**Sprawozdanie Projektowe**

**Bazy Danych 2**

1. Dane zespołu
   * Patryk Grzywacz Gr.2ID12A 090111
   * Dominik Grudzień Gr.2ID12A 090798
2. Opis Problemu
   * Temat „Sieć Gabinetów Lekarskich”

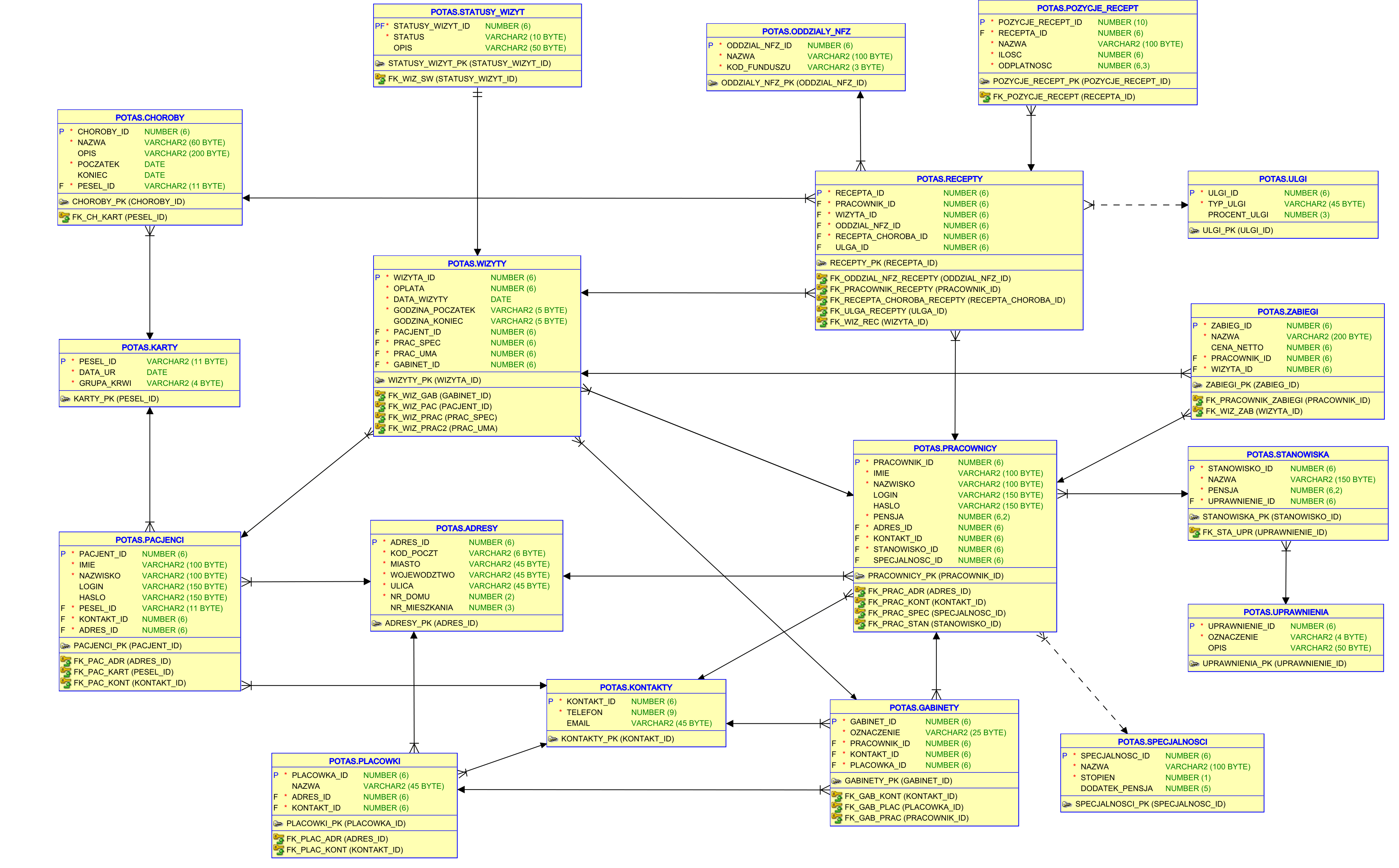
Podeszliśmy do tematu racjonalnie, projektując bazę danych ,która mogłaby faktycznie służyć w prawdziwych okolicznościach jako narzędzie do przechowywania danych ,potrzebnych do działania takich placówek lekarskich.

Rozdzieliliśmy sekcję gabinetów na placówki do, których one przynależą, a przynajmniej w prawdziwym świecie tak jest.

Baza danych zbudowana jest z 18 tabel ,które przechowują istotne elementy, od danych pacjentów, lekarzy i innych pracowników ,po przeróżne wizyty , zabiegi oraz wystawione recepty na leki .

Projektując relacje pomiędzy poszczególnymi tabelami staraliśmy się kierować logiką , aby nie tworzyć niepotrzebne połączenia , czy jak się tyczy samych połączeń, aby ich typy były zgodne z logiką ich późniejszych zastosowań , stosując 2 z 3 dostępnych typów relacji tj: 1-1 czy też 1-n .

1. Schemat ERD Bazy



1. Opis Tabel
   * Adresy – Przechowuje dane dotyczące adresu fizycznego domu/budynku/placówki etc.
   * Kontakty – Przechowuje dane odnośnie możliwości kontaktowych zarówno osób jak i placówek medycznych.
   * Specjalności – Przechowuje dane o specjalnościach lekarzy.
   * Uprawnienia – Znajdują się w niej dane o uprawnieniach przypisywanych do stanowisk.
   * Stanowiska – Przechowują dane odnośnie stanowisk przydzielanych do poszczególnych pracowników.
   * Pracownicy – Przechowują dane osobiste jak i firmowe o pracownikach , a także odnośniki do adresów, kontaktów, stanowisk oraz ewentualnie do specjalności.
   * Placówki – Zawierają dane o placówkach medycznych ,adresie oraz kontakcie z nimi.
   * Gabinety – Zawierają dane o gabinetach lekarskich mieszczących się w placówkach ,do których przypisany jest odpowiedni pracownik, najczęściej jakiś lekarz.
   * Karty – Przechowują dane osobowe tj. data urodzenia czy też grupę krwi.
   * Pacjenci – Przechowują dane osobowe jak i klienckie, odnośniki do kart ,adresów oraz kontaktów.
   * Choroby – Przechowują dane odnośnie choroby/chorób przypisanych do danej Karty pacjenta
   * Ulgi – Zawierają dane ulg przeznaczonych do wykorzystania przy receptach.
   * Pozycje\_Recept – Zawierają dane materialne odnośnie leków przepisanych w ramach danej recepty.
   * Oddzialy\_NFZ – Zawierają dane specjalistyczne wymagane na każdej recepcie.
   * Statusy\_Wizyt – Zawierają dane określające status danej wizyty.
   * Wizyty – Przechowuje dane o wizytach pacjentów ,o dacie i czasie trwania, miejscu, opłacie a także kto umówił/przyjmował pacjenta.
   * Recepty – Zawiera same odnośniki do pracownika ,który ją wystawił , do wizyty podczas ,której została wystawiona a także do choroby , ulgi i oddziału nfz.
   * Zabiegi – Zawiera dane o zabiegu wykonanym w ramach danej wizyty , jego cenie oraz pracowniku, który go przeprowadzał.
2. Opis procesu transformacji bazy do hurtowni

W celu transformacji relacyjnej bazy danych do hurtowni posłużyliśmy schematem konstelacyjnym, w którym mogą się znaleźć różne rodzaje faktów, a pewne wymiary są współdzielone. Do naszych tabel faktów zaliczyły się tabele Pozycje\_Recept oraz Wizyty. Fakt Pozycje\_Recept posiada klucze obce do tabel: Recepty, Leki, Ulgi oraz miary: ilość (danego leku), procent\_ulgi i odplatnosc.

Fakt Wizyty ma klucze obce do tabel: Pracownicy, Statusy\_Wizyt, Daty\_Wizyt, Zabiegi, Pacjenci, Gabinety, Recepty oraz miary: oplata (za wizyte) i cena\_netto\_za\_zabieg.

Opis wymiarów faktu Pozycje\_recept:

* Ulgi – wymiar określający nazwy typów ulg. Kolumna procent\_ulgi została przeniesiona do faktu pozycje\_recept i pełni rolę miary w tym fakcie.
* Leki – wymiar określający nazwy leków.
* Recepty – wymiar współdzielony z tabelą faktów Wizyty. Określa receptę, która jest powiązana z lekami i wizytą.

