Cloudera Developer Training for Spark & Hadoop (DSH)

ID DSH **Preis** auf Anfrage **Dauer** 4 Tage

Voraussetzungen

This course is designed for developers and engineers who have programming experience, but prior knowledge of Hadoop and/or Spark is not required.

- Apache Spark examples and hands-on exercises are presented in Scala and Python. The ability to program in one of those languages is required.
- Basic familiarity with the Linux command line is assumed.
- Basic knowledge of SQL is helpful

Kursziele

Hands-on exercises take place on a live cluster, running in the cloud. A private cluster will be built for each student to use during the class.

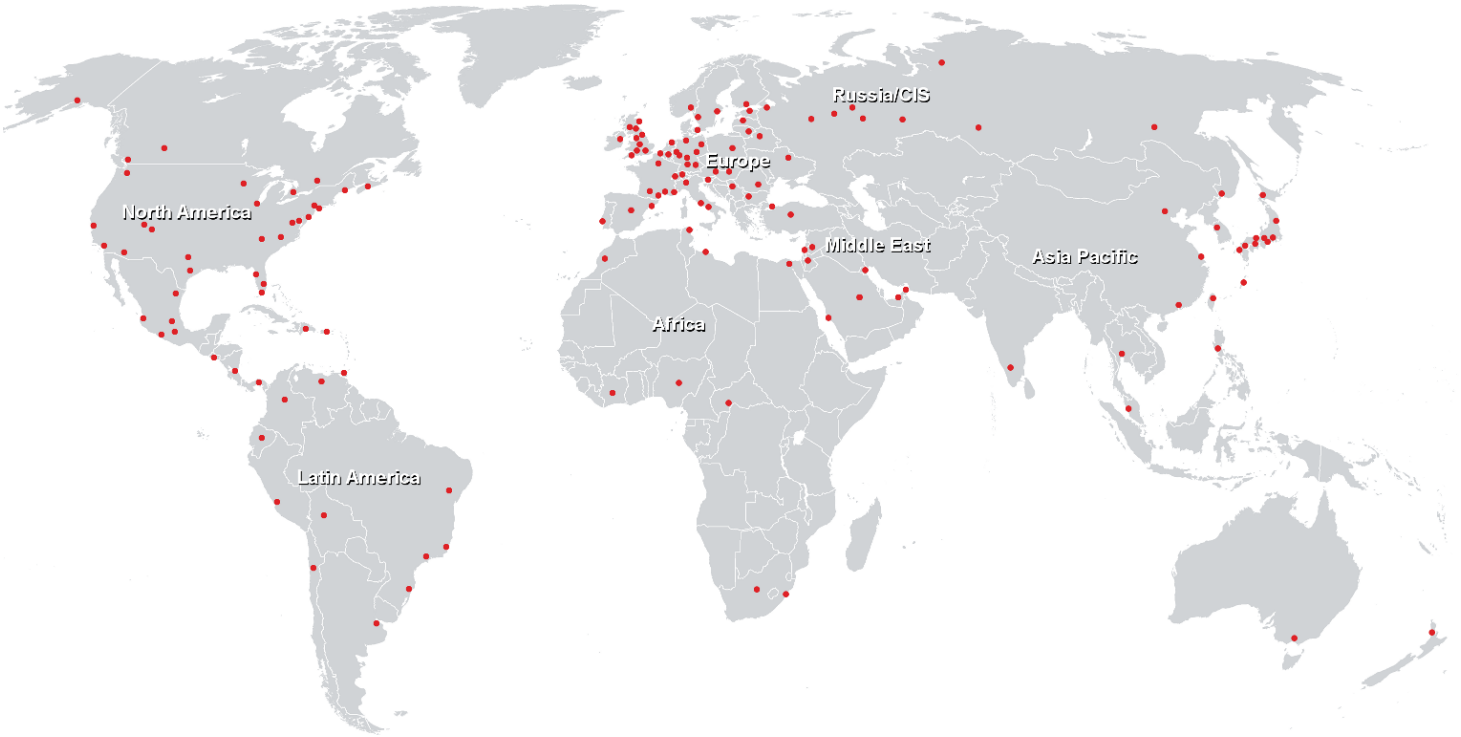
Through instructor-led discussion and interactive, hands-on exercises, participants will navigate the Hadoop ecosystem, learning how to::

- Distribute, store, and process data in a Hadoop cluster
- Write, configure, and deploy Spark applications on a cluster
- Use the Spark shell for interactive data analysis
- Process and query structured data using Spark SQL
- Use Spark Streaming to process a live data stream

Kursinhalt

- Introduction to Apache Hadoop and the Hadoop Ecosystem
- Apache Hadoop File Storage
- Distributed Processing on an Apache Hadoop Cluster
- Apache Spark Basics
- Working with DataFrames and Schemas
- Analyzing Data with DataFrame Queries
- RDD Overview
- Transforming Data with RDDs
- Aggregating Data with Pair RDDs
- Querying Tables and Views with Apache Spark SQL
- Working with Datasets in Scala
- Writing, Configuring, and Running Apache Spark Applications
- Distributed Processing
- Distributed Data Persistence
- Common Patterns in Apache Spark Data Processing
- Apache Spark Streaming: Introduction to DStreams
- Apache Spark Streaming: Processing Multiple Batches
- Apache Spark Streaming: Data Sources

Weltweite Trainingscenter



Fast Lane Institute for Knowledge Transfer GmbH

Husacherstrasse 3
CH-8304 Wallisellen
Tel. +41 44 832 50 80

info@flane.ch, https://www.flane.ch