KARTA MODUŁU / KARTA PRZEDMIOTU

Kod modułu	**Z-ID-608b**
Nazwa modułu	**Bazy danych typu Big Data**
Nazwa modułu w języku angielskim	**Big Data Databases**
Obowiązuje od roku akademickiego	**2015/2016**

A. USYTUOWANIE MODUŁU W SYSTEMIE STUDIÓW

Kierunek studiów	**Inżynieria danych**
Poziom kształcenia	**I stopień**
Profil studiów	**Praktyczny**
Forma i tryb prowadzenia studiów	**Studia stacjonarne**
Specjalność	**Inżynieria zasobów danych**
Jednostka prowadząca moduł	**Katedra Informatyki i Matematyki Stosowanej**
Koordynator modułu	**Dr hab. Marzena Nowakowska**
Zatwierdził	

B. OGÓLNA CHARAKTERYSTYKA PRZEDMIOTU

Przynależność do grupy/bloku przedmiotów	**Specjalnościowy**
Status modułu	**Obowiązkowy**
Język prowadzenia zajęć	**Polski**
Usytuowanie modułu w planie studiów - semestr	**Semestr VI**
Usytuowanie realizacji przedmiotu w roku akademickim	**Semestr letni**
Wymagania wstępne	**Bazy danych, Programowanie baz danych w środowisku RAD, Zarządzanie bazami danych - SQL**
Egzamin (TAK/NIE)	**NIE**
Liczba punktów ECTS	**4**

Forma prowadzenia zajęć	**wykład w**	**ćwiczenia ć**	**laboratorium l**	**projekt p**	**inne i**
Liczba godzin w semestrze	**10**		**30**		

C. EFEKTY KSZTAŁCENIA I METODY SPRAWDZANIA EFEKTÓW KSZTAŁCENIA

Cel modułu	Zapoznanie z kluczowymi technologiami informatycznymi stosowanymi do składowania, przetwarzania i analizowania dużych zbiorów danych typu Big data. Nabycie praktycznych umiejętności w zakresie budowania prostych aplikacji analitycznych na platformach typu Big Data oraz stosowania podstawowych narzędzi do wizualizacji dużych zbiorów danych.

Symbol efektu	Efekty kształcenia	Forma prowadzenia zajęć	Odniesienie do efektów kierunkowych	Odniesienie do efektów obszarowych
W_01	Student ma podstawową wiedzę na temat środowisk programistycznych używanych do budowy rozproszonych baz danych i przetwarzania rozproszonego danych.	w	K_W12 K_W13	T1P_W04 T1P_W06 T1P_W07 X1P_W04 inzP_W01 inzP_W02 inzP_W03 inzP_W04
W_02	Student ma wiedzę na temat narzędzi programowych do wizualizacji danych.	l	K_W13	T1P_W04 T1P_W06 X1P_W04 inzP_W01 inzP_W02 inzP_W03
U_01	Student potrafi pozyskiwać informacje z artykułów naukowych, podręczników i opracowań dotyczących technologii typu Big Data oraz integrować uzyskane informacje w celu dokonania ich interpretacji i oceny, a także wyciągać wnioski oraz formułować i uzasadniać opinie.	l	K_U01	T1P_U01 T1P_U03 T1P_U06 X1P_U07 inzP_U03
U_02	Student potrafi posługiwać się narzędziami programowymi do wizualizacji dużych zbiorów danych.	l	K_U06	T1P_U05 T1P_U08 T1P_U09 T1P_U13 T1P_U16 inzP_U02 inzP_U03 inzP_U07
U_03	Student potrafi posługiwać się platformami programistycznymi do budowy rozproszonych baz danych i przetwarzania rozproszonego danych, potrafi dobrać odpowiednie narzędzia programistyczne do zadanego problemu.	l	K_U12 K_U09	T1P_U14 T1P_U15 T1P_U16 T1P_U17 inzP_U03 inzP_U06 inzP_U07 inzP_U08
K_01	Ma świadomość odpowiedzialności za pracę własną oraz gotowość podporządkowania się zasadom pracy w zespole i ponoszenia odpowiedzialności za wspólnie realizowane zadania	l	K_K04	T1P_K03 T1P_K04 X1P_K02 InzP_K02
K_02	Rozumie potrzebę ciągłego doskonalenia i podnoszenia kompetencji zawodowych.	w, l	K_K01	T1P_K01 X1P_K01 X1P_K05 InzP_K01 InzP_K02

