



Program kursu

Data Engineer

kurs weekendowy

90 h zajęć z trenerem na żywo

40 h pracy własnej

70% / 30% praktyki / teorii

Postaw na
praktykę

Kurs Data Engineer

Buduj pipeline'y, na których stoi cała analityka.

Jeśli Twoją pasją jest budowanie solidnych fundamentów pod systemy Big Data, a nie tylko analiza wyników, to szkolenie jest dla Ciebie. To ścieżka dla osób z zacięciem technicznym, które chcą wyjść poza ramy prostych baz danych i zająć się projektowaniem skalowalnych potoków (pipelines) oraz nowoczesnych ekosystemów danych.

Co będziesz potrafić po kursie?

- ✓ Projektujesz pipeline'y ETL/ELT
- ✓ Pracujesz z hurtowniami danych i Big Data (Spark, Databricks)
- ✓ Orkiestrujesz przepływy w Airflow
- ✓ Budujesz strumienie danych z Kafką

Liczba godzin i forma zajęć

Zajęcia odbywają się w soboty i niedziele, zależnie od edycji w godzinach 11:30–14:30 lub 8:00–11:00.

10 h

Prework

szacowany czas na przygotowanie do kursu

90 h

Zajęcia z trenerem

zdalnie na żywo, zakończone certyfikatem

30 h

Praca własna

szacowany czas na pracę indywidualną między zajęciami



Nagrania z zajęć

Nagrania wideo z zajęć dostępne po zakończeniu kursu.



Narzędzia i technologie

W praktyce opanujesz: AWS, AWS S3, SQL, Kubernetes, Docker, Python, Pandas, NumPy, Polars, Hadoop, Spark, DynamoDB, Kafka. Zakres opiera się na realnych case studies.



Certyfikat

Kurs kończysz zaliczeniem potwierdzającym umiejętności oraz certyfikatem ukończenia.



AI w kursie

Uczysz się korzystać z AI jak nowoczesny specjalista — do przyspieszania pracy, debugowania i analizy — zawsze z weryfikacją wyniku przez człowieka.

Data Engineer

14 modułów · 90 h



MODUŁ 0 · PREWORK

Pework

Moduł przygotowawczy budujący fundament przed kursem: podstawy kontroli wersji Git, składni Pythona oraz pracy z bazami danych w SQL.

ZAKRES MODUŁU

- ✓ Git: podstawy kontroli wersji i śledzenie zmian w projekcie
- ✓ Git: korzystanie z repozytoriów i współpraca z innymi programistami
- ✓ Python: podstawy składni — zmienne, pętle i funkcje
- ✓ SQL: podstawowe operacje na bazach danych
- ✓ SQL: projektowanie prostych schematów bazy danych



MODUŁ 1

Warsztat Data Engineera

Praktyczne zastosowania Dockera w pracy Data Engineera: konteneryzacja aplikacji danych, automatyzacja wdrożeń i monitorowanie infrastruktury.

ZAKRES MODUŁU

- ✓ Praktyczne zastosowania Dockera w pracy Data Engineera
- ✓ Konfiguracja, zarządzanie obrazami i uruchamianie kontenerów dla aplikacji danych
- ✓ Rola Dockera w procesach integracji, przetwarzania i zarządzania danymi
- ✓ Techniki skalowania i automatyzacji wdrożeń w środowiskach produkcyjnych
- ✓ Narzędzia i metody monitorowania oraz optymalizacji infrastruktury danych

2

MODUŁ 2

Wprowadzenie do AWS

Konfiguracja i zarządzanie podstawowymi usługami AWS, integracja usług oraz optymalizacja kosztów i wydajności w środowisku chmurowym.

ZAKRES MODUŁU

- ✓ Konfiguracja i zarządzanie podstawowymi usługami AWS, takimi jak EC2 i S3
- ✓ Zasady zarządzania dostępem i bezpieczeństwem w środowisku chmurowym
- ✓ Integracja różnych usług AWS w celu tworzenia kompleksowych rozwiązań
- ✓ Optymalizacja wydajności usług AWS oraz monitorowanie zasobów
- ✓ Konfiguracja sieci i storage w AWS
- ✓ Zarządzanie kosztami w AWS i optymalizacja zasobów chmurowych

3

MODUŁ 3

Database i zaawansowany SQL

Budowanie skomplikowanych zapytań SQL z funkcjami okna i CTE, optymalizacja i indeksowanie oraz projektowanie schematów dla różnych scenariuszy biznesowych.

ZAKRES MODUŁU

- ✓ Budowanie skomplikowanych zapytań SQL z wykorzystaniem funkcji okna i CTE
- ✓ Zaawansowane techniki optymalizacji zapytań i indeksowania baz danych
- ✓ Projektowanie i implementacja schematów baz danych dla różnych scenariuszy biznesowych
- ✓ Zarządzanie dużymi zbiorami danych i optymalizacja ich przechowywania
- ✓ Techniki partycjonowania danych w bazach SQL
- ✓ Analiza i interpretacja danych za pomocą zaawansowanych zapytań SQL

4

MODUŁ 4

Praca z nierelacyjnymi bazami danych (DynamoDB)

Projektowanie baz danych w modelu NoSQL na przykładzie DynamoDB: konfiguracja, skalowanie, optymalizacja zapytań i zarządzanie danymi w środowiskach rozproszonych.

ZAKRES MODUŁU

- ✓ Projektowanie baz danych z wykorzystaniem modeli NoSQL
- ✓ Podstawy pracy z DynamoDB — konfiguracja i zarządzanie bazami
- ✓ Skalowanie baz NoSQL w zależności od potrzeb aplikacji
- ✓ Optymalizacja wydajności zapytań i operacji w bazach NoSQL
- ✓ Różnice między modelami relacyjnymi a nierelacyjnymi i ich zastosowania
- ✓ Zarządzanie danymi w środowiskach rozproszonych z wykorzystaniem NoSQL

5

MODUŁ 5

Python w przetwarzaniu danych

Przetwarzanie dużych zbiorów danych w Pythonie z użyciem Pandas, NumPy i Polars, integracja źródeł danych przez API oraz optymalizacja kodu.

ZAKRES MODUŁU

- ✓ Biblioteki Pandas, NumPy i Polars do przetwarzania dużych zbiorów danych
- ✓ Zaawansowane techniki programowania w Pythonie — zarządzanie pamięcią i optymalizacja kodu
- ✓ Tworzenie skryptów Pythona do automatyzacji zadań przetwarzania danych
- ✓ Praca z API do pobierania i integrowania danych z zewnętrznych źródeł
- ✓ Integracja różnych źródeł danych w aplikacjach Pythonowych
- ✓ Narzędzia do monitorowania i optymalizacji wydajności aplikacji Pythonowych

6

MODUŁ 6

PySpark

Przetwarzanie Big Data w PySpark: praca z RDD i DataFrame, lazy evaluation, skalowanie w środowiskach rozproszonych i optymalizacja przetwarzania równoległego.

ZAKRES MODUŁU

- ✓ Podstawy pracy z RDD i DataFrame'ami w PySpark
- ✓ Lazy evaluation i Transformations w PySpark
- ✓ Skalowanie aplikacji PySpark w środowiskach rozproszonych
- ✓ Techniki zarządzania pamięcią i optymalizacji przetwarzania w PySpark
- ✓ Podstawy równoległości i przetwarzania równoległego w PySpark
- ✓ Implementacja projektów Big Data z wykorzystaniem PySpark

7

MODUŁ 7

Streaming processing (Apache Kafka)

Projektowanie systemów przetwarzania strumieniowego w Apache Kafka: konfiguracja klastrów, zarządzanie topikami, producentami i konsumentami oraz monitorowanie wydajności.

ZAKRES MODUŁU

- ✓ Konfiguracja i zarządzanie klastrami Apache Kafka
- ✓ Tworzenie i zarządzanie topikami, producentami i konsumentami w Kafka
- ✓ Projektowanie systemów przetwarzania strumieniowego o wysokiej wydajności
- ✓ Monitorowanie i diagnostyka wydajności systemów Kafka
- ✓ Techniki skalowania i zarządzania obciążeniem w środowiskach strumieniowych
- ✓ Architektura rozproszonych systemów przetwarzania strumieniowego

8

MODUŁ 8

Tworzenie i sterowanie przepływem danych (ETL i ELT)

Projektowanie procesów ETL/ELT: transformacja danych z wielu źródeł, automatyzacja i monitorowanie oraz integracja z hurtowniami i jeziorami danych.

ZAKRES MODUŁU

- ✓ Projektowanie procesów ETL do przetwarzania danych z różnych źródeł
- ✓ Techniki transformacji danych w celu uzyskania optymalnych rezultatów
- ✓ Automatyzacja i monitorowanie procesów ETL w dużych organizacjach
- ✓ Implementacja narzędzi ETL w różnych środowiskach technologicznych
- ✓ Walidacja i czyszczenie danych przed ich ładowaniem do baz danych
- ✓ Integracja procesów ETL z hurtowniami danych i jeziorami danych

9

MODUŁ 9

Data Warehouses & Lakes

Projektowanie i implementacja hurtowni oraz jezior danych w chmurze: integracja źródeł, zarządzanie i optymalizacja przechowywania pod raporty i analizy biznesowe.

ZAKRES MODUŁU

- ✓ Projektowanie i implementacja schematów hurtowni i jeziora danych
- ✓ Integracja różnych źródeł danych z hurtowniami i jeziorami danych
- ✓ Metody zarządzania hurtowniami i jeziorami danych w chmurze
- ✓ Rola hurtowni i jezior danych w tworzeniu raportów i analiz biznesowych
- ✓ Optymalizacja przechowywania danych w hurtowniach i jeziorach dla lepszej wydajności

10

MODUŁ 10

Orkiestracja zadań (Apache Airflow)

Automatyzacja i orkiestracja przepływów danych w Apache Airflow: konfiguracja środowiska, definiowanie DAGs, integracja z innymi narzędziami i monitorowanie.

ZAKRES MODUŁU

- ✓ Konfiguracja i uruchamianie środowiska Apache Airflow
- ✓ Definiowanie i zarządzanie DAGs dla przepływów pracy
- ✓ Automatyzacja złożonych przepływów danych za pomocą Airflow
- ✓ Integracja Airflow z innymi narzędziami i systemami danych
- ✓ Monitorowanie i zarządzanie przepływami pracy w Airflow
- ✓ Użycie Airflow do zarządzania projektami danych w organizacji

11

MODUŁ 11

Databricks

Efektywne przetwarzanie danych w środowisku Databricks: tworzenie i wdrażanie modeli uczenia maszynowego, integracja z narzędziami analitycznymi oraz współpraca zespołowa.

ZAKRES MODUŁU

- ✓ Efektywne przetwarzanie danych w środowisku Databricks
- ✓ Tworzenie i wdrażanie modeli uczenia maszynowego w Databricks
- ✓ Testowanie i walidacja modeli w chmurze z użyciem Databricks
- ✓ Integracja Databricks z innymi narzędziami analitycznymi i platformami danych
- ✓ Narzędzia do współpracy zespołowej dostępne w Databricks
- ✓ Zarządzanie projektami danych w środowisku chmurowym Databricks

12

MODUŁ 12

Snowflake

Konfiguracja i zarządzanie środowiskiem Snowflake: architektura oddzielająca przetwarzanie od przechowywania, zarządzanie dużymi zbiorami danych oraz narzędzia analityczne.

ZAKRES MODUŁU

- ✓ Konfiguracja i zarządzanie środowiskiem Snowflake
- ✓ Kluczowe komponenty i architektura Snowflake
- ✓ Oddzielenie przetwarzania od przechowywania danych w Snowflake
- ✓ Zarządzanie dużymi zbiorami danych w Snowflake
- ✓ Czyszczenie i walidacja danych w Snowflake
- ✓ Narzędzia analityczne dostępne w Snowflake

13

MODUŁ 13

Data Mesh

Koncepcja i zasady Data Mesh: decentralizacja zarządzania danymi, definiowanie i zarządzanie domenami danych oraz utrzymanie spójności i jakości w modelu rozproszonym.

ZAKRES MODUŁU

- ✓ Koncepcja i zasady działania modelu Data Mesh
- ✓ Planowanie i implementacja decentralizacji zarządzania danymi
- ✓ Definiowanie i zarządzanie domenami danych w organizacji
- ✓ Utrzymanie spójności i jakości danych w rozproszonym modelu zarządzania
- ✓ Technologie wspierające implementację Data Mesh
- ✓ Wyzwania w implementacji Data Mesh i strategię ich rozwiązywania

Sfinansuj kurs tak, jak Ci wygodnie

W infoShare Academy oferujemy różne formy sfinansowania kursu — bez względu na to, który pakiet wybierzesz. Sprawdź, która forma jest dla Ciebie najlepsza!

01 Raty PayU 0%

Najszybszy sposób na rozłożenie płatności na niskie, comiesięczne raty. Do wyboru 5, 10 lub 15 rat 0% albo 24, 36, a nawet 50 rat nisko oprocentowanych.

02 Płatność jednorazowa

Płacisz i zaczynasz naukę. Jednorazowa płatność często wiąże się z dużym rabatem — zapytaj o niego nasze konsultantki.

03 Płatność ratałna

Możliwość rozłożenia płatności na wygodne raty wewnątrz Akademii. Skontaktuj się z nami, aby poznać szczegółowy harmonogram płatności.

04 Dofinansowanie z Urzędu Pracy

Warunkiem jest status osoby bezrobotnej oraz gwarancja zatrudnienia od przyszłego pracodawcy. Wniosek składasz osobiście w swoim Urzędzie Pracy.

05 Baza Usług Rozwojowych (BUR)

Nawet do 80% dofinansowania na kursy w infoShare Academy — dla pracowników mikro, małych i średnich firm, osób prowadzących działalność oraz sektora publicznego i NGO.

06 Pożyczki na kształcenie

Nieoprocentowana pożyczka spłacana do 36 miesięcy, finansuje do 100% kosztów (maks. 75 000 zł). Możliwe umorzenie od 25% do 50% kwoty.

07 Krajowy Fundusz Szkoleniowy (KFS)

Inicjatywa ze środków Funduszu Pracy wspierająca pracodawców inwestujących w kształcenie ustawiczne pracowników oraz własne.

Stawiamy na jakość

2015

uczymy IT
od tego roku

9970+

przeszkolonych
osób

330+

trenerów
w naszej kadrze

Łączymy kompetencje techniczne z praktyką biznesową, a każdy kurs prowadzą trenerzy pracujący na co dzień w branży. Dzięki temu uczysz się tego, co realnie przydaje się w pracy.



MASZ PYTANIA?

Sylwia Liedtke

Doradczyni ds. kursów

+48 730 822 825 · sylwia.liedtke@infohareacademy.com