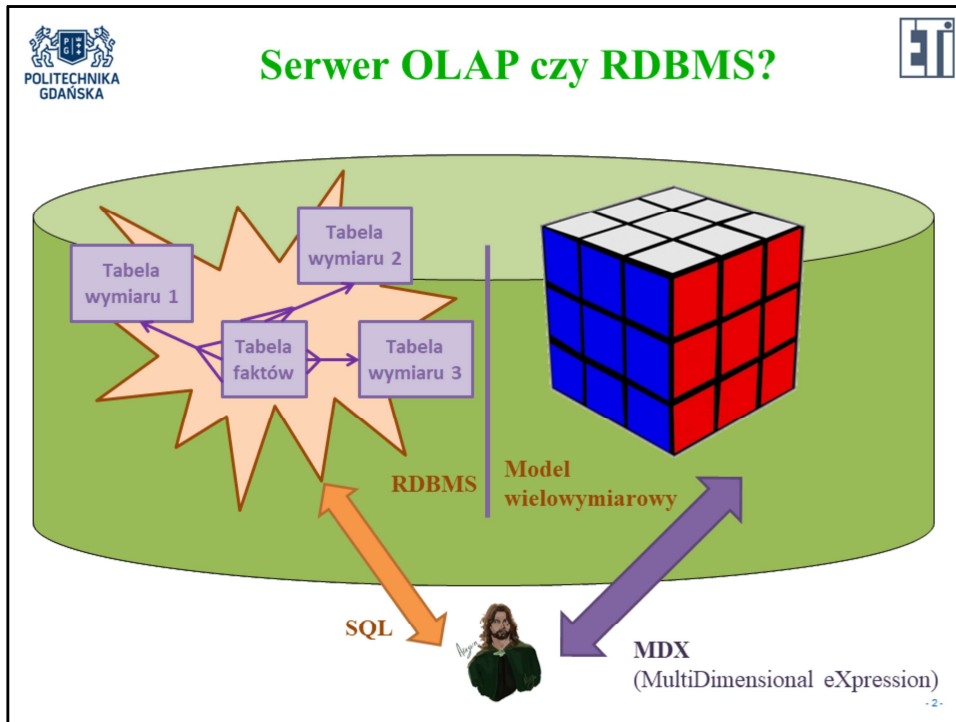


Serwery OLAP

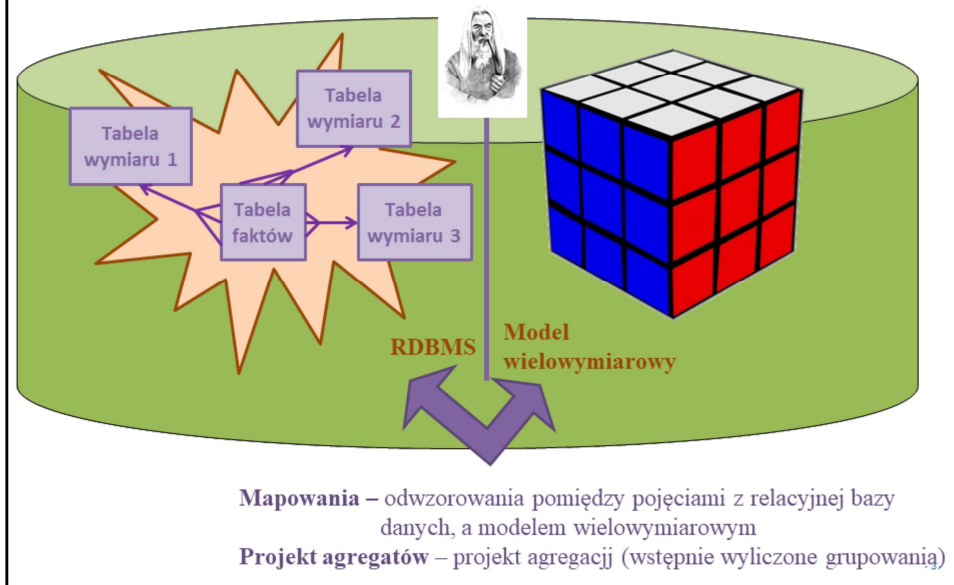


- 1 -

W poprzednich wykładach dużo mówiliśmy o modelu wielowymiarowym. Model relacyjnych (schemat gwiazdy, płątka śniegu, konstelacja) jest oczywiście przechowywany w systemach RDBMS. Gdzie zaś jest zapisywany model wielowymiarowy? Model wielowymiarowy jest przechowywany na serwerach OLAP.