Treści kształcenia

1. Treści kształcenia w zakresie wykładu

Nr wykładu	Treści kształcenia	Odniesienie do efektów kształcenia dla modułu
1	**Wprowadzenie do Big Data** Źródła danych masowych. Nowe technologie i trendy w przetwarzaniu danych masowych. Perspektywy analiz biznesowych z wykorzystaniem danych masowych.	W_01, K_02
2	**Wstępne przetwarzanie danych** Znaczenie jakości danych. Oczyszczanie i standaryzacja danych. Podstawowe techniki i narzędzia. Wizualizacja danych.	W_01, W_02, K_02
3-4	**Wprowadzenie do przetwarzania rozproszonego** Obliczenia rozproszone jako szczególny przypadek obliczeń równoległych. Model programowania *map-reduce*. Projekt Apache Hadoop jako wolna (open source) implementacja paradygmatu *map-reduce*. Stos programowy Hadoop'a. Hadoop MapReduce: cechy, interfejsy, przykłady zadań typu *map* i *reduce*.	W_01, W_02
5	**Zarządzanie przetwarzaniem danych** Pliki wsadowe, potoki danych, zarządzanie cyklami zadań przetwarzania wsadowego na platformie Hadoop. Dane ustrukturyzowane a dane semistrukturalne. Ograniczenia relacyjnych baz danych w przetwarzaniu danych typu Big Data. Hbase - implementacja rozproszonej bazy danych typu NoSQL na platformie Hadoop. Architektura i właściwości.	W_01, W_02

2. Treści kształcenia w zakresie ćwiczeń.

3. Treści kształcenia w zakresie zadań laboratoryjnych.

Nr lab.	Treści kształcenia	Odniesienie do efektów kształcenia dla modułu
1	Wstępna analiza danych. Czyszczenie danych. Ćwiczenia praktyczne.	U_01, U_02, U_03, K_01, K_02
2	Standaryzacja danych. Wizualizacja danych dla celów raportowania analitycznego. Ćwiczenia praktyczne.	U_01, U_02, U_03, K_01, K_02
3	Rozproszony systemu plików na platformie Hadoop (Hadoop Distributed File System - HDFS). Cechy, dostęp, architektura. Ćwiczenia praktyczne.	U_01, U_03, K_01, K_02
4	Hadoop MapReduce: interfejsy użytkownika i konfiguracja. Ćwiczenia praktyczne.	U_01, U_03, K_01, K_02
5-6	Hadoop MapReduce: programowanie. Ćwiczenia praktyczne.	U_01, U_03, K_01, K_02
7	Zarządzanie klastrem Apache Hadoop. Typy instalacji. Instalacja wybranej dystrybucji. Konfiguracja. Formatowanie HDFS. Uruchomienie.	U_01, U_03, K_01, K_02
8-9	Podłączanie do HBase za pomoca Java API. Ćwiczenia praktyczne w pisaniu prostych aplikacji dla HBase.	U_01, U_03, K_01, K_02
9-10	Programowanie zadań przetwarzania potokowego z użyciem języka Pig Latin. Ćwiczenia praktyczne w programowaniu operacji typu agregowanie, filtrowanie, sortowanie, łączenie danych.	U_01, U_03, K_01, K_02
11-12	System magazynu danych Hive. Architektura i właściwości. Ćwiczenia praktyczne w tworzeniu zapytań i programowaniu zadań MapReduce w języku HivQL.	U_01, U_03, K_01, K_02
13-14	Integracja systemu SAS z platformą Hadoop. Ładowanie danych z HDFS. Wizualizacja dużych zbiorów danych.	U_01, U_03, K_01, K_02
15	Kolokwium zaliczeniowe.	

