

# ĆWICZENIE 9

Dominika Folta

Ćwiczenie 9 polegało na przeprowadzeniu eksperymentu sprawdzającego wydajność dla złączeń i zagnieżdżeń skorelowanych dla systemów zarządzania bazami danych SQL server i PostgreSQL. Ćwiczenie to zostało wykonane na podstawie artykułu Łukasza Jajeńnica i Adama Piórkowskiego (Akademia Górniczo–Hutnicza, Katedra Geoinformatyki i Informatyki Stosowanej) pod tytułem “WYDAJNOŚĆ ZŁĄCZEŃ I ZAGNIEŻDŻEŃ DLA SCHEMATÓW ZNORMALIZOWANYCH I ZDENORMALIZOWANYCH”.

Aby przeprowadzenie testów było możliwe, została utworzona baza danych o nazwie Tabela\_geochronologiczna. Obrazuje przebieg historii Ziemi na podstawie następstwa procesów i warstw skalnych. W tabeli geochronologicznej znalazły się jednostki geochronologiczne przedstawione w postaci tabel mające wymiar czasowy

- eon
- era
- okres
- epoka
- piętro

oraz odpowiadające im jednostki stratygraficzne. Konstrukcja tabeli geochronologicznej powstała za pomocą schematu zdenormalizowanego (schematu gwiazdy). Formę tą osiągnięto tworząc jedną tabelę o nazwie GeoTabela zawierającą wszystkie dane z powyższych tabel. Dokonano tego za pomocą złączenia naturalnego:

```
CREATE TABLE GeoTabela AS (SELECT * FROM GeoPietro NATURAL JOIN  
GeoEpoka NATURAL JOIN GeoOkres NATURAL JOIN Geo Era NATURAL JOIN  
GeoEon );
```

Utworzenie tabeli GeoTabela umożliwia szybki dostęp do wszystkich danych tabeli geochronologicznej za pomocą jednego zapytania.

Następnie należało przeprowadzić test wydajności porównujący wydajność złączeń oraz zapytań zagnieżdżonych wykonanych na tabelach o dużej ilości danych.

Została utworzona tabela o nazwie Milion, którą wypełniono kolejnymi liczbami naturalnymi od 1 do 999 999, a także tabela Dziesiec (wypełniona liczbami od 0 do 9), na podstawie której powstała tabela Milion.

## KONFIGURACJA SPRZĘTOWA

Testy zostały wykonane na komputerze o podanej konfiguracji sprzętowej i programowej:

- CPU: Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz, rdzenie 6
- RAM: Pamięć 8GB (SO-DIMM DDR4, 2666 MHz),
- SSD: 512GB
- System operacyjny: Windows 10

Jako systemy zarządzania bazami danych wybrano oprogramowanie wolno dostępne:

- SQL server 2019 wersja 18.9
- PostgreSQL, wersja 13.2

Kolejnym krokiem było przeprowadzenie zapytań, sprawdzających wydajność złączeń i zagnieżdżeń z tabelą geochronologiczną. Na początku zapytania były przeprowadzone bez indeksów nałożonych na kolumny danych, następnie zostały nałożone indeksy na wszystkie kolumny biorące udział w złączeniu. Zastosowano cztery poniższe zapytania:

- Zapytanie 1 (1 ZL)

```
SELECT COUNT(*) FROM Milion INNER JOIN GeoTabela ON  
(mod(Milion.liczba,68)=(GeoTabela.id_pietro));
```

- Zapytanie 2 (2 ZL)

-

```
SELECT COUNT(*) FROM Milion INNER JOIN GeoPietro ON  
(mod(Milion.liczba,68)=GeoPietro.id_pietro) NATURAL JOIN GeoEpoka NATURAL  
JOIN GeoOkres NATURAL JOIN GeoEra NATURAL JOIN GeoEon;
```

- Zapytanie 3 (3 ZG)

```
SELECT COUNT(*) FROM Milion WHERE mod(Milion.liczba,  
68)=(SELECT id_pietro FROM GeoTabela WHERE  
mod(Milion.liczba,68)=(id_pietro));
```

- Zapytanie 4 (4 ZG)

```
SELECT COUNT(*) FROM Milion WHERE mod(Milion.liczba,68)=  
(SELECT GeoPietro.id_pietro FROM GeoPietro NATURAL JOIN GeoEpoka  
NATURAL JOIN GeoOkres NATURAL JOIN GeoEra NATURAL JOIN GeoEon;
```

