

Preparing with Cloudera Data Engineering

powered by Apache Spark, Hive, and Airflow

Dieser praxisorientierte Kurs vermittelt die wichtigsten Konzepte und Kenntnisse, die Entwickler benötigen, um mit Apache Spark hochleistungsfähige, parallele Anwendungen auf der Cloudera Data Platform (CDP) zu entwickeln.

Praktische Übungen ermöglichen, das Schreiben von Spark-Anwendungen zu üben, die mit CDP-Kernkomponenten integriert werden. Sie lernen die Verwendung von Spark SQL zur Abfrage strukturierter Daten, die Verwendung von Hive-Funktionen zum Einlesen und Denormalisieren von Daten und die Arbeit mit "Big Data", die in einem verteilten Dateisystem gespeichert sind.

Nach diesem Kurs sind Sie in der Lage, sich realen Herausforderungen zu stellen. Sie können Anwendungen erstellen, um schnellere und bessere Entscheidungen zu treffen. Weiterhin können Sie Interaktive Analysen durchführen, die auf eine Vielzahl von Anwendungsfällen, Architekturen und Branchen angewendet werden.

Kursinhalt

- HDFS Introduction
- YARN Introduction
- Working with RDDs
- Working with DataFrames
- Introduction to Apache Hive
- Working with Apache Hive
- Hive and Spark Integration
- Distributed Processing Challenges
- Spark Distributed Processing
- Spark Distributed Persistence
- Data Engineering Service
- Workload XM
- Appendix: Working with Datasets in Scala

E-Book Sie erhalten die englischen Original-Unterlagen als E-Book (pdf).

Zielgruppe

Dieser Kurs ist für Entwickler und Dateningenieure gedacht.

Voraussetzungen

Es wird erwartet, dass Sie über grundlegende Linux-Kenntnisse und Grundkenntnisse der Programmiersprachen Python oder Scala verfügen. Grundkenntnisse in SQL sind hilfreich. Vorkenntnisse in Spark und Hadoop sind nicht erforderlich.

Hierzu empfehlen wir Ihnen auch unsere Trainings im Bereich Programmiersprachen und Softwareentwicklung sowie Linux.

Kursziel

- Verteilen, Speichern und Verarbeiten von Daten in einem CDP-Cluster
- Schreiben, Konfigurieren und Bereitstellen von Apache Spark-Anwendungen
- Verwendung der Spark-Interpreter und Spark-Anwendungen zur Erkundung, Verarbeiten und Analysieren von verteilten Daten
- Abfragen von Daten mit Spark SQL, DataFrames und Hive-Tabellen
- Bereitstellen einer Spark-Anwendung auf dem Data Engineering Service

Dieser Kurs im Web



Alle tagesaktuellen Informationen und Möglichkeiten zur Bestellung finden Sie unter dem folgenden Link: www.experteach.de/go/CLSP

Vormerkung

Sie können auf unserer Website einen Platz kostenlos und unverbindlich für 7 Tage reservieren. Dies geht auch telefonisch unter 06074 4868-0.

Garantierte Kurstermine

Für Ihre Planungssicherheit bieten wir stets eine große Auswahl garantierter Kurstermine an.

Ihr Kurs maßgeschneidert

Diesen Kurs können wir für Ihr Projekt exakt an Ihre Anforderungen anpassen.

Training		Preise zzgl. MwSt.	
Termine in Deutschland	4 Tage	€ 3.495,-	
Online Training	4 Tage	€ 3.495,-	
Termin/Kursort	<input type="checkbox"/> Online	Kurssprache Englisch	<input type="checkbox"/> Online
20.05.-23.05.24	<input type="checkbox"/> Online	05.11.-08.11.24	<input type="checkbox"/> Online
01.10.-04.10.24	<input type="checkbox"/> Online		

Stand 24.04.2024



Unser Trainingsangebot für Sie:



Classroom Training

Das Live-Trainingserlebnis in unseren Training Centern oder bei Ihnen vor Ort.



Online Training

Nehmen Sie online am Kurs teil – ohne Reise- und Hotelaufwände.



Hybrid Training

Classroom & online in einem Kurs – Sie wählen, wie Sie teilnehmen möchten.



Inhouse-Schulungen

Für Ihr Projekt erstellen wir genau passende Trainingskonzepte.



Garantierte Kurstermine

Die ExperTeach Garantietermine geben Ihnen Sicherheit für Ihre Planung.

Auszeichnungen für ExperTeach



ExperTeach Gesellschaft für Netzwerkkompetenz mbH

Waldstr. 94 · 63128 Dietzenbach · Telefon: +49 6074 4868-0 · Fax: +49 6074 4868-109 · info@exper teach.de · www.exper teach.de