

# Szkolenie Databricks

## Opis

Szkolenie Databricks ma na celu wprowadzenie i zapoznanie uczestników z środowiskiem pracy, jakim jest Databricks. Uczestnicy dostaną dostęp do platformy, a w ramach szkolenia nauczą się tworzyć klastry obliczeniowe i notebooki, jak również korzystać z hurtowni danych SQL i automatyzować procesy.

**Grupa docelowa:** Analitycy danych, Inżynierowie Danych, Data Scientist

**Środowisko pracy:** Przygotowanie środowiska pracy na jednym z dostawców chmury: Microsoft Azure, Amazon Web Services (wkrótce), Google Cloud Platform (wkrótce)

**Wymagana wiedza:** dobra znajomość języka Python, opcjonalnie – znajomość języka Scala, dobra znajomość SQL

## Czego Cię nauczymy

### Wprowadzenie do Databricks

Nauczysz się korzystać z platformy Databricks, zdobywając wiedzę na temat struktury kosztów i efektywnego zarządzania zasobami. Tworzenie podstawowego klastra i uruchamianie pierwszego notebooka pozwoli Ci szybko zintegrować się z środowiskiem pracy.

### Zaawansowane Techniki Notebooków

Opanujesz zaawansowane techniki pracy z notebookami, w tym tworzenie interaktywnych widgetów i importowanie modułów. Dzięki temu będziesz w stanie efektywnie organizować kod, tworzyć interaktywne analizy i korzystać z funkcji importowania kodu z innych notebooków.

### Zarządzanie Klastrami i Zasobami

Zdobędziesz umiejętności zarządzania klastrami, w tym konfigurację polityk, autoskalowanie, korzystanie z interfejsu Spark UI oraz zarządzanie bibliotekami i zmiennymi środowiskowymi. Terminal stanie się dla Ciebie narzędziem do zaawansowanego zarządzania środowiskiem pracy.

### Automatyzacja Procesów i Praca z Danymi

Poznasz techniki automatyzacji poprzez ustawianie harmonogramów notebooków i wprowadzanie workflow. Zrozumiesz, jak efektywnie przekazywać dane pomiędzy zadaniami, co znacząco usprawni i zoptymalizuje procesy pracy. Dzięki temu nabędziesz umiejętność skutecznego wykorzystania Databricks do automatyzacji powtarzalnych zadań i pracy z danymi w hurtowni SQL za pomocą narzędzia Data Explorer..

# Program szkolenia

## 1. Wprowadzenie

- Wprowadzenie Omówienie kosztów
- Utworzenie podstawowego klastra
- Uruchomienie pierwszego notebooka

## 2. Notebooki

- Tworzenie widgetów
- Importowanie modułów/innych notebooków
- Moduł dbutils
- Korzystanie z innych języków
- Dostęp do danych

## 3. Cluster

- Polityki
- Autoskalowanie
- Spark UI
- Biblioteki i zmienne środowiskowe

- Terminal

## 4. Automatyizacja

- Ustawienie harmonogramu notebooka
- Wprowadzenie do workflow
- Przekazywanie danych pomiędzy zadaniami

## 5. SQL Warehouse + Data Explorer

- Tworzenie hurtowni danych
- Automatyizacja pracy z danymi za pomocą SQL
- Struktury danych (skupienie się na Delta)
- Dostęp do danych za pomocą Data Explorer

## **Czas trwania**

3 dni | 24 godziny zajęć

## **Certyfikat**

Uczestnicy szkolenia otrzymują imienne certyfikaty sygnowane przez Expose Sp. z o.o.

## **Cena szkolenia**

2 990 PLN netto (VAT 23%) za osobę (szkolenie grupowe)

## **Cena szkolenia zawiera**

- ✔ zapewnienie autorskich materiałów szkoleniowych dla uczestników szkolenia
- ✔ wystawienie certyfikatów po zakończonym szkoleniu
- ✔ rekomendacje dla uczestników szkolenia w zakresie dalszej pracy w obszarze szkolenia
- ✔ pakiet konsultacji z wykładowcą po zakończonym szkoleniu w razie jakichkolwiek niejasności przez okres 3 miesięcy
- ✔ całodzienny serwis kawowy oraz lunch