

## **Optymalizacja Hurtowni Danych – Raport**

### **1. Cel laboratorium**

Cel zadania polega na przedstawieniu problemów związanych z różnymi modelami kostek danych oraz projektowaniem agregacji.

### **2. Wstępne założenia**

#### **Rozmiar bazy danych (Hurtownia danych):**

Hurtownia zawiera:

- 500 krotek z wymiaru Stacja
- 1000 krotek z wymiaru Trasa
- 50000 – 200000 krotek z faktu Kurs
- 500000 krotek z faktu Przewoz\_Pasazera
- Wymiar Czas zawiera jeden rekord na każdą minutę w dobie (1440 krotek)
- Wymiar Data zawiera jeden rekord na każdy dzień 2023 roku (365 krotek)
- Wymiar Junk zawiera wszystkie możliwe kombinacje atrybutów: Sposób zakupu, Klasa i Ulga (72 krotki)

#### **Środowisko do testowania:**

- Windows 10 Pro
- AMD Ryzen 5 3600 6-Core
- RAM 16GB
- Microsoft Visual Studio 2019
- Microsoft SQL Server Management Studio 2018
- Microsoft SQL Server Profiler 2018

### 3. Testowanie

Testowanie czasów wykonywania zapytań dla różnych modeli, zarówno zdefiniowanych, jak i bez zdefiniowanych agregacji. Testowanie czasów przetwarzania kostki danych w tych samych ustawieniach testowych.

Zapytania testowe:

1. Porównaj liczbę sprzedanych biletów w poszczególne dni tygodnia w danym miesiącu.
2. Porównaj przychód ze sprzedaży biletów na poszczególnych trasach w tym miesiącu w porównaniu z poprzednim.
3. Porównaj liczbę sprzedanych biletów ulgowych w tym miesiącu w porównaniu z poprzednim.

	MOLAP		ROLAP		HOLAP	
	Aggr.	No aggr.	Aggr.	No aggr.	Aggr.	No aggr.
Querying speed (for 3 different queries) [ms]	2,6	6,3	932,4	944,1	2,7	933,1
	27	26,9	207,7	219,8	210,8	205,5
	4,5	14,3	615,5	610,8	4,9	607,3
Processing time [ms]	4473	4301	1958	2102	2011	1947
Total size [MB]	59,03	59,22	44,61	44,61	43,41	44,61

### 4. Wnioski

Model MOLAP zarówno w teorii jak i w praktyce wypadł najlepiej, jeśli chodzi o czas wykonania zapytań. Zdefiniowanie agregacji dodatkowo przyspieszyło czas wykonywania zapytań. Ze względu na to, że model MOLAP przechowuje wszystkie dane dotyczące miar i agregacji w bazie analitycznej, czas przetwarzania kostki oraz jej rozmiar są największe ze wszystkich modeli, co potwierdziły nasze testy.

Model ROLAP z kolei nie przechowuje danych miar i agregacji w bazie analitycznej, dlatego jego czas przetwarzania oraz rozmiar są najmniejsze. Z tego powodu przy przetwarzaniu zapytań musi on pobierać dane z relacyjnej bazy

danych hurtowni co negatywnie wpływa na czas ich wykonywania. Z tego samego powodu dodatkowo definiowane agregacje nie mają wpływu na przetwarzanie zapytań.

Model HOLAP jest kompromisem pomiędzy modelami MOLAP i ROLAP. W bazie analitycznej przechowuje on tylko informacje o agregacjach. Oznacza to, że w przypadku braku zdefiniowanych agregacji, czas wykonania zapytań jest porównywalny z modelem ROLAP, natomiast zastosowanie agregacji zbliża go czasowo do modelu MOLAP. Czas przetwarzania i rozmiar kostki są zbliżone do modelu ROLAP.

W przypadku zapytania numer 2, dodatkowe agregacje nie wpłynęły pozytywnie na czas wykonywania zapytań. Może być to spowodowane dużą ilością agregowanych atrybutów (Trasa.Nazwa). W przypadku agregowanych atrybutów z pierwszego oraz trzeciego zapytania (Data.Dzien\_Tygodnia i Junk.Ulga) liczba była znacznie mniejsza.