

Akademia Górniczo-Hutnicza w Krakowie
Wydział Geologii, Geofizyki i Ochrony Środowiska



*Badanie wydajności złączy i zagnieżdżeń dla schematów
znormalizowanych i zdenormalizowanych*

projekt w ramach przedmiotu

Bazy Danych

Filip Giermek

geoinformatyka

Kraków, 01.06.2021

Spis treści:

| | |
|---|---|
| Problematyka..... | 2 |
| Konfiguracja sprzętowa i systemowa..... | 2 |
| Zapytania..... | 2 |
| Przygotowanie niezbędnych tabel..... | 3 |
| Przeprowadzenie testów..... | 4 |
| Interpretacja wyników..... | 8 |
| Obserwacje i wnioski..... | 9 |

Problematyka

Poniższy projekt ma na celu odtworzenie oraz przeprowadzenie analizy wydajności złączeń tabel z indeksami i bez indeksów w tabelach znormalizowanych oraz zdenormalizowanych. Całość bazuje na artykule naukowym autorstwa Łukasza Jajeńnicy oraz Adama Piórkowskiego pt. *„Wydajność Złączeń i Zagnieżdżeń Dla Schematów Znormalizowanych i Zdenormalizowanych.”*

Całość operacji bazuje na tabeli geochronologicznej, czyli tabeli opisującej przebieg historii Ziemi, ustalonej przez Międzynarodową Komisję Stratygrafii (ICS). Pozwala ona na ujednolicenie terminologii i ułatwienie wymiany informacji na polu międzynarodowym.

Konfiguracja sprzętowa i systemowa

CPU: Intel Core i5-8250U 1.80 GHz

RAM: 16GB

HDD: WDC PC SN520 SDAPMUW-256G-1101, 1GB cache

S.O.: Windows 10, 64-bitowy

System zarządzania bazami danych:

MySQL 8.0.25

Microsoft SQL Server 2019 (RTM) - 15.0.2000.5

Zapytania

1 – złączenie tablicy Milion z GeoTabelą, do warunku złączenia dodano operację modulo, dopasowującą zakresy wartości złączanych kolumn.

2 – złączenie tablicy Milion ze złączeniem pięciu tabel znormalizowanych.

3 – złączenie tablicy Milion z GeoTabelą, złączenie jest wykonywane poprzez zagnieżdżenie skorelowane.

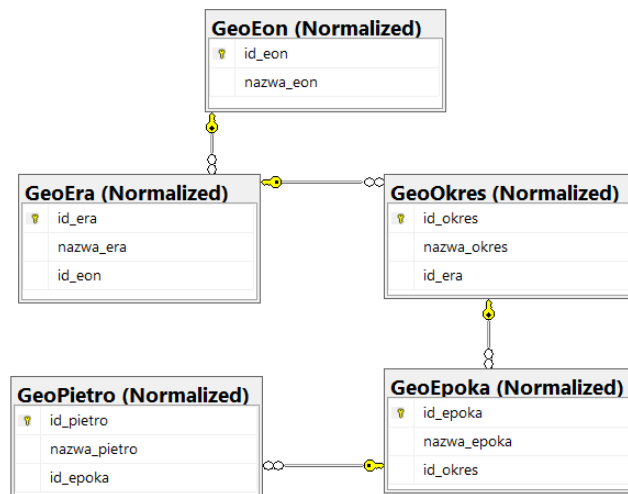
4 -złączenie tablicy Milion z tablicą stratygraficzną, przy czym złączenie jest wykonywane poprzez zagnieżdżenie skorelowane, zapytanie wewnętrzne jest złączeniem tabel znormalizowanych.

Przygotowanie niezbędnych tabel

Idea projektu bazuje na skorzystaniu z dwóch schematów *hurtowni danych*. Każdy z nich ma swoje zalety, a obydwa zostały zrealizowane w oparciu o podstawową tabelę stratygraficzną (geochronologiczną).

Struktura tabeli stratygraficznej pozwala na stworzenie jej bazodanowego odpowiednika na dwa sposoby:

- *znormalizowany* (schemat płątka śniegu)



Realizacja schematu według modelu relacyjnych baz danych pozwala na przechowywanie informacji o znacznych rozmiarach i dużej złożoności.

