

Hurtownie danych

Wprowadzenie do systemów typu *Business Intelligence*

Krzysztof Goczyła

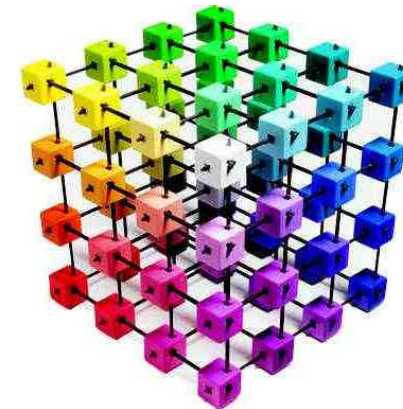
Teresa Zawadzka

Katedra Inżynierii Oprogramowania

Wydział Elektroniki, Telekomunikacji i Informatyki

Politechnika Gdańska

{kris, tegra}@eti.pg.gda.pl



Pobudka



Średnia z kolokwium w roku akademickim
2015/2016 w I podejściu – **11.85/20**

Średnia osób nieuczęszczających na wykłady:
10.74/20

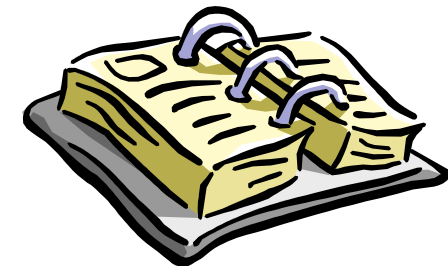
Średnia osób uczęszczających na wykłady:
14.14/20

Zasady zaliczenia

1. Ocena jest wyznaczana na podstawie punktów uzyskanych z wykładu i laboratorium.
2. Połowę oceny z przedmiotu stanowią punkty uzyskane z wykładu, zaś drugą połowę punkty uzyskane z laboratorium.
3. Zaliczenie wykładu (kolokwium) i zaliczenie laboratorium jest obowiązkowe.
4. Ocena z przedmiotu:
 - $\geq 100\%$ - Dokument wyróżniający
 - $\geq 90\%$ - 5.0
 - $\geq 80\%$ - 4.5
 - $\geq 70\%$ - 4.0
 - $\geq 60\%$ - 3.5
 - $\geq 50\%$ - 3.0

Wybrana literatura

- D. Mendrala, M. Szeliga: „SQL 2008. Usługi biznesowe. Analiza i eksploracja danych”. Helion 2009.
- W. H. Inmon: „Building the Data Warehouse”. J. Wiley&Sons, 2002.**
- R. Kimball: „ The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition”. J. Wiley&Sons, 2013.**
- V. Poe, P. Klauer, S. Brebst: „Tworzenie hurtowni danych”, WNT 2000
- P. Ponniah: „Data Warehousing”. J. Wiley&Sons, 2001.



Kierunkowe efekty kształcenia (1)

[K_U08] śledzi i optymalizuje wydajność bazy danych, administruje systemami baz danych i ich odtwarzania po awarii, waliduje wymagania niefunkcjonalne aplikacji bazodanowych

Student potrafi dokonać analizy źródeł danych - ich zawartości i ich formatu. Wie, jak zaprojektować i przeprowadzić proces ETL. Student projektuje hurtownię danych - model logiczny w postaci schematu gwiazdy lub płatka śniegu i model fizyczny (stosując różne modele pamięci). Formułuje zapytania w języku MDX albo w rozszerzeniach wielowymiarowych SQL. Analizuje uzyskane wyniki. Posługuje się prostymi narzędziami typu Business Intelligence.

Kierunkowe efekty kształcenia (2)

[K_U10] posługuje się pakietami oprogramowania do obliczeń symbolicznych i numerycznych, narzędziami business intelligence, notacjami modelowania systemów informatycznych, prowadzi wnioski dedukcyjne i indukcyjne, formułuje zapytania w wybranych językach

Student potrafi korzystać z podstawowych narzędzi typu Business Intelligence, w szczególności wchodzących w skład standardowych komercyjnych pakietów oprogramowania hurtowni danych.

Kierunkowe efekty kształcenia (3)

[K_W04] zna protokoły komunikacji, modele przetwarzania danych, technologie klasycznego i semantycznego Internetu, działanie serwera www, aplikacje internetowe i wzorce programowe oraz budowę aplikacji hostowanych w przeglądarce.

Student zna i potrafi stosować modele danych stosowane w hurtowniach danych oraz budować hurtownie danych zgodne z tymi modelami.

Kierunkowe efekty kształcenia (4)

[K_W01] zna metody przeciwdziałania zagrożeniom bezpieczeństwa sieci komputerowej, protokoły kryptograficzne oraz metody kontroli dostępu do baz danych, systemów i sieci oraz techniki ich zabezpieczeń

Student zna zabezpieczenia baz danych, w tym hurtowni danych, przed nieautoryzowanym dostępem.

Wprowadzenie do systemów typu typu *Business Intelligence*

**Czym jest system typu BI?
Czarodziej to, czy nie czarodziej?**



Doradca, informator, egzekutor?



Kto podejmuje decyzję?



