1. **Krótki opis problemu:**

W pierwszej części wykonywanego projektu sieci gabinetów lekarskich należało wykonać bazę danych którą należało zasilić danymi przez co mogliśmy zacząć operację na tych danych.

Po przeanalizowaniu bazy danych stwierdziliśmy że najważniejszą tabelą w naszym projekcie jest tabela „Wizyty”. Jest to kluczowy aspekt dla każdej z sieci gabinetów lekarskich. Dlatego więc użyliśmy jej do stworzenia podstawowej struktury hurtowni danych.

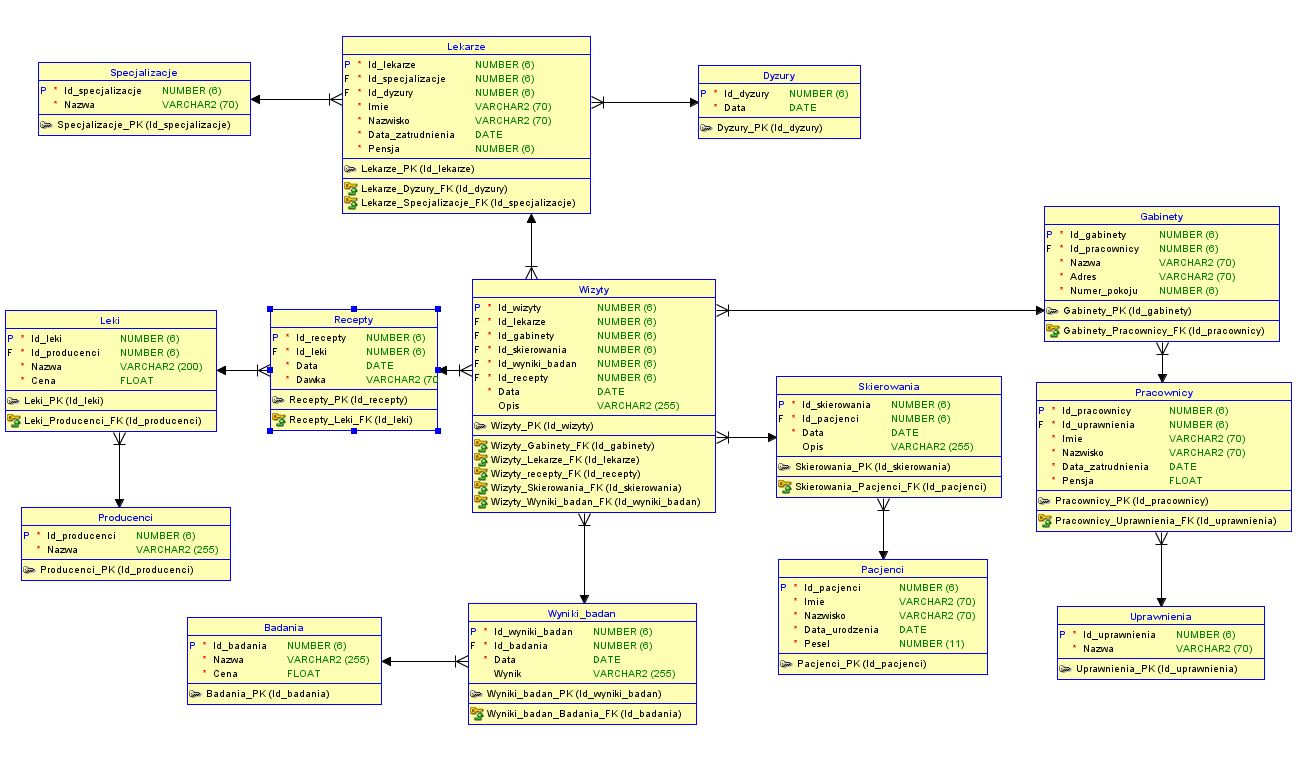
Drugą częścią naszego projektu jest stworzenie zapytań do bazy danych oraz hurtowni danych które będą dawały taki sam efekt. Skonstruowanie prawidłowej hurtowni danych poprzez wydanie napisanych przez nas zapytań jest w stanie zaoszczędzić znaczne ilości czasu przy dużych ilościach danych w hurtowni.

Ostatnią częścią projektu jest przeniesienie wyników do sprawozdania oraz opracowanie wniosku.

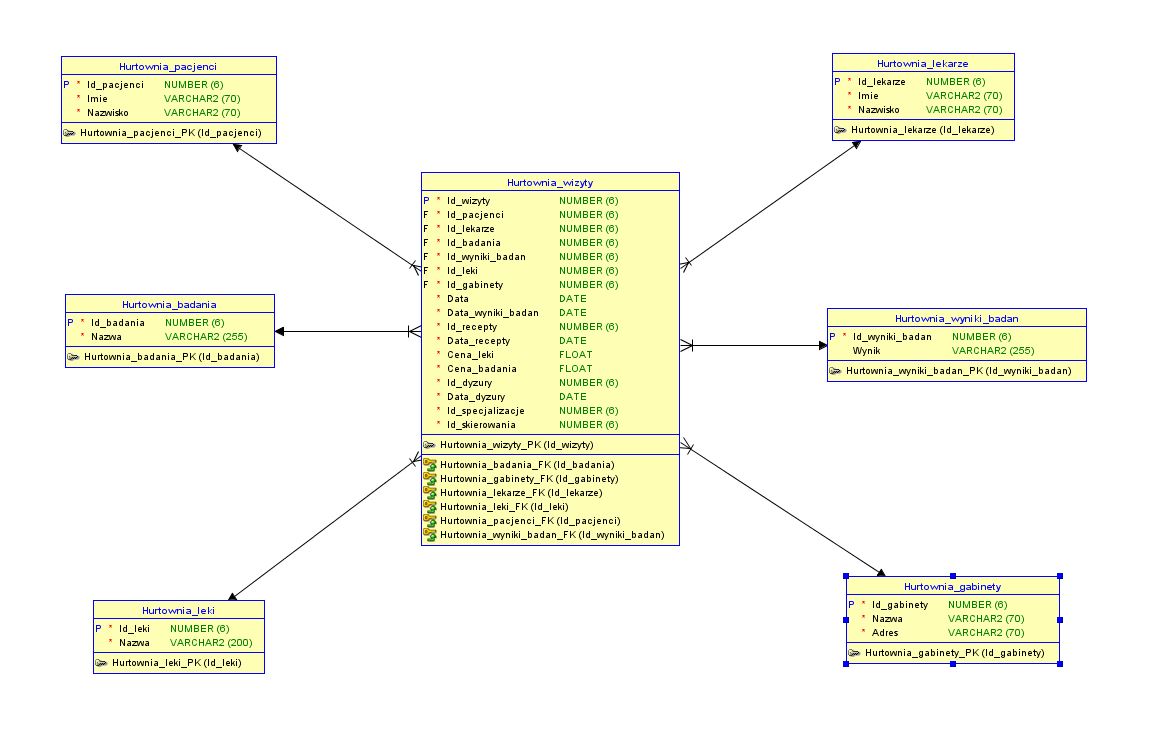
1. **Schemat ERD bazy i hurtowni:**

Schematy zostały stworzone przy użyciu programu Oracle DataModeler.

**BAZA DANYCH**

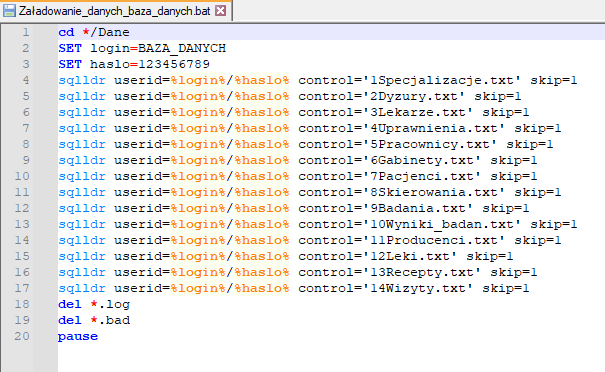


**HURTOWNIA DANYCH**

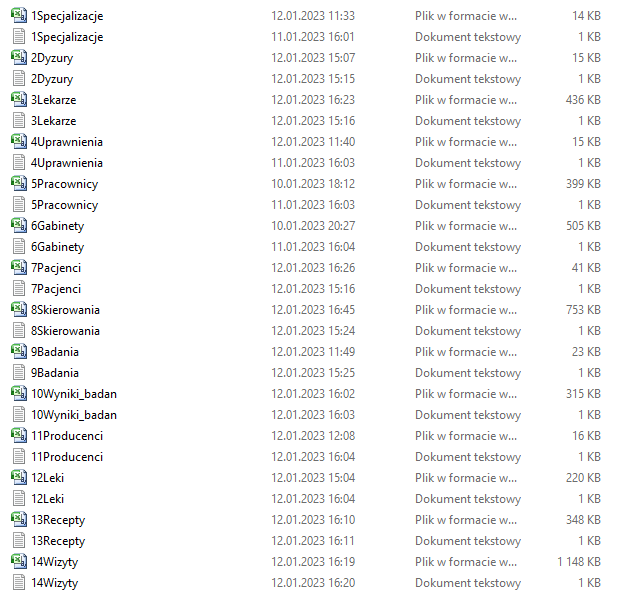


1. **Opis procesu ładowania i transformacji danych z bazy do hurtowni:**

Dane które znajdują się w plikach CSV w folderze „Dane” z baza danych zostały wygenerowane przy pomocy strony internetowej <https://www.mockaroo.com/>. Dzięki wyborze plików CSV jesteśmy w stanie łatwo załadować te pliki dzięki SQL LOADER. Również w folderze „Dane” znajdują się pliki kontrolne .txt które załadują nam konkretne dane do odpowiednich kolumn. Aby załadować dane do bazy danych musimy zmienić w pliku „Załadowanie\_danych\_baza\_danych.bat” dwa parametry „login” oraz „haslo”.