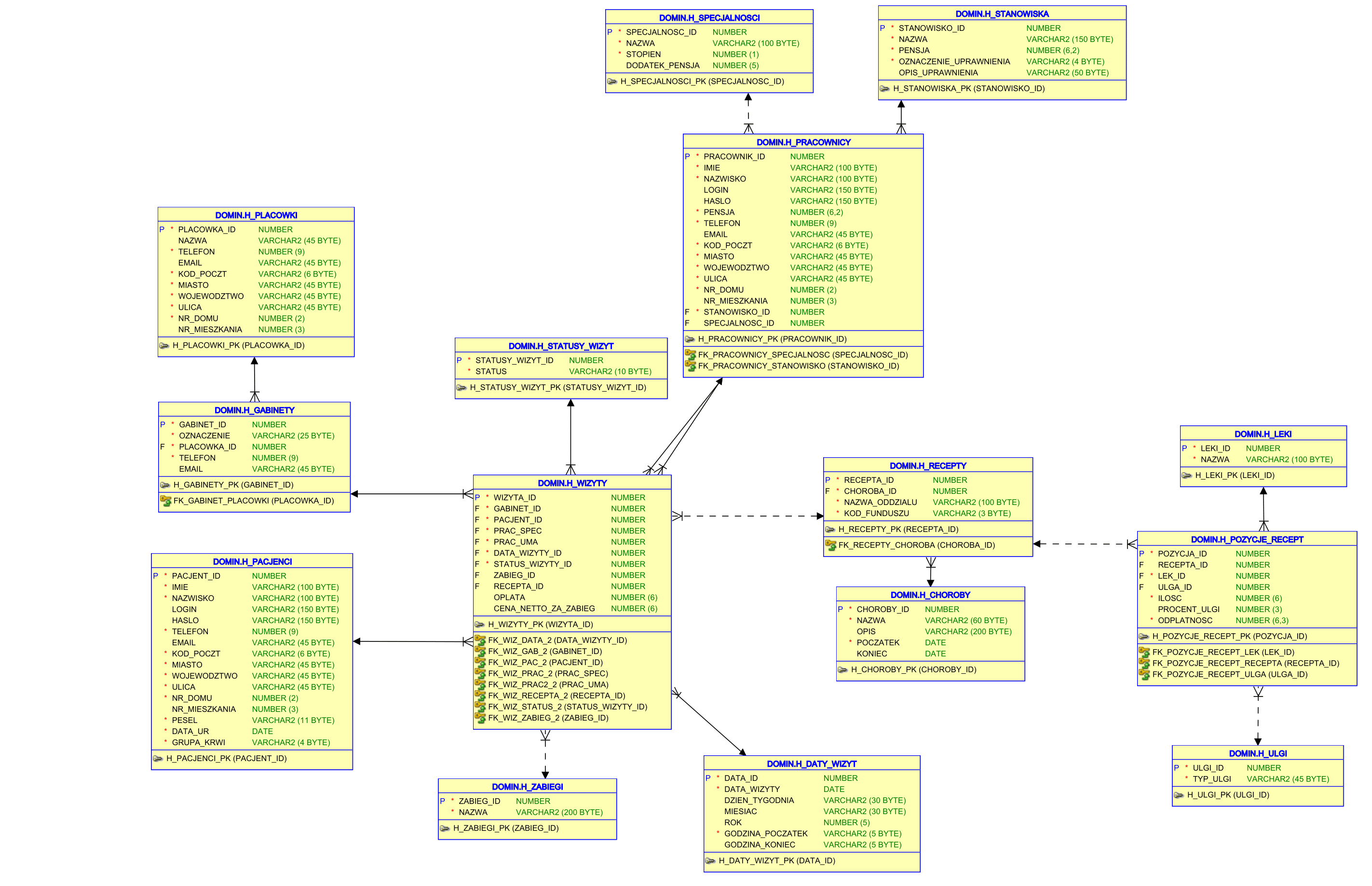
Opis wymiarów faktu Wizyty:

* Statusy\_Wizyt – wymiar opisujący status wizyty.
* Pracownicy – wymiar opisujący pracownika związanego z wizytą. Zostały do niego wcielone dane z dwóch tabel: Kontakty i Adresy.
* Recepty - wymiar współdzielony z tabelą faktów Pozcyje\_Recept. Określa receptę, która jest powiązana z lekami i wizytą.
* Daty\_wizyt – wymiar opisujący dokładny czas odbycia się wizyty.
* Zabiegi – wymiar określający nazwę zabiegu związanego z wizytą.
* Pacjenci – wymiar opisujący pacjenta związanego z wizytą. Zostały do niego wcielone dane z trzech tabel: Kontakty, Adresy oraz Karty.
* Gabinety – wymiar opisujący gabinet, w którym odbyła się wizyta. Została do niego włączona tabela kontakty.

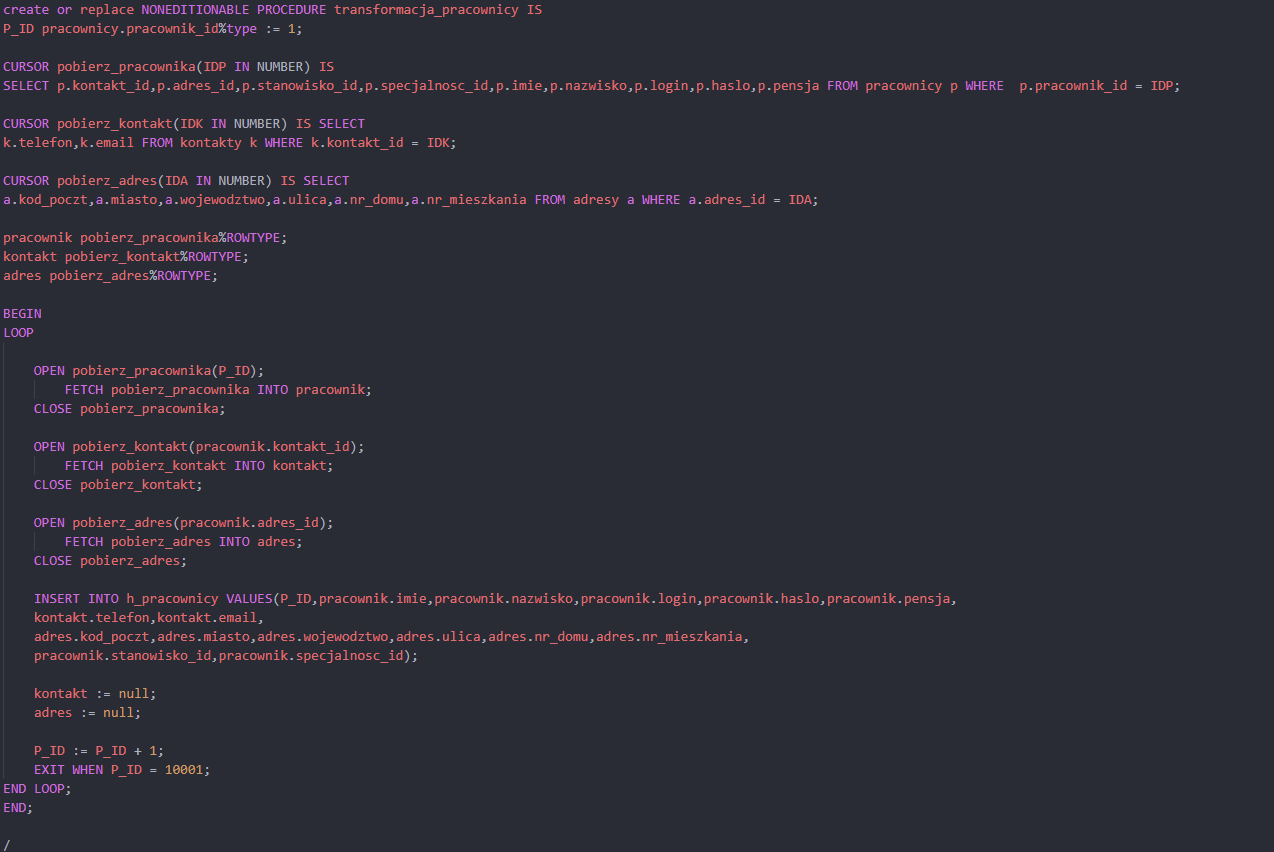
Wymiary Pracownicy, Gabinety i Recepty posiadają zewnętrzne tabele wymiarów. Mają one na celu stworzyć hierarchię tabel wymiarów. Wymiar Pracownicy posiada dwie tabele zewnętrzne: specjalności i stanowiska (do stanowisk zostały włączone dane z tabeli uprawnienia). Wymiar Gabinety ma z kolei jedną tabele zewnętrzną placówki (do placówek zostały włączone dane z tabel: Adresy i Kontakty). Natomiast wymiar Recepty posiada zewnętrzną tabele Choroby.

W schemacie konstelacyjnym współdzielonym wymiarem została tabela Recepty. Zostały do niej włączone dane z tabeli oddzialy\_nfz.

1. Schemat ERD Hurtowni

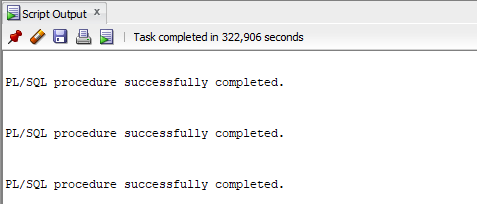


1. Opis procesu ładowania i transformacji danych z bazy do hurtowni
   * Do załadowania i przetransformowania danych z bazy do hurtowni użyliśmy własnoręcznie napisanych procedur, każda procedura wypełnia 1 tabelę z hurtowni danych, dla przykładu procedura „transformacja\_pracownicy”

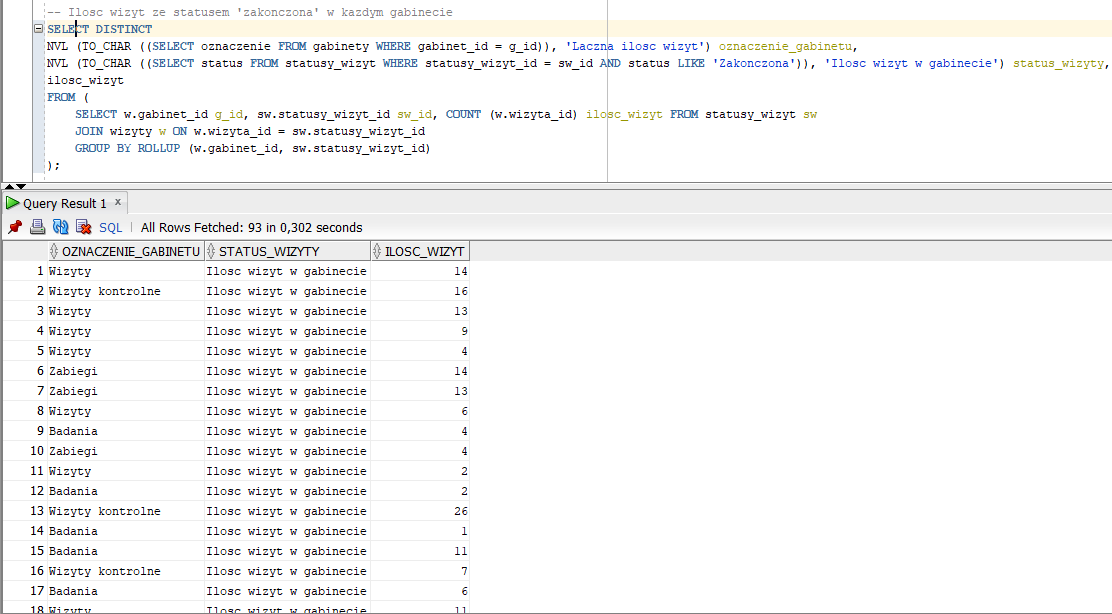


W tej procedurze deklarujemy 4 zmienne oraz 3 kursory, Iterator po tablicy pracownicy z bazy, kursory ,które pobierają wszystkie dane z tabeli podstawowej oraz tych ,które zostały wcielone w hurtowni do danej tabeli oraz 3 zmienne typu zwracanego rekordu kursora . Następnie w pętli iterujemy po tablicy pracownicy i wyciągamy interesujące nas dane poprzez kursor pobierz\_pracownika ,później pobieramy dane z adresów i kontaktów za pomocą kluczy obcych wyjętych 1 kursorem i następnie łączymy te wszystkie potrzebne dane w instrukcji insert do tabeli h\_pracownicy ,wypełniając ją danymi z bazy.

