

## ***AW272 Building Batch Data Analytics Solutions on AWS***

### **Kurzbeschreibung:**

In diesem Kurs **AW272 Building Batch Data Analytics Solutions on AWS** lernen Sie, wie Sie Batch Data Analytics-Lösungen mit Amazon EMR aufbauen – einem verwalteten Enterprise-Service für Apache Spark und Apache Hadoop. Sie lernen, wie Amazon EMR mit Open-Source-Projekten wie Apache Hive, Hue und HBase sowie mit AWS-Services wie AWS Glue und AWS Lake Formation integriert wird. Der Kurs **AW272 Building Batch Data Analytics Solutions on AWS** behandelt die Komponenten Datenerfassung, Datenaufnahme, Katalogisierung, Speicherung und Verarbeitung im Kontext von Spark und Hadoop. Sie lernen, EMR Notebooks sowohl für Analytics- als auch für Machine-Learning-Workloads zu verwenden. Darüber hinaus lernen Sie, Best Practices zu Sicherheit, Performance und Kostenmanagement beim Betrieb von Amazon EMR anzuwenden.

Dieser Kurs **AW272 Building Batch Data Analytics Solutions on AWS** umfasst Präsentationen, interaktive Demos, praktische Übungen (Labs), Diskussionen und Kursaktivitäten.

### **Zielgruppe:**

Der Kurs **AW272 Building Batch Data Analytics Solutions on AWS** richtet sich an:

- Data Platform Engineers
- Architekten und Betreiber, die Data-Analytics-Pipelines aufbauen und betreiben

### **Voraussetzungen:**

Um an dem Kurs **AW272 Building Batch Data Analytics Solutions on AWS** bei qSkills teilnehmen zu können, sollten Sie folgende AWS Trainings besucht haben:

- [AW110 AWS Technical Essentials](#)  
oder
- [AW200 Architecting on AWS](#)
- [AW270 Building Data Lakes on AWS](#)

Teilnehmende mit mindestens einjähriger Erfahrung im Management von Open-Source-Datenframeworks wie Apache Spark oder Apache Hadoop profitieren besonders von diesem Kurs.

### **Sonstiges:**

**Dauer:** 1 Tage

**Preis:** 750 Euro plus Mwst.

### **Ziele:**

In diesem Kurs **AW272 Building Batch Data Analytics Solutions on AWS** lernen Sie:

- Die Funktionen und Vorteile von Data Warehouses, Data Lakes und modernen Datenarchitekturen zu vergleichen

- Eine Batch Data Analytics-Lösung zu entwerfen und zu implementieren
- Geeignete Techniken – einschließlich Komprimierung – zur Optimierung der Datenspeicherung zu identifizieren und anzuwenden
- Geeignete Optionen zur Datenaufnahme, -transformation und -speicherung auszuwählen und bereitzustellen
- Die passenden Instanz- und Knotentypen, Cluster, Auto Scaling-Mechanismen und Netzwerktopologien für einen bestimmten geschäftlichen Anwendungsfall auszuwählen
- Zu verstehen, wie Datenspeicherung und -verarbeitung Analyse- und Visualisierungsmechanismen beeinflussen, die zur Gewinnung geschäftsrelevanter Erkenntnisse erforderlich sind
- Daten im Ruhezustand und bei der Übertragung zu schützen
- Analytics-Workloads zu überwachen, um Probleme zu identifizieren und zu beheben
- Best Practices im Kostenmanagement anzuwenden

## Inhalte/Agenda:

- **◆ Überblick über Data Analytics und die Datenpipeline**
  - ◆ Anwendungsfälle für Data Analytics
  - ◆ Nutzung der Datenpipeline für Analytics
- **◆ Einführung in Amazon EMR**
  - ◆ Einsatz von Amazon EMR in Analytics-Lösungen
  - ◆ Architektur von Amazon EMR-Clustern
  - ◆ Interaktive Demo 1: Starten eines Amazon EMR-Clusters
  - ◆ Strategien zum Kostenmanagement
- **◆ Data Analytics-Pipeline mit Amazon EMR: Datenaufnahme und -speicherung**
  - ◆ Speicheroptimierung mit Amazon EMR
  - ◆ Techniken zur Datenaufnahme
- **◆ Hochleistungs-Batch Data Analytics mit Apache Spark auf Amazon EMR**
  - ◆ Anwendungsfälle für Apache Spark auf Amazon EMR
  - ◆ Warum Apache Spark auf Amazon EMR?
  - ◆ Spark-Konzepte
  - ◆ Interaktive Demo 2: Verbindung zu einem EMR-Cluster und Ausführung von Scala-Befehlen mit der Spark-Shell
  - ◆ Transformation, Verarbeitung und Analyse
  - ◆ Verwendung von Notebooks mit Amazon EMR
  - ◆ Practice Lab 1: Low-Latency Data Analytics mit Apache Spark auf Amazon EMR
- **◆ Verarbeitung und Analyse von Batch-Daten mit Amazon EMR und Apache Hive**
  - ◆ Verwendung von Amazon EMR mit Hive zur Verarbeitung von Batch-Daten
  - ◆ Transformation, Verarbeitung und Analyse
  - ◆ Practice Lab 2: Batch-Datenverarbeitung mit Amazon EMR und Hive
  - ◆ Einführung in Apache HBase auf Amazon EMR
- **◆ Serverlose Datenverarbeitung**
  - ◆ Serverlose Datenverarbeitung, -transformation und -analyse
  - ◆ Verwendung von AWS Glue mit Amazon EMR-Workloads
  - ◆ Practice Lab 3: Orchestrierung der Datenverarbeitung in Spark mit AWS Step Functions
- **◆ Sicherheit und Monitoring von Amazon EMR-Clustern**
  - ◆ Absicherung von EMR-Clustern
  - ◆ Interaktive Demo 3: Clientseitige Verschlüsselung mit EMRFS
  - ◆ Monitoring und Fehlerbehebung von Amazon EMR-Clustern
  - ◆ Demo: Auswertung des Apache Spark-Clusterverlaufs
- **◆ Design von Batch Data Analytics-Lösungen**
  - ◆ Anwendungsfälle für Batch Data Analytics
  - ◆ Kursaktivität: Design eines Workflows für Batch Data Analytics
- **◆ Entwicklung moderner Datenarchitekturen auf AWS**
  - ◆ Moderne Datenarchitekturen