Kod szkolenia: **SPARK/ANA**

Tytuł szkolenia: **Analiza danych z użyciem Apache Spark**

Data analysis with Apache Spark

Dni: 2

Opis:

Adresaci szkolenia

Osoby pracujące z danymi chcące pozyskać umiejętności pozwalające na analizę dużych zbiorów danych przy użyciu Apache Spark.

Cel szkolenia

Celem szkolenia jest zdobycie praktycznych umiejętności i wiedzy pozwalających na wykonywanie analiz dużych zbiorów danych z wykorzystaniem Apache Spark. Uczestnicy szkolenia w trakcie ćwiczeń zapoznają się z problemami przetwarzania, czyszczenia oraz eksploracyjnej analizy danych, a także zagadnieniami pokrewnymi, jak chociażby wykorzystaniem algorytmów uczenia maszynowego na przygotowanym zbiorze.

Mocne strony szkolenia

Wprowadzenie do świata analizy danych Big Data. Koncentracja na użytkowaniu Sparka nie przytłoczy mnogością szczegółów technicznych. Duży nacisk na część warsztatową i pisanie kodu pozwoli na szybkie wykorzystanie zdobytych umiejętności w praktyce. Wykorzystanie głównie Spark SQL pozwoli na intuicyjną pracę z dużymi zbiorami danych.

Wymagania

Podstawowa znajomość Pythona oraz SQLa

Specjalne wymagania techniczne

Ubuntu, Docker lub VirtualBox

Parametry szkolenia

2 * 8 godzin (2 * 7 godzin netto) warsztatów i wykładów. Zdecydowana przewaga warsztatów. Wielkość grupy: 8-10 osób.

sages

Program szkolenia:

1. Wprowadzenie do Apache Spark
    - Architektura
    - Moduły
    - Miejsce w ekosystemie Big Data
2. Środowisko pracy
    - Jupyter
    - Przegląd możliwości i udogodnień
3. Spark
    - Spark Context i Spark Session
    - RDD
    - Akcje i transformacje
    - DataFrame
    - Źródła danych
    - Składnia Spark SQL
    - Statystyki zmiennych
    - Grupowanie i agregacja danych
    - Funkcje analityczne i okienne
4. Wizualizacja danych
    - Podstawy wizualizacji
    - Typy wykresów
    - Wizualizacja dużych zbiorów danych
5. Spark ML
    - Wektory gęste i rzadkie
    - Przekształcanie zbioru do postaci wektorowej
    - Przegląd dostępnych transformacji zmiennych
    - Przegląd dostępnych algorytmów uczenia maszynowego
        - modele klasyfikacyjne
        - modele regresyjne
        - algorytmy klastrowania
    - Przetwarzanie potokowe (pipeline)

Sages sp. z o.o., ul. Nowogrodzka 62c, 02-002 Warszawa
tel. +48 22 203 56 00 fax: +48 22 203 56 01
e-mail: biuro@sages.com.pl web: www.sages.com.pl