

# 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](http://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.

Stand 19.07.2024

Training		Preise zzgl. MwSt.	
<b>Termine in Deutschland</b>	<b>3 Tage</b>	<b>€ 2.195,-</b>	
<b>Online Training</b>	<b>3 Tage</b>	<b>€ 2.195,-</b>	
<b>Termin/Kursort</b>	Kursprache Deutsch		
28.10.-30.10.24	Frankfurt	28.04.-30.04.25	Frankfurt
28.10.-30.10.24	Online	28.04.-30.04.25	Online



# Inhaltsverzeichnis

## Big Data – Einblick in Hadoop und andere Frameworks

<b>1 Was ist Big Data?</b>	<b>4 Big Data Technologien</b>	<b>6.5.1</b> Sicherheit der Daten
<b>1.1</b> Der große Berg Daten	<b>4.1</b> Das Hadoop Ecosystem	<b>6.5.2</b> Qualität der Daten
<b>1.2</b> Anwendungsgebiete von BIG DATA	<b>4.2</b> Pig	<b>6.6</b> Definitions- und Aussagerisiko
<b>1.3</b> Die Definition von Big Data: 3-5 „V“s	<b>4.3</b> Hive	<b>6.7</b> Risikovermeidung
<b>1.3.1</b> Volume	<b>4.4</b> Mahout	<b>6.7.1</b> Faktor Daten
<b>1.3.2</b> Velocity: Geschwindigkeit und Aktualität der Daten	<b>4.5</b> HBase	<b>6.7.2</b> Faktor Data Management
<b>1.3.3</b> Variety: Vielfältigkeit der Daten	<b>4.6</b> Sqoop	<b>6.7.3</b> Faktor Organisation
<b>1.3.4</b> Veracity: Gültigkeit und Zuverlässigkeit der Daten	<b>4.7</b> Flume	<b>6.7.4</b> Faktor Prozess
<b>1.3.5</b> Das ist Big Data	<b>4.8</b> Chukwa	<b>6.7.5</b> Faktor Kunde als Betroffener
<b>1.4</b> Der Ursprung von Big Data	<b>4.9</b> Flink	<b>6.8</b> Herausforderungen
<b>2 Big Data Grundlagen</b>	<b>4.10</b> Oozie	<b>7 Herausforderungen im Betrieb von Big Data Lösungen</b>
<b>2.1</b> Die BIG DATA Value Chain	<b>4.11</b> Zookeeper	<b>7.1</b> Wo ist der Anfang?
<b>2.2</b> Quellen für BigData Datenanalysen	<b>4.12</b> Ambari	<b>7.2</b> Hadoop unternehmensweit betreiben
<b>2.3</b> Die Architektur	<b>4.13</b> R Connector	<b>7.2.1</b> Physische Infrastruktur
<b>2.4</b> SQL: feste, vordefinierte Tabellenschemata	<b>4.14</b> Cassandra	<b>7.2.2</b> Datenhaltung
<b>2.5</b> Normalisierung von Tabellen	<b>4.15</b> SAP HANA	<b>7.2.3</b> Datenzugriff
<b>2.6</b> NoSQL	<b>5 Anwendungsbeispiele für Big Data</b>	<b>7.2.4</b> Datenintegration
<b>2.6.1</b> Key-Value Stores	<b>5.1</b> Limitierungen klassischer analytischer Anwendungen	<b>7.2.5</b> IT-Sicherheit
<b>2.6.2</b> In-Memory Key-Value Stores	<b>5.2</b> Einsatzszenarien von Big Data	<b>7.2.6</b> Weitere Betriebskriterien
<b>2.6.3</b> Document Stores	<b>5.2.1</b> Clickstream Analyse	<b>7.2.7</b> Wirtschaftliche Kriterien
<b>2.6.4</b> Graph Databases	<b>5.2.2</b> Stimmungsanalysen aus Social Media	<b>7.3</b> Echtzeitanalysen für Streaming-Daten
<b>2.6.5</b> Column Stores	<b>5.2.3</b> Analyse von Log-Daten	<b>8 Ausblick</b>
<b>2.7</b> CAP-Theorem	<b>5.2.4</b> Analyse von Sensordaten	<b>8.1</b> Aktueller Stand
<b>2.7.1</b> Kombination CA des CAP-Theorems	<b>5.2.5</b> Analyse von Texten	<b>8.2</b> Technische Entwicklungen
<b>2.7.2</b> Kombination CP des CAP-Theorems	<b>5.2.6</b> Analyse von Video- und Sprachdaten	<b>8.3</b> Markt-Entwicklungen
<b>2.7.3</b> Kombination AP des CAP-Theorems	<b>5.3</b> Weltweite Immobilienvermietung am Beispiel von AirBnB	<b>8.4</b> Business-Entwicklungen
<b>3 Hadoop und Spark</b>	<b>5.4</b> Internationale Datacenterverknüpfung am Beispiel CERN	<b>8.5</b> Diskussion der Ergebnisse
<b>3.1</b> Hadoop	<b>5.5</b> Beispiel: US FINRA	<b>9 Hadoop Installation &amp; Configuration &amp; Go!</b>
<b>3.2</b> MapReduce	<b>5.6</b> Beispiel: Jobbörse MONSTER	<b>9.1</b> Installationsschema für APACHE HADOOP 3.1.3
<b>3.2.1</b> Hauptkonzepte – MapReduce	<b>5.7</b> Beispiel: Google BigTable	<b>9.2</b> Hadoop 3.1.3 auf Github
<b>3.2.2</b> Beispiel: Wörter zählen	<b>6 Data Governance + Risiken</b>	<b>9.3</b> Die Experteach Labor-Umgebung
<b>3.3</b> HDFS	<b>6.1</b> Die 3 Säulen von Data Governance	<b>9.4</b> Anpassung der Konfigurationsdateien
<b>3.3.1</b> HDFS – Hauptkomponenten	<b>6.2</b> Was kann ich tun, um meine Daten zu schützen?	<b>9.5</b> Übersicht der Filestrukturen im Labor
<b>3.3.2</b> HDFS – Architektur	<b>6.3</b> Risiken	<b>9.6</b> Erster Start des HDFS
<b>3.4</b> YARN	<b>6.4</b> Data Compliance Risiko	<b>9.7</b> Syntax und Ablauf von Zählaufgaben
<b>3.5</b> Apache Spark	<b>6.4.1</b> National und in Europa	<b>9.8</b> Output beim MAPREDUCE Vorgang
<b>3.5.1</b> Resilient Distributed Dataset	<b>6.4.2</b> International	<b>9.9</b> Hadoop Cockpit
<b>3.5.2</b> Spark SQL	<b>6.4.3</b> Sozialrisiko	<b>9.10</b> Wordcount Abfrage via PIG:
<b>3.5.3</b> Spark Streaming	<b>6.5</b> Datenrisiken	<b>9.11</b> RATING – Datensätze filtern (25 Mio)
<b>3.5.4</b> MLlib		
<b>3.5.5</b> Machine Learning		
<b>3.5.6</b> GraphX		

