|  |  |
| --- | --- |
| **Bazy Danych 2 - projekt** | |
| **Politechnika Świętokrzyska** | |
| **Prowadzący** | dr. inż. Jarosław Wikarek |
| **Autorzy** | Adam Furmanek  Tomasz Gębski |
| **Grupa** | 2ID12B |

**1. Opis problemu**

Dla tematu „Sklep muzyczny” należy:

1. Zaprojektować schemat bazy danych (3 postać normalna, więzy typu check). Baza danych powinna zawierać minimum 10 tabel podstawowych.

2. Zasilić bazę danymi (tabele podstawowe po około 100 rekordów, Przynajmniej 1 tabela będąca wynikiem rozbicia relacji wiele do wielu 10 000 rekordów.

3. Napisać przykładowe zapytania raportujące dane z bazy.

4. W oparciu o opracowanych schemat bazy danych opracować projekt hurtowni danych.

5. Zasilić hurtownię danymi (poprzez pliki tekstowe csv).

6. Napisać po 5 zapytań dla operatorów:

a) ROLLUP;

b) CUBE;

c) Partycje obliczeniowe;

d) Okna;

e) Funkcje rankingowe;

7. Opracować w formie elektronicznej sprawozdanie zawierające:

a) Dane zespołu;

b) Schemat bazy danych oraz schemat hurtowni (schematy ERD);

c) Skrypty zakładające bazę oraz hurtownie;

