

Introduction to Apache Cassandra

This 3-day class cover the essentials of Cassandra DB for DBAs Architects and Java Developers

By the end of this course you will master Apache Cassandra installation administration and development using the Datastax Java Driver for Cassandra.

Apache Cassandra is a free and open-source, distributed, wide column store, NoSQL database management system designed to handle large amounts of data across many commodity servers, providing high availability with no single point of failure. Cassandra offers robust support for clusters spanning multiple datacenters, with asynchronous master-less replication allowing low latency operations for all clients.

Note: A 2 day class that focuses on architecture and administration is available

123Completed Experts assisted numerous customers in Architecting the Cassandra cluster and Data Schema for achieving zero data loss, maximize performance and gain availability

AUDIENCE

Architects
DBAs
Developers

KNOWLEDGE REQUIREMENTS

SQL
RDBMS
Java

LENGTH

3 Days

BONUS

Hands-on lab sessions

SYLLABUS

Day 1

Course Introduction
Cassandra Introduction
Data Modeling with Cassandra
Data Manipulation
Query Advanced Topics
Cassandra Read Path
Cassandra Cluster
Consistency

Day 2

Lightweight Transactions
Disks usage and data sizing
Data Center Aware Keyspace
Configuration Files
Snitch and The Gossip Protocol
Nodetool
Monitoring Cassandra
Backup (Snapshot) and Recovery
Compaction

Day 3

The Java Driver
Simple Statements
Prepared Statements
Querying JSON
Paging
Query Builder
Advanced Java Driver Topics
Summary

HARDWARE AND SOFTWARE REQUIREMENTS

Computer Requirements

- RAM: minimum 8 GB of RAM required for exercises and platform to operate, 16 GB and up recommended.
- Disk Space: At least 40 GB of free disk space
- Internet connection
- All machines connected to same Network

Supported Operating Systems

- Linux Ubuntu VM is provided.
- Machine should be able to run VMware Player and host the Ubuntu VM

Additional Software Requirements

- PDF Reader
- Zip software

Class HW required (If delivered on customer site)

- Projector 1024*768 minimum resolution (HD preferred)
- White Board
- Erasable Markers
- Desktops or Laptops (see HW Requirements)
- 12-24 ports Switch
- Internet connectivity
- Electricity outlets for all computers/monitors and other equipment.
- At least 3 electricity outlets next to instructor location.

DAY 1 - AGENDA

Lesson 1: Course Introduction

Duration: 0.5 hour Including Lab

- Course Introduction
- Courseware walkthrough
- Course environment
- Documentation
- Lab

Lesson 2: Cassandra Introduction

Duration: 1.5 hours Including Lab

- What is Apache Cassandra?
- Getting Started with Cassandra
- Introduction to CQL - Cassandra Query Language
- CQLSH – Cassandra CLI
- Data Modeling with Cassandra
- Basic CRUD operations
- Introduction to Cassandra Cluster
- Introduction to Nodetool
- Cassandra Terminology

Lesson 3: Data Modeling with Cassandra

Duration: 1.5 hours Including Lab

- Keyspaces
- Cassandra Tables
- Table Components - MemTable and SSTable
- Cassandra Write Path the basics
- Google Big Table
- Primary Key
- Partition Key
- Clustering Columns
- Table Columns
- Cassandra Cell
- CQL vs SQL
- Data Types
- Collections
- User Define Types
- Indexes
- JSON Query
- Materialized Views and should you use them?

Lesson 4: Data Manipulation

Duration: 1 hour Including Lab

- Insert
- Update
- Delete
- Counters
- Write Path Basics
- Commit Log
- Tombstones

- **Compaction Basics**

Lesson 5: Query Advanced Topics**Duration: 1 hour Including Lab**

- Querying non-partition key columns
- Indexes dos and don'ts
- Aggregations and Cassandra

Lesson 6: Cassandra Read Path**Duration: 0.5 hour**

- Table structure revisited
- Bloom Filters
- Key cache
- Row cache
- Full read path scenario

Lesson 7: Cassandra Cluster**Duration: 1 hour Including Lab**

- Cassandra Ring
- Tokens
- Partitioning
- Keyspace Replication Factor
- Keyspace Network Topology
- SimpleStrategy
- NetworkTopologyStrategy

Lesson 8: Consistency**Duration: 1 hour Including Lab**

- Tunable consistency
 - Read Consistency
 - Write Consistency
 - Read or Write Failure
 - Hinted Handoff
 - Time To Live
 - Tombstones
-

DAY 2 - AGENDA

Lesson 9: Lightweight Transactions

Duration: 0.75 hours
Including Lab

- Compare and set
- Transaction considerations
- Paxos

Lesson 10: Disks usage and data sizing

Duration: 0.75 hour

- Cassandra files
- Data disk sizing methodology
- Log files

Lesson 11: Datacenter aware Keyspace

Duration: 1 hour Including
Lab

- NetworkTopologyStrategy
- Data Center Configuration
- Multi DC R/W consistency
- Step by step Multi DC cluster configuration

Lesson 12: Configuration Files

Duration: 1 hour

- Installation Directory
- Cassandra.yaml
- Commonly used configurations
- Performance related configuration
- Advanced configurations
- Security configuration (overview)

Lesson 13: Snitch and The Gossip Protocol

Duration: 0.5 hour

- Gossip
- Snitch

Lesson 14: Nodetool

Duration: 1 hour Including
Lab

- Nodetool
- Most common commands
- Nodetool repair
- Adding a node to the cluster
- Removing a node from a cluster
- Rebuilding

Lesson 15: Monitoring Cassandra

Duration: 1 hours Including
Lab

- Monitoring Introduction
- Monitoring Cassandra Using Nodetool

Lesson 16: Backup (Snapshot) and Recovery

Duration: 1 hours Including
Lab

- Cassandra Snapshots
- Incremental and full Snapshots
- Recover from a Snapshot

Lesson 17: Compaction

Duration: 0.5 hour Including
Lab

- Compaction process
- Forcing compaction

DAY 3 - AGENDA**Lesson 18: The Java Driver**

Duration: 1 hour Including Lab

- Datastax Java Driver
- Driver Classes
- Connecting to the cluster
- Address Resolution
- Authentication
- Session

Lesson 19: Simple Statements

Duration: 1.5 hour Including Lab

- Simple Statements
- Select
- Insert
- Update
- Delete

Lesson 20: Prepared Statements

Duration: 1.5 hour Including Lab

- Prepared Statements
- Build Statement
- Batch Statement

Lesson 21: Querying JSON

Duration: 0.5 hour including lab

- What is JSON
- Querying JSON

Lesson 22: Paging

Duration: 0.75 hour including lab

- Paging
- Fetch Size
- Result Set

Lesson 23: Query Builder

Duration: 0.1 hour Including Lab

- Query builder
- CRUD with Query Builder
- Retry policy
- Additional properties

Lesson 24: Advanced Java Driver Topics

Duration: 1.5 hour Including Lab

- Load Balancing
- Failover and Retry policy
- Statement TTL, TIMESTAMP and LIMIT
- Consistency Level
- Lightweight Transactions

Lesson 18: Summary

Duration: 0.5 hour Including Lab

- Summary