Learn Cassandra in 1 Day

By Krishna Rungta

Table Of Content

Chapter 5: Create, Alter & Drop Keyspace in Cassandra with Example

1. How to Create Keyspace
2. Alter Keyspace
3. Drop/Delete Keyspace

Chapter 6: Cassandra Table: Create, Alter, Drop & Truncate (with Example)

1. How to Create Table
2. Cassandra Alter table
3. Drop Table
4. Truncate Table

Chapter 7: Cassandra Query Language(CQL): Insert Into, Update, Delete (Example)

1. Insert Data
2. Upsert Data
3. Update Data
4. Cassandra Delete Data
5. What Cassandra does not support
6. Cassandra Where Clause

Chapter 8: Create & Drop INDEX in Cassandra

1. Cassandra Create Index
2. Cassandra Drop Index

Chapter 9: Cassandra CQL Data Types & Data Expiration using TTL (Example)

1. Cassandra Data Types
2. Cassandra Automatic Data Expiration using Time to Live (ttl)

Chapter 10: Cassandra Collection: Set, List, Map with Example

1. Cassandra Set
2. Cassandra List
3. Cassandra Map

Chapter 11: Cassandra Cluster Setup on Multiple Nodes (Machines)

1. Prerequisites for Cassandra Cluster
2. Enterprise Edition Installation
3. Starting Cassandra Node

Chapter 12: DataStax DevCenter & OpsCenter Installation Guide

1. DevCenter Installation
2. OpsCenter Installation

Chapter 13: Cassandra Security: Create User & Authentication with JMX

1. What is Internal Authentication and Authorization
2. Configure Authentication and Authorization
3. Logging in
4. Create New User
5. Authorization
6. Configuring Firewall
7. Enabling JMX Authentication

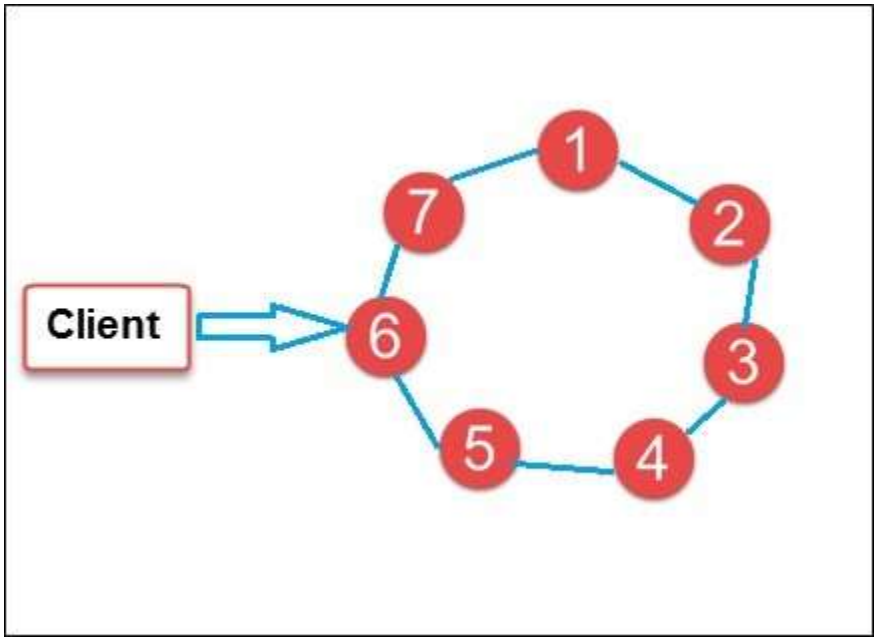
Chapter 1: What is Apache Cassandra?

What is Apache Cassandra?

Cassandra is a distributed database management system designed for handling a high volume of structured data across commodity servers

Cassandra handles the huge amount of data with its distributed architecture. Data is placed on different machines with more than one replication factor that provides high availability and no single point of failure.

In the image below, circles are Cassandra nodes and lines between the circles shows distributed architecture, while the client is sending data to the node.



Cassandra History

- Cassandra was first developed at Facebook for inbox search. Facebook
- open sourced it in July 2008.
- Apache incubator accepted Cassandra in March 2009. Cassandra is a
- top level project of Apache since February 2010. The latest version of
- Apache Cassandra is 3.2.1.

First let's understand what NoSQL database is.

Nosql Cassandra Database

NoSQL databases are called "Not Only SQL" or "Non-relational" databases. NoSQL databases store and retrieve data other than tabular relations such as relation databases.

NoSQL databases include MongoDB, HBase, and Cassandra. There are

following properties of NoSQL databases.

- Design Simplicity
- Horizontal Scaling
- High Availability

Data structures used in Cassandra are more specified than data structures used in relational databases. Cassandra data structures are faster than relational database structures.

NoSQL databases are increasingly used in Big Data and real-time web applications. NoSQL databases are sometimes called Not Only SQL i.e. they may support SQL-like query language.

Nosql Cassandra Database Vs Relational databases

Here are the differences between relation databases and NoSQL databases in a tabular format.

Relational Database	**NoSQL Database**
Handles data coming in low velocity	Handles data coming in high velocity
Data arrive from one or few locations	Data arrive from many locations
Manages structured data	Manages structured unstructured and semi- structured data.
Supports complex transactions (with joins)	Supports simple transactions
single point of failure with failover	No single point of failure
Handles data in the moderate volume.	Handles data in very high volume
Centralized deployments	Decentralized deployments
Transactions written in one location	Transaction written in many locations
Gives read scalability	Gives both read and write scalability
Deployed in vertical fashion	Deployed in Horizontal fashion

Apache Cassandra Features

There are following features that Cassandra provides.

- **Massively Scalable Architecture:** Cassandra has a masterless design where all nodes are at the same level which provides operational simplicity and easy scale out.
- **Masterless Architecture:** Data can be written and read on any node.
- **Linear Scale Performance:** As more nodes are added, the performance of Cassandra increases.
- **No Single point of failure:** Cassandra replicates data on

different nodes that ensures no single point of failure.

- **Fault Detection and Recovery:** Failed nodes can easily be restored and recovered.
- **Flexible and Dynamic Data Model:** Supports datatypes with Fast writes and reads.
- **Data Protection:** Data is protected with commit log design and build in security like backup and restore mechanisms.
- **Tunable Data Consistency:** Support for strong data consistency across distributed architecture.
- **Multi Data Center Replication:** Cassandra provides feature to replicate data across multiple data center.
- **Data Compression:** Cassandra can compress up to 80% data without any overhead.
- **Cassandra Query language:** Cassandra provides query language that is similar like SQL language. It makes very easy for relational database developers moving from relational database to Cassandra.

Cassandra Use Cases/Application

Cassandra is a non-relational database that can be used for different types of applications. Here are some use cases where Cassandra should be preferred.

- **Messaging**

  Cassandra is a great database for the companies that provides Mobile phones and messaging services. These companies have a huge amount of data, so Cassandra is best for them.

- **Internet of things Application**

  Cassandra is a great database for the applications where data is coming at very high speed from different devices or sensors.

- **Product Catalogs and retail apps**

  Cassandra is used by many retailers for durable shopping cart protection and fast product catalog input and output.

- **Social Media Analytics and recommendation engine**

  Cassandra is a great database for many online companies and social media providers for analysis and recommendation to their customers.