

APACHE SPARK



Kategoria	Czas trwania	Termin	Cena
DATA	16h / 2 dni	ustalamy indywidualnie	ustalamy indywidualnie

Program szkolenia:

Poniżej przedstawiamy przykładowy program szkolenia, który może zostać zmodyfikowany zgodnie z oczekiwaniami oraz poziomem grupy szkoleniowej. Przed przygotowaniem docelowego programu szkolenia, przeprowadzamy rozmowę techniczną, w której bierze udział trener oraz osoba techniczna lub cały zespół developerów reprezentujący klienta, w celu ustalenia szczegółów szkolenia.

Dzień 1: Wprowadzenie do Apache Spark i podstawy przetwarzania danych

Wprowadzenie do Apache Spark

- Historia i rozwój Apache Spark
- Architektura i główne komponenty (RDD, DataFrame, Spark SQL)

Instalacja i konfiguracja środowiska

- Instalacja Apache Spark i zależności
- Konfiguracja środowiska pracy (Standalone, Hadoop, AWS)

Podstawy przetwarzania danych w Apache Spark

- Praca z plikami: JSON, CSV, XML, TXT, Parquet, AVRO
- Omówienie czym jest Transformacja, a Akcja (Lazy evaluation)

Dzień 2: Zaawansowane techniki i praktyczne zastosowania

Zaawansowane przetwarzanie danych z użyciem DataFrame i Spark SQL

- Tworzenie i zarządzanie DataFrame
- Użycie Spark SQL do zapytań na dużych zbiorach danych

Transformacja danych

- Sposoby sortowania, grupowania i filtrowania danych
- Wykonywanie transformacji danych poprzez funkcje map, flatMap, UDF
- Funkcje okna/analityczne

Warsztat: Przetwarzanie i analiza danych z użyciem DataFrame

- Implementacja operacji na DataFrame i zapytań SQL
- Analiza dużych zbiorów danych z użyciem Spark SQL

Optymalizacja i tuning wydajności

- Techniki optymalizacji zapytań i działania Spark
- Zarządzanie pamięcią i rozdzielanie zasobów
- Partycjonowanie i zapis danych

Deployowanie aplikacji Apache Spark

- Przygotowanie i eksportowanie aplikacji Spark
- Wdrożenie aplikacji w środowisku produkcyjnym

KONTAKT


Jesteś zainteresowany kursem wieczorowym lub
dedykowanym szkoleniem dla Twojej firmy?

Skontaktuj się z Przemkiem!



PRZEMYSŁAW WOŁOSZ

Key Account Manager

 (+48) 730 830 801

 przemyslaw.wolosz@infoShareAcademy.com