

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Hurtownie danych i przetwarzanie analityczne		Kod 1010512321010513977
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 2
Ścieżka obieralności/specjalność Technologie przetwarzania danych	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) niestacjonarna	
Godziny Wykłady: 16 Ćwiczenia: - Laboratoria: 8 Projekty/seminaria: 16	Liczba punktów 5	
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (ogólnouczelniany, z innego kierunku) kierunkowy z danego kierunku		
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne nauki techniczne		Podział ECTS (liczba i %) 5 100% 5 100%
Odpowiedzialny za przedmiot / wykładowca: Odpowiedzialny za przedmiot / wykładowca:		
dr hab. inż. Robert Wrembel email: robert.wrembel@cs.put.poznan.pl tel. 6652991 Informatyki ul. Piotrowo 2		Robert Wrembel email: robert.wrembel@cs.put.poznan.pl tel. 6652991 Informatyki ul. Piotrowo 2
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z systemów baz danych (relacyjny model danych, SQL, indeks drzewiasty, schematy koncepcyjne i logiczne, zarządzanie transakcjami i współbieżnym dostępem do danych), języków programowania, systemów operacyjnych.
2	Umiejętności:	Student powinien umieć: dokonać analizy wymagań dla systemu bazy danych, zainstalować system zarządzania bazą danych i skonfigurować bazę danych, utworzyć jej użytkownika, utworzyć podstawowe obiekty, wykonać podstawowe polecenia SQL. Ponadto, powinien umieć korzystać z różnych źródeł literaturowych w celu przeanalizowania istniejących rozwiązań i narzędzi dla zadanego problemu. Powinien rozumieć konieczność poszerzania wiedzy, umiejętności i kompetencji ponieważ technologie hurtowni danych rozwijają się bardzo szybko. Powinien być gotowy do podjęcia współpracy w ramach zespołu i do prowadzenia rozmów z klientem (zaj. projektowe).
3	Kompetencje społeczne	Student musi prezentować takie postawy jak: uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi, chęć niesienia pomocy zawodowej, zaangażowanie w realizowane przedsięwzięcie.
Cel przedmiotu:		
1. Wskazanie praktycznych problemów projektowania, implementowania, wdrażania i utrzymania systemów hurtowni danych (HD), klasycznych i big data.		
2. Przekazanie wiedzy dotyczącej projektowania systemów HD, w zakresie: architektur technicznych, modelowania danych, projektowania warstwy integrującej i zasilającej - ETL, struktur fizycznych, optymalizacji zapytań analitycznych, technologii przetwarzania danych masywnych.		
3. Przedstawienie problematyki implementowania HD i aplikacji do analizy danych, w zakresie: rozszerzeń SQL do analizy danych, wykorzystania struktur fizycznych (indeksy, partycje, perspektywy zmaterializowane) w procesie optymalizacji zapytań analitycznych.		
4. Rozwijanie umiejętności rozwiązywania problemów, w zakresie: projektowania i implementowania systemów HD, oceny przydatności technologii HD i analizy danych do konkretnego zastosowania, testowania zaproponowanego rozwiązania pod kątem jego efektywności i funkcjonalności.		
5. Kształtowanie umiejętności pracy zespołowej w ramach projektów. Kształtowanie umiejętności realizowania projektów praktycznych z zakresu HD i analizy danych.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		

