

Analiza danych

Wprowadzenie do systemów baz danych

Andrzej Łuksza

Przetwarzanie analityczne on-line

(On-Line Analytical Processing – OLAP)

- Hurtownie Danych (*Data Warehouses*)
- Magazyny Danych (*Data Marts*)
- Systemy Wspomagania Decyzji (*Decision Support Systems* – DSS)
- Systemy Informowania Kierownictwa (*Executive Information Systems* – EIS)
- Systemy Zarządzania Informacją (*Management Information System* – MIS)
- Zarządzanie kontaktami z klientami (*Customer Relationship Management* – CRM)
- Eksploracja danych (*Data Mining*) – zgłębianie danych
- Big data



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz Społeczny



Pojemność współczesnych baz danych

- Dysk twardy o pojemności 2TB kosztuje około 300 zł
- Przyjmując, że 1 strona maszynopisu zajmuje 2KB uzyskamy, że 2TB = miliard stron maszynopisu
- Przyjmując, że 1 fotografia zajmuje 2MB uzyskamy, że 2TB = milion fotografii
- Przyjmując, że 1 s rozmowy telefonicznej 1kB uzyskamy, że 2TB = 555 tysięcy godzin
- Zakup sprzętu do przechowywania ogromnych ilości danych nie stanowi obecnie większego problemu, znacznie większym problemem jest pozyskanie danych i ich wprowadzenie do systemu

Projekt „SezAM wiedzy, kompetencji i umiejętności” jest współfinansowany przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Wiedza Edukacja Rozwój



Charakterystyka

- OLTP (*Online Transaction Processing*)
 - Dane ciągle aktualizowane
 - Wykorzystywane do aktywnego przetwarzania transakcji
 - Ograniczone do dziesiątek gigabajtów
 - Posiadają małą ilość danych historycznych
 - Zoptymalizowane do efektywnej aktualizacji danych (przetwarzania transakcji)
 - Zorientowane na działanie
- OLAP (*On-Line Analytical Processing*)
 - Dane historyczne, aktualizowane w ustalonym czasie np. co noc
 - Wykorzystywane do analizy w dłuższym przedziale czasowym
 - Zbiory rzędu terabajtów
 - Przechowuje dane przez wiele lat
 - Zoptymalizowane do wyszukiwania danych
 - Zorientowane na cel



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz Społeczny



Operacje analityczne

- **Konsolidacja** – łączenie danych pochodzących z różnych źródeł
 - Np. połączenie danych z różnych uczelni
- **Drażenie** – operacja przeciwna do konsolidacji polegająca na wydobywaniu danych szczegółowych
 - Np. wydobywanie danych studentów urodzonych w konkretnym roku
- **Obracanie** – przedstawianie tych samych danych z różnych punktów widzenia („cięcie i rzutowanie” danych)

Wielowymiarowe struktury danych

- Wielowymiarowa struktura danych jest przedstawiana jako kostka (sześcian) danych
- Każdy bok reprezentuje jeden wymiar
- Przecięcie wymiarów tworzy komórkę, w której jest przechowywana wartość danej, jest to zwykle wartość funkcji agregującej
- Nie ma ograniczeń na liczbę wymiarów struktury danych ani na liczbę poziomów agregacji



