

Kod szkolenia: **HADOOP/ADM**

Tytuł szkolenia: **Administracja Apache Hadoop**

Dni: **3**



Partner merytoryczny

Opis:

Adresaci szkolenia:

Szkolenie jest adresowane do architektów i administratorów dużych klastrów obliczeniowych opartych o Apache Hadoop, którzy chcą zarówno tworzyć jak i utrzymywać systemy, w których wolumen przetwarzanych danych ma najwyższy priorytet i przekracza możliwości tradycyjnych architektur i systemów takich jak relacyjne bazy danych czy nawet hurtownie danych.

Cel szkolenia:

Uczestnicy szkolenia zdobędą niezbędną wiedzę do samodzielnego budowania i utrzymywania dużych klastrów obliczeniowych opartych o Apache Hadoop. Zostaną omówione tematy związane z przygotowaniem samych maszyn, budowę klastra, instalację systemów operacyjnych, instalację poszczególnych komponentów środowiska Apache Hadoop, podstawy ich obsługi oraz zarządzania nimi podczas codziennej pracy.

Mocne strony szkolenia:

Program obejmuje zarówno ogólne wprowadzenie w tematykę jak i całościowe przedstawienie stosu produktowego wokół Apache Hadoop. Szkolenie jest unikalne, gdyż tematyka poruszana w jego trakcie nie jest wyczerpująco ujęta w dostępnej literaturze, a wiedza na ten temat jest mocno rozproszona. Program jest ciągle uaktualniany ze względu na szybki rozwój rozwiązań, których dotyczy szkolenie.

Wymagania:

Od uczestników wymagana jest podstawowa znajomość obsługi systemu Linux.

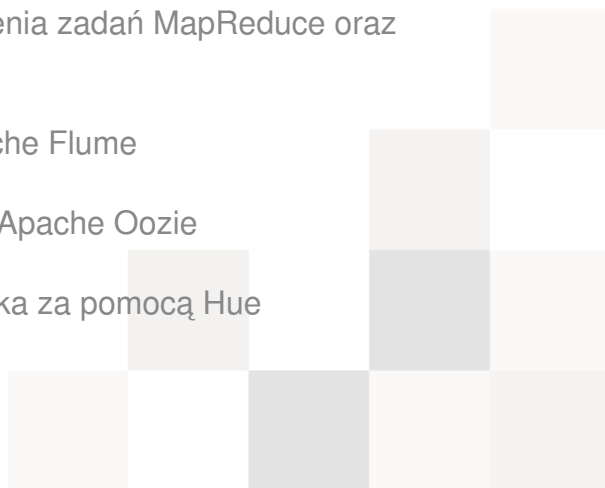
Parametry szkolenia:

3*8 godzin (3*7 netto) wykładów i warsztatów, z wyraźną przewagą warsztatów. W trakcie

warsztatów uczestnicy zbudują swój własny klaster obliczeniowy oraz nauczą się z niego korzystać i nim zarządza. Wielkość grupy: maks. 8-10 osób.

Program szkolenia:

1. Wprowadzenie do Big Data i ekosystemu Apache Hadoop
2. Organizacja infrastruktury i topologii klastra obliczeniowego
3. Budowa klastra obliczeniowego
 - Automatyczna instalacja systemów przez sieć (PXE)
 - Konfiguracja systemu operacyjnego
4. Zarządzanie i administracja Hadoop
 - Instalacja i konfiguracja HDFS
 - Korzystanie z HDFS z linii komend
 - Rozproszone kopiowanie danych za pomocą DistCP
 - Montowanie HDFS'a w systemie operacyjnym
 - Tworzenie Snapshotów i kopii bezpieczeństwa dla HDFS
 - Konfiguracja i zarządzanie YARN oraz MapReduce
 - Różnice pomiędzy podstawowymi typami plików
 - Kompresja danych
5. Zarządzanie i administracja Apache HBase
 - Instalacja na klastrze
 - Korzystanie z HBase Shell i narzędzi administracyjnych
 - Migracja danych z innych źródeł
 - Tworzenie i przywracanie kopii bezpieczeństwa
 - Monitorowanie i diagnostyka
6. Instalacja Hive oraz Pig do szybkiego tworzenia zadań MapReduce oraz optymalizacja za pomocą Apache Tez
7. Agregacja logów za pomocą narzędzia Apache Flume
8. Zarządzanie przepływami pracy za pomocą Apache Oozie
9. Dodawanie graficznego interfejsu użytkownika za pomocą Hue



10. Instalacja Apache Spark i integracja z YARN i HDFS
11. Automatyczna instalacja komponentów
 - Ambari
 - Pakiety deb rpm
 - Gotowe dystrybucje (Hortonworks, Cloudera, MapR, etc)
 - Puppet i Chef
12. Konfiguracja klastra obliczeniowego pod kątem wydajności systemu
13. System wysokiej niezawodności (SWN) / High availability (HA) w ekosystemie Hadoop
14. Rozwiązywanie typowych problemów
15. Dodawanie kolejnych węzłów obliczeniowych
16. Bezpieczeństwo danych i uprawnienia
17. Monitorowanie i zarządzanie
 - Ambari
 - Ganglia
 - Nagios

