Building Batch Data Analytics Solutions on AWS (BBDAS)

ID BBDAS **Preis** CHF 850.– (exkl. MwSt.) **Dauer** 1 Tag

Zielgruppe

Dieser Kurs richtet sich an:

- Ingenieure für Datenplattformen
- Architekten und Betreiber, die Datenanalyse-Pipelines aufbauen und verwalten

Empfohlenes Training für die Zertifizierung zum

AWS Certified Data Engineer - Associate (AWCDEA)

Voraussetzungen

Teilnehmer mit mindestens einem Jahr Erfahrung im Umgang mit Open-Source-Datenframeworks wie Apache Spark oder Apache Hadoop profitieren von diesem Kurs.

Kursziele

In diesem Kurs werden Sie lernen:

- Vergleichen Sie die Funktionen und Vorteile von Data Warehouses, Data Lakes und modernen Datenarchitekturen
- Entwurf und Implementierung einer Lösung für die Batch-Datenanalyse
- Identifizierung und Anwendung geeigneter Techniken, einschliesslich Komprimierung, zur Optimierung der Datenspeicherung
- Auswahl und Einsatz geeigneter Optionen zur Aufnahme, Umwandlung und Speicherung von Daten
- Auswahl der geeigneten Instanz- und Knotentypen, Cluster, der automatischen Skalierung und der Netzwerktopologie für einen bestimmten geschäftlichen Anwendungsfall
- Verstehen, wie sich Datenspeicherung und -verarbeitung auf die Analyse- und Visualisierungsmechanismen auswirken, die für die Gewinnung verwertbarer Geschäftseinblicke erforderlich sind
- Sichere Daten im Ruhezustand und bei der Übertragung
- Überwachen von Analyse-Workloads, um Probleme zu erkennen und zu beheben
- Anwendung bewährter Kostenmanagementverfahren

Kursinhalt

Modul A: Überblick über Datenanalyse und die Datenpipeline

- Anwendungsfälle der Datenanalyse
- Nutzung der Datenpipeline für Analysen

Modul 1: Einführung in Amazon EMR

- Verwendung von Amazon EMR in Analyselösungen
- Amazon EMR-Cluster-Architektur
- Interaktive Demo 1: Starten eines Amazon EMR-Clusters
- Strategien für das Kostenmanagement

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

- Speicheroptimierung mit Amazon EMR
- Techniken für die Datenübernahme

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

- Apache Spark auf Amazon EMR Anwendungsfälle
- Warum Apache Spark auf Amazon EMR
- Spark-Konzepte
- Interaktive Demo 2: Verbinden mit einem EMR-Cluster und Ausführen von Scala-Befehlen mit der Spark-Shell
- Umwandlung, Verarbeitung und Analyse
- Verwendung von Notebooks mit Amazon EMR
- Praxisübung 1: Datenanalyse mit niedriger Latenz mit Apache Spark auf Amazon EMR

Modul 4: Verarbeitung und Analyse von Batch-Daten mit Amazon EMR und Apache Hive

- Verwendung von Amazon EMR mit Hive zur Verarbeitung von Stapeldaten
- Umwandlung, Verarbeitung und Analyse
- Praxisübung 2: Batch-Datenverarbeitung mit Amazon EMR und Hive
- Einführung in Apache HBase auf Amazon EMR

Modul 5: Serverlose Datenverarbeitung

- Serverlose Datenverarbeitung, -umwandlung und -analyse

- Verwendung von AWS Glue mit Amazon EMR-Arbeitslasten
- Praxisübung 3: Orchestrierung der Datenverarbeitung in Spark mit AWS Step Functions

Modul 6: Sicherheit und Überwachung von Amazon EMR-Clustern

- Sicherung von EMR-Clustern
- Interaktive Demo 3: Client-seitige Verschlüsselung mit EMRFS
- Überwachung und Fehlerbehebung von Amazon EMR-Clustern
- Demo: Überprüfung der Apache Spark-Cluster-Historie