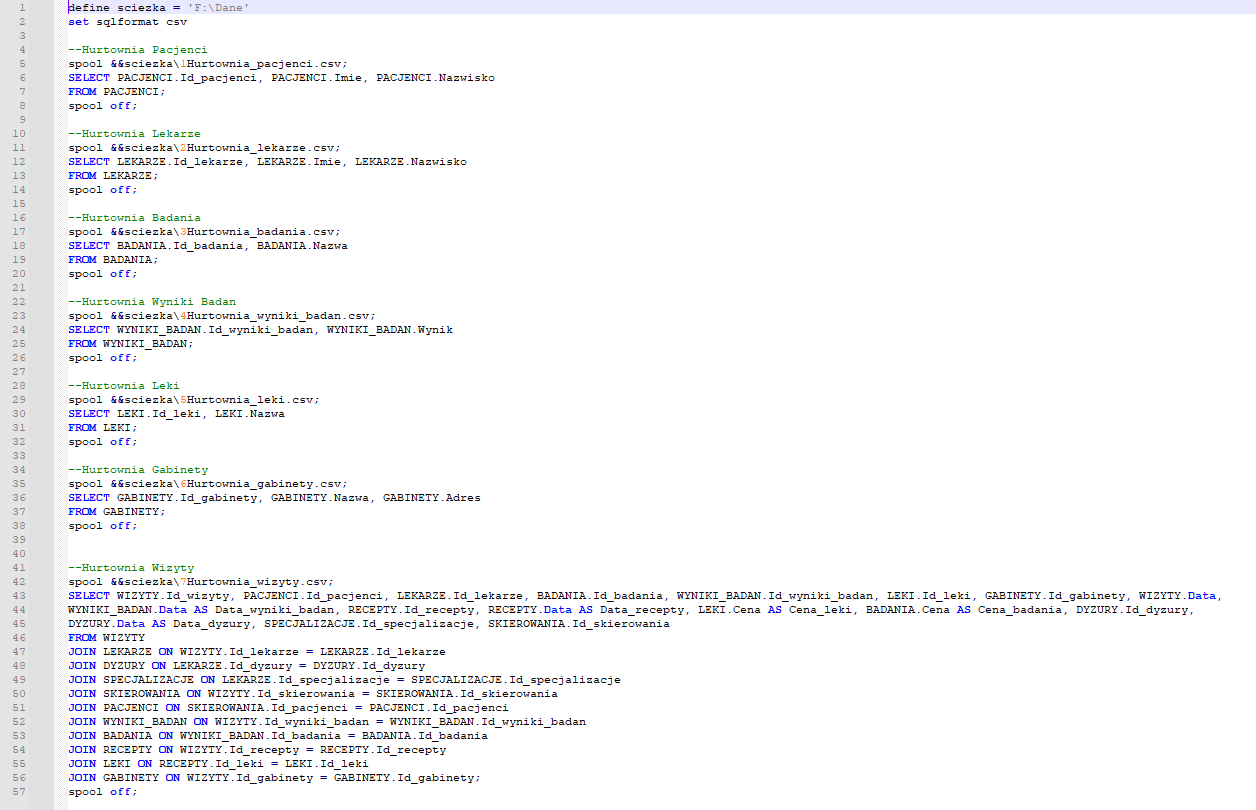
**PLIK „Załadowanie\_danych\_baza\_danych.bat”**



**Pliki CSV oraz Pliki kontrolne Bazy danych**

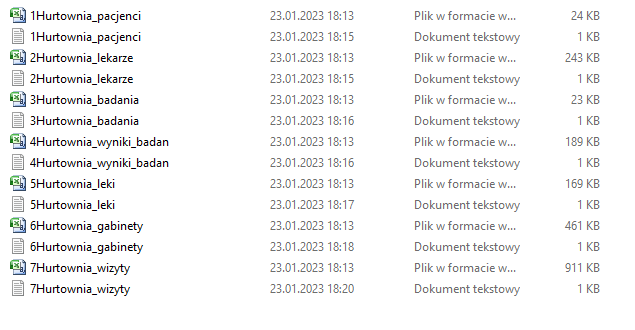
Po utworzeniu bazy danych i wczytaniu ich należy wygenerować dane które przeniesiemy do Hurtowni danych. Za tą czynność odpowiada plik „Gnerowanie\_danych\_z\_bazy\_danych\_do\_hurtowni.sql” który należy wykonać w danej bazie danych gdzie są już są wczytane dane z wcześniejszych plików CSV.

Aby generowanie danych zostało wykonane musimy zmienić jeden parametr w tym pliku którym jest „sciezka” ,aby dane zostały wygenerowane w naszym projekcie Dane w katalogu Hurowania.



**Plik „Generowanie\_danych\_z\_bazy\_danych\_do\_hurtowni.sql”**

Po wykonaniu poprzedniego kroku należy utworzyć Hurtownie danych oraz analogicznie do Bazy danych załadować pliki CSV.

