



KATEDRA
INŻYNIERII
OPROGRAMOWANIA

Serwery OLAP

Krzysztof Goczyła
Teresa Zawadzka
Wojciech Waloszek

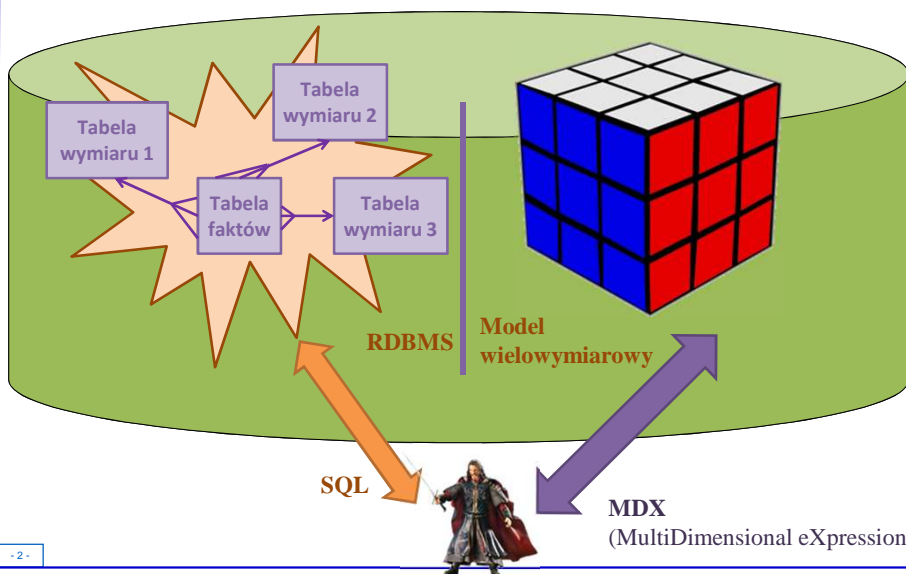
*Katedra Inżynierii Oprogramowania
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska
{kris, tegra}@eti.pg.gda.pl*



