

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: 20 Ćwiczenia: - Laboratoria: 16 Projekty/seminaria: -	Liczba punktów 4	
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		Podział ECTS (liczba i %) 4 100%
Odpowiedzialny za przedmiot / wykładowca: dr hab. inż. Robert Wrembel, prof. nadzw. email: Robert.Wrembel@put.poznan.pl tel. 61 6652991 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	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 www.fc.put.poznan.pl Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z systemów baz danych i języków programowania.
2	Umiejętności:	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 www.fc.put.poznan.pl Powinien posiadać umiejętność rozwiązywania podstawowych problemów z projektowania systemów informatycznych, administrowania system. baz danych, formułowania poleceń w języku SQL oraz umiejętność pozyskiwania inform. ze wskazanych źródeł. Powinien również rozumieć konieczność poszerzania swoich kompetencji / mieć gotowość do podjęcia współpracy w ramach zespołu.
3	Kompetencje społeczne	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 www.fc.put.poznan.pl 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: Wskazanie praktycznych problemów jakie rozwiązuje się projektując, implementując i wdrażając systemy hurtowni danych. Przekazanie studentom wiedzy dotyczącej projektowania hurtowni danych, w zakresie: architektury systemów hurtowni danych, modelowania danych, projektowania warstwy integrującej i zasilającej - ETL, struktur fizycznych, optymalizacji zapytań analitycznych, zarządzania metadanymi, trendów rozwojowych hurtowni danych i systemów BI. Przedstawienie problematyki implementowania hurtowni danych i aplikacji klasy BI, w zakresie: rozszerzeń SQL dla aplikacji klasy BI, wykorzystania struktur fizycznych (indeksy, partycje, perspektywy zmaterializowane) w procesie optymalizacji zapytań analitycznych. Rozwijanie u studentów umiejętności rozwiązywania problemów, w zakresie: projektowania i implementowania systemu hurtowni danych, oceny przydatności technologii HD i BI do konkretnego zastosowania, testowania efektywności technologii HD i BI. Kształtowanie u studentów umiejętności pracy zespołowej w ramach budowy systemów hurtowni danych. Kształtowanie u studentów umiejętności realizowania projektów HD i BI.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		

1. ma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną w zakresie architektury systemów komputerowych, baz danych (w szczególności systemu rozproszonych baz danych i architektur heterogenicznych systemów informatycznych) - [K_W4]
2. ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami z zakresu informatyki, takimi jak: architektury i technologie integracji danych, architektury i technologie replikacji danych, zarządzanie transakcjami rozproszonymi, techniki optymalizacji zapytań rozproszonych - [K_W5]
3. ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami z zakresu informatyki, takimi jak: algorytmy alokacji danych w węzłach systemu rozproszonej bazy danych, zarządzanie blokadami w systemie rozproszonej bazy danych - [K_W5]
4. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w informatyce - [K_W6]
5. ma podstawową wiedzę o cyklu życia systemów informatycznych sprzętowych i programowych - [K_W7]
6. zna podstawowe metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich z wybranego obszaru informatyki (projektowania i implementowania architektury integracji danych, projektowania i implementowania systemu rozproszonych baz danych) - [K_W8]

Umiejętności:

1. potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł (w języku ojczystym i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie (także przygotowywać prezentacje i raporty techniczne) - [K_U1]
2. potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia - [K_U5]
3. potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody symulacyjne i eksperymentalne - [K_U9]
4. potrafi ? przy formułowaniu i rozwiązywaniu zadań inżynierskich ? integrować wiedzę z różnych obszarów informatyki (a w razie potrzeby także wiedzę z innych dyscyplin naukowych) oraz zastosować podejście systemowe, uwzględniające także aspekty pozatechniczne - [K_U10]
5. potrafi formułować i testować hipotezy związane z problemami inżynierskimi i prostymi problemami badawczymi - [K_U12]
6. potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych (w szczególności systemów zarządzania bazami danych) - [K_U13]

Kompetencje społeczne:

1. rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe - [K_K1]
2. zna przykłady i rozumie przyczyny wadliwie działających systemów informatycznych - [K_K4]
3. potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania - [K_K6]
4. potrafi współdziałać i pracować w grupie, przyjmując w niej różne role - [K_K5]

Sposoby sprawdzenia efektów kształcenia

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

- a) w zakresie wykładów:
- na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach,
- b) w zakresie laboratoriów:
- na podstawie oceny bieżącego postępu realizacji zadań,

Ocena podsumowująca:

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

- ocenę wiedzy i umiejętności wykazanych na egzaminie pisemnym o charakterze problemowym (student może korzystać z dowolnych materiałów dydaktycznych) i w formie testu jedno- lub wielokrotnego wyboru;

Egzamin składa się z 7 zadań problemowych i 16 pytań testowych jedno lub wielokrotnego wyboru. Maksymalnie można uzyskać 50 punktów, z czego 16 za pytania testowe. Nie przyznaje się punktów ułamkowych. Egzamin uznaje się zaliczony od 26 punktów. Przyjmuje się następującą skalę ocen i punktów:

0-25: nast., 26-30: dst, 31-35: dst+, 36-40: db, 41-45: db+, 46-50: bdb

- omówienie wyników egzaminu;

b) w zakresie laboratoriów / ćwiczeń weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocenę wiedzy i umiejętności związanych z realizacją zadań laboratoryjnych poprzez 1 kolokwium w semestrze;
- ocenę wiedzy i umiejętności związanych z realizacją zadań projektowych poprzez ocenę i obronę przez studenta sprawozdania z realizacji projektu;

