

**Akademia Górniczo-Hutnicza   
im. Stanisława Staszica   
w Krakowie**

**WYDAJNOŚĆ ZŁĄCZEŃ I ZAGNIEŻDŻEŃ DLA SCHEMATÓW**

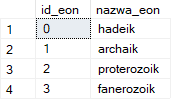
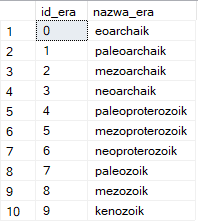
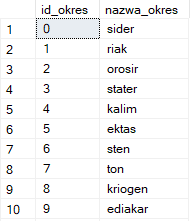
**ZNORMALIZOWANYCH I ZDENORMALIZOWANYCH**

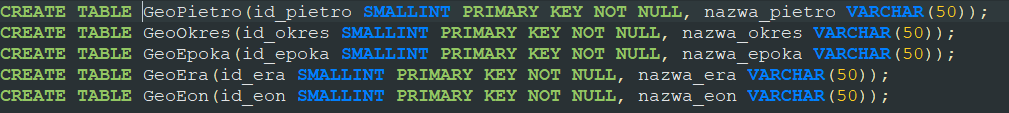
14.06.2022

Dawid Makowski (gr. lab. 2) 407894

1. **Konstrukcja wymiaru geochronologicznego.** Do realizacji zadania potrzebne są dane ilustrujące wymiar geochronologiczny, w tym celu skonstruowano schematy zdenormalizowane (jedna tabela) oraz znormalizowane w postaci kilu tabel:

* Schemat znormalizowany:

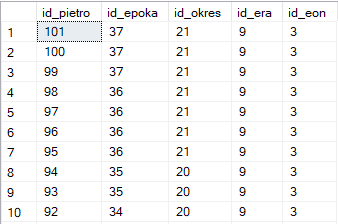
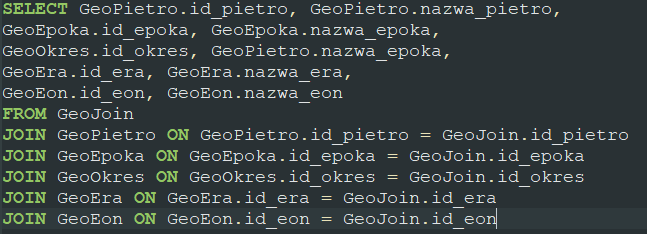




* Schemat zdenormalizowany

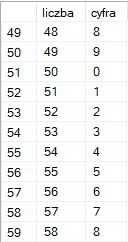
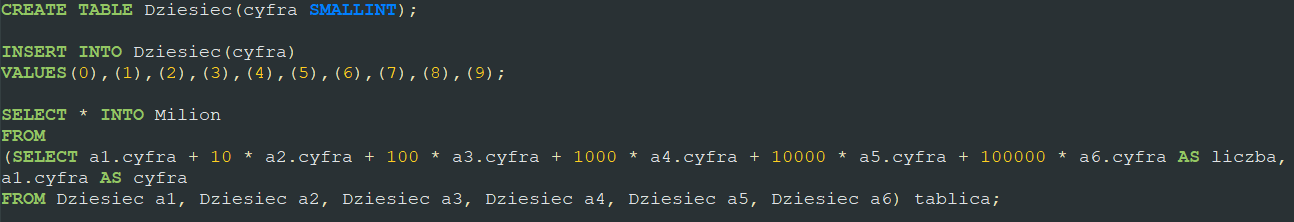


W celu skonstruowania tabeli GeoTabela, najpierw stworzono tabelę GeoJoin obrazującą powiązania między głównymi kluczami



1. **Testy wydajności.**

Do badania wydajności zastosowano dwa środowiska: PostgreSQL oraz SQL Server 2017. Utworzono również dwie tabele, ‘Dziesiec’ i ‘Milion’. Tabela ‘Dziesiec’ zawiera cyfyr 0-9, posługując się nią tworzymy tabelę ‘Milion’.



