

## Zaawansowane bazy danych i hurtownie danych

Wydział Informatyki  
Politechnika Białostocka

wiosna 2014

**Prowadzacy:** Agnieszka Onisko-Drużdżel, Marek J. Drużdżel  
**pokój:** 207, Wiejska 45A  
**telefon:** 85-746- 9086  
**E-mail:** [a.onisko@pb.edu.pl](mailto:a.onisko@pb.edu.pl)  
**WWW:** <http://aragorn.pb.bialystok.pl/~aonisko>  
**konsultacje:** środa, godz. 14.15-16.00, czwartek, 16.30-17.30  
lub poprzez indywidualne ustalenie terminu

**Wykład:** czwartek, 10.15-12.00, sala 130  
**Pracownia specjalistyczna:** PS1: czwartek, 12.15-14.00, sala 236  
PS2: czwartek, 14.15-16.00, sala 236  
PS3: czwartek, 17.40-19.15, sala 236

### **Strony WWW przedmiotu:**

<http://aragorn.pb.bialystok.pl/~aonisko/ZBDiHD/ZBDiHD-wyklad.html>  
<http://aragorn.pb.bialystok.pl/~aonisko/ZBDiHD/ZBDiHD-ps.html>

### **Prerekwizyty:**

Znajomość relacyjnych baz danych oraz języka SQL. Umiejętność projektowania relacyjnych baz danych oraz programowania w języku SQL.

### **Opis przedmiotu:**

Przedmiot składa się z kilku bloków tematycznych, które obejmują tematykę hurtowni danych, optymalizację zapytań w praktyce, bazy danych typu NoSQL oraz analizę danych w oparciu o sieci bayesowskie. Większość z tych zagadnień będzie również realizowana w ramach zajęć z pracowni specjalistycznej. Szczegółowy plan tematów zawiera harmonogram zamieszczony na końcu tego dokumentu oraz na stronie WWW przedmiotu.

### **Wymagane oprogramowanie:**

Serwer bazodanowy:

Oracle 11g Express Database System

Klient SQL:

SquirrelSQL 3.4, Oracle SQL Developer

Program **GeNIe** dostępny na stronie <http://genie.sis.pitt.edu/>

## **Literatura:**

- Connolly, T., Begg, C. *Database Systems: A Practical Approach to Design, Implementation, and Management*, 5th edition, Addison-Wesley, 2009.
- Casteel, J., *Oracle 11g: SQL*, Course Technology, 2010.
- Todman, Ch. *Projektowanie hurtowni danych. Wspomaganie zarządzania relacjami z klientami*.  
Seria: Kanon informatyki, Księgarnia Techniczna, 2011.
- Kimball, R., Caserta, J. *The Data Warehouse ETL Toolkit (2nd edition)*. New York: Wiley, 2008.
- Inmon, B., Strauss, D., Neushloss, G. *DW 2.0 – Architecture for the Next Generation of Data Warehousing*. Elsevier Press, 2008.
- Adamson, Ch., Kimball, R. *Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance*. John Wiley & Sons, 2006.
- Walmsley, P. *Xquery*. O'Reilly Media, 2007.
- Silvers, F. *Building and Maintaining a Data Warehouse*. Auerbach Publications, 2008.
- Projekt "Opracowanie programów nauczania na odległość na kierunku studiów wyższych – Informatyka." Dostępny na stronie WWW:  
[http://wazniak.mimuw.edu.pl/index.php?title=Zaawansowane\\_systemy\\_baz\\_danych](http://wazniak.mimuw.edu.pl/index.php?title=Zaawansowane_systemy_baz_danych)

Dokumentacja ORACLE 11g Release 2:

<http://www.oracle.com/pls/db112/homepage>

ORACLE SQL Language Reference:

[http://docs.oracle.com/cd/E11882\\_01/server.112/e26088/toc.htm](http://docs.oracle.com/cd/E11882_01/server.112/e26088/toc.htm)

ORACLE PL/SQL Language Reference:

[http://docs.oracle.com/cd/E11882\\_01/appdev.112/e25519/toc.htm](http://docs.oracle.com/cd/E11882_01/appdev.112/e25519/toc.htm)

## **Ocenianie:**

### ***Wykład:***

Ocenę z wykładu stanowi ocena z egzaminu końcowego przeprowadzonego w sesji na koniec semestru. Egzamin będzie obejmował materiał przedstawiony na wykładach. Będzie to egzamin pisemny bez możliwości korzystania z dodatkowych pomocy w postaci podręczników i innych materiałów. Egzamin powinien być napisany przez każdego studenta samodzielnie bez pomocy innych osób. W przypadku stwierdzenia niesamodzielnej pracy studenta w czasie egzaminu, praca oceniona zostanie na ocenę niedostateczną. Data egzaminu końcowego jest podana w harmonogramie przedmiotu. Wyniki egzaminu będą zamieszczone w systemie USOS najpóźniej dwa dni po przeprowadzonym egzaminie.

***Pracownia specjalistyczna:***

Zaliczenie pracowni specjalistycznej odbędzie się na podstawie zadań, wykonanych w czasie zajęć, kolokwium z SQL'a, oraz zaimplementowanego i zaprezentowanego projektu.