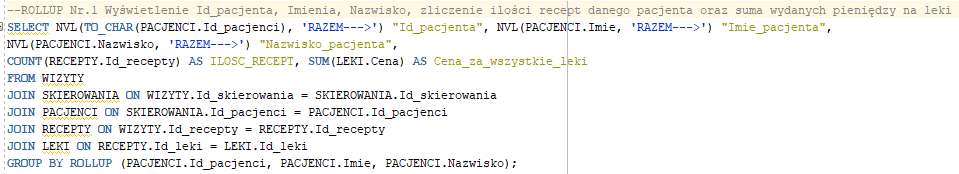


**Pliki CSV oraz Pliki kontrolne Hurtowni danych**

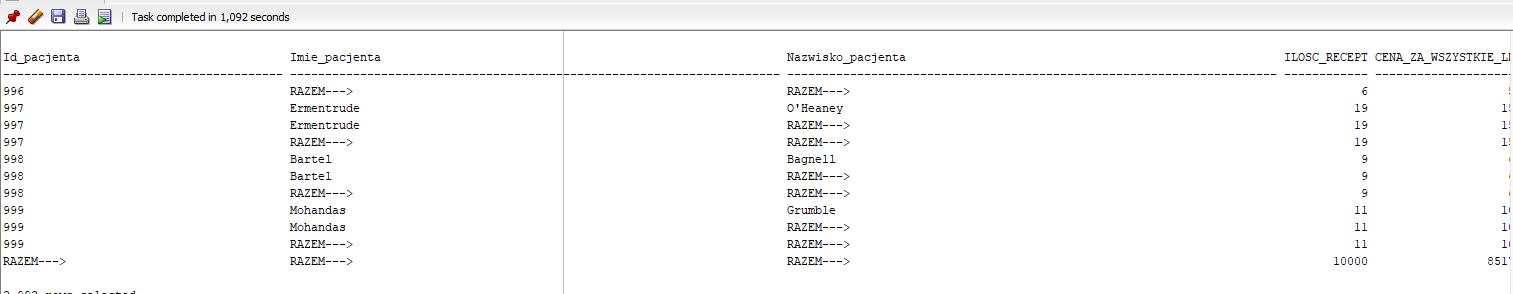
1. **Zapytania:**

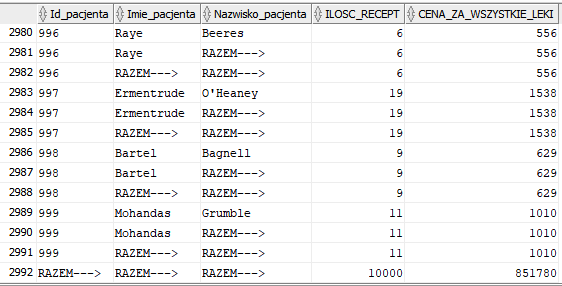
* **01 – ROLLUP:**

**Zapytanie nr.1:**

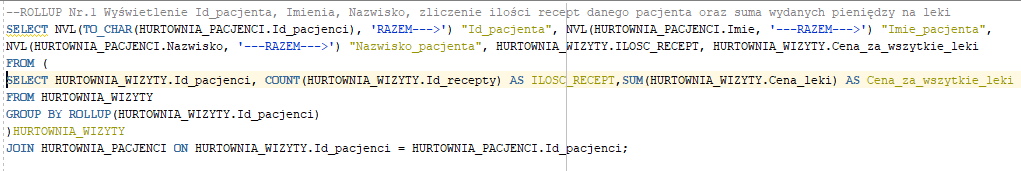
****

**Zapytanie nr.1 baza danych**

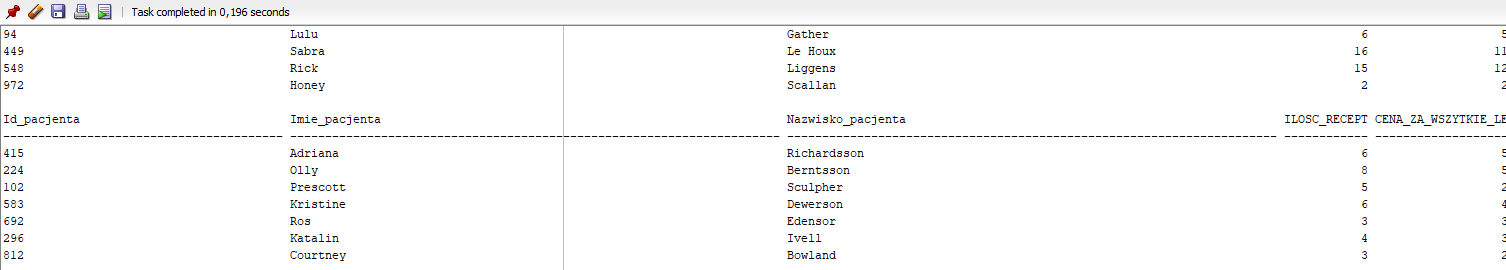
****

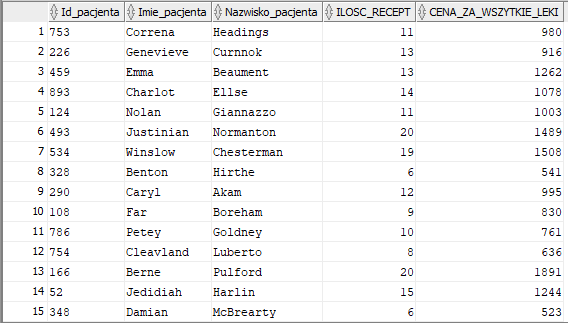
****

**Odpowiedź nr.1 baza danych**

****

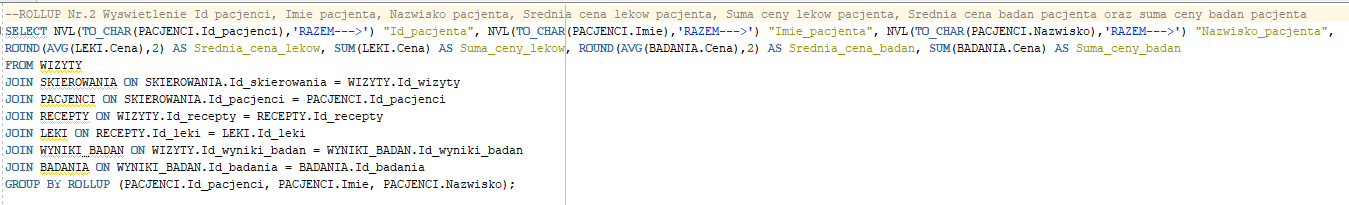
**Zapytanie nr.1 hurtownia danych**

****

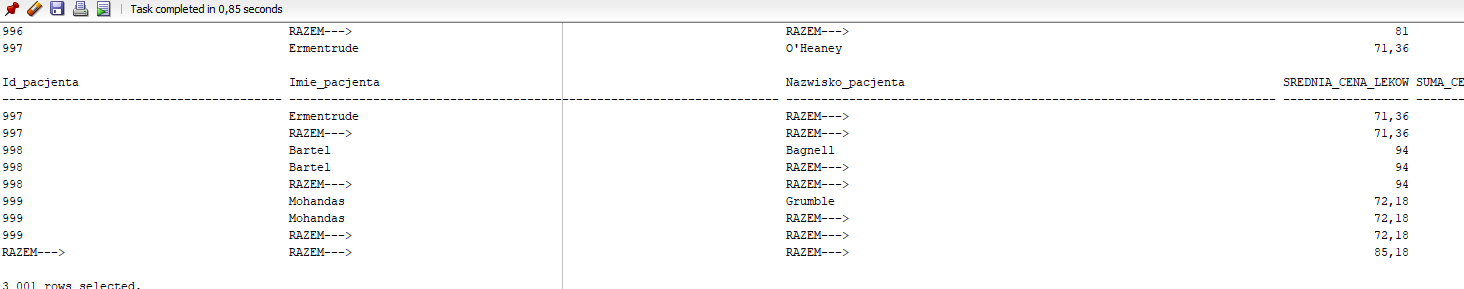
****

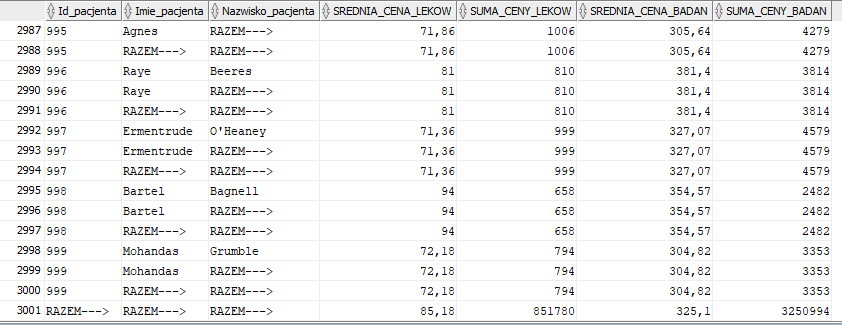
**Odpowiedź nr.1 hurtownia danych**

**Zapytanie nr.2:**

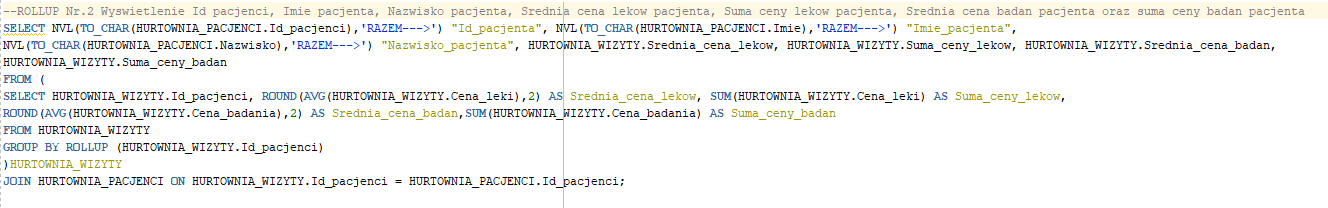
****

**Zapytanie nr.2 baza danych**

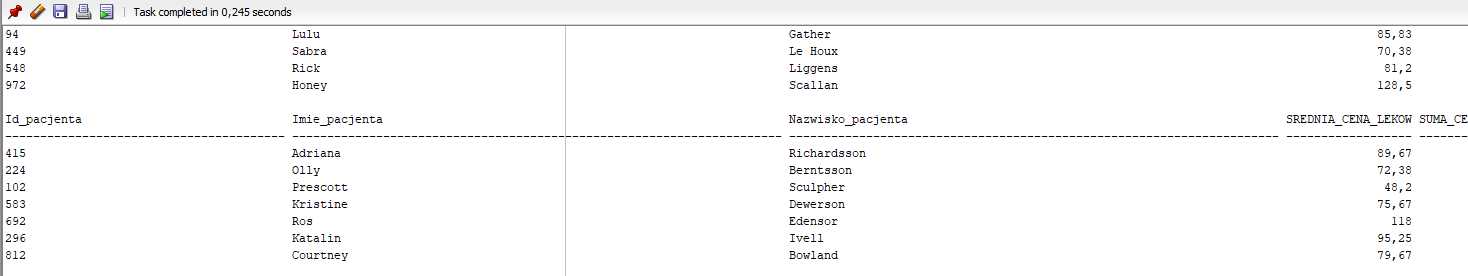
****

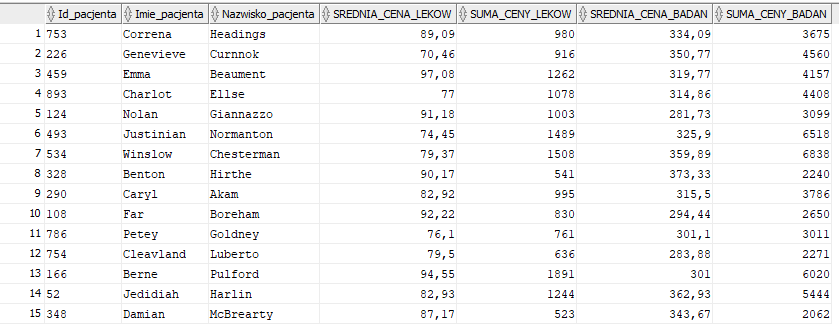
****

**Odpowiedź nr.2 baza danych**

****

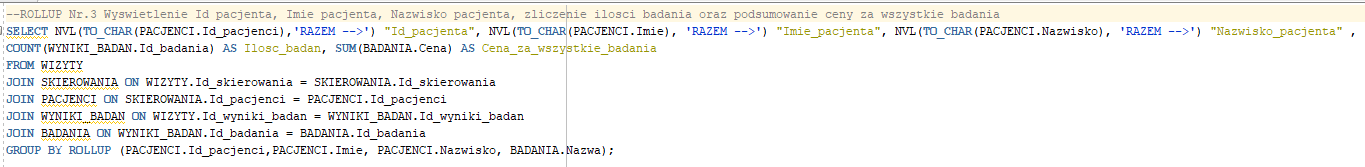
**Zapytanie nr.2 hurtownia danych**

****

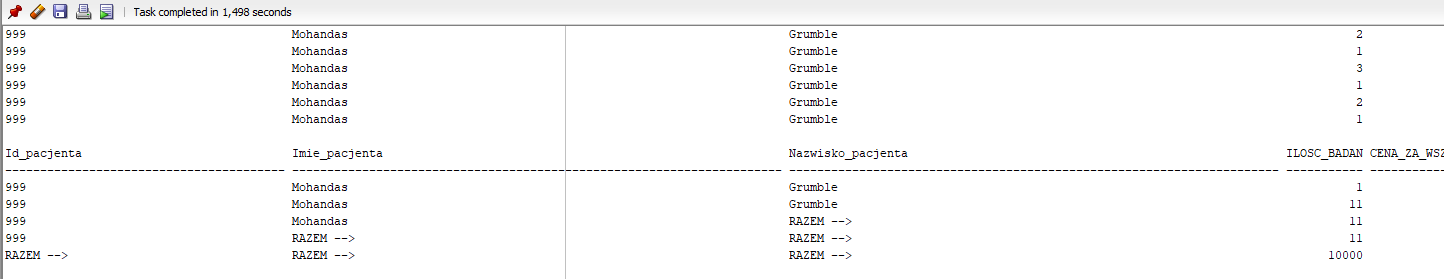
****

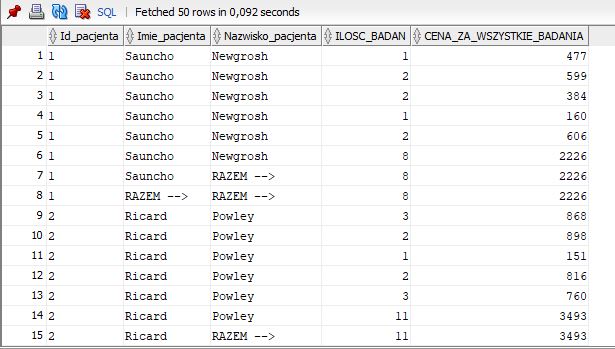
**Odpowiedź nr.2 hurtownia danych**

**Zapytanie nr.3:**

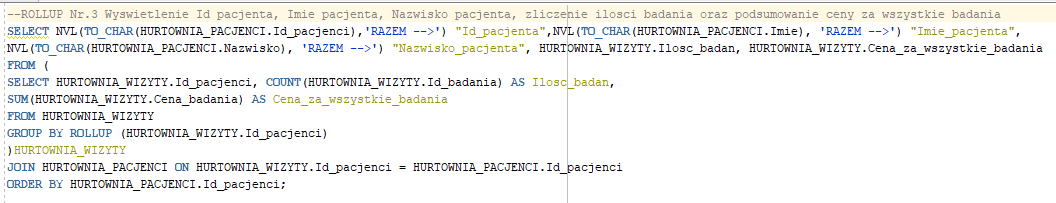
****

**Zapytanie nr.3 baza danych**

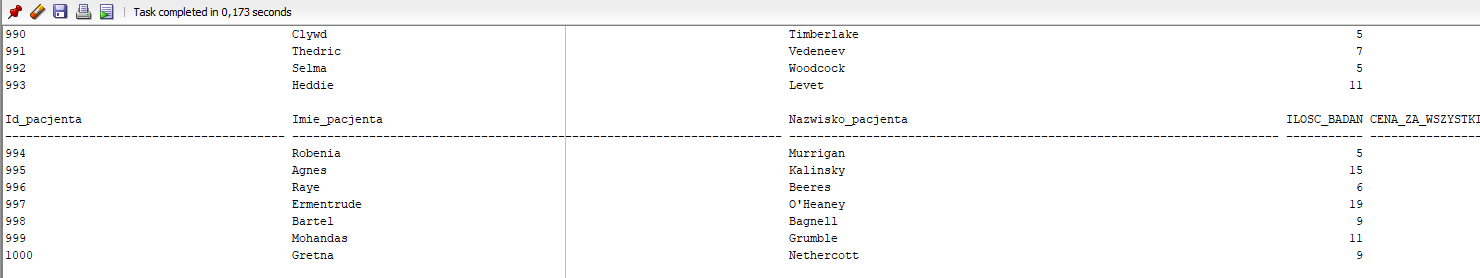
****

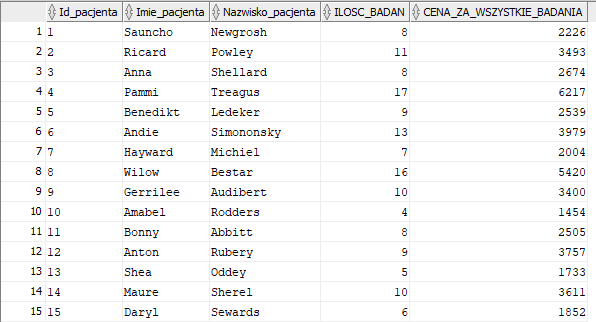
****

**Odpowiedź nr.3 baza danych**

****

**Zapytanie nr.3 hurtownia danych**

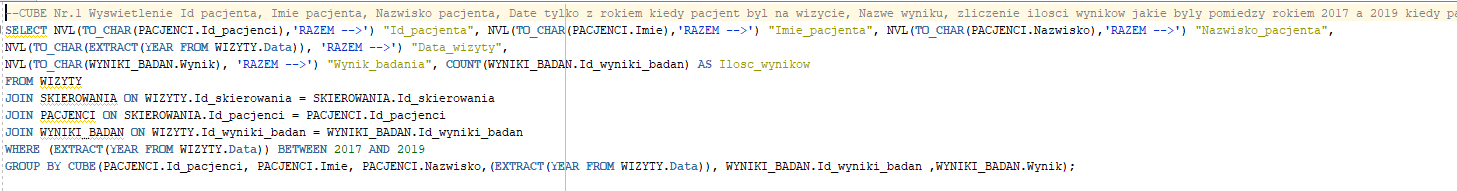
