NoSQL for Developers (Cassandra)

Course Summary

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

Format: Lectures and hands on labs. (50% lecture + 50% labs). Pace of the class is determined by the students.

Topics
- Introduction to Big Data / NoSQL
- Cassandra Basics
- Data Modeling – part 1
- Data Modeling – part 2
- Data Modeling Labs: Group design session
- Cassandra drivers
- Cassandra Internals
- Administration
- Bonus Lab (time permitting)

Audience
This course was designed for developers.

Prerequisites
Before taking this course, students should have the following skills:
- Be comfortable with Java programming language
- Be comfortable in Linux environment (navigating command line, editing files with vi / nano)

Duration
Three days
NoSQL for Developers (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 and columns
    D. Partitioning, replication, tokens
    E. Quorum and consistency levels
    F. Labs : interacting with cassandra using CQLSH

III. Data Modeling – part 1
     A. Introduction to CQL
     B. CQL Datatypes
     C. Creating keyspaces & 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 / 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)

V. Data Modeling Labs : Group design session
   A. Multiple use cases from various domains are presented
   B. Students work in groups to come up designs and models
   C. Discuss various designs, analyze decisions
   D. Labs : creating and using indexes; modeling time series data

VI. Cassandra drivers
    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

VII. Cassandra Internals
     A. Understand Cassandra design under the hood
     B. Sstables, memtables, commit log
     C. Read path / write path
     D. Caching
     E. Vnodes

VIII. Administration
      A. Hardware selection
      B. Cassandra distributions
      C. Cassandra best practices (compaction, garbage collection,)
      D. troubleshooting tools and tips
      E. Lab : students install Cassandra, run benchmarks

IX. Bonus Lab (time permitting)
     A. Implement a music service like Pandora / Spotify on Cassandra

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.