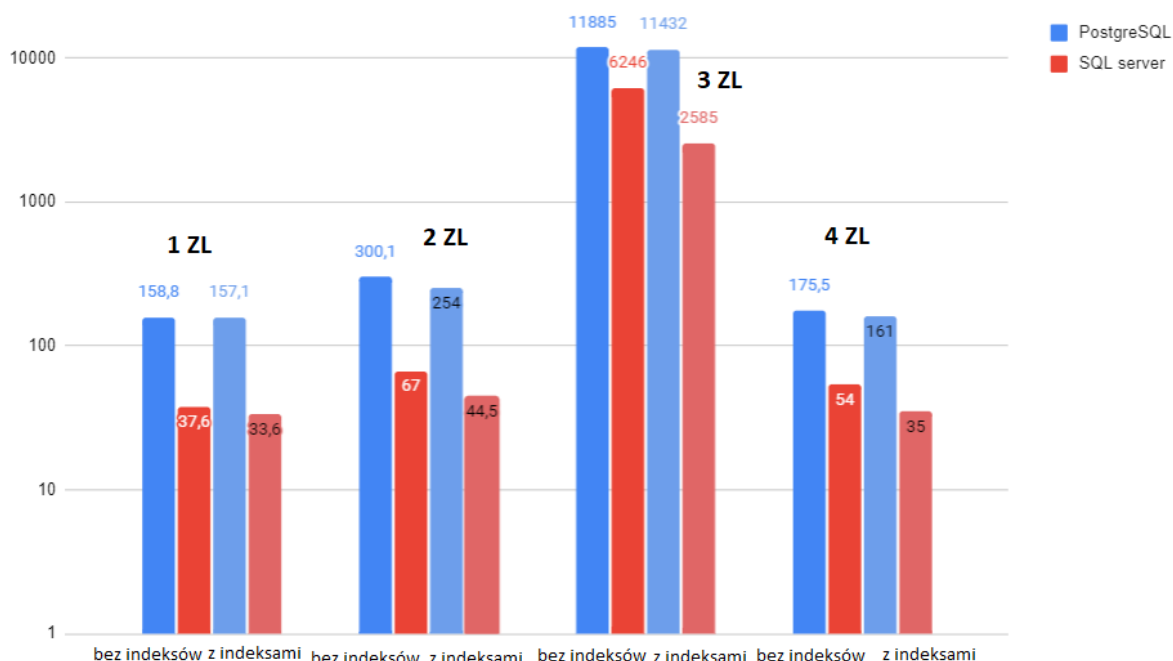
Wyniki czasów wykonywania tych zapytań zostały wpisane do tabeli.

Czasy wykonania zapytań 1 ZL, 2 ZL, 3 ZG i 4 ZG [ms]

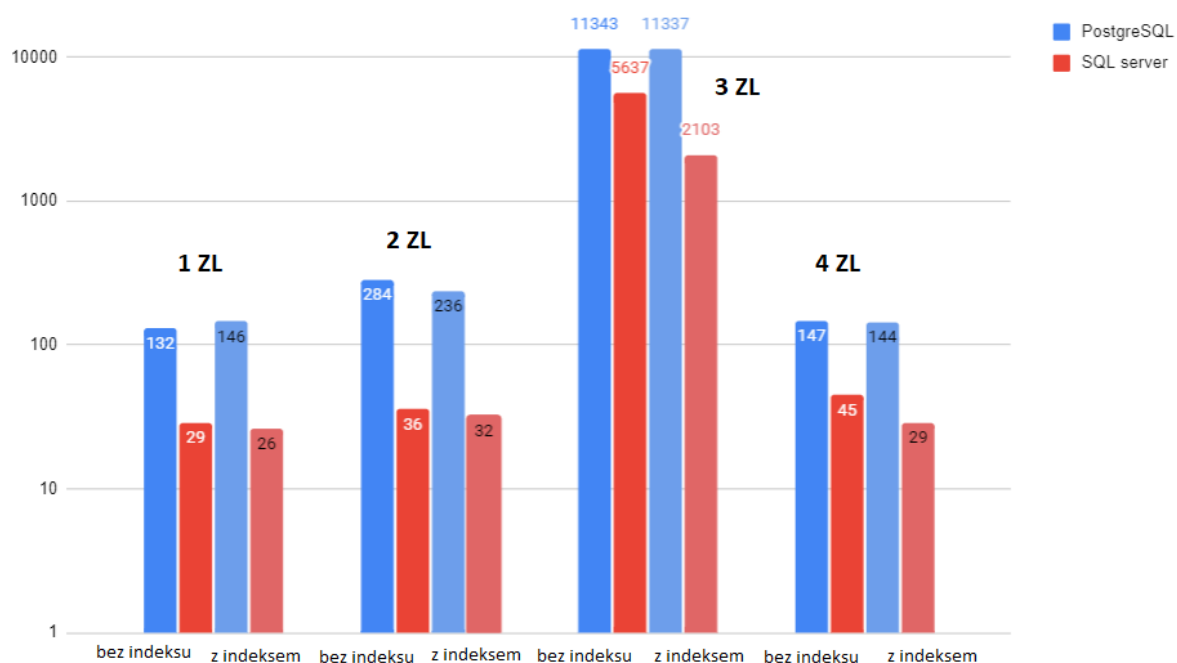
	1 ZL		2 ZL		3 ZG		4 ZG	
Bez indeksów	MIN	ŚR	MIN	ŚR	MIN	ŚR	MIN	ŚR
PostgreSQL	132	158,8	284	300,1	11343	11885	147	175,5
SQL server	29	37,6	36	67	5637	6246	45	54
Z indeksami								
PostgreSQL	128	157,1	236	254	11337	11432	144	161
SQL server	26	33,6	32	44,5	2103	2585	29	35

Można zauważyć, że wyniki czasów wykonywania zapytań są znacznie większe w Postgresie niż w SQL serverze. Dodatkowo po dodaniu indeksów czas także się zmniejszył. Największe różnice w czasie można dostrzec podczas wykonywania zapytania trzeciego w SQL serverze, ponieważ czas ten jest ponad połowę mniejszy.

Średnie czasy wykonywania dla tabeli bez indeksów i z indeksami



Minimalne czasy wykonywania dla tabeli bez indeksów i z indeksami



Podsumowując, dodanie indeksów do tabeli zmniejsza czas dodawania zapytań.