

Apache Cassandra Development and Administration

Course Summary

Description

The "Apache Cassandra – Development and Administration" class focuses on imparting the core concepts, architecture and design of Apache Cassandra and develops its students to become successful Cassandra professionals. The class starts with explaining how and where Cassandra fits in a business ecosystem, and how does it impact the Total Cost of Operations (TCO) of a Big Data analytical solution.

Class Highlights

- The Cassandra Architecture is discussed in detail to prepare students to fathom & follow the rest of the class.
- Data Structures are explained along with Data Partitioning and the various choices available.
- Scheme design is taught with focus on De-normalization techniques & Eventual Consistency.
- Indexing mechanisms are also explained.
- The Read and Write paths are explored & Deletes covered.
- The various Consistency levels are explained.
- Cassandra's core strengths in Availability are Failure Handling are described in great details.
- Compaction & Ring Management concepts are also explained.

From an Administration perspective, the class:

- Details the process of installing a Cassandra cluster,
- How to manage the cluster's data,
- Explains the various tools, utilities & client APIs available in the industry today, along with what to use when and why.

The Live demos showcase:

- How to setup a Cassandra cluster
- Run a 4 node Cassandra cluster,
- How to use Opscenter,
- Run Nodetool commands,
- Use the CLI and CQL shell

The hands-on labs include:

- Setting up a cluster, creating the schema & manipulating data in it.
- Dynamically extending the cluster, running a constant load on the cluster and observing the impact of nodes going down.
- Running "Repair" and monitoring the cluster using Opscenter.

At the end of this two day class, students will be ready to start engaging themselves in a productive and meaningful way with Cassandra for their professional journeys beyond.

About Apache Cassandra:

Apache Cassandra is a highly scalable, high performance and fault tolerant distributed data infrastructure. Cassandra solves both real time and analytical big data problems, from write intensive workloads to sub millisecond caching layer reads to analytical workloads involving petabytes of data using MapReduce. Offering distribution of data across multiple data centers and incremental scalability with no single points of failure, Cassandra is the logical choice when you need reliability without compromising performance.

Audience

This course is designed for Developers, Database administrators, Data Analytics professionals, Data architects, Managers.

Prerequisites

Students should have a basic understanding of database and ACID transactions

Duration

Two 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