Introduction to Cassandra

Course Summary

Description

This course focuses on advanced concepts of Cassandra, walking the attendees from the basics to the more advanced concepts. It addresses both the administration of Cassandra, Clustering and use of Cassandra for solving big data problems including how to use the Java API. The internals of Column Orientation and Column Architected Databases is also addressed in this course with discussions on the various storage models.

Each section introduces and explains a concept or technology, followed, as relevant and necessary, by an interactive coding section to help you understand the concept. This course combines concepts and hands-on labs in good proportion to illustrate key concepts and practical applications of these concepts.

Topics

- Column Architected Databases
- Cassandra Architecture
- Cassandra Query Language
- Advanced Querying
- Advanced Cassandra Concepts
- Cassandra Java API

Audience

This course is designed for DBAs, Java Developers and Architects who have a basic understanding of data modeling, data architecture and big data.

Prerequisites

Required
- Basic SQL experience
- Basic Java programming skills

Recommended
- Knowledge and understanding of Big Data

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

Introduction to Cassandra

Course Outline

I. Column Architected Databases
- A. Understanding Big Data
- B. SQL and Relational Database limitations
- C. Benefits of Column Orientation and Column Architected Databases

II. Cassandra Architecture
- A. Ring topology, Nodes, Virtual Nodes, Tokens and Gossip Protocol
- B. Keyspaces
- C. Data distribution, Partitioners
- D. Tokens, Partitioners, and data distribution

III. Cassandra Query Language
- A. Managing Keyspace (Create, Use, Drop, etc.)
- B. Managing Tables, Column Families
- C. Insert, Update and Delete
- D. Importing and Exporting data
- E. CQL data types
- F. Querying with CQL
- G. Batched execution

IV. Advanced Querying
- A. Partition Key and Clustering columns
- B. Order By, Clustering Order By, multiple clustering columns
- C. Filtering using CQL
- D. Composite Partition keys
- E. Collections including sets, lists and maps
- F. Secondary indexes, benefits and limitations
- G. CQL Counters
- H. Batching support in CQL
- I. CQL 1.2 vs 2.0

V. Advanced Cassandra Concepts
- A. Aggregation in Cassandra
- B. Materialization, Denormalization and other concepts
- C. Handling Schema changes in Cassandra
- D. The Cassandra Storage Model
- E. Advanced Replication concepts including Replication Factor
- F. Cross-data center Partition and Replication
- G. Additional Clustering concepts
- H. Bloom Filters role in reading data

VI. Cassandra Java API
- A. The DataStax based Driver
- B. Cluster Builder, Query Builder, Session, ResultSet, PreparedStatement, BoundStatement, etc.
- C. Async Querying

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