**Wszystkie testy przeprowadzono na komputerze o parametrach:**

* CPU: Intel Core i7 10750H 2,6 GHz,
* RAM: Pamięć DDR4 8GB
* S.O.: Windows 10 Pro

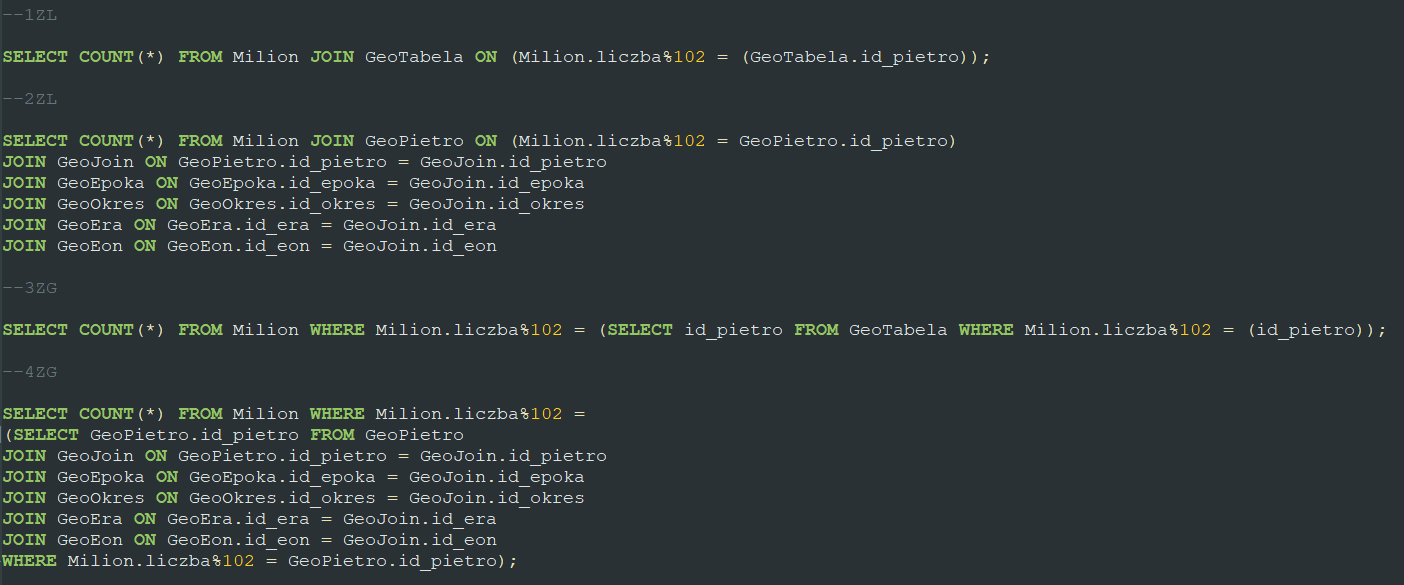
**Jako systemy zarządzania bazami danych wybrano oprogramowanie wolno dostępne:**

* SQL Server 2017
* PostgreSQL 14.4

Testy wykonano dziesięciokrotnie dla każdego kryterium z i bez indeksów.

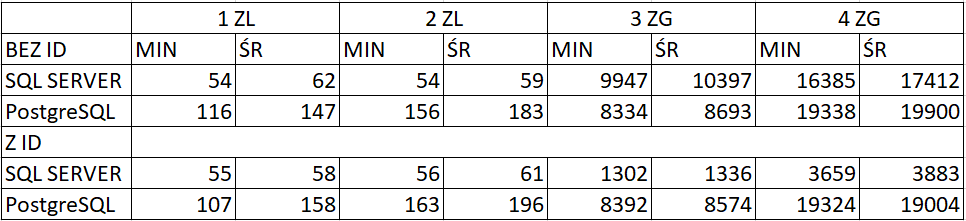
**Jako kryteria zostały wykonane cztery zapytania:**