4. Charakterystyka zadań projektowych.

5. Charakterystyka zadań w ramach innych typów zajęć dydaktycznych.

Metody sprawdzania efektów kształcenia

Symbol efektu	**Metody sprawdzania efektów kształcenia** *(sposób sprawdzenia, w tym dla umiejętności – odwołanie do konkretnych zadań projektowych, laboratoryjnych, itp.)*
W_01	Kolokwium zaliczeniowe w formie testu. Zadanie do samodzielnego wykonania w grupie, dyskusja na ćwiczeniach laboratoryjnych.
W_02	Kolokwium zaliczeniowe w formie testu. Zadanie do samodzielnego wykonania w grupie, dyskusja na ćwiczeniach laboratoryjnych.
U_01	Zadanie do samodzielnego wykonania w grupie, dyskusja na ćwiczeniach laboratoryjnych.
U_02	Zadanie do samodzielnego wykonania w grupie, dyskusja na ćwiczeniach laboratoryjnych.
U_03	Zadanie do samodzielnego wykonania w grupie, dyskusja na ćwiczeniach laboratoryjnych.
K_01	Zadanie do samodzielnego wykonania w grupie, dyskusja na ćwiczeniach laboratoryjnych.

D. NAKŁAD PRACY STUDENTA

Bilans punktów ECTS			
Lp.	**Rodzaj aktywności**	**Obciążenie studenta**	**Jednostka**
1.	Udział w wykładach	**10**	h
2.	Udział w ćwiczeniach		
3.	Udział w laboratoriach	**30**	h
4.	Udział w zajęciach projektowych		
5.	Udział w konsultacjach (2-3 razy w semestrze)	**5**	h
6.	Konsultacje projektowe		
7.	Udział w egzaminie		
8.			
9.	**Liczba godzin realizowanych przy bezpośrednim udziale nauczyciela akademickiego**	**45**	h
10.	**Liczba punktów ECTS, którą student uzyskuje na zajęciach wymagających bezpośredniego udziału nauczyciela akademickiego** *(1 punkt ECTS=27 godzin obciążenia studenta)*	**1,7**	ECTS
11.	Samodzielne studiowanie tematyki wykładów	**6**	h
12.	Samodzielne przygotowanie się do ćwiczeń		
13.	Samodzielne przygotowanie się do kolokwiów	**10**	h
14.	Samodzielne przygotowanie się do laboratoriów	**16**	h
15.	Wykonanie sprawozdań	**18**	h
16.	Przygotowanie do kolokwium końcowego z laboratorium	**12**	h
17.	Wykonanie projektu lub dokumentacji		
18.	Przygotowanie do egzaminu		
19.			
20.	**Liczba godzin samodzielnej pracy studenta**	**62**	h
21.	**Liczba punktów ECTS, którą student uzyskuje w ramach samodzielnej pracy** *(1 punkt ECTS=27 godzin obciążenia studenta)*	**2,3**	ECTS
22.	**Sumaryczne obciążenie pracą studenta**	**107**	h
23.	**Punkty ECTS za moduł** *1 punkt ECTS=27 godzin obciążenia studenta*	**4**	ECTS
24.	**Nakład pracy związany z zajęciami o charakterze praktycznym** *Suma godzin związanych z zajęciami praktycznymi*	**91**	h
25.	**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym** *1 punkt ECTS=27 godzin obciążenia studenta*	**3,4**	ECTS

E. LITERATURA

Wykaz literatury	1. Capriolo E., *Programming Hive*, O'Reilly 2012, ISBN 978-1449319335. 2. George L., *HBase: The Definitive Guide*, O'Reilly 2011, ISBN 978-1449396107 3. Russell J., *Zwinna analiza danych. Apache Hadoop dla każdego*, Helion, Katowice 2015, ISBN 9788324699445. 4. Summer E., *Hadoop operations*, O'Reilly 2012, ISBN 978-1449327057. 5. *The Apache Software Foundation* [online]. URL: apache.org.
Witryna WWW modułu/przedmiotu	