

Technologie big data - opis przedmiotu

Informacje ogólne	
Nazwa przedmiotu	Technologie big data
Kod przedmiotu	11.3-WE-BEP-TBD
Wydział	Wydział Informatyki, Elektrotechniki i Automatyki
Kierunek	Biznes elektroniczny
Profil	praktyczny
Rodzaj studiów	pierwszego stopnia z tyt. inżyniera
Semestr rozpoczęcia	semestr zimowy 2017/2018

Informacje o przedmiocie	
Semestr	3
Liczba punktów ECTS do zdobycia	6
Typ przedmiotu	obowiązkowy
Język nauczania	polski
Sylabus opracował	prof. dr hab. inż. Dariusz Uciński

Formy zajęć					
Forma zajęć	Liczba godzin w semestrze (stacjonarne)	Liczba godzin w tygodniu (stacjonarne)	Liczba godzin w semestrze (niestacjonarne)	Liczba godzin w tygodniu (niestacjonarne)	Forma zaliczenia
Wykład	30	2	18	1,2	Zaliczenie na ocenę
Projekt	30	2	18	1,2	Zaliczenie na ocenę

Cel przedmiotu

Głównym celem przedmiotu jest zaznajomienie studentów z istniejącymi rozwiązaniami do składowania oraz przetwarzaniu dużych zbiorów danych (big data) oraz wykształceniu w nich umiejętności racjonalnego doboru istniejących technologii pod względem dopasowania do rozpatrywanego problemu. Studenci zapoznają się z następującymi technologiami: SQL, Apache Hadoop, Apache Spark, Elasticsearch oraz SAS. Rozpatrywane przykłady użycia technologii będą dotyczyły takich problemów jak: transfer, składowanie oraz przetwarzanie big data, koszt wdrożenia oraz utrzymania infrastruktury, skalowalność oraz niezawodność rozwiązań.

Wymagania wstępne

Projektowanie i programowanie obiektowe.

Zakres tematyczny

Geneza oraz charakterystyka big data. Zastosowanie big data w inteligencji biznesowej oraz analityce. Techniki składowania dużych wolumenów danych. Systemy relacyjne typu SQL (MySQL). Dedykowane technologie oraz platformy do przetwarzania oraz analizy big data (Apache Hadoop, Apache Spark, Elasticsearch oraz SAS). Przetwarzanie równoległe z wykorzystaniem paradygmatu MapReduce. Rozproszone bazy kolumnowe typu NoSQL (Hbase).

Przykłady wykorzystania technologii big data. Analiza strumieni danych, składowanie oraz przetwarzanie dużych wolumenów danych w trybie wsadowym, analityka w czasie rzeczywistym (w pamięci).

Analiza funkcjonalności platform big data. Dzielenie, repliki, balansowanie obciążeniem, odporność na uszkodzenia podczas składowania/udostępniania danych. Techniki dostępu do danych. Wizualizacja oraz eksploracja danych w czasie rzeczywistym z wykorzystaniem platformy Elasticsearch, Apache Spark oraz SAS Visual Analytics.

Omówienie problemów związanych z kosztem wdrożenia oraz utrzymania infrastruktury, skalowalności oraz niezawodności omawianych rozwiązań.

Zrozumienie zasad pracy rozproszonych baz danych typu NoSQL (Hadoop oraz Elasticsearch). Stworzenie aplikacji (w Javie) strumieniującej dane do powyższych systemów. Stworzenie prostego algorytmu MapReduce do równoległego przetwarzania danych.

Metody kształcenia

Wykład - wykład konwencjonalny z wykorzystaniem wideoprojektora.

Laboratorium - zajęcia praktyczne w laboratorium komputerowym.

Efekty uczenia się i metody weryfikacji osiągnięcia efektów uczenia się

Opis efektu	Symbole efektów	Metody weryfikacji	Forma zajęć
Rozumie w jaki sposób funkcjonują rozproszone systemy składowania danych typu NoSQL. Potrafi scharakteryzować podstawowe funkcje oferowane przez te systemy.	- K_W04 - K_W15	przygotowanie projektu	Projekt
Potrafi dokonać analizy porównawczej systemów składowania typu SQL oraz NoSQL.	K_W04	sprawdzian	Wykład
Zna kluczowe pojęcia z zakresu big data	- K_W06 - K_W19	sprawdzian	Wykład

Opis efektu	Symbole efektów	Metody weryfikacji	Forma zajęć
Zna metodologie związane z przetwarzaniem big data oraz potrafi je odpowiednio dobrać do rozpatrywanego problemu.	• K_W04 • K_W06	• sprawdzian	• Wykład
Potrafi napisać program klienta korzystającego ze strumienia danych do komunikacji z omawianymi systemami oraz wykorzystać technikę MapReduce do przetwarzania rozproszonego przesyłanych danych	• K_U02 • K_U14	• przygotowanie projektu	• Projekt

Warunki zaliczenia

Wykład - sprawdzian w formie pisemnej i/lub ustnej, realizowany na koniec semestru.

Laboratorium – ocena końcowa stanowi sumę ważoną ocen uzyskanych za realizację poszczególnych ćwiczeń laboratoryjnych oraz sprawdzianów kontrolnych weryfikujących przygotowanie merytoryczne do ćwiczeń.

Ocena końcowa = 50 % oceny zaliczenia z formy zajęć wykład + 50 % oceny zaliczenia z formy zajęć laboratorium.

Literatura podstawowa

1. White, T., Hadoop: The Definitive Guide, 3rd Edition, O'Reilly Media/Yahoo Press, 2012.
2. George, L., HBase: The Definitive Guide, O'Reilly Media, 2011.

Literatura uzupełniająca

1. Stanton, J.M.: Introduction to Data Science, e-book, 2013.

Uwagi

Zmodyfikowane przez dr hab. inż. Marcin Mrugalski, prof. UZ (ostatnia modyfikacja: 07-04-2017 14:01)

Wygenerowano automatycznie z systemu SylabUZ