



# Building Batch Data Analytics Solutions on AWS

## Building Batch Data Analytics Solutions on AWS

In diesem Kurs lernen Sie, wie Sie Batch-DatenanalySELösungen mit Amazon EMR erstellen, einem Apache Spark- und Apache Hadoop-Managed-Service der Enterprise-Klasse. Sie erfahren, wie sich Amazon EMR in Open-Source-Projekte wie Hive, Hue und HBase sowie in AWS-Services wie AWS Glue und AWS Lake Formation integrieren lässt.

Der Kurs befasst sich mit Datenerfassung, Erfassung, Katalogisierung, Speicherung und Verarbeitung von Komponenten im Kontext von Spark und Hadoop. Sie lernen, EMR-Notebooks zu verwenden, um sowohl Analyse- als auch maschinelle Lernworkloads zu unterstützen. Sie lernen auch, Best Practices für Sicherheit, Leistung und Kostenmanagement auf den Betrieb von Amazon EMR anzuwenden.

Dieser Kurs umfasst Präsentationen, interaktive Demos, Übungslabore, Diskussionen und Klassenübungen.

### Kursinhalt

- Module A: Overview of Data Analytics and the Data Pipeline
- Module 1: Introduction to Amazon EMR
- Module 2: Data Analytics Pipeline Using Amazon EMR: Ingestion and Storage
- Module 3: High-Performance Batch Data Analytics Using Apache Spark on Amazon EMR
- Module 4: Processing and Analyzing Batch Data with Amazon EMR and Hive
- Module 5: Serverless Data Processing
- Module 6: Security and Monitoring of Amazon EMR Clusters
- Module 7: Designing Batch Data Analytics Solutions
- Module B: Developing Modern Data Architectures on AWS

Auf die Labs haben Sie nach dem Kurs noch weitere 4 Wochen Zugriff. So können Sie Übungen wiederholen oder individuell vertiefen.

**E-Book** Die englischsprachigen Original-Unterlagen von Amazon Web Services erhalten Sie als E-Book.

### Zielgruppe

- Dieser Kurs richtet sich an:
- Data Platform Engineers
  - Architects und Operators, die Datenanalyse-Pipelines erstellen und verwalten

Personen mit mindestens einem Jahr Erfahrung in der Verwaltung von Open-Source-Datenframeworks wie Apache Spark oder Apache Hadoop werden von diesem Kurs profitieren.

### Voraussetzungen

Wir empfehlen den AWS Hadoop Fundamentals-Kurs für diejenigen, die eine Auffrischung zu Apache Hadoop benötigen.

Wir empfehlen den vorherigen Kursbesuch von:

- AWS Technical Essentials oder Architecting on AWS
- Building Data Lakes on AWS oder Getting Started with AWS Glue

### Dieser Kurs im Web



Alle tagesaktuellen Informationen und Möglichkeiten zur Bestellung finden Sie unter dem folgenden Link: [www.experteach.de/go/AWBB](http://www.experteach.de/go/AWBB)

### 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>1 Tag</b>	<b>€ 795,-</b>
<b>Online Training</b>	<b>1 Tag</b>	<b>€ 795,-</b>
<b>Termin/Kursort</b>	Kurssprache Deutsch	
24.02.-24.02.25	25.08.-25.08.25	

Stand 25.10.2024



# Inhaltsverzeichnis

## Building Batch Data Analytics Solutions on AWS

### Module A: Overview of Data Analytics and the Data Pipeline

- Data analytics use cases
- Using the data pipeline for analytics

### Module 1: Introduction to Amazon EMR

- Using Amazon EMR in analytics solutions
- Amazon EMR cluster architecture
- Interactive Demo 1: Launching an Amazon EMR cluster

- Cost management strategies

### Module 2: Data Analytics Pipeline Using Amazon EMR: Ingestion and Storage

- Storage optimization with Amazon EMR
- Data ingestion techniques

### Module 3: High-Performance Batch Data Analytics Using Apache Spark on Amazon EMR

- Apache Spark on Amazon EMR use cases
- Why Apache Spark on Amazon EMR
- Spark concepts
- Interactive Demo 2: Interactive analytics using Apache Spark on Amazon EMR

- Transformation, processing, and analytics

- Using notebooks with Amazon EMR

### Module 4: Processing and Analyzing Batch Data with Amazon EMR and Hive

- Using Amazon EMR with Hive to process batch data
- Transformation, processing, and analytics

- Practice Lab 2: Batch data processing using Amazon EMR with Hive

### Module 5: Serverless Data Processing

- Introduction to HBase on Amazon EMR
- Serverless data processing, transformation, and analytics
- Using AWS Glue with Amazon EMR workloads

- Practice Lab 3: Orchestrate data processing in Spark using AWS Step Functions

### Module 6: Security and Monitoring of Amazon EMR Clusters

- Securing EMR clusters
- Interactive Demo 3: Encrypting data at rest in Amazon EMR

- Monitoring and troubleshooting EMR clusters
- Demo: Reviewing Apache Spark cluster history

- Monitoring and troubleshooting Amazon EMR clusters

### Module 7: Designing Batch Data Analytics Solutions

- Batch data analytics use cases
- Activity: Designing a batch data analytics workflow

### Module B: Developing Modern Data Architectures on AWS

- Modern data architectures