Tabela faktów

Ocena
IDStudenta
IDEgzaminatora
IDPrzedmiotu
Ocena
Data



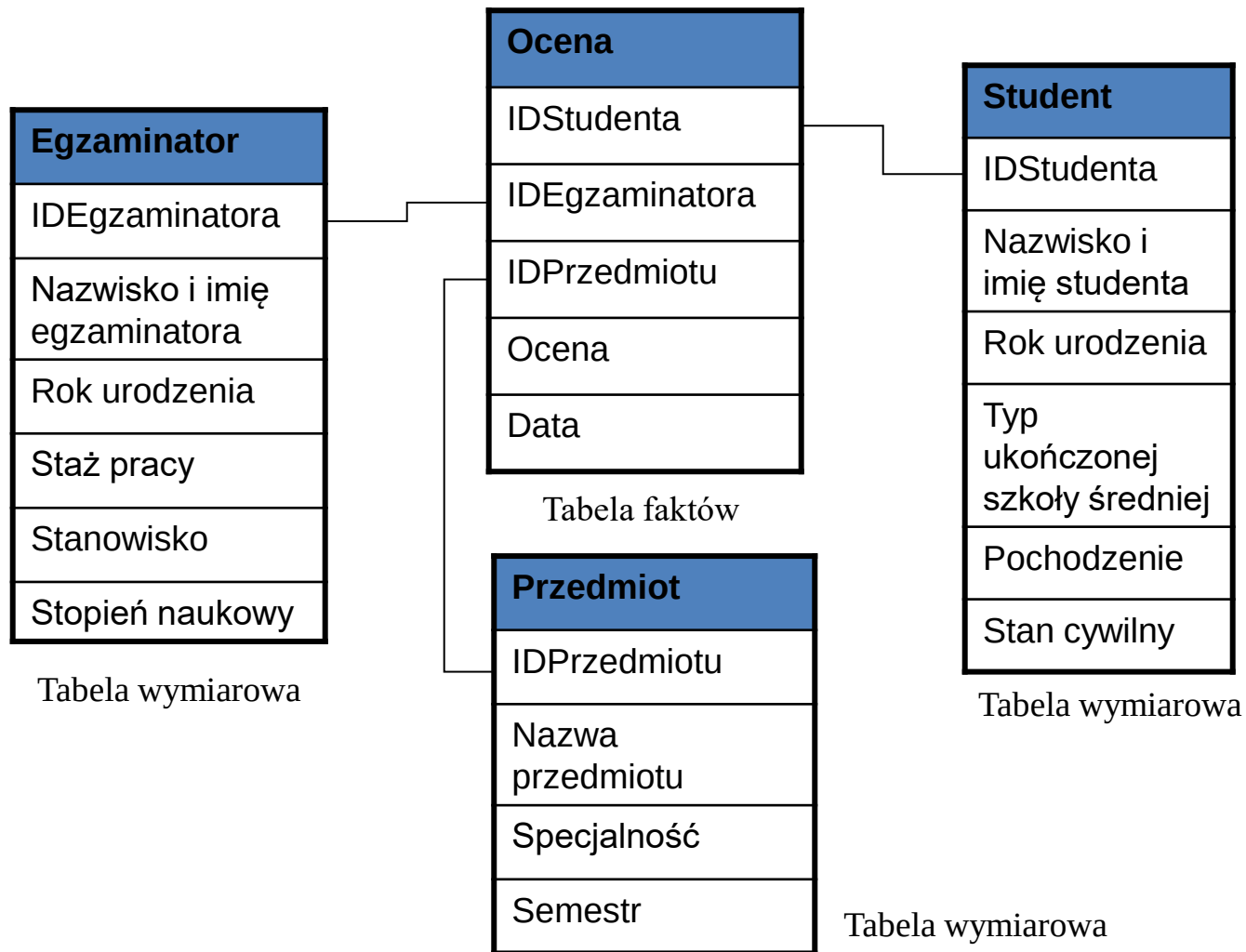
Tabele wymiarowe

Student
IDStudenta
Nazwisko i imię
Rok urodzenia
Typ ukończonej szkoły średniej
Pochodzenie
Stan cywilny

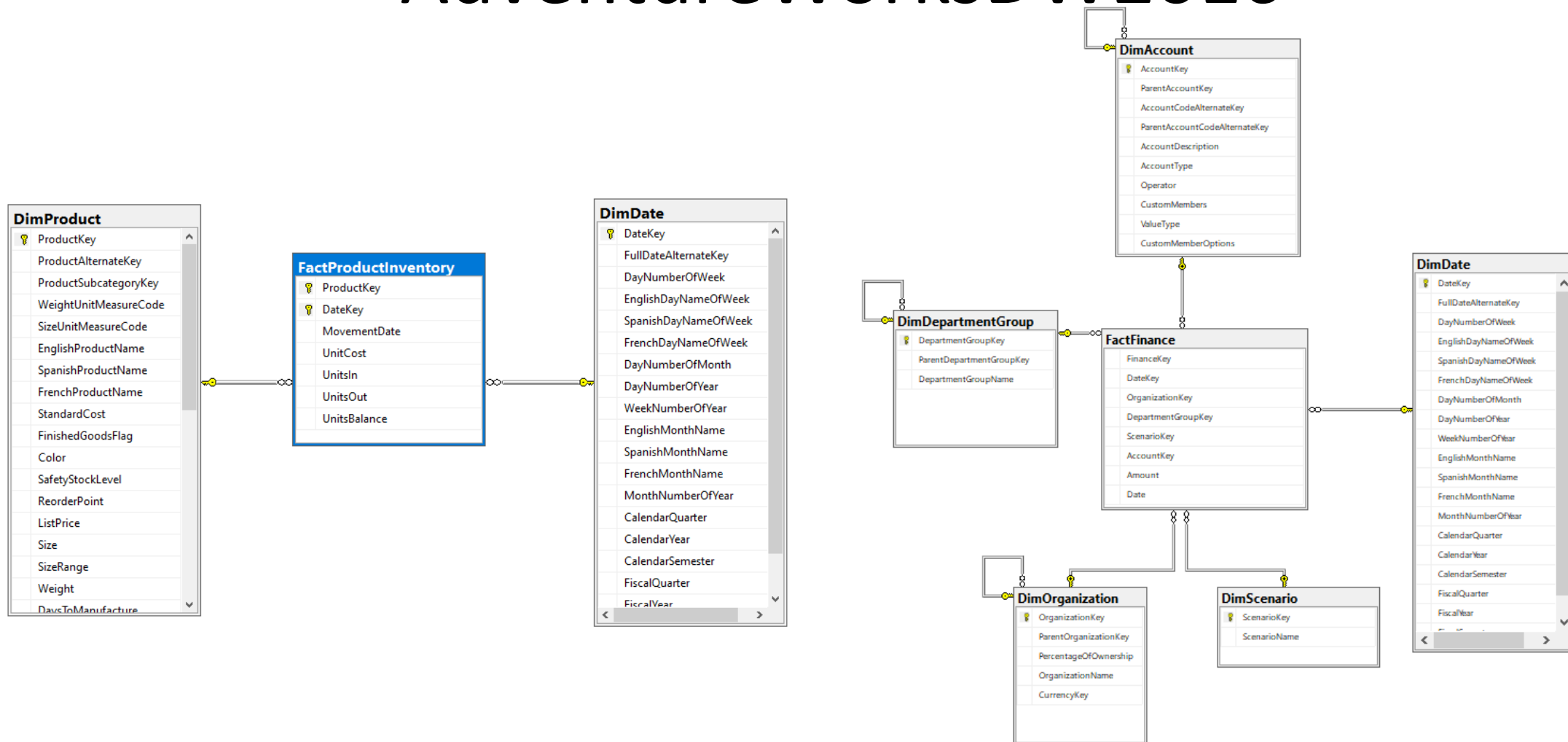
Egzaminator
IDEgzaminatora
Nazwisko i imię egzaminatora
Rok urodzenia
Staż pracy
Stanowisko
Stopień naukowy



