ProTech protechtraining.com *"Charting the Course ... ... to Your Success!"*

Apache Cassandra

Course Summary

Description

Apache Cassandra is an open-source project and a second-generation distributed NoSQL database. It is the best choice for higher availability and scalability of the database. Cassandra supports replication across multiple datacenters. It makes the write and read processes highly scalable by offering tunable consistency.

This Apache Cassandra training course offered provides an overview of the following:

- Fundamentals of big data and NoSQL databases
- Cassandra and its features
- Architecture and data model of Cassandra
- Installation, configuration and monitoring of Cassandra
- Hadoop ecosystem of products around Cassandra

Objectives

By the end of this course, participants will be able to:

- Describe the need for big data and NoSQL
- Explain the fundamental concepts of Cassandra
- Describe the architecture of Cassandra
- Demonstrate data model creation in Cassandra
- Use Cassandra database interfaces
- Demonstrate Cassandra database configuration

Topics

- Course Overview
- Introduction to big data and No-SQL Databases
- Introduction to Cassandra
- Architecture of Cassandra
- Installation and Configuration of Cassandra
- Cassandra Data Model
- Cassandra Interfaces
- Advanced Architecture and Cluster Management
- Hadoop Ecosystem around Cassandra

Audience

- Professionals aspiring for a career in NoSQL databases and Cassandra
- Analytics professionals
- Research professional
- IT developers
- Testers
- Project managers

Prerequisites

- Fundamental knowledge of any programming language
- Basic understanding of any database, SQL and query language for databases
- Working knowledge of Linux or UNIX based systems (not mandatory)

Duration

Six days

ProTech protechtraining.com *"Charting the Course ... ... to Your Success!"*

Apache Cassandra

Course Outline

I. Course Overview
- A. Target Audience
- B. Course Prerequisites
- C. Value of Professionals
- D. Lessons Covered

II. Introduction to big data and No-SQL Databases
- A. 3Vs of big data
- B. Data evolution and characteristics of big data
- C. Industry Examples
- D. Apache Hadoop - HDFS and Mapreduce
- E. Brewer's CAP principle
- F. Types of NoSQL database databases

III. Introduction to Cassandra
- A. Introducing Cassandra
- B. Cassandra Origin
- C. Main features of Cassandra
- D. Cassandra Example and Command line interface
- E. Cassandra Advantages and limitations
- F. PuTTY
- G. WinSCP
- H. Demo: Installing Ubuntu VM and connecting with Putty

IV. Architecture of Cassandra
- A. Cassandra Architecture
- B. Cassandra write process
- C. Cassandra read process
- D. Data partitions and replication
- E. Network topology and snitches
- F. Gossip protocol and seed nodes
- G. Virtual nodes and token generator
- H. Failure scenarios

V. Installation and Configuration of Cassandra
- A. Cassandra versions
- B. Preparation for installation
- C. Installation of Cassandra
- D. Configuration of Cassandra
- E. Installing on CentOS
- F. Demo: Installing and configuring Cassandra

VI. Cassandra Data Model
- A. Keyspaces, tables and columns
- B. Cassandra DDL: CREATE TABLE, ALTER TABLE, DROP TABLE
- C. Cassandra DML: INSERT, UPDATE, SELECT, DELETE
- D. SELECT statements restrictions
- E. Demo 1 - Data Definition and Data Manipulation Statements
- F. Demo 2 - Create a Table with Composite Key
- G. Demo 3 - Collection Columns in Cassandra

VII. Cassandra Interfaces
- A. Cassandra Command Line Interface
- B. CQLSH options
- C. Java interfaces
- D. Compiling a Java Program for Cassandra
- E. ODBC interface

VIII. Advanced Architecture and Cluster Management
- A. Partitions
- B. Replication strategy
- C. Consistency
- D. Monitoring and Administration tools

IX. Hadoop Ecosystem around Cassandra
- A. Apache Storm, Storm Architecture, Data Model, and Topology
- B. Apache Kafka, Kafka Architecture, Data Model
- C. Real -time analysis platform
- D. Apache Spark, Spark Architecture
- E. Spark and Scala