Course Outline

ProTech Professional Technical Services, Inc.

Cassandra For Developers

Course Summary

Description

Modern, large-scale applications involve dealing with Big Data, which is often larger than what traditional databases (RDBMS) can handle.

The Cassandra (C*) is a massively scalable NoSQL database that provides high availability and fault tolerance.

This hands-on course will introduce Cassandra, concepts, data modeling, and CQL (Cassandra Query Language). The focus is practical aspects of working with C* effectively. We will also cover “anti-patterns” and best practices, that will lead to optimal C* implementations in high-performance production systems.

Topics

- Introduction to Big Data / NoSQL
- Cassandra Essentials
- Data Modeling – part 1
- Data Modeling – part 2
- C* Java API
- C* Internals
- C* Admin
- C* Best Practices
- C* Case Studies
- C* Data Modeling labs
- C* Workshop (Time permitting)

Audience

This course is designed for Developers, Architects, and Database admins.

Prerequisites

- Comfortable with Java programming language
- Comfortable in Linux environment (navigating command line, running commands)

Duration

Three days

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically

Course Outline

I. Introduction to Big Data / NoSQL

A. Big Data challenges vs RDBMS
B. NoSQL overview
C. CAP theorem
D. When is NoSQL appropriate
E. Columnar storage
F. NoSQL ecosystem

II. Cassandra Essentials

A. C* architecture overview
B. C* clusters, rings, nodes
C. Keyspaces, tables, rows and columns
D. Partitioning, replication, tokens
E. Quorum and consistency levels
F. Labs: installing Cassandra, interacting with Cassandra using CQLSH

III. Data Modeling – part 1

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
I. CQL updates
J. Collections (list, map, and set)
K. Labs: various data modeling exercises using CQL; experimenting with queries and supported data types

IV. Data Modeling – part 2

A. Creating and using secondary indexes
B. Composite keys (partition keys and clustering keys)
C. Time series data
D. Best practices for time series data
E. Counters
F. Lightweight transactions (LWT)
G. Labs: creating and using indexes; modeling time series data

V. C Java API*

A. Introduction to Java driver
B. CRUD (Create / Read / Update, Delete) operations using Java client
C. Asynchronous queries
D. Labs: using Java API for Cassandra

VI. C Internals*

A. Understand Cassandra design under the hood
B. Partitioners, gossip protocols, snitches
C. sstables, memtables, commit log
D. Read path, write path
E. Deletions, compactions, tombstones
F. Failure handling
G. Caching

VII. C Admin*

A. Hardware selection
B. Software dependencies
C. Cassandra distributions
D. Lab: students install Cassandra, run benchmarks

VIII. C Best Practices*

A. C* best practices
B. Performance tuning
C. Troubleshooting tools and tips
D. “Anti-patterns” – how NOT to use C*

IX. C Case Studies*

A. We will look at some C* use cases in the industry. Study their system architecture, best practices, and recommendations. This gives attendees a good sense of how C* is being used in real-world use cases.

X. C Data Modeling labs*

A. Attendees will work as teams
B. Multiple use cases from various domains are presented
C. Students work in groups to come up with designs and models, discuss various designs, analyze decisions

XI. C Workshop (Time permitting)*

A. In this section, attendees will implement a real-world use case using C*
B. Attendees will work as teams
C. Each team will come up with data models for C* and implement them and test them
D. Also, teams are encouraged to present their solution to the class. We will discuss, provide feedback and learn from each other
E. Possible project ideas:
F. Implement a Slack-like messaging system. Come up with data models for users, messages and group chats
G. Implement a music service like Spotify. Come up with data models for songs, users, ratings
H. Implement a stock quotes tracking system. Come up with models for stock tickets, prices (time series data)
I. Browse our courses
J. We offer instructor-led courses (onsite and online)
K. Looking for team training?
L. Up-skill your team with a customized, private training
M. Public Classes
N. Suitable for small teams and individuals

Due to the nature of this material, this document refers to numerous hardware and software products by their trade names. References to other companies and their products are for informational purposes only, and all trademarks are the properties of their respective companies. It is not the intent of ProTech Professional Technical Services, Inc. to use any of these names generically