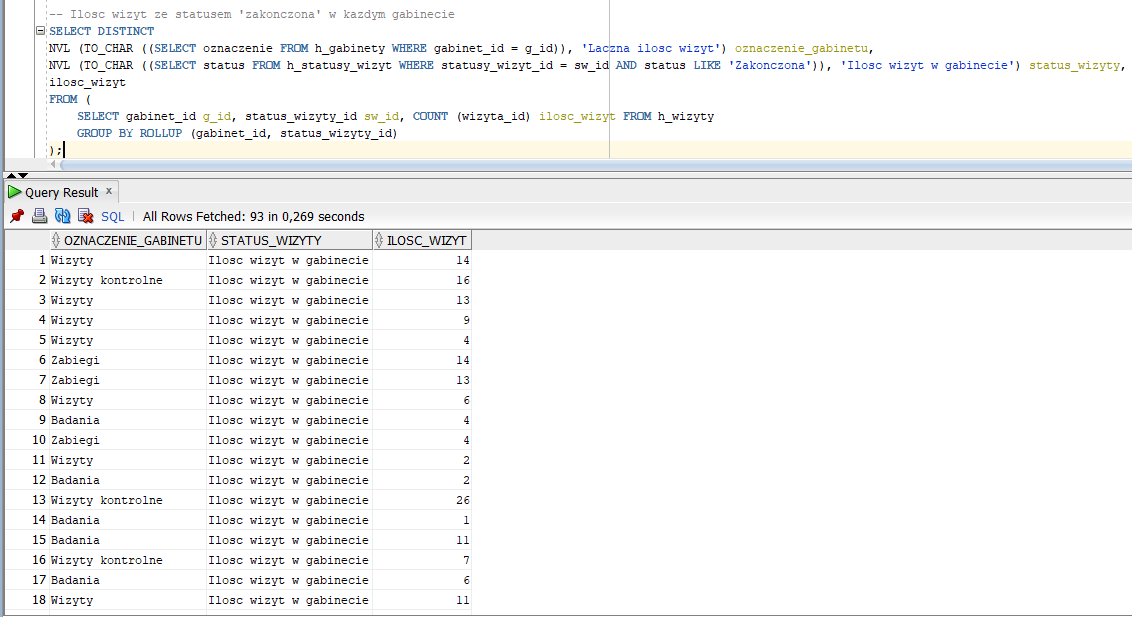
Wszystkie procedury transformacji danych trwają około 5 minut i 23 sekund.



1. Porównanie wykonania zapytań
   * Rollup nr1.
     + Baza

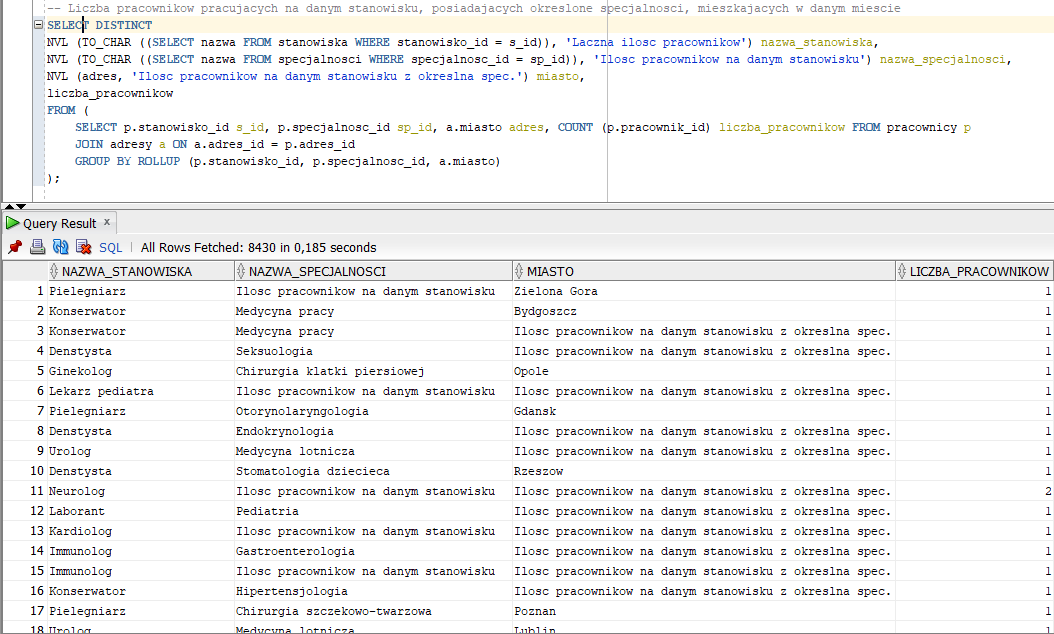


* + - Hurtownia

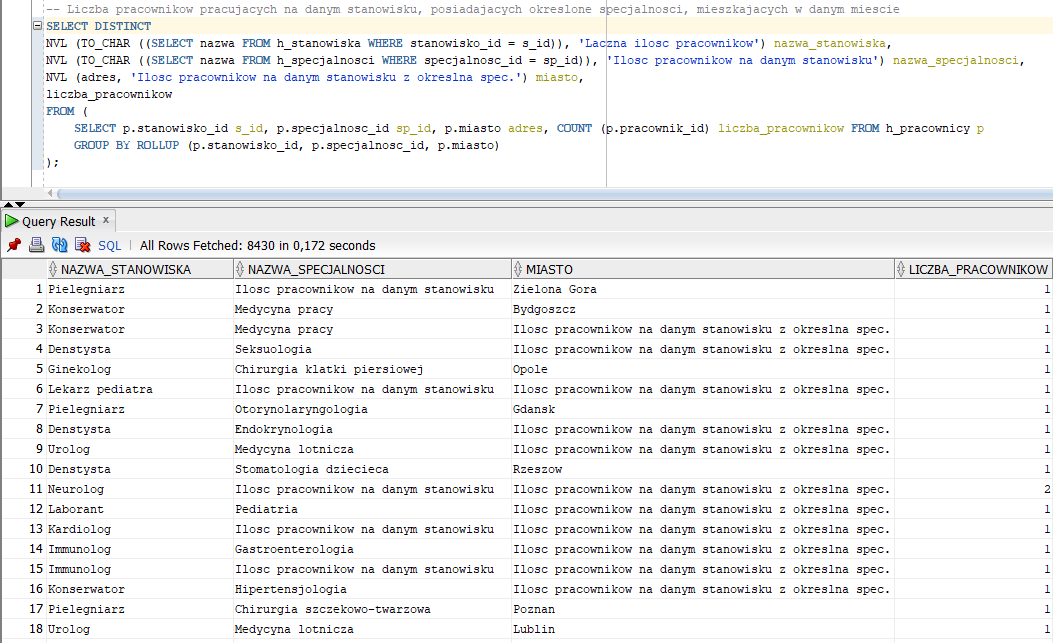


Różnicą między zapytaniem z bazy, a zapytaniem z hurtowni jest to ,że w bazie wymaga ono o jedno więcej połączenie JOIN. W hurtowni to zapytanie wykonuje się o 0,033s szybciej.

