

Data Engineering on AWS

To kompleksowe, 3-dniowe szkolenie prowadzone przez instruktora zapewnia szczegółowe wprowadzenie do praktyk i rozwiązań data engineering na platformie Amazon Web Services (AWS). Uczestnicy nauczą się, jak projektować, budować, optymalizować i zabezpieczać rozwiązania data engineering z wykorzystaniem usług AWS.

Zakres obejmuje zarówno podstawowe koncepcje, jak i praktyczne wdrożenia:

- data lakes
- data warehouses
- procesów przetwarzania danych wsadowych (batch) i strumieniowych (streaming)

Szkolenie wyposaża specjalistów danych w umiejętności potrzebne do projektowania i zarządzania nowoczesnymi rozwiązaniami danych w dużej skali.



Odbiorcy szkolenia

Odbiorcy szkolenia:

- Data engineer
- Solutions architect
- DevOps engineer
- Specjalista IT
- Data analyst chcący rozwijać się w kierunku data engineering



Korzyści

W ramach kursu nauczysz się:

- Projektować i wdrażać skalowalne data lakes oraz data warehouses na AWS

- Budować, optymalizować i zabezpieczać procesy przetwarzania danych wsadowych (batch)
- Tworzyć i zarządzać rozwiązaniami przetwarzania danych w czasie rzeczywistym (streaming)
- Stosować najlepsze praktyki w zakresie zarządzania danymi i bezpieczeństwa
- Automatyzować procesy data engineering z wykorzystaniem usług AWS
- Wdrażać kontrolę dostępu i mechanizmy bezpieczeństwa dla rozwiązań danych



Program szkolenia

Dzień 1

Course Introduction

Moduł 1: Data Engineering Roles and Key Concepts

- Rola data engineer
- Odkrywanie danych w systemach analitycznych
- Usługi AWS wykorzystywane w przetwarzaniu danych
- Ciągła integracja i ciągłe dostarczanie (CI/CD)
- Aspekty sieciowe

Moduł 2: Designing and Implementing Data Lakes

- Wprowadzenie do data lakes
- Przechowywanie danych w data lake
- Pobieranie danych (ingest)
- Katalogowanie danych
- Transformacja danych
- Udostępnianie danych

Lab: Konfiguracja Data Lake na AWS

Moduł 3: Optimizing and Securing Data Lake Solutions

- Optymalizacja wydajności
- Bezpieczeństwo z wykorzystaniem AWS Lake Formation
- Zarządzanie uprawnieniami w Lake Formation
- Bezpieczeństwo i governance danych
- Rozwiązywanie problemów

Lab: Automatyzacja tworzenia Data Lake z użyciem AWS Lake Formation Blueprints

Dzień 2

Moduł 4: Data Warehouse Architecture and Design Principles

- Wprowadzenie do data warehouses
- Przegląd Amazon Redshift
- Ładowanie danych do Amazon Redshift
- Przetwarzanie danych
- Udostępnianie danych

Lab: Konfiguracja hurtowni danych z wykorzystaniem Amazon Redshift Serverless

Moduł 5: Performance Optimization Techniques for Data Warehouses

- Monitorowanie i optymalizacja
- Optymalizacja danych w Amazon Redshift
- Optymalizacja zapytań w Amazon Redshift
- Orkiestracja danych

Moduł 6: Security and Access Control for Data Warehouses

- Uwierzytelnianie i kontrola dostępu w Amazon Redshift
- Bezpieczeństwo danych w Amazon Redshift

Lab: Praca z Amazon Redshift

Moduł 7: Designing Batch Data Processing

- Wprowadzenie do przetwarzania wsadowego
- Projektowanie procesów przetwarzania wsadowego
- Pobieranie danych wsadowych

Moduł 8: Implementing Batch Data Processing Strategies

- Przetwarzanie i transformacja danych
- Transformacja formatów danych
- Integracja danych
- Katalogowanie danych
- Udostępnianie danych

Lab: Dzień z życia Data Engineera

Dzień 3

Moduł 9: Optimizing, Orchestrating, and Securing Batch Processing

- Optymalizacja przetwarzania wsadowego
- Orkiestracja przetwarzania
- Zabezpieczanie przetwarzania

Lab: Orkiestracja przetwarzania danych w Spark z wykorzystaniem AWS Step Functions

Moduł 10: Streaming Data Architecture Patterns

- Wprowadzenie do przetwarzania strumieniowego
- Pobieranie danych ze źródeł strumieniowych
- Przechowywanie danych strumieniowych
- Przetwarzanie danych strumieniowych
- Analiza danych strumieniowych

Lab: Analiza strumieniowa z wykorzystaniem Amazon Managed Service for Apache Flink

Moduł 11: Optimizing and Securing Streaming Solutions

- Optymalizacja rozwiązań strumieniowych
- Zabezpieczanie przetwarzania strumieniowego

Lab: Kontrola dostępu z wykorzystaniem Amazon Managed Streaming for Apache Kafka

Moduł 12: Compliance and Cost Optimization

- Wymagania zgodności (compliance)
- Optymalizacja kosztów

Moduł 13: Course Wrap-Up



Oczekiwane przygotowanie uczestnika

Zalecane:

- Podstawowa znajomość usług AWS
- Znajomość koncepcji baz danych
- Podstawowa znajomość programowania lub skryptów
- Zrozumienie podstaw przetwarzania danych



Szkolenie obejmuje

Szkolenie obejmuje prezentacje, demonstracje oraz ćwiczenia praktyczne (labs).



Czas trwania

3 dni / 24 godzin

Język

Szkolenie: polski

Materiały: angielski