Szkolenie: Cloudera

Cloudera Data Analyst Training: Using Pig, Hive, and Impala with Hadoop

FORMA SZKOLENIA	MATERIAŁY SZKOLENIOWE	CENA	CZAS TRWANIA
Stacjonarne	Cyfrowe	2180 EUR NETTO*	4 dni
Stacjonarne	Tablet CTAB	2330 EUR NETTO*	4 dni
Metoda dlearning	Cyfrowe	2180 EUR NETTO*	4 dni
Metoda dlearning	Tablet CTAB	2180 EUR NETTO*	4 dni

* (+VAT zgodnie z obowiązującą stawką w dniu wystawienia faktury)

LOKALIZACJE

Kraków - ul. Tatarska 5, II piętro, godz. 9:00 - 16:00
Warszawa - ul. Bielska 17, godz. 9:00 - 16:00

DOSTĘPNE TERMINY

2019-07-22 | 4 dni | Warszawa
2019-07-22 | 4 dni | Warszawa
2019-09-16 | 4 dni | Warszawa
2019-09-16 | 4 dni | Warszawa
2019-10-28 | 4 dni | Warszawa
2019-10-28 | 4 dni | Warszawa

Cel szkolenia:

Czterodniowy kurs **Cloudera Data Analyst Training: Using Pig, Hive, and Impala with Hadoop** skupia się na zagadnieniach i technologiach przydatnych w pracy każdego analityka tj. **Apache Pig**, **Hive** i **Cloudera Impala**, które umożliwią uczestnikom wykorzystanie tradycyjnych podejść i metod analitycznych praktywanych dotychczas na wykorzystanie ich w technologii **BigDATA**. W trakcie kursu **Cloudera Data Analyst** prezentowane są profesjonalne narzędzia pozwalające na uzyskanie dostępu, zmianę, transformacje i analizę skomplikowanych struktur danych umieszczonych na klastrze **Hadoop**, przy użyciu języków skryptowych zawierających podobieństwa do SQL.

Apache Hive pozwala rzucić nowe spojrzenie na skomplikowane duże struktury danych, co przekłada się na możliwość wykonania na nich niezbędnej analityki. Narzędzie idealne dla analityków, administratorów oraz wszystkich innych tych, którzy nie posiadają wiedzy i doświadczenia nt języka programowania **Java**. **Apache Pig** dodaje możliwość użycia prostego i łatwo przyswajalnego języka skryptowego do klastra hurtownii danych Hadoop. **Cloudera Impala** z kolei to narzędzie ułatwiające analitykę na danych w hurtownii w czasie rzeczywistym, zbliżonym do tego jaki spotyka się w relacyjnych bazach danych, przy wykorzystaniu natywnego języka **SQL**.

W trakcie szkolenia prowadzony jest wykład przeplatany dyskusją, burzą mózgów, wykonywaniem

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Bielska 17, 02-394 Warszawa, tel.: (22) 417 41 70
Alior Bank Spółka Akcyjna
Oddział w Krakowie
Nr konta: 19 2490 0005 0000 4530 5733 7313

ćwiczeń praktycznych, uczestnicy poruszać będą m.in. takie tematy w ramach technologii około **Hadoop** jak:

- Funkcjonalności narzędzi Pig, Hive oraz Impala, pozwalające na zbieranie danych, zapisywanie wyników i analitykę
- Podstawowa wiedza nt Apache Hadoop i jego narzędzi oraz procesu ETL (extract, transform, load)
- Jak Pig, Hive, i Impala pozwolą podnieść wydajność dla typowych i codziennych zadań analitycznych
- Łączenie różnych zestawów danych, aby uzyskać cenne i wartościowe wartości biznesowe i wyciągać wnioski
- Wykonywanie złożonych zapytań na zbiorach danych

Plan szkolenia:

- Wprowadzenie
- Podstawy nt Hadoop
  - Dlaczego Hadoop?
  - Przegląd technologii
  - HDFS to miejsce gdzie trzymamy nasze dane
  - Rozproszone procesowanie danych z wykorzystaniem: YARN, MapReduce i Spark
  - Procesowanie i analiza danych: Pig, Hive i Impala
  - Ładowanie danych z wykorzystaniem narzędzia: Sqoop
  - Inne narzędzia Hadoop
  - Opiś środowiska szkoleniowego, na którym będziemy pracować w trakcie kursu
- Wprowadzenie do Pig
  - Czym jest Pig?
  - Funkcjonalności narzędzia Pig
  - Przypadki użycia narzędzia Pig
  - Praca z narzędziem Pig
- Podstawowa analiza danych z wykorzystaniem narzędzia Pig
  - Składnia języka Pig Latin
  - Ładowanie danych
  - Proste typy danych
  - Definicje pól
  - Reprezentacja danych wynikowych
  - Podgląd schematu
  - Filtrowanie i sortowanie danych

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Bielska 17, 02-394 Warszawa, tel.: (22) 417 41 70
Alior Bank Spółka Akcyjna
Oddział w Krakowie
Nr konta: 19 2490 0005 0000 4530 5733 7313

- Powszechnie wykorzystywane funkcje
- Procesowanie skomplikowanych/złożonych danych z wykorzystaniem narzędzia Pig
  - Formaty przechowywania danych
  - Złożone/zagnieżdżone typy danych
  - Grupowanie danych
  - Przegląd wbudowanych funkcji możliwych do użycia w przypadku złożonych typów danych
  - Iterowanie i dostęp do zgrupowanych danych
- Operacje na wielu zbiorach danych z wykorzystaniem narzędzia Pig
  - Techniki łączenia zbiorów danych
  - Łączenie danych z wykorzystaniem narzędzia Pig
  - Inne typy łączenia danych (cross, union)
  - Podział zbioru danych na mniejsze porcje
- Rozwiązywanie problemów i optymalizacja w Pig
  - Rozwiązywanie problemów
  - Podejście do logów
  - Wykorzystanie interfejsów przeglądarkowych w narzędziach technologii Hadoop
  - Próbkowanie danych, usuwanie błędów
  - Podejście do wydajności
  - Plan zapytań i jego przydatność
  - Dobre praktyki, wskazówki i podpowiedźi w osiąganiu lepszej wydajności przy wykonywaniu zadań z Pig
- Wprowadzenie do Hive oraz Impala
  - Czym jest Hive?
  - Czym jest Impala?
  - Jak wygląda schemat, struktura oraz przechowywanie danych
  - Porównanie Hive do tradycyjnych relacyjnych baz danych Databases
  - Przypadki użycia Hive
- Wykonywanie zapytań w Hive i Impala
  - Bazy danych i tabele
  - Podstawy języka pisania zapytań w Hive i Impala
  - Typy danych
  - Różnica języka SQL pomiędzy Hive i Impala
  - Wykorzystanie interfejsu przeglądarkowego Hue do wykonywania zapytań
  - Przypadki użycia Impala Shell
- Zarządzanie danymi
  - Przechowywanie danych

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Bielska 17, 02-394 Warszawa, tel.: (22) 417 41 70
Alior Bank Spółka Akcyjna
Oddział w Krakowie
Nr konta: 19 2490 0005 0000 4530 5733 7313

- Tworzenie baz danych i tabel
- Ładowanie danych
- Zmiana schematu baz danych i tabel
- Uproszenie zapytań z wykorzystaniem widoków
- Zapisywanie wyników zapytań

- Przechowywanie danych i wydajność
  - Partycjonowanie tabel
  - Wybór formatu plikowego
  - Zarządzanie meta danymi
  - Kontrola dostępu do danych
- Analizy z wykorzystaniem narzędzi Hive i Impala
  - Łączenie zbiorów danych
  - Popularne wbudowane funkcje
  - Agregacja i użycie okienkowości
- Praca z narzędziem Impala
  - Jak Imapala uruchamia zadane zapytania
  - Rozszerzenie narzędzia o tzw. User-Defined Functions (UDF)
  - Co zrobić aby było wydajniej?
- Analiza tekstu i złożonych typów danych z wykorzystaniem narzędzia Hive
  - Złożone wartości w Hive
  - Użycie wyrażeń regularnych w Hive
  - Analiza nastroju oraz rekomendacje z użyciem N-Gram
  - Podsumowanie
- Optymalizacja Hive
  - Podejście do zrozumienia wydajności zapytania
  - Kontrola planu zapytania
  - Bucketing
  - Zakładanie indeksów
- Rozszerzenie Hive
  - SerDes
  - Transformacja danych z wykorzystaniem skryptów (python, perl, ...)
  - User-Defined Functions (UDF)
  - Parametryzacja zapytań
- Wybór najlepszego narzędzia dla danego zadania
  - Porównanie MapReduce, Pig, Hive, Impala i relacyjnych baz danych
  - Które narzędzie wybrać?

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Bielska 17, 02-394 Warszawa, tel.: (22) 417 41 70
Alior Bank Spółka Akcyjna
Oddział w Krakowie
Nr konta: 19 2490 0005 0000 4530 5733 7313

- Podsumowanie

Wymagania:

- Szkolenie dedykowane jest dla analityków, secjalistów BI, deweloperów, architektów systemowych i administratorów baz danych.
- Znajomość języka SQL oraz podstawy Linux-a, to umiejętności które bedą przydatne do uczestnictwa w kursie.
- Ponadto wiedza nt przynajmniej jednego z wymienionych języków skryptowych: Bash scripting, Perl, Python, Ruby, byłaby bardzo pomocna na zajęciach ale nie jest wymagana.
- Wiedza w zakresie technologii Apache Hadoop nie jest wymagana.

Poziom trudności

Certyfikaty:

Uczestnicy szkolenia otrzymają **certyfikat** ukończenia kursu z patronatem i autoryzacją **Cloudera**.

Prowadzący:

Certyfikowany instruktor Cloudera.

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Bielska 17, 02-394 Warszawa, tel.: (22) 417 41 70

Alior Bank Spółka Akcyjna
Oddział w Krakowie
Nr konta: 19 2490 0005 0000 4530 5733 7313