

Kraków, 24.05.2023

„Analiza wydajności złączeń i zagnieżdżeń dla schematów znormalizowanych i zdenormalizowanych”

Bazy danych- ćwiczenie 9

Zuzanna Sabat

410176

Geoinformatyka II

1. Wstęp

Celem zadania było odtworzenie badania wydajności złączeń i zagnieżdżeń w schematach znormalizowanych i zdenormalizowanych przeprowadzonego w pracy o tym samym tytule autorstwa Łukasza Jajeśnicy i Adama Piórkowskiego. Do wykonania zadania stworzono dwie tabele w języku SQL przedstawiające tabelę geochronologiczną zawierającą następujące jednostki czasowe: eon, era, okres, epoka i piętro. Pierwsza z nich została stworzona na podstawie schematu płatka śniegu (forma znormalizowana), a druga na podstawie schematu gwiazdy (forma zdenormalizowana). Analiza została wykonana w dwóch środowiskach: MySQL oraz PostgreSQL.

2. Konfiguracja sprzętowa i programowa

I. Badania przeprowadzono na laptopie o następującej specyfikacji:

-Procesor: Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz 2.71 GHz

-Pamięć RAM: 8,00 GB

-Dysk SSD- pojemność: 128 GB

-System operacyjny: 64-bitowy system operacyjny, procesor x64 Windows 10 Home

II. Do utworzenia bazy danych użyto systemów:

-MySQL Workbench 8.0 CE/ MySQL Shell

-PostgreSQL 15.3/ pgAdmin4

3. Stworzenie bazy danych

Na początku dla schematu „stratygrafia” wykonano pięć znormalizowanych tabel przedstawiających poszczególne jednostki czasowe: GeoEon, GeoEra, GeoOkres, GeoEpoka i GeoPiętro. Tabele stworzono w oparciu o schemat płatka śniegu.

```
CREATE TABLE stratygrafia.GeoEon(  
    id_eon INTEGER PRIMARY KEY,  
    nazwa_eon CHAR(10) NOT NULL);
```

3.1 Kod dla tabeli GeoEon wykonany w środowisku PostgreSQL

Aby wykonać zdenormalizowaną tabelę „GeoTabela” użyto złączeń naturalnych wspomnianych wcześniej tabel. Tabela ta bazuje na schemacie gwiazdy.

```

CREATE TABLE stratygrafia.GeoTabela
    AS(
SELECT * FROM stratygrafia.GeoPietro

    NATURAL JOIN stratygrafia.GeoEpoka
    NATURAL JOIN stratygrafia.GeoOkres
    NATURAL JOIN stratygrafia.GeoEra
    NATURAL JOIN stratygrafia.GeoEon );

```

3.2 Kod dla tabeli GeoTabela w środowisku PostgreSQL.

4. Zapytania testowe

Do zbadania wydajności złączeń i zagnieżdżeń wykonano cztery testy. Aby je przeprowadzić przygotowano dwie tabelę w schemacie „test”: „dziesięć” oraz „milion”. Pierwsza z nich zawierała cyfry od 0 do 9 włącznie, natomiast druga od 0 do 999999.

```

CREATE TABLE test.dziesiec(
    cyfra INTEGER NOT NULL,
    bitt INTEGER NOT NULL
);

INSERT INTO test.dziesiec VALUES
    (0, 1),
    (1, 1),
    (2, 1),
    (3, 1),
    (4, 1),
    (5, 1),
    (6, 1),
    (7, 1),
    (8, 1),
    (9, 1);

```

4.1 Kod dla tabeli „dziesięć” w środowisku PostgreSQL

```
CREATE TABLE test.milion(
    liczba INTEGER NOT NULL,
    cyfra INTEGER NOT NULL,
    bitt INTEGER NOT NULL);
```

```
INSERT INTO test.milion SELECT
a1.cyfra+10*a2.cyfra+100*a3.cyfra+1000*a4.cyfra+10000*a5.cyfra+
100000*a6.cyfra
    AS liczba , a1.cyfra AS cyfra, a1.bitt AS bitt
FROM test.dziesiec a1, test.dziesiec a2, test.dziesiec a3,
    test.dziesiec a4, test.dziesiec a5, test.d
```

4.2 Kod dla tabeli „milion” w środowisku PostgreSQL

I. TEST 1 (zapytanie 1: ZL_1)

Celem testu było złączenie tablicy „milion” z tabelą geochronologiczną „GeoTabela” w postaci zdenormalizowanej. Aby dopasować zakresy wartości złączanych kolumn użyto operacji modulo.

```
SELECT COUNT(*) AS ZL_1
FROM test.milion INNER JOIN stratygrafia.GeoTabela ON
    MOD(test.milion.liczba, 75) = GeoTabela.id_pietro;
```

