

[Home](#)

🌟 [Shikha Sharma](#) — Published On March 2, 2023

[Beginner](#) [Data Engineering](#) [Database](#) [Interview Questions](#) [NoSQL](#)

Introduction

Cassandra is an Apache-developed free and open-source distributed [NoSQL database management system](#). It manages huge volumes of data across many commodity servers, ensures fault tolerance with the swift transfer of data, and provides high availability with no single point of failure.



Java-written Apache Cassandra is highly scalable for Big Data models and comprises flexible schemas. It is a hybrid of column-oriented and key-value store databases, initially designed by Facebook.

In this blog, we discuss Cassandra's questions deeply, which are beneficial for beginners and experts.

Learning Objectives

Below is what we'll learn after reading this blog thoroughly:

- A common understanding of what Cassandra is and its role in the technical era.
- Knowledge of Cassandra-related terms like Data Replication, Commit Log, Composite Key, Consistency, Cluster, etc.
- An understanding of the CAP theorem.
- An understanding of Eventual and Strong tunable consistencies.
- Insights into the concept of when Cassandra is useful and when we need to avoid its usage.

Overall, by reading this blog, we will gain a comprehensive understanding of managing a large volume of data. We will be equipped with the knowledge and ability to use this technique effectively and handle the coming flow of interview questions.

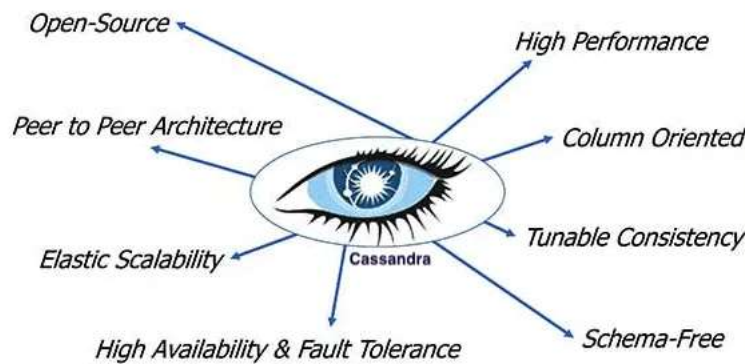
This article was published as a part of the [Data Science Blogathon](#).

Table of Contents

1. [Q1. Why use Cassandra Despite Having Many Traditional Databases?](#)
2. [Q2. Explain the Following Terms in Cassandra.](#)
3. [Q3. Define CAP Theorem in Cassandra.](#)
4. [Q4. Explain Tunable Consistency in Cassandra.](#)

Top 5 Interview Questions on Cassandra

We have many reasons to prove the consideration of Apache Cassandra that are enough to replace traditional databases. Some are:



Real-time Performance: Apache Cassandra simplifies the job of many Software Engineers, Developers, Administrators, and Data Analysts by providing a near real-time performance that is not available in usual databases.

Peer-to-peer Architecture: It assures no failure just because of its peer-to-peer architecture, whereas in traditional databases, we still use the master-slave architecture. In any data center, Cassandra allows the insertion of multiple nodes into any cluster, which assures phenomenal flexibility. It allows clients to forward its request to any server.

Scalability: When it comes to scalability, Cassandra allows us to scale up and down easily per the user requirements, facilitating extensible scalability. At the time of scaling, we don't have to restart this NoSQL application specifically with high throughput for read and write operations.

Data Replication: Cassandra facilitates vital data replication on node capability by allowing users to access the data in another location if one node fails. It offers data storage at multiple locations. Users can choose the number of replicas they want to create as per their requirements.

Massive Dataset: Cassandra is often called the most preferable [NoSQL database](#) as it offers outstanding performance when used for massive datasets.

Column-Oriented: Cassandra is a column-oriented database that makes data access and retrieval efficient and fastens the slicing process.

Schema-Free data model: As Cassandra follows the schema-optional data model, we are not bound to show all the columns of an application; we can avoid unwanted data.

Q2. Explain the Following Terms in Cassandra.

