CLOUDERA Training Partner

# Cloudera Data Scientist Training

**Dieser Workshop befasst sich mit Enterprise Data Science und maschinenunterstütztem Lernen mit Apache Spark in Cloudera Data Science Workbench (CDSW). Die Teilnehmer verwenden Spark SQL, um Daten zu laden, zu untersuchen, zu bereinigen, zu verbinden und zu analysieren. Weiterhin wird Spark MLlib genutzt, um Pipelines für maschinenunterstütztes Lernen zu spezifizieren, zu trainieren, zu bewerten, abzustimmen und einzusetzen. Sie tauchen in die Grundlagen der Spark-Architektur und des Ausführungsmodells ein, die notwendig sind um ihre Spark-Anwendungen effektiv zu konfigurieren, zu überwachen und abzustimmen. Die Teilnehmer lernen auch, wie Spark mit Schlüsselkomponenten der Cloudera-Plattform wie HDFS, YARN, Hive, Impala und Hue sowie mit ihren bevorzugten Python- oder R-Paketen integriert wird.**

**Die Lerninhalte werden in Form einer Reihe von kurzen Vorlesungen, interaktiven Vorführungen, umfangreichen praktischen Übungen und Diskussionen vermittelt. Die Vorführungen und Übungen zu Apache Spark werden mit Python (mit PySpark) und R (mit sparklyr) in der Cloudera Data Science Workbench (CDSW) durchgeführt.**

**Kursinhalt**
- Data Science Overview
- Cloudera Data Science Workbench (CDSW)
- Science Workbench
- Workbench Works
- Workbench
- Case Study
- Apache Spark
- Summarizing and Grouping DataFrames
- Window Functions
- Exploring DataFrames
- Apache Spark Job Execution
- Processing Text and Training and Evaluating Topic Models
- Training and Evaluating Recommender Models
- Running a Spark Application from (CDSW)
- Columns of a DataFrame
- Inspecting a Spark SQL DataFrame
- Transforming DataFrames
- Monitoring, Tuning, and Configuring Spark Applications
- Machine Learning Overview
- Training and Evaluating Regression Models
- Working with Machine Learning Pipelines
- Deploying Machine Learning Pipelines
- Transforming DataFrame Columns
- Complex Types
- User-Defined Functions
- Reading and Writing Data
- Combining and Splitting DataFrames
- Training and Evaluating Classification Models
- Tuning Algorithm Hyperparameters Using Grid Search
- Training and Evaluating Clustering Models
- Overview of sparklyr
- Introduction to Additional CDSW Features

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

**Zielgruppe**
Der Workshop wird für Data Scientists angeboten, die Python oder R einsetzen, um kleinere Datensätze auf einem einzigen Endgerät zu bearbeiten und die ihre Analysen und Modelle für maschinenunterstütztes Lernen nach oben skalieren müssen, um mit großen Datensätzen und verteilten Clustern arbeiten zu können. Mitarbeiter, die für Datenverarbeitung verantwortlich sind, sowie Entwickler mit Grundkenntnissen in Data Science und maschinenunterstütztem Lernen profitieren ebenfalls von diesem Workshop.

**Voraussetzungen**
Die Teilnehmer sollten Grundkenntnisse zu Python oder R sowie zu Erarbeitung und Analyse von Daten und der Entwicklung statistischer und maschinenunterstützter Lernmodelle haben. Kenntnisse zu Hadoop oder Spark werden nicht vorausgesetzt.

**Dieser Kurs im Web**
Alle tagesaktuellen Informationen und Möglichkeiten zur Bestellung finden Sie unter dem folgenden Link: www.experteach.de/go/**CLDS**

**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,-** |
| Termine auf Anfrage | | |

Stand 23.04.2024

Cloudera Data Scientist Training

Cloudera

ExperTeach

# 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@experteach.de · www.experteach.de