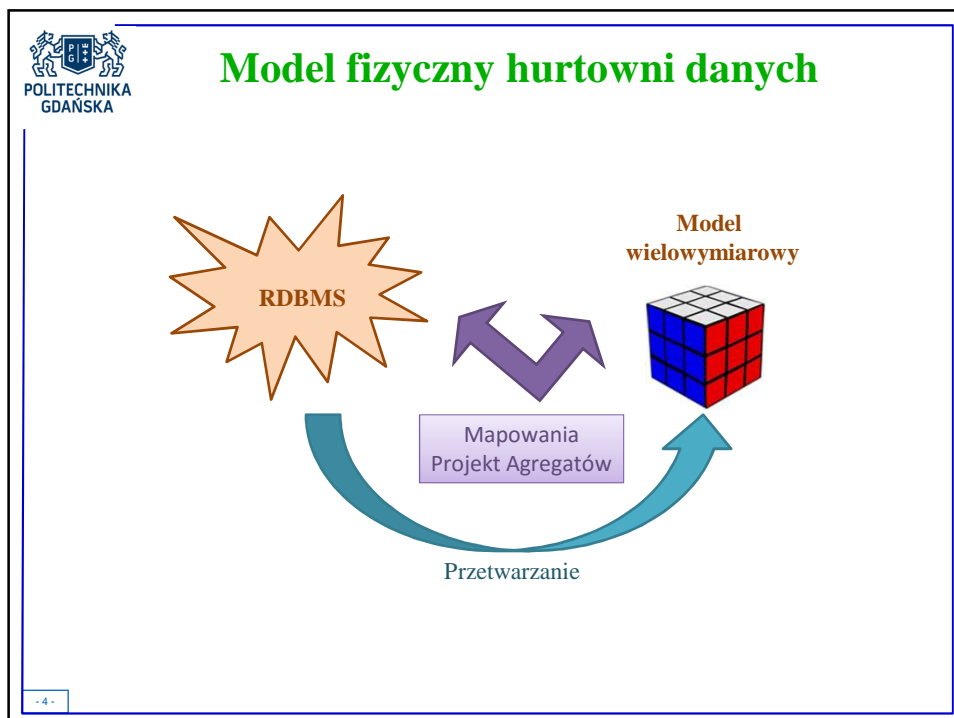
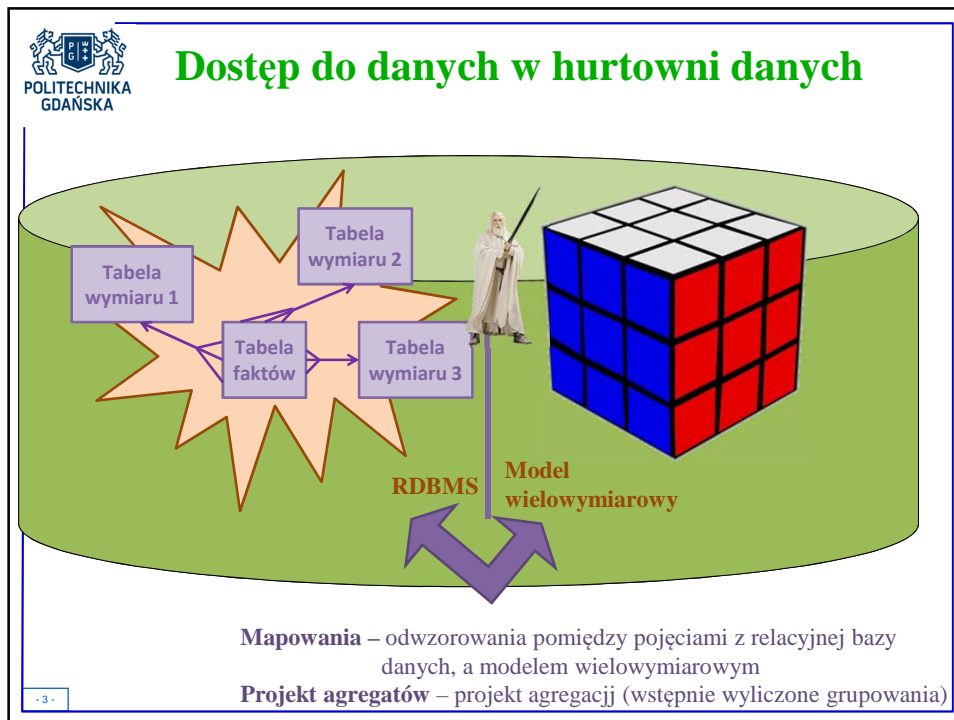
- 1 -