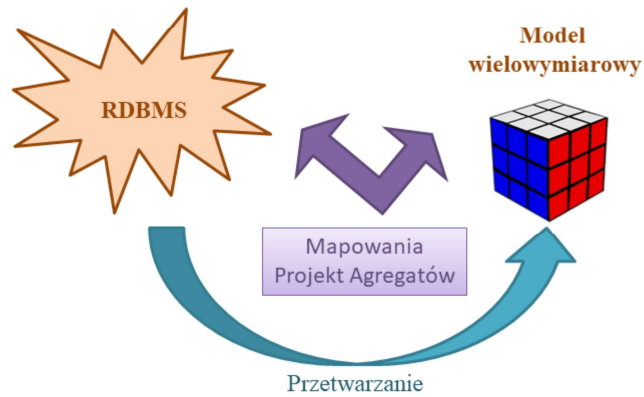


Dostęp do systemów RDBMS zapewnia język SQL. W przypadku serwerów OLAP dostęp zapewniony jest poprzez język MDX.



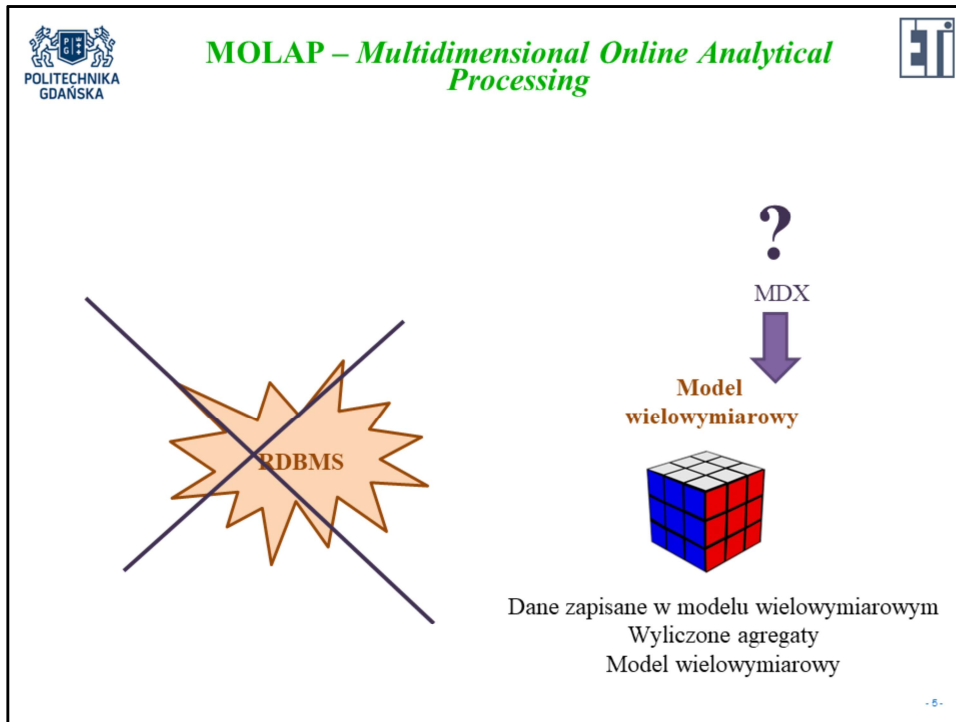
Jak już Państwo wiedzą elementy modelu wielowymiarowego są skojarzone z odpowiednimi elementami modelu relacyjnego (np.: wymiary z tabelami wymiarów). Odwzorowania pomiędzy pojęciami z modelu wielowymiarowego, a modelu relacyjnego nazwijmy mapowaniami. Ponadto, możliwe (nie konieczne) jest zdefiniowanie projektu agregatów. Projekt agregatów to nic innego jak wstępnie wyliczone zagregowane wartości miar. Definiujemy, które wartości miar mają być wstępnie zagregowane w odniesieniu do jakich wymiarów.