d) Skrypty umożliwiające załadowanie danych do bazy

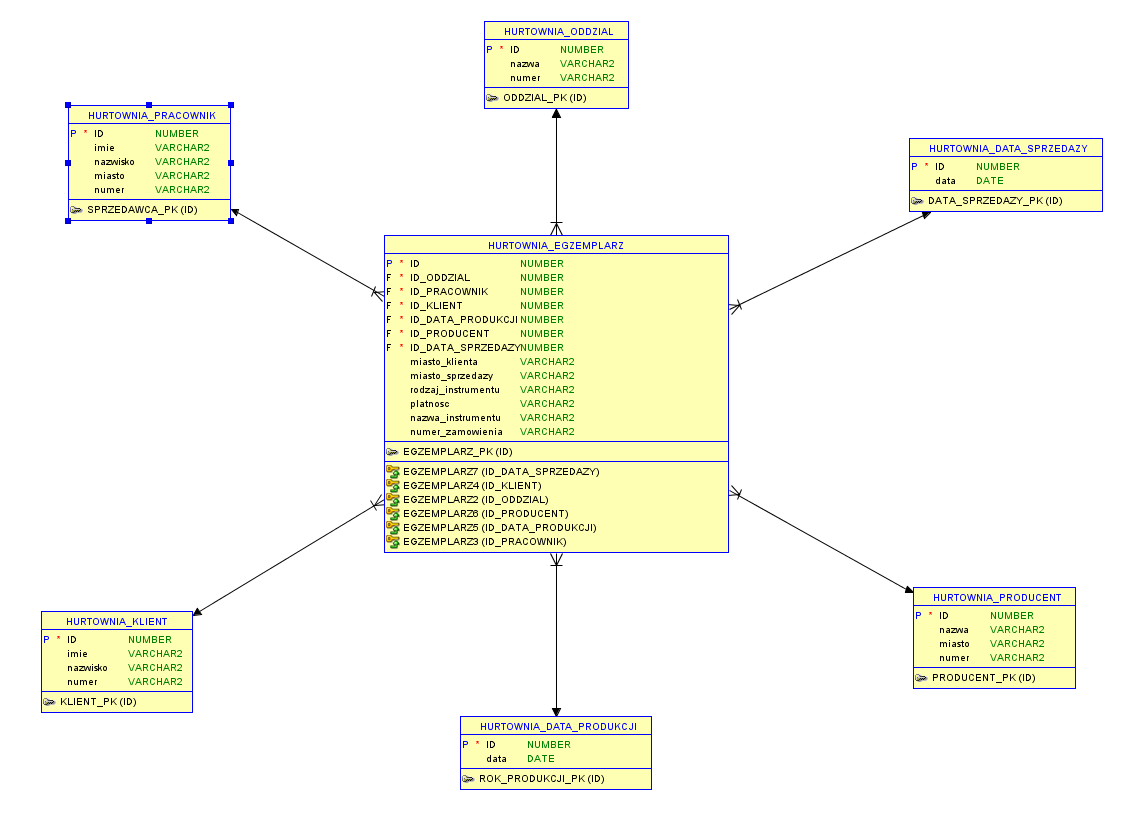
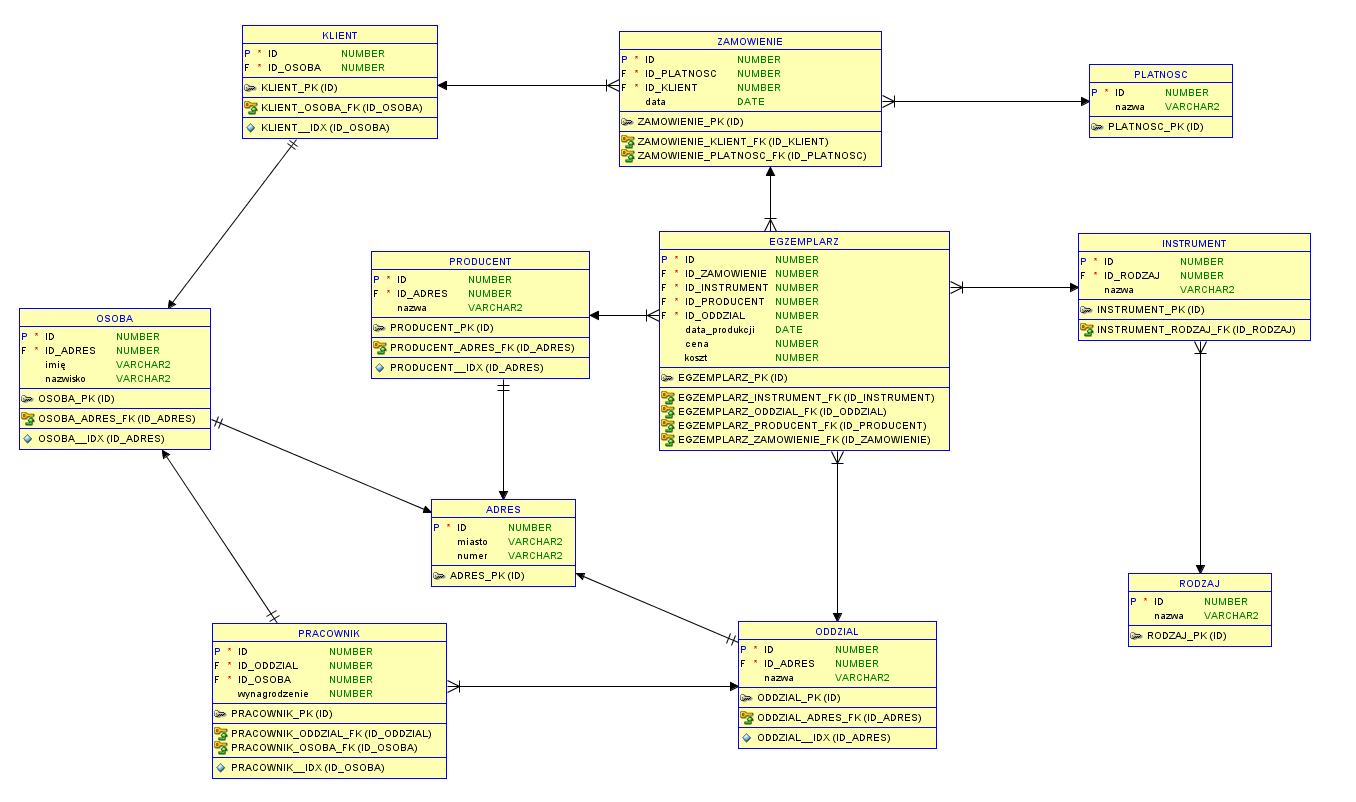
e) wyeksportowanie danych z bazy, zasilenie hurtowni.

f) Pliki csv z wyeksportowanymi danymi

g) Napisane zapytania (wraz z wynikami)