1. Data Replication: Cassandra supports the data replication feature to ensure data redundancy and fault tolerance in the database. Data Replication is basically an operation in which data from one node is copied to other nodes in the cluster. Data replication comprises two components: the replication factor, which decides the count of copies, and the replication strategy, which decides the nodes in which the data is copied.

2. Commit Log: Commit Log is a mechanism that is used at the time of database crashes to recover data. We can recover the data from the commit log easily because every operation that is carried out is saved/defined in the commit log.

3. Composite Key: Cassandra's composite keys are made up of a row key and column name, used to declare the column family

Top 5 Interview Questions on Cassandra

6. SSTable: SSTable stands for the Sorted String Table, a data file that accepts the regular Mem Tables.

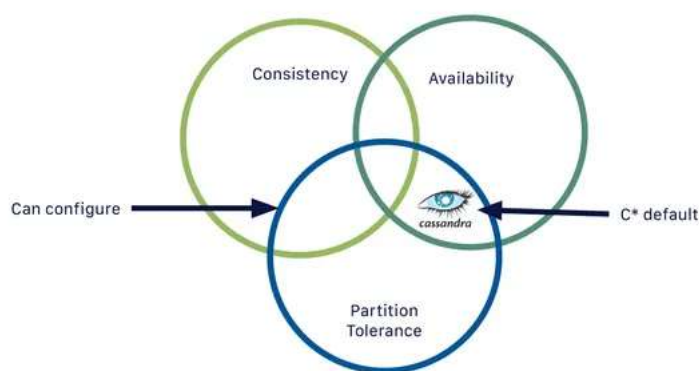
7. Data Center: As the name suggests, the Data center is a collection of all the data that is available in the Cluster.

8. YAML file in Cassandra: The main configuration file of Cassandra is Cassandra.yaml file; we have to restart the node to see the changes just after updating any properties in this Cassandra.yaml file.

9. Clusters: Clusters are basically the containers for the Keyspaces. They are the outermost structure in Cassandra and are often known as rings because the data to the cluster node is arranged in a circular ring.

Q3. Define CAP Theorem in Cassandra.

CAP stands for Consistency, Availability, and Partition Tolerance, this theorem plays a significant role in managing the scaling strategy by the time it's required to scale systems when additional resources are needed.



Source: cassandra.apache.org

[CAP theorem](#) is an efficient method to handle scaling in distributed systems like Cassandra. According to the CAP theorem, users can take advantage of only two out of these three characteristics by sacrificing one. We have two possibilities for the characteristics: AP (Availability and Partition Tolerance) and CP (Consistency and Partition Tolerance).

The characteristics are defined as follows:

Consistency: It ensures the user's return of the most recent write.

Availability: It ensures a rational response within a minimum time.

Partition Tolerance: It ensures that the system will continue its operations whenever the network partition occurs. The two options available are AP and CP.

Q4. Explain Tunable Consistency in Cassandra.

Tunable consistency is nothing but Cassandra's phenomenal characteristic, making it a famous choice in competition with other traditional databases. The synchronized and up-to-date data rows among the replicas are what consistency refers to here. In order to choose the best-suited consistency level for our use cases, we rely on its Tunable Consistency. Tunable Consistency supports two types of consistencies:-

- 1. Eventual Consistency:** When we don't have any new update or change over the given data items, we refer to the eventual consistency where all accesses eventually return the last updated value.

Top 5 Interview Questions on Cassandra

N: Number of replicas

W: Number of nodes that have to agree for a successful write

R: Number of nodes that have to agree for a successful read

So, we can say that strong consistency is when the number of replicas is smaller than the summation of the number of nodes with successful reads and the number of nodes with successful writes.

Q5. When to use Cassandra, and When not to use?

When it comes to the real-time usage of Cassandra, it plays a significant role in solving the need for heavy write systems by offering an efficient, responsive reporting system. Cassandra is a member of the NoSQL family, and it helps in cases when we're looking for quick responses. You can see the usage in web analytics where the log data of requests is stored, and we desire to create an analytical platform around that for counting the hits by the IP, by browser, by the hour, etc. in a real-time format.



