

Apache Spark Application Performance Tuning

Dieser Kurs vermittelt die wichtigsten Konzepte und Fachkenntnisse, die Entwickler benötigen, um die Leistung ihrer Apache Spark-Anwendungen zu verbessern. Während des Kurses lernen die Teilnehmer, wie sie häufige Ursachen für schlechte Performance in Spark-Anwendungen identifizieren, Techniken zu deren Vermeidung sowie Best Practices für die Überwachung von Spark-Anwendungen finden.

Der Kurs stellt die Architektur und Konzepte von Apache Spark und der zugrunde liegenden Datenplattform vor und baut dann auf diesem grundlegenden Verständnis auf, indem die Schüler lernen, wie der Spark-Anwendungscode optimiert wird.

Der Schwerpunkt des Kurses liegt auf Demonstrationen unter Anleitung des Kursleiters, die sowohl Leistungsprobleme als auch die entsprechenden Techniken veranschaulichen, gefolgt von praktischen Übungen, die den Teilnehmern die Möglichkeit geben, das Gelernte in einer interaktiven Notebook-Umgebung zu üben.

Der Kurs gilt für Spark 2.4, führt aber auch in das Spark 3.0 Adaptive Query Execution Framework ein.

Kursinhalt

- Spark Architecture
- Data Sources and Formats
- Inferring Schemas
- Dealing With Skewed Data
- Catalyst and Tungsten Overview
- Mitigating Spark Shuffles
- Partitioned and Bucketed Tables
- Improving Join Performance
- Pyspark Overhead and UDFs
- Caching Data for Reuse
- Workload XM (WXM) Introduction
- What's New in Spark 3.0?

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

Zielgruppe

Dieser Kurs richtet sich an Softwareentwickler, Ingenieure und Datenwissenschaftler, die Erfahrung mit der Entwicklung von Spark-Anwendungen haben und lernen möchten, wie sie die Leistung ihres Codes verbessern können. Dies ist keine Einführung in Spark.

Voraussetzungen

Spark-Beispiele und praktische Übungen werden in Python präsentiert und die Fähigkeit, in dieser Sprache zu programmieren, ist erforderlich. Grundlegende Vertrautheit mit der Linux-Befehlszeile wird vorausgesetzt. Grundlegende Kenntnisse von SQL sind hilfreich.

Kursziel

Wenn Sie diesen Kurs erfolgreich abgeschlossen haben, werden Sie in der Lage sein:

- Die Architektur von Apache Spark, die Job-Ausführung und wie Techniken wie Lazy Execution und Pipelining die Laufzeitleistung verbessern können, zu verstehen,
- Die Leistungsmerkmale von Kerndatenstrukturen wie RDD und DataFrames zu bewerten,
- Dateiformate auszuwählen, die die beste Leistung für Ihre Anwendung bieten,
- Leistungsprobleme zu identifizieren und zu lösen, die durch Datenschieflage verursacht werden,
- Partitionierungs-, Bucketing- und Join-Optimierungen zur Verbesserung der SparkSQL-Leistung zu verwenden,
- den Leistungs-Overhead von Python-basierten RDDs, DataFrames und benutzerdefinierten Funktionen zu verstehen,
- Die Vorteile des Caching für eine bessere Anwendungsleistung zu nutzen,
- zu verstehen, wie die Optimierer Catalyst und Tungsten arbeiten,
- zu verstehen, wie Workload XM bei der Fehlerbehebung und proaktiven Überwachung der Leistung von Spark-Anwendungen helfen kann,
- die neuen Funktionen in Spark 3.0 zu kennen und insbesondere, wie die Adaptive Query Execution Engine die Leistung verbessert.

Stand 07.05.2025

Dieser Kurs im Web



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

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 Österreich	3 Tage	€ 2.695,-
Online Training	3 Tage	€ 2.695,-
Termine auf Anfrage		

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 Training & Consulting GmbH

Handelskai 94-96 · 1200 Wien · Telefon: +43 1 2350 383-0 · Fax: +43 1 2350 383-19 · info@experteach.at · www.experteach.at