

## Introduction to Apache Cassandra

### Course Summary

#### Description

This course will introduce Cassandra –a popular NoSQL database. It will cover Cassandra principles, architecture, and data models. Students will learn data modeling in CQL (Cassandra Query Language) by participating in interactive, hands-on labs. This session also discusses Cassandra internals as well as some admin topics.

#### Topics

- Introduction to Big Data / NoSQL
- Cassandra Basics
- Data Modeling (Part One)
- Data Modeling (Part Two)
- Data Modeling Labs – Group Design Session
- Cassandra Drivers
- Cassandra Internals
- Administration
- Bonus Lab (If Time Permits)

#### Prerequisites

Class participants should be comfortable with Java programming language and comfortable in the Linux environment (i.e., navigating command line, editing files with vi / nano).

#### Duration

Three days

## Introduction to Apache Cassandra

### Course Outline

- I. Introduction to Big Data / NoSQL**
  - A. NoSQL overview
  - B. CAP Theorem
  - C. When is NoSQL Appropriate?
  - D. Columnar Storage
  - E. NoSQL Ecosystem
- II. Cassandra Basics**
  - A. Design and Architecture
  - B. Cassandra Nodes, Clusters, Datacenters
  - C. Keyspaces, Tables, Rows, Columns
  - D. Partitioning, Replication, Tokens
  - E. Quorum and Consistency Levels
  - F. Labs
- III. Data Modeling (Part One)**
  - A. Introduction to CQL
  - B. CQL Datatypes
  - C. Creating Keyspaces and Tables
  - D. Choosing Columns and Types
  - E. Choosing Primary Keys
  - F. Data Layout for Rows and Columns
  - G. Time to Live (TTL)
  - H. Querying with CQL oCQL Updates
  - I. Collections (list / map / set)
  - J. Labs
- IV. Data Modeling (Part Two)**
  - A. Creating and Using Secondary Indexes
  - B. Composite Keys – Partition and Clustering
  - C. Time Series Data
  - D. Best Practices for Time Series Data
  - E. Counters
  - F. Lightweight Transactions (LWT)
- V. Data Modeling Labs – Group Design Session**
  - A. Multiple Use Cases from Various Domains Are Presented
  - B. Students Work in Groups to Come Up with Designs and Models
  - C. Discuss Various Designs, Analyze Decisions
  - D. Lab: Implement One of the Scenarios
- VI. Cassandra Drivers**
  - A. Introduction to Java Driver
  - B. CRUD (Create / Read / Update / Delete) Operations Using Java Client
  - C. Asynchronous Queries
  - D. Labs
- VII. Cassandra Internals**
  - A. Cassandra Design Under the Hood
  - B. SSTables, Memtables, CommitLog/Read/Write Path
  - C. Caching
  - D. Vnodes
- VIII. Administration**
  - A. Hardware Selection
  - B. Cassandra Distributions
  - C. Cassandra Best Practices (Compaction, Garbage Collection)
  - D. Troubleshooting Tools and Tips
  - E. Lab: Install Cassandra, Run Benchmarks
- IX. Bonus Lab (If Time Permits)**
  - A. Implement a Music Service like Pandora or Spotify on Cassandra