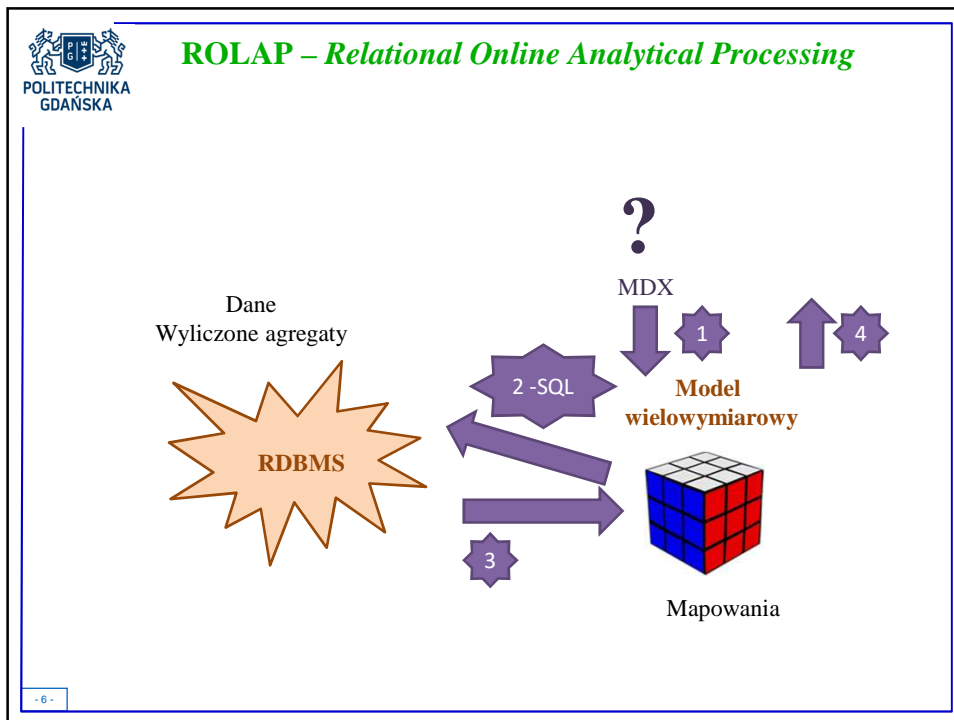
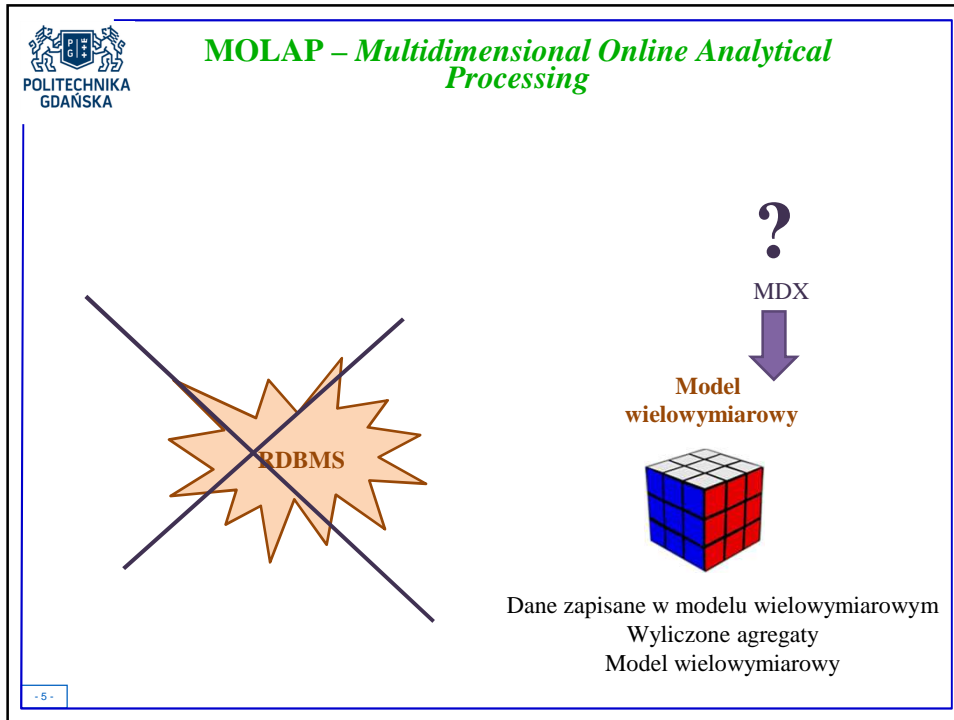


Serwer OLAP czy RDBMS?



- 2 -







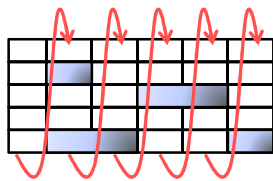
HOLAP – Hybrid Online Analytical Processing

- Połączenie modelu ROLAP i MOLAP;
- Część danych jest przechowywana w strukturach relacyjnych, część w strukturach wielowymiarowych;
- Konkretnie rozwiązania zależne od dostawców rozwiązań typu BI.