****

****

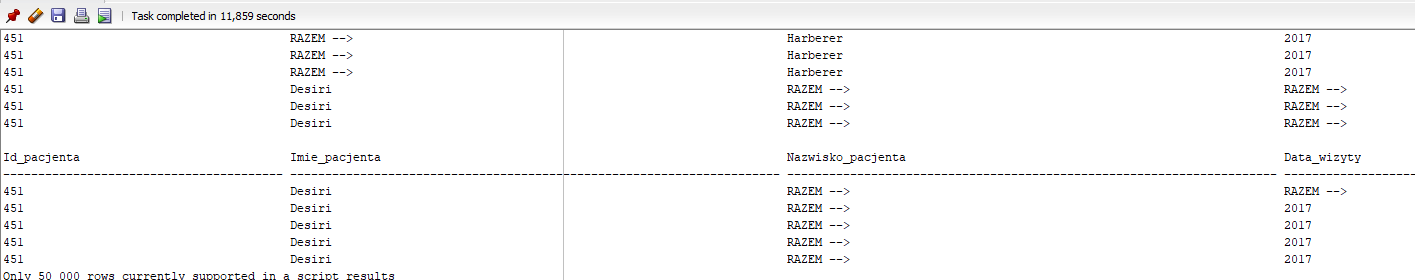
**Odpowiedź nr.3 hurtownia danych**

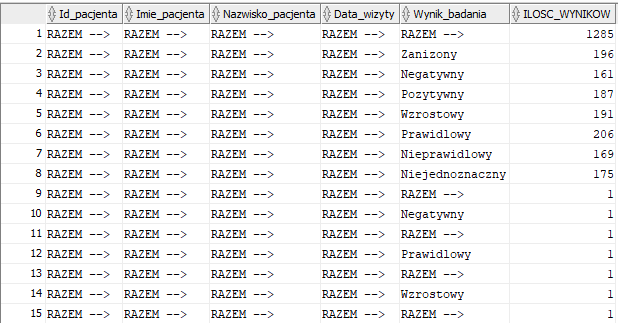
* **02 – CUBE:**

**Zapytanie nr.1:**

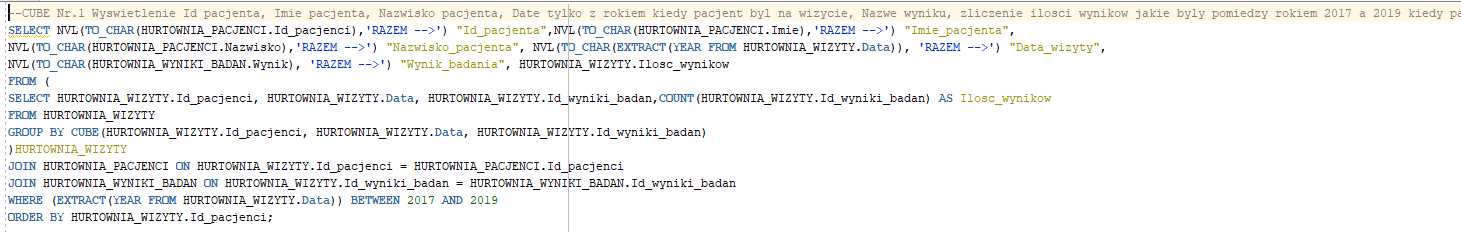


**Zapytanie nr.1 baza danych**

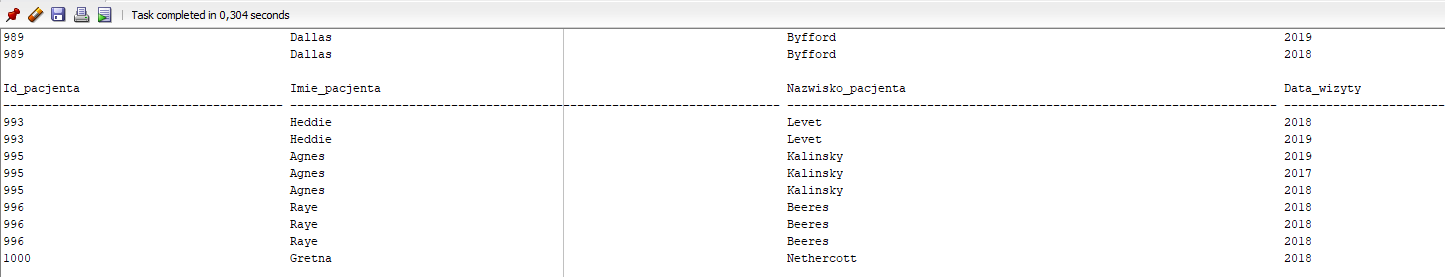
****

****

**Odpowiedź nr.1 baza danych**

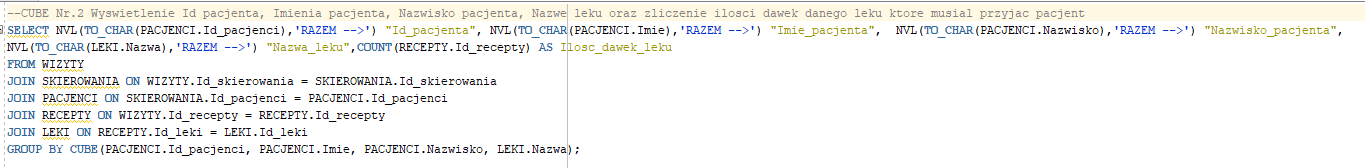
****

**Zapytanie nr.1 hurtownia danych**

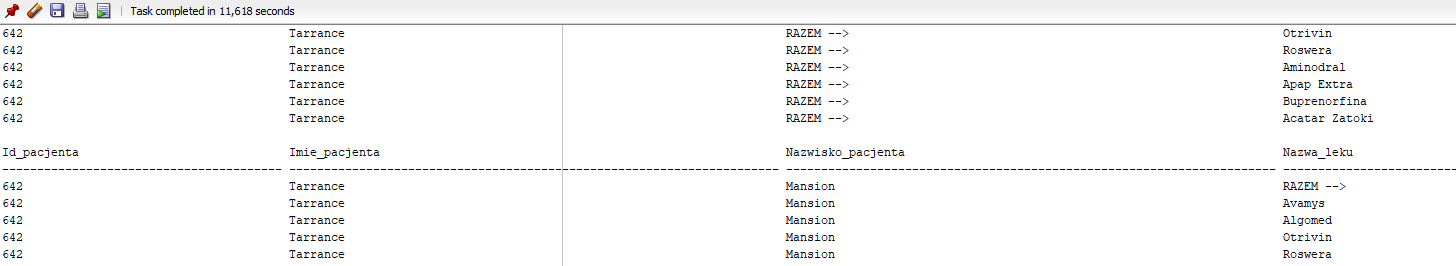
****

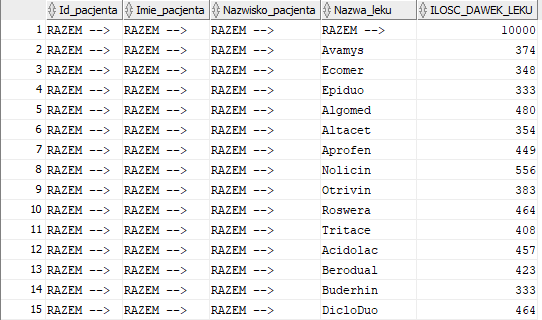
****

**Zapytanie nr.2:**

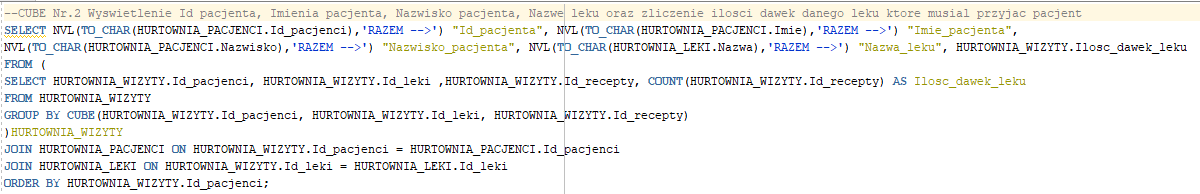
****

**Zapytanie nr.2 baza danych**

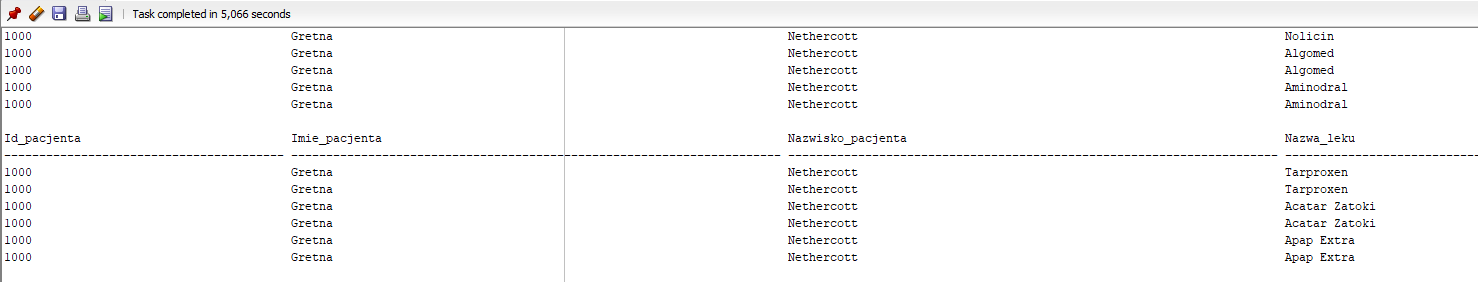




**Odpowiedź nr.2 baza danych**

****

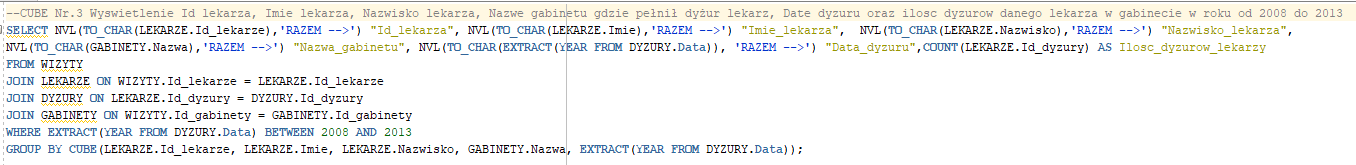
**Zapytanie nr.2 hurtownia danych**

****

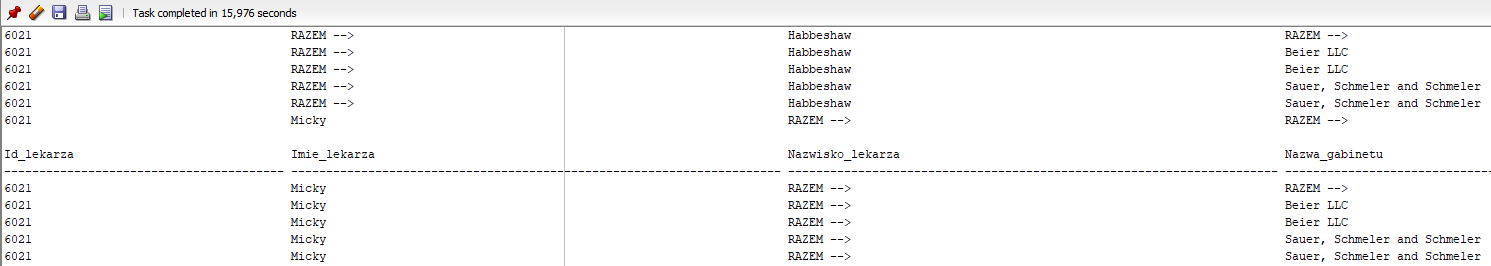
****

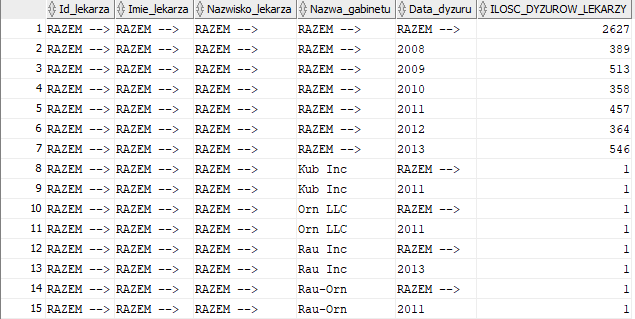
**Odpowiedź nr.2 hurtownia danych**

**Zapytanie nr.3:**

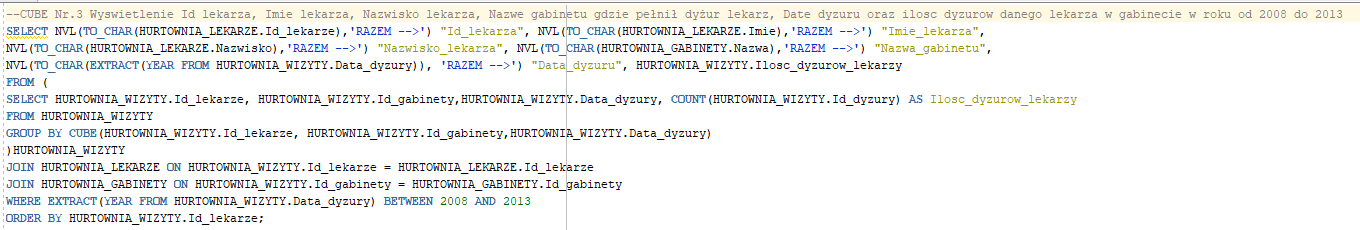
****

**Zapytanie nr.3 baza danych**

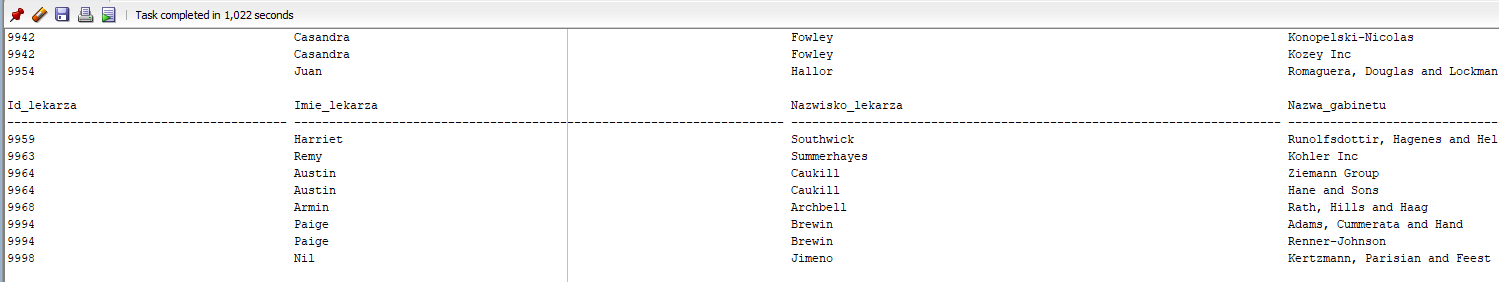
****

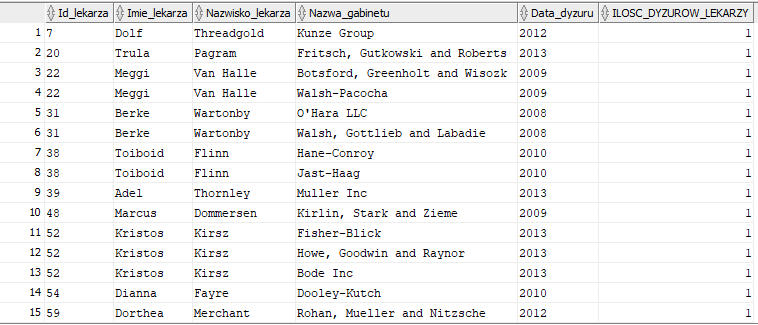
****

**Odpowiedź nr.3 baza danych**

****

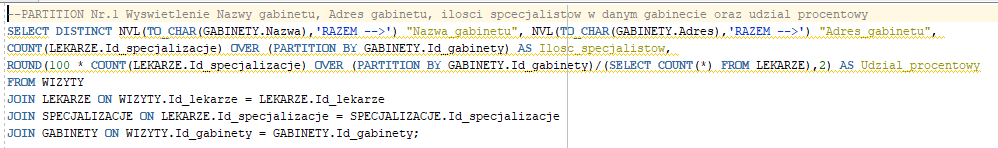
**Zapytanie nr.3 hurtownia danych**

****

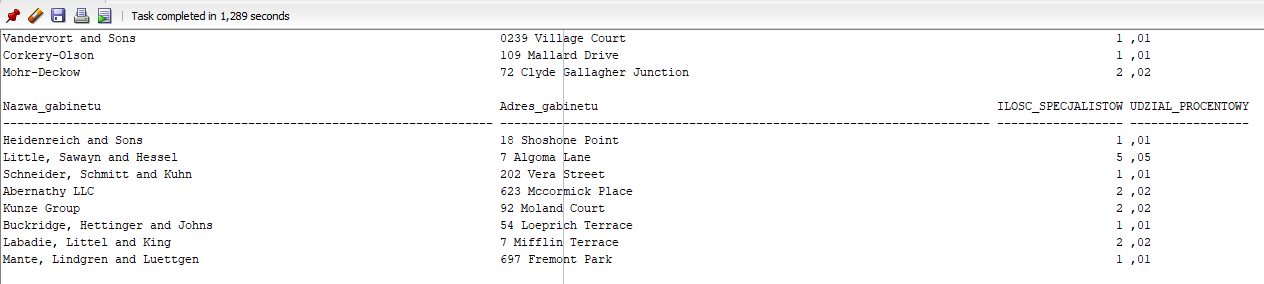
****

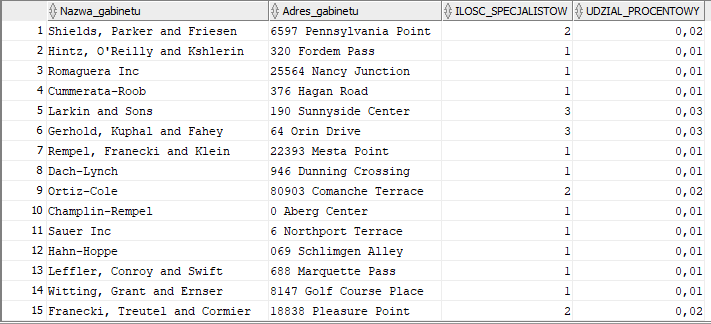
* **03 – PARTYCJE OBLICZENIOWE:**