4.1 Test 1 w środowisku PostgreSQL.

II. TEST 2 (zapytanie 2: ZL_2)

Celem testu było złączenie tablicy „milion” z tabelą geochronologiczną w postaci znormalizowanej przez złączenia pięciu tabel: Eon, Epoka, Okres, Era i Piętro.

```
SELECT COUNT(*) AS ZL_2
FROM test.milion
INNER JOIN stratygrafia.GeoPietro ON MOD(test.milion.liczba, 75)
    = stratygrafia.GeoPietro.id_pietro
NATURAL JOIN stratygrafia.GeoEpoka
NATURAL JOIN stratygrafia.GeoOkres
NATURAL JOIN stratygrafia.GeoEra
NATURAL JOIN stratygrafia.GeoEon;
```

4.2. Test 2 w środowisku PostgreSQL .

III. TEST 3 (zapytanie 3: ZG_3)

Celem testu było złączenie tabeli „milion” z „GeoTabelą” poprzez zagnieżdżenie skorelowane.

```
SELECT COUNT(*) AS ZG_3
FROM test.milion
WHERE MOD(test.milion.liczba, 75) = (SELECT id_pietro FROM
stratygrafia.GeoTabela
WHERE MOD(test.milion.liczba, 75) = id_pietro);
```

4.3 Test 3 w środowisku PostgreSQL.

IV. TEST 4 (zapytanie 4: ZG_4)

Celem testu było złączenie tablicy „milion” z tabelą geochronologiczną w postaci znormalizowanej przez złączenia pięciu tabel. Przy czym, złączenie to było wykonywane poprzez zagnieżdżenie skorelowane.

```
SELECT COUNT(*) AS ZG_4
FROM test.milion
WHERE MOD(test.milion.liczba, 75) IN
(SELECT stratygrafia.GeoPietro.id_pietro
FROM stratygrafia.GeoPietro
NATURAL JOIN stratygrafia.GeoEpoka
NATURAL JOIN stratygrafia.GeoOkres
NATURAL JOIN stratygrafia.GeoEra
NATURAL JOIN stratygrafia.GeoEon);
```

4.4 Test 4 w środowisku PostgreSQL.

Każdy z testów został wykonany na danych bez założonych indeksów oraz na danych z założonymi indeksami na każdej z kolumn.

```
CREATE UNIQUE INDEX id_eon_idx ON stratygrafia.GeoEon(id_eon);
```

4.5 Stworzenie indeksu na tabeli GeoEon w środowisku PostgreSQL.

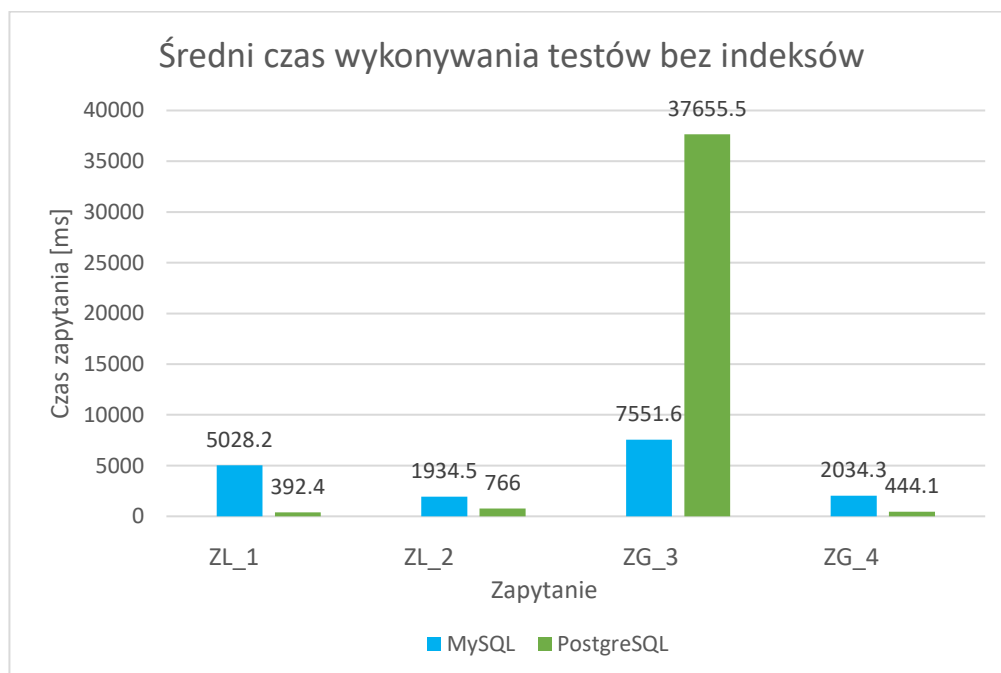
5. Wyniki testów.

Wszystkie testy powtórzono dziesięć-krotnie, a następnie została obliczona średnia arytmetyczna wyników, aby ukazać najbardziej typowy czas wykonania. Dodatkowo zostały zapisane wartości skrajne- minimalne. Czas został wyrażony w milisekundach.