BI

Podstawowa architektura rozwiązań typu BI

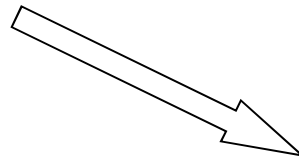


...

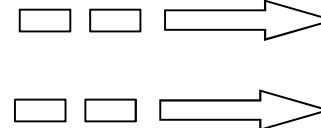


Źródła danych

Proces ETL



Hurtownia danych

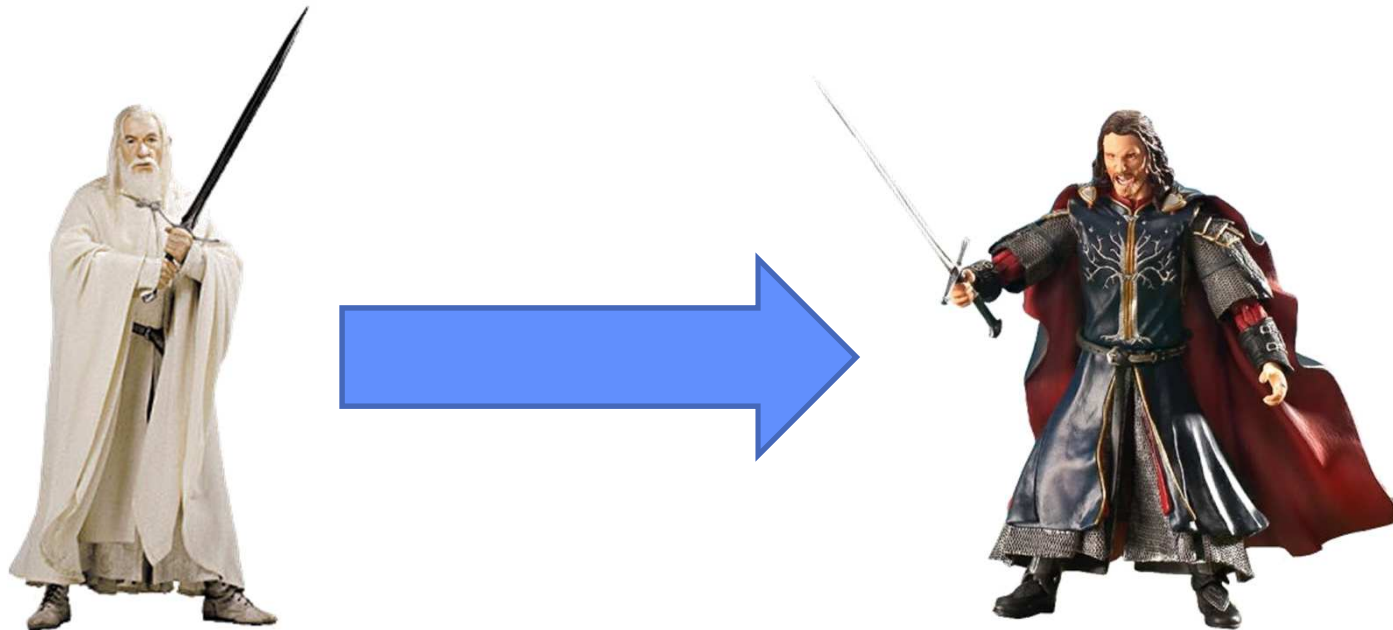


Prezes/Analityk



Kto zna?

- Wydarzenia?
- Powiązania między wydarzeniami?



Hurtownia danych czy zdarzeń?



Dobra analogia?

NIE

**Gandalf potrafi
myśleć!!!**

Systemy typu Business Intelligence

Systemy wspomagania decyzji o charakterze strategicznym dla firmy

(**DSS**, *Decision Support System*; **BI**, *Business Intelligence*):

- łatwy dostęp do informacji
- silne narzędzia analityczne
- wydajne przetwarzanie ogromnych ilości danych
- przyjazne dla użytkownika prezentowanie spójnych wyników
- podatny na zmiany
- ukierunkowany na procesy biznesowe



Narzędzie analityczne



1. **Sprawdzenie wystąpienia danego zdarzenia (czynności)**, np.: „czy dany produkt został sprzedany danemu klientowi”?
2. **Porównanie**, np.: „porównaj wartość zakupów danych dwóch klientów w ramach ostatnich 6 miesięcy” lub „porównaj liczbę sprzedanych egzemplarzy danego produktu względem jego kategorii” .
3. **Analiza trendu**, np.: “czy w ciągu ostatnich 12 miesięcy jest trend wzrostowy sprzedaży danego produktu?”.
4. **Rankingi**, np.: “podaj 10 najlepiej sprzedających się produktów w ciągu ostatniego roku”.
5. **Analiza statystyczna**, np.: “podaj średni zysk z ostatnich 6 miesięcy”.

Online Analytical Processing – OLAP

Przetwarzanie danych, którego celem są analizy trendów, analizy przekrojowe i inne analizy o charakterze strategicznym.

Przykłady:

<http://www2.microstrategy.com/dashboardgallery/Dashboards/InteractiveBubbleChart.asp>

<http://www2.microstrategy.com/dashboardgallery/Dashboards/California.asp>

Proces biznesowy

Seria powiązanych ze sobą działań lub zadań, które rozwiązują określony problem lub prowadzą do osiągnięcia określonego efektu.

1. Procesy biznesowe są zazwyczaj wyrażane za pomocą czasowników ponieważ reprezentują aktywność organizacji.
2. Procesy biznesowe są często wspierane przez systemy, jak np.: system fakturowania.
3. Procesy biznesowe generują metryki liczbowe, które podlegają analizie.

FASMI (*Fast Analysis of Shared Multidimensional Information*) test

1. FAST – system dostarcza użytkownikowi większość odpowiedzi w czasie około 5 sekund.
2. ANALYSIS – system jest dostosowany do logiki biznesowej i statystyki analitycznej.
3. SHARED – system implementuje politykę bezpieczeństwa dostępu do danych.
4. **MULTIDIMENSIONAL** – system musi dostarczać wielowymiarowego spojrzenia na dane.
5. INFORMATION – system musi dostarczać informacji.

Cechy OLAP

	OLTP	OLAP
Wielkość		
Historia		
Liczba użytkowników		
Złożoność zapytań		
Normalizacja		
Dodawanie		
Usuwanie		
Aktualizacja		
Transakcyjność		

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Cechy OLAP

	OLTP	OLAP
Wielkość	Od małej do dużej	Ogromna
Historia		
Liczba użytkowników		
Złożoność zapytań		
Normalizacja		
Dodawanie		
Usuwanie		
Aktualizacja		
Transakcyjność		

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Cechy OLAP

	OLTP	OLAP
Wielkość	Od małej do dużej	Ogromna
Historia	Czasami	Zawsze
Liczba użytkowników		
Złożoność zapytań		
Normalizacja		
Dodawanie		
Usuwanie		
Aktualizacja		
Transakcyjność		

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Cechy OLAP

	OLTP	OLAP
Wielkość	Od małej do dużej	Ogromna
Historia	Czasami	Zawsze
Liczba użytkowników	Duża	Mała
Złożoność zapytań		
Normalizacja		
Dodawanie		
Usuwanie		
Aktualizacja		
Transakcyjność		

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Cechy OLAP

	OLTP	OLAP
Wielkość	Od małej do dużej	Ogromna
Historia	Czasami	Zawsze
Liczba użytkowników	Duża	Mała
Złożoność zapytań	Mała lub średnia	Duża
Normalizacja		
Dodawanie		
Usuwanie		
Aktualizacja		
Transakcyjność		

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Cechy OLAP

	OLTP	OLAP
Wielkość	Od małej do dużej	Ogromna
Historia	Czasami	Zawsze
Liczba użytkowników	Duża	Mała
Złożoność zapytań	Mała lub średnia	Duża
Normalizacja	Tak	Nie
Dodawanie		
Usuwanie		
Aktualizacja		
Transakcyjność		

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Cechy OLAP

	OLTP	OLAP
Wielkość	Od małej do dużej	Ogromna
Historia	Czasami	Zawsze
Liczba użytkowników	Duża	Mała
Złożoność zapytań	Mała lub średnia	Duża
Normalizacja	Tak	Nie
Dodawanie	Tak	Tak (BULK)
Usuwanie		
Aktualizacja		
Transakcyjność		

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Cechy OLAP

	OLTP	OLAP
Wielkość	Od małej do dużej	Ogromna
Historia	Czasami	Zawsze
Liczba użytkowników	Duża	Mała
Złożoność zapytań	Mała lub średnia	Duża
Normalizacja	Tak	Nie
Dodawanie	Tak	Tak (BULK)
Usuwanie	Tak	Nie
Aktualizacja		
Transakcyjność		

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Cechy OLAP

	OLTP	OLAP
Wielkość	Od małej do dużej	Ogromna
Historia	Czasami	Zawsze
Liczba użytkowników	Duża	Mała
Złożoność zapytań	Mała lub średnia	Duża
Normalizacja	Tak	Nie
Dodawanie	Tak	Tak (BULK)
Usuwanie	Tak	Nie
Aktualizacja	Tak	Rzadko
Transakcyjność		

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Cechy OLAP

	OLTP	OLAP
Wielkość	Od małej do dużej	Ogromna
Historia	Czasami	Zawsze
Liczba użytkowników	Duża	Mała
Złożoność zapytań	Mała lub średnia	Duża
Normalizacja	Tak	Nie
Dodawanie	Tak	Tak (BULK)
Usuwanie	Tak	Nie
Aktualizacja	Tak	Rzadko
Transakcyjność	Tak	Nie

**Przetwarzanie transakcyjne (w bazach danych):
OLTP (*On-Line Transactional Processing*)**

Co każdy student wiedzieć powinien...

1. Wytłumaczyć, czym jest system typu BI?
2. Opisać ogólną architekturę systemów typu BI.
3. Wytłumaczyć, czym jest przetwarzanie analityczne.
4. Umieć porównać przetwarzanie analityczne z przetwarzaniem transakcyjnym

