**Nazwa przedmiotu:**

Przetwarzanie danych w platformach Big Data

**Koordynator przedmiotu:**

Dr hab. inż. Maciej Grzenda, Dr inż Jarosław Legierski

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Informatyka i Systemy Informacyjne

**Grupa przedmiotów:**

Wspólne

**Kod przedmiotu:**

1120-INPAD-MSP-XXXX

**Semestr nominalny:**

2 / rok ak. 2018/2019

**Liczba punktów ECTS:**

4

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

.

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

.

**Język prowadzenia zajęć:**

polski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

.

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład: | 15h |
| Ćwiczenia: | 0h |
| Laboratorium: | 30h |
| Projekt: | 0h |
| Lekcje komputerowe: | 0h |

**Wymagania wstępne:**

Bazy danych

**Limit liczby studentów:**

.

**Cel przedmiotu:**

Celem przedmiotu jest przekazanie wiedzy na temat architektury i przykładowych środowisk Big Data umożliwiających pozyskiwanie, składowanie i analizę danych w trybie wsadowym i strumieniowym.

**Treści kształcenia:**

Wykład:
Przegląd zagadnień oraz oprogramowania nurtu Big Data, w tym projektów takich jak Apache Hadoop, Apache Spark, czy też Apache Flume.
Architektura systemów Big Data: pozyskiwanie danych, składowanie danych, przekazywanie danych w rozproszonej architekturze systemu, analiza danych, w tym analiza z wykorzystaniem metod uczenia maszynowego.
Przetwarzanie wsadowe a przetwarzanie strumieni danych.
Programowanie rozwiązań wykorzystujących platformy Big Data z uwzględnieniem m.in. wzorca MapReduce, użycia równoległego i rozproszonego przetwarzania danych oraz środowisk ułatwiających tworzenie kodu przetwarzania danych w trybie wsadowym i strumieniowym.
Wzorce projektowe i architektoniczne np. architektura Lambda i Kappa.
Laboratorium:
Konfiguracja pozyskiwania danych z różnorodnych źródeł. Programowanie zadań MapReduce. Programowanie przetwarzania wsadowego. Programowanie przetwarzania strumieniowego. Projektowanie architektury złożonego systemu.

**Metody oceny:**

Zaliczenie przedmiotu oparte jest o wyniki realizacji zadań punktowanych w trakcie laboratorium (60%) oraz kolokwium końcowe (40%). Maksymalna liczba dostępnych punktów wynosi 100. Wyniki zadań punktowanych są ogłaszane na stronie internetowej prowadzącego zajęcia w danej grupie laboratoryjnej lub rozsyłane do uczestników drogą mailową.
W końcowej części semestru student może skorzystać z terminu poprawkowego, w trakcie którego może poprawić jedno z zadań punktowanych lub kolokwium końcowe:
- termin poprawkowy jest ogłaszany studentom z wyprzedzeniem co najmniej jednego tygodnia,
- każdy ze studentów ma zagwarantowaną możliwość udziału w jednym terminie poprawkowym, przy założeniu, iż był obecny w oryginalnym terminie zadania/kolokwium lub przedstawił w terminie dokument uzasadniający nieobecność w tym terminie,
- najpóźniej tydzień przed terminem poprawkowym, studenci zainteresowani poprawianiem zadania/kolokwium, zobligowani są zadeklarować, które z zadań chcieliby poprawiać w terminie poprawkowym,
- ocena zadania uzyskana w terminie poprawkowym zastępuje oryginalną ocenę z tego zadania lub kolokwium.
Ocena końcowa zależy od łącznej liczby punktów uzyskanych z zadań punk-towanych oraz kolokwium i jest wyznaczana zgodnie z poniższymi regułami: 0-50 pkt. - 2.0, 51-60 pkt. - 3.0, 61-70 pkt. - 3.5, 71-80 pkt. - 4.0, 81-90 pkt. - 4.5, 91-100 pkt. - 5.0.

**Egzamin:**

nie

**Literatura:**

1. B. Ellis, Real-Time Analytics. Techniques to Analyse and Visualise Streaming Data, Wiley, 2014
2. A. Holmes, Hadoop in practice, Manning Publications, 2013
3. N. Marz, J. Warren, Big Data. Principles and best practices of scalable realtime data systems, Manning, 2015
4. F. Provost, T. Facett, Data Science for Business. What you need to know about data mining and data-analytic thinking, O’Reilly, 20

**Witryna www przedmiotu:**

e.mini.pw.edu.pl

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W2\_01:**

Zna sposób programowania przetwarzania danych Big Data z wykorzystaniem uznanych środowisk klastrowych zapewniających równoległe i rozproszone wykonywanie kodu

Weryfikacja:

ocena kolokwium

**Powiązane charakterystyki kierunkowe:** PD\_W15, PD\_W04, PD\_W05, PD\_W11

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W2\_02:**

Zna wzorce architektoniczne i zasady użycia metod uczenia maszynowego w ramach środowisk Big Data

Weryfikacja:

ocena kolokwium

**Powiązane charakterystyki kierunkowe:** PD\_W04, PD\_W10

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U2\_01:**

Umie korzystać z dokumentacji projektów Big Data dostępnej w języku angielskim

Weryfikacja:

zadania punktowane

**Powiązane charakterystyki kierunkowe:** PD\_U03

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U2\_02:**

Potrafi stworzyć kod prostej analizy danych z wykorzystaniem środowisk Big Data i metod statystycznych oraz uczenia maszynowego

Weryfikacja:

zadania punktowane

**Powiązane charakterystyki kierunkowe:** PD\_U06

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U2\_03:**

Potrafi zaprojektować architekturę systemu Big Data dla zadanego zagadnienia z uwzględnieniem składowania i analizy danych oraz sposobów zapewnienia koniecznej wydajności i eliminacji opóźnień w przetwarzaniu danych

Weryfikacja:

zadania punktowane

**Powiązane charakterystyki kierunkowe:** PD\_U07, PD\_U09, PD\_U12, PD\_U17

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U2\_04:**

Potrafi zaimplementować kod równoległej agregacji, filtrowania i analizy danych z wykorzystaniem środowisk Big Data

Weryfikacja:

zadania punktowane

**Powiązane charakterystyki kierunkowe:** PD\_U11

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka K2\_01:**

Potrafi posługiwać się językiem angielskim w stopniu umożliwiającym wykorzystanie dokumentacji systemów Big Data i rozumie rolę komunikacji w zespole projektowym w projektach Big Data

Weryfikacja:

zadania punktowane

**Powiązane charakterystyki kierunkowe:** PD\_K08

**Powiązane charakterystyki obszarowe:**