

[ZBD] Week 2 - PostgreSQL database

Zacharczuk Jakub

jz418488

1 A brief justification why the actual design has been chosen.

Zdecydowałem się stworzyć dwie dodatkowe tabele: **country** oraz **jurisdiction**. Zauważyłem, że kolumna **country_codes** może jednoznacznie definiować wartość kolumny **countries**. W tym celu napisałem skrypt w Pythonie załączony w folderze w pliku **analise.py** oraz uruchamiany z odpowiednimi argumentami w **main.py**, analizujący podany w ścieżce plik csv, pod kątem średniej liczby bajtów elementów w kolumnie, medianie, najkrótszej i najdłuższej liczbie znaków w kolumnie. Pomogło mi to dobrać odpowiednie typy i wielkości pól. Wyniki działania skryptu znajdują się w pliku **statistics**. Dodatkowo funkcja **are_columns_dependant()** sprawdziła czy dla pewnego **country_code** w danej tabeli, znajdują się różne wartości **countries**. Analogicznie dla **jurisdiction**. Okazało się że kod kraju/jurysdykcji jednoznacznie definiuje jej opis, zatem możemy wydzielić tę kolumnę do nowej tabeli oszczędzając pamięć.

2 Zapytanie o graf

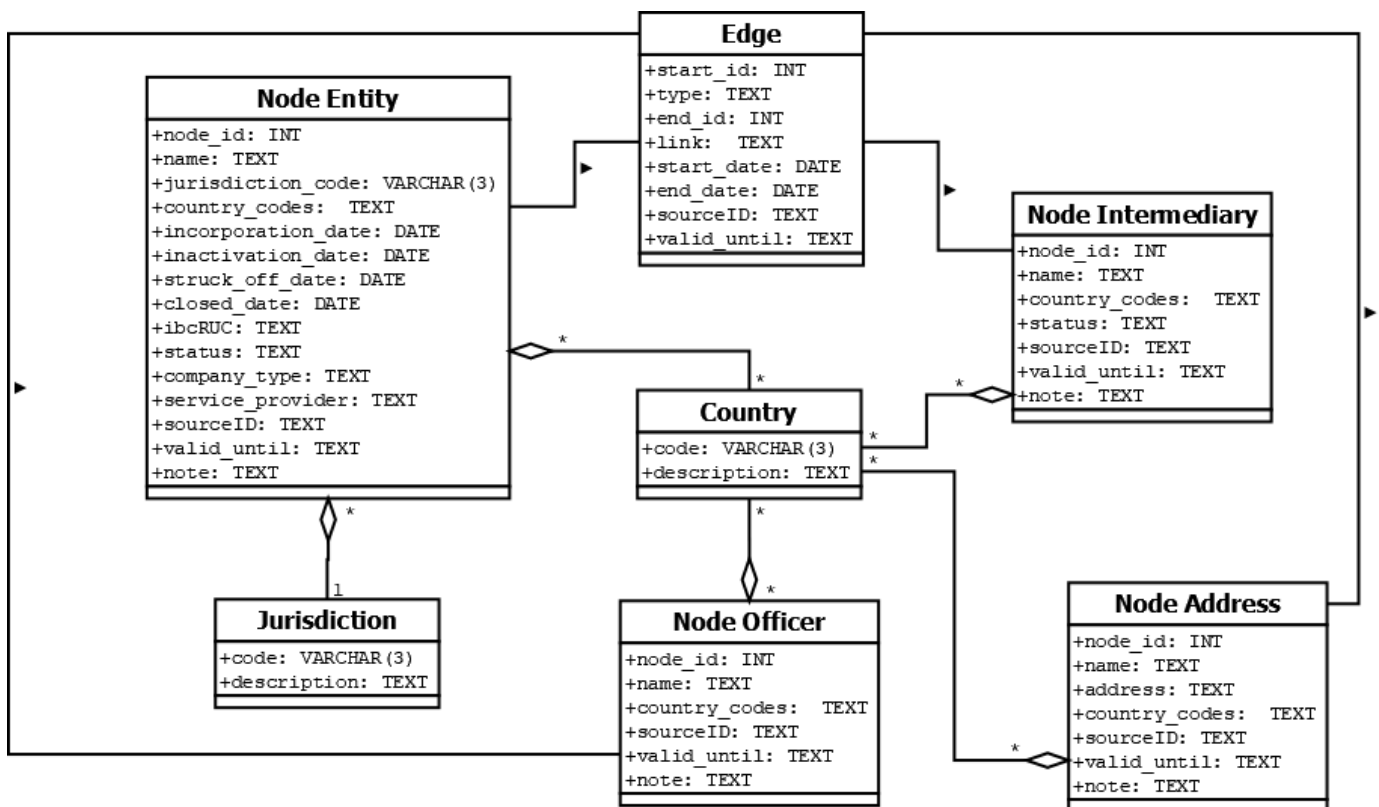


Diagram UML został wykonany przed analizą. Ostatecznie typy TEXT zostały zastąpione przez VARCHAR(n) z odpowiednią wartością.

3 A script to create the database.

Znajduje się w oddzielnym pliku **skrypt.sql**

4 A PDF report with the information on the time spent during import, the size of the database in GB, the count of rows in each table

address = 93454 rows

edge = 674102 rows

entity = 213634 rows

country = 312 rows

officer = 238402 rows

intermediary = 14110 rows

jurisdiction = 21 rows

Czas importowania:

real 0m4.874s

user 0m0.184s

sys 0m0.096s

Wielkość bazy danych: 199 MB