* + Rollup nr2.
    - Baza

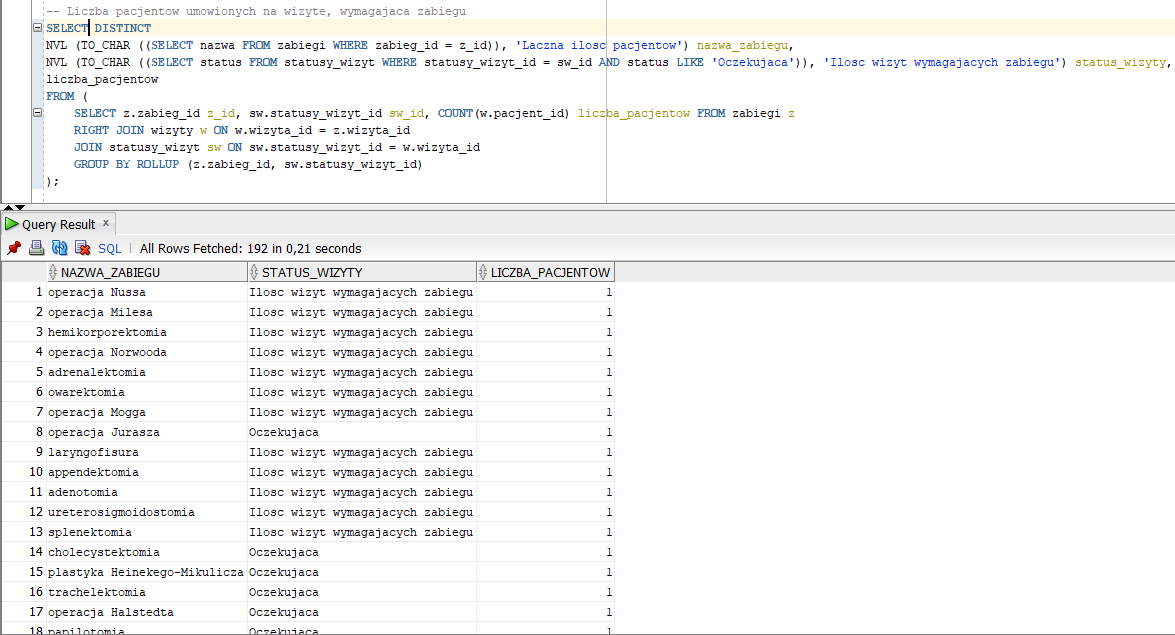


* + - Hurtownia

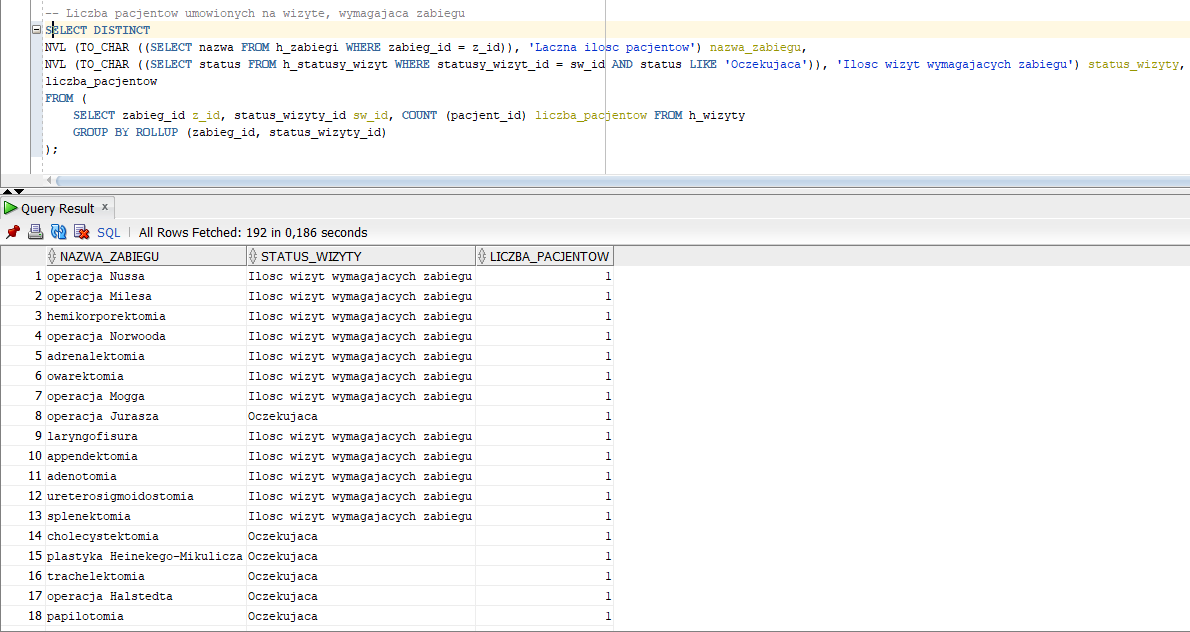


Różnicą między zapytaniem z bazy, a zapytaniem z hurtowni jest to, że w bazie wymaga ono o jedno więcej połączenie JOIN. W hurtowni to zapytanie wykonuje się o 0,013s szybciej.

* + Rollup nr3.
    - Baza

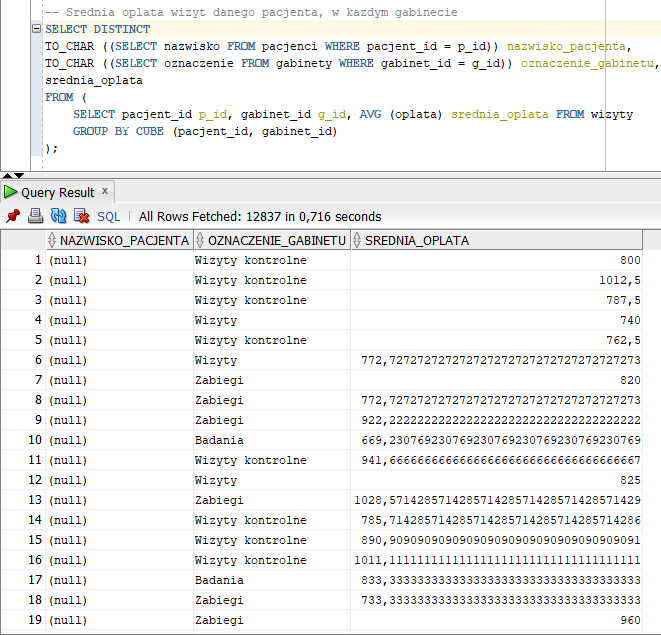


* + - Hurtownia

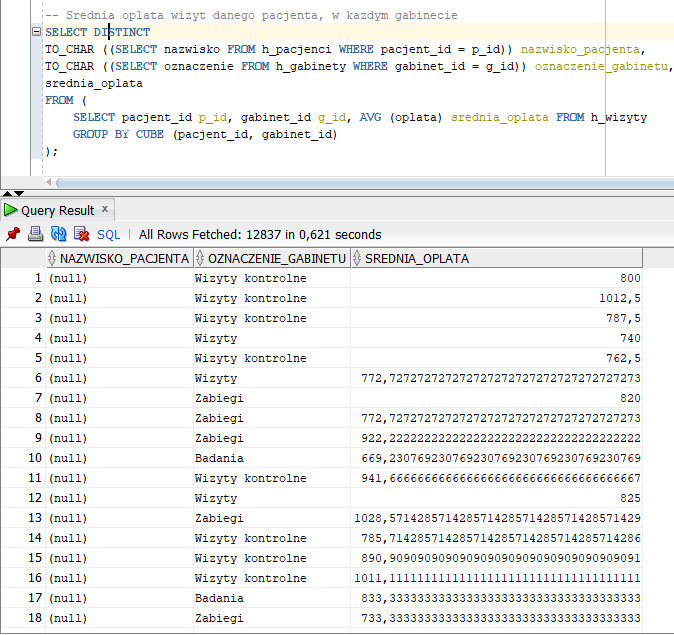


Różnicą między zapytaniem z bazy, a zapytaniem z hurtowni jest to, że w bazie wymaga ono o 2 połączenia JOIN więcej. W hurtowni to zapytanie wykonuje się o 0,024s szybciej.

* + Cube nr1.
    - Baza

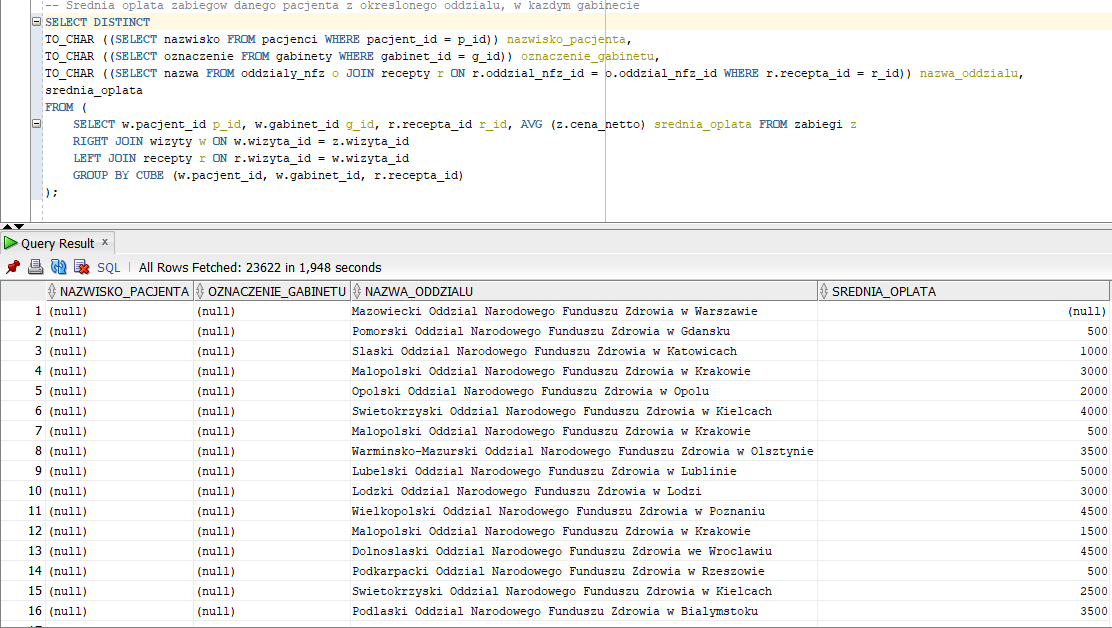


* + - Hurtownia

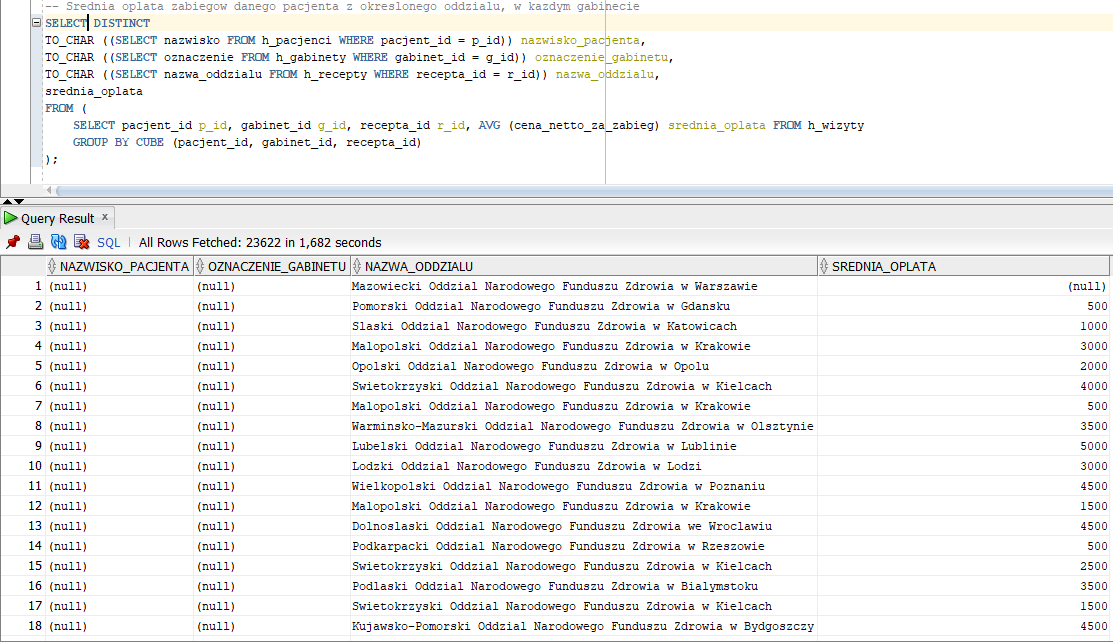


Wersja dla hurtowni jak i dla bazy jest taka sama, mimo to zapytanie do hurtowni danych wykonało się o 0,095s szybciej

