
Ćwiczenie 9

Spis treści

Cel ćwiczenia	3
Specyfikacja.....	3
Opis	3
Kryteria testów.....	4
Wyniki testów	5
Wyniki testów: MS SQL Server	5
Wyniki testów: PostgreSQL	6
Wyniki testów: MS SQL Server oraz PostgreSQL	7
Wnioski.....	8

Cel ćwiczenia

Celem przedmiotowego ćwiczenia było przeprowadzenie szeregu testów wydajnościowych złączeń oraz zapytań zagnieżdżonych, wykonywanych na tabelach zawierających dużą ilość danych. Aby prawidłowo wykonać ćwiczenie, musieliśmy zgłębić naszą wiedzę z zakresu posługiwania się językiem SQL oraz poznać sposób działania środowisk w celu wykonania naszego zadania.

Specyfikacja

Specyfikacja sprzętu, na którym wykonano testy:

- System Operacyjny: Windows 10
- Procesor: Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz, 2304 MHz, Rdzenie: 4, Procesory logiczne: 8
- RAM: 8GB
- HDD: ST3500418AS ATA Device

Jako systemy zarządzania bazami danych wybrano oprogramowanie wolno dostępne:

- Microsoft SQL Server 18.9.1
- PostgreSQL, wersja 13. 2

Opis

Przed przeprowadzeniem testów utworzono bazę danych **TabelaStratygraficzna**, zawierającą tabelę geochronologiczną, zawierającą informację o wymiarach czasowych i odpowiadające im jednostki stratygraficzne.

Skład schematu znormalizowanej tabeli geologicznej:

- GeoEon
- GeoEra
- GeoOkres
- GeoEpoka
- GeoPietro

Oraz tabele pomocnicze:

- Dziesiec – wypełniona liczbami od 0 do 9, zastosowana przy tworzeniu tabeli *Milion*,
- Milion – tabela wypełniona liczbami naturalnymi od 0 do 999 999, powstała w oparciu o tabele *Dziesiec*,
- GeoTabela – forma zdenormalizowana tabeli geochronologicznej.

Kryteria testów

Testy były wykonywane w seriach po 10 powtórzeń, sprawdzających wydajność złączeń i zagnieźdżeń z tabelą geochronologiczną w wersji zdenormalizowanej i znormalizowanej.

- **Zapytanie 1 (1 ZL)** - celem jest złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci zdenormalizowanej, przy czym do warunku złączenia dodano operację modulo, dopasowującą zakresy wartości złączanych kolumn:

```
SELECT COUNT(*) FROM Milion INNER JOIN GeoTabela ON  
(mod(Milion.liczba,77)=(GeoTabela.id_pietro));
```

- **Zapytanie 2 (2 ZL)** - celem jest złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci znormalizowanej, reprezentowaną przez złączenia pięciu tabel:

```
SELECT COUNT(*) FROM Milion INNER JOIN GeoPietro ON (mod(Milion.liczba,77)=GeoPietro.id_pietro)  
NATURAL JOIN GeoEpoka NATURAL JOIN GeoOkres NATURAL JOIN GeoEra NATURAL JOIN GeoEon;
```

- **Zapytanie 3 (3 ZG)** - celem jest złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci zdenormalizowanej, przy czym złączenie jest wykonywane poprzez zagnieźdzenie skorelowane:

```
SELECT COUNT(*) FROM Milion WHERE mod(Milion.liczba,77)=  
(SELECT id_pietro FROM GeoTabela WHERE mod(Milion.liczba,77)=(id_pietro));
```

- **Zapytanie 4 (4 ZG)** - celem jest złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci znormalizowanej, przy czym złączenie jest wykonywane poprzez zagnieźdzenie skorelowane, a zapytanie wewnętrzne jest złączeniem tabel poszczególnych jednostek geochronologicznych:

```
SELECT COUNT(*) FROM Milion WHERE mod(Milion.liczba,77) IN  
(SELECT GeoPietro.id_pietro FROM GeoPietro NATURAL JOIN GeoEpoka NATURAL JOIN GeoOkres  
NATURAL JOIN GeoEra NATURAL JOIN GeoEon;
```

Wyniki testów

Poniżej zestawiono tabelę zawierającą wyniki minimalnych oraz średnich czasów testów zapytań [ms], kutyh celem było złączenie syntetycznej tabeli *Milion* z tabelą geologiczną:

- W postaci zdenormalizowanej: ZL1 oraz ZG3, przy czym ZG3 jest zapytaniem o złączeniu wykonywanym przez zagnieżdżenie skorelowane.
- W postaci znormalizowanej: ZL2 ORAZ ZG4, przy czym ZG3 jest zapytaniem o złączeniu wykonywanym przez zagnieżdżenie skorelowane.

