

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.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 	
27.10.-29.10.25  Frankfurt	28.04.-30.04.26  Frankfurt	
27.10.-29.10.25  Online	28.04.-30.04.26  Online	

Stand 23.05.2025



EXPERTeach



Inhaltsverzeichnis

Big Data – Einblick in Hadoop und andere Frameworks

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