* + Cube nr2.
    - Baza

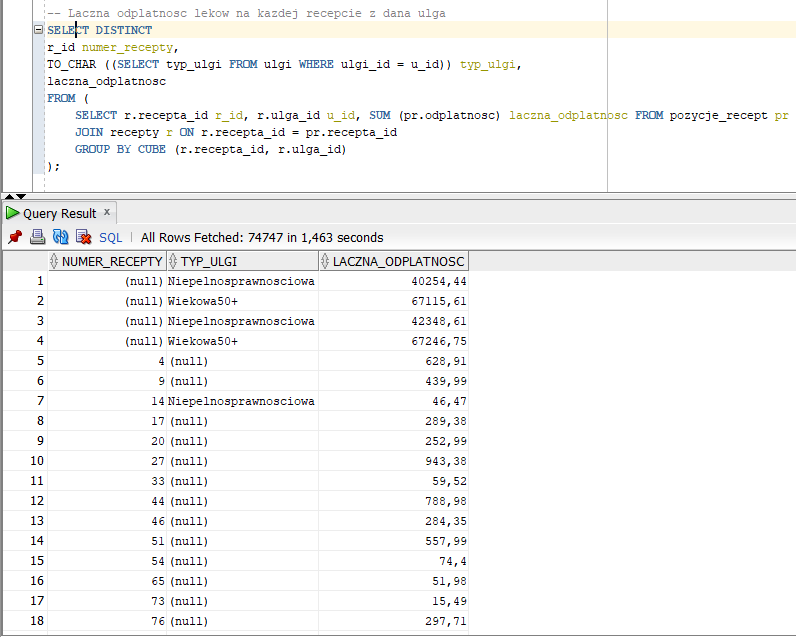


* + - Hurtownia

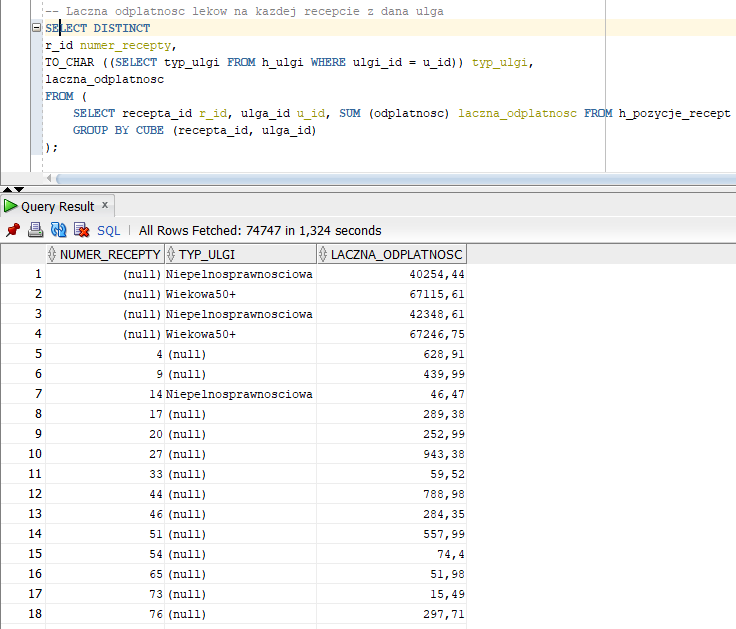


Porównując te dwa zapytania widzimy ,że wersja dla hurtowni jest prostsza i posiada o 2 złączenia tabel mniej i zlicza średnią opłat prosto z tabeli faktów. Dodatkowo wykonuje się o 0,266s szybciej niż zapytanie do bazy danych

* + Cube nr3.
    - Baza

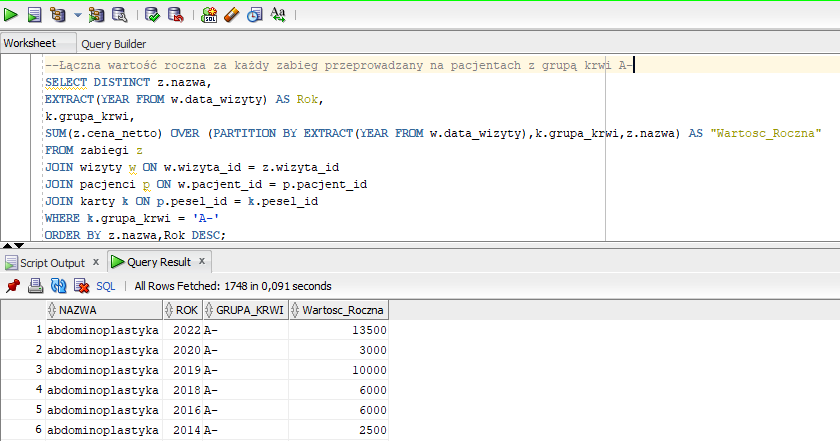


* + - Hurtownia

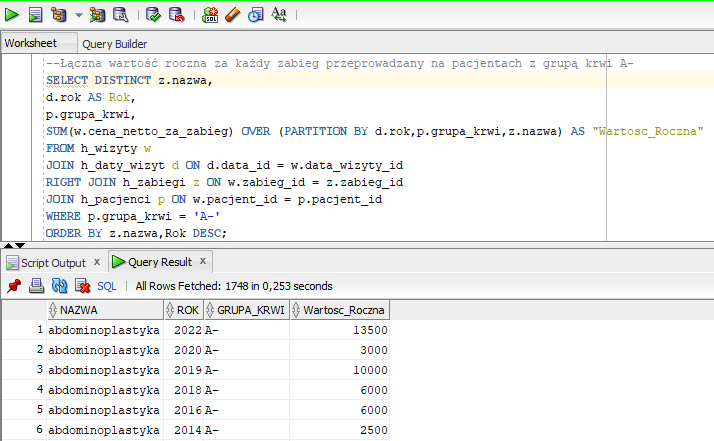


Różnicą między zapytaniem z bazy, a zapytaniem z hurtowni jest to, że w bazie wymaga ono o jedno więcej połączenie JOIN. W hurtowni to zapytanie wykonuje się o 0,139s szybciej.

* + Partycje Obliczeniowe nr1.
    - Baza

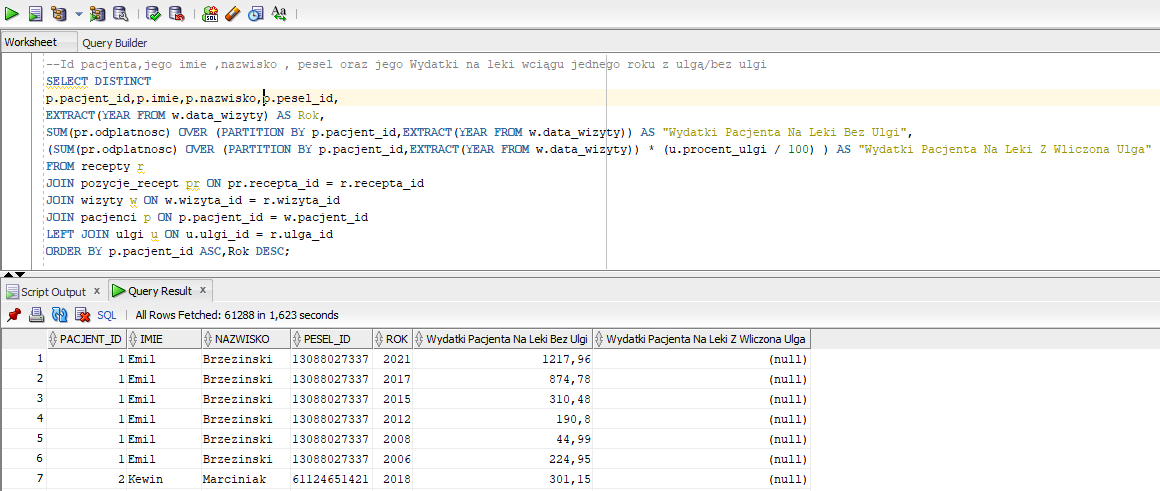


* + - Hurtownia

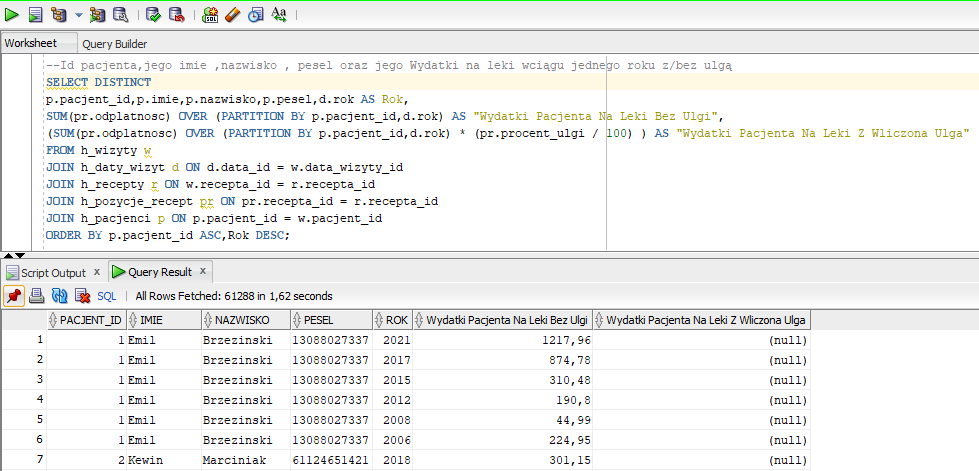


Te dwie wersje zapytania praktycznie nie różnią się złożonością , lecz różnią się tym skąd dane są wyciągane z tabel, np.: w hurtowni Suma za zabiegi jest liczona z danych zawartych w tabeli faktów a partycjonowanie jest w pierwszej kolejności po polu rok z tabeli h\_daty\_wizyt a nie jak w wersji dla zwykłej bazy ,gdzie partycjonowane po roku ,który jest wyłuskiwany z pola typu Date z wizyt. Końcowo zapytanie w hurtowni wykonuje się 178% wolniej niż w zwykłej bazie.

