

Apache Cassandra

Course Offer 2013



Geroba Data Technologies e.U.

DI Gernot R. Bauer +43 650 26 44 013

DI Philipp Potisk +43 660 46 02 632

bigdata@geroba.com

Kaiserfeldgasse 17/25

8010 Graz

What is Apache Cassandra?

The amount of data that must be processed by applications has multiplied in the past few years. Experience shows that traditional database systems such as Oracle or Microsoft SQL Server have problems scaling to large data sizes and require expensive solutions. At the same time, applications need to provide good performance even with large amounts of data.

Apache Cassandra is a representative of the so-called NoSQL database family. This type of novel database systems are characterized by the fact that they are designed for large amounts of data, high availability, good performance and simple, dynamic and cost-effective scalability. This allows applications to process large amounts of data without having to resort to expensive and highly specific server hardware.

Apache Cassandra is a successful and often used representative of NoSQL databases. Cassandra was developed by Facebook and combines ideas from Google's "Bigtable" database and Amazon's Dynamo database. Cassandra has proved to be a flexible database that can be operated with little effort as a cluster. This permits the creation of highly available applications that are still fulfilling their job if individual servers or even entire data centers break. Under load, Cassandra reaches read and especially write speeds that can be difficult (if not impossible) to achieve with traditional database systems.

Apache Cassandra is a field-proven database system and is used successfully by companies such as Twitter, eBay, Cisco or Walmart.

Course Offer

Overview

Our two-part course provides a mix of theory, hands-on examples and use cases from practice to system architects, developers, database administrators and IT managers. The course gives the trainees insights into the architecture and theory behind Cassandra and the advantages and disadvantages compared to traditional database solutions. With hands-on examples, trainees can implement and consolidate their newly acquired knowledge. Use cases and practical experience will complete the training and will give participants the resources to successfully carry out projects with Cassandra.

Target Audience

The courses are aimed at software architects, software developers, database administrators and IT managers.

Trainers

Since finishing their studies of software engineering and business economics at Graz University of technology, Gernot R. Bauer and Philipp Potisk work as a software architects and Java EE developers. They have practical experience with Apache Cassandra in several projects since 2010. They are both active members in the Cassandra community and always interested in the progress of the open source project.

Beginners' Course

The focus of the first part of the course are the basic concepts of NoSQL databases and their properties, especially in comparison to traditional SQL database systems. Building on that, the course gives an overview of the main elements of Cassandra: the data model, partitioning and replication mechanisms and consistency levels of Cassandra with its ring architecture. Furthermore, real-life use cases will be presented.

In workshop format, the trainees will perform installation and configuration of Cassandra and will perform basic tasks to interact with the database. In addition, several client frameworks are developed using practical examples. The main focus lies in Java client frameworks which are used for simple implementation tasks. This also highlights the strength and weaknesses of each of those frameworks.

The objective of this basic section is to provide developers and operational staff the mindset that is behind NoSQL stores and especially behind Cassandra. This happens especially in comparison to established SQL databases. The Hands-On allows developers to test the theoretical principles with practical examples extensively.

Agenda

- Introduction to BigData and NoSQL databases
- Properties of Cassandra and real-life use cases
- Cassandra data model
- Hands-on: installation, configuration, command line-interface
- Cassandra system architecture
- Overview of client frameworks
- Java examples with thrift, hector, astyanax, kundera

Duration

1 day (about 8 hours including breaks)

Documentation

Trainees will receive the presentation along with the source code of the training examples.

"They have the hands-on experience to answer real-life questions. The course features well prepared code snippets & examples and an exciting experiment."

(Trainee feedback)

Interested in our courses?

If you would like to inquire about course details or want to discuss a tailored course agenda, please do contact us.

Geroba Data Technologies e.U.

DI Gernot Bauer

+43 650 2644013

bigdata@geroba.at

Kaiserfeldgasse 17/25

8010 Graz

Advanced Course

The advanced course covers various aspects of Cassandra in-depth. Some of these aspects are the processing steps of read and write operations (read/write path) and related mechanisms such as: tombstones, compaction, read repair, hinted handoff, etc.

Regarding data modelling, the course presents best-practices and typical modelling patterns. The course features several selected case studies from real-world examples. In group discussions, the trainees will discuss modeling approaches and their advantages and disadvantages together with the trainers.

The first afternoon session covers the new application interface to Cassandra (CQL 3.0) alongside with training examples. Furthermore, the trainees will develop a simple Java application using the CQL Driver of DataStax. This example can be tested on a live cluster of 4 nodes provided by us.

In addition, the course presents several tools for administrations and operations of of Cassandra clusters. Furthermore tuning tips, patterns and anti-patterns are discussed.

The aim of this course is to give developers and operational staff a deeper insight into Cassandra and to cover individual aspects in detail. With the help of use cases from practice, "Thinking in Cassandra" mindset from the beginners' course will be used to solve common problems of software application engineering.

Agenda

- Use Case eBay
- Cassandra internals (read and write paths, ...)
- Operating tools (nodetool, JMX, OpsCenter) with live demo in a cluster
- Hands-on introduction to Cassandra Query Language 3.0
- Implementation of an example Java application and deployment on live cluster
- Tuning, patterns, anti-patterns
- Optional: customer-specific use cases
- Optional: Cassandra & Hadoop, Cassandra & Solr

Duration

1 day (about 8 hours including breaks)

Documentation

Trainees will receive the presentation along with the source code of the training examples.

"The guys' style is fresh combined with deep technical knowledge."
(Trainee feedback)

Interested in our courses?

If you would like to inquire about course details or want to discuss a tailored course agenda, please do contact us.

Geroba Data Technologies e.U.
DI Gernot Bauer
+43 650 2644013
bigdata@geroba.at
Kaiserfeldgasse 17/25
8010 Graz