

<b>KARTA OPISU MODUŁU KSZTAŁCENIA</b>		
Nazwa modułu/przedmiotu <b>Hurtownie danych</b>		Kod <b>1010515321010510515</b>
Kierunek studiów <b>Informatyka</b>	Profil kształcenia (ogólnoakademicki, praktyczny) <b>ogólnoakademicki</b>	Rok / Semestr <b>1 / 2</b>
Ścieżka obieralności/specjalność <b>Informatyka w procesach biznesowych</b>	Przedmiot oferowany w języku: <b>polski</b>	Kurs (obligatoryjny/obieralny) <b>obieralny</b>
Stopień studiów: <b>II stopień</b>	Forma studiów (stacjonarna/niestacjonarna) <b>niestacjonarna</b>	
Godziny Wykłady: <b>16</b> Ćwiczenia: - Laboratoria: <b>16</b> Projekty/seminaria: -		Liczba punktów <b>3</b>
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (ogólnouczelniany, z innego kierunku) <b>kierunkowy z danego kierunku</b>		
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki <b>nauki techniczne</b>		Podział ECTS (liczba i %) <b>3 100%</b>
<b>Odpowiedzialny za przedmiot / wykładowca:</b>  dr inż. Krzysztof Jankiewicz email: Krzysztof.Jankiewicz@cs.put.poznan.pl tel. 61 6652960 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		
<b>Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:</b>		
1	<b>Wiedza:</b>	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z systemów baz danych i języków programowania.
2	<b>Umiejętności:</b>	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	<b>Kompetencje społeczne</b>	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.
<b>Cel przedmiotu:</b> Wskazanie praktycznych problemów jakie rozwiązuje się projektując, implementując i wdrażając systemy hurtowni danych. Przekazanie studentom wiedzy dotyczącej projektowania systemów hurtowni danych, w zakresie: architektury, 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. Kształtowanie u studentów umiejętności korzystania ze źródeł wiedzy (np. publikacje anglojęzyczne) i samokształcenia się oraz zarządzania czasem.		
<b>Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia</b>		
<b>Wiedza:</b>		

<p>1. ma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną w zakresie architektury systemów informatycznych, w szczególności hurtowni danych i baz danych - [K_W4]</p> <p>2. ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami z zakresu informatyki, takimi jak: architektury i technologie integracji danych, metodyka projektowania systemu informatycznego (w szczególności hurtowni danych), - [K_W5]</p> <p>3. ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami z zakresu informatyki, takimi jak: technologie analizy danych, techniki optymalizacji zapytań analitycznych, struktury fizyczne dla przetwarzania analitycznego, modele składowania danych - [K_W5]</p> <p>4. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w dziedzinie przetwarzania danych i hurtowni danych - [K_W6]</p> <p>5. ma podstawową wiedzę o cyklu życia systemów informatycznych, w szczególności systemów hurtowni danych - [K_W7]</p> <p>6. zna podstawowe metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich z wybranego obszaru informatyki (projektowania i implementowania systemu hurtowni danych, projektowania i implementowania schematów hurtowni danych i struktur fizycznych, projektowania zapytań analitycznych) - [K_W8]</p>
<b>Umiejętności:</b>
<p>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]</p> <p>2. potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia - [K_U5]</p> <p>3. potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody eksperymentalne - [K_U9]</p> <p>4. potrafi - przy formułowaniu i rozwiązywaniu zadań inżynierskich - integrować wiedzę z różnych obszarów informatyki (np. inżynieria oprogramowania, administrowanie systemami informatycznymi, bazy danych) oraz zastosować podejście systemowe, uwzględniające także aspekty pozatechniczne - [K_U10]</p> <p>5. potrafi formułować i testować hipotezy związane z problemami inżynierskimi i prostymi problemami badawczymi - [K_U12]</p> <p>6. potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych (w szczególności systemów zarządzania bazami danych, środowisk programistycznych ETL) - [K_U13]</p>
<b>Kompetencje społeczne:</b>
<p>1. rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe - [K_K1]</p> <p>2. zna przykłady i rozumie przyczyny wadliwie i właściwie działających systemów informatycznych (w szczególności systemów hurtowni danych) - [K_K4]</p> <p>3. potrafi odpowiednio określić priorytety służące realizacji zadania określonego przez siebie lub innych - [K_K6]</p> <p>4. potrafi współdziałać i pracować w grupie, przyjmując w niej różne role (m.in., kierownika projektu, analityka, programisty) - [K_K5]</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 na koniec semestru.
  - uzyskiwanie punktów dodatkowych za aktywność podczas zajęć, a szczególnie za:
  - omówienia dodatkowych aspektów zagadnienia,
  - uwagi związane z udoskonaleniem materiałów dydaktycznych.

### 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, płątka 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 gwiazdzistych),
- 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 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.

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

<b>Literatura podstawowa:</b>		
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		
<b>Literatura uzupełniająca:</b>		
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		
<b>Bilans nakładu pracy przeciętnego studenta</b>		
<b>Czynność</b>	<b>Czas (godz.)</b>	
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	4	
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		
<b>Obciążenie pracą studenta</b>		
<b>forma aktywności</b>	<b>godzin</b>	<b>ECTS</b>
Łączny nakład pracy	78	3
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	40	2
Zajęcia o charakterze praktycznym	28	1