

Wydajność złączeń i zagnieżdżeń dla schematów znormalizowanych i zdenormalizowanych

