Top 23 Cassandra Interview Questions & Answers

1) Explain what is Cassandra?

Cassandra is an open source data storage system developed at Facebook for inbox search and designed for storing and managing large amounts of data across commodity servers. It can server as both

- Real time data store system for online applications
- Also as a read intensive database for business intelligence system

2) What is the use of Cassandra and why to use Cassandra?

Cassandra was designed to handle big data workloads across multiple nodes without any single point of failure. The various factors responsible for using Cassandra are

- It is fault tolerant and consistent
- Gigabytes to petabytes scalabilities
- It is a column-oriented database
- No single point of failure
- No need for separate caching layer
- Flexible schema design
- It has flexible data storage, easy data distribution, and fast writes
- It supports ACID (Atomicity, Consistency, Isolation, and Durability)properties
- Multi-data center and cloud capable
- Data compression

3) Explain what is composite type in Cassandra?

In Cassandra, composite type allows to define key or a column name with a concatenation of data of different type. You can use two types of Composite Type

- Row Key
- Column Name

4) How Cassandra stores data?

- All data stored as bytes
- When you specify validator, Cassandra ensures those bytes are encoded as per requirement
- Then a comparator orders the column based on the ordering specific to the encoding
- While composite are just byte arrays with a specific encoding, for each component it stores a two byte length followed by the byte encoded component followed by a termination bit.

5) Mention what are the main components of Cassandra Data Model?

The main components of Cassandra Data Model are

- Cluster
- Keyspace
- Column
- Column & Family

6) Explain what is a column family in Cassandra?

Column family in Cassandra is referred for a collection of Rows.

7) Explain what is a cluster in Cassandra?

A cluster is a container for keyspaces. Cassandra database is segmented over several machines that operate together. The cluster is the outermost container which arranges the nodes in a ring format and assigns data to them. These nodes have a replica which takes charge in case of data handling failure.

8) List out the other components of Cassandra?

The other components of Cassandra are

- Node
- Data Center
- Cluster
- Commit log
- Mem-table
- SSTable
- Bloom Filter

9) Explain what is a keyspace in Cassandra?

In Cassandra, a keyspace is a namespace that determines data replication on nodes. A cluster consist of one keyspace per node.

10) What is the syntax to create keyspace in Cassandra?

Syntax for creating keyspace in Cassandra is

CREATE KEYSPACE <identifier> WITH <properties>

11) Mention what are the values stored in the Cassandra Column?

In Cassandra Column, basically there are three values

- Column Name
- Value
- Time Stamp

12) Mention when you can use Alter keyspace?

ALTER KEYSPACE can be used to change properties such as the number of replicas and the durable_write of a keyspace.

13) Explain what is Cassandra-Cqlsh?

Cassandra-Cqlsh is a query language that enables users to communicate with its database. By using Cassandra cqlsh, you can do following things

- Define a schema
- Insert a data and
- Execute a query

14) Mention what does the shell commands "Capture" and "Consistency" determines?

There are various Cqlsh shell commands in Cassandra. Command "Capture", captures the output of a command and adds it to a file while, command "Consistency" display the current consistency level or set a new consistency level.

15) What is mandatory while creating a table in Cassandra?

While creating a table primary key is mandatory, it is made up of one or more columns of a table.

16) Mention what needs to be taken care while adding a Column?

While adding a column you need to take care that the

- Column name is not conflicting with the existing column names
- Table is not defined with compact storage option

17) Mention what is Cassandra- CQL collections?

Cassandra CQL collections help you to store multiple values in a single variable. In Cassandra, you can use CQL collections in following ways

- **List**: It is used when the order of the data needs to be maintained, and a value is to be stored multiple times (holds the list of unique elements)
- **SET**: It is used for group of elements to store and returned in sorted orders (holds repeating elements)
- **MAP**: It is a data type used to store a key-value pair of elements

18) Explain how Cassandra writes data?

Cassandra writes data in three components

- Commitlog write

- Memtable write
- SStable write

Cassandra first writes data to a commit log and then to an in-memory table structure memtable and at last in SStable

19) Explain what is Memtable in Cassandra?

- Cassandra writes the data to a in memory structure known as Memtable
- It is an in-memory cache with content stored as key/column
- By key Memtable data are sorted
- There is a separate Memtable for each ColumnFamily, and it retrieves column data from the key

20) Explain what is SStable consist of?

SStable consist of mainly 2 files

- Index file (Bloom filter & Key offset pairs)
- Data file (Actual column data)

21) Explain what is Bloom Filter is used for in Cassandra?

A bloom filter is a space efficient data structure that is used to test whether an element is a member of a set. In other words, it is used to determine whether an SSTable has data for a particular row. In Cassandra it is used to save IO when performing a KEY LOOKUP.

22) Explain how Cassandra writes changed data into commitlog?

- Cassandra concatenate changed data to commitlog
- Commitlog acts as a crash recovery log for data
- Until the changed data is concatenated to commitlog write operation will be never considered successful

Data will not be lost once commitlog is flushed out to file.

23) Explain how Cassandra delete Data?

SSTables are immutable and cannot remove a row from SSTables. When a row needs to be deleted, Cassandra assigns the column value with a special value called Tombstone. When the data is read, the Tombstone value is considered as deleted.

Refer our [Cassandra Tutorials](#) for an extra edge in your interview.

Guru99 Provides   FREE ONLINE TUTORIAL   on Various courses like

Java	MIS	MongoDB	BigData	Cassandra

Web Services	SQLite	JSP	Informatica	Accounting

SAP Training	Python	Excel	ASP Net	HBase

Project Management	Test Management	Business Analyst	Ethical Hacking	PMP

Live Project	SoapUI	Photoshop	Manual Testing	Mobile Testing

Data Warehouse	R Tutorial	Tableau	DevOps	AWS

Jenkins	Agile Testing	RPA	JUnit	Software Engineering

Selenium	CCNA	AngularJS	NodeJS	PLSQL