-7-



Serwery OLAP – kolumnowa organizacja danych



Kolumnowe bazy danych różnią się sposobem przechowywania danych: przechowują dane w sposób *kolumnowy*.

Pobyt pacjenta w szpitalu

Id_pacjenta

Data przyjęcia

Data wypisania

liczba operacji

liczba procedur

koszt pobytu

refundacja_NFZ

-8-



Przechowywanie danych wiersz-po-wierszu

Id pacjenta	Id szpitala	Data przyjęcia	Data wypisania	Liczba operacji	Liczba procedur	Koszt pobytu	refundacja
1	5	2012.08.09	2012.08.30	1	12	1200,-	1200,-
1	5	2012.10.10	2012.10.15	0	3	400,-	350,-
2	5	2012.08.09	2012.08.11	0	3	100,-	100,-
3	6	2012.09.14	2012.09.20	1	15	1900,-	1300,-
4	6	2012.10.10	2012.10.11	0	2	50,-	50,-
5	5	2012.10.11	2012.10.30	1	7	1000,-	1000,-
...	

Pobyt pacjenta w szpitalu (
 (1, 5, 2012.08.09, 2012.08.30, 1, 0, 12, 1200, 1200),
 (1, 5, 2012.10.10, 2012.10.15, 0, 3, 400, 350),
 (2, 5, 2012.08.09, 2012.08.11, 0, 3, 100, 100),
 (3, 6, 2012.09.14, 2012.09.20, 1, 15, 1900, 1300),
 ...)

- 9 -



Przechowywanie danych kolumna-po-kolumnie

Pobyt pacjenta w szpitalu (
 (1, 5, 2012.08.09, 2012.08.30, 1, 0, 12, 1200, 1200),
 (1, 5, 2012.10.10, 2012.10.15, 0, 3, 400, 350),
 (2, 5, 2012.08.09, 2012.08.11, 0, 3, 100, 100),
 (3, 6, 2012.09.14, 2012.09.20, 1, 15, 1900, 1300),
 ...)

Pobyt pacjenta w szpitalu (
 (1, 1, 2, 3, 4, 5, ...),
 (5, 5, 5, 6, 6, 5, ...),
 (2012.08.09, 2012.10.10, 2012.08.09, 2012.09.14, 2012.10.10, 2012.10.11, ...),
 (2012.08.30, 2012.10.15, 2012.08.11, 2012.09.20, 2012.10.11, 2012.10.30, ...),
 (1, 0, 0, 1, 0, 1, ...),
 (12, 3, 3, 15, 2, 7, ...),
 (1200, 400, 100, 1900, 50, 1000, ...),
 (1200, 350, 100, 1300, 50, 1000, ...)
)

- 10 -



Po co?

1. W relacyjnej bazie (nieco upraszczając) istnieją dwa sposoby dostępu do danych:
 - przez indeks,
 - przez skanowanie.

2. Większość zapytań analitycznych musi skanować całą zawartość tabeli, np.:

```
SELECT ID_Pacjenta, SUM(liczba_procedur)
FROM Pobyty_pacjenta_w_szpitalu
GROUP BY ID_Pacjenta
```

- 11 -



Kolumnowe = szybkość

```
SELECT ID_Pacjenta, SUM(liczba_procedur)
FROM Pobyty_pacjenta_w_szpitalu
GROUP BY ID_Pacjenta
```

1. **Baza wierszowa** musi przejrzeć całość tabeli (ze względu na specyfikę odczytu z dysku nie ma sensu „omijanie” fragmentów rekordów)
2. **Baza kolumnowa** przejrzę tylko potrzebne kolumny (Id_pacjenta i liczba_procedur).

- 12 -



Co każdy student wiedzieć powinien...

Jakie są modele przetwarzania analitycznego?
Czym jest kolumnowa organizacja danych?