Model fizyczny hurtowni danych

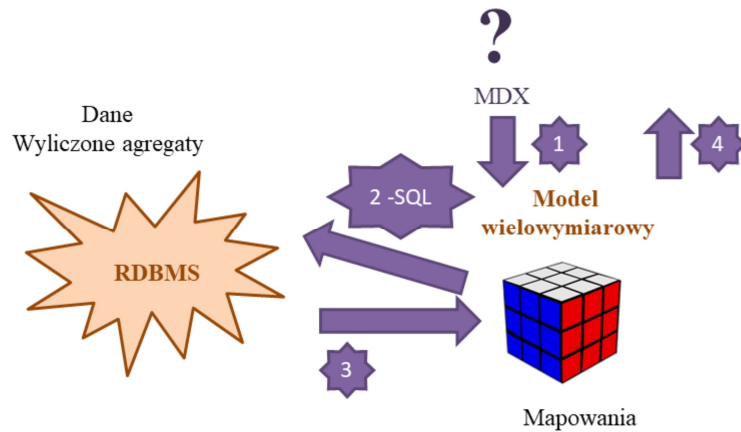


- 4 -

Mając zdefiniowany schemat relacyjny oraz określone mapowania i projekt agregatów możliwe jest przetworzenie kostki. Przetworzenie kostki to nic innego jak utworzenie modelu wielowymiarowego na serwerze OLAP. Model ten jest budowany w oparciu o schemat relacyjny, mapowania i projekt agregatów. Na serwerze OLAP utworzony model wielowymiarowy może być modelem ROLAP, MOLAP lub HOLAP.



Model MOLAP zakłada, że wszystkie dane są przechowywane w modelu wielowymiarowym. Oznacza to, że na serwerze OLAP przechowywany jest schemat kostki, mapowania, wszystkie dane oraz wyliczone agregaty. Jeżeli zadamy zapytanie MDX do serwera OLAP, a przyjęliśmy model MOLAP kostki, to nie musimy mieć dostępnego systemu RDBMS. Zapytanie jest wykonywane wyłącznie po stronie serwera OLAP.



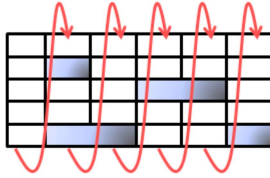
W przypadku modelu ROLAP na serwerze OLAP przechowywane są tylko metadane kostki. Zadane zapytanie MDX skierowane do serwera OLAP, przy założeniu że kostka zbudowana jest w modelu ROLAP, jest wykonywane w następujący sposób:

1. Zapytanie MDX jest zadawane w terminach modelu wielowymiarowego.
2. Serwer OLAP na podstawie metadanych kostki przekształca zapytanie MDX do zapytania SQL i wysyła to zapytanie SQL do systemu RDBMS.
3. Wynik zapytania SQL jest zwracany do serwera OLAP, a następnie jest przekształcany do terminów znanych w modelu wielowymiarowym.
4. Wynik w terminach modelu wielowymiarowego jest zwracany użytkownikowi końcowemu.

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

Istnieją jeszcze modele typu HOLAP, będące połączeniem modelu ROLAP i MOLAP. Istnieje wiele implementacji takiego modelu.

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

Serwery OLAP charakteryzują się kolumnową organizacją danych. Czerwone strzałki wskazują w jaki sposób dane są przechowywane (w odróżnieniu od sposobu przechowywania danych w systemach relacyjnych).

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 -

Dane w relacyjnej bazie danych są przechowywane wiersz po wierszu.

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 -

W kolumnowej bazie danych są zapisywane kolumna po kolumnie. Oznacza to, że najpierw przechowywane są wszystkie id pacjenta, później wszystkie id szpitala, daty przyjęcia itd.

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 Pobyt_pacjenta_w_szpitalu
GROUP BY ID_Pacjenta
```

Aby wykonać zapytanie analityczne silnik relacyjnej bazy danych musi odczytać całą zawartość tabeli, kiedy tak naprawdę (dla przykładowego zapytania) potrzebne są tylko wartości kolumny ID_Pacjenta i liczba procedur. Ze względu na fakt, że tabele faktów są bardzo duże odczytywanie całych wierszy jest mało wydajne. Stąd serwery OLAP przechowują dane w tabelach o organizacji kolumnowej.

Kolumnowe = szybkość



```
SELECT ID_Pacjenta, SUM(liczba_procedur)
FROM Pobyt_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** przejrzy tylko potrzebne kolumny (Id_pacjenta i liczba_procedur).

Dzięki reprezentacji kolumnowej serwer OLAP może odczytać jedynie potrzebne kolumny.

Co każdy student wiedzieć powinien?



- **Jakie mamy modele przetwarzania danych?**
- **Na czym polega kolumnowa organizacja danych?**

