Maxym Besarab 164042

Karyna Lukashova 187714

Optymalizacja hurtowni danych

1. Cel laboratorium

Celem zadania jest przedstawienie zagadnień związanych z różnymi modelami fizycznymi kostki i projektowaniem agregatów.

2. Założenia wstępne

Rozmiar bazy danych:

Tabele faktów	**Tabele wymiarów**
Wykonania przelotów – 145 000 krotek	Samoloty – 1060 krotek
	Serwisy/Warsztaty naprawcze – 40 krotek
	Daty – każda data od 01/01/1970 do 31/12/2023 ma oddzielny wiersz. Razem 18993
Wystąpienia awarii – 12 000 wierszy	Czas – każda godzina od 00:00:00 do 23:59:59 ma oddzielny wiersz. Razem 86400
	Trasy – 50 krotek
	LotJunk – 20 wierszy
	AwariaJunk – 32 wiersze

Środowisko testowe:

CPU: AMD Ryzen 5 4600H 3,00 GHz	OS: Windows 10 Home (64-bit)
RAM: 16GB (DDR4)	Microsoft Visual Studio 2019 Microsoft SQL Server Management Studio 2018 Microsoft SQL Server Profiler 2018
GPU: NVIDIA GeForce 1650 Ti 4GB SSD: 512GB (Intel 660p Series)	

3. Testowanie

Testowanie polegało na pomiarze czasów wykonania zapytań dla różnych modeli, ze zdefiniowanymi agregacjami lub bez tych agregacji. Dodatkowo, mierzony był czas procesowania kostki dla tych samych ustawień.

Krótki opis zapytań:

#1 - zapytanie z agregacją na datach

Sklasyfikuj miesiące roku i liczbę awarii, która w nich wystąpiła.

#2 - zapytanie dla konkretnego atrybutu wymiaru

Ile awarii miał samolot o konkretnym numerze seryjnym w ostatnim roku?

#3 - zapytanie ogólne

Jaki był czas opóźnień (w godzinach) w tym i w poprzednim miesiącu?

	MOLAP		ROLAP		HOLAP	
	Aggregacje	**Brak aggr.**	**Aggregacje**	**Brak aggr.**	**Aggregacje**	**Brak aggr.**
Czas zapytania #1	24 ms	37 ms	57 ms	85 ms	49 ms	69 ms
Czas zapytania #2	17 ms	20 ms	49 ms	50 ms	16 ms	41 ms
Czas zapytania #3	899 ms	1154 ms	1223 ms	1334 ms	988 ms	1217 ms
Czas przetwarzania kostki	976 ms	897 ms	357 ms	350 ms	386 ms	326 ms
Rozmiar bazy danych	26,56 MB	20,01 MB	20,65 MB	19,62 MB	21,30 MB	19,62 MB

4. Dyskusja (porównanie teorii z uzyskanymi wynikami)

Model MOLAP zgodnie z założeniami teoretycznymi osiągał najszybszy czas wykonywania zapytań. Jego wadą jest jednak dużo dłuższy czas przeprocesowania kostki oraz większy narzut pamięciowy (z powodu przechowywania duplikatów danych). Większe zużycie pamięci jest szczególnie widocznie, gdy zdefiniujemy dla niego agregaty.

Model ROLAP w teorii miał być dużo wolniejszy od MOLAP i tak jest w rzeczywistości. Wynika to z faktu, że na serwerze OLAP przechowywane są tylko dane dotyczące kostki, a reszta informacji w relacyjnej bazie danych. Jedyną zaletą tego rozwiązania jest szybszy czas procesowania kostki.

Model HOLAP, rozwiązanie pośrednie pomiędzy MOLAP i ROLAP, w przypadku zdefiniowania agregatów uzyskał zbliżone czasy wykonania dla dwóch z trzech zapytań. Dodając do tego mniejsze zużycie pamięci i szybszy czas przetwarzania kostki, może także być dobrym wyborem, o ile dobrze zdefiniujemy agregacje. W przypadku braku agregatów – jest dużo wolniejszy od MOLAP, zgodnie z teorią.

Zdefiniowanie odpowiednich agregatów pozwoliło przyspieszyć znacznie czasy wykonania zapytań dla modelów MOLAP i HOLAP. Różnica jest mniej widoczna dla modelu ROLAP. Stosowanie agregacji powoduje jednak wzrost zużycia pamięci.