

Data Warehousing on AWS



Odbiorcy szkolenia

Ten kurs jest przeznaczony dla:

- Architektów baz danych
- Administratorów baz danych
- Programistów baz danych
- Analityków danych
- Naukowców danych



Korzyści

W tym kursie nauczysz się:

- Omówić podstawowe pojęcia związane z magazynowaniem danych oraz interakcję między magazynowaniem danych a rozwiązaniami big data.
- Uruchomić klaster Amazon Redshift i wykorzystać jego komponenty, funkcje i możliwości do implementacji magazynu danych w chmurze.
- Korzystać z innych usług AWS związanych z danymi i analizą, takich jak Amazon DynamoDB, Amazon EMR, Amazon Kinesis i Amazon S3, w celu wspierania rozwiązania magazynu danych.
- Projektować architekturę magazynu danych.
- Identyfikować problemy z wydajnością, optymalizować zapytania i dostrajać bazę danych w celu uzyskania lepszej wydajności.
- Korzystać z Amazon Redshift Spectrum do analizy danych bezpośrednio z koszyka Amazon S3.
- Używać Amazon QuickSight do przeprowadzania analiz danych i wizualizacji zadań na podstawie magazynu danych.



Program szkolenia

Dzień 1

Moduł 1: Wprowadzenie do hurtowni danych

- Bazy danych relacyjne
- Koncepcje hurtowni danych
- Splot hurtowni danych z rozwiązaniami big data
- Przegląd zarządzania danymi w AWS
- Laboratorium 1: Wprowadzenie do Amazon Redshift

Moduł 2: Wprowadzenie do Amazon Redshift

- Przegląd koncepcyjny
- Przykłady zastosowań w rzeczywistości
- Laboratorium 2: Uruchamianie klastra Amazon Redshift

Moduł 3: Uruchamianie klastrów

- Budowanie klastra
- Łączenie z klastrem
- Kontrola dostępu
- Bezpieczeństwo bazy danych
- Ładowanie danych
- Laboratorium 3: Optymalizacja schematów baz danych

Dzień 2

Moduł 4: Projektowanie schematu bazy danych

- Schematy i typy danych
- Kompresja kolumnowa
- Style dystrybucji danych
- Metody sortowania danych

Moduł 5: Identyfikacja źródeł danych

- Przegląd źródeł danych
- Amazon S3
- Amazon DynamoDB
- Amazon EMR
- Amazon Kinesis Data Firehose
- AWS Lambda Database Loader dla Amazon Redshift
- Laboratorium 4: Ładowanie danych w czasie rzeczywistym do bazy danych Amazon Redshift

Moduł 6: Ładowanie danych

- Przygotowywanie danych
- Ładowanie danych za pomocą COPY
- Utrzymywanie tabel
- Równoczesne operacje zapisu
- Rozwiązywanie problemów z ładowaniem

- Laboratorium 5: Ładowanie danych za pomocą komendy COPY

Dzień 3

Moduł 7: Pisanie zapytań i optymalizacja wydajności

- Amazon Redshift SQL
- Funkcje zdefiniowane przez użytkownika (UDF)
- Czynniki wpływające na wydajność zapytań
- Komenda EXPLAIN i plany zapytań
- Zarządzanie obciążeniem (WLM)
- Laboratorium 6: Konfiguracja zarządzania obciążeniem

Moduł 8: Amazon Redshift Spectrum

- Amazon Redshift Spectrum
- Konfigurowanie danych dla Amazon Redshift Spectrum
- Zapytania w Amazon Redshift Spectrum
- Laboratorium 7: Korzystanie z Amazon Redshift Spectrum

Moduł 9: Utrzymywanie klastrów

- Dzienniki audytu
- Monitorowanie wydajności
- Wydarzenia i powiadomienia
- Laboratorium 8: Audyt i monitorowanie klastrów
- Zmiana rozmiaru klastrów
- Tworzenie kopii zapasowych i przywracanie klastrów
- Etykietowanie zasobów oraz limity i ograniczenia
- Laboratorium 9: Tworzenie kopii zapasowych, przywracanie i zmiana rozmiaru klastrów

Moduł 10: Analiza i wizualizacja danych

- Potęga wizualizacji
- Tworzenie pulpitów nawigacyjnych
- Wydania i funkcje Amazon QuickSight.



Oczekiwane przygotowanie uczestnika

Zalecamy, aby uczestnicy tego kursu mieli:

- Ukończony kurs AWS Technical Essentials (lub posiadali równoważne doświadczenie z AWS)
- Znajomość baz danych relacyjnych oraz pojęć związanych z projektowaniem baz danych.



Szkolenie obejmuje

Kurs ten obejmuje prezentacje, ćwiczenia grupowe oraz laboratoria praktyczne.



Czas trwania

3 dni / 21 godzin

Język

- Szkolenie: polski
- Materiały: angielski.