

# Building Batch Data Analytics Solutions on AWS – Intensive Training («AWSB05»)

In diesem Kurs lernen Sie den Aufbau von Batch-DatenanalySELösungen mit Amazon EMR, einem verwalteten Apache Spark- und Apache Hadoop-Service der Enterprise-Klasse.

**Dauer:** 1 Tag

**Preis:** 900.– zzgl. 8.1% MWST

**Kursdokumente:** Digitale Original-AWS-Kursunterlagen

## Inhalt

Erfahren Sie, wie Amazon EMR mit Open-Source-Projekten wie Apache Hive, Hue und HBase und mit AWS-Services wie AWS Glue und AWS Lake Formation integriert wird. Der Kurs behandelt die Komponenten Datenerfassung, -aufnahme, -katalogisierung, -speicherung und -verarbeitung im Kontext von Spark und Hadoop. Sie lernen die Verwendung von EMR Notebooks zur Unterstützung von Analyse- und Machine-Learning-Arbeitslasten. Sie werden auch lernen, die besten Praktiken für Sicherheit, Leistung und Kostenmanagement auf den Betrieb von Amazon EMR anzuwenden.

**Inhalt:**

### Modul A: Überblick über Datenanalyse und die Datenpipeline

- Anwendungsfälle der Datenanalyse
- Verwendung der Datenpipeline für die Datenanalyse

### Modul 1: Einführung in Amazon EMR

- Verwendung von Amazon EMR in Analyselösungen
- Architektur von Amazon-EMR-Clustern

### Interaktive Demo 1: Starten eines Amazon-EMR-Clusters

- Strategien zur Kostenverwaltung

### Modul 2: Datenanalyse-Pipeline mit Amazon EMR: Ingestion und Speicherung

- Speicheroptimierung mit Amazon EMR
- Techniken zur Datenübernahme

### Modul 3: Leistungsstarke Batch-Datenanalyse mit Apache Spark auf Amazon EMR

- Anwendungsfälle für Apache Spark auf Amazon EMR
- Warum Apache Spark auf Amazon EMR
- Spark-Konzepte
- Interaktive Demo 2: Verbinden Sie sich mit einem EMR-Cluster und führen Sie Scala-Befehle mit der Spark-Shell
- Transformation, Verarbeitung und Analyse
- Verwendung von Notebooks mit Amazon EMR
- Practice Lab 1: Datenanalyse mit niedriger Latenz mit Apache Spark auf Amazon EMR

### Modul 4: Verarbeiten und Analysieren von Batch-Daten mit Amazon EMR und Apache Hive

- Verwendung von Amazon EMR mit Hive zur Verarbeitung von Stapeldaten
- Transformation, Verarbeitung und Analyse
- Practice Lab 2: Batch-Datenverarbeitung mit Amazon EMR und Hive

- Einführung in Apache HBase auf Amazon EMR

## Modul 5: Serverlose Datenverarbeitung

- Serverlose Datenverarbeitung, -transformation und -analytik
- Verwendung von AWS Glue mit Amazon-EMR-Arbeitslasten
- Practice Lab 3: Orchestrierung der Datenverarbeitung in Spark mit AWS Step Functions

## Modul 6: Sicherheit und Überwachung von Amazon EMR-Clustern

- Sichern von EMR-Clustern
- Interaktive Demo 3: Client-seitige Verschlüsselung mit EMRFS
- Überwachung und Fehlersuche bei Amazon-EMR-Clustern
- Demo: Überprüfung des Verlaufs von Apache-Spark-Clustern

## Modul 7: Entwerfen von Batch-Datenanalyiselösungen

- Anwendungsfälle für die Batch-Datenanalytik
- Aktivität: Entwerfen eines Arbeitsablaufs für die Batch-Datenanalyse

## Modul B: Entwickeln moderner Datenarchitekturen auf AWS

- Moderne Datenarchitekturen

## Key Learnings

- Vergleich der Funktionen und Vorteile von Data Warehouses, Data Lakes und modernen Datenarchitekturen
- Entwerfen und Implementieren einer Lösung für die Batch-Datenanalyse
- Identifizieren und Anwenden geeigneter Techniken, einschliesslich Komprimierung, zur Optimierung der Datenspeicherung
- Auswählen und Bereitstellen geeigneter Optionen für die Aufnahme, Umwandlung und Speicherung von Daten
- Auswahl der geeigneten Instanz- und Knotentypen, Cluster, automatische Skalierung und Netzwerktopologie für einen bestimmten geschäftlichen Anwendungsfall
- Verstehen, wie sich Datenspeicherung und -verarbeitung auf die Analyse- und Visualisierungs-Mechanismen auswirken, die für die Gewinnung verwertbarer Geschäftserkenntnisse erforderlich sind
- Sichern von Daten im Ruhezustand und bei der Übertragung
- Überwachung von Analyse-Workloads, um Probleme zu erkennen und zu beheben
- Anwendung von Best Practices für das Kostenmanagement

## Methodik & Didaktik

Dieser Kurs beinhaltet Präsentationen, interaktive Demos, praktische Übungen, Diskussionen und Klassenübungen.

## Zielpublikum

Dieser Kurs richtet sich an folgende Jobrollen:

- Data Engineer

Wir empfehlen, dass die Teilnehmer dieses Kurses die folgenden Voraussetzungen mitbringen:

- Mindestens ein Jahr Erfahrung mit der Verwaltung von Open-Source-Daten-Frameworks wie Apache Spark oder Apache Hadoop
- Den Kurs [AWS Hadoop Fundamentals](#) für diejenigen, die eine Auffrischung von Apache Hadoop benötigen

und den folgenden Kurs besucht haben (oder über gleichwertige Kenntnisse verfügen):

- [Building Data Lakes on AWS – Intensive Training \(«AWSB04»\)](#)
- [Architecting on AWS – Intensive Training \(«AWSA01»\)](#)
- [AWS Technical Essentials – Intensive Training \(«AWSE01»\)](#)

## Weiterführende Kurse

- [Building Streaming Data Analytics Solutions on AWS – Intensive Training \(«AWSB08»\)](#)

## Haben Sie Fragen oder möchten Sie einen Firmenkurs buchen?

Wir beraten Sie gerne unter 044 447 21 21 oder [info@digicomp.ch](mailto:info@digicomp.ch). Detaillierte Infos zu den Terminen finden Sie unter [www.digicomp.ch/weiterbildung-it-provider/amazon-web-services-aws/aws-data-engineer/kurs-building-batch-data-analytics-solutions-on-aws-intensive-training](http://www.digicomp.ch/weiterbildung-it-provider/amazon-web-services-aws/aws-data-engineer/kurs-building-batch-data-analytics-solutions-on-aws-intensive-training)