

Big Data

Einblick in Hadoop und andere Frameworks

Das Thema Big Data ist lange schon den Kinderschuhen entwachsen. Erfahrungen und Informationen sind längst das Kapital vieler Unternehmen und die Analyse und Strukturierung riesiger Datenmengen daher unternehmenskritisch. Wer Trends und Zusammenhänge schneller erkennt als die Marktbegleiter hat einen entscheidenden Wettbewerbsvorteil. Big-Data-Lösungen sprießen daher wie Pilze aus dem Boden. Dieser Kurs vermittelt, was sich hinter diesem Hype versteckt, welche Technologien zum Einsatz kommen und wie diese arbeiten.

Kursinhalt

- Was versteckt sich hinter Big Data?
- Anwendungsbeispiele für Big Data
- Speicherung großer Datenmengen in verteilten Dateisystemen
- MapReduce-Verfahren und CAP-Theorem
- NoSQL-Datenbanken
- Software-Lösungen für Big Data: Hadoop, Spark und Flink
- Data Analytics
- IT-Architekturen für Big Data
- Big Data und Data Analytics Praxis-Demo
- Blick in die Zukunft

E-Book Sie erhalten das ausführliche deutschsprachige Unterlagenpaket aus der Reihe ExperTeach Networking – Print, E-Book und personalisiertes PDF! Bei Online-Teilnahme erhalten Sie das E-Book sowie das personalisierte PDF.

Zielgruppe

Der Kurs richtet sich an all diejenigen, die Big-Data-Lösungen planen, bewerten oder implementieren möchten.

Voraussetzungen

Es werden keine speziellen Vorkenntnisse vorausgesetzt. Wer Interesse an Big-Data-Lösungen in Bezug auf Anwendungsszenarien sowie technische Realisierung hat, wird von diesem Kurs profitieren.

Dieser Kurs im Web



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

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	3 Tage	€ 2.195,-
Online Training	3 Tage	€ 2.195,-
Termin/Kursort	Kurssprache Deutsch	
28.10.-30.10.24	Frankfurt	28.10.-30.10.24
		Online