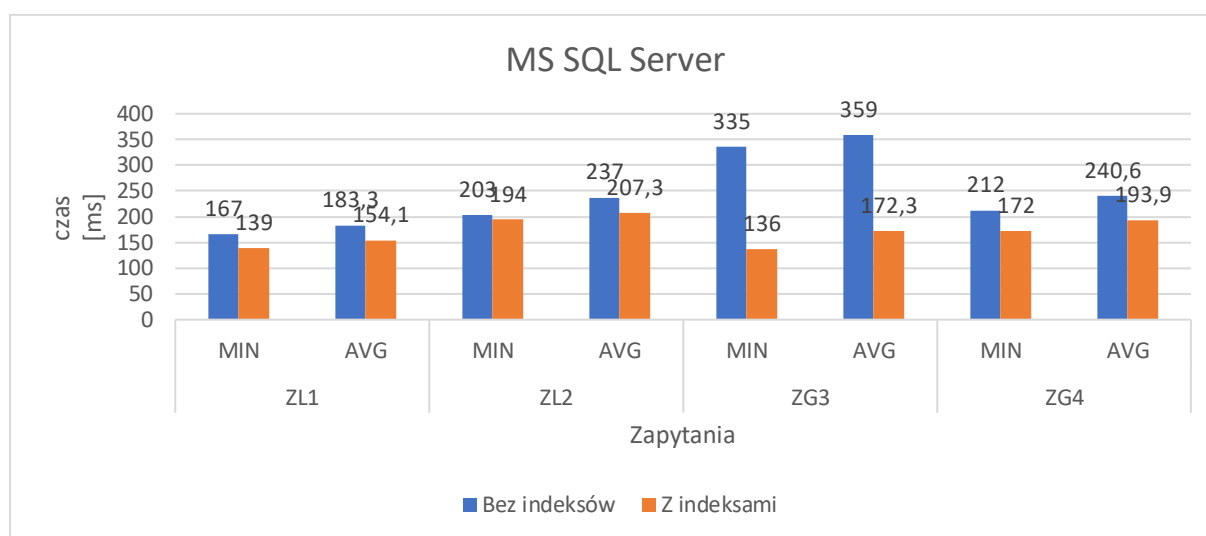
WYNIKI								
	ZL1		ZL2		ZG3		ZG4	
	MIN	AVG	MIN	AVG	MIN	AVG	MIN	AVG
Bez indeksow								
MS SQL Server	167	183,3	203	237	335	359	212	240,6
PostgreSQL	179	224,3	354	378,1	10546	10772	173	194,6
Z indeksami								
MS SQL Server	139	154,1	194	207,3	136	172,3	172	193,9
PostgreSQL	173	202	245	265,1	10567	10721,9	194,6	194,5

Tab1: Zestawienie średnich oraz minimalnych czasów [ms] z indeksami oraz bez dla MS SQL Server oraz PostgreSQL

Opracowanie: Wiktoria Drożdż

Wyniki testów: MS SQL Server

Analizując czas zapytań dla testów przeprowadzonych w MS SQL Server, oparto się o zestawienie uzyskanych wyników w postaci wykresów w skali liniowej.



