Updated 04/11/2024

Sign up

Cassandra training

3 days (21 hours)

Presentation

Cassandra is a distributed NoSQL database known for its ability to manage very large quantities of structured data without failure, and is used in particular through Spark in the new SMACK architectures (Spark / Mesos / Akka / Cassandra / Kafka).

This open-source database management system is highly scalable and benefits from a peer-to-peer architecture to achieve a low risk of failure.

What's more, this tool boasts extensive analytics capabilities, including integrated Solr search, batch analytics with Hadoop, and real-time analytics.

This Cassandra training course will give you the knowledge and practical experience to implement and administer Apache Cassandra on a daily basis. You'll discover how to model your data, perform monitoring tasks and optimize Cassandra administration by learning best practices.

As with all our training courses, this one will introduce you to the latest stable release (Cassandra 5).

Objectives

- Understand the architecture and operation of the Apache Cassandra NoSQL DBMS
- Installing and configuring a Cassandra cluster
- Create a database and manipulate its objects
- Discover the principles of optimization and development
- Familiarize yourself with Cassandra administration and monitoring tools

Target audience

- Developers
- Lead Developer
- Architects
- System administrators

Prerequisites

Basic knowledge of a Unix system (knowledge of how to use the terminal) and of the Java language.

Cassandra training program

Introduction to Cassandra & NOSQL (Day 1 - Morning)

- Architecture: from yesterday to today "BIG DATA" ('SQL' => 'NOSQL')
- Polygloc Persistence*. Typical architecture (past and present)
- Introduction to Cassandra (Concepts, ACID => CAP)
- Cassandra Architecture (Understanding how it works to make the most of it: commit log, data storage, services, flows)
- Distribution
- Consistency & Replication (Hint, Repair, Tombstones)

Data modeling with Cassandra : Become a super modeler! (Day 1 - Afternoon)

- The keyspace
- Big Table: Key / Value, the key is in the key :)
- Relational in NOSQL => De-normalization
- Don't be afraid of writes!
- Wide Row (Clustering columns)
- Counters
- Do you need a Transaction?
- Table options (compactions, gc graces, ttl ...)
- Best practices: what to think about (row size, partitions, clustering or not clustering columns)
- TP (Use-cases & Discussions)

Monitoring: what to monitor (Day 2 - Morning)

- Throughput, read and write requests
- Latency, read and write latency
- Disk usage, disk space on each node
- Garbage collection frequency and duration
- Errors and overruns, especially unavailable exceptions which indicate failed requests due to unavailability of nodes in the cluster
- Tools (nodetool, JMX, Datastax OpsCenter (DSE) ...

Cassandra Tuning (Day 2 - Afternoon)

- Cassandra hardware prerequisites
- cassandra.yaml (properties)
- Major properties (tuning)
- JVM options (tuning)
- Use-cases & Discussions

Best Practices (Day 3 - Morning)

- Java Driver (description, configuration, instantiation, etc.)
- Writes : best practises
- Reads : best practises
- Error handling (exception)
- TP
- Use-cases & Discussions

Administration (Day 3 - Afternoon)

- Object Mapping
- Integration testing (Cassandra embedded)
- TP
- Cassandra cluster administration (overview): possibilities, dos and don'ts
- Feedback, to go further (combining other technologies?)

Further information

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.