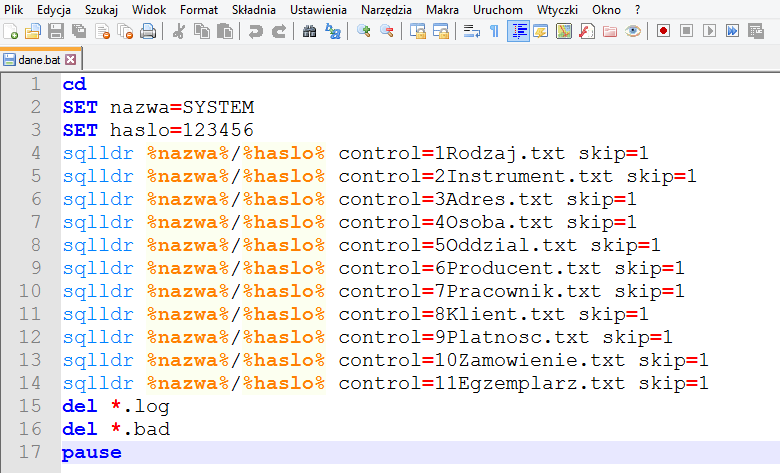
**2. Schemat ERD bazy danych oraz schemat hurtowni**

Schematy zostały stworzone przy użyciu programu Oracle DataModeler.