Wyk.1 : Wyniki testów z indeksacją oraz bez indeksacji w MS SQL Server

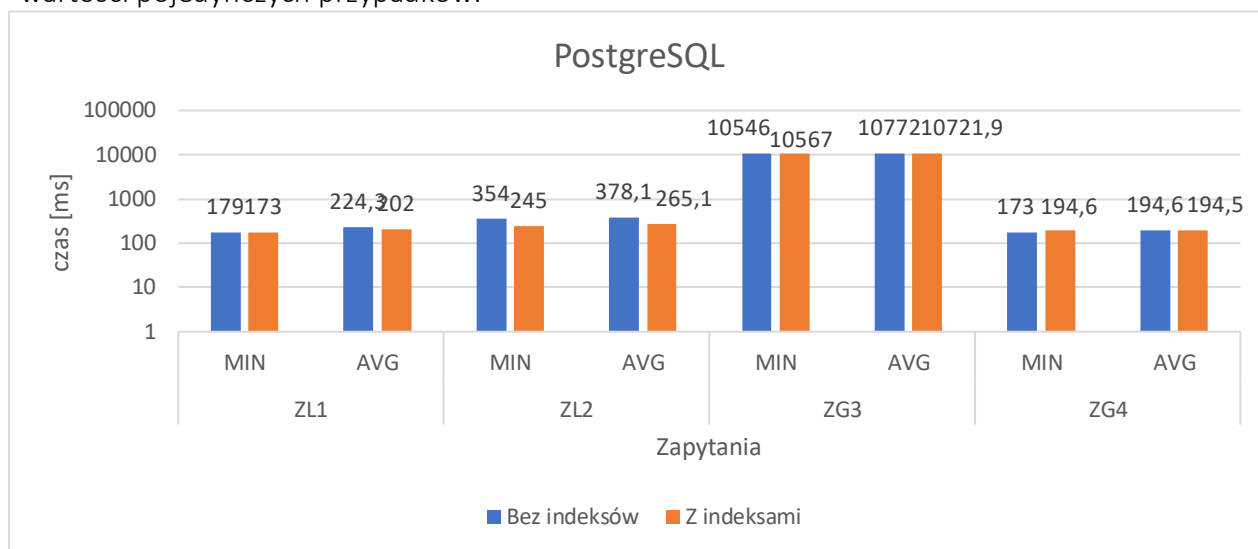
Opracowanie: Wiktoria Drożdż

Uzyskane wyniki potwierdzają w sposób jednoznaczny, iż testy przeprowadzone dla zapytań kolumn niezawierających indeksów okazały się wolniejsze w stosunku do zapytań przeprowadzonych dla kolumn zawierających indeksy.

Szczególną różnicę widać w zestawieniu wyników dla zapytania ZG3, które jest zapytaniem do postaci zdenormalizowanej tabeli geologicznej o złączeniu wykonywanym przez zagnieżdżenie skorelowane oraz dla zapytania ZG4, które jest zapytaniem do postaci znormalizowanej o złączeniu wykonywanym przez zagnieżdżenie skorelowane. Czas testu zapytania ZG3 przeprowadzonego na kolumnach nieposiadających indeksów jest najwolniejszy.

Wyniki testów: PostgreSQL

Analizując czas zapytań testów przeprowadzonych w PostgreSQL, dla lepszej czytelności zostały one oparte na postaci wykresów w skali logarytmicznej, ze względu na dość duże wartości pojedynczych przypadków.



