

# ETAP 7

## Optymalizacja hurtowni danych

### raport

#### 1. Cel

Celem zadania jest przedstawienie wyzwań związanych z różnymi fizycznymi modelami kostek danych oraz procesem projektowania agregacji.

#### 2. Założenia przed testem

Liczba wierszy w tabeli faktów: 1 250 000

Testowane środowisko i uruchomione aplikacje:

- Procesor: 12th Gen Intel® Core™ i7-1260P 2.10 GHz
- Zainstalowana pamięć RAM: 16,0 GB (dostępne: 15,7 GB)
- Typ systemu: 64-bitowy system operacyjny, procesor x64
- Wersja: Windows 11 Pro

Uruchomione aplikacje:

- Microsoft Visual Studio 2022
- Microsoft SQL Server Management Studio 20
- Google Chrome
- SQL Server Profiler 20
- Discord

#### 3. Założenia teoretyczne

Testowanie czasów wykonywania zapytań dla różnych modeli, z uwzględnieniem i bez zdefiniowanych agregacji.

Testowanie czasów przetwarzania kostek w tych samych warunkach testowych. Po każdej instrukcji usuwany był cache

#### **4. Wybrane zapytania**

##### **1. [dates]**

Zapytanie wyświetla średnią wartość ocen dla każdego półrocza w roku 2024.

SELECT

{[Measures].[ŚredniaWartoscOcen]} ON COLUMNS,

Descendants([Data Wystawienia].[Półrocze], [Data Wystawienia].[2024]) ON ROWS

FROM [HD Etap 4]

##### **2. [one for a particular dimension attribute]**

Zapytanie wyświetla 5 nauczycieli, którzy wystawili najwięcej ocen.

SELECT

{[Measures].[Liczba wystawionych ocen]} ON COLUMNS,

TOPCOUNT(

[Nauczyciel].[Nazwisko Imie].MEMBERS,

5,

[Measures].[Liczba wystawionych ocen]

) ON ROWS

FROM [HD Etap 4]

### 3. [general one]

Zapytanie wyświetla średnią wartość ocen dla każdego przedmiotu.

SELECT

{[Measures].[SredniaWartoscOcen]} ON COLUMNS,

{ORDER(

[Nauczyciel].[Przedmiot].MEMBERS,

[Measures].[SredniaWartoscOcen],

DESC

)} ON ROWS

FROM [HD Etap 4]

### 5. Testowanie

Tabela 4.1. Czas zapytań dla MOLAP, ROLAP i HOLAP z agregacjami i bez agregacji.

		Cube Processing			Query 1			Query 2			Query 3		
		Molap	Rolap	Holap	Molap	Rolap	Holap	Molap	Rolap	Holap	Molap	Rolap	Holap
BEZ AGREGACJI		9927	6013	8781	105	653	514	154	760	170	83	513	613
		9384	4027	8684	54	283	594	88	590	280	84	170	167
		10617	3903	8992	85	514	273	83	361	179	40	165	208
		13350	3917	5120	101	813	542	149	388	160	101	192	355
		10253	3923	5560	47	267	420	81	290	155	106	309	571
		9560	4736	6823	79	767	272	80	311	184	84	534	401
		9715	4793	5136	105	423	493	128	432	191	109	199	537
		11253	3598	6923	187	653	293	76	161	292	132	175	640
		9739	3681	6795	68	280	284	79	237	149	58	411	613
		12153	3983	5723	102	273	473	102	565	142	62	219	475

Z AGREGACJĄ	11624	4303	5163	3	922	5	120	195	3	10	170	6
	10007	3252	7572	6	281	9	1	227	4	7	250	6
	9040	2824	4966	7	398	4	2	337	4	7	181	7
	8560	2671	8457	3	882	3	3	583	4	3	335	5
	11804	3382	3239	4	815	6	2	292	5	4	287	4
	8558	3282	4225	6	583	6	7	245	3	4	216	3
	8351	3617	6828	6	281	3	3	769	4	5	161	8
	8465	4189	5430	6	272	6	3	932	4	2	738	5
	9176	3017	3852	5	280	3	2	247	4	4	331	4
	9132	4840	3972	4	272	5	3	339	3	4	197	6

Tabela 4.2. Uśredniony czas zapyta, procesowania kostki i odchylenia standardowego dla MOLAP, ROLAP i HOLAP z agregacjami i bez agregacji

	MOLAP		ROLAP		HOLAP	
	bez agregacji	z agregacją	bez agregacji	z agregacją	bez agregacji	z agregacją
<b>QUERY1 (avg)</b>	93,3	5	492,6	498,6	415,8	5
<b>QUERY2 (avg)</b>	103,64	3,1	409,5	416,6	190,2	3,8
<b>QUERY3 (avg)</b>	85,9	5	288,7	286,6	458	5,4
<b>Processing time</b>	10595,1	9471,7	4257,4	3537,7	6853,7	5370,4
<b>Standard deviation for processing time</b>	1227,147	1211,743459	696,1879	664,4594871	1433,679	1637,825217

## 6. Cache i ustawienia agregacji

- Cache

```
<ClearCache xmlns="http://schemas.microsoft.com/analysiservices/2003/engine">
  <Object>
    <DatabaseID>HD2024Cube</DatabaseID>
  </Object>
</ClearCache>
```

## - Agregacje

	A0	A1	A2	A3	A4	A5	A6
Properties							
Status	✓	✓	✓	✓	✓	✓	✓
Junk							
ID Junk (14)	✓			✓			✓
Rodzaj Oceny	✓	✓	✓	✓	✓	✓	✓
Czy Zatwierdzona	✓	✓	✓	✓	✓	✓	✓
Uczen							
ID Ucznia (377)	✓	✓	✓	✓	✓	✓	✓
Nazwisko Imie	✓	✓	✓	✓	✓	✓	✓
Czy Absolwent	✓	✓	✓	✓	✓	✓	✓
Is Current	✓	✓	✓	✓	✓	✓	✓
Pesel	✓	✓	✓	✓	✓	✓	✓
Data Wystawienia							
ID Data (1462)	✓	✓	✓	✓	✓	✓	✓
Data	✓	✓	✓	✓	✓	✓	✓
Rok (5)	✓	✓	✓	✓	✓	✓	✓
Miesiac	✓	✓	✓	✓	✓	✓	✓
Dzien	✓	✓	✓	✓	✓	✓	✓
Polrocz	✓	✓	✓	✓	✓	✓	✓
Dzien Pracujacy	✓	✓	✓	✓	✓	✓	✓
Dzien Tygodnia	✓	✓	✓	✓	✓	✓	✓
Sala							
ID Sali (112)	✓	✓	✓	✓	✓	✓	✓
Numer Sali	✓	✓	✓	✓	✓	✓	✓
Rodzaj Sali	✓	✓	✓	✓	✓	✓	✓
Nauczyciel							
ID Nauczyciela (50)	✓	✓	✓	✓	✓	✓	✓
Nazwisko Imie	✓	✓	✓	✓	✓	✓	✓
Stopien Naukowy	✓	✓	✓	✓	✓	✓	✓
Email	✓	✓	✓	✓	✓	✓	✓
Przedmiot	✓	✓	✓	✓	✓	✓	✓

## 8. Wydajność obsługi zapytań

Najlepszą wydajność w przetwarzaniu zapytań osiągnął MOLAP, co wynika z faktu, że nie korzysta on z relacyjnej bazy danych. Dane są przechowywane w wielowymiarowej strukturze analitycznej, co minimalizuje liczbę operacji dostępowych. Najgorsze czasy zapytań zaobserwowano w ROLAP oraz HOLAP (bez agregacji), ponieważ dane są pobierane z relacyjnej bazy danych, co generuje dodatkowe koszty dostępu. W przypadku ROLAP, nawet z agregacjami, poprawa była minimalna, ponieważ są one przechowywane jako widoki w relacyjnej bazie, co nadal wymaga ich obliczenia. HOLAP z agregacjami osiągnął wydajność zbliżoną do MOLAP, ponieważ również przechowuje zagregowane dane w strukturach wielowymiarowych.

## 9. Czas przetwarzania kostki

ROLAP przetwarzał kostkę najszybciej, ponieważ dane pozostają w relacyjnej bazie i nie wymagają dodatkowej obróbki. HOLAP zajmował drugie miejsce, a MOLAP był zdecydowanie najwolniejszy, gdyż procesuje dane i zapisuje je w wielowymiarowej strukturze. Kostki z włączonymi agregacjami wymagały więcej czasu na przetwarzanie, co szczególnie było widoczne w przypadku HOLAP, który buduje zagregowane dane w wielowymiarowej bazie.

## 10. Wnioski

1. **MOLAP** zapewnia najlepszą wydajność zapytań, ale wymaga najwięcej czasu na przetwarzanie danych.
2. **ROLAP** charakteryzuje się najniższą wydajnością zapytań, ale przetwarza kostkę najszybciej.
3. **HOLAP** łączy cechy MOLAP i ROLAP – oferuje wydajność zapytań zbliżoną do MOLAP przy włączonych agregacjach, jednak kosztem dłuższego przetwarzania danych.

Wybór właściwego modelu powinien zależeć od priorytetów użytkownika – czy większe znaczenie ma szybkość zapytań (MOLAP), krótki czas przetwarzania kostki (ROLAP), czy może kompromis pomiędzy tymi cechami (HOLAP).