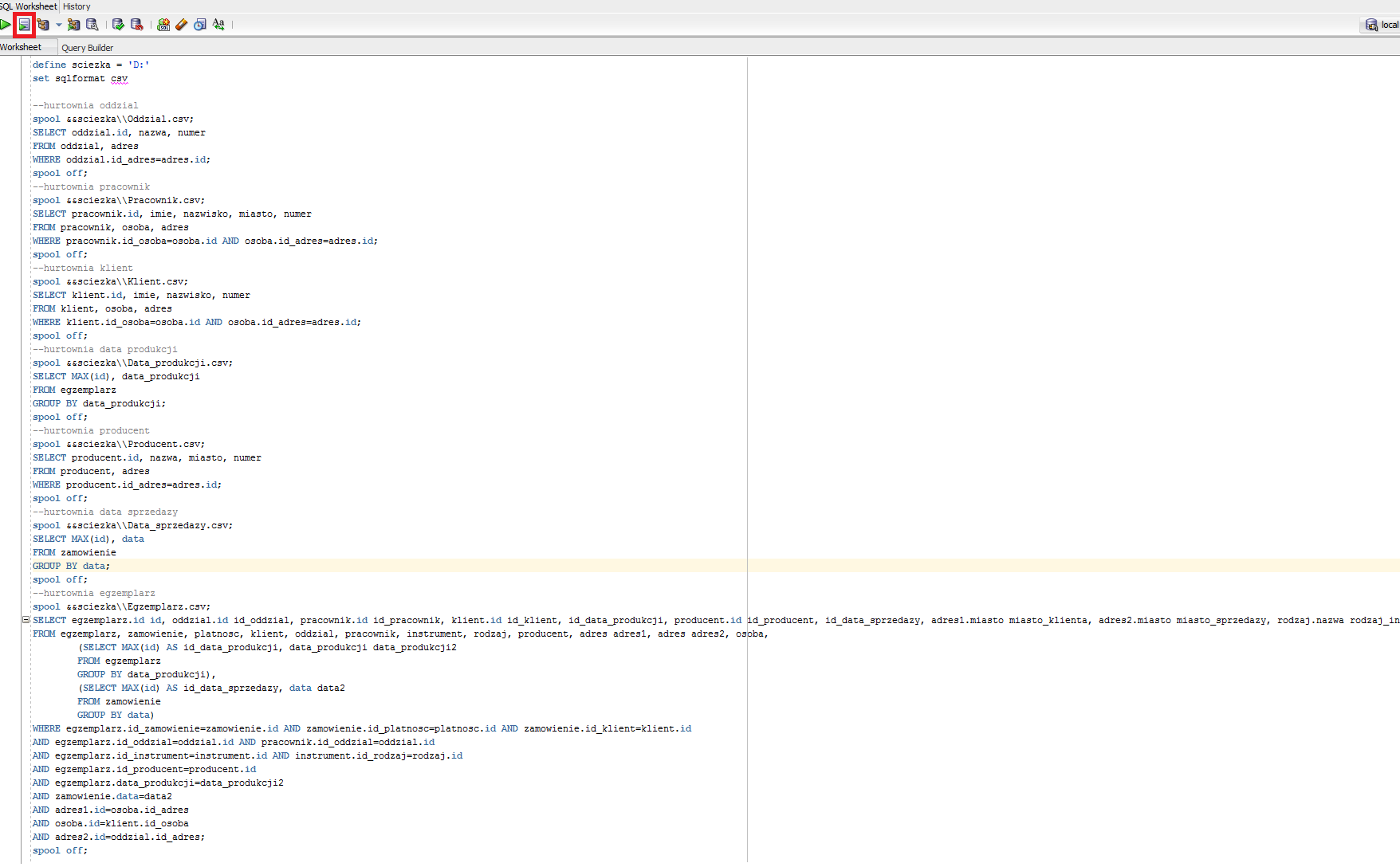


**3. Proces ładowania i transformacji danych**

Po utworzeniu bazy oraz hurtowni danych należy załadować do nich dane. Najpierw robi się to dla bazy przy pomocy SQL Loadera używając pliku wsadowego dane.bat, w którym można ustawić nazwę oraz hasło bazy przed uruchomieniem. Należy jedynie ustawić odpowiednie wartości dla „nazwa” i „haslo”.



Następnie dane z bazy trzeba odpowiednio przekonwertować i wyprowadzić do plików csv dla hurtowni. Służy do tego skrypt „Transformacja.sql”, który tworzy pliki csv w wybranym miejscu przez użytkownika. Należy użyć tutaj funkcji „Run Script” w programie Oracle SQL Developer.



Utworzone pliki csv można teraz wprowadzić do hurtowni, podobnie jak w przypadku bazy, używając dane.bat. Tutaj również można wprowadzić odpowiednią nazwę oraz hasło dla bazy.

**4. Porównanie zapytań**

Tabele przedstawiają wyniki porównań zapytań bazy i hurtowni. W obu przypadkach wynik jest takim sam, choć kolejność rekordów może być różna. Jeśli w kolumnie liczba wyszukowanich tabel dodawane są dwie liczby, oznacza to, że kwerenda korzysta z podzapytania.

**Rollup**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Opis zapytania** | **Porównanie czasów** | **Liczba przeszukiwanych tabel (wpływa na złożoność)** | | **Przystępność/czytelność kodu** |
| **Hurtownia** | **Baza** |
| Ile sprzedano jakich instrumentów i jakiego rodzaju. | Około 20ms na korzyść hurtowni. | 1 | 3 | Prostszy kod w zapytaniu hurtowni. |
| Ile sprzedano instrumentów, w jakim mieście i klientom jakich miast. | Około 50ms na korzyść hurtowni. | 1 | 7 | Bardzo prosty kod w zapytaniu hurtowni i bardzo trudny w zapytaniu bazy. |
| Ile sprzedano, w jakim roku i w jakim miesiącu. | Zbliżone czasy | 2 | 2 | Kody tej samej trudności. |
| Ile sprzedano, w jakim oddziale. | Zbliżone czasy | 2 | 2 | Kody tej samej trudności. |
| Ile wydano na instrumenty od jakiego producenta. | Zbliżone czasy | 2 | 2 | Kody tej samej trudności. |

**Cube**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Opis zapytania** | **Porównanie czasów** | **Liczba przeszukiwanych tabel (wpływa na złożoność)** | | **Przystępność/czytelność kodu** |
| **Hurtowna** | **Baza** |
| Ile kupił, jaki klient. | Około 50ms na korzyść hurtowni. | 1+1 | 4+1 | Dużo rostszy kod w zapytaniu hurtowni. |
| Ile sprzedano, przy jakim zamówieniu. | Zbliżone czasy | 1 | 2 | Prostszy kod w zapytaniu hurtowni. |
| Ile sprzedano, w jakich miesiącach i jakich dniach. | Zbliżone czasy | 2 | 2 | Kody tej samej trudności. |
| Ile kupił klient, z jakiego miasta i jakiej płatności używał. | Około 30ms na korzyść hurtowni. | 1 | 6 | Bardzo prosty kod w zapytaniu hurtowni i bardzo trudny w zapytaniu bazy. |
| Ile wydał klient, używając jakiej płatności z jakiego jest miasta. | Około 30ms na korzyść hurtowni. | 1 | 6 | Bardzo prosty kod w zapytaniu hurtowni i bardzo trudny w zapytaniu bazy. |

