

Optymalizacja hurtowni danych - raport

Jakub Nowak 197860, Piotr Staszko 197938

1. Cel laboratorium

Celem zadania jest przedstawienie problemów dotyczących różnych **fizycznych modeli kostek** takich jak MOLAP (Multidimensional OLAP) czy ROLAP (Relational OLAP).

2. Specyfikacja

liczba faktów - 225 288

rozmiar hurtowni danych - 240,3 MB

Pomiary zostały wykonane na laptopie wyposażonym w procesor **AMD Ryzen 7 6800H**, **32 GB pamięci RAM** oraz **1 TB pamięci wewnętrznej**. Używaliśmy systemu **Windows 11**. Do oceny czasu przetwarzania kostki (ang. cube) użyliśmy **SQL Server Management Studio (SSMS)** z rozszerzeniem **SQL Server Profiler**.

3. Teoretyczne założenia

	MOLAP	ROLAP
Czas wykonywania zapytań	Krótki	Długi
Czas procesowania kostki	Długi	Krótki
Całkowity rozmiar	Duży	Mały

4. Testowanie

Testowaniu podlegają 3 różne zapytania mdx oraz czas przetwarzania kostki. Brane pod uwagę są dwa modele: MOLAP i ROLAP

1. Ilu uczniów miało wynik z matury <15%?

```
SELECT
NON EMPTY { [Measures].[Wynik z Matury] } ON COLUMNS,
NON EMPTY {
    FILTER(
        (
            [Dim Ucen].[Imie Nazwisko].[Imie Nazwisko].ALLMEMBERS * EXCEPT(
                [ID Daty Matury].[Rok].[Rok].ALLMEMBERS,
                [ID Daty Matury].[Rok].[Rok].[Unknown]
            ) * [Dim Przedmiot].[Nazwa].[Nazwa].ALLMEMBERS
        ),
        [Measures].[Wynik z Matury] > 0 AND [Measures].[Wynik z Matury] < 15
    )
}
ON ROWS
FROM [Szkola30]
```

2. Czy uczniowie, którzy wcześniej zaczęli chodzić do szkoły (są młodsi) lepiej / gorzej napisali maturę?

```
SELECT
NON EMPTY { [Measures].[sredni_wynik_z_matury] } ON COLUMNS,
NON EMPTY {
    FILTER(
        (
            [Dim Ucen].[Wiek].[Wiek].ALLMEMBERS * [Dim Ucen].[Imie Nazwisko].[Imie Nazwisko].ALLMEMBERS
        ),
        [Measures].[sredni_wynik_z_matury] > 0
    )
}
ON ROWS
FROM [Szkola30]
```

3. Jaki był najwyższy wynik z matury dla ucznia z oceną 2?

```
SELECT
NON EMPTY { [Measures].[Wynik z Matury] } ON COLUMNS,
NON EMPTY {
    GENERATE(
        EXCEPT(
            [ID Daty Matury].[Rok].[Rok].ALLMEMBERS,
            [ID Daty Matury].[Rok].[Rok].[Unknown]
        ),
        TOPCOUNT(
            [ID Daty Matury].[Rok].CurrentMember * [Dim Przedmiot].[Nazwa].[Nazwa].ALLMEMBERS *
            [Dim Ucen].[Imie Nazwisko].[Imie Nazwisko].ALLMEMBERS,
            1,
            [Measures].[Wynik z Matury]
        )
    )
} ON ROWS
FROM [Szkola30]
WHERE ([Dim Junk].[Ocena].&[2])
```

Testy wykonane bez agregacji. Czas podany w [ms].

Procesowanie kostki	Zapytanie 1 (mdx 2)	Zapytanie 2 (mdx 3)	Zapytanie 3 (mdx11)
---------------------	---------------------	---------------------	---------------------

MOLAP	ROLAP	MOLAP	ROLAP	MOLAP	ROLAP	MOLAP	ROLAP
1772	373	1142	1670	807	1662	692	2025
1203	294	1045	1364	849	1588	654	896
1287	303	1072	1424	1019	1614	577	885
1142	390	1033	1580	823	1640	664	754
1208	287	980	1438	877	1542	832	1853
1201	299	970	1606	847	1496	743	789
1164	282	950	1411	847	1634	720	2677
1120	279	948	1379	903	1606	676	943
1201	280	976	1384	859	1534	734	1429
1256	294	1012	1505	880	1588	687	791

Testy wykonane przy użyciu agregacji. Czas podany w [ms].