- *zdenormalizowany* (schemat gwiazdy)

| GeoTabela (nonNormalized) | |
|---------------------------|--|
| id_pietro | |
| nazwa_pietro | |
| id_epoka | |
| nazwa_epoka | |
| id_okres | |
| nazwa_okres | |
| id_era | |
| nazwa_era | |
| id_eon | |
| nazwa_eon | |

Ten prosty schemat zwiększa wydajność operacji - nie trzeba wykonywać szeregu złączeń w momencie, gdy wszystkie niezbędne informacje są w jednym miejscu.

**tabele Dziesięć i Milion*

Do przeprowadzenia testów potrzebna jest tabela złożona z miliona wierszy. Do utworzenia tabeli Milion posłużyła tabela Dziesięć wypełniona cyframi od 0 do 9 i poddana operacji szeregu autozłączeń.

Przeprowadzenie testów

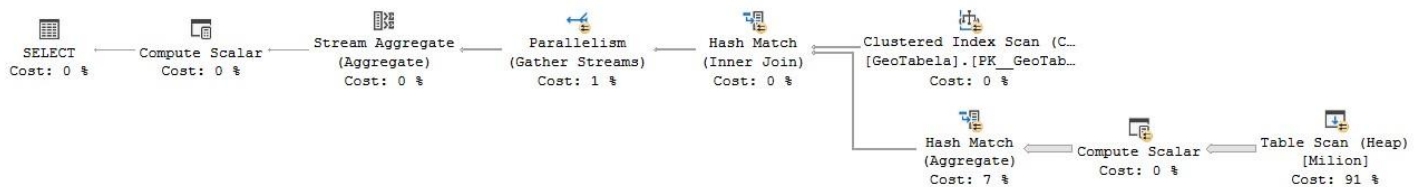
Wykonywane złączenia są bazowane na ww. artykule naukowym. Wprowadzone zmiany spowodowane są różnicami w składni między systemami bazodanowymi¹.

Testy zostały przeprowadzone na dwa sposoby: bez indeksów (oprócz kluczy głównych) oraz z indeksami na kolumnach uczestniczących w złączeniu.

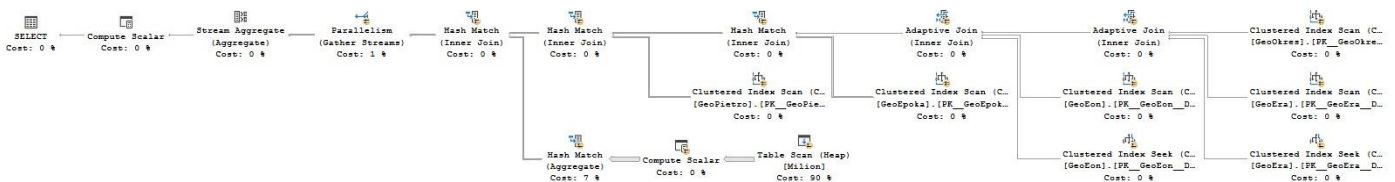
Poniżej znajdują się schematy operacji wraz ze statystykami czasowymi.

Dla SQL Server bez indeksów:

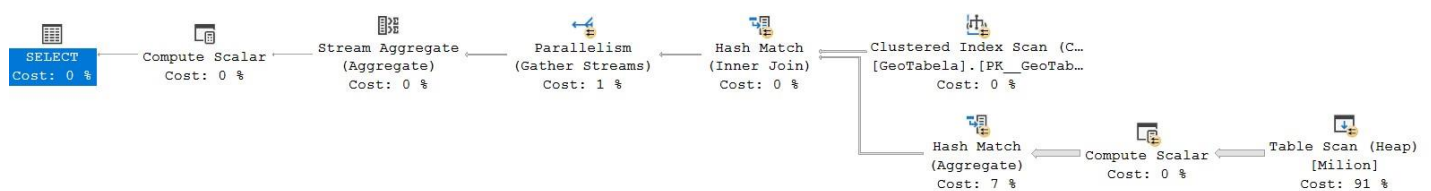
- 1 złączenie



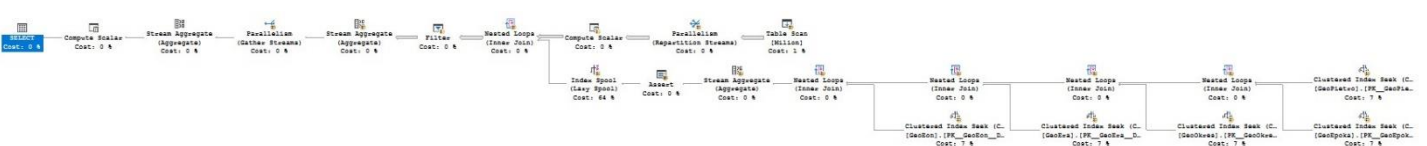
- 2 złączenie



- 3 złączenie



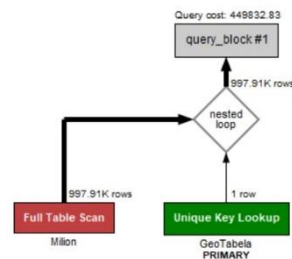
- 4 złączenie



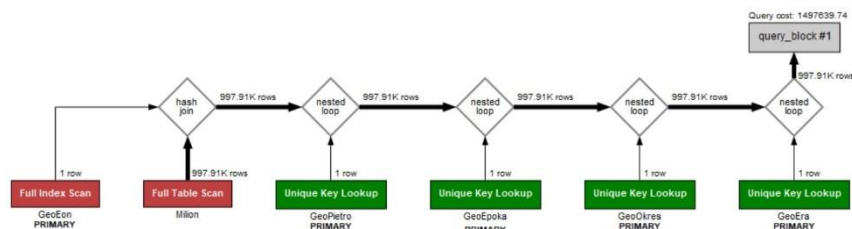
¹ Pełny skrypt do wglądu w repozytorium na GitHub

Dla MySQL bez indeksów:

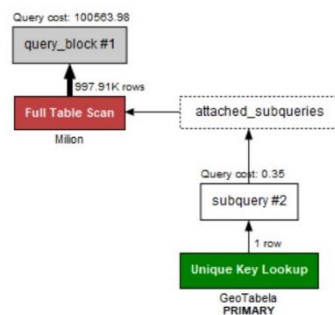
- 1 złączenie



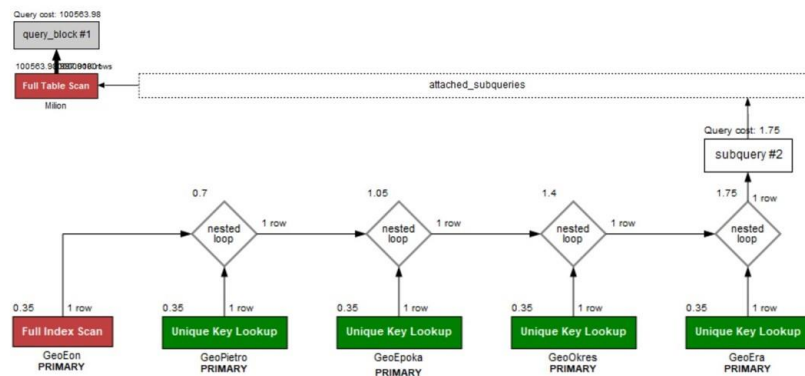
- 2 złączenie



- 3 złączenie

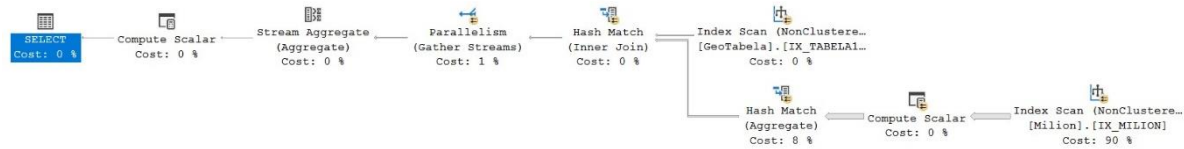


- 4 złączenie

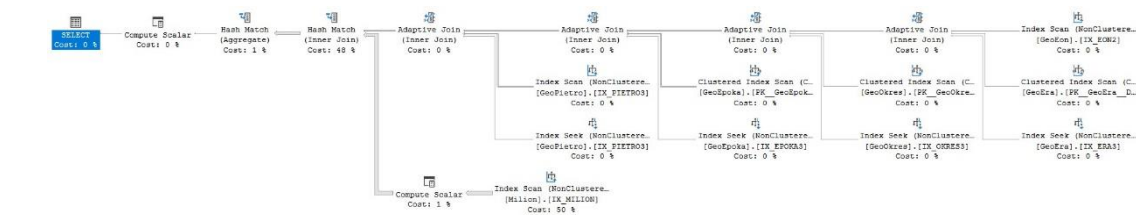


Dla SQL Server z indeksami

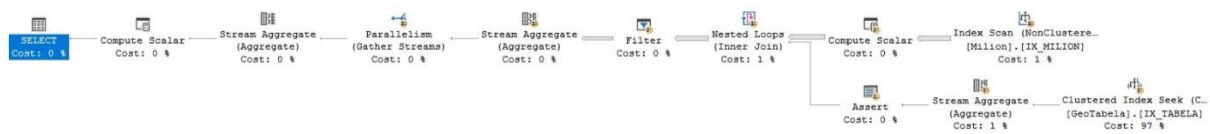
- 1 złączenie



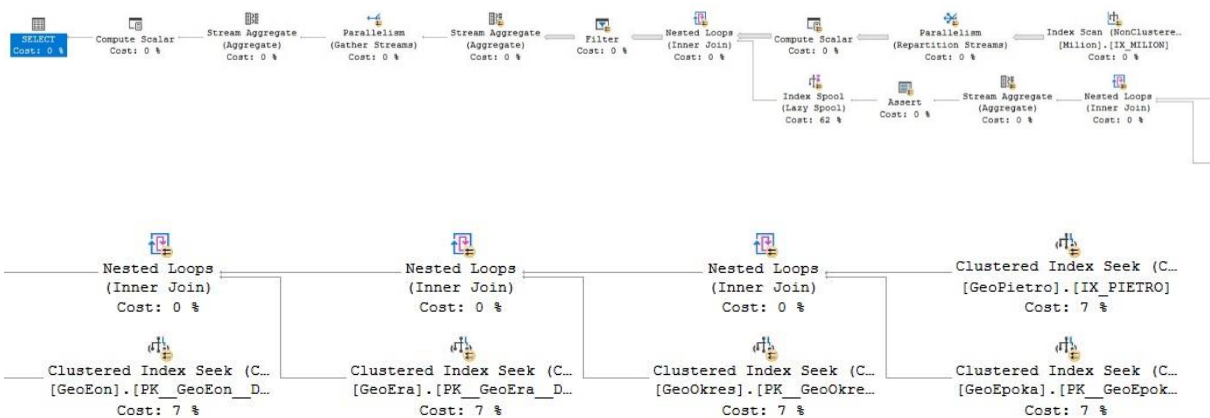
- 2 złączenie



- 3 złączenie



- 4 złączenie



(schemat został podzielony na dwa zrzuty kolejno lewa i prawa część)

- 1 złączenie



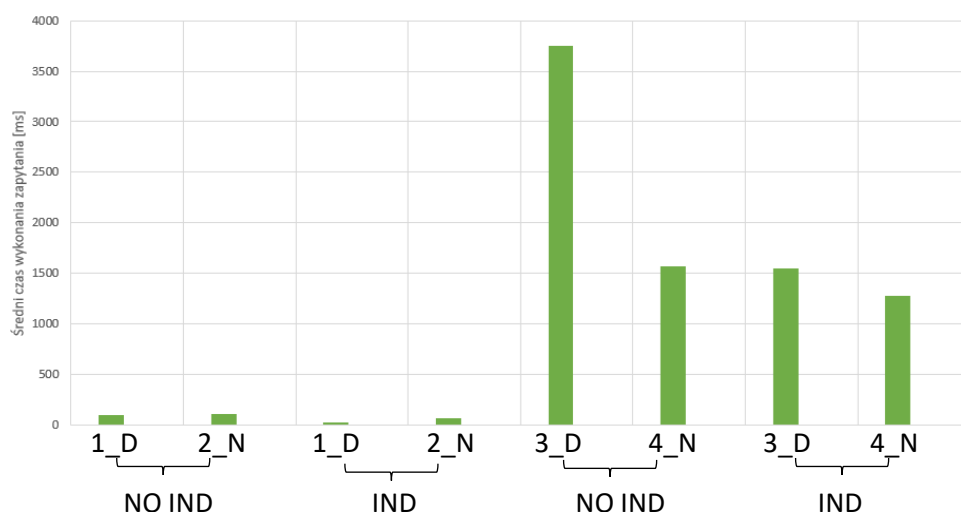
Interpretacja wyników

W czasie przeprowadzanych testów dla obu systemów kilukrotnie pojawiły się wartości wyraźnie odstające od reszty zgodnych z tendencją pomiarów. Z powodu rzadkiego występowania zostały one odrzucone. Dla każdej konfiguracji zostało wykonanych 20 pomiarów. Następnie policzona została wartość średnia oraz odchylenie standardowe. Wyniki prezentuje poniższa tabela oraz wykresy zestawiające średni czas wykonania zapytań.