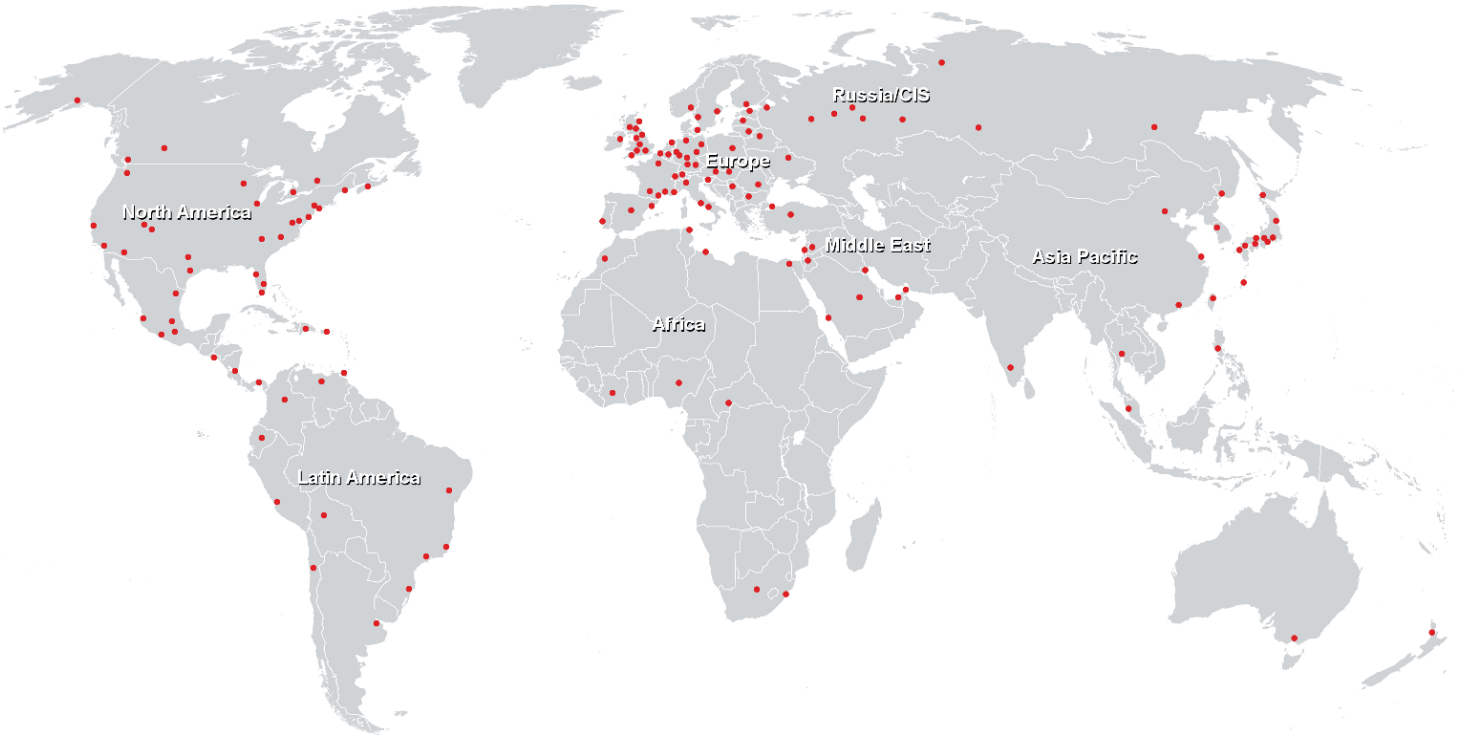
Modul 7: Entwurf von Batch Data Analytics-Lösungen

- Anwendungsfälle der Batch-Datenanalyse
- Tätigkeit: Entwurf eines Arbeitsablaufs für die Batch-Datenanalyse

Modul B: Entwicklung von modernen Datenarchitekturen auf AWS

- Moderne Datenarchitekturen

Weltweite Trainingscenter



Fast Lane Institute for Knowledge Transfer GmbH

Husacherstrasse 3
CH-8304 Wallisellen
Tel. +41 44 832 50 80

info@flane.ch, https://www.flane.ch