**Partycje obliczeniowe**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Opis zapytania** | **Porównanie czasów** | **Liczba przeszukiwanych tabel (wpływa na złożoność)** | | **Przystępność/czytelność kodu** |
| **Hurtownia** | **Baza** |
| Pokazuje jaki odsetek konkretnego rodzaju stanowi konkretny instrument. | Około 10ms na korzyść hurtowni | 1 | 3 | Prostszy kod w zapytaniu hurtowni. |
| Pokazuje procentowy udział miast klientów, którzy kupowali w miastach oddziałów. | Około 90ms na korzyść hurtowni | 1 | 7 | Bardzo prosty kod w zapytaniu hurtowni i bardzo trudny w zapytaniu bazy. |
| Pokazuje procentowy udział miesięcy w sprzedaży rocznej. | Zbliżone czasy | 2 | 2 | Kody tej samej trudności. |
| Pokazuje procentowy udział miast dla danych płatności. | Około 50ms na korzyść hurtowni. | 1 | 5 | Dużo rostszy kod w zapytaniu hurtowni. |
| Pokazuje procentowy udział płatności dla danych miast. | Około 40ms na korzyść hurtowni. | 1 | 5 | Dużo rostszy kod w zapytaniu hurtowni. |

**Ruchome okno obliczeniowe**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Opis zapytania** | **Porównanie czasów** | **Liczba przeszukiwanych tabel (wpływa na złożoność)** | | **Przystępność/czytelność kodu** |
| **Hurtownia** | **Baza** |
| Miesięczne zestawienie zysku sprzedanych instrumentów licząc od początku roku. | Zbliżone czasy | 2 | 2 | Kody tej samej trudności. |
| Miesięczne zestawienie ilości sprzedanych instrumentów licząc od początku roku. | Zbliżone czasy | 2 | 2 | Kody tej samej trudności. |
| Miesięczne zestawienie sprzedanych instrumentów w oddziałach licząc od początku roku. | Zbliżone czasy | 3 | 3 | Kody tej samej trudności. |
| Roczne zestawienie produkcji instrumentów od danych producentów od początku istenienia. | Zbliżone czasy | 3 | 3 | Kody tej samej trudności. |
| Roczne zestawienie zakupów klientów licząc od początku istnienia sklepu. | Zbliżone czasy | 3+1 | 3+3 | Prostszy kod w zapytaniu hurtowni. |

**Funkcje rankingowe**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Opis zapytania** | **Porównanie czasów** | **Liczba przeszukiwanych tabel (wpływa na złożoność)** | | **Przystępność/czytelność kodu** |
| **Hurtownia** | **Baza** |
| Ranking klientów według ilości kupionych instrumentów. | Około 20ms na korzyść hurtowni | 1+1 | 3+3 | Dużo rostszy kod w zapytaniu hurtowni. |
| Ranking sprzedawców według ilości sprzedanych instrumentów. | Około 20ms na korzyść hurtowni | 1+1 | 3+3 | Dużo rostszy kod w zapytaniu hurtowni. |
| Ranking instrumentów według ich sprzedanej ilości. | Zbliżone czasy | 1 | 2 | Prostszy kod w zapytaniu hurtowni. |
| Ranking metody płatności według ilości ich użyć. | Około 10ms na korzyść hurtowni | 1 | 2 | Prostszy kod w zapytaniu hurtowni. |
| Ranking oddziałów według ilości sprzedanych instrumentów. | Zbliżone czasy | 1+1 | 2+2 | Prostszy kod w zapytaniu hurtowni. |

**5. Uwagi i wnioski**

- z powodu długich nazw w hurtowni (w nazwie każdej tabeli dodany jest przedrostek „hurtownia\_”) jej zapytania wydają się nieczytelne, choć w rzeczywistości są one łatwiejsze do napisania i odczytania niż zapytania dla bazy;

- ponieważ dane w bazie i hurtowni są losowe, nie oddają one dokładnie rzeczywistych warunków. Przykładowo: 50 klientów kupiło łącznie 10000 instrumentów, a robili to jeżdząc po całej Polsce, nawet kiedy mieli oddział firmy w swoim rodzinnym mieście.

- podczas testowania czasów wykonania zapytań zauważono, że przy ponownym uruchomienie tej samej kwerendy, czas wykonania mocno spada, dlatego pomiary były wykonywane kilka razy z rzędu lub naprzemiennie z jej drugim odpowiednikiem. Oba sposoby pokazywały podobne różnice w czasach między bazą a hurtownią.

- prównywanie czasów było przeprowadzane dla wszystkich wierszy wybieranych przez zapytania, przy użyciu skrótu klawiszowego ctrl+end, który przechodził do ostatniego wiersza wyniku.

- hurtownia danych jest bardzo dobra do analizowania ogromnych ilości danych, gdyż nawet na tak małym przykładzie różnice w czasie obliczania i złożoności zapytań są widoczne.