Struktura gwiazdy



AdventureWorksDW2016





Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz Społeczny



Przykład wielowymiarowej struktury danych

- Informacja o ocenach
 - W komórkach znajduje się ocena maksymalna lub średnia
- Wymiary
 - Osoba otrzymująca ocenę (nazwisko i imię)
 - Osoba wystawiająca ocenę (nazwisko i imię)
 - Staż pracy wystawiającego ocenę
 - Stanowisko wystawiającego ocenę
 - Nazwa przedmiotu
 - Specjalność
 - Rok akademicki
 - Semestr studiów
 - Rok urodzenia studenta, egzaminatora
 - ...

Projekt „SezAM wiedzy, kompetencji i umiejętności” jest współfinansowany przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Wiedza Edukacja Rozwój

Zobrazowanie wyników

- Człowiek może analizować wielowymiarowe struktury danych i wyciągać, na ich podstawie, właściwe wnioski, jednak analiza taka musi składać się z serii analiz cząstkowych
 - Najłatwiej jest analizować i wizualizować struktury 2 wymiarowe – dodanie 3 wymiaru stwarza już trudności
 - W celu uzyskania 2, 3 wymiarowych struktur danych dokonuje się cięcia (rzutu) struktur wielowymiarowych, co umożliwia oddzielną analizę różnych aspektów złożonego zagadnienia

Przykłady różnych analiz

- Średnia i odchylenie standardowe ocen wystawianych przez pracowników w funkcji stażu pracy
- Średnia i odchylenie standardowe ocen wystawianych przez pracowników w zależności od semestru
- Średnia i odchylenie standardowe ocen wystawianych przez pracowników w kolejnych latach – nie powinno być zbyt dużego rozrzutu
- Średnia ocen studentów w zależności od semestru
- Średnia ocen studentów w kolejnych latach akademickich
- W biznesie analizy dotyczą głównie finansów – maksymalizacji zysków

Kwerenda podsumowująca (grupująca)

- Umożliwia agregację danych np.
 - Obliczenie sumy zamówień klientów
 - Obliczenie sumy zamówień klientów w poszczególnych kwartałach
 - Obliczenie średnich ocen studentów
 - Obliczenie średnich ocen studentów w poszczególnych semestrach
- Grupowanie może następować po kilku kategoriach, których nazwy pojawią się jako wartości w kolumnach
- Wyniki podsumowania są wypisywane w kolumnie
- Liczba podsumowań nie jest ograniczona

Przykład kwerendy podsumowującej

```
SELECT      student.nazwisko, oceny.semestr, AVG(oceny.ocena) AS Srednia,  
            COUNT(oceny.ocena) AS [Liczba ocen]  
FROM        student INNER JOIN  
            oceny ON student.id_studenta = oceny.id_studenta  
GROUP BY    student.nazwisko, oceny.semestr  
ORDER BY    student.nazwisko, oceny.semestr
```

Wynik kwerendy podsumowującej

Nazwisko	Semestr	AVG(Ocena)	Count(Ocena)
Student1	1	3.333	9
Student1	2	4.000	12
Student1	3	3.800	13
Student1	4	4.200	12
Student2	1	3.500	8
Student2	2	3.900	12
Student2	3	4.000	13



Kwerenda krzyżowa

- Podobna do kwerendy podsumowującej
- Wybrane kategorie grupowania pojawiają się jako nagłówki kolumn
- Wybrane kategorie grupowania pojawiają się jako wartości w pierwszych kolumnach tabeli wynikowej – nagłówki wierszy
- Wyniki podsumowań pojawiają się jako wartości w wierszach
- MS Access i SQL Server posiada narzędzia ułatwiające tworzenie kwerend krzyżowych



Wynik kwerendy krzyżowej

Nazwisko	Sem. 1	Sem. 2	Sem. 3	Sem. 4
Student1	3.333	4.000	3.800	4.200
Student2	3.500	3.900	4.000	4.300
Student3	4.000	4.300	4.500	4.800
Student4	3.350	4.100	3.900	4.100
Student5	3.700	3.900	4.300	4.200
Student6	4.300	4.300	4.500	4.800
Student7	3.200	4.100	3.500	4.200



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz Społeczny



Przykład AdventureWorks2014

Zapytanie grupujące

```
USE AdventureWorks ;  
GO  
SELECT DaysToManufacture,  
       AVG(StandardCost) AS AverageCost  
FROM Production.Product  
GROUP BY DaysToManufacture;
```

