

# 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** Das ausführliche deutschsprachige digitale Unterlagenpaket, bestehend aus PDF und E-Book, ist im Kurspreis enthalten.

### 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.at/go/BIGD](http://www.experteach.at/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.
<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	
27.10.-29.10.25	Frankfurt	27.10.-29.10.25
		Online



# Inhaltsverzeichnis

## Big Data – Einblick in Hadoop und andere Frameworks

<b>1 Was ist Big Data?</b>	<b>4.9</b> Flink	<b>8.5</b> Datenrisiken
<b>1.1</b> Der große Berg Daten	<b>4.10</b> Oozie	<b>8.5.1</b> Sicherheit der Daten
<b>1.2</b> Anwendungsgebiete von BIG DATA	<b>4.11</b> Zookeeper	<b>8.5.2</b> Qualität der Daten
<b>1.3</b> Die Definition von Big Data: 3-5 „V“s	<b>4.12</b> Ambari	<b>8.6</b> Definitions- und Aussagerisiko
<b>1.3.1</b> Volume	<b>4.13</b> R Connector	<b>8.6.1</b> Faktor Daten
<b>1.3.2</b> Velocity: Geschwindigkeit und Aktualität der Daten	<b>4.14</b> Cassandra	<b>8.6.2</b> Faktor Data Management
<b>1.3.3</b> Variety: Vielfältigkeit der Daten	<b>4.15</b> SAP HANA	<b>8.6.3</b> Faktor Organisation
<b>1.3.4</b> Veracity: Gültigkeit und Zuverlässigkeit der Daten		<b>8.6.4</b> Faktor Prozess
<b>1.3.5</b> Das ist Big Data		<b>8.6.5</b> Faktor Kunde als Betroffener
<b>1.4</b> Der Ursprung von Big Data		<b>8.7</b> Herausforderungen
		<b>8.8</b> EU-AI-Act: Erste Regulierungen von KI
<b>2 Big Data Grundlagen</b>	<b>5 Anforderungen an die Datenbankentwicklung</b>	<b>9 Herausforderungen im Betrieb von Big Data Lösungen</b>
<b>2.1</b> Die BIG DATA Value Chain	<b>5.1</b> Die Entwicklung von SQL zu NoSQL	<b>9.1</b> Wo ist der Anfang?
<b>2.2</b> Quellen für BigData Datenanalysen	<b>5.2</b> Facebook und Twitter: höhere Anforderungen	<b>9.2</b> Hadoop unternehmensweit betreiben
<b>2.3</b> Die Architektur	<b>5.2.1</b> Read-Only-Replicas zur Überbrückung	<b>9.2.1</b> Physische Infrastruktur
<b>2.4</b> SQL: feste, vordefinierte Tabellenschemata	<b>5.2.2</b> Multi-Master Datenbanken	<b>9.2.2</b> Datenhaltung
<b>2.5</b> Normalisierung von Tabellen	<b>5.3</b> ACID vs BASE	<b>9.2.3</b> Datenzugriff
<b>2.6</b> NoSQL	<b>5.4</b> Vorteile von NoSQL (BASE)	<b>9.2.4</b> Datenintegration
<b>2.6.1</b> Key-Value Stores	<b>5.5</b> Der ELK Stack	<b>9.2.5</b> IT-Sicherheit
<b>2.6.2</b> In-Memory Key-Value Stores	<b>5.5.1</b> Elasticsearch Abfragen auf der Konsole	<b>9.2.6</b> Weitere Betriebskriterien
<b>2.6.3</b> Document Stores	<b>5.5.2</b> Kibana - das Dashboard für grafische Anzeigen	<b>9.2.7</b> Wirtschaftliche Kriterien
<b>2.6.4</b> Graph Databases	<b>5.6</b> ELK: Sharding und Reverse Indexing	<b>9.3</b> Betriebsphasen einer neuen Big-Data Landschaft
<b>2.6.5</b> Column Stores	<b>5.7</b> MongoDB - ein beliebter Document Store	<b>9.4</b> Erfolgsfaktoren
<b>2.7</b> CAP-Theorem		
<b>2.7.1</b> Kombination CA des CAP-Theorems	<b>6 Anwendungsbeispiele für Big Data</b>	<b>10 Ausblick</b>