Kolokwium zaliczeniowe składa się z 7 zadań problemowych (student może korzystać z dowolnych materiałów dydaktycznych), za które można otrzymać maksymalnie 25 punktów. Rozwiązanie zadania projektowego wraz z prezentacją wyników i sprawozdaniem są oceniane maksymalnie na 25 punktów. Zatem, łącznie można zdobyć 50 punktów. Laboratorium uznaje się zaliczone od 26 punktów, przy czym należy uzyskać przynajmniej 13 punktów z kolokwium i 13 z projektu. Przyjmuje się następującą skalę ocen i punktów:

0-25: ndst

26-30: dst

31-35: dst+

36-40: db

41-45: db+

46-50: bdb

Uzyskiwanie punktów dodatkowych za aktywność podczas zajęć (ujęte w ocenie końcowej), a szczególnie za:

- omówienia dodatkowych aspektów zagadnienia - rozszerzenie zakresu projektu,
- umiejętność współpracy w ramach zespołu praktycznie realizującego zadanie szczegółowe w laboratorium.

Treści programowe

Pięć wykładów w semestrze jest prowadzonych przez ekspertów z przemysłu. Są to wykłady obejmujące:

- Składowanie dużych wolumenów danych (firma BCC),
- System HD i BI - case study (firma BCC),
- Hadoop i technologie składowania dużych wolumenów danych (firma Roche),
- Projektowanie mobilnych aplikacji klasy BI (firma Roche),
- Hurtownie danych w chmurze (firma Microsoft).

Program pozostałych wykładów obejmuje następujące zagadnienia:

- wprowadzenie do problematyki analizy danych, hurtowni danych i business intelligence (uzasadnienie potrzeby wykorzystywania tych technologii),
- 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),
- problematyka modelowania wymiarów,
- problematyka integracji danych dla hurtowni danych,
- projektowanie warstwy ETL,
- struktury fizyczne (indeks B-drzewo, bitmapowy, połączeniowy, bitmapowy połączeniowy, sklastrowany, klaster wielowymiarowy),
- optymalizacja zapytań analitycznych (tzw. zapytań gwiazdzistych),
- optymalizacja operacji grupowania (algorytmy optymalizujące),
- problematyka materializowania wyników zapytań i wyboru zbioru zmaterializowanych wyników,
- kompresja danych i indeksów,
- aplikacje klasy business intelligence,
- aktywne hurtownie danych,
- trendy rozwojowe hurtowni danych.

Zajęcia laboratoryjne prowadzone są w formie piętnastu 2-godzinnych zajęć, odbywających się w laboratorium. Program zajęć podzielono na dwie części. Część pierwsza, obejmująca 5 spotkań, jest prowadzona w formie zajęć ćwiczeniowych przy komputerach, przy czym każdy student pracuje samodzielnie. Każde ćwiczenie jest poprzedzone wprowadzeniem prowadzącego. Następnie, omówione zagadnienia są ćwiczone w praktyce na hurtowni danych. Ta część programu laboratorium obejmuje następujące zagadnienia:

- rozszerzenia analityczne języka SQL (m.in., nowe schematy grupowania, funkcje agregujące w oknach, przewidywanie trendów ilustrowane regresją liniową),
- mechanizmy partycjonowania i ich wykorzystanie w optymalizacji zapytań analitycznych,
- modelowanie wymiarów,
- funkcjonalność perspektyw zmaterializowanych i ich implementacja,
- wykorzystywanie perspektyw zmaterializowanych w procesie optymalizacji zapytań analitycznych metodą przepisywania zapytań.

Część druga laboratorium, obejmująca 10 spotkań, jest prowadzona w formie projektów. Studenci rozwiązują praktyczne zadania projektowe z zakresu m.in. projektowania warstwy ETL, oceny efektywności struktur danych w różnych systemach zarządzania hurtowniami danych (m.in. Oracle, DB2, SQL Server, przetwarzanie danych w chmurze), zaawansowanej analizy danych. Zadania projektowe są realizowane w grupach 2-4 osobowych, w zależności od złożoności zadania.

Część wymienionych wyżej treści programowych realizowana jest w ramach pracy własnej studenta.

Metody dydaktyczne:

1. wykład: prezentacja multimedialna, prezentacja jest uzupełniana krótkimi przykładami prezentowanymi w sposób tradycyjny z wykorzystaniem tablicy, rozwiązywanie problemów i omawianie rozwiązań na 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
3. projekt: realizowany zgodnie z zasadami sztuki; obejmuje: analizę wymagań, projekt rozwiązania, implementację, testowanie, lub wykonanie eksperymentu symulacyjnego (jeśli wymagane w projekcie); projekt jest konsultowany na bieżąco przez prowadzącego, są prowadzone dyskusje nad przyjętym sposobem rozwiązania problemu

Literatura podstawowa:

1. Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P.: Fundamentals of Data Warehouses. Springer, 2010, ISBN-13: 978-3642075643
2. 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 laboratoryjnych / ćwiczeniach	16	
2. realizacja projektów (poza czasem laboratoryjnym)	30	
3. udział w konsultacjach związanych z realizacją procesu kształcenia, w szczególności projektu	4	
4. udział w wykładach	20	
5. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi w czasie realizacji projektu (10 stron tekstu naukowego = 1 godz.), 100 stron	10	20
6. przygotowanie do egzaminu i obecność na egzaminie (2 godz):		
Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	100	4
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	42	2
Zajęcia o charakterze praktycznym	3650	1