

Szkolenie: Cloudera  
Cloudera Administrator Training for Apache 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

2019-09-02 | 4 dni | Warszawa

2019-09-02 | 4 dni | Warszawa

2019-10-14 | 4 dni | Warszawa

2019-10-14 | 4 dni | Warszawa

## Cel szkolenia:

Czterodniowe szkolenie **Cloudera Administrator Training for Apache Hadoop** dedkowane administratorom pozwoli uczestnikom w pełni zapoznać się i zrozumieć wszystkie niezbędne kroki do obsługi i utrzymania klastra **Hadoop**. Od instalacji i konfiguracji przez tunowanie polegające na równoważeniu obciążenia i zmiany ustawień do osiągnięcia najlepszej wydajności i pojemności platformy. Szkolenie Cloudera to najlepsze rozwiązanie dla administratorów na co dzień pracujących z systemami **Hadoop**, oraz tych którzy chcieli by rozpocząć swoją przygodę z tą platformą technologiczną.

W trakcie szkolenia prowadzony jest wykład przeplatany dyskusją, burzą mózgow, wykonywaniem ćwiczeń praktycznych, uczestnicy poruszać będą m.in. takie tematy w ramach technologii około Hadoop jak:

- YARN, MapReduce i HDFS
- Przystosowanie prawidłowej konfiguracji sprzętowej i infrastruktury dla docelowego klastra
- Rekomendowana konfiguracja klastra i umieszczenie technologii Hadoop w aktualnie istniejącym centrum danych
- Jak załadować dane do klastra z dynamicznie generowanych plików za pomocą kolektora

FLUME, czy z RDBMS wykorzystując narzędzie Sqoop?

- Konfiguracja FairScheduler dla zachowania odpowiedniego podziału zasobów na klastrze w środowisku wielu użytkowników
- Najlepsze praktyki przygotowania i utrzymania środowiska Apache Hadoop do użycia w firmie
- Rozwiązywanie problemów, diagnozowanie, tunowanie Hadoop

## Plan szkolenia:

- Wprowadzenie
- Ogólnie o Apache Hadoop
  - Dlaczego Hadoop?
  - Kluczowe komponenty Hadoop
  - Podstawowe pojęcia
- HDFS
  - Funkcjonalności HDFS
  - Zapis i odczyt plików
  - Zawartość pamięci NameNode
  - Rzut okiem na bezpieczeństwo HDFS
  - Przypadki użycia przeglądarkowego interfejsu NameNode
  - Przypadki użycia Hadoop File Shell
- Ładowanie danych do HDFS
  - Przyjmowanie danych z zew. źródeł za pomocą kolektora Flume
  - Przyjmowanie danych z relacyjnych baz danych za pomocą narzędzia Sqoop
  - Interfejsy REST
  - Najlepsze praktyki ładowania danych
- YARN and MapReduce
  - Czym jest MapReduce?
  - Proste przypadki użycia MapReduce
  - Architektura klastra YARN
  - Zarządzanie zasobami
  - Przypadki odzyskiwania funkcjonalności po awarii
  - Przypadki użycia przeglądarkowego interfejsu YARN
  - MapReduce w wersji pierwszej
- Planowanie własnego klastra Hadoop
  - Poruszenie podstawowych kwestii nt planowanego klastra Hadoop
  - Jak wybrać/dobrać prawidłową konfigurację sprzętową w zależności od potrzeb
  - Połączenia sieciowe dla planowanego klastra

- Konfiguracja Node-ów
- Zarządzanie klastrem
- Instalacja i startowa konfiguracja klastra Hadoop
  - Możliwości wykonania wdrożenia
  - Instalacja Hadoop
  - Określenie konfiguracji Hadoop
  - Startowa/początkowa konfiguracja Hadoop
  - Startowa/początkowa konfiguracja YARN i MapReduce
  - Zarządzanie logowaniem
- Instalacja i konfiguracja Hive, Impala i Pig
  - Hive
  - Impala
  - Pig
- Klienci Hadoop
  - Co oznacza bycie klientem Hadoop?
  - Instalacja i konfiguracja klientów Hadoop
  - Instalacja i konfiguracja narzędzia Hue
  - Autentykacja i autoryzacja w Hue
- Cloudera Manager
  - Dlaczego Cloudera Manager?
  - Możliwości i funkcjonalności Cloudera Manager
  - Jaką wersję wybrać? Express czy Enterprise?
  - Toplogia Cloudera Manager
  - Instalacja Cloudera Manager
  - Instalacja klastra Hadoop przy użyciu Cloudera Manager
  - Przypadki użycia Cloudera Manager w kontekście administracyjnym – zarządzanie, utrzymanie, rozwój
- Zaawansowana konfiguracja klastra
  - Zaawansowane parametry konfiguracji
  - Konfiguracja portów dla poszczególnych elementów technologii Hadoop
  - Zarządzanie dostępnością Node-ów dla klastra
  - Konfiguracja skryptów Rack Awareness dla HDFS
  - Konfiguracja wysokiej dostępności dla HDFS - HA
- Bezpieczeństwo Hadoop
  - Co oznacza priorytet bezpieczeństwa w Hadoop?
  - Możliwości systemów bezpieczeństwa w Hadoop

- Czym jest Kerberos i gdzie można go wykorzystać?
- Jak zabezpieczyć swój klastr w oparciu o technologię Kerberos
- Zarządzanie i cykliczne uruchamianie zadań
  - Zarządzanie uruchomionymi zadaniami
  - Harmonogramowanie zadań
  - Konfiguracja FairScheduler
  - Zarządzanie harmonogramowaniem zadań w Impala
- Utrzymanie klastra
  - Sprawdzanie stanu działania HDFS
  - Kopiowanie danych pomiędzy różnymi klastrami
  - Dodawanie, usuwanie Node-ów do/z klastra
  - Procesy pozwalające zrównoważyć obciążenie i wydajność klastra
  - Podnoszenie wersji klastra
- Rozwiązywanie problemów i monitoring klastra
  - Podstawowy system monitoringu
  - Jak monitorować klastr?
  - Wspólne rozwiązywanie problemów z klastrem
  - Wspólne szukanie i określanie typowych błędów konfiguracji
- Podsumowanie

## Wymagania:

- Szkolenie jest dedykowane dla **administratorów systemów** oraz **menadżerów IT**, którzy posiadają podstawową wiedzę w zakresie technologii **Linux**. Wiedza w zakresie technologii **Apache Hadoop** nie jest wymagana.

## Poziom trudności



## Certyfikaty:

Uczestnicy szkolenia otrzymają **certyfikat ukończenia kursu** z patronatem i autoryzacją **Cloudera**. Ponadto szkolenie przygotowuje do **egzaminu certyfikacyjnego Cloudera Certified Administrator for Apache Hadoop (CAHA)** <https://www.cloudera.com/about/training.html>. **Certyfikacja** jest doskonałym wyróżnikiem, pomaga ustanowić się jako specjalistę w dziedzinie, zapewniając pracodawcom i klientom namacalny dowód swoich umiejętności i wiedzy.

## Prowadzący:

Certyfikowany instructor Cloudera.