

Optymalizacja Hurtowni Danych - raport

1. Cel laboratorium

Celem zadania jest pokazanie zagadnień dotyczących różnych fizycznych modeli kostek oraz projektów agregacji.

2. Założenia wstępne

Wielkość hurtowni danych: 1936,00 MB

Środowisko testowe:

Programy:

Visual Studio Enterprise 2022

SQL Server Management Studio 2019

MSSQL Server 2022

Sprzęt:

Procesor: 11th Gen Intel(R) Core(TM) i7-1165G7

RAM: 16 GB

3. Testy

a) (one with aggregations on dates)

```
--2 Porównaj różne rodzaje interwencji w bieżącym miesiącu w
stosunku do poprzednich miesięcy.
WITH
SET [Kradzieże] AS
    FILTER(
        [Szczegóły Interwencji].[Rodzaj].Members,
        [Szczegóły Interwencji].[Rodzaj].CurrentMember IS [Szczegóły
Interwencji].[Rodzaj].&[kradzież]
    )

SET [Zniszczenia mienia] AS
    FILTER(
        [Szczegóły Interwencji].[Rodzaj].Members,
        [Szczegóły Interwencji].[Rodzaj].CurrentMember IS [Szczegóły
Interwencji].[Rodzaj].&[zniszczenie mienia]
    )
```

```

SET [Zastraszenia] AS
    FILTER(
        [Szczegoly Interwencji].[Rodzaj].Members,
        [Szczegoly Interwencji].[Rodzaj].CurrentMember IS [Szczegoly
Interwencji].[Rodzaj].&[zastraszanie]
    )

SET [Rozboje] AS
    FILTER(
        [Szczegoly Interwencji].[Rodzaj].Members,
        [Szczegoly Interwencji].[Rodzaj].CurrentMember IS [Szczegoly
Interwencji].[Rodzaj].&[rozboj]
    )

SET [Sytuacje przeciwko zyciu czlowieka] AS
    FILTER(
        [Szczegoly Interwencji].[Rodzaj].Members,
        [Szczegoly Interwencji].[Rodzaj].CurrentMember IS [Szczegoly
Interwencji].[Rodzaj].&[przeciwko zyciu czlowieka]
    )

SET [Spowodowania Uszczerbku na Zdrowiu] AS
    FILTER(
        [Szczegoly Interwencji].[Rodzaj].Members,
        [Szczegoly Interwencji].[Rodzaj].CurrentMember IS [Szczegoly
Interwencji].[Rodzaj].&[spowodowanie uszczerbku na zdrowiu]
    )

SET [Oszustwa] AS
    FILTER(
        [Szczegoly Interwencji].[Rodzaj].Members,
        [Szczegoly Interwencji].[Rodzaj].CurrentMember IS [Szczegoly
Interwencji].[Rodzaj].&[oszustwo]
    )

SET [Inne] AS
    FILTER(
        [Szczegoly Interwencji].[Rodzaj].Members,
        [Szczegoly Interwencji].[Rodzaj].CurrentMember IS [Szczegoly
Interwencji].[Rodzaj].&[inne]
    )

MEMBER [Measures].[Liczba kradziezy] AS

```

```

SUM(
    [Kradzieze],
    [Measures].[Liczba Faktow Interwencji]
)

MEMBER [Measures].[Liczba zniszczen mienia] AS
SUM(
    [Zniszczenia mienia],
    [Measures].[Liczba Faktow Interwencji]
)

MEMBER [Measures].[Liczba zastraszen] AS
SUM(
    [Zastraszenia],
    [Measures].[Liczba Faktow Interwencji]
)

MEMBER [Measures].[Liczba rozbojow] AS
SUM(
    [Rozboje],
    [Measures].[Liczba Faktow Interwencji]
)

MEMBER [Measures].[Liczba Przeciwno zyciu czlowieka] AS
SUM(
    [Sytuacje przeciwno zyciu czlowieka],
    [Measures].[Liczba Faktow Interwencji]
)

MEMBER [Measures].[Spowodowania Uszczerbku na Zdrowiu] AS
SUM(
    [Spowodowania Uszczerbku na Zdrowiu],
    [Measures].[Liczba Faktow Interwencji]
)

MEMBER [Measures].[Oszustwa] AS
SUM(
    [Oszustwa],
    [Measures].[Liczba Faktow Interwencji]
)

MEMBER [Measures].[Inne] AS
SUM(

```

```

        [Inne],
        [Measures].[Liczba Faktow Interwencji]
    )

SELECT {[Measures].[Liczba kradziezy], [Measures].[Liczba
zniszczen mienia], [Measures].[Liczba zastraszen],
[Measures].[Liczba rozbojow], [Measures].[Liczba Przeciwno zyciu
czlowieka], [Measures].[Spowodowania Uszczerbku na Zdrowiu],
[Measures].[Oszustwa], [Measures].[Inne]} ON COLUMNS,
{Descendants([Data].[HierarchiaDaty].[Rok].Members,
[Data].[HierarchiaDaty].[Miesiac])} ON ROWS
FROM [Police DW]

```

b) (one for a particular dimension attribute)

```

--3 Czy częstotliwość konserwacji broni i sprzętu koreluje z
powodzeniem akcji?
SELECT
    {[Measures].[Ogolna Skuteczność Interwencji]} ON COLUMNS,
    {[Wyposazenie].[Czestotliwosc Konserwacji].Members} ON ROWS
FROM [Police DW];

```

c) (ogólny)

```

--1 Czy różne rodzaje broni i sprzętu są bardziej skuteczne w
zależności od rodzaju interwencji?
WITH
SET [Bron] AS
    FILTER(
        [Wyposazenie].[Rodzaj].Members,
        [Wyposazenie].[Rodzaj].CurrentMember IS
[Wyposazenie].[Rodzaj].&[bron]
    )

SET [Pojazdy] AS
    FILTER(
        [Wyposazenie].[Rodzaj].Members,
        [Wyposazenie].[Rodzaj].CurrentMember IS
[Wyposazenie].[Rodzaj].&[pojazd]
    )

SET [Srodki Laczności] AS
    FILTER(

```

```

        [Wyposazenie].[Rodzaj].Members,
        [Wyposazenie].[Rodzaj].CurrentMember IS
[Wyposazenie].[Rodzaj].[srodki_lacznosci]
    )

SET [Inne] AS
    FILTER(
        [Wyposazenie].[Rodzaj].Members,
        [Wyposazenie].[Rodzaj].CurrentMember IS
[Wyposazenie].[Rodzaj].[inne]
    )

MEMBER [Measures].[Bron] AS
    SUM(
        [Bron],
        [Measures].[Ogolna Skuteczność Interwencji]
    )

MEMBER [Measures].[Pojazdy] AS
    SUM(
        [Pojazdy],
        [Measures].[Ogolna Skuteczność Interwencji]
    )

MEMBER [Measures].[Srodki Lacznosci] AS
    SUM(
        [Srodki Lacznosci],
        [Measures].[Ogolna Skuteczność Interwencji]
    )

MEMBER [Measures].[Inne] AS
    SUM(
        [Inne],
        [Measures].[Ogolna Skuteczność Interwencji]
    )

SELECT {[Measures].[Bron], [Measures].[Pojazdy],
[Measures].[Srodki Lacznosci], [Measures].[Inne]} ON COLUMNS
FROM [Police DW]

```

	MOLAP		ROLAP		HOLAP	
	Agr.	Bez	Agr.	Bez	Agr.	Bez
Szybkość odczytu	30	96	423	622	29	475
	270	275	1 595	1881	1 514	1 574
	262	336	3 506	3 498	3 475	3 467
Czas przetwarzania	7234	15138	6845	10687	13569	7419
Całkowita wielkość	116,20 MB	115,72 MB	108,45 MB	108,46 MB	109,19 MB	108,45 MB

4. Wnioski

	MOLAP	HOLAP	ROLAP
Querying time	Short	Moderate (short with well designed aggregations)	Long
Processing time	Long	Moderate (if no aggregations are designed, it will be short)	Short
Total size	Big (size of the measure group is much smaller if no aggregations are designed for them)	Moderate	Small

Według powyższej tabeli czasy wykonywania zapytań dla modelu MOLAP powinny być krótkie. Zgadza się to z naszymi wynikami. Dodatkowo czas procesowania kostki i rozmiar są znacznie większe niż przy innych metodach, co również zgadza się z założeniami teoretycznymi. Te dwie ostatnie cechy niewiele się różnią od siebie w zależności od tego, czy był to MOLAP z agregacją czy bez. Wadą tego modelu jest to, że zawiera duplikaty w bazie analitycznej przez co zarówno hurtownia jak i baza analityczna zawierają duże ilości tych samych danych - to jest możliwy powód takich dużych czasów.

W ROLAP szybkość wykonywania zapytań jest około 7 razy dłuższa niż dla MOLAP. Natomiast zmniejszył się rozmiar całkowity kostki oraz czas jej przetwarzania. Największą zaletą tej metody jest to, że znacznie zmniejszył się całkowity czas przetwarzania - jest on mniejszy ze wszystkich uzyskanych innymi metodami czasów.

Model HOLAP według wzorców teoretycznych powinien być rozwiązaniem pośrednim między wyżej opisanymi metodami. Model zawiera w bazie analitycznej wyłącznie agregacje, wobec tego dla dobrze zaprojektowanych agregacji znacznie zmniejszają one czas wykonywania zapytania. Widać to dobrze na przykładzie pierwszego zapytania, który ma o około 16 razy lepszy czas dla zapytania z agregacją niż bez. Ale ogólnie czasy uzyskane tą metodą były zbliżone do tych uzyskanych przez ROLAP. Dodatkowo czas przetwarzania dla tego modelu z agregacjami był nieporównywalnie większy od pozostałych.