Wynik zapytania

DaysToManufacture	AverageCost
-----	-----
0	5.0885
1	223.88
2	359.1082
4	949.4105



Przykład AdventureWorks

Zapytanie krzyżowe

```
-- Pivot table with one row and five columns
SELECT 'AverageCost' AS Cost_Sorted_By_Production_Days,
[0], [1], [2], [3], [4]
FROM
(SELECT DaysToManufacture, StandardCost
     FROM Production.Product) AS SourceTable
PIVOT
(
AVG(StandardCost)
FOR DaysToManufacture IN ([0], [1], [2], [3], [4])
) AS PivotTable;
```

Wynik

Cost_Sorted_By_Production_Days	0	1	2	3	4
AverageCost	5.0885	223.88	359.1082	NULL	949.4105

Przykład AdventureWorks

Zapytanie krzyżowe

```
USE AdventureWorks2008R2;
GO
SELECT VendorID, [250] AS Emp1, [251] AS Emp2, [256] AS Emp3, [257]
      AS Emp4, [260] AS Emp5
FROM
  (SELECT PurchaseOrderID, EmployeeID, VendorID
  FROM Purchasing.PurchaseOrderHeader) p
PIVOT
  (
    COUNT (PurchaseOrderID)
  FOR EmployeeID IN
    ( [250], [251], [256], [257], [260] )
  ) AS pvt
ORDER BY pvt.VendorID;
```

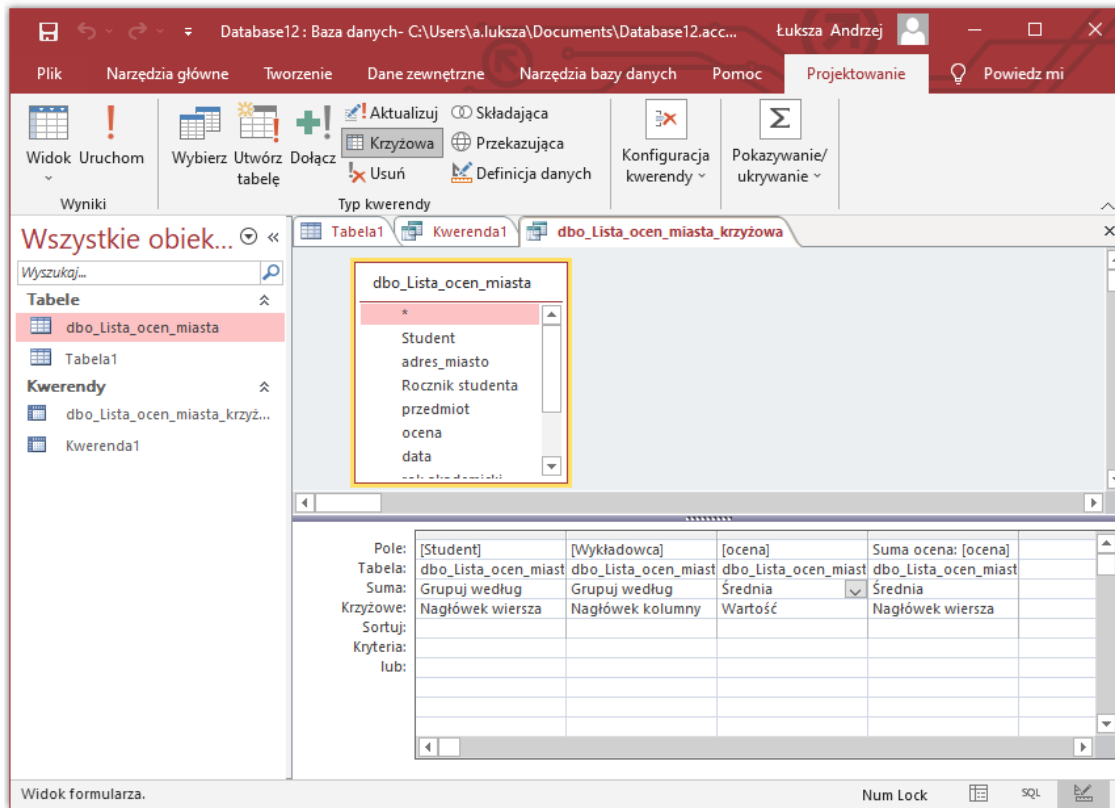
Wynik zapytania

VendorID	Emp1	Emp2	Emp3	Emp4	Emp5
1492	2	5	4	4	4
1494	2	5	4	5	4
1496	2	4	4	5	5
1498	2	5	4	4	4
1500	3	4	4	5	4
1504	2	5	5	4	5
1506	2	4	5	5	5
1508	2	4	4	6	5
1510	2	4	4	5	5
1514	2	4	4	5	4
1516	1	5	5	4	5
1518	0	0	1	0	0
1520	0	0	0	0	1



