

AW480 Building Data Analytics Solutions using Amazon Redshift

Kurzbeschreibung:

In diesem Kurs erstellen Sie eine Datenanalyselösung mit Amazon Redshift, einem Cloud-Data-Warehouse-Dienst. Der Kurs konzentriert sich auf die Datenerfassungs-, Aufnahme-, Katalogisierungs-, Speicherungs- und Verarbeitungskomponenten der Analysepipeline. Sie lernen, Amazon Redshift in einen Data Lake zu integrieren, um sowohl Analyse- als auch Machine-Learning-Workloads zu unterstützen. Sie erfahren auch, wie Sie Best Practices für Sicherheit, Leistung und Kostenmanagement auf den Betrieb von Amazon Redshift anwenden.

Zielgruppe:

Der Kurs richtet sich an:

- Data warehouse engineers
- Data platform engineers
- Architekten und Betreiber von data analytics pipelines

Voraussetzungen:

Um an dem Kurs Building Data Analytics Using Amazon Redshift bei qSkills teilnehmen zu können, sollten Sie die folgenden AWS-Trainings besucht haben:

- "AWS Technical Essentials"
- oder
- "Architecting on AWS"
- "Building Data Lakes auf AWS"

Erfahrung im Umgang mit AWS-Technologien zur Unterstützung von Data Lakes und anderen datengesteuerten Workloads ist zusätzlich empfehlenswert.

Sonstiges:

Dauer: 1 Tage

Preis: 750 Euro plus Mwst.

Ziele:

In diesem Kurs lernen Sie:

- Vergleich von Nutzen und Funktionen von Data Warehouses, Data Lakes und modernen Datenarchitekturen
- Design und Implementierung einer Lösung zum analysieren von Data Warehouses
- Identifikation und Anwendung entsprechender Techniken wie z.B. Kompression zur Optimierung der Datenspeicherung
- Auswahl und Einsatz angemessener Optionen zum Aufnehmen, transformieren und Speichern von Daten

- Auswahl angemessener Instanzen und Node-Typen, Cluster, Automatische Skalierung sowie der Netzwerk Topologie für spezielle Unternehmens Use-Cases
- Zu verstehen, wie sich die Datenspeicherung und Datenverarbeitung auf die Analyse und Visualisierungsmechanismen auswirkt
- Sichern von Daten im Ruhezustand und während der Übertragung
- Überwachen von Analyse-Workloads, um Probleme zu erkennen und zu beheben
- Best Practices für das Kostenmanagement anwenden

Inhalte/Agenda:

- Modul A: Überblick zur Datenanalyse und der Datenpipeline
 - ◆ Anwendungsfälle der Datenanalyse
 - ◆ Anwendung der Datenpipeline für die Analyse
- Modul 1: Die Nutzung von Amazon Redshift mit der Data Analytics Pipeline
 - ◆ Warum Amazon Redshift für Data Warehouses?
 - ◆ Überblick Amazon Redshift
- Modul 2: Einführung in Amazon Redshift
 - ◆ Die Architektur von Amazon Redshift
 - ◆ Interaktive Demo 1: Rundgang in der Amazon Redshift Konsole
 - ◆ Amazon Redshift Funktionen
 - ◆ Lab 1: Laden und Abfragen von Daten in Amazon Redshift Clustern
- Modul 3: Aufnahme und Speicherung
 - ◆ Aufnahme
 - ◆ Interaktive Demo 2: Verbindung von Amazon Redshift Clustern mit Jupyter Notebook und API
 - ◆ Datenverteilung und -Speicherung
 - ◆ Interaktive Demo 3: Analyse von teilstrukturierten Daten mit dem SUPER Datentyp
 - ◆ Abfragen von Daten in Amazon Redshift
 - ◆ Lab 2: Datenanalyse mit Amazon Redshift Spectrum
- Modul 4: Verarbeitung und Optimierung von Daten
 - ◆ Datentransformation
 - ◆ Fortgeschrittene Abfragen
 - ◆ Lab 3: Datentransformation und Abfragen in Amazon Redshift
 - ◆ Ressourcenmanagement
 - ◆ Interaktive Demo 4: Anwenden des gemischten Workload-Managements in Amazon Redshift
 - ◆ Automatisierung und Optimierung
 - ◆ Interaktive Demo 5: Amazon Redshift Cluster abändern von dc2.large zu ra3.xplus Cluster
- Modul 5: Sicherheit und Überwachung von Amazon Redshift Clustern
 - ◆ Sichern des Amazon Redshift Cluster
 - ◆ Überwachung und Troubleshooting zu Amazon Redshift Clustern
- Modul 6: Gestalten von Data Warehouse Analytics Solutions
 - ◆ Rückblick zu Data Warehouse Anwendungsfällen
 - ◆ Übung: Gestalten eines Data Warehouse analytics workflow
- Modul B: Entwicklung von Modern Data Architectures auf AWS
 - ◆ Moderne Datenarchitekturen