**Zapytanie nr.1:**

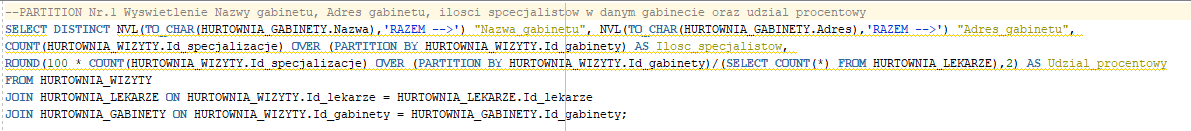
****

**Zapytanie nr.1 baza danych**

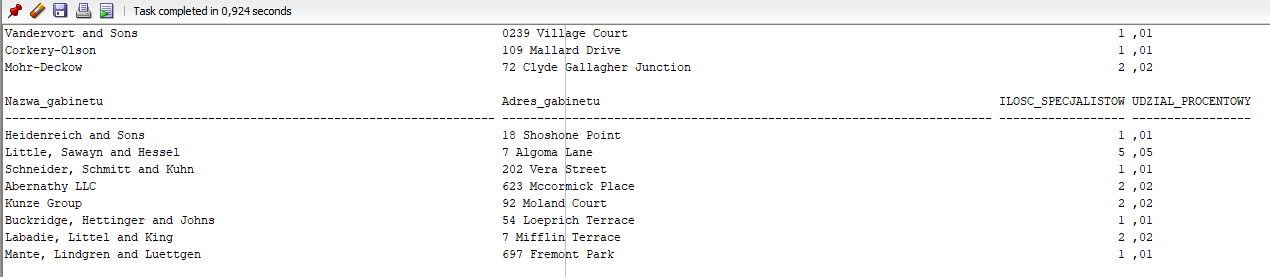
****

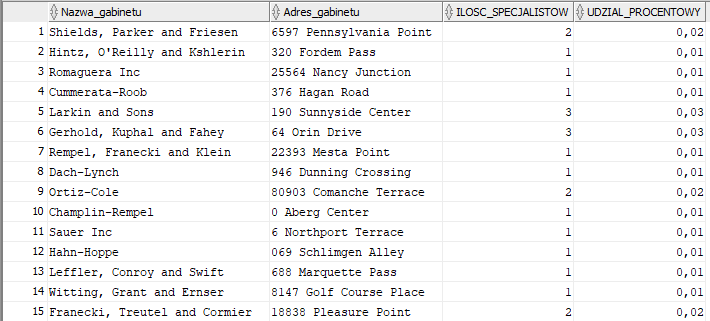
****

**Odpowiedz nr.1 baza danych**

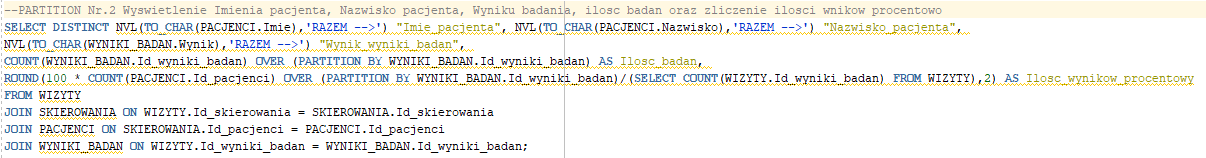
****

**Zapytanie nr.1 hurtownia danych**

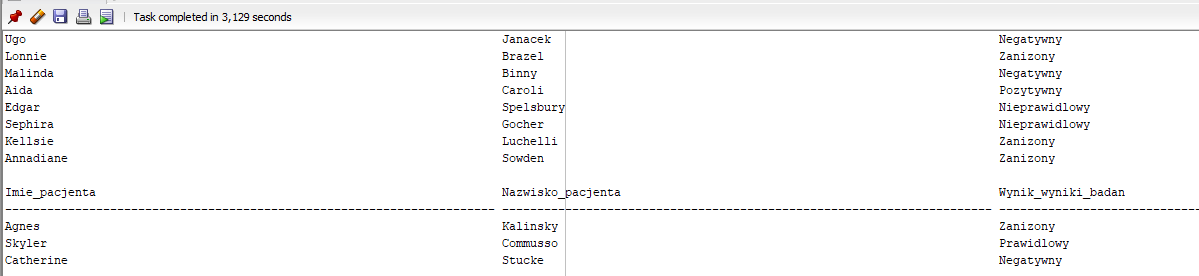
****

****

**Zapytanie nr.2:**

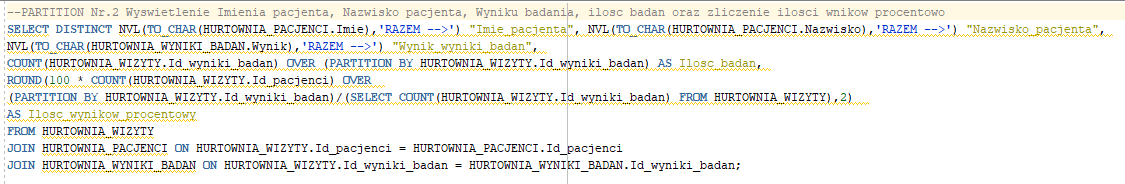
****

**Zapytanie nr.2 baza danych**

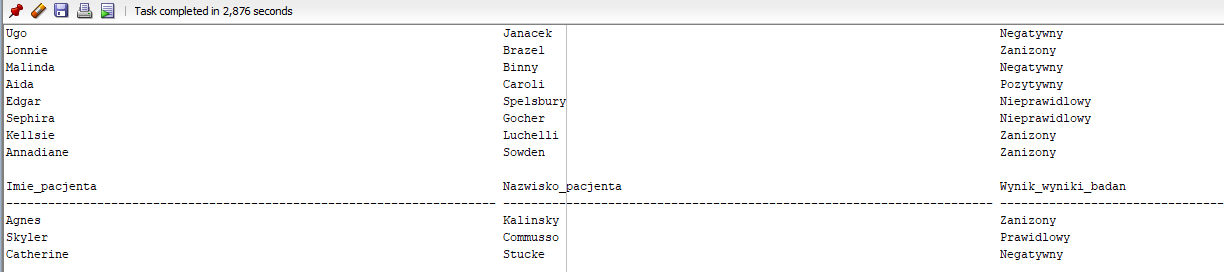
****

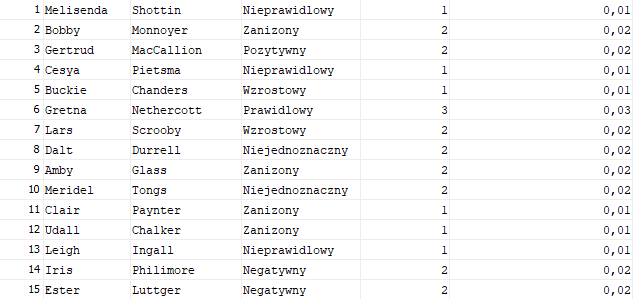
****

**Odpowiedz nr.2 baza danych**

****

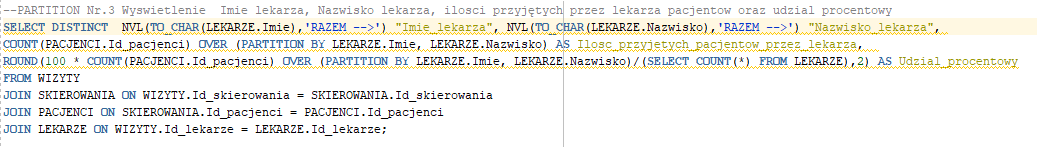
**Zapytanie nr.2 hurtownia danych**

****

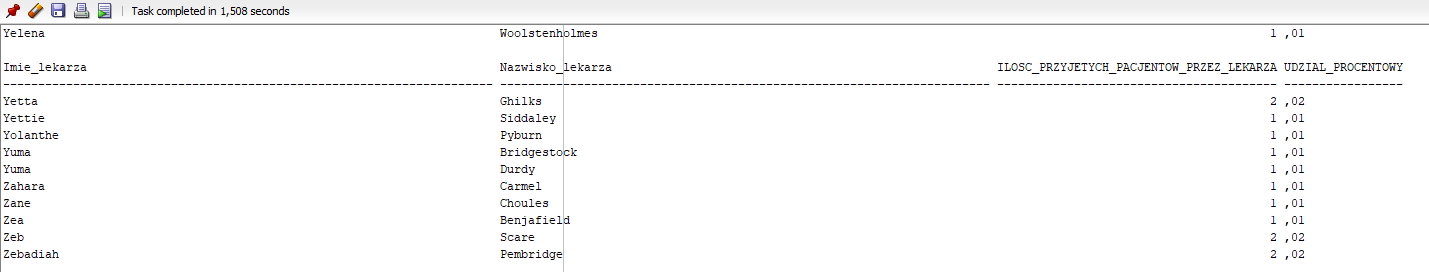
****

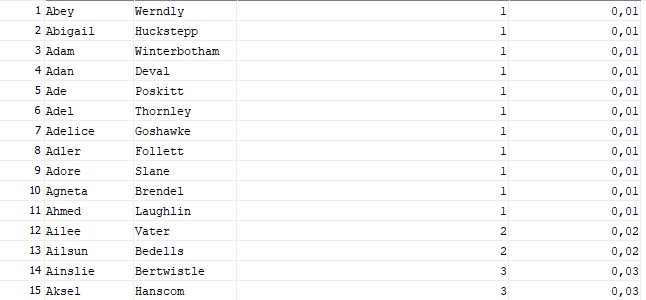
**Odpowiedz nr.2 hurtownia danych**

**Zapytanie nr.3:**

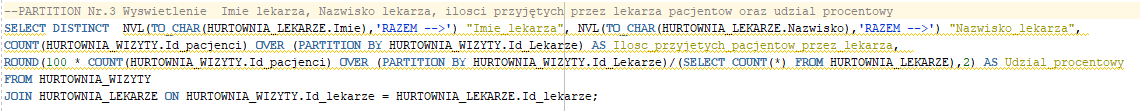
****

**Zapytanie nr.3 baza danych**

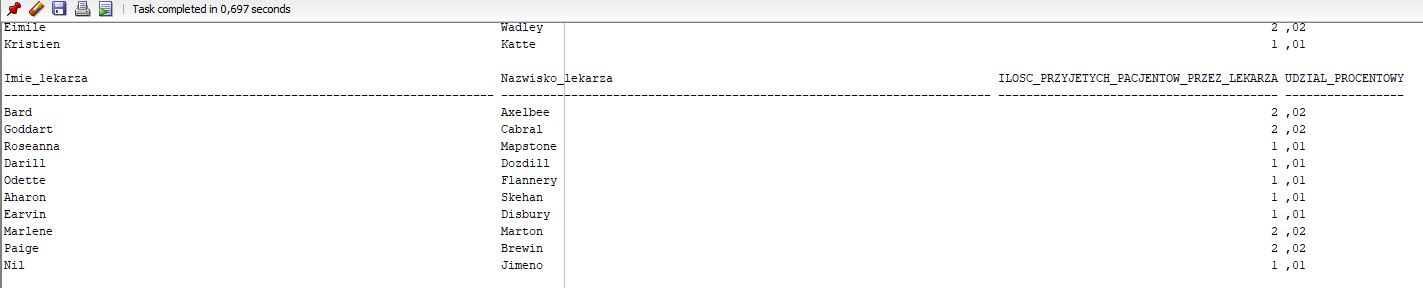
****

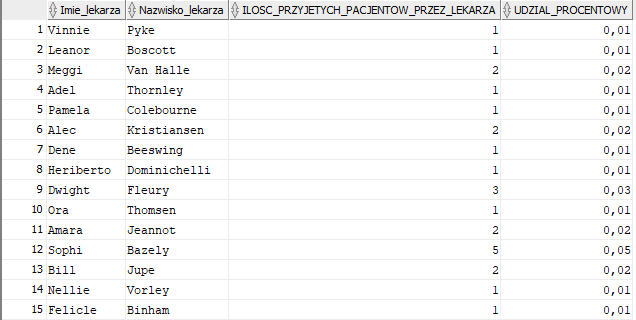
****

**Odpowiedz nr.3 baza danych**

****

**Zapytanie nr.3 hurtownia danych**

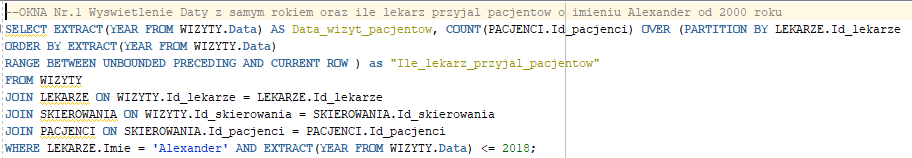
****

****

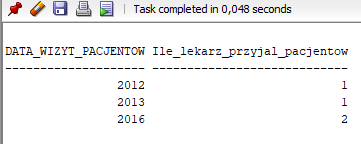
**Odpowiedz nr.3 hurtownia danych**

* **04 – OKNA:**

**Zapytanie nr.1:**

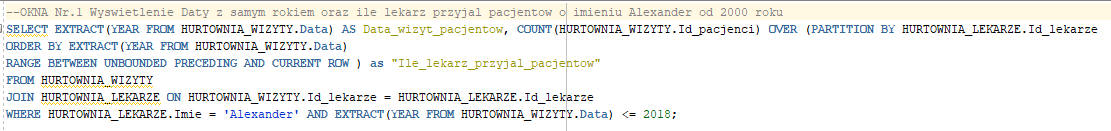
****

**Zapytanie nr.1 baza danych**

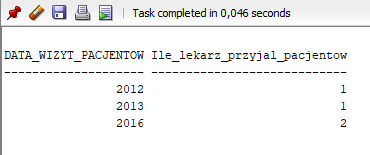