* + Partycje Obliczeniowe nr2.
    - Baza

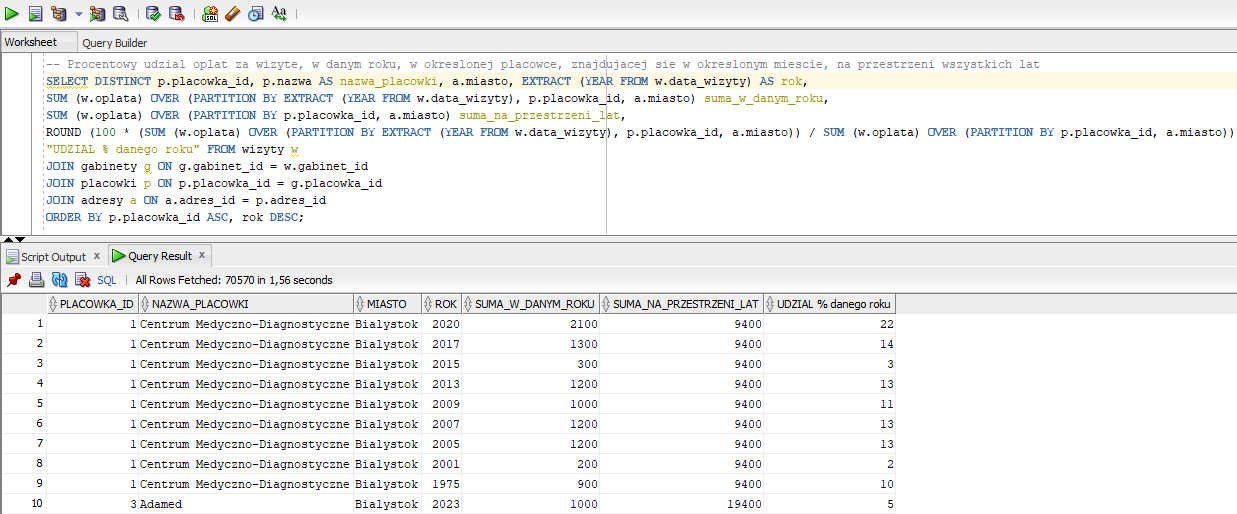


* + - Hurtownia

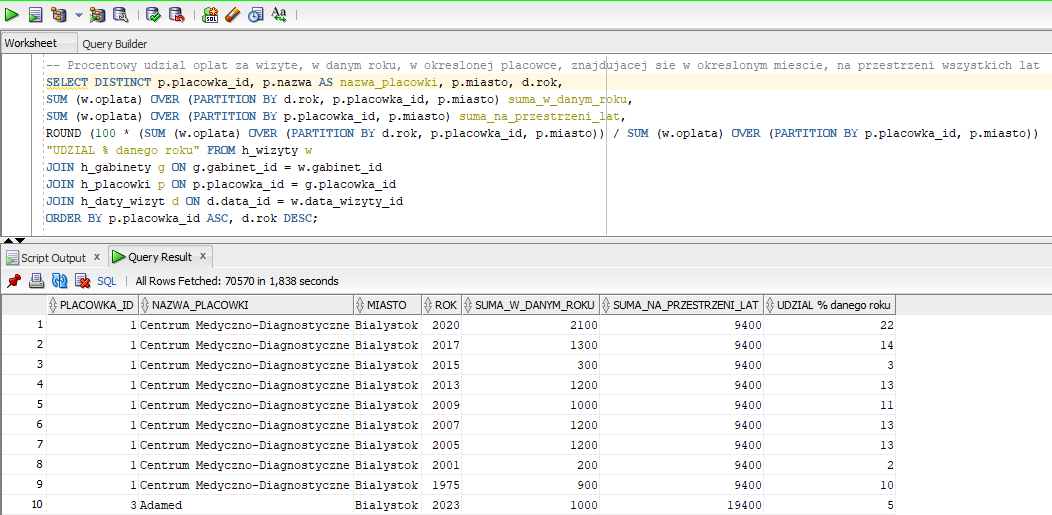


Zapytanie to w obu wersjach praktycznie nie różni się czasem wykonania, zawiera tyle samo JOIN’ów ,jedynie czym się różni wersja hurtowni od zwykłej jest prostota wyciągania kolumn po ,których się wyciąga dane i partycjonuje sumowanie.

* + Partycje Obliczeniowe nr3.
    - Baza

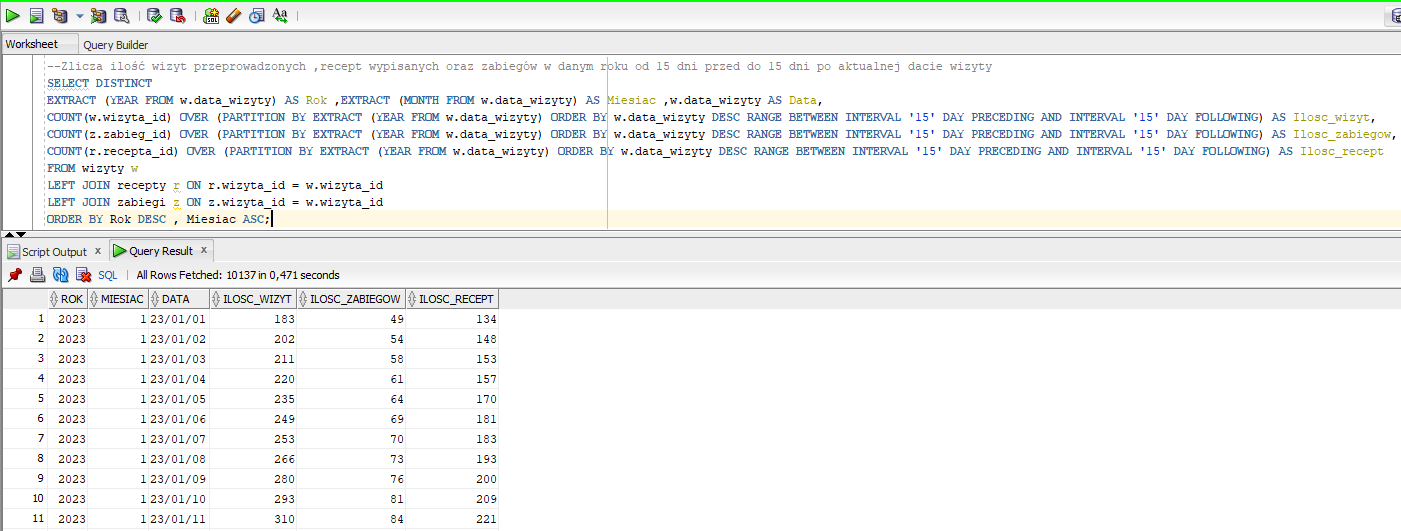


* + - Hurtownia

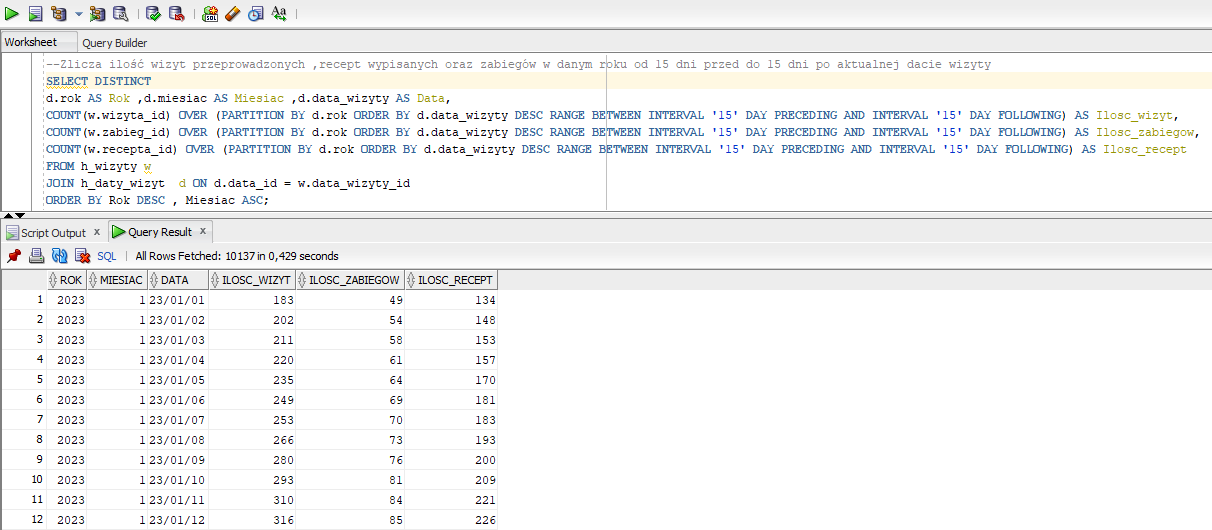


Te dwie wersje zapytania posiadają tyle samo złączeń tabel, różnią się tym ,że wersja hurtowni ma prostszy zapis lecz wykonuje się prawie 18% dłużej.

