



## KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Zaawansowane technologie baz danych

### Przedmiot

Kierunek studiów

Informatyka

Studia w zakresie (specjalność)

Internet Przedmiotów

Poziom studiów

drugiego stopnia

Forma studiów

stacjonarne

Rok/semestr

1/1

Profil studiów

ogólnoakademicki

Język oferowanego przedmiotu

polski

Wymagalność

obligatoryjny

### Liczba godzin

Wykład

15

Laboratoria

30

Inne (np. online)

Ćwiczenia

Projekty/seminaria

### Liczba punktów ECTS

3

### Wykładowcy

Odpowiedzialny za przedmiot/wykładowca:

prof. dr hab. inż. Zbyszko Królikowski

email: Zbyszko.Krolikowski@cs.put.poznan.pl

tel: 61 665 2907

wydział: Informatyki i Telekomunikacji

adres: ul. Piotrowo 2, 60-965 Poznań

Odpowiedzialny za przedmiot/wykładowca:

dr inż. Paweł Boiński

email: pawel.boinski@cs.put.poznan.pl

tel: 61 665 2965

wydział: Informatyki i Telekomunikacji

adres: ul. Piotrowo 2, 60-965 Poznań

### Wymagania wstępne

Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K\_W1-2, K\_W4, K\_W6-15, weryfikowane w procesie rekrutacji na studia 2 stopnia – efekty te prezentowane są w serwisie internetowym wydziału.

Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K\_U1-2, K\_U4, K\_U7-8, K\_U14-20, K\_U22-23, K\_U26, weryfikowane w procesie rekrutacji na studia 2 stopnia – efekty te prezentowane są w serwisie internetowym wydziału.

Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K\_K1-9, weryfikowane w procesie rekrutacji na studia 2 stopnia – efekty te prezentowane są w serwisie internetowym wydziału. Ponadto w zakresie kompetencji społecznych student musi prezentować takie



postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.

### Cel przedmiotu

Przekazanie studentom wiedzy dotyczącej projektowania i implementacji hurtowni danych i aplikacji klasy BI, w zakresie: architektury, modelowania danych, projektowania warstwy integrującej i zasilającej - ETL, rozszerzeń SQL dla aplikacji klasy BI, struktur fizycznych i indeksowych hurtowni danych, trendów rozwojowych hurtowni danych i systemów BI. Przekazanie studentom podstawowej wiedzy dotyczącej baz NoSQL.

Rozwijanie u studentów umiejętności rozwiązywania problemów, w zakresie: projektowania i implementowania systemu hurtowni danych, oceny przydatności technologii hurtowni danych i BI oraz eksploracji danych do konkretnego zastosowania.

Kształtowanie u studentów umiejętności pracy zespołowej w ramach budowy systemów hurtowni danych. Kształtowanie u studentów umiejętności realizowania projektów hurtowni danych i BI oraz eksploracji danych. Kształtowanie u studentów umiejętności korzystania ze źródeł wiedzy (np. publikacje anglojęzyczne) i samokształcenia się oraz zarządzania czasem.

### Przedmiotowe efekty uczenia się

#### Wiedza

ma zaawansowaną i pogłębioną wiedzę z zakresu hurtowni danych oraz baz NoSQL, podstaw teoretycznych ich budowania oraz metod, narzędzi i środowisk programistycznych wykorzystywanych do ich implementacji

ma zaawansowaną wiedzę szczegółową dotyczącą hurtowni danych i baz NoSQL

ma wiedzę o trendach rozwojowych baz NoSQL

ma zaawansowaną i szczegółową wiedzę o procesach zachodzących w cyklu życia hurtowni danych zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich i prowadzeniu prac badawczych w zakresie hurtowni danych i analizy danych

#### Umiejętności

potrafi pozyskiwać informacje na temat zaawansowanych systemów baz danych z literatury oraz innych źródeł (w języku ojczystym i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i problemów badawczych metody eksperymentalne

potrafi - przy formułowaniu i rozwiązywaniu zadań inżynierskich z zakresu zaawansowanych systemów baz danych - integrować wiedzę z różnych obszarów informatyki (np. inżynieria oprogramowania, administrowanie systemami informatycznymi, bazy danych) oraz zastosować podejście systemowe potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych (w szczególności systemów zarządzania bazami danych, środowisk programistycznych ETL)

potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia (m.in. wykorzystując dostępne przewodniki po narzędziach ETL)



### Kompetencje społeczne

rozumie, że w informatyce wiedza i umiejętności z zakresu baz danych bardzo szybko stają się przestarzałe

rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu baz danych w rozwiązywaniu problemów badawczych i praktycznych (m.in. dobór odpowiednich narzędzi i metod dla zasilania i odświeżania hurtowni danych)

### Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

W zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez ocenę wiedzy i umiejętności wykazanych na kolokwium zaliczeniowym z pytaniami (5-7) problemowymi i otwartymi. Kolokwium uznaje się za zaliczony po uzyskaniu ponad 50% możliwych do zdobycia punktów.

Przyjmuje się następującą skalę ocen i punktów:

<0;50%>: ndst., (50%;60%>: dst, (60%;70%>: dst+, (70%;80%>: db, (80%;90%>: db+, (90%;100%>: bdb

W zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocenę realizacji zadań zleczanych na każdych zajęciach,
- ocenę wiedzy i umiejętności związanych z realizacją zadań laboratoryjnych poprzez rozwiązanie sprawdzianu (w formie testu) na koniec semestru.
- uzyskiwanie punktów dodatkowych za aktywność podczas zajęć, a szczególnie za:
- omówienia dodatkowych aspektów zagadnienia,
- uwagi związane z udoskonaleniem materiałów dydaktycznych.