****

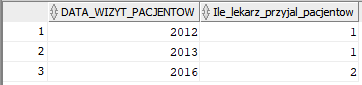
****

**Odpowiedz nr.1 baza danych**

****

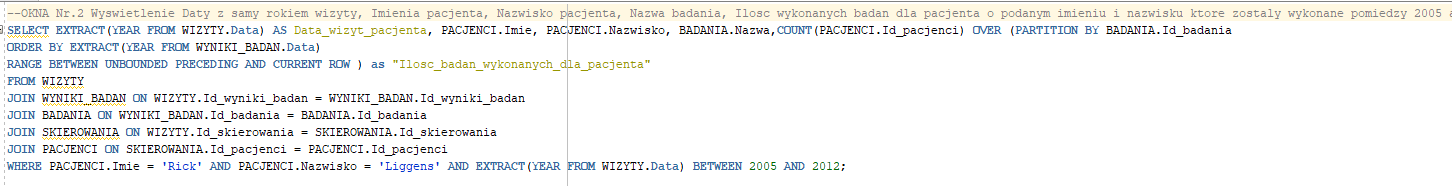
**Zapytanie nr.1 hurtownia danych**

****

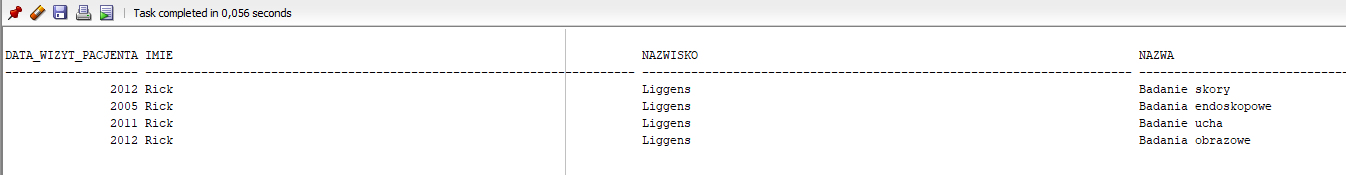
****

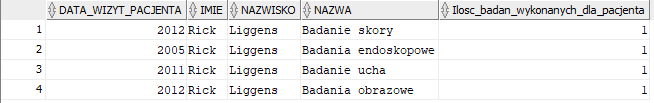
**Odpowiedz nr.1 hurtownia danych**

**Zapytanie nr.2:**

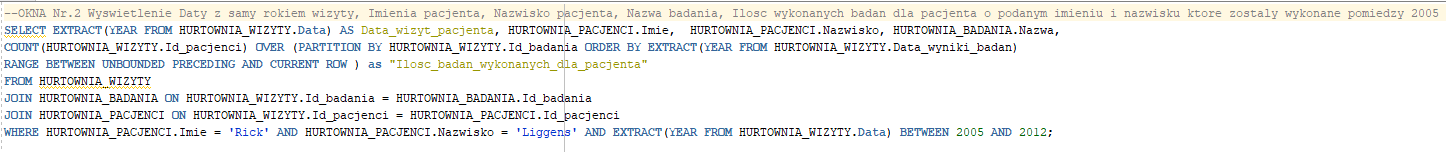
****

**Zapytanie nr.2 baza danych**

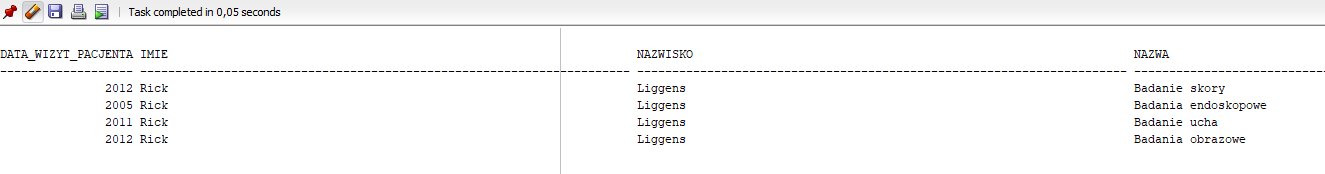
****

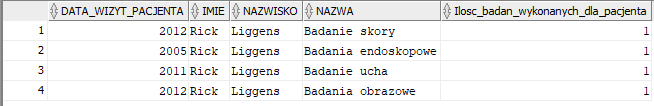
****

**Odpowiedz nr.2 baza danych**

****

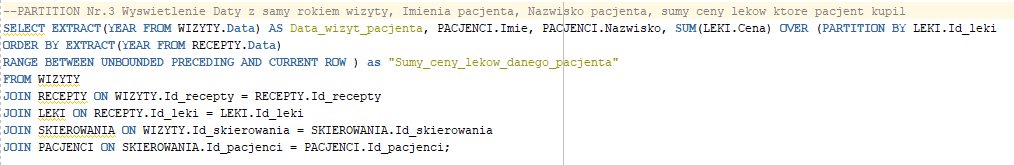
**Zapytanie nr.2 hurtownia danych**

****

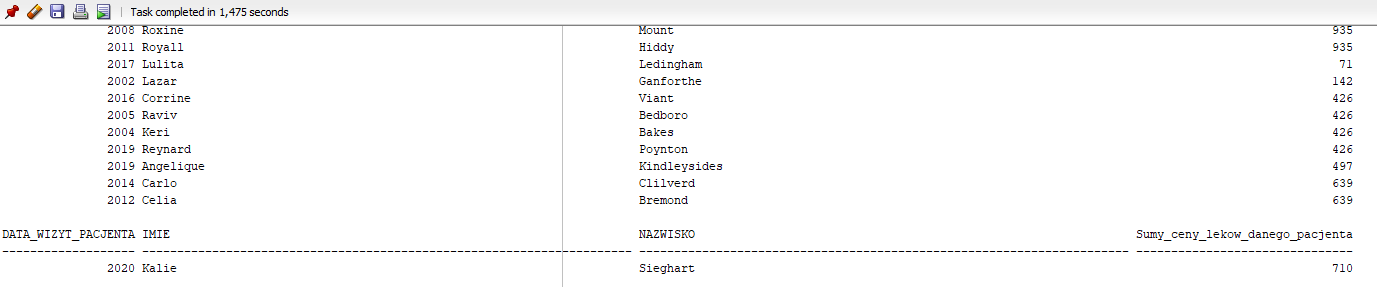
****

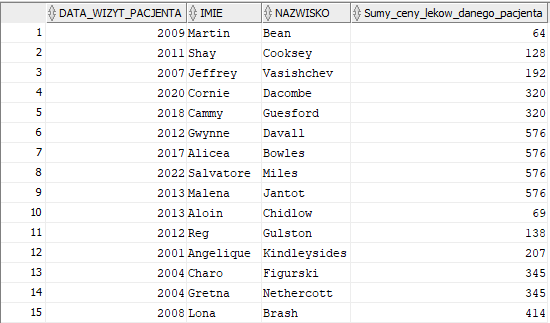
**Odpowiedz nr.2 hurtownia danych**

**Zapytanie nr.3:**

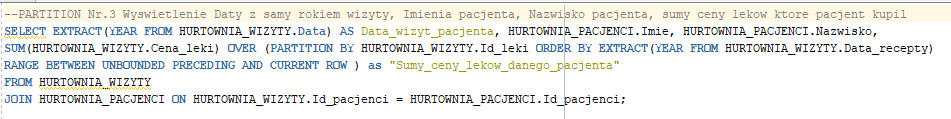
****

**Zapytanie nr.3 baza danych**

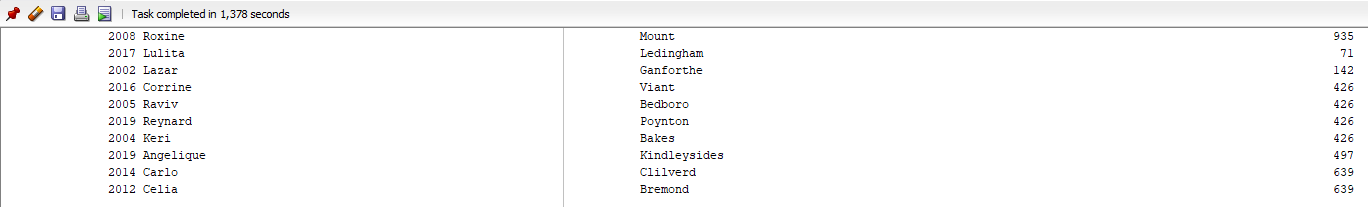
****

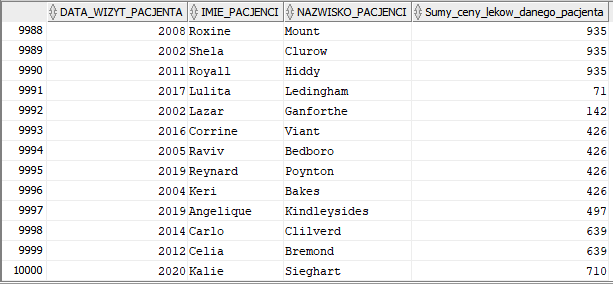
****

**Odpowiedz nr.3 baza danych**

****

**Zapytanie nr.3 hurtownia danych**

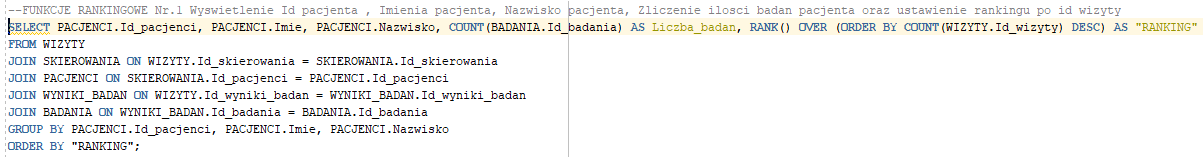
****

****

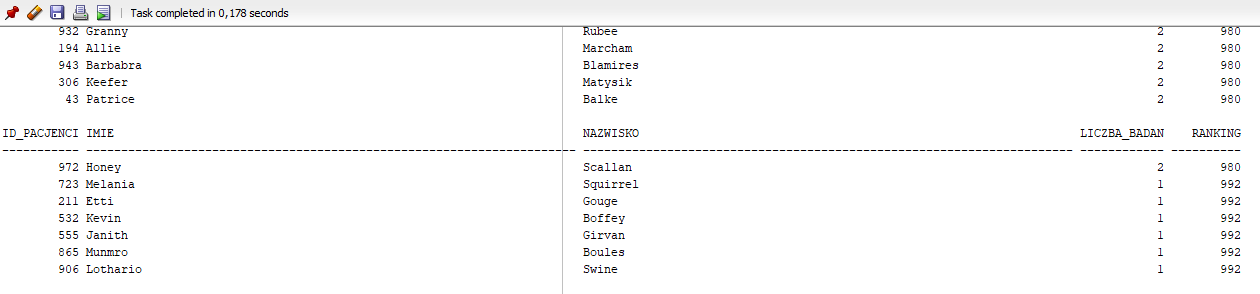
**Odpowiedź nr.3 hurtownia danych**

* **05 – FUNKCJE RANKINGOWE:**

**Zapytanie nr.1:**

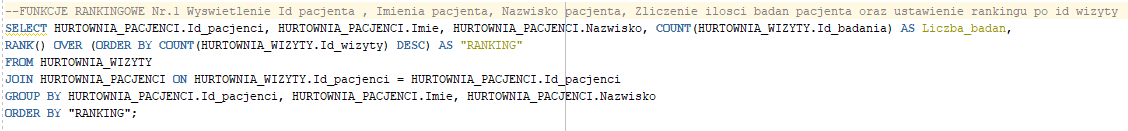
****

**Zapytanie nr.1 baza danych**

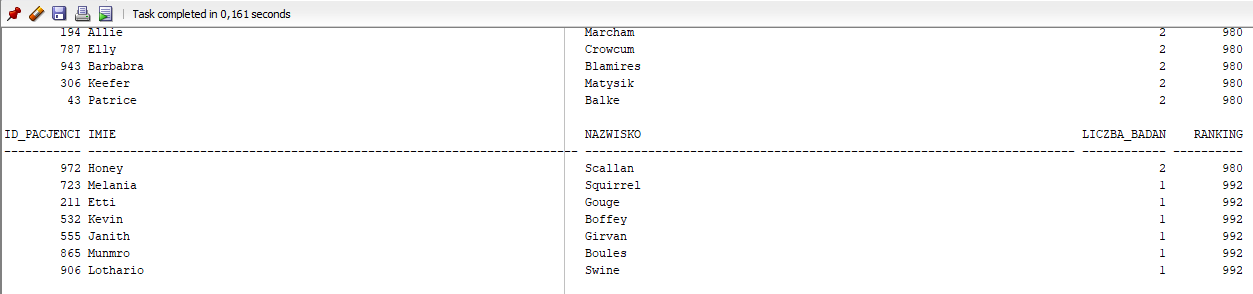
****

****

**Odpowiedz nr.1 baza danych**

****

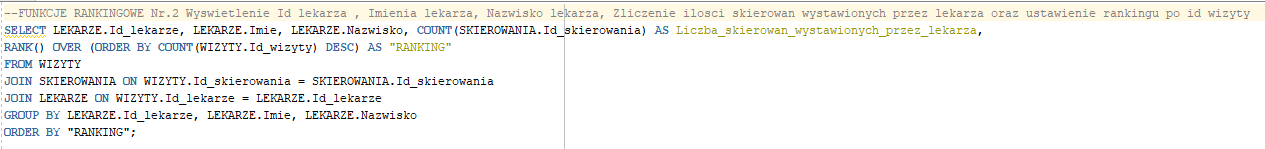
