together with vnodes. As same as above, it is also functional as for even distribution. MD5 hash values partition key with Range: 0 to 2127-1

Byte Ordered Partitioner: Byte Ordered Partitioner is a system that are beneficially to organize the location of the keys in the Cassandra. raw byte array value in Byte ordered Partitioner of the row key checks and make the decision regarding the storage of rows on the nodes.

Q18. Give the differences between the different types of Primary Keys in Cassandra?

- ○ **Single primary key**

In this case, only one column is used as a primary key. This column is also referred to as partitioning key which is used to partition the data. By virtue of the partition key, data has been spread on various nodes.

- ○ **Compound Primary Key**

In this, the data is partitioned and then grouped. race_name is referred to as partitioning key while the race_position is referred to as clustering key. Former decides the partition of data and the latter decides the clustering of data.

Q19. .What do you understand by Logging in Cassandra?

It is in the logging directory where logs are written to the system.log and debug.log file. It is the simplest way to check what's happening in the database just by changing the logging level. We can configure it by programmatically or by manually.

Q20. What are different logging levels in Cassandra?

There are many levels which are described below:

1. **All**: It includes all levels addition with custom levels.
2. **DEBUG**: To debug an application, it designates fine-grained informational events.
3. **WARN**: Detects potentially harmful conditions.
4. **INFO**: Indicates informational messages to how the progress.
5. **ERROR**: Specifies error events

These are some question which will help you to crack your interview. Of Course, you should also prepare well in this field to get a highly payable job.

Q21. What do you understand by the term Snitch in Cassandra? Give some example.

It is the work of snitch that determines to which nodes belong. It can belong to data centers and racks. It provides the information to Cassandra about the replication strategy and network topology for replication schemes. There are several examples of snitches, some of these are:

Simple Snitch	Property File Snitch	Ec2Snitch	Cloud stack Snitch
Dynamic snitching	Rack Inferring Snitch	Gossiping Property File	Google Cloud Snitch

Q22. What is rack in cassandra?

Apache Cassandra Rack is a classified set of servers and the architecture of Cassandra manages racks so that no duplicate is stored redundantly inside a single rack, guaranteeing that replicas are spread throughout different racks if one rack stops working. Within a data center, there could be various racks with multiple servers.

Q23. Cassandra was developed by Whom?

Cassandra (NoSQL database management system) was developed by **Apache Software Foundation**.

Please Visit OnlineInterviewquestions.com to download more pdfs