Wyk.2: Wyniki testów z indeksacją oraz bez indeksacji w PostgreSQL

Opracowanie: Wiktoria Drożdż

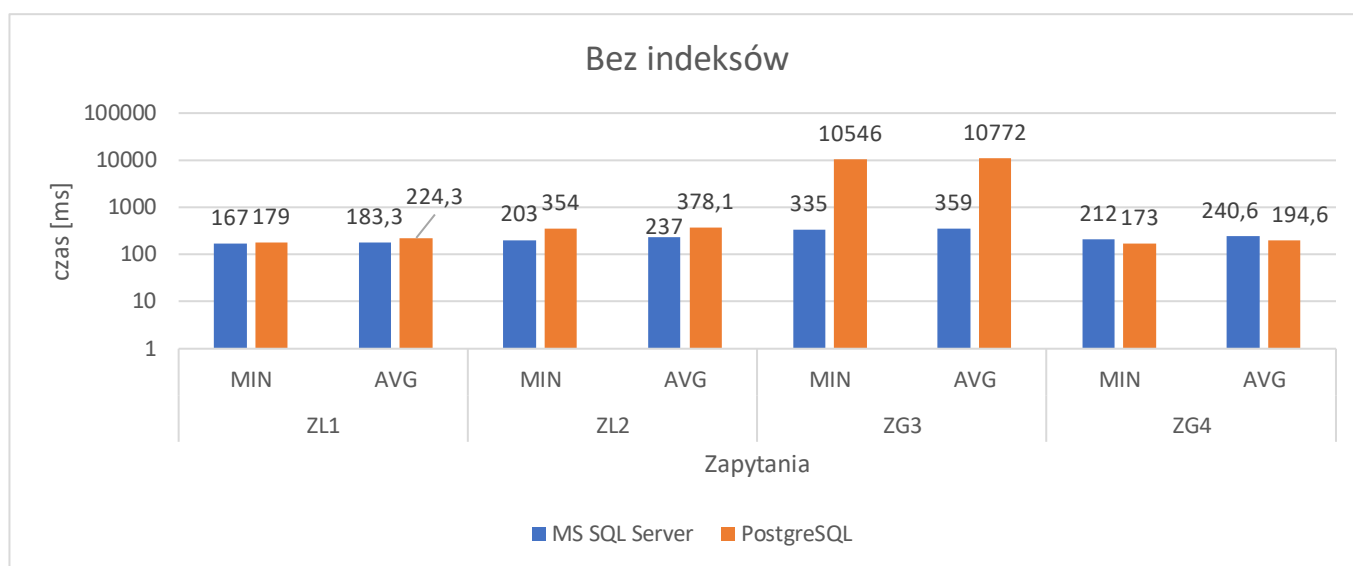
W przypadku wyników testów przeprowadzonych w PostgreSQL można stwierdzić, że pierwsze dwa zapytania (pierwsze jest zapytaniem do postaci zdenormalizowanej tabeli stratygraficznej, a drugie do postaci znormalizowanej) zostały szybciej przeprowadzone dla kolumn z naniesionymi indeksami.

W przypadku zapytania do postaci zdenormalizowanej (ZG3 oraz ZG4), które są zapytaniami do postaci zdenormalizowanej i znormalizowanej tabeli, o złączeniach przeprowadzonych poprzez zagnieżdżenia skorelowane, widać minimalne różnice przy przeprowadzaniu testów na kolumnach bez oraz z naniesionymi indeksami, jednakże ze względu na dość małe wartości można przyjąć, że wartości średnie są zbliżone, identyczne.

Najwyższe wartości czasów są osiągnięte dla testów zapytania ZG3.

Wyniki testów: MS SQL Server oraz PostgreSQL

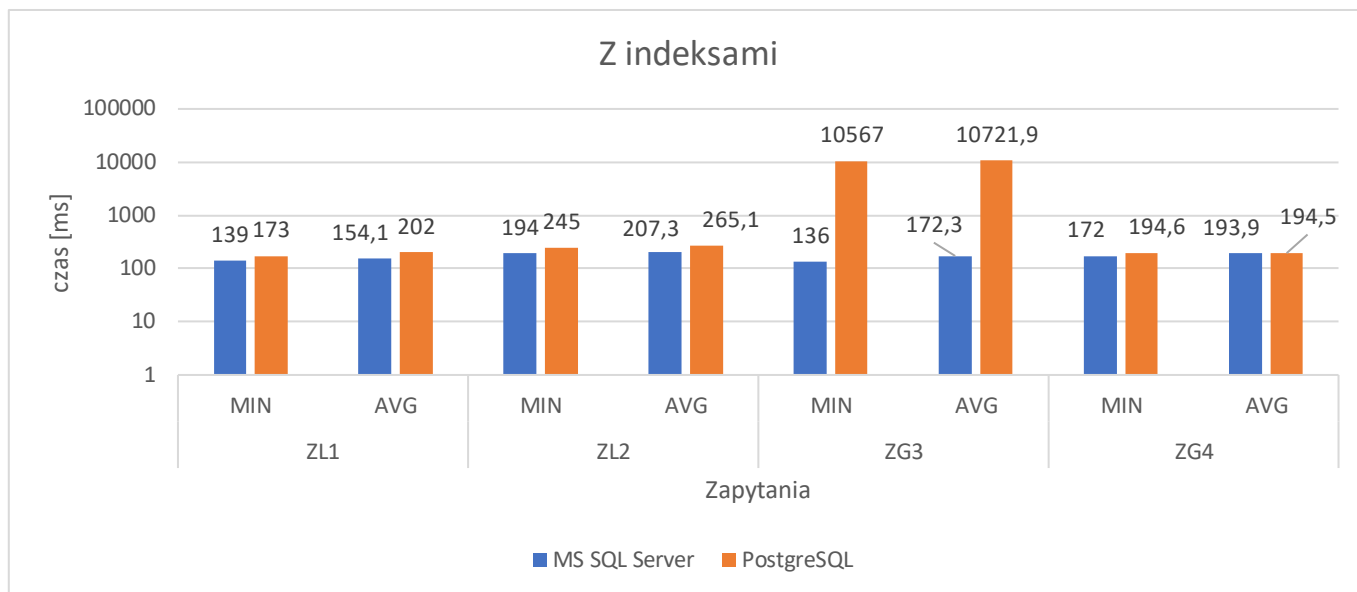
Analizując czas zapytań dla testów przeprowadzonych w PostgreSQL oraz MS SQL Server ponownie oparto się na przedstawieniu uzyskanych wyników w postaci wykresów w skali logarytmicznej, ze względu na dość duże wartości pojedynczych przypadków. Osobno zestawiono wyniki dla zapytań przeprowadzonych na kolumnach nieposiadających indeksów oraz dla kolumn zawierających indeksy.



