

Big Data -

podstawy analizy danych opartej o duże zbiory danych

Cele szkolenia

Celem szkolenia jest poznanie jest zapoznanie się z podstawami analizy danych opartych na dużych zbiorach danych. Szkolenie jest przeglądem technik i narzędzi dostępnych na rynku, pozwalających na optymalną pracę z Big Data.

Umiejętności

Dzięki szkoleniu uczestnik:

- Dowie się, co to jest Big Data i jak wygląda architektura systemów Big Data
- Nauczy się zbierać i składować dane
- Pozna sposoby przetwarzania dużych danych
- Będzie potrafił wizualizować dane
- Dowie się, czym jest skalowalność i wydajność systemów Big Data
- Oraz jak wygląda bezpieczeństwo danych

Profil uczestników

Szkolenie przeznaczone jest dla analityków danych, architektów struktur danych, programistów oraz osób biorących udział w projektach związanych z Big Data.

Przygotowanie uczestników

Od uczestników szkolenia wymagana są podstawowej wiedzy na temat danych oraz ich analizy.

Szczegółowy program szkolenia

1. Sesja wprowadzająca do big data
 - Definicja big data i jej znaczenie w dzisiejszym świecie.
 - Wyzwania i możliwości związane z analizą dużych zbiorów danych.
 - Przykłady zastosowań big data w różnych dziedzinach.
2. Architektura systemów big data
 - Przegląd popularnych technologii i narzędzi do przechowywania, przetwarzania i analizy dużych zbiorów danych, takich jak Hadoop, Spark, NoSQL, Apache Kafka itp.
 - Porównanie różnych modeli architektonicznych, takich jak przetwarzanie wsadowe (batch processing) i przetwarzanie strumieniowe (stream processing).

3. Zbieranie i składowanie danych
 - Strategie zbierania danych i ich przechowywanie.
 - Bazy danych relacyjne i nierelacyjne.
 - Skalowalne systemy przechowywania danych, takie jak Hadoop Distributed File System (HDFS) i Amazon S3.
4. Przetwarzanie danych
 - Wprowadzenie do przetwarzania wsadowego i przetwarzania strumieniowego.
 - Wykorzystanie narzędzi takich jak Apache Spark do przetwarzania dużych zbiorów danych.
 - Techniki transformacji i wstępnego przetwarzania danych.
5. Analiza i wizualizacja danych
 - Narzędzia i techniki analizy dużych zbiorów danych.
 - Wprowadzenie do algorytmów uczenia maszynowego i technik eksploracji danych.
 - Wykorzystanie narzędzi do wizualizacji danych, takich jak Tableau lub Python z bibliotekami takimi jak Matplotlib i Seaborn.
6. Skalowalność i wydajność
 - Wyzwania związane z skalowalnością i wydajnością systemów big data.
 - Optymalizacja zapytań i przetwarzania równoległego.
 - Zarządzanie zasobami i monitorowanie wydajności.
7. Bezpieczeństwo i prywatność danych
 - Zagadnienia związane z bezpieczeństwem i prywatnością danych w kontekście big data.
 - Metody szyfrowania, uwierzytelniania i kontroli dostępu.
8. Studium przypadku i praktyczne ćwiczenia
 - Przeprowadzenie studium przypadku wykorzystującego narzędzia i techniki omówione podczas szkolenia.

 - Praktyczne ćwiczenia, które pozwolą uczestnikom na zdobycie praktycznych umiejętności w zakresie pracy z dużymi zbiorami danych.

Metoda realizacji szkolenia

Szkolenie realizowane jest w formie naprzemiennie następujących po sobie wykładów oraz ćwiczeń praktycznych. Szkolenie łączy w sobie wiedzę merytoryczną z praktycznymi przykładami jej wykorzystania w środowisku pracy.

Liczba dni, liczba godzin szkoleniowych

2 dni, 16 godzin szkoleniowych

Ścieżka rozwoju po szkoleniu

Big Data i Data Science - zastosowanie algorytmów sztucznej inteligencji