Inhaltsverzeichnis

Big Data – Einblick in Hadoop und andere Frameworks

- 1 Was ist Big Data?**
 - 1.1 Der große Berg Daten
 - 1.2 Anwendungsgebiete von BIG DATA
 - 1.3 Die Definition von Big Data: 3-5 „V“s
 - 1.3.1 Volume
 - 1.3.2 Velocity
 - 1.3.3 Variety
 - 1.3.4 Das vierte V – Veracity
 - 1.3.5 Das ist Big Data
 - 1.4 Der Ursprung von Big Data
 - 2 Big Data Grundlagen**
 - 2.1 Die BIG DATA Value Chain
 - 2.2 Quellen für BigData Datenanalysen
 - 2.3 Die Architektur
 - 2.4 SQL: feste, vordefinierte Tabellenschemata
 - 2.5 Normalisierung von Tabellen
 - 2.6 NoSQL
 - 2.6.1 Key-Value Stores
 - 2.6.2 In-Memory Key-Value Stores
 - 2.6.3 Document Stores
 - 2.6.4 Graph Databases
 - 2.6.5 Column Stores
 - 2.7 CAP-Theorem
 - 2.7.1 Kombination CA des CAP-Theorems
 - 2.7.2 Kombination CP des CAP-Theorems
 - 2.7.3 Kombination AP des CAP-Theorems
 - 3 Hadoop und Spark**
 - 3.1 Hadoop
 - 3.2 MapReduce
 - 3.2.1 Hauptkonzepte – MapReduce
 - 3.2.2 MapReduce – Datenfluss
 - 3.2.3 Beispiel: Wörter zählen
 - 3.2.4 Map Reduce – Hinweise
 - 3.3 HDFS
 - 3.3.1 HDFS – Hauptkomponenten
 - 3.3.2 HDFS – Architektur
 - 3.4 YARN
 - 3.5 Apache Spark
 - 3.5.1 Resilient Distributed Dataset
 - 3.5.2 Spark SQL
 - 3.5.3 Spark Streaming
 - 3.5.4 MLlib
 - 3.5.5 Machine Learning
 - 3.5.6 GraphX
 - 4 Big Data Technologien**
 - 4.1 Das Hadoop Ecosystem
 - 4.2 Pig
 - 4.3 Hive
 - 4.4 Mahout
 - 4.5 HBase
 - 4.6 Sqoop
 - 4.7 Flume
 - 4.8 Chukwa
 - 4.9 Flink
 - 4.10 Oozie
 - 4.11 Zookeeper
 - 4.12 Ambari
 - 4.13 R Connector
 - 4.14 Cassandra
 - 4.15 SAP HANA
 - 5 Anwendungsbeispiele für Big Data**
 - 5.1 Limitierungen klassischer analytischer Anwendungen
 - 5.2 Einsatzszenarien von Big Data
 - 5.2.1 Clickstream Analyse
 - 5.2.2 Stimmungsanalysen aus Social Media
 - 5.2.3 Analyse von Log-Daten
 - 5.2.4 Analyse von Sensordaten
 - 5.2.5 Analyse von Texten
 - 5.2.6 Analyse von Video- und Sprachdaten
 - 5.2.7 Business Intelligence (BI) und Big Data
 - 5.2.8 Hybrid-Lösung mit einem Data Warehouse
 - 5.3 Fazit
 - 6 Data Governance + Risiken**
 - 6.1 Die 3 Säulen von Data Governance
 - 6.2 Was kann ich tun, um meine Daten zu schützen?
 - 6.3 Risiken
 - 6.4 Data Compliance Risiko
 - 6.4.1 National und in Europa
 - 6.4.2 International
 - 6.4.3 Sozialrisiko
 - 6.5 Datenrisiken
 - 6.5.1 Sicherheit der Daten
 - 6.5.2 Qualität der Daten
 - 6.6 Definitions- und Aussagerisiko
 - 6.7 Risikovermeidung
 - 6.7.1 Faktor Daten
 - 6.7.2 Faktor Data Management
 - 6.7.3 Faktor Organisation
 - 6.7.4 Faktor Prozess
 - 6.7.5 Faktor Kunde als Betroffener
 - 6.8 Herausforderungen
- 7 Herausforderungen im Betrieb von Big Data Lösungen**
 - 7.1 Wo ist der Anfang?
 - 7.2 Hadoop unternehmensweit betreiben
 - 7.2.1 Physische Infrastruktur
 - 7.2.2 Datenhaltung
 - 7.2.3 Datenzugriff
 - 7.2.4 Datenintegration
 - 7.2.5 IT-Sicherheit
 - 7.2.6 Weitere Betriebskriterien
 - 7.2.7 Wirtschaftliche Kriterien
 - 7.3 Echtzeitanalysen für Streaming-Daten
- 8 Ausblick**
 - 8.1 Aktueller Stand
 - 8.2 Technische Entwicklungen
 - 8.3 Markt-Entwicklungen
 - 8.4 Business-Entwicklungen
 - 8.5 Diskussion der Ergebnisse
- 9 Hadoop Installation & Configuration & Go!**
 - 9.1 Installationsschema für APACHE HADOOP 3.1.3
 - 9.2 Hadoop 3.1.3 auf Github
 - 9.3 Die Expertech Labor-Umgebung
 - 9.4 Anpassung der Konfigurationsdateien
 - 9.5 Übersicht der Filestrukturen im Labor
 - 9.6 Erster Start des HDFS
 - 9.7 Syntax und Ablauf von Zählaufgaben
 - 9.8 Output beim MAPREDUCE Vorgang
 - 9.9 Hadoop Cockpit
 - 9.10 Wordcount Abfrage via PIG:
 - 9.11 RATING - Datensätze filtern (25 Mio)