W zakresie laboratorium przyjmuje się następującą skalę ocen w zależności od liczby uzyskanych punktów: <0;50%>: ndst., (50%;60%>: dst, (60%;70%>: dst+, (70%;80%>: db, (80%;90%>: db+, (90%;100%>: bdb.

### Treści programowe

Program wykładów obejmuje następujące zagadnienia:

- problematyka przetwarzania BigData, w tym: BI;
- architektury systemu hurtowni danych (podstawowa ETL, z warstwą ODS, z warstwą data mart, architektura ELT, BigData),
- modelowanie danych (konceptualny model wielowymiarowy, implementacja relacyjna - schematy gwiazdy, płątka śniegu, konstelacji faktów wraz z ich oceną, implementacja wielowymiarowa),
- struktury fizyczne plików hurtowni danych oraz struktury indeksowe,
- bazy NoSQL do zastosowań web-owych i chmurowych (Dlaczego NoSQL? Bazy klucz – wartość, bazy dokumentów, bazy grafowe, ...)

Program zajęć laboratoryjnych podzielono na następujące części:

1. Wprowadzenie do środowiska ćwiczeniowego
  - studium przypadku,
  - źródła danych, schemat hurtowni danych,
  - podstawy metodyki Agile BI.
2. Wprowadzenie do obsługi narzędzia Pentaho Data Integration



- podstawowe pojęcia,
  - repozytorium,
  - transformacja oparta na jednym źródle danych,
  - transformacja podrzędna.
3. Obsługa wielu źródeł danych
- rozbudowa istniejących transformacji i transformacji podrzędnych o dodatkowe źródło danych,
  - sterowanie ścieżką przepływu danych,
  - metody łączenia danych.
4. Dodatkowe transformacje
- metody eliminowania duplikatów,
  - automatyczne generowanie danych dla wymiarów,
  - zasilanie tabeli faktów.
  - podstawy metodyki Agile BI.
5. Zaawansowane transformacje
- źródła danych oparte na plikach CSV, wykrywanie zmian w źródłach danych,
  - operacyjna składnica danych, odświeżanie hurtowni danych.
6. Nowoczesne źródła danych
- dokumenty XML, usługi sieciowe.
7. Profilowanie i czyszczenia danych, dane historyczne
- wykrywanie błędów w danych (dane referencyjne, wzorce danych),
  - automatyczne poprawianie błędów, naprawianie błędów w źródłach danych,
  - modyfikacja transformacji w celu przechowywania danych historycznych dla zmieniających się wymiarów.
8. Poprawa wydajności procesu ETL, tematyczne hurtownie danych
- masowe ładowanie danych (Oracle, PostgreSQL, MySQL)
  - wyliczanie agregatów z danych, przykład tematycznej hurtowni danych.
9. Przetwarzanie danych w hurtowniach danych za pomocą języka SQL i jego rozszerzeń.

Zajęcia są prowadzone w formie zajęć ćwiczeniowych przy komputerach, przy czym każdy student pracuje samodzielnie. Każde zadanie jest poprzedzone krótką prezentacją a następnie omówione zagadnienia są ćwiczone w praktyce.

### Metody dydaktyczne

1. wykład: prezentacja multimedialna, prezentacja jest uzupełniana krótkimi przykładami prezentowanymi w sposób tradycyjny z wykorzystaniem tablicy.
2. laboratoria: prezentacja multimedialna, prezentacja jest uzupełniana krótkimi przykładami prezentowanymi w sposób tradycyjny z wykorzystaniem tablicy, wykonywanie ćwiczeń w hurtowni danych, omawianie trudniejszych ćwiczeń przy tablicy, odpowiedzi na pytania na bieżąco, rozwiązywanie problemów na bieżąco

### Literatura



Podstawowa

1. Z.Królikowski, Hurtownie danych – Struktury logiczne i fizyczne, Wydawnictwo Naukowe PP, 2008
2. A.Chodkowska-Gyurics, Hurtownie danych, Teoria i praktyka, PWN 2014
3. P.J.Sadalage, M.Flower, „NoSQL - Kompendium wiedzy”, Helion, 2015

Uzupełniająca

1. Jiang B.: Constructing Data Warehouses with Metadata-driven Generic Operators, and more: Architecture, Methodology, and Paradigm; Concepts, Algorithms, and Operators; Principles, Recommendations, and Exercises. DBJ Publishing, 2011, ISBN-13: 978-3033029200
2. Dokumentacja Pentaho Data Integration <https://wiki.pentaho.com/>
3. Matt Casters, Roland Bouman, Jos Van Dongen: Pentaho Kettle Solutions, John Wiley & Sons 2010
4. Jiang B.: Constructing Data Warehouses with Metadata-driven Generic Operators, and more: Architecture, Methodology, and Paradigm; Concepts, Algorithms, and Operators; Principles, Recommendations, and Exercises. DBJ Publishing, 2011, ISBN-13: 978-3033029200
5. Dokumentacja Pentaho Data Integration <https://wiki.pentaho.com/>
6. Matt Casters, Roland Bouman, Jos Van Dongen: Pentaho Kettle Solutions, John Wiley & Sons 2010
7. A.Pelikant, Hurtownie danych – od przetwarzania analitycznego do raportowania, Wyd. Helion, 2011

**Bilans nakładu pracy przeciętnego studenta**

	Godzin	ECTS
Łączny nakład pracy	75	3,0
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	45	2,0
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych, przygotowanie do sprawdzianów, wykonanie zadań <sup>1</sup> )	30	1,0

<sup>1</sup> niepotrzebne skreślić lub dopisać inne czynności