Przykład kwerendy krzyżowej Access

Projekt kwerendy krzyżowej



Wynik

Student	Suma ocena	dr Czikonsek Darr	dr Kowalski	dr Michał Sa	majs_ Woźn	mgr Antoni
Analfabeta Maciej	4,125					4,
Antoni Wolowski	4					
Artowski Kacper	4,5					
Aston Marian	4			3,5		
Bambi Jan	4,1	5			4	
Binos Alojzy	3,5		2	5		
Bisior Arek	5			5		
Biznesowy Janusz	2,66666666666667		2	2		
Blendzior Jan	3					
Blok Kacper	5		5			
Błaszczkowski Kubaś	4					
Bobas Tomek	3,9		4	5		3
Boguń Kacper	5		5			
Bogusz Rafał	3					
Brajan Nosacz	4,25			4		
Bret Albercik	4,72222222222222	4,5	5	3,5	5	
Broziewicz Alojzy	4,27777777777778	4,5		4,5	4,5	4,
Bster Adrian	3,69230769230769	3,66666666666667	2,5	4		



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz Społeczny



Silny klient

- Komputery PC stały się tanie i mają duże moce obliczeniowe
- Część obliczeń związanych z analizą danych może być wykonana na komputerze klienckim odciążając serwer
- Dane są przechowywane na serwerze i przesyłane klientowi, który dokonuje ich analizy przy pomocy programu działającego na komputerze klienckim
- Najpopularniejszymi programami dokonującymi analizy danych są arkusze kalkulacyjne

Raport tabeli przestawnej

- Realizowany przez arkusz kalkulacyjny np. MS Excel, LibreOffice, ...
- Dane pochodzą z samego arkusza lub są pobierane z bazy danych
- Jeśli dane są pobierane z bazy danych możliwa jest ich aktualizacja

	A	B	C	D	E	F	G	H	I
1	Średnia z ocena	Etykiety kolumn							
2	Etykiety wierszy	1	2	3	4	5	6	7	Suma końcowa
3	▢ Aabik Jan	3,50	3,50		3,50				3,50
4	Basen	3,50	3,50		3,50				3,50
5	Malarstwo piękne	3,50							3,50
6	Wyciskanie sztangi	3,50							3,50
7	▢ Albini Adrian	4,50	4,50	4,50	3,00	3,00	4,00		4,18
8	Basen	5,00	4,50	4,50	3,00				4,25
9	Fizyka	3,50							3,50
10	Matematyka					3,00			3,00
11	Piłka siatkowa	4,50							4,50
12	Plastyka	5,00							5,00
13	Teorie Spiskowe	5,00							5,00
14	TI	4,00							4,00
15	Wyciskanie sztangi						4,00		4,00
16	▢ Alferd Garbus	5,00	4,00		5,00				4,67
17	Basen	5,00							5,00
18	Bazy danych		4,00						4,00
19	TI				5,00				5,00
20	▢ Analfabeta Maciej	4,25	3,50	4,50					4,13
21	Basen	4,25	3,50	4,50					4,13



SQL Server Analysis Services

- Tworzy wielowymiarowe struktury danych na podstawie źródeł danych OLTP
 - MOLAP (*Multidimensional OLAP*) – dane są pobierane z innych baz danych po czym wielowymiarowa struktura danych zapisywana jest na trwałym nośniku, przez serwer OLAP
 - ROLAP (*Relational OLAP*) – na serwerach OLTP tworzone są tabele zbiorcze zawierające pogrupowane dane, na ich podstawie serwer OLAP tworzy wielowymiarową strukturę danych ale jej nie przechowuje (nie zapisuje na dysku)
 - HOLAP (*Hybrid OLAP*) – połączenie technik MOLAP i ROLAP – centralna baza pozostaje na serwerze OLTP a zagregowane dane są przechowywane na serwerze OLAP



