



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Technologie Big Data [S1DSwB1>TBD]

Przedmiot

Kierunek studiów

Data Science w biznesie

Rok/Semestr

3/6

Studia w zakresie (specjalność)

–

Profil studiów

ogólnoakademicki

Poziom studiów

pierwszego stopnia

Język oferowanego przedmiotu

polski

Forma studiów

stacjonarne

Wymagalność

obligatoryjny

Liczba godzin

Wykład

30

Laboratorium

0

Inne

0

Ćwiczenia

30

Projekty/seminaria

0

Liczba punktów ECTS

5,00

Koordynatorzy

dr hab. Grzegorz Pawłowski

grzegorz.pawlowski@put.poznan.pl

Wykładowcy

Wymagania wstępne

Podstawowa wiedza za zakresu programowania oraz obsługi baz danych.

Cel przedmiotu

Przekazanie podstawowej wiedzy i umiejętności w zakresie organizacji, zarządzania i przetwarzania bardzo dużych, zmiennych i różnorodnych zbiorów danych typu Big Data.

Przedmiotowe efekty uczenia się

Wiedza:

1. Charakteryzuje metody analizy i przetwarzania dużych zbiorów danych, w tym techniki eksploracji, uczenia maszynowego i wizualizacji danych w środowisku Big Data [DSB1_W01].
2. Opisuje architekturę systemów rozproszonych oraz baz danych NoSQL wykorzystywanych w technologii Big Data [DSB1_W02].
3. Wyjaśnia koncepcję przetwarzania danych w modelu MapReduce oraz wykorzystanie Apache Spark do analizy Big Data [DSB1_W03].
4. Analizuje kluczowe aspekty bezpieczeństwa, zarządzania prywatnością i regulacji prawnych w zakresie danych wielkoskalowych [DSB1_W05].

Umiejętności:

1. Dobiera odpowiednie metody i narzędzia analizy danych Big Data, uwzględniając ich efektywność i skalowalność [DSB1_U02].
2. Projektuje i przeprowadza eksperymenty analityczne w środowisku Big Data, wykorzystując rozproszone systemy obliczeniowe [DSB1_U03].
3. Tworzy procesy analizy i przetwarzania danych z użyciem Spark SQL i PySpark, optymalizując zarządzanie dużymi zasobami danych [DSB1_U08].
4. Stosuje techniki uczenia maszynowego w analizie dużych zbiorów danych, wdrażając modele predykcyjne i eksploracyjne [DSB1_U09].
5. Efektywnie współpracuje w interdyscyplinarnych zespołach analitycznych, integrując wiedzę z zakresu technologii chmurowych i systemów rozproszonych [DSB1_U14].

Kompetencje społeczne:

1. Podejmuje inicjatywy biznesowe związane z wdrażaniem technologii Big Data, wykorzystując innowacyjne metody analizy danych [DSB1_K04].

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Ocena podsumowująca z wykładu wystawiana na podstawie wyniku procentowego z kolokwium. Pytania i zadania sprawdzające zrozumienie przedmiotowych zagadnień. Próg zaliczeniowy - 50%. Ocena formująca z laboratorium składa się z ocen, które student otrzymuje za wykonanie poszczególnych zadań podczas zajęć. Ocena podsumowująca z laboratorium wystawiana jest jako średnia z tych ocen. Do oceny uwzględnia się poprawność i kompletność osiągniętych rezultatów.

Treści programowe

Wykład: Definicja, znaczenie, zastosowania oraz wyzwania Big Data. Wprowadzenie do rozproszonych systemów plików i baz danych NoSQL. Model programowania MapReduce, architektura i komponenty Sparc, pobieranie i przesyłanie strumieniowe danych. Magazynowanie danych i wykorzystanie eksploracji danych oraz uczenia maszynowego w analizie i wizualizacji danych. Idea przetwarzania w chmurze i charakterystyka największych platform: MS Azure, Amazon Web Service (AWS) i Google Cloud Platform (GCP) oraz ich usług. Zarządzanie danymi, bezpieczeństwo i odpowiedzialne korzystanie z danych.

Laboratorium: Zapoznanie studentów z platformą Apache Hadoop celem operowanie na rozproszonych systemach plików. Wprowadzenie do silników przetwarzania wsadowego, optymalizacji i dekompozycji danych na przykładzie MapReduce. Relacyjne przetwarzanie danych przy użyciu Spark SQL. Wykorzystanie PySparc do zarządzania dużymi zasobami danych.

Tematyka zajęć

1. Podstawy Big Data
2. Charakterystyka zasobów Big Data
3. Metody przetwarzania dużych zbiorów danych
4. Analiza i wizualizacja Big Data
5. Platformy chmur obliczeniowych i ich usługi
6. Wybrane systemy zarządzania Big Data
7. Bezpieczne zarządzanie prywatnymi danymi.

Metody dydaktyczne

Wykłady: wykład informacyjny, prezentacja multimedialna, wykład problemowy.

Laboratoria: metoda laboratoryjna, metoda przypadków (case study), metoda warsztatowa.

Literatura

Podstawowa:

Marz N., Warren J., Big Data. Najlepsze praktyki budowy skalowalnych systemów obsługi danych w czasie rzeczywistym, Helion, 2016.

White T., Hadoop. Kompletny przewodnik. Analiza i przechowywanie danych, Helion, 2015.

Zaharia M., Chambers B., Spark: The Definitive Guide, O'Reilly Media / Helion, 2018.
Chalkiopoulos A., Programming MapReduce with Scalding. A practical guide to designing, testing, and implementing complex MapReduce applications in Scala, Helion, 2014.

Uzupełniająca:

Garcia-Molina H., Ullman J.D., Widom J., Systemy baz danych. Kompletny podręcznik, Helion, 2015.
Ryza S., Lasersson U., Owen S., Wills J., Spark. Zaawansowana analiza danych, Helion, Gliwice, 2015.

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	125	5,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	62	2,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwίων/egzaminu, wykonanie projektu)	63	2,50