

## AW415 Big Data on AWS

### Kurzbeschreibung:

Der Kurs Big Data on AWS vermittelt den Teilnehmern Details zu cloudbasierten Big Data-Lösungen.

### Zielgruppe:

- Solutions Architects
- SysOps-Administratoren
- Big-Data Scientists
- Datenanalysten
- Big-Data Manager

### Voraussetzungen:

Die Teilnehmer sollten folgende Voraussetzungen erfüllen:

- Teilnahme am Kurs Big Data Technology Fundamentals oder eine vergleichbare Weiterbildung
- Teilnahme am Kurs AWS Technical Essentials oder eine vergleichbare Weiterbildung
- Basiswissen rund um Big Data-Technologien einschließlich Apache Hadoop-, MapReduce-, HDFS- und SQL/NoSQL-Abfragen
- Praxiserfahrung im Umgang mit wichtigen AWS-Services und der Implementierung von öffentlichen Clouds
- Verständnis von Data Warehouses, relationalen Datenbanksystemen und Datenbank-Design

### Sonstiges:

**Dauer:** 3 Tage

**Preis:** 2685 Euro plus Mwst.

### Ziele:

- Integration von AWS-Lösungen in ein Big Data-Ökosystem
- Einsatz von Apache Hadoop im Umfeld von Amazon EMR
- Konfiguration und Start eines Amazon EMR-Clusters und Identifizierung der einzelnen Komponenten
- Umgang mit üblichen Programmierungs-Frameworks, die Amazon EMR zur Verfügung stellt, z.B. Pig, Hive oder Streaming
- Steigerung der Benutzerfreundlichkeit durch den Einsatz von Hadoop User Experience (Hue)
- Einsatz von In-Memory-Analysen mit Apache Spark
- Auswahl passender AWS-Datenspeicheroptionen
- Effiziente Speicherung und Analyse von Daten mit Hilfe von Amazon Redshift
- Verständnis für Kosten und Sicherheit einer Big-Data-Lösung und deren Verwaltung
- Definieren der Optionen für das Einlesen, Übertragen und Komprimieren von Daten
- Einsatz von Amazon Athena, um Analysen mit Adhoc-Abfragen zu generieren
- Verwendung von AWS Glue, um eine Automatisierung beim Extrahieren, Transformieren und Laden von Arbeitslasten zu erreichen

- Einsatz von Amazon QuickSight für die Abbildung von Daten und Abfragen

## Inhalte/Agenda:

- Dazu gehören unter anderem Amazon (EMR), Amazon Redshift, Amazon Kinesis sowie alle weiteren Komponenten der AWS-Plattform für Big Data. Neben einer Einführung in Amazon Elastic MapReduce (EMR) zeigt dieses Training wie Amazon EMR bei der Datenverarbeitung durch den Einsatz von Hadoop-Tools wie beispielsweise Pig und Hive eingesetzt werden kann.  
Das Erstellen von sicheren und wirtschaftlichen Big-Data-Umgebungen, unter Berücksichtigung bewährter Methoden und das Arbeiten mit Amazon DynamoDB, Amazon Redshift, Amazon Quicksight, Amazon Athena und Amazon Kinesis runden dieses Training ab.

Dieser Kurs setzt sich aus einer Präsentation und Übungen zusammen, um das Erlernte praktisch anzuwenden.

Die Kursunterlagen (E-Book) sind in englischer Sprache, die Kurssprache ist deutsch.

- Der Kurs Security Engineering on AWS unterstützt Sie bei der Vorbereitung auf folgende Prüfungen:

- ◆ AWS Certified Data Analytics Specialty

- **Tag 1**

- ◆ Überblick über Big Data und Apache Hadoop
- ◆ Big Data-Daten Eingabe und Übertragung
- ◆ Vorteile, Architektur und Einsatz von Amazon EMR
- ◆ Amazon EMR-Clusters Starten und Nutzen
- ◆ Programmierungs-Frameworks für Apache Hadoop
- ◆ Lab: Einsatz von Amazon Athena zur Abfrage von Protokolldaten aus Amazon S3

- **Tag 2**

- ◆ Apache Hadoop Programmier-Frameworks
- ◆ Lab: Daten in Amazon DynamoDB Speicherung und Abfrage
- ◆ Einsatz von Streaming für biowissenschaftliche Analysen
- ◆ Amazon EMR
- ◆ Lab: Verarbeitung von Serverprotokollen mit Hive auf Amazon EMR
- ◆ Kostenmanagement für Amazon EMR
- ◆ Web-Schnittstellen auf Amazon EMR
- ◆ Übersicht über die Amazon EMR-Sicherheit
- ◆ Untersuchen der Amazon EMR-Sicherheit
- ◆ Lab: Ausführung von Pig-Skripten in Hue auf Amazon EMR
- ◆ Eingabe, Übertragung und Komprimierung von Daten

- **Tag 3**

- ◆ Amazon Kinesis für die Verarbeitung von Big Data in Echtzeit
- ◆ Visualisieren und Orchestrieren von Big Data
- ◆ Lab: Einsatz von TIBCO Spotfire zur Visualisierung von Data
- ◆ Optionen für die Datenspeicherung in AWS
- ◆ Verwenden von DynamoDB mit Amazon EMR
- ◆ Übersicht: Amazon Redshift und Big Data
- ◆ Verwenden von Amazon Redshift für Big Data
- ◆ Visualisieren von Big Data mit Tableau Desktop oder Jaspersoft BI