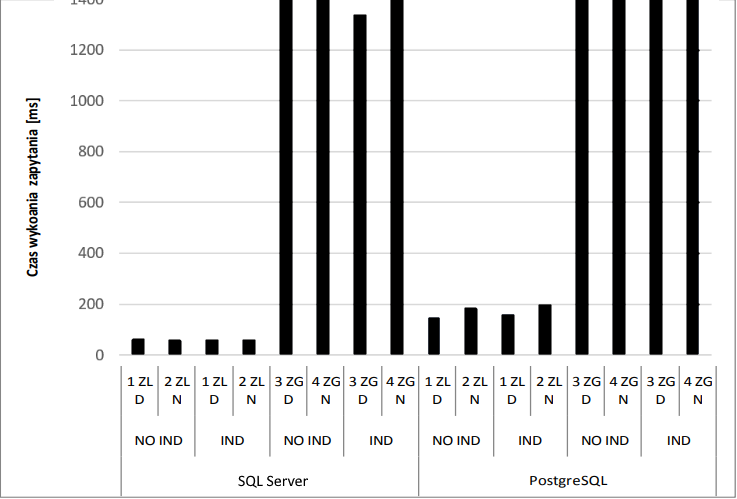
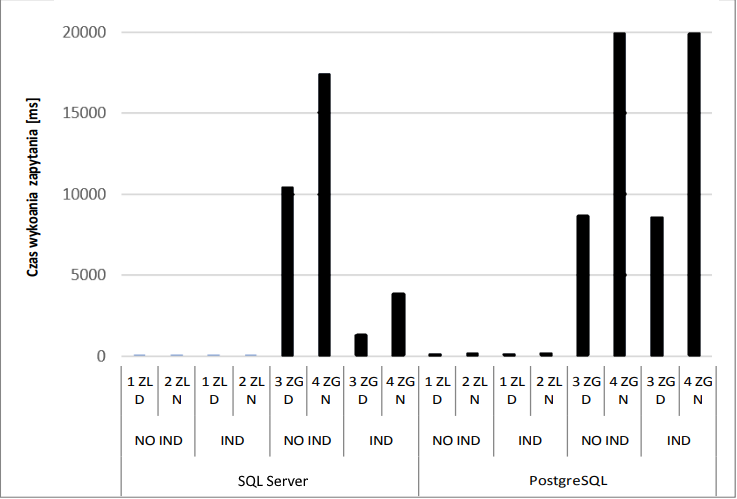
* 1ZL - złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci zdenormalizowanej, przy czym do warunku złączenia dodano operację modulo, dopasowującą zakresy wartości złączanych kolumn
* 2ZL - złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci znormalizowanej, reprezentowaną przez złączenia pięciu tabel
* 3ZG - złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci zdenormalizowanej, przy czym złączenie jest wykonywane poprzez zagnieżdżenie skorelowane
* 4ZG - złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci znormalizowanej, przy czym złączenie jest wykonywane poprzez zagnieżdżenie skorelowane, a zapytanie wewnętrzne jest złączeniem tabel poszczególnych jednostek geochronologicznych



1. **Wyniki.**

Poniżej prezentowane są wyniki uzyskane w wyniku testów





1. **Podsumowanie.**

Z powyższych danych można wyciągnąć następujące wnioski:

* Uśredniając PostgreSQL spisał się gorzej niż SQL Server
* Postać zdenormalizowana jest praktycznie zawsze wydajniejsza, jedynie przy naprawdę szybkich operacjach ciężko dostrzec różnicę.
* Zagnieżdżenia skorelowane są dużo wolniejsze w wykonaniu niż złączenia, wydłużają czas działania programu do kilkunastu sekund
* Użycie indeksów w systemie SQL Server prowadzi do znacznego przyspieszenia w przypadku zagnieżdżeń skorelowanych poniżej 5 sekund, w przypadku Postgres’a indeksacja nie miała znacznego wpływu na działanie programu

Ogólnie rzecz biorąc w większości przypadków SQL Server spisuje się lepiej, jednak bez indeksacji i Postgres jest czasem lepszy, na przykład przy zagnieżdżaniu skorelowanym działa podobnie a nawet i szybciej w przypadku kryterium czwartego.

Normalizacja prowadzi, w większości, do spadku wydajności, jednak dzięki niej w bazie jest łatwiej się odnaleźć i panuje większe uporządkowanie.