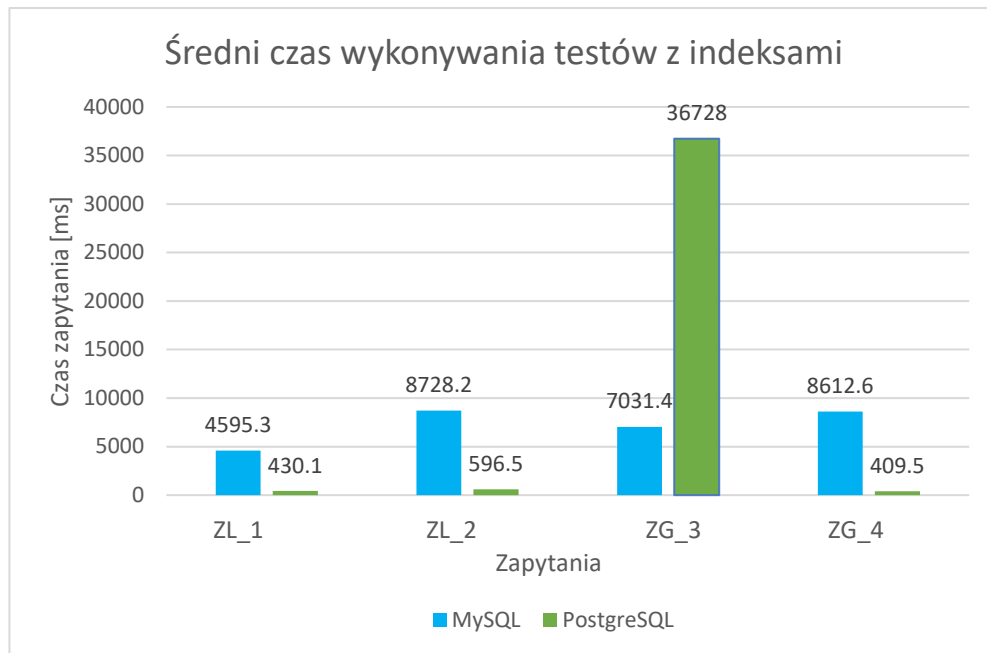
| | ZL_1 | | ZL_2 | | ZG_3 | | ZG_4 | |
|--------------|------|---------|------|---------|-------|---------|------|---------|
| Bez indeksów | MIN | ŚREDNIA | MIN | ŚREDNIA | MIN | ŚREDNIA | MIN | ŚREDNIA |
| PostgreSQL | 314 | 392,4 | 677 | 766 | 35121 | 37655.5 | 377 | 444.1 |
| MySQL | 4828 | 5028.2 | 1890 | 1934.5 | 7250 | 7551.6 | 1875 | 2034.3 |
| Z indeksami | | | | | | | | |
| PostgreSQL | 357 | 430.1 | 513 | 596.5 | 35098 | 36728 | 318 | 409.5 |
| MySQL | 4391 | 4595.3 | 8172 | 8728.2 | 6703 | 7031.4 | 8266 | 8612.6 |

5.1 Zestawienie czasu wykonywania zapytań z indeksami i bez indeksów w milisekundach.

Wyniki średnich pomiarów czasów przedstawiono graficznie za pomocą wykresów słupkowych stworzonych w programie Excel.



5.2 Wykres przedstawiający średni czas wykonywania testów bez indeksów w środowiskach MySQL oraz PostgreSQL.



5.2 Wykres przedstawiający średni czas wykonywania testów z indeksami w środowiskach MySQL oraz PostgreSQL.

6. Wnioski

Na podstawie przeprowadzonych testów można wysnuć następujące wnioski:

- I. Zapytania dla tabeli o schemacie płotka śniegu (znormalizowane) były wydajniejsze bez założonych indeksów zarówno w środowisku MySQL jak i PostgreSQL.
- II. Test dla tabeli o schemacie gwiazdy (zdenormalizowane) kompilowały się szybciej po nałożeniu indeksów na wszystkie kolumny. Wyjątek stanowiło zapytanie stworzone w środowisku PostgreSQL.
- III. Średni czas wykonywania wszystkich zapytań był najkrótszy dla bazy MySQL bez indeksów, a najdłuższy dla PostgreSQL bez indeksów.

Nie da się jednoznacznie stwierdzić, które środowisko jest wydajniejsze. W większości przypadków zapytania wykonane w PostgreSQL kompilowały się najkrócej, jednak należy zwrócić uwagę na to, że trzeci test w tym środowisku wykonywał się około pięć razy dłużej niż w środowisku MySQL zarówno dla zapytań z indeksami jak i bez.

7. Bibliografia

-Artykuł: „Wydajność złączeń i zagnieźdzeń dla schematów znormalizowanych i zdenormalizowanych”- mgr. inż. Łukasz Jajeńska, dr.hab. inż. Adam Piórkowski