Składowe elementy zaliczenia pracowni specjalistycznej to:

Zadania	120pkt
Wejściówka	10pkt
Kolokwium SQL	90pkt
Projekt	90pkt
<hr/>	
	310pkt

<u>Zakres:</u>	<u>Ocena:</u>
(285, 310>	5.0
(266, 285>	4.5
(245, 265>	4.0
(215, 245>	3.5
(185, 215>	3.0
< 0, 185>	2.0

Zadania:

Uwaga: Warunkiem koniecznym na otrzymanie maksymalnej liczby punktów za poszczególne zadania rozwiązywane w czasie zajęć jest wysłanie sprawozdania na koniec zajęć. Sprawozdanie wysłane w tym samym dniu (ale po zajęciach) wiąże się z utratą co najmniej 8 punktów, każdy kolejny dzień spóźnienia: utratą kolejnych 6 punktów.

Zadania zostaną ocenione w ciągu tygodnia od ich wysłania. Jeśli uważasz, że Twoje rozwiązania zadań zostały źle ocenione, możesz zgłosić się do prowadzącego o ich ponowne sprawdzenie.

Kolokwium (SQL):

Kolokwium z SQL będzie przeprowadzone na ósmym zajęciach z pracowni specjalistycznej. W czasie kolokwium będzie można korzystać z kompilatora bazodanowego. Kolokwium będzie trwało 45 minut. Maksymalna liczba punktów, jaką można zdobyć na kolokwium to 100 punktów. Kolokwium poprawkowe nie jest przewidziane. Jeśli student nie może uczestniczyć w kolokwium, będzie przeprowadzone kolokwium dodatkowe, o ile student udokumentuje swoją niobecność na kolokwium przeprowadzonym w pierwszym terminie. W czasie kolokwium będzie można korzystać z dokumentacji ORACLE zamieszczonej na stronie przedmiotu.

Projekt:

Na szóstym zajęciach z pracowni specjalistycznej zaprezentowane zostaną propozycje projektów, które trzeba będzie wybrać w ciągu tygodnia. Projekty mogą być realizowane w grupach dwu-osobowych. Wyznaczone również zostały etapy w tworzeniu projektu zamieszczone w harmonogramie pracowni specjalistycznej.

Prezentacja projektów odbędzie się na ostatnich zajęciach. Każda prezentacja powinna trwać maksymalnie 10 minut.

**Obecność na zajęciach:**

Obecność na wykładzie nie jest obowiązkowa, aczkolwiek gorąco zachęcam do aktywnego uczestnictwa w wykładach. Z kolei udział w zajęciach z pracowni specjalistycznej jest obowiązkowy. Spóźnianie się na zajęcia rozprasza pozostałych studentów, dlatego proszę o punktualne przychodzenie zarówno na wykłady, jak i na zajęcia z pracowni specjalistycznej.

**Opuszczone zajęcia:**

Studenci nieobecni na wykładzie/pracowni specjalistycznej są odpowiedzialni za zapoznanie się z materiałem przedstawionym na tych zajęciach.

## Wstępny harmonogram wykładu

<u>Data</u>	<u>Temat</u>
06 marca	Wprowadzenie od przedmiotu. Proceduralny język SQL
13 marca	Zaawansowane obiekty baz danych
20 marca	Optymalizacja zapytań w praktyce
27 marca	Modele danych w hurtowniach danych
03 kwietnia	Architektura hurtowni danych
10 kwietnia	Operacje w hurtowniach danych
17 kwietnia	Hurtownie danych: Funkcje analityczne
24 kwietnia	Integracja danych
28 kwietnia	Bazy danych typu NoSQL (BD oparte o wartości kluczowe, dokumentowe BD, kolumnowe BD, grafowe BD)
08 maja	Wprowadzenie do MongoDB
15 maja	Technika Map/Reduce na przykładzie Apache Hadoop (prof. Marek Drużdżel)
22 maja	Analiza danych: Sieci bayesowskie (prof. Marek Drużdżel)
29 maja	Analiza danych: Uczenie sieci bayesowskich, odkrywanie przyczynowości (prof. Marek Drużdżel)
05 czerwca	TBA
12 czerwca	TBA
26 czerwca	<b>Egzamin końcowy</b>

## Wstępny harmonogram pracowni specjalistycznej

<u>Data</u>	<u>Temat</u>
06 marca	Proceduralny język SQL (ORACLE)
13 marca	Proceduralny język SQL. Zaawansowane obiekty baz danych (ORACLE) <b>Wejściówka</b>
20 marca	Optymalizacja zapytań w praktyce (ORACLE) <b>Zadania #1 (do zrealizowania w czasie zajęć)</b>
27 marca	Projektowanie modelu hurtowni danych I
03 kwietnia	Projektowanie modelu hurtowni danych II <b>Zadania #2 (do zrealizowania w czasie zajęć)</b>
10 kwietnia	Hurtownie danych: Podstawowe operacje (SQL, ORACLE), termin wybrania projektu
17 kwietnia	Hurtownie danych: Funkcje analityczne (SQL, ORACLE)
24 kwietnia	<b>Zaawansowany SQL – kolokwium</b>
28 kwietnia	Integracja danych (ORACLE) <b>Zadania #3 (do zrealizowania w czasie zajęć)</b>
08 maja	Bazy danych typu NoSQL, cz. I, projekt – prezentacja postępów
15 maja	Bazy typu NoSQL, cz. II, stworzenie bazy noSQL: np. MongoDB, Neo4j <b>Zadania #4 (do zrealizowania w czasie zajęć)</b>
22 maja	Analiza danych: Sieci bayesowskie (GeNIe) <b>Zadania #5 (do zrealizowania w czasie zajęć)</b>
29 maja	Analiza danych: Odkrywanie przyczynowości (GeNIe) <b>Zadania #6 (do zrealizowania w czasie zajęć)</b>
05 czerwca	Tworzenie dokumentacji w technologii MediaWiki
12 czerwca	<b>Prezentacja projektów</b>