

# Introduction to Cassandra / Cassandra Quick Start - TTDS6776

Explore Core Skills Required to Create Solid Models with Cassandra (C\*), Explore CQL, Querying & More

**Duration:** 2 Days

**Skill Level:** Introductory

**Available Format:** Instructor-Led Online; Instructor-Led, Onsite In Person ; Blended; On Public Schedule

Introduction to Cassandra is a two day, hands-on course designed to teach attendees the basics of how to create good data models with Cassandra. This technical course has a focus on the practical aspects of working with C\*, and introduces essential concepts needed to understand Cassandra, including enough coverage of internal architecture to make good decisions.

## What You'll Learn

### Overview

The Cassandra (C\*) database is a massively scalable NoSQL database that provides high availability and fault tolerance, as well as linear scalability when adding new nodes to a cluster. It has many powerful capabilities, such as tunable and eventual consistency, that allow it to meet the needs of modern applications, but also introduce a new paradigm for data modeling that many organizations do not have the expertise to use in the best way.

**Introduction to Cassandra** is a two day, hands-on course designed to teach attendees the basics of how to create good data models with Cassandra. This technical course has a focus on the practical aspects of working with C\*, and introduces essential concepts needed to understand Cassandra, including enough coverage of internal

architecture to make good decisions. It is hands-on, with labs that provide experience in core functionality. Students will also explore CQL (Cassandra Query Language), as well as some of the "anti-patterns" that lead to non-optimal C\* data models and be ready to work on production systems involving Cassandra.

## Objectives

The goal of this course is to enable technical students new to Cassandra to begin working with Cassandra in an optimal manner. This course combines engaging instructor-led presentations and useful demonstrations with valuable hands-on labs and engaging group activities. Throughout the course you will learn to:

- Understand the Big Data needs that C\* addresses
- Be familiar with the operation and structure of C\*
- Be able to install and set up a C\* database
- Use the C\* tools, including cqlsh, nodetool, and ccm (Cassandra Cluster Manager)
- Be familiar with the C\* architecture, and how a C\* cluster is structured
- Understand how data is distributed and replicated in a C\* cluster
- Understand core C\* data modeling concepts, and use them to create well-structured data models
- Be familiar with the C\* eventual consistency model and use it intelligently
- Be familiar with consistency mechanisms such as read repair and hinted handoff
- Understand and use CQL to create tables and query for data
- Know and use the CQL data types (numerical, textual, uuid, etc.)
- Be familiar with the various kinds of primary keys available (simple, compound, and composite primary keys)
- Be familiar with the C\* write and read paths
- Understand C\* deletion and compaction
- Optional: Get introduced to using Cassandra and IntelliJ

If your team requires different topics, additional skills or a custom approach, our team will collaborate with you to adjust the course to focus on your specific learning objectives and goals.

## Audience

This introductory-level course is geared for data engineers, database administrators, system architects, and software developers, or those who are new to or have basic familiarity with NoSQL databases and are interested in building robust, scalable data-driven applications. Professionals who are tasked with managing or designing distributed data systems, working in industries where data scalability and availability are of high importance, will find this course particularly useful. Furthermore, any individual involved in decision-making processes around technology choices, architecture or data modeling would benefit from the unique insights and practical skills developed in this hands-on course, ensuring optimal usage of the Cassandra database in production environments.

## Pre-Requisites

To ensure a smooth learning experience and maximize the benefits of attending this course, you should have the following prerequisite skills:

- Since Cassandra is a type of database, it is crucial that participants have some fundamental knowledge about databases. Knowing SQL would be beneficial. This includes understanding concepts such as tables, records, indexes, and queries.
- While not specific to any one language, participants should be comfortable with general programming concepts like variables, data types, loops, conditionals, and functions.
- Some of the operations with Cassandra will require using CLI tools. Therefore, attendees should be comfortable with using a command line interface on their chosen operating system.
- Though the course will dive deep into data modeling with Cassandra, having a basic understanding of data modeling concepts such as entities, relationships, and schema design would provide a strong foundation and enrich the learning experience.

## Agenda

*Please note that this list of topics is based on our standard course offering, evolved from typical industry uses and trends. We'll work with you to tune this course and level of coverage to target the skills you need most. Topics, agenda and labs are subject to change, and may adjust during live delivery based on audience skill level, interests and participation.*

## **Cassandra Overview**

- Why We Need Cassandra - Big Data Challenges vs RDBMS
- High level Cassandra Overview
- Cassandra Features
- Basic Cassandra Installation and Configuration

## **Cassandra Architecture and CQL Overview**

- Cassandra Architecture Overview
- Cassandra Clusters and Rings
- Nodes and Virtual Nodes
- Data Replication in Cassandra
- Introduction to CQL
- Defining Tables with a Single Primary Key
- Using cqlsh for Interactive Querying
- Selecting and Inserting/Upserting Data with CQL
- Data Replication and Distribution
- Basic Data Types (including uuid, timeuuid)

## **Data Modeling and CQL Core Concepts**

- Defining a Compound Primary Key
- CQL for Compound Primary Keys
- Partition Keys and Data Distribution
- Clustering Columns
- Overview of Internal Data Organization
- Overview of Other Querying Capabilities
- ORDER BY, CLUSTERING ORDER BY, UPDATE , DELETE, ALLOW FILTERING
- Batch Queries
- Data Modeling Guidelines
- Denormalization
- Data Modeling Workflow
- Data Modeling Principles
- Primary Key Considerations
- Composite Partition Keys
- Defining with CQL
- Data Distribution with Composite Partition Key

- Overview of Internal Data Organization
- Lab: Composite Partition Key (Substantial lab)

### **Additional CQL Capabilities**

- Indexing
- Primary/Partition Keys and Pagination with token()
- Secondary Indexes and Usage Guidelines
- Cassandra collections
- Collection Structure and Uses
- Defining and Querying Collections (set, list, and map)
- Materialized View
- Usage Guidelines

### **Data Consistency In Cassandra**

- Overview of Consistency in Cassandra
- CAP Theorem
- Eventual (Tunable) Consistency in C\* - ONE, QUORUM, ALL
- Choosing CL ONE
- Choosing CL QUORUM
- Achieving Immediate Consistency
- Overview of Other Consistency Levels
- Supportive Consistency Mechanisms
- Writing / Hinted Handoff
- Read Repair
- Nodetool repair

### **Internal Mechanisms**

- Ring Details
- Partitioners
- Gossip Protocol
- Snitches
- Write Path
- Overview / Commit Log

- Memtables and SSTables
- Write Failure
- Unavailable Nodes and Node Failure
- Requirements for Write Operations
- Read Path Overview
- Read Mechanism
- Replication and Caching
- Deletion/Compaction Overview
- Delete Mechanism
- Tombstones and Compaction

#### **OPTIONAL: Working with IntelliJ**

- Configuring JDBC Data Source for Cassandra
- Reading Schema Information
- Querying and Editing Tables

All applicable course software, digital courseware files or course notes, labs, data sets and solutions, live coaching support channels and rich extended learning and post training resources are provided for you in our “easy access, no install required” online **Learning Experience Platform (LXP)**, remote lab and content environment. Access periods vary by course. We’ll collaborate with you to ensure your team is set up and ready to go well in advance of the class. Please inquire about set up details and options for your specific course of interest.

## **For More Information**

Please [contact us](#) or call 844-475-4559 toll free for more information about our training services (instructor-led, self-paced or blended), coaching and mentoring services, public course enrollment or questions, partner programs, courseware licensing options and more.