Procesowanie kostki		Zapytanie 1 (mdx 2)		Zapytanie 2 (mdx 3)		Zapytanie 3 (mdx11)	
MOLAP	ROLAP	MOLAP	ROLAP	MOLAP	ROLAP	MOLAP	ROLAP
1642	354	894	1406	755	1752	637	823
1245	291	973	1438	827	1637	765	853
1221	343	953	1387	777	1483	680	843
1163	332	894	1220	782	1772	787	828
1211	279	1002	1291	778	1669	730	789
1421	342	954	1283	855	1485	700	722
1286	463	913	1439	880	1537	624	788
1144	368	1016	1232	885	1567	685	1974
1201	280	949	1393	967	1714	618	799
1256	294	921	1222	891	1839	675	780

Uśrednione wyniki testów [ms]

	Procesowanie kostki		Zapytanie 1 (mdx 2)		Zapytanie 2 (mdx 3)		Zapytanie 3 (mdx11)	
	MOLAP	ROLAP	MOLAP	ROLAP	MOLAP	ROLAP	MOLAP	ROLAP
Średnia (bez agregacji)	1255.4	308.1	1012.8	1476.1	871.1	1590.4	697.9	1304.2
Średnia (z agregacją)	1279	334.6	946.9	1331.1	839.7	1645.5	690.1	919.9
Średnia	1267.2	321.4	984.4	1403.6	855.4	1618	694	1112

5. Cache i agregacje

Agregacje:

Cube Objects	Default	Full	None	Unrestricted
Dim Uczeń	1	1	4	0
ID Ucznia				
Pesel				
Imie Nazwisko				
Wiek				
Start Date				
End Date				
Dim Klasa	1	1	3	0
ID Klasy				
Nazwa				
Wielkosc Klasy				
Start Date				
End Date				

Cache:

```
<ClearCache xmlns="http://schemas.microsoft.com/analysiservices/2003/engine">
  <Object>
    <DatabaseID>SzkołaDataWarehouse1</DatabaseID>
  </Object>
</ClearCache>
```

6. Podsumowanie

a) Przetwarzanie kostki

Średni czas przetwarzania kostki: **MOLAP - 1267.2 ms**, **ROLAP - 321.4 ms**. Można łatwo zaobserwować, że wykorzystując model **ROLAP** kostka przetwarza się **4 razy szybciej**. Wynika to z faktu, że ROLAP operuje bezpośrednio na danych w relacyjnej bazie i nie wymaga przechowywania wszystkich agregatów. MOLAP natomiast przetwarza pełne dane, agregaty i metadane w swojej strukturze kostki, co wydłuża czas przetwarzania, zwłaszcza przy dużej ilości danych.

b) Wykonywanie zapytań

	MOLAP [ms]	ROLAP [ms]
Zapytanie 1	984.4	1403.6
Zapytanie 2	855.4	1618
Zapytanie 3	694	1112

MOLAP wykonuje zapytania około **2 razy szybciej** niż ROLAP. Dzieje się tak, ponieważ MOLAP na etapie wykonywania zapytania ma już pobrane wszystkie potrzebne dane (po przetworzeniu kostki), nie musi odpytywać bazy o informacje i może natychmiast obliczać wynik.

c) Agregacje

	Procesowanie kostki		Zapytanie 1 (mdx 2)		Zapytanie 2 (mdx 3)		Zapytanie 3 (mdx11)	
	MOLAP	ROLAP	MOLAP	ROLAP	MOLAP	ROLAP	MOLAP	ROLAP
Średnia (bez agregacji)	1255.4	308.1	1012.8	1476.1	871.1	1590.4	697.9	1304.2
Średnia (z agregacją)	1279	334.6	946.9	1331.1	839.7	1645.5	690.1	919.9

Po przeanalizowaniu wyników można zauważyć, że użycie agregacji **przyspiesza** wykonanie zapytań zarówno dla MOLAP jak i ROLAP. Agregacje wykonują się tylko jeden raz, dzięki czemu w dalszych operacjach **zmniejszana** jest złożoność obliczeniowa, co bezpośrednio wpływa na **wydajność**. W obu modelach procesowanie kostki trwa **dłużej** przy korzystaniu z agregacji, ponieważ w tym czasie są one obliczane.

7. Wnioski

Przeprowadzone testy dostarczają wielu wartościowych informacji przedstawiających charakterystykę i działanie różnych fizycznych modeli kostek oraz wpływu agregacji na wykonywanie procesów. Obserwacje były zgodne z założeniami teoretycznymi:

- **MOLAP:** Dłuższy czas procesowania kostki, szybsze wykonywanie zapytań
- **ROLAP:** Krótszy czas procesowania kostki, wolniejsze wykonywanie zapytań
- **Agregacje:** Zwiększają wydajność wykonywania zapytań, ale wydłużają czas procesowania kostki