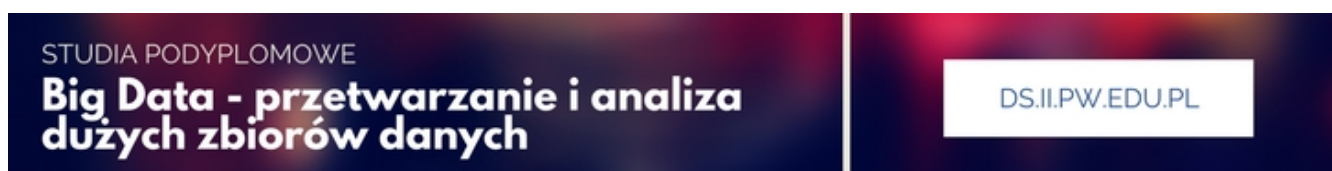


Kod szkolenia: **HADOOP/ANA**

Tytuł szkolenia: **Big Data dla analityków**

Dni: 2

Opis:



Adresaci szkolenia:

Analitycy danych oraz programiści, którzy chcą rozpocząć swoją przygodę z analizą dużych zbiorów danych.

Cel szkolenia:

Przekrojowe szkolenie mające na celu zapoznanie się z narzędziami przeznaczonymi dla analityka big data. Szkolenie skupia się na płynnym wejściu w podstawy każdego narzędzia, tak aby analityk danych mógł w przyszłości bez problemu poruszać się po ekosystemie Hadoop.

Mocne strony szkolenia:

Zapoznanie z wieloma narzędziami i językami programowania, szkolenie ma na celu pokazanie jak łatwo można analizować dane bez użycia konsoli i narzędzi IDE.

Wymagania:

Podstawy SQL, podstawowa umiejętność programowania, najlepiej w: Python, R lub Java

Parametry szkolenia:

Wielkość grupy: maks. 8-10 osób.

Program szkolenia:

1. Wprowadzenie do Big Data
 - I. Definicja
 - II. Geneza Big Data
 - III. Klasyfikacja problemu
 - IV. Przegląd ekosystemu Hadoop
 - V. Dystrybucje Big Data
2. Wprowadzenie do MapReduce na przykładzie platformy Hadoop
 - I. Architektura
 - II. HDFS and YARN
 - III. Map Reduce Framework
 - IV. Map Reduce Streaming
 - V. Warsztat Map Reduce
3. Wprowadzenie do analizy danych na przykładzie Hive
 - I. Architektura
 - II. Tryby pracy
 - III. Typy danych
 - IV. Składnia
 - V. Formaty danych
 - VI. Wydajność
 - VII. Warsztat Hive
 - VIII. Ładowanie danych
 - IX. Zapytania
 - X. Partycjonowanie
 - XI. Formaty danych
 - XII. Silnik przetwarzania
 - XIII. Funkcje okienkowe
 - XIV. UDF
4. Wprowadzenie do Spark
 - I. Architektura
 - II. Ekosystem Spark
 - III. Transformacje i akcje RDD
 - IV. Dataframe i SparkSQL
 - V. Spark ML
 - VI. Warsztat
 - o RDD
 - o Spark SQL
 - o Spark ML z wykorzystaniem wybranych algorytmów