Zastosowania

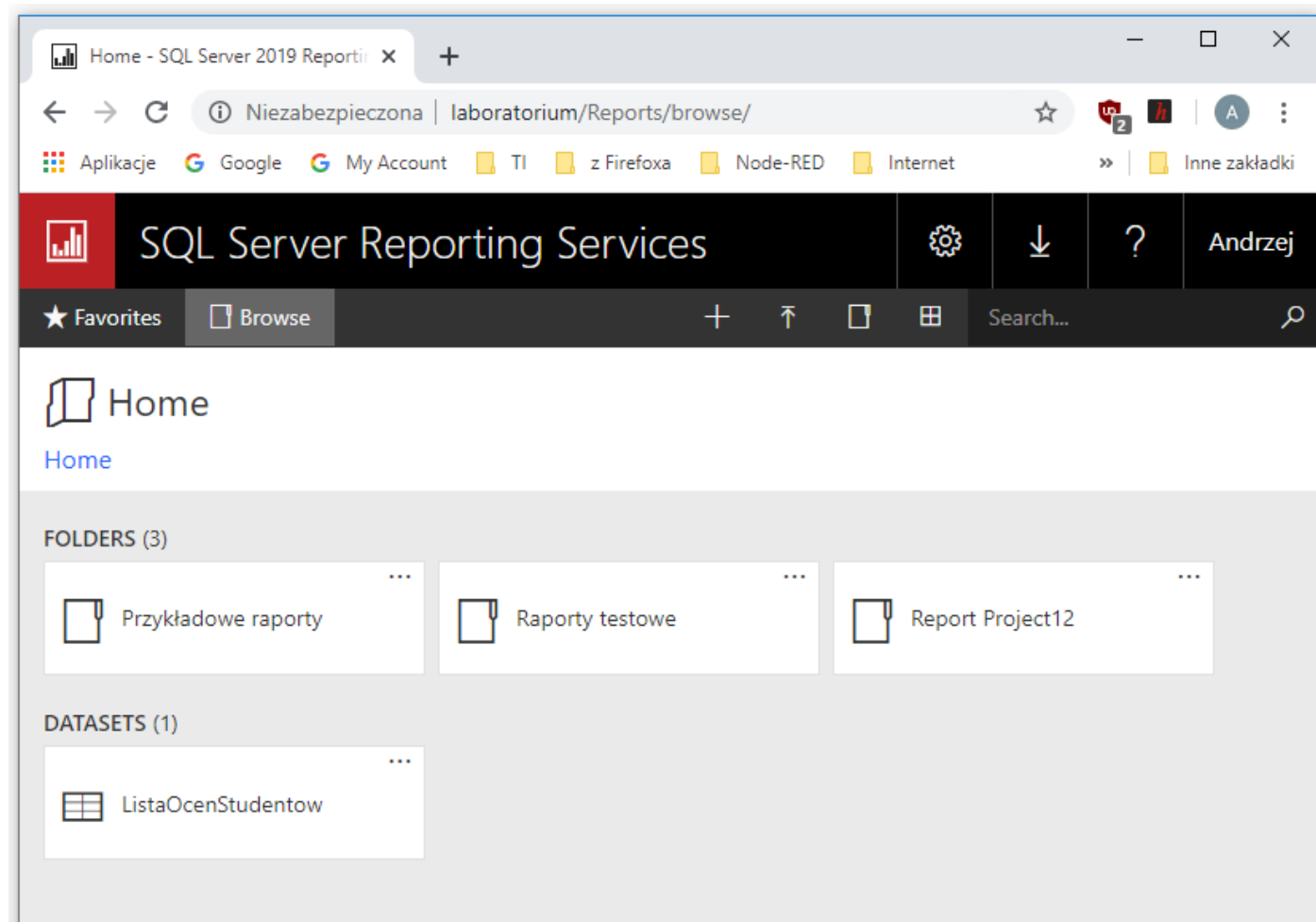
- Analiza przyczyn obserwowanych zjawisk ekonomicznych, społecznych itp.
- Analiza upodobań konsumentów
- Walka z przestępczością i terroryzmem
- Przewidywanie potencjalnych zagrożeń – ekonomicznych czy związanych z bezpieczeństwem
- Planowanie strategii rozwoju organizacji
- Badania socjologiczne – pisanie rozpraw naukowych
- Przewidywanie przyszłości ;), opracowywanie prognoz
- Oddziaływanie na opinię publiczną, analiza skuteczności PR



Eksploracja danych

- Eksploracja danych jest procesem wydobywania nieznanych wcześniej informacji z dużych baz danych i wykorzystania ich przy podejmowaniu decyzji
- Techniki eksploracji danych
 - Modelowanie predykcyjne (modelownie sposobu uczenia)
 - Klasyfikacja i predykcja wartości
 - Segmentacja bazy danych
 - Analiza powiązań
 - Wykrywanie odchyleń
- Każda technika może być realizowana przy pomocy różnych algorytmów
- Istnieje wiele produktów realizujących różne algorytmy eksploracji danych

<http://lukan.sytes.net/reports/browse/raporty>



Projekt „SezAM wiedzy, kompetencji i umiejętności” jest współfinansowany przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Wiedza Edukacja Rozwój

SQL Server Reporting Services (SSRS)

Konfiguracja Serwera Raportów

Report Server Configuration Manager: LABORATORIUM\SSRS

Report Server Configuration Manager

Connect

- LABORATORIUM\SSRS
- Service Account
- Web Service URL**
- Database
- Web Portal URL
- E-mail Settings
- Execution Account
- Encryption Keys
- Subscription Settings
- Scale-out Deployment
- Power BI Service (cloud)

Web Service URL

Configure a URL used to access the Report Server. Click Advanced to define multiple URLs for a single Report Server instance, or to specify additional parameters on the URL.

Report Server Web Service Virtual Directory

Virtual Directory:

Report Server Web Service Site identification

IP Address:

TCP Port:

HTTPS Certificate:

HTTPS Port:

Report Server Web Service URLs

URLs: <http://LABORATORIUM:80/ReportServer>
<https://localhost:443/ReportServer>

Results



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz Społeczny



SAP Crystal Reports, version for Visual Studio

Link do strony, z której można pobrać pliki instalacyjne, po założeniu konta

<https://www.sap.com/cmp/td/sap-crystal-reports-visual-studio-trial.html>

Link do strony z której można pobrać pliki instalacyjne bez zakładania konta.
Strona zawiera szczegółową instrukcję instalacji

<https://www.tektutorialshub.com/crystal-reports/download-crystal-reports-for-visual-studio-2019/>

Aplikacja Cristal Report

Utwórz nowy projekt

Ostatnie szablony projektów

- Aplikacja internetowa platformy ASP.NET (.NET Framework) C#
- Aplikacja Crystal Reports C#
- Aplikacja internetowa ASP.NET Core C#
- Aplikacja konsolowa C#
- Aplikacja konsoli (.NET Framework) C#
- Aplikacja internetowa platformy ASP.NET (.NET Framework) Visual Basic

C# Windows Wszystkie typy projektów

Aplikacja WPF programu Crystal Reports
Projekt do tworzenia aplikacji WPF języka C# z interfejsem użytkownika systemu Windows

