

Optymalizacja hurtowni

Proces zakupu i wykorzystania karnetów przez klientów

Krzysztof Nasuta 193328, Filip Dawidowski 193433

1. Wstęp

Celem raportu jest pokazanie wpływu różnych fizycznych modeli kostki oraz sposobu agregacji na wydajność.

2. Założenia wstępne

2.1. Wielkość hurtowni danych

- Rozmiar bazy danych: 136MB (+264MB log)
- Ilość wierszy w tabeli faktu zjazdu: 1 599 884

2.2. Środowisko testowe

Maszyna wirtualna QEMU z systemem Windows 11, 16GB RAM, 8 vCPU (4 rdzenie, 8 wątków), procesor AMD Ryzen 5 3600, dysk SSD.

3. Testowanie

Sprawdzanie czasu wykonania zapytań i czasu procesowania kostki dla różnych modeli fizycznych kostki, z oraz bez agregacji.

3.1. Opis zapytań

3.1.1. Agregacja dat:

```
SELECT
    {
        [Ride Date].[DateHierarchy].[Year]
    } ON ROWS,
    {
        [Measures].[AverageRideCountPerCard], [Measures].[Ride Count]
    } ON COLUMNS
FROM
    [Ski Center Data Warehouse]
```

3.1.2. Wymiar:

```
SELECT
    {
        [Pass].[Price].[Price].MEMBERS
    } ON ROWS,
    {
        [Measures].[AverageRideCountPerCard], [Measures].[Ride Count]
    } ON COLUMNS
FROM
    [Ski Center Data Warehouse]
```

3.1.3. Ogólne:

```

WITH MEMBER [Measures].[Average Rides Per Month] AS
    AVG(
        [Ride Date].[Year].[Year].MEMBERS,
        [Measures].[Ride Count]
    )
SELECT
    NON EMPTY {
        [Measures].[Average Rides Per Month]
    } * {
        [Ride Date].[Month].[Month].ALLMEMBERS
    } ON COLUMNS,
    NON EMPTY {
        [Pass].[Price].[Price].ALLMEMBERS
    } ON ROWS
FROM
    [Ski Center Data Warehouse]

```

3.2. Wyniki

Czas podany w milisekundach, średnia z 10 pomiarów

	MOLAP		HOLAP		ROLAP	
	Agregacja	Bez agr.	Agregacja	Bez agr.	Agregacja	Bez agr.
Czas zapytania (3 zapytania)	106.5	192.67	101.375	362.4	353.71	350.6
	123.375	198.33	170.33	169.33	164.5	159.75
	23.75	31	89.67	100.75	101.56	89.33
Czas procesowania	6652.25	6534,25	2715.33	2303.25	2450.83	2346.67
Łączny rozmiar	16,75 MB	16,46 MB	15,06 MB	14,77 MB	14,76 MB	14,77 MB

4. Wnioski

Dla testowanej hurtowni danych najlepsze wyniki czasowe dla zapytań osiągnięto generalnie dla modelu MOLAP. Model ROLAP okazał się najwolniejszy, co jest zgodne z oczekiwaniami, ponieważ dane pobierane są z relacyjnej bazy danych. Zastosowanie agregacji przyspieszyło czas zapytań - szczególnie w przypadku modelu MOLAP, dla ROLAP agregacje zwiększyły czas zapytań. Czas procesowania kostki jest wysoki dla MOLAP i niższy dla HOLAP i ROLAP, co jest zgodne z oczekiwaniami. Łączny rozmiar kostki jest największy dla MOLAP i najmniejszy dla ROLAP, co potwierdza teorię modeli fizycznych kostki. Agregacje nie wpłynęły znacząco na rozmiar kostki.