* + Okna Czasowe nr1.
    - Baza

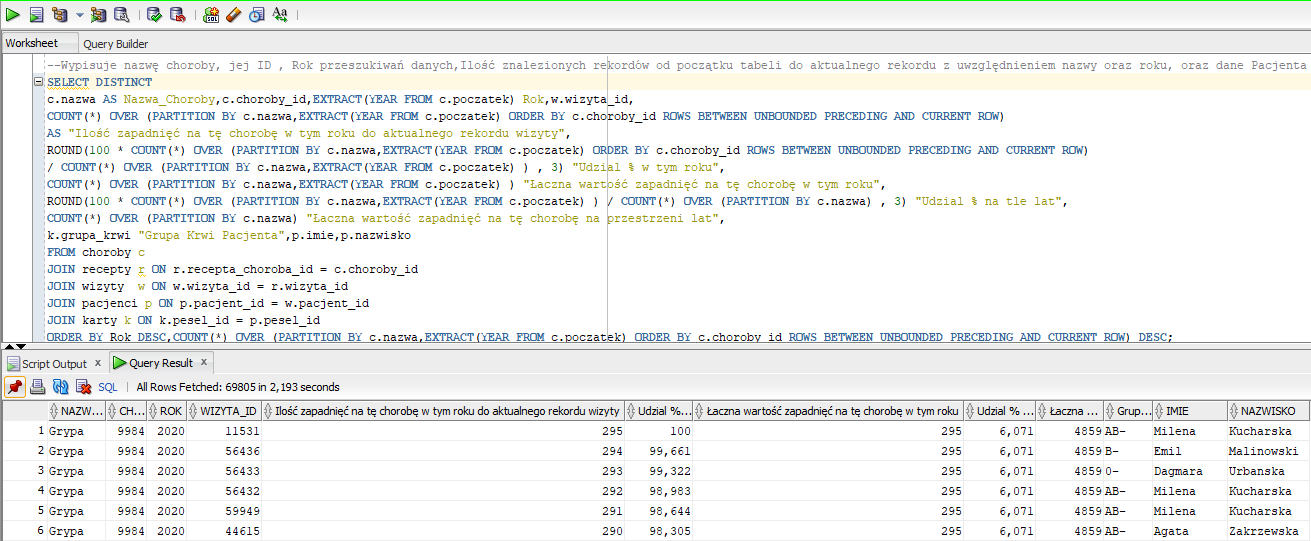


* + - Hurtownia

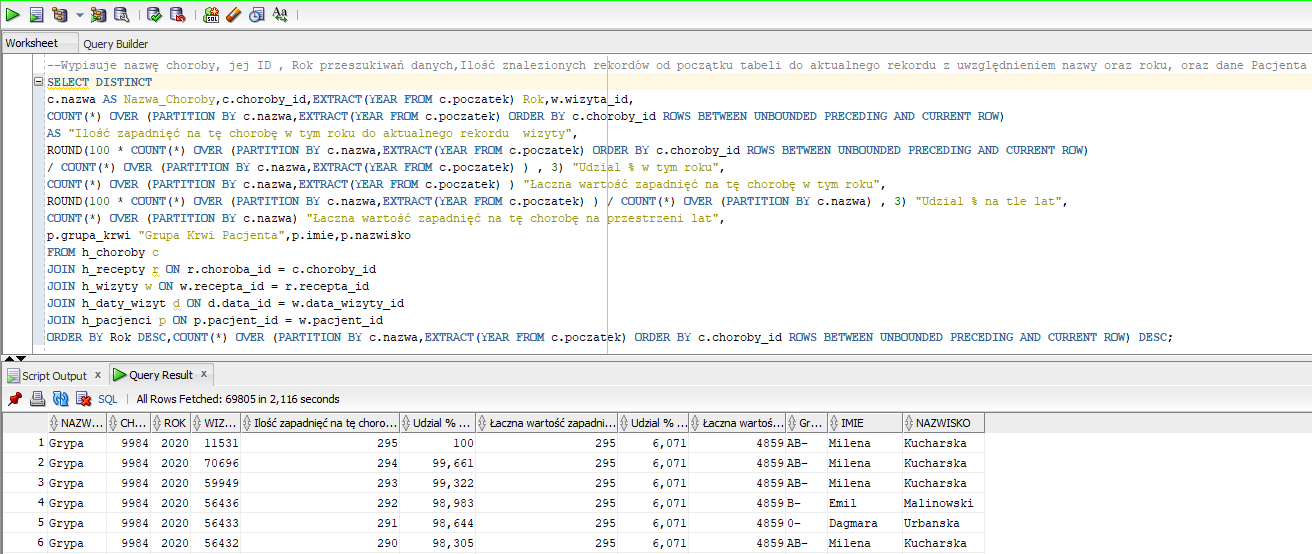


Te dwie wersje zapytań różnią się tym ,że zapytanie w hurtowni wymaga o 1 złączenie mniej i mniej zabiegów przy wyciąganiu danych, ponieważ nie potrzeba wyłuskiwać z Daty rok i miesiąc, także zapytanie dla hurtowni wykonuje się 9% szybciej.

* + Okna Czasowe nr2.
    - Baza

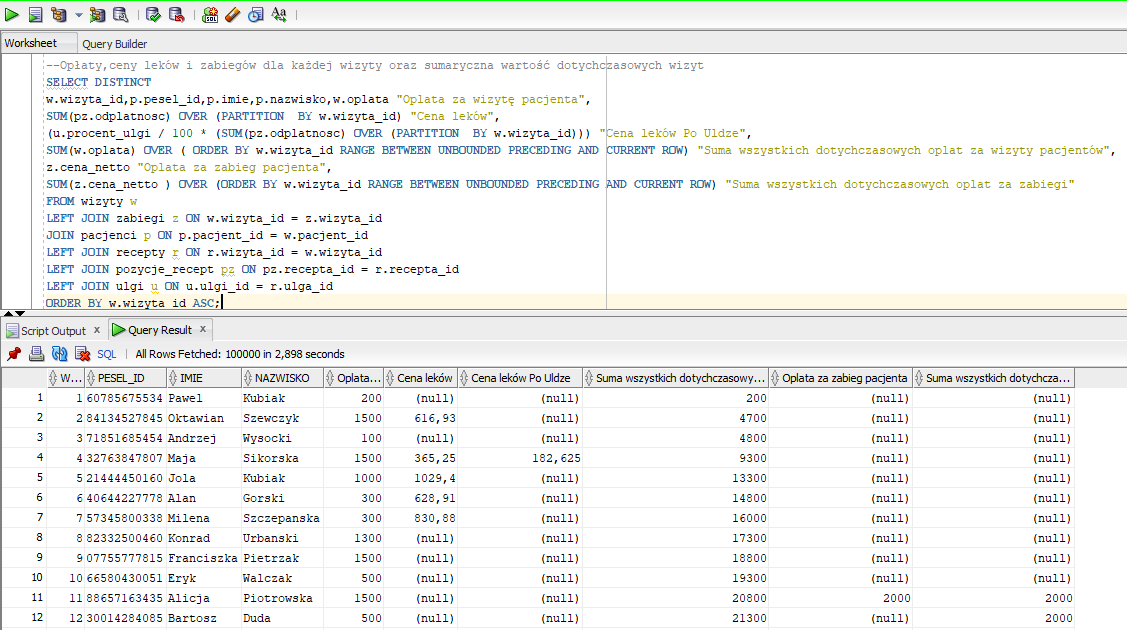


* + - Hurtownia

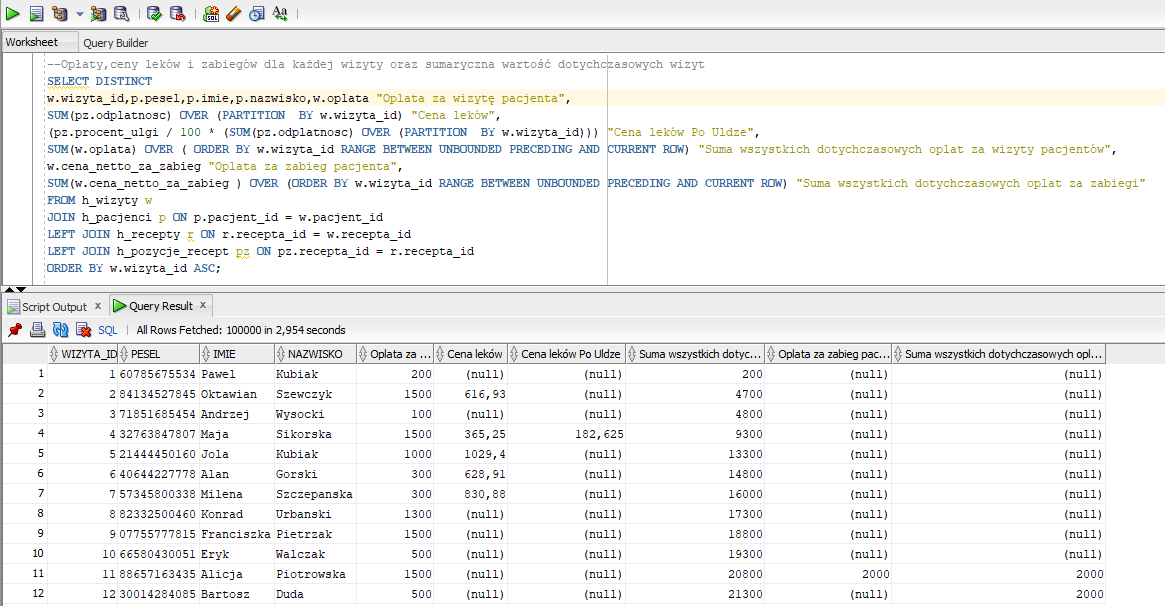


Te dwa zapytania są praktycznie identyczne pod względem złożoności zapytania ,natomiast wersja dla hurtowni wykonuje się około 4.5% szybciej od wersji dla bazy.