Wyk.3: Wyniki testów bez indeksacji w MS SQL Server oraz PostgreSQL

Opracowanie: Wiktoria Drożdż

W przypadku wyników testów przeprowadzonych w MS SQL Server oraz PostgreSQL na kolumnach nieposiadających indeksów, czasy ich przeprowadzenia w MS SQL Server okazały się szybsze dla dwóch pierwszych zapytań odnoszących się do tabeli geologicznej w postaci zdenormalizowanej oraz znormalizowanej. Inaczej jest w przypadku testów zapytań, w których występują zagnieżdżenia skorelowane. W przypadku testu zapytania ZG3, które jest zapytaniem do tabeli stratygraficznej postaci zdenormalizowanej, czas wykonania jest znacznie szybszy niż czas przeprowadzenia tego testu w PostgreSQL, jednakże w przypadku zapytania ZG4, które jest zapytaniem z do tabeli w postaci znormalizowanej, czas przeprowadzonego w PostgreSQL jest szybszy niż czas wykonania tego samego testu w MS SQL Server.



Wyk.4: Wyniki testów z indeksacją w MS SQL Server oraz PostgreSQL

Opracowanie: Wiktoria Drożdż

W przypadku wyników testów przeprowadzonych w MS SQL Server oraz PostgreSQL na kolumnach posiadających indeksy, czas ich przeprowadzenia w MS SQL Server okazały się szybsze dla dwóch pierwszych zapytań do tabeli geologicznej w postaci zdenormalizowanej oraz znormalizowanej. Również dla testu zapytania ZG3, które jest zapytaniem do zdenormalizowanej tabeli geologicznej, czas wykonania jest znacznie szybszy niż czas przeprowadzenia tego testu w PostgreSQL, lecz i w tym przypadku testu zapytania ZG4, które jest zapytaniem do znormalizowanej tabeli stratygraficznej, czas przeprowadzonego w PostgreSQL jest szybszy niż czas wykonania tego samego testu w MS SQL Server, jednakże różnice są minimalne, wręcz wyniki są do siebie zbliżone.

Wnioski

Na podstawie uzyskanych wyników można wyciągnąć następujące wnioski:

- Wydajniejsza jest postać znormalizowana, wyjątkiem jest postać znormalizowana o zagnieżdżeniu skorelowanym (przypadek zapytania ZG3).
- Zapytania wykonywały się znacznie szybciej w MS SQL Server niż w PostgreSQL, wyjątkiem są zapytania znormalizowane o złączeniu przeprowadzonym zagnieżdżeniem skorelowanym (zapytanie ZG4).
- Złączenia w PostgreSQL w postaci znormalizowanej wykonują się równie szybko co z w postaci zdenormalizowanej.
- Zagnieżdżenia skorelowane są wolniejsze niż złączenia.