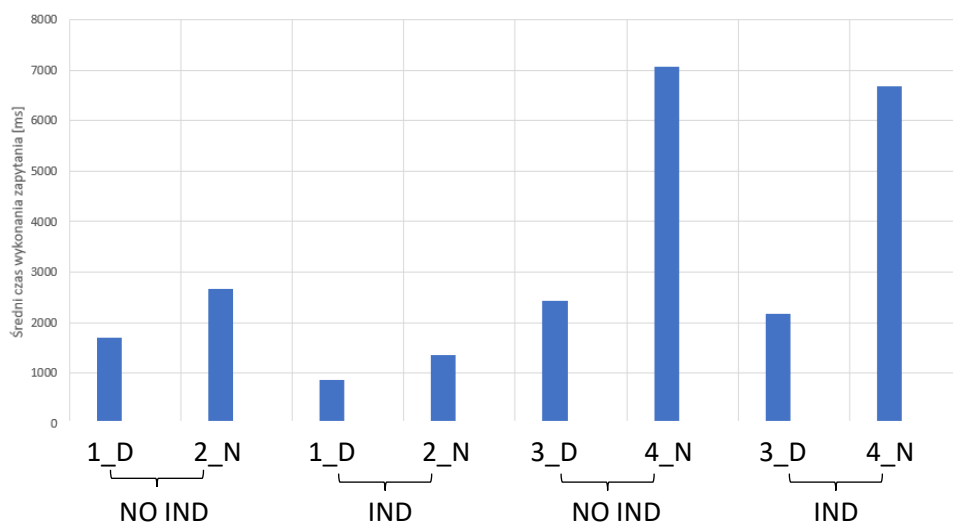
| bez indeksów | 1 | | 2 | | 3 | | 4 | |
|--------------|------|--------|------|--------|------|--------|-------|--------|
| | SD | AVG | SD | AVG | SD | AVG | SD | AVG |
| SQL Server | 35.7 | 90.6 | 48.3 | 107.1 | 80.8 | 3751.8 | 38.4 | 1570.9 |
| MySQL | 55.5 | 1700.0 | 82.0 | 2671.1 | 71.1 | 2424.2 | 141.9 | 7061.8 |

| z indeksami | | | | | | | | |
|-------------|------|-------|------|--------|------|--------|-------|--------|
| SQL Server | 2.6 | 19.9 | 3.8 | 62.6 | 34.8 | 1548.6 | 53.1 | 1274.4 |
| MySQL | 42.0 | 879.6 | 80.8 | 1362.5 | 38.2 | 2170.3 | 173.7 | 6672.9 |

SQL SERVER



MySQL



Obserwacje i wnioski

Zestawienie wyników pozwala na zaobserwowanie konkretnych wniosków na temat użyteczności postaci tabel, indeksacji oraz systemów bazodanowych:

- Postać znormalizowana była szybsza jedynie dla SQL Server w przypadku zagnieżdżeń
- Dla SQL Server różnice między postacią znormalizowaną i zdenormalizowaną dla zapytań jest minimalna, czego nie da się zaobserwować dla pozostałych przypadków
- Indeksowanie przyspiesza wykonanie operacji w obu systemach dla złączeń oraz zagnieżdżeń
- MySQL jest znacznie wolniejszy od SQL Server
- Wyjątkowo dużo czasu dla SQL Server zajmuje operacja 3 (zagnieżdzenie) dla postaci zdenormalizowanej.
- Podobna znaczna różnica zachodzi dla postaci znormalizowanej dla zagnieżdżeń w MySQL

Z powyższych obserwacji wynika, że SQL Server wydaje się być słusznym wyborem w zestawieniu z MySQL, gdy kryterium jest prędkość wykonywanych operacji.

Postać zdenormalizowana zgodnie ze wstępem teoretycznym pozwala na szybsze wykonanie operacji. Nie oznacza to jednak, że jest ona lepszym wyborem – w zestawieniu z postacią znormalizowaną nie wykazuje cech, które w dalszej perspektywie prawdopodobnie okażą się znacznie cenniejsze, jak przechowywanie złożonych informacji oraz zachowywanie porządku i przejrzystości danych.

Bibliografia:

- Materiały z zajęć autorstwa dr inż. Michała Lupy
- Łukasz Jajeńska, Adam Piórkowski,
*Wydajność złączeń i zagnieżdżeń dla schematów
znormalizowanych i zdenormalizowanych*, AGH 2010
- *SQL Notes for Professionals* – goalkicker.com
- *Microsoft SQL Server Notes for Professionals* – goalkicker.com
- *MySQL Notes for Professionals* – goalkicker.com