

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Hurtownie danych		Kod 1010515321010510515
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 2
Ścieżka obieralności/specjalność Informatyka w procesach biznesowych	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obieralny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) niestacjonarna	
Godziny Wykłady: 16 Ćwiczenia: - Laboratoria: 16 Projekty/seminaria: -		Liczba punktów 3
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) kierunkowy		(ogólnouczelniany, z innego kierunku) z danego kierunku
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne nauki techniczne		Podział ECTS (liczba i %) 100 3% 100 3%
Odpowiedzialny za przedmiot / wykładowca: dr inż. Krzysztof Jankiewicz email: Krzysztof.Jankiewicz@cs.put.poznan.pl tel. 61 6652960 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z systemów baz danych i języków programowania.
2	Umiejętności:	Powinien posiadać umiejętność rozwiązywania podstawowych problemów z projektowania systemów informatycznych, administrowania systemami baz danych, formułowania poleceń w języku SQL oraz umiejętność pozyskiwania informacji ze wskazanych źródeł.
3	Kompetencje społeczne	Powinien również rozumieć konieczność poszerzania swoich kompetencji / mieć gotowość do podjęcia współpracy zespołowej. 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 są rozwiązywane podczas projektowania, implementacji i wdrażania systemów hurtowni danych (HD). Przekazanie studentom wiedzy dotyczącej projektowania systemów HD, w zakresie: architektury, modelowania danych, projektowania warstwy integrującej i zasilającej - ETL, struktur fizycznych, optymalizacji zapytań analitycznych, zarządzania metadanymi, trendów rozwojowych HD i systemów klasy Business Intelligence (BI). Przedstawienie problematyki implementowania hurtowni danych i aplikacji klasy BI, w zakresie: rozszerzeń SQL dla aplikacji klasy BI, wykorzystania struktur fizycznych (indeksy, partycje, perspektywy materializowane) 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 HD. Kształtowanie u studentów umiejętności realizowania projektów HD i BI. Kształtowanie u studentów umiejętności korzystania ze źródeł wiedzy (np. publikacje anglojęzyczne) i samokształcenia się oraz zarządzania czasem.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		

<p>1. ma uporządkowaną i podbudowaną teoretycznie wiedzę ogólną związaną z technologiami hurtowni danych i systemów klasy Business Intelligence - [K2st_W2]</p> <p>2. ma zaawansowaną wiedzę szczegółową dotyczącą wybranych zagadnień związanych z technologiami hurtowni danych i systemów klasy Business Intelligence - [K2st_W3]</p> <p>3. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w zakresie projektowania i implementacji hurtowni danych - [K2st_W4]</p> <p>4. zna zaawansowane metody, techniki i narzędzia stosowane w projektowaniu, implementacji i wdrażaniu hurtowni danych oraz systemów klasy Business Intelligence a także w prowadzeniu prac badawczych w tym obszarze informatyki - [K2st_W6]</p>
<p>Umiejętności:</p> <p>1. potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł (w języku polskim i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie - [K2st_U1]</p> <p>2. potrafi planować i przeprowadzać testy wydajnościowe systemów hurtowni danych, interpretować uzyskane wyniki i wyciągać wnioski, proponować rozwiązania zwiększające wydajność, w tym odpowiednie typy indeksów, perspektyw materializowanych itp. - [K2st_U3]</p> <p>3. potrafi wykorzystywać do projektowania hurtowni danych i systemów klasy BI, różne metody symulacyjne oraz eksperymentalne - [K2st_U4]</p> <p>4. potrafi - przy projektowaniu, implementacji oraz wdrażaniu hurtowni danych - integrować wiedzę z różnych obszarów informatyki (a w razie potrzeby także wiedzę z innych dyscyplin naukowych), uwzględniając także aspekty pozatechniczne - [K2st_U5]</p> <p>5. potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych z zakresu rozwiązań dla hurtowni danych oraz systemów klasy BI - [K2st_U6]</p> <p>6. potrafi rozwiązywać złożone problemy związane z projektowaniem hurtowni danych, w tym przypadki/problemy nietypowe oraz zawierające komponent badawczy - [K2st_U10]</p> <p>7. potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia, w tym innych osób - [K2st_U16]</p>
<p>Kompetencje społeczne:</p> <p>1. rozumie, że w zakresie projektowania i implementacji hurtowni danych wiedza i umiejętności bardzo szybko stają się przestarzałe - [K2st_K1]</p> <p>2. rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu projektowania i implementacji hurtowni danych w rozwiązywaniu problemów badawczych i praktycznych - [K2st_K2]</p>

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 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 w formie testu z pytaniami otwartymi oraz pytaniami jedno- lub wielokrotnego wyboru;

Test składa się z szeregu pytań (15-30) otwartych oraz jedno lub wielokrotnego wyboru. Test 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

- omówienie wyników testu;

- b) w zakresie laboratoriów / ćwiczeń 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 jednego kolokwium (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:

- 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, płatka śniegu, konstelacji faktów wraz z ich oceną, implementacja wielowymiarowa),
- problematyka modelowania wymiarów ze szczególnym uwzględnieniem problematyki wymiarów wolnozmiennych,
- problematyka integracji danych dla hurtowni danych,
- projektowanie warstwy ETL,
- struktury fizyczne (indeks B-drzewo, bitmapowy, połączeniowy, bitmapowy połączeniowy),
- optymalizacja zapytań analitycznych (tzw. zapytań gwiazdzystych),
- optymalizacja operacji grupowania (algorytmy optymalizujące),
- problematyka materializowania wyników zapytań i wyboru zbioru zmaterializowanych wyników,
- kompresja danych i indeksów,
- partycjonowanie danych.

Zajęcia laboratoryjne prowadzone są w formie czterech 4-godzinnych zajęć, odbywających się w laboratorium. Program zajęć podzielono na osiem 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 istniejącej 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.

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, 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		
Literatura podstawowa:		
1. R. Kimball, J. Caserta: The Data Warehouse ETL Toolkit. Wiley, 2004, ISBN 0-7645-6757-8-54500		
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	16	
2. realizacja zadań (poza czasem laboratoryjnym):	12	
3. udział w konsultacjach (mogą być realizowane drogą elektroniczną) związanych z realizacją procesu kształcenia, w szczególności projektu	2	
4. przygotowanie do sprawdzianów / kolokwium	8	
5. udział w wykładach	16	
6. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi w czasie realizacji projektu (10 stron tekstu naukowego = 1 godz.), 100 stron	10	
7. przygotowanie do testu zaliczającego wykład (w tym: udział w teście ? 2 godz.)	2	
8. omówienie wyników testu	2	
Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	76	3
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	34	1
Zajęcia o charakterze praktycznym	28	1