**Zapytanie nr.1 hurtownia danych**

****

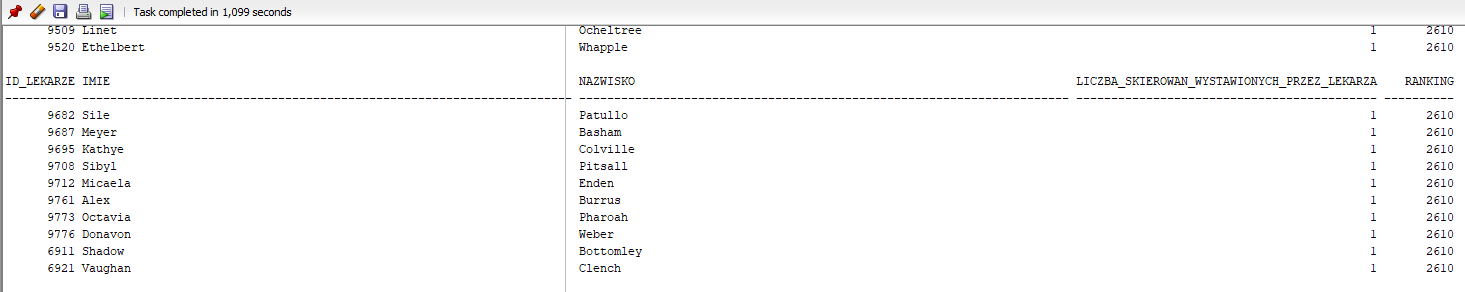
****

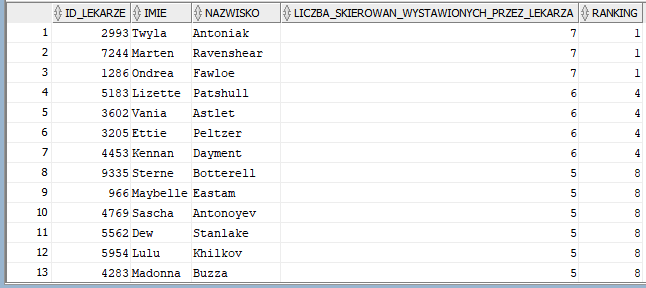
**Odpowiedz nr.1 hurtownia danych**

**Zapytanie nr.2:**

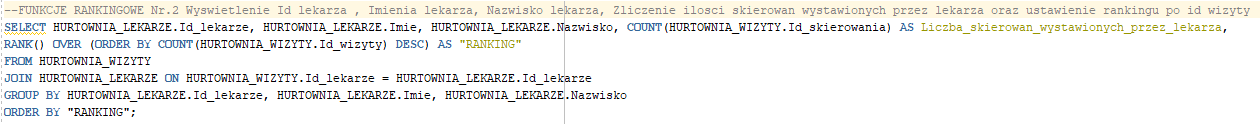
****

**Zapytanie nr.2 baza danych**

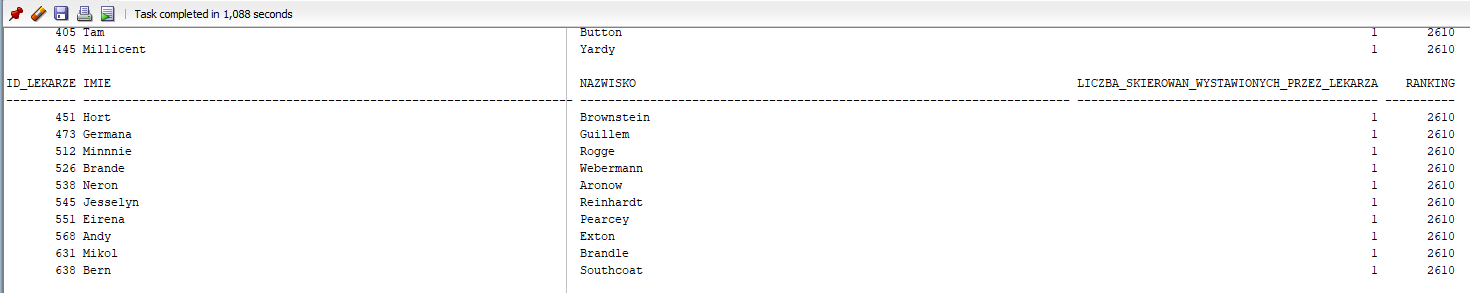
****

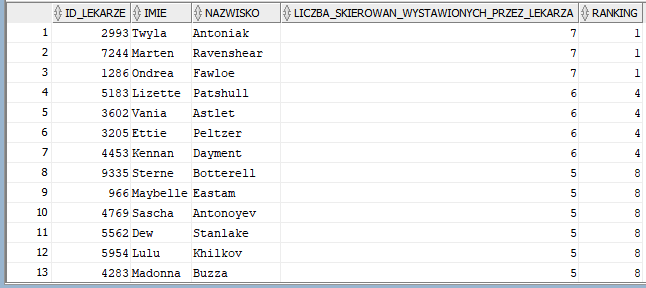
****

**Odpowiedź nr.2 baza danych**

****

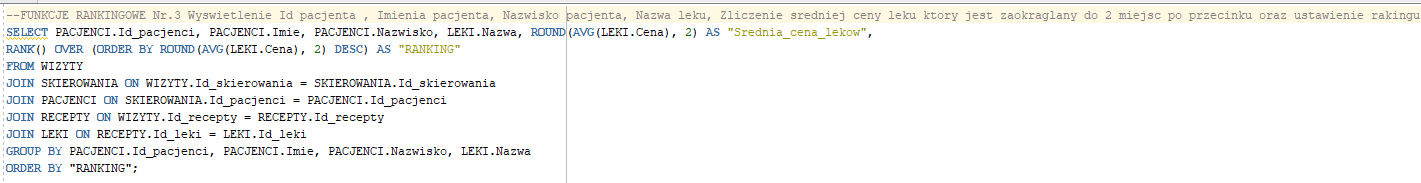
**Zapytanie nr.2 hurtownia danych**

****

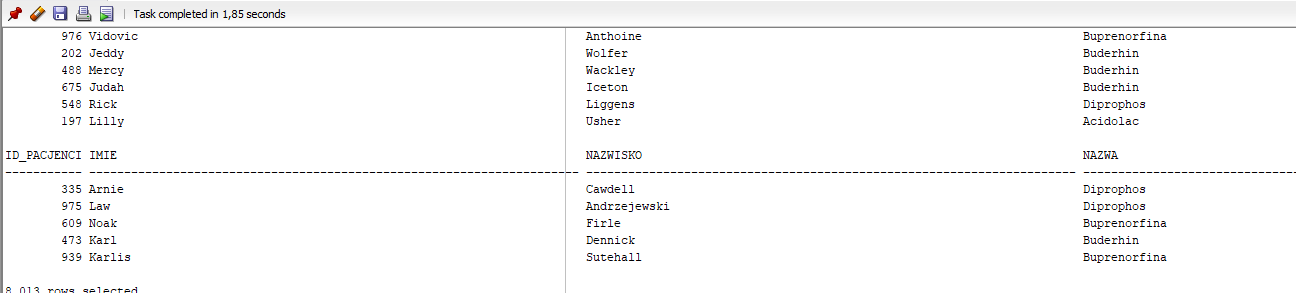
****

**Odpowiedz nr.2 hurtownia danych**

**Zapytanie nr.3:**

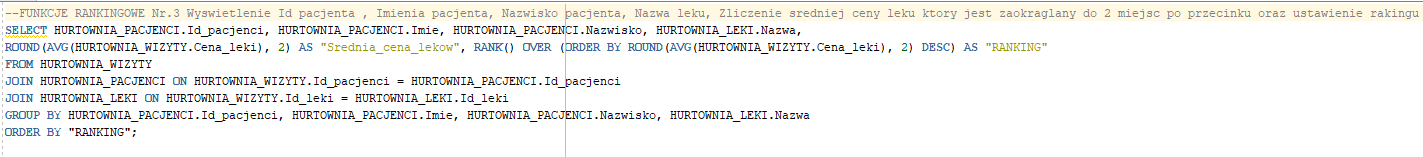
****

**Zapytanie nr.3 baza danych**

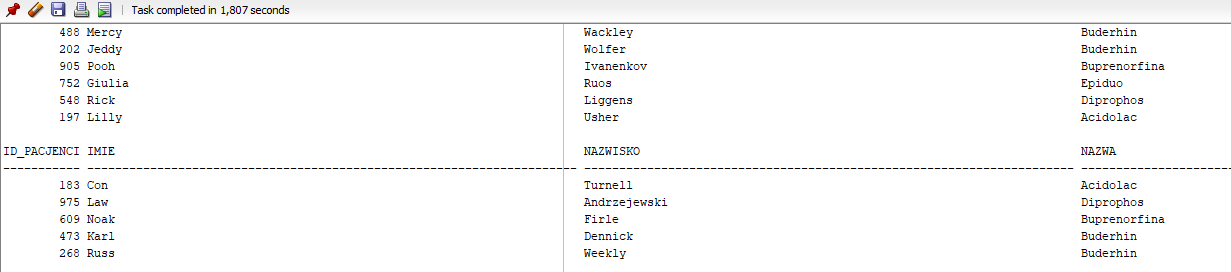
****

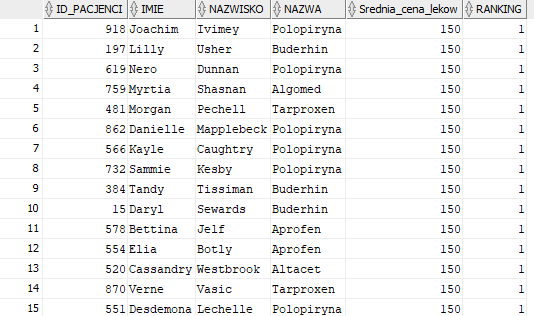
****

**Odpowiedz nr.3 baza danych**

****

**Zapytanie nr.3 hurtownia danych**

****

****

**Odpowiedz nr.3 hurtownia danych**

1. **Podsumowanie:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ZAPYTANIE NR.1 | ZAPYTANIE NR.2 | ZAPYTANIE NR.3 |
| BAZA DANYCH  **ROLLUP** | **1,092 sek** | **0,850 sek** | **1,498 sek** |
| HURTOWNIA DANYCH  **ROLLUP** | **0,196 sek** | **0,245 sek** | **0,173 sek** |
| BAZA DANYCH  **CUBE** | **11,859 sek** | **11,618 sek** | **15,976 sek** |
| HURTOWNIA DANYCH  **CUBE** | **0,304 sek** | **5,066 sek** | **1,022 sek** |
| BAZA DANYCH  **PARTYCJE**  **OBLICZENIOWE** | **1,289 sek** | **3,129 sek** | **1,508 sek** |
| HURTOWNIA DANYCH  **PARTYCJE**  **OBLICZENIOWE** | **0,924 sek** | **2,876 sek** | **0,697 sek** |
| BAZA DANYCH  **OKNA** | **0,048 sek** | **0,056 sek** | **1,475 sek** |
| HURTOWNIA DANYCH  **OKNA** | **0,046 sek** | **0,050 sek** | **1,378 sek** |
| BAZA DANYCH  **FUNKCJE**  **OBLICZENIOWE** | **0,178 sek** | **1,099 sek** | **1,850 sek** |
| HURTOWNIA **FUNKCJE**  **OBLICZENIOWE** | **0,161 sek** | **1,088 sek** | **1,807 sek** |

**Tabela podsumowująca wyniki zapytań**

1. **Uwagi i wnioski:**

Nie wszystkie wygenerowane dane poprzez stronę internetową oddają rzeczywiste warunki. Gdzie w mojej hurtowni przykładowo jeden lekarz miał maksymalnie trzy wizyty przez okres prawie 20 lat.

W tabeli z podsumowaniem wynika że hurtownia danych zawsze jest szybsza od bazy danych. Wynik to z:

* Jakości sprzętu na którym pracujemy,
* Łącza internetowego,
* Odległości od której znajduje się baza danych,
* Wykonanego zapytania do bazy,
* Serwera na którym znajduje się baza danych,

Projekt ten udowadnia, że hurtownia danych jest dobrym rozwiązaniem dla dużych ilości danych. Po upływie czasu dane w hurtowni wciąż przybywają dzięki czemu hurtownia zyskuje przewagę pod względem wydajności. W moim przypadku gdzie operowaliśmy na małej ilości danych czasy wykonania zapytań są bardzo podobne.