<b>2.7.2</b> Kombination CP des CAP-Theorems	<b>6.1</b> Limitierungen klassischer analytischer Anwendungen	<b>10.1</b> Prognose der Big Data Markt-Kapitalisierung
<b>2.7.3</b> Kombination AP des CAP-Theorems	<b>6.2</b> Einsatzszenarien von Big Data	<b>10.2</b> Marktanteile führender Unternehmen im 4. Quartal 2023
	<b>6.2.1</b> Clickstream Analyse	<b>10.3</b> Orchestrierung in der Cloud
	<b>6.2.2</b> Stimmungsanalysen aus Social Media	<b>10.4</b> KI und ML
	<b>6.2.3</b> Analyse von Log-Daten	<b>10.5</b> Environment-, Social- und Governance-Faktoren (ESG)
	<b>6.2.4</b> Analyse von Sensordaten	<b>10.6</b> Quantencomputing und Big Data
	<b>6.2.5</b> Analyse von Texten	<b>10.7</b> Natural Language Processing (NLP)
	<b>6.2.6</b> Analyse von Video- und Sprachdaten	<b>10.8</b> Big-Data- und ML/KI Fachkräftebedarf
	<b>6.3</b> Weltweite Immobilienvermietung am Beispiel von Airbnb	<b>10.9</b> Aktueller Stand
	<b>6.4</b> Internationale Datacenterverknüpfung am Beispiel CERN	<b>10.10</b> Technische Entwicklungen
	<b>6.5</b> Beispiel: US FINRA	<b>10.11</b> Markt-Entwicklungen
	<b>6.6</b> Beispiel: Jobbörse MONSTER	<b>10.12</b> Business-Entwicklungen
	<b>6.7</b> Beispiel: Google BigTable	<b>10.13</b> Diskussion der Ergebnisse
	<b>7 Pipelines und ETL Prozesse</b>	<b>11 Hadoop Installation &amp; Configuration &amp; Go!</b>
	<b>7.1</b> DB vs. Datawarehouse vs. Data Lake	<b>11.1</b> Installationsschema für APACHE HADOOP 3.1.3
	<b>7.2</b> Snowflake-Schemas	<b>11.2</b> Hadoop 3.1.3 auf Github
	<b>7.3</b> Extract - Transform - Load (ETL)	<b>11.3</b> Die Expertech Labor-Umgebung
	<b>7.4</b> Pivot Tabellen / Klassische Reports	<b>11.4</b> Anpassung der Konfigurationsdateien
	<b>7.5</b> Pipeline für einen ETL Prozess	<b>11.5</b> Übersicht der Filestrukturen im Labor
	<b>7.6</b> Data-Lakes	<b>11.6</b> Erster Start des HDFS
		<b>11.7</b> Syntax und Ablauf von Zählaufgaben
		<b>11.8</b> Output beim MAPREDUCE Vorgang
		<b>11.9</b> Hadoop Cockpit
		<b>11.10</b> Wordcount Abfrage via PIG:
		<b>11.11</b> RATING – Datensätze filtern (25 Mio)
<b>3 Hadoop und Spark</b>		
<b>3.1</b> Hadoop		
<b>3.2</b> MapReduce		
<b>3.2.1</b> Hauptkonzepte – MapReduce		
<b>3.2.2</b> Beispiel: Wörter zählen		
<b>3.3</b> HDFS		
<b>3.3.1</b> HDFS – Hauptkomponenten		
<b>3.3.2</b> HDFS – Architektur		
<b>3.4</b> YARN		
<b>3.5</b> Apache Spark		
<b>3.5.1</b> Resilient Distributed Dataset		
<b>3.5.2</b> Spark SQL		
<b>3.5.3</b> Spark Streaming		
<b>3.5.4</b> MLlib		
<b>3.5.5</b> Machine Learning		
<b>3.5.6</b> GraphX		
<b>4 Big Data Technologien</b>		
<b>4.1</b> Das Hadoop Ecosystem		
<b>4.2</b> Pig		
<b>4.3</b> Hive		
<b>4.4</b> Mahout		
<b>4.5</b> HBase		
<b>4.6</b> Sqoop		
<b>4.7</b> Flume		
<b>4.8</b> Chukwa		
	<b>8 Data Governance + Risiken</b>	
	<b>8.1</b> Die 3 Säulen von Data Governance	
	<b>8.2</b> Was kann ich tun, um meine Daten zu schützen?	
	<b>8.3</b> Risiken	
	<b>8.4</b> Data Compliance Risiko	
	<b>8.4.1</b> National und in Europa	
	<b>8.4.2</b> Inhalte der DSGVO	
	<b>8.4.3</b> International	
	<b>8.4.4</b> Sozialrisiko	