C# Windows Klasyczny Reporting

Aplikacja Crystal Reports
Projekt do tworzenia aplikacji języka C# z interfejsem użytkownika systemu Windows i przykładowym raportem programu Crystal Reports

C# Windows Klasyczny Reporting

Witryna sieci Web ASP.NET programu Crystal Reports
Witryna sieci Web ASP.NET z przykładowym raportem programu Crystal Reports

C# Windows Internet Reporting

Inne wyniki oparte na wyszukiwaniu

Aplikacja WPF programu Crystal Reports
Projekt do tworzenia aplikacji WPF języka VB z interfejsem użytkownika systemu Windows

Visual Basic Windows Klasyczny Reporting

Dalej

Crystal Reports — galeria

Tworzenie nowego raportu programu Crystal Reports

☒ Za pomocą Kreatora raportów

☐ Jako pustego raportu

☐ Z istniejącego raportu

Wybierz eksperta

Standardowy

Macierz

Etykieta poczty



Pomaga w tworzeniu typowego raportu.

OK

Pomoc

Wybór źródła danych

Kreator raportów standardowych

Dane
Wybierz dane, których ma dotyczyć raport.

Dostępne źródła danych:

- + Dane projektu
- Moje połączenia
 - + 127.0.0.1
 - + 127.0.0.1_1
 - argo.am.gdynia.pl
 - + Dodaj polecenie
 - studenci
 - dbo
 - + Tabele
 - lista_ocen
 - Widoki
 - do_normalizacji
 - lista_ocen
 - Lista_ocen_miasta
 - raportoceny
 - View_1
 - + Procedury wbudowane
 - + INFORMATION_SCHEMA
 - + sys

Wybrane tabele:

- argo.am.gdynia.pl
 - lista_ocen

< Wstecz Dalej > Zakończ Anuluj Pomoc

Kreator raportów standardowych

Pola
Wybierz informacje do wyświetlenia w raporcie.

Dostępne pola:

- lista_ocen
 - Student
 - Rocznik studenta
 - przedmiot
 - ocena
 - data
 - rok akademicki
 - specjalność
 - semestr
 - Wykładowca

Pola do wyświetlenia:

- lista_ocen.Student
- lista_ocen.semestr
- lista_ocen.przedmiot
- lista_ocen.ocena
- lista_ocen.data
- lista_ocen.Wykładowca

Przeglądaj dane... Znajdź pole...

< Wstecz Dalej > Zakończ Anuluj Pomoc

Grupowanie i podsumowania

Kreator raportów standardowych

Grupowanie
(Opcjonalnie) Zgrupuj informacje zawarte w raporcie.

Dostępne pola:

- Pola raportu
 - lista_ocen.Student
 - lista_ocen.semestr
 - lista_ocen.przedmiot
 - lista_ocen.ocena
 - lista_ocen.data
 - lista_ocen.Wykładowca
- lista_ocen
 - Student
 - Rocznik studenta
 - przedmiot
 - ocena
 - data
 - rok akademicki
 - specjalność
 - semestr

Grupuj według:

- lista_ocen.Student - A
- lista_ocen.semestr - A

Przeglądaj dane... Znajdź pole...

w kolejności rosnącej.

< Wstecz Dalej > Zakończ Anuluj Pomoc

Kreator raportów standardowych

Podsumowania
(Opcjonalnie) Dodaj do raportu informacje sumaryczne.

Dostępne pola:

- Pola raportu
 - lista_ocen.Student
 - lista_ocen.semestr
 - lista_ocen.przedmiot
 - lista_ocen.ocena
 - lista_ocen.data
 - lista_ocen.Wykładowca
- lista_ocen
 - Student
 - Rocznik studenta
 - przedmiot
 - ocena
 - data
 - rok akademicki
 - specjalność
 - semestr

Pola podsumowań:

- lista_ocen.Student
 - Średnia z lista_ocen.ocena
- lista_ocen.semestr
 - Średnia z lista_ocen.ocena

Average

Przeglądaj dane... Znajdź pole...

< Wstecz Dalej > Zakończ Anuluj Pomoc



Szablon raportu

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
	▼ Section1 (Nagłówek raportu)																				
	▼ Section2 (Nagłówek strony)																				
	Data wydru																				
	Studen					semestr		przedmio						ocena		data			Wykładow		
	▼ GroupHeaderSection1 (Nagłówek grupy nr 1: lista_ocen.Student – A)																				
	Nazwa grupy nr 1																				
	▼ GroupHeaderSection2 (Nagłówek grupy nr 2: lista_ocen.semestr – A)																				
	Nazwa																				
	▼ Section3 (Szczegóły)																				
	Student					semestr		przedmiot						ocena		data			Wykładowca		
	▼ GroupFooterSection2 (Stopka grupy nr 2: lista_ocen.semestr – A)																				
	ednia z lista_ocen.ocena Nazwa																				
	▼ GroupFooterSection1 (Stopka grupy nr 1: lista_ocen.Student – A)																				
	Nazwa grupy nr 1 i.ocena																				
	▼ Section4 (Stopka raportu)																				
	Suma łączna: en.ocena																				
	▼ Section5 (Stopka strony)																				



