

Karta usługi została zablokowana. Podmiot nie dopełnił wymagań związanych z publikacją karty usługi w Bazie Usług Rozwojowych. Zapis na usługę nie jest możliwy.

# Developer Training For Spark and Hadoop

## Usługa archiwalna

Usługa została zablokowana przez Administratora Bazy.

## Informacje o usłudze

<b>Czy usługa może być dofinansowana?</b>	Tak
<b>Sposób dofinansowania</b>	<ul style="list-style-type: none"><li>wsparcie dla osób indywidualnych</li><li>wsparcie dla przedsiębiorców i ich pracowników</li></ul>
<b>Rodzaj usługi</b>	Usługa szkoleniowa
<b>Podrodzaj usługi</b>	Usługa szkoleniowa
<b>Dostępność usługi</b>	Otwarta

Numer usługi	<b>2019/04/30/10100/388588</b>		
Cena netto	<b>9 374,00 zł</b>	Cena brutto	<b>11 530,02 zł</b>
Cena netto za godzinę	<b>334,79 zł</b>	Cena brutto za godzinę	<b>411,79</b>
Usługa z możliwością dofinansowania	<b>Tak</b>		
Liczba godzin usługi	<b>28</b>		
Termin rozpoczęcia usługi	<b>2019-12-02</b>	Termin zakończenia usługi	<b>2019-12-05</b>

Termin rozpoczęcia rekrutacji	<b>2019-04-30</b>	Termin zakończenia rekrutacji	<b>2019-12-02</b>
Maksymalna liczba uczestników	12		
Kategoria główna KU	<b>Informatyka i telekomunikacja</b>		
Kategorie dodatkowe KU	<b>Informatyka i telekomunikacja</b>		
Podstawa uzyskania wpisu w zakresie świadczenia usług współfinansowanych	<b>Certyfikaty:</b> Znak Jakości Małopolskich Standardów Usług Edukacyjno-Szkoleniowych (MSUES)		
Czy usługa pozwala na uzyskanie kwalifikacji lub części kwalifikacji zarejestrowanych w ZRK?	<b>Nie</b>		
Czy usługa pozwala na uzyskanie kwalifikacji innych niż kwalifikacje zarejestrowane w ZRK?	<b>Nie</b>		
Czy usługa prowadzi do nabycia kompetencji?	<b>Tak</b>		

## Informacje o podmiocie świadczącym usługę

Nazwa podmiotu	<b>Compendium - Centrum Edukacyjne Spółka z o.o.</b>		
Osoba do kontaktu	<b>Michał Dobrzański</b>	Telefon	<b>12 29 84 777</b>
E-mail	<b>michal.dobrzanski@compendium.pl</b>		

## Cel usługi

### Cel edukacyjny

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 procesowane na klastrze Hadoop? Jak używać Sqoop i Flume do przyjmowania danych na klaster? Jak przetwarzać dystrybuowane dane przy pomocy Apache Spark? 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?

# Szczegółowe informacje o usłudze

---

## Ramowy program usługi

- 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
  - 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
    - 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
- 

Efekty usługi (produkty), efekty uczenia się/kształcenia

Nabywanie umiejętności IT

---

Grupa docelowa

Usługa również adresowana dla uczestników projektu **Kierunek Kariera**

**Harmonogram**

---

LP	Przedmiot / Temat zajęć	Data realizacji zajęć	Godzina rozpoczęcia	Godzina zakończenia	Liczba godzin
Brak danych					

## Osoby prowadzące usługę

---

## Lokalizacja usługi

---

Adres: <b>Bielska 17</b> <b>02-394 Warszawa, woj. mazowieckie</b> Szczegóły miejsca realizacji usługi:	Warunki logistyczne: <b>Klimatyzacja, Laboratorium komputerowe, Wi-fi</b>
---	--