* + Okna Czasowe nr3.
    - Baza

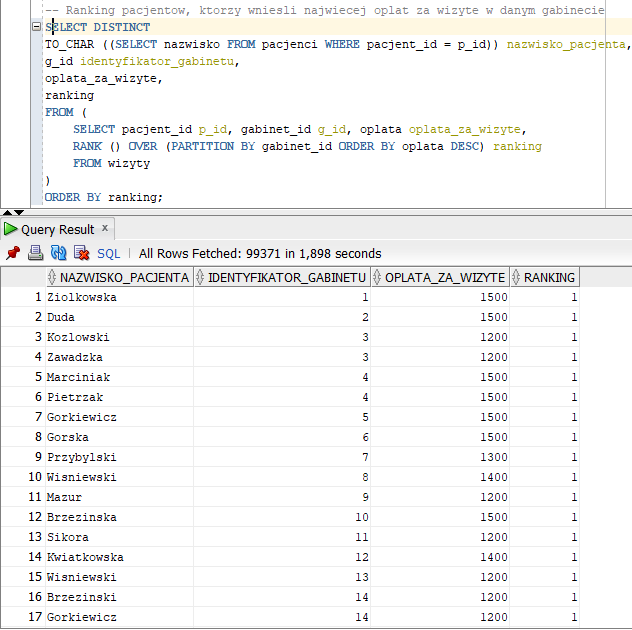


* + - Hurtownia

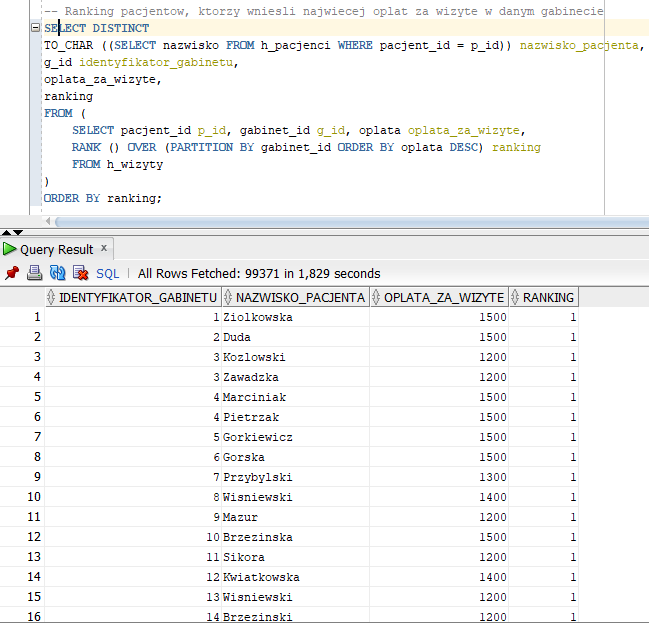


Te dwie wersje zapytań różnią się tym, że baza posiada o 2 złączenia więcej ,natomiast hurtownia wykonuje się prawie 2% wolniej.

* + Funkcje Rankingowe nr1.
    - Baza

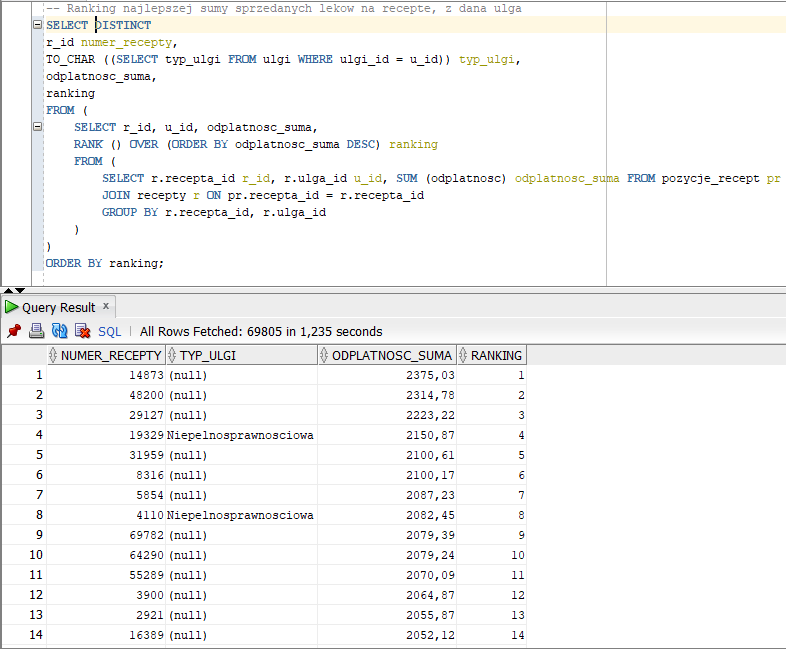


* + - Hurtownia

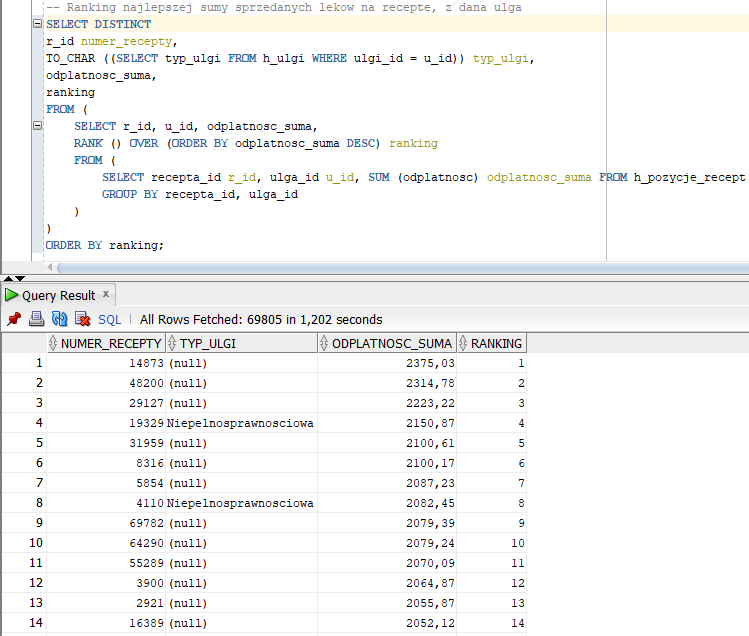


Struktura zapytania dla hurtowni jak i dla bazy jest taka sama, mimo to zapytanie do hurtowni danych wykonało się o 0,069s szybciej

* + Funkcje Rankingowe nr2.
    - Baza

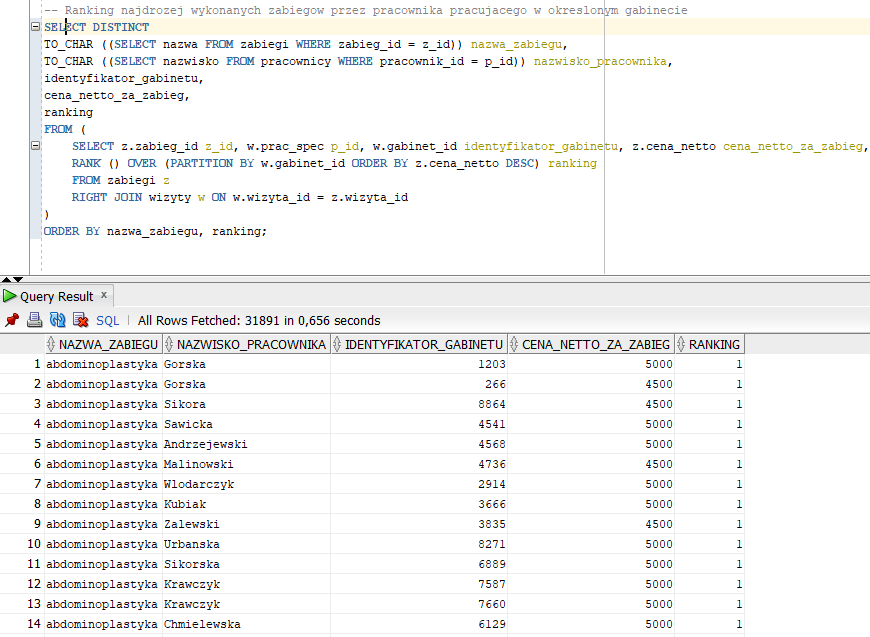


* + - Hurtownia

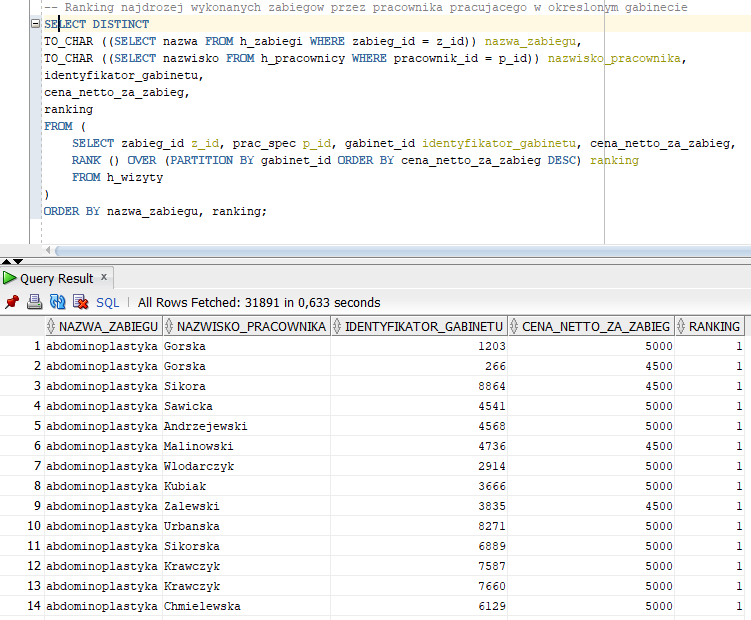


Różnicą między zapytaniem z bazy, a zapytaniem z hurtowni jest to ,że w bazie wymaga ono o jedno więcej połączenie JOIN. W hurtowni to zapytanie wykonuje się o 0,033s szybciej.

* + Funkcje Rankingowe nr3.
    - Baza



* + - Hurtownia



Różnicą między zapytaniem z bazy, a zapytaniem z hurtowni jest to ,że w bazie wymaga ono o jedno więcej połączenie JOIN. W hurtowni to zapytanie wykonuje się o 0,023s szybciej.

1. Uwagi i wnioski
   * Zdecydowaliśmy się na schemat konstelacji z dwoma tabelami faktów, ponieważ w ten sposób możemy osobno badać fakty wizyt i pozycji recept (leków), co przekłada się na wydajniejsze operacje ,kiedy badamy tylko konkretną tabelę faktów.
   * Wszystkie zapytania do hurtowni danych wykonują się szybciej niż do bazy danych. Największe różnice były widoczne przy zapytaniach z operatorem CUBE.
   * Struktura zapytań do hurtowni danych nie posiada żadnych złączeń typu JOIN, gdyż nie były one konieczne.
   * Czas trwania procedur transformacyjnych jest tak duży, ponieważ w naszej bazie danych znajduje się bardzo duża ilość rekordów w tabelach , i z tego powodu system potrzebuje więcej czasu na przerobienie danych i wstawienie do tabel w hurtowni.
   * Do wygenerowania tak dużej ilości rekordów użyliśmy własnoręcznie napisanego programu generującego w języku Java . Znaczna ilość danych w tabelach jest powiązana dosyć logicznie poprzez generator , niemniej jednak spora liczba mniej istotnych danych jest losowa.