Początkowy wygląd raportu

Form1

1 /1+ SAP CRYSTAL REPORTS

Raport główny

2023-05-24

Student	semestr	przedmiot	ocena	data
Aabik Jan				
Aabik Jan				
Aabik Jan				
Aabsolwent Imiennik				
Aabsolwent Imiennik				
Aabsolwent Imiennik				
Aaron Jan				
Aaron Jan	1	1 J. angielski	2,00	2015.12.22
Aaron Jan	1	1 Plastyka	4,00	2015.12.22
Aaron Jan	1	1 J. angielski	3,00	2015.12.22
	3,00	1		
	2			
Aaron Jan	2	2 WF	4,00	2015.12.22
Aaron Jan	2	2 WF	2,00	
Aaron Jan	2	2 WF	3,00	2019.11.13
Aaron Jan	2	2 Plastyka	4,00	2015.12.22
	3,25	2		
	3			

Nr bieżącej strony: 1 Całkowita liczba stron: 1+ Współczynnik powiększenia: 100%

Projekt „SezAM wiedzy, kompetencji i umiejętności” jest współfinansowany przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Wiedza Edukacja Rozwój



Fundusze Europejskie
Wiedza Edukacja Rozwój

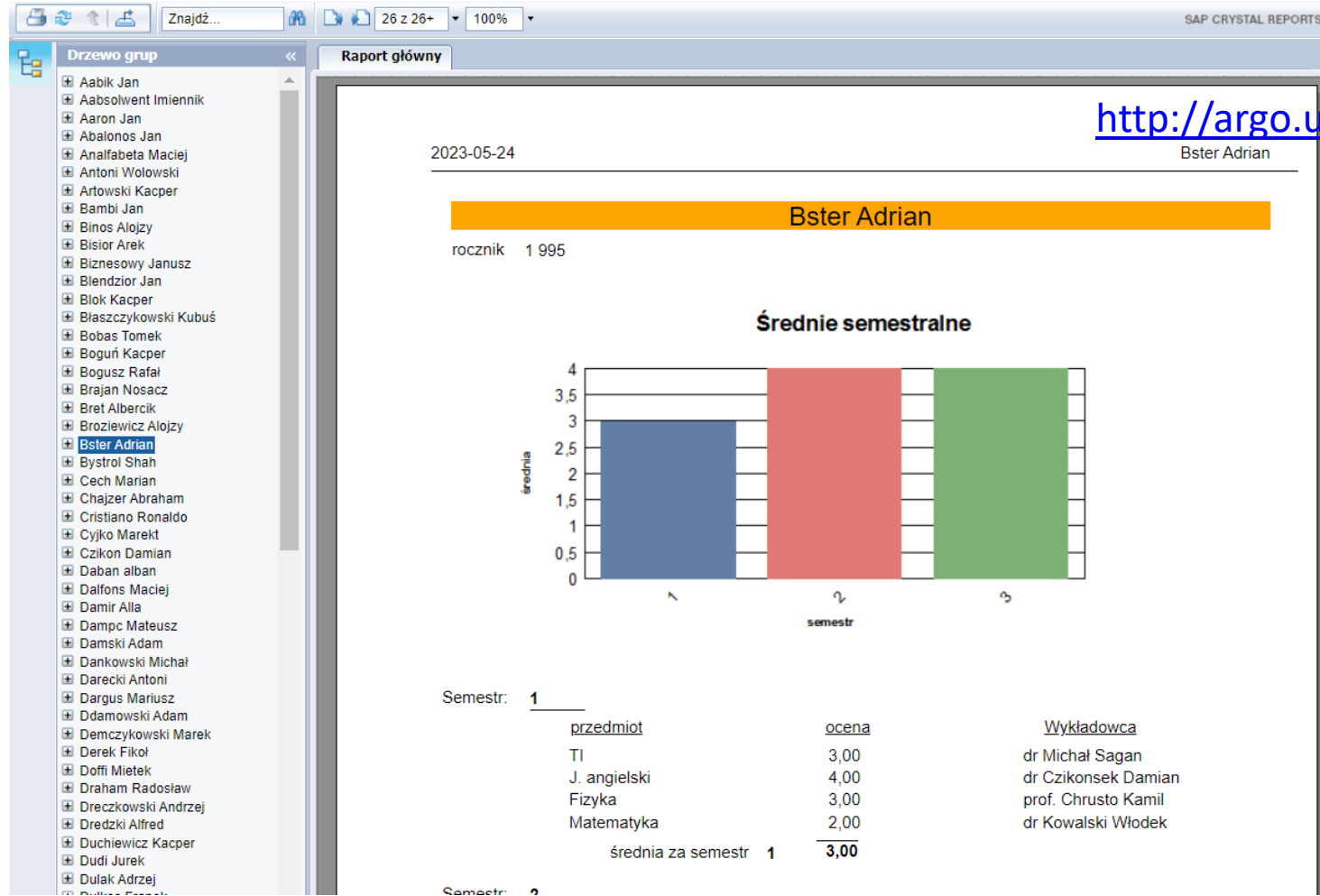


Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz Społeczny



Raporty po dopracowaniu



<http://argo.umg.edu.pl/www/aplikacje/cw17/>

Projekt „SezAM wiedzy, kompetencji i umiejętności” jest współfinansowany przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Wiedza Edukacja Rozwój



Przeglądarka raportów

