Preparing with Cloudera Data Engineering (PCDE)

ID PCDE **Preis** US$ 5'250.– (exkl. MwSt.) **Dauer** 4 Tage

Zielgruppe

Dieser Kurs ist für Entwickler und Dateningenieure gedacht. Von allen Teilnehmern wird erwartet, dass sie über grundlegende Linux-Erfahrungen und Grundkenntnisse in den Programmiersprachen Python oder Scala verfügen.

Voraussetzungen

Grundkenntnisse in SQL sind hilfreich. Vorkenntnisse in Spark und Hadoop sind nicht erforderlich.

Kursziele

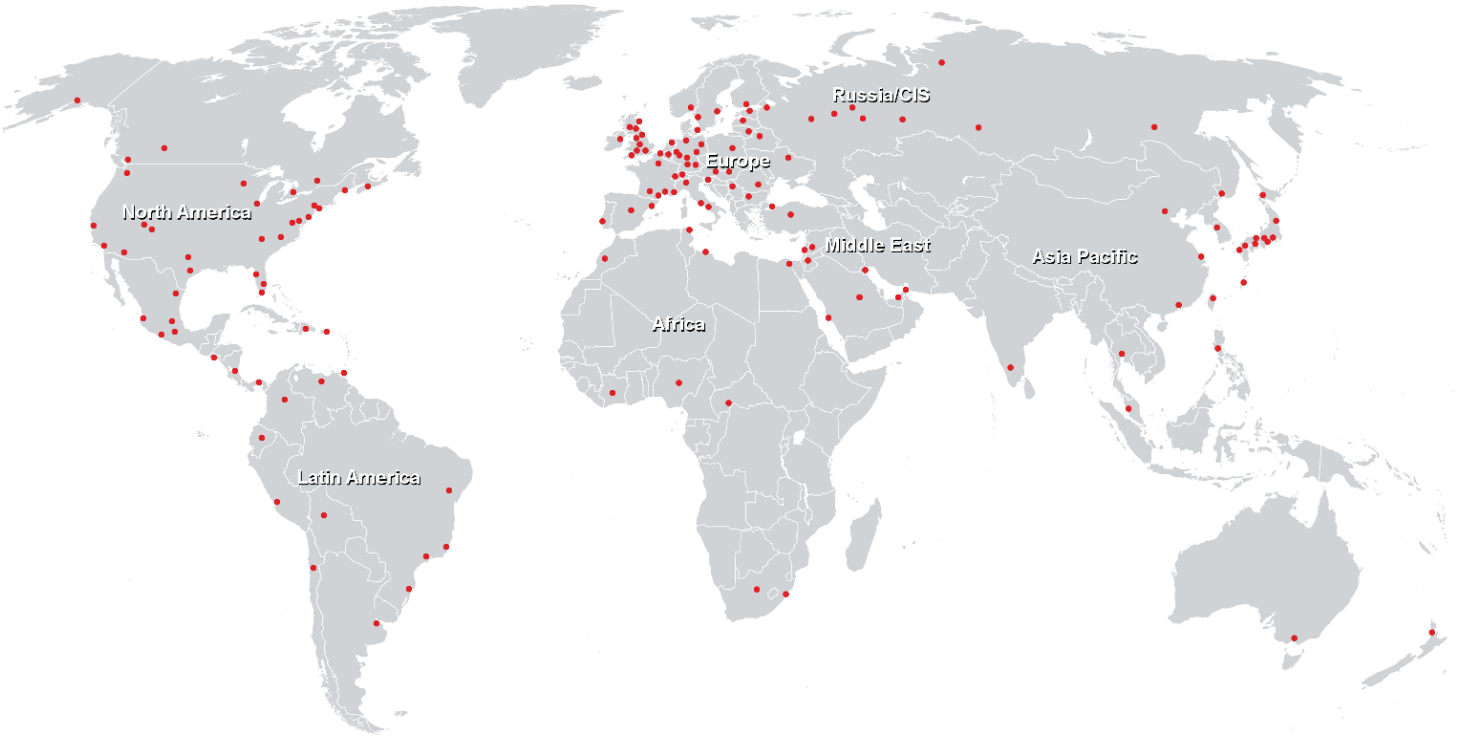
In diesem Kurs werden Sie lernen, wie man:

- Verteilen, Speichern und Verarbeiten von Daten in einem CDP-Cluster
- Schreiben, Konfigurieren und Bereitstellen von Apache Spark-Anwendungen
- Verwenden Sie die Spark-Interpreter und Spark-Anwendungen, um verteilte Daten zu untersuchen, zu verarbeiten und zu analysieren
- Abfrage von Daten mit Spark SQL, DataFrames und Hive-Tabellen
- Bereitstellen einer Spark-Anwendung auf dem Data Engineering Service

Kursinhalt

- HDFS-Einführung
- YARN-Einführung
- Arbeiten mit RDDs
- Arbeiten mit DataFrames
- Einführung in Apache Hive
- Arbeiten mit Apache Hive
- Integration von Hive und Spark
- Herausforderungen bei der verteilten Verarbeitung
- Verteilte Spark-Verarbeitung
- Verteilte Persistenz von Spark
- Daten-Engineering-Dienst
- Arbeitsbelastung XM
- Anhang: Arbeiten mit Datensätzen in Scala

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