Apache Cassandra

Description

Apache Cassandra is a free, open-source project and a second-generation distributed NoSQL database and is considered to be the best choice for high availability and scalability databases, particularly when dealing with large amounts of data. Cassandra supports replication across multiple datacenters, while also making the write and read processes highly scalable by offering tunable consistency. This Apache Cassandra training course will provide you with an overview of the fundamentals of Big Data and NoSQL databases, an understanding of Cassandra and its features, architecture and data model, its role in the Hadoop Big Data ecosystem, and show you how to install, configure and monitor Cassandra.

The large volume and variety of data that today's businesses process require the need for a highly available, low latency database. Apache Cassandra provides this solution by permitting high-speed reads and writes across a replicated, distributed system. This Apache Cassandra training course provides data modeling experience to take advantage of the linearly scalable peer-to-peer design of Cassandra.

Audience

Professionals aspiring for a career in NoSQL databases and Cassandra

- Analytics professionals
- Research professionals
- IT developers
- Testers
- Project managers

Prerequisites

- Knowledge of databases and SQL
- Java programming

Outline

Understanding NoSQL and Cassandra Basics

- Introduction to NoSQL databases
- Overview of Apache Cassandra
- Key features and advantages of Cassandra
- Understanding the CAP theorem in the context of Cassandra

Data Modeling in Cassandra

- Cassandra data model: keyspace, column family, columns, and rows
- Understanding the importance of denormalization
- Designing primary keys and composite keys
- Time-series data modeling in Cassandra

Cassandra Architecture

- Cassandra's distributed architecture
- Gossip protocol and failure detection
- Read and write paths in Cassandra
- Hinted Handoff and read repair mechanisms

Cassandra Operations

- Installation and configuration of Cassandra
- Cluster setup and configuration
- Node operations: adding, removing, and replacing nodes
- Data replication and consistency levels

Advanced Cassandra Concepts

- Cassandra Query Language (CQL) basics
- Secondary indexes in Cassandra
- Materialized views and SASI indexes
- Batch operations and atomicity in Cassandra

Best Practices and Case Studies

- Best practices for data modeling in Cassandra
- Tuning Cassandra for performance
- Backup and restore strategies
- Real-world case studies of Cassandra implementations