<p>1. Zaawansowana wiedza z zakresu: (1) architektur systemów hurtowni danych (HD) (klasycznych i big data), (2) teorii modelowania danych dla zastosowań analitycznych, (3) struktur danych dla HD, (4) technik optymalizacji zapytań analitycznych, (5) narzędzi i środowisk programistycznych wykorzystywanych do budowy HD - [K2st_W1]</p> <p>2. Wiedza szczegółowa z zakresu systemów HD (architektury, techniki i narzędzia integracji danych, modele logiczne i implementacyjne, struktury fizyczne, optymalizacja zapytań gwiazdzystych, strojenie wydajności, super-serwery) - [K2st_W3]</p> <p>3. Wiedza o trendach rozwojowych architektur i technologii HD. Wiedza o istniejących problemach dotyczących projektowania i budowania systemów HD - [K2st_W4]</p> <p>4. Zaawansowana wiedza o w cyklu projektowania i życia systemów HD - [K2st_W5]</p> <p>5. Zaawansowana wiedza dotycząca architektur, metod, technik, struktur fizycznych w rozwiązywaniu zadań projektowania systemów HD i rozwiązywaniu nietypowych zadań analizy danych (por. zajęcia projektowe) - [K2st_W6]</p>
<p>Umiejętności:</p> <p>1. Pozyskuje informacje z różnych źródeł wiedzy technicznej i naukowej (w j. pol. i ang.) na temat zagadnień objętych programem przedmiotu. Potrafi integrować i konfrontować te informacje, dokonywać ich interpretacji i krytycznej oceny. Potrafi uzasadnić wybór rozwiązania dla zadanego problemu (por. zaj. projektowe) - [K2st_U1]</p> <p>2. Posługuje się technikami informacyjno-komunikacyjnymi do realizowania projektów - [K2st_U2]</p> <p>3. Potrafi: (1) projektować i przeprowadzać eksperymenty, (2) interpretować uzyskane wyniki i wyciągać wnioski, (3) formułować i weryfikować hipotezy, w zakresie zadań deweloperskich systemów HD. Potrafi zrealizować projekty techniczne w oparciu o technologie HD (por. zaj. projektowe) - [K2st_U3]</p> <p>4. Potrafi wykorzystać metody analityczne, symulacyjne i eksperymentalne do formułowania i rozwiązywania zadań technicznych w zakresie systemów HD - [K2st_U4]</p> <p>5. Przy rozwiązywaniu zadań technicznych i badawczych integruje wiedzę z różnych obszarów informatyki (bazy danych, hurtownie danych, systemy operacyjne, systemy rozproszone, języki programowania, złożoność obliczeniowa, technologie internetowe) - [K2st_U5]</p> <p>6. Ocenia przydatność i możliwość wykorzystania nowych koncepcji, technologii, i oprogramowania dla systemów hurtowni danych (np. systemy klasy NoSQL, technologie strumieniowe, Hadoop, Spark) - [K2st_U6]</p> <p>7. Potrafi ocenić przydatność i możliwość wykorzystania metod i narzędzi służących do rozwiązania zadania technicznego/inżynierskiego, polegającego na: zaprojektowaniu, zaimplementowaniu, lub ocenie wybranych komponentów systemu HD - [K2st_U9]</p> <p>8. Stosując nowe techniki, technologie, oprogramowanie, rozwiązuje złożone zadania projektowania, implementowania, wdrożenia wybranych komponentów systemu HD, często o charakterze badawczym (por. zaj. proj.) - [K2st_U10]</p> <p>9. Na podstawie otrzymanej lub opracowanej samodzielnie analizy wymagań, potrafi zaprojektować system HD lub jego fragment. W tym celu wykorzystuje/przystosowuje właściwe dla problemu metody, techniki, oprogramowanie lub opracowuje własne rozwiązanie (por. zaj. proj.) - [K2st_U11]</p> <p>10. Potrafi pracować w zespole, przyjmując w nim różne role (por. zaj. proj.) - [K2st_U15]</p>
<p>Kompetencje społeczne:</p> <p>1. Rozumie, że obszary technologii hurtowni danych (podobnie jak w innych obszarach informatyki) wiedza i umiejętności bardzo szybko stają się nieaktualne i wymagają ciągłego uczenia się - [K2st_K1]</p> <p>2. Rozumie konieczność korzystania z najnowszej wiedzy i rozwiązań z zakresu technologii hurtowni danych (np. systemy klasy NoSQL, architektury przetwarzania równoległego Spark, Hadoop, architektury przetwarzania danych strumieniowych), w rozwiązywaniu problemów technicznych - [K2st_K2]</p>

Sposoby sprawdzenia efektów kształcenia

Przedstawione wyżej efekty kształcenia są weryfikowane w następujący sposób:

a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocenę wiedzy i umiejętności wykazanych na egzaminie pisemnym o charakterze problemowym i testowym (student może korzystać z dowolnych materiałów dydaktycznych);

Egzamin składa się z 5-6 zadań problemowych i 6-8 pytań testowych jedno lub wielokrotnego wyboru. Maksymalnie można uzyskać 40 punktów, z czego 6-8 za pytania testowe. Nie przyznaje się punktów ułamkowych. Egzamin uznaje się jako zaliczony od 21 punktów. Przyjmuje się następującą skalę ocen i punktów:

0-20: ndst., 21-24: dst, 25-28: dst+, 29-32: db, 33-36: db+, 37-40: bdb

- omówienie wyników egzaminu;

b) w zakresie projektów weryfikowanie założonych efektów kształcenia jest realizowane przez:

- ocenę wiedzy i umiejętności związanych z realizacją zadań projektowych poprzez spotkania ze studentami w terminach zjazdów;

- okresowe prezentacje studentów z postępów w realizacji projektów;

- okresowe weryfikowanie jakości i zawartości dokumentacji technicznej;

- obronę projektu przez studentów;

- ocenę wyniku projektu;

- ocenę dokumentacji projektowej.

Maksymalnie można uzyskać 100 punktów, z czego 50 za wynik projektu, 40 - za dokumentację techniczną, 10 - za prezentację końcową. Nie przyznaje się punktów ułamkowych. Projekt uznaje się za zaliczony od 51 punktów. Przyjmuje się następującą skalę ocen i punktów: 0-50: ndst., 51-60: dst, 61-70: dst+, 71-80: db, 81-90: db+, 91-100: bdb

c) w zakresie zajęć laboratoryjnych weryfikowanie efektów kształcenia jest realizowane w postaci zaliczenia pisemnego, które składa się z 4 zadań polegających na wykazaniu się umiejętnością zastosowania zdobytej wiedzy do rozwiązania zadanych problemów. Za część pisemną maksymalnie można uzyskać 20 punktów. Nie przyznaje się punktów ułamkowych.

Laboratorium uznaje się za zaliczone od 11 punktów. Przyjmuje się następującą skalę ocen i punktów: 0-10: ndst., 11-12: dst, 13-14: dst+, 15-16: db, 17-18: db+, 19-20: bdb

Treści programowe

- Problematyka integracji danych
- Architektury integracji danych
- Architektury systemów hurtowni danych dla zastosowań klasycznych
- Zasilanie hurtowni danych - ETL/ELT
- Modelowanie hurtowni danych
- Wymiary ewoluujące
- Struktury fizyczne dla hurtowni danych
- Optymalizacja zapytań gwiazdzystych
- Systemy klasy main-memory
- Problematyka przetwarzania gigadanych
- Architektury przetwarzania gigadanych

W ramach zajęć projektowych studenci rozwiązują praktyczne zadania projektowe. Zakres tematów jest każdego roku inny. Tematyka projektów obejmuje m.in. projektowanie warstwy ETL, ocenę efektywności struktur danych w różnych systemach zarządzania hurtowniami danych (m.in. Oracle, DB2, SQL Server), zaawansowaną analizę danych, projektowanie i implementowanie aplikacji analitycznych, składowanie i analizę danych NoSQL. Zadania projektowe są realizowane w grupach 2-4 osobowych, w zależności od złożoności zadania. Zaliczenie części projektowej polega na: (1) przedstawieniu zaimplementowanego rozwiązania lub eksperymentu, (2) omówieniu problemu i jego rozwiązania w formie prezentacji dla grupy i (3) przygotowania dokumentacji technicznej.

Literatura podstawowa:

1. Vaisman A., Zimanyi E.: Data Warehouse Systems - Design and Implementation. Springer Verlag, 2014
2. Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P.: Fundamentals of Data Warehouses. Springer, 2010, ISBN-13: 978-3642075643
3. Golfarelli M., Rizzi S.: Data Warehouse Design: Modern Principles and Methodologies. McGraw-Hill Osborne, 2009, ISBN-13: 978-0071610391

Literatura 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

Bilans nakładu pracy przeciętnego studenta

Czynność		Czas (godz.)
1. udział w zajęciach projektowych		16
2. realizacja projektów (poza czasem laboratoryjnym)		24
3. udział w konsultacjach związanych z realizacją procesu kształcenia, w szczególności projektu		2
4. udział w wykładach		16
5. udział w zajęciach laboratoryjnych		8
6. przygotowanie do laboratorium		16
7. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi w czasie realizacji projektu (10 stron tekstu naukowego = 1 godz.), 250 stron		25
8. przygotowanie do egzaminu		16
Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	123	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	42	2
Zajęcia o charakterze praktycznym	48	2