Source: [data-flair.training](#)

Although Cassandra is the most famous and reliable technique, when it comes to the use cases like Financial data, it won't suit due to the unavailability of ACID property and relational databases. It is based on the [NoSQL database](#), and when our demand is to manage the data concerning ACID properties, the scenario becomes hard. We can perform that process here, but we have to write lots of application codes to manage the [ACID property](#) and will lose time to market badly. Also, handling these codes in Cassandra would be hard and tedious for us.

Conclusion

This blog covers some of the frequently asked Apache Cassandra interview questions that could be asked in data science and big data developer interviews. Using these interview questions as a reference, you can better understand the concept of Apache Cassandra and start formulating effective answers for upcoming interviews. The key takeaways from this blog are:

1. Apache Cassandra is a Java-written, [NoSQL](#) database management system that can manage large volumes of data and ensures the fault-tolerance.
2. Although we have many traditional databases, the availability of features like real-time performance, peer-to-peer architecture, scalability, data replication, schema-free model, etc., makes it a unique and irreplaceable technique.
3. We learn some of the commonly asked definitions of rapid-fire interview rounds in this blog, including SSTable, Memtable, Commit Log, Data Replication, Consistency, etc.
4. It offers Eventual and Strong tunable consistencies which keep the data up-to-date.

Top 5 Interview Questions on Cassandra

[Apache](#) [blogathon](#) [Cassandra](#) [database](#) [databases](#) [interview questions](#) [NoSQL](#)

About the Author



[Shikha Sharma](#)

Our Top Authors



[view more](#)

Download

Analytics Vidhya App for the Latest blog/Article



Previous Post

[What is Quantum Machine Learning?](#)

Next Post

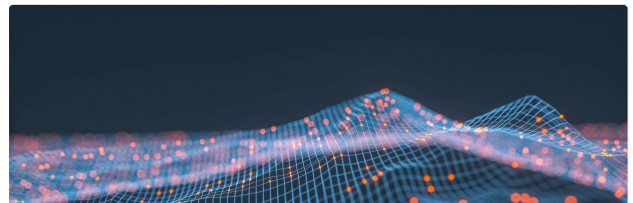
[Understanding the Basics of Database Normalization](#)

Top Resources



[Understand Random Forest Algorithms With Examples \(Updated 2023\)](#)

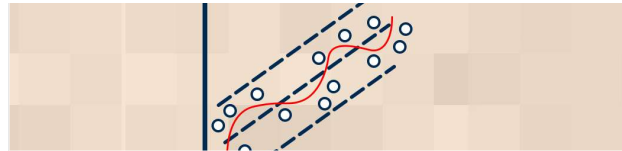
[Sruthi ER](#) - JUN 17, 2021



[Decision Tree Algorithm – A Complete Guide](#)

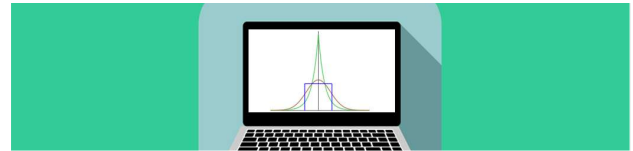
[Anshul Saini](#) - AUG 29, 2021

Top 5 Interview Questions on Cassandra



[Everything you need to Know about Linear Regression!](#)

👤 [KAVITA MALI](#) - OCT 04, 2021



[Skewness and Kurtosis: Quick Guide \(Updated 2023\)](#)

👤 [Suvarna Gawali](#) - MAY 02, 2021

Download App



Analytics Vidhya

[About Us](#)

[Our Team](#)

[Careers](#)

[Contact us](#)

Companies

[Post Jobs](#)

[Trainings](#)

[Hiring Hackathons](#)

[Advertising](#)

Data Scientists

[Blog](#)

[Hackathon](#)

[Discussions](#)

[Apply Jobs](#)

Visit us



© Copyright 2013-2023 Analytics Vidhya.

[Privacy Policy](#) [Terms of Use](#) [Refund Policy](#)