Szkolenie: Cloudera
Developer Training For Spark and 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

2019-06-10 | 4 dni | Warszawa
2019-06-10 | 4 dni | Warszawa
2019-11-04 | 4 dni | Warszawa
2019-11-04 | 4 dni | Warszawa
2019-12-02 | 4 dni | Warszawa
2019-12-02 | 4 dni | Warszawa

Cel szkolenia:

Developer Training For Spark and Hadoop to 4-dniowe praktyczne szkolenie przedstawia uczestnikom kluczowe założenia teoretyczne i rozwija umiejętności przechwytywania i procesowania danych na platformie **Hadoop**, używając najaktualniejszych narzędzi i technik. Poprzez wykorzystanie ekosystemu aplikacji **Hadoop**, takich jak: **Spark**, **Hive**, **Flume**, **Sqoop** i **Impala**, szkolenie to doskonale przygotowuje do mierzenia się z praktycznymi wyzwaniami, z jakimi borykają się programiści **Hadoop**. Uczestnicy uczą się, jak poprawnie wybrać odpowiednie narzędzia w konkretnych sytuacjach oraz nabywają praktycznego doświadczenia w ich używaniu.

Poprzez prowadzone przez instruktora dyskusje i interaktywne ćwiczenia praktyczne, uczestnicy poznają technologię **Apache Spark** oraz uczą się, jak ją integrować z całym ekosystemem aplikacji **Hadoop**. Wiedza ta dostarcza im odpowiedzi na pytania:

- Jak dane są dystrybuowane, zapisywane i procesowanie na klastrze Hadoop?
- Jak używać Sqoop i Flume do przyjmowania danych na klaster?
- Jak przetwarzać dystrybuowane dane przy pomocy Apache Spark?

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

- Jak modelować strukturyzowane dane jako tabele w Impala i Hive?
- Jak wybrać najbardziej odpowiedni format zapisu danych dla różnych przypadków użycia?
- Jakie są dobre praktyki zapisu danych?

Plan szkolenia:

- Wprowadzenie do Hadoop i ekosystemu aplikacji Hadoop
  - Problemy z tradycyjnymi systemami dużej skali
  - Hadoop!
  - Ekosystem Hadoop
- Architektura Hadoop i HDFS
  - Rozproszone przetwarzanie na klastrze
  - Przechowywanie danych: Architektura HDFS
  - Przechowywanie danych: Użycie HDFS
  - Zarządzanie zasobami: Architektura YARN
  - Zarządzanie zasobami: Praca z YARN
- Import relacyjnych danych używając Apache Sqoop
  - Wprowadzenie do Sqoop
  - Podstawowe importy i eksporty
  - Ograniczanie wyników
  - Optymalizacja wydajności w Sqoop’s
  - Sqoop 2
- Wprowadzenie do Impala i Hive
  - Wprowadzenie do Impala i Hive
  - Kto używa Impala i Hive?
  - Porównanie Hive z tradycyjnymi bazami danych
  - Przypadki użycia Hive
- Modelowanie i zarządzanie danymi w Impala i Hive
  - Wstęp do zapisu danych
  - Tworzenie baz danych i tabel
  - Umieszczanie danych w tabelach
  - HCatalog
  - Cache'owanie metadanych w Impala
- Formaty plików
  - Wybór formatu plików
  - Wsparcie narzędzi Hadoop dla różnych formatów plików

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

- Schematy Avro
- Używanie Avro z Hive i Sqoop
- Ewolucja schematów Avro
- Kompresja

- Partycjonowanie danych
  - Wstęp do partycjonowania
  - Partycjonowanie w Impala i Hive
- Zbieranie danych używając Apache Flume
  - Co to jest Apache Flume?
  - Podstawowa architektura Flume
  - Flume Sources
  - Flume Sinks
  - Flume Channels
  - Konfiguracja Flume
- Podstawy Spark
  - Co to jest Apache Spark?
  - Wykorzystanie Spark Shell
  - RDDs (Resilient Distributed Datasets)
  - Programowanie funkcyjne w Spark
- Używanie RDDs w Spark
  - Bardziej szczegółowo o RDDs
  - Klucz-wartość Pair RDDs
  - MapReduce
  - Inne operacje na Pair RDD
- Tworzenie i wdrażanie aplikacji Spark
  - Aplikacje Spark VS Spark Shell
  - Tworzenie SparkContext
  - Budowanie aplikacji Spark (Scala i Java)
  - Uruchamianie aplikacji Spark
  - Web UI aplikacji Spark
  - Konfigurowanie ustawień Spark
  - Logowanie
- Programowanie współbieżne w Spark
  - Przypomnienie: Spark uruchamiany na klastrze
  - Partycje w RDD
  - Partycjonowanie RDD opartych na plikach

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

- HDFS i Data Locality
- Wykonywanie równoległych operacji
- Stages i Tasks

- Cache'owanie i trwały zapis danych w Spark
  - RDD Lineage
  - Wstęp do cache'owania
  - Rozproszony trwały zapis danych
- Powszechne wzorce w procesowaniu danych przy użyciu Spark
  - Popularne przypadki użycia Spark
  - Iteracyjne algorytmy w Spark
  - Przetwarzanie i analiza grafów
  - Uczenie maszynowe
  - Przykład: k-średnie
- Przegląd: Spark SQL
  - Spark SQL i SQL Context
  - Tworzenie DataFrames
  - Transformacje i wykonywanie zapytań na DataFrames
  - Zapis DataFrames
  - Porównanie Spark SQL z Impala
- Podsumowanie

Wymagania:

Kurs ten adresowany jest do programistów i inżynierów oprogramowania, którzy mają doświadczenie programistyczne. Przykłady w **Apache Spark** i ćwiczeniach praktycznych są przedstawiane w **Scala** i **Python**, stąd też umiejętność programowania w jednym z tych języków jest wymagana. Plan kursu zakłada również podstawową znajomość linii poleceń w **Linux**. Pomocna będzie też podstawowa wiedza z zakresu **SQL**. Natomiast uprzednia wiedza z zakresu **Hadoop** nie jest wymagana.

Poziom trudności

Certyfikaty:

Uczestnicy otrzymają **certyfikaty** uczestnictwa w szkoleniu sygnowane przez **Cloudera**.

Ponadto szkolenie to jest doskonałym sposobem rozpoczęcia przygotowań do **certyfikatu CCP: Data Engineer**. Jakkolwiek późniejsza nauka zagadnień poruszonych podczas szkolenie jest wymagana

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

przed podejściem do tego **egzaminu**.

Prowadzący:

Certyfikowany instruktor Cloudera.

Informacje dodatkowe:

Po zakończeniu kursu rekomendujemy, aby uczestnicy rozważyli uczestnictwo w szkoleniu: **Cloudera's Developer Training for Spark and Hadoop II: Advanced Techniques**, które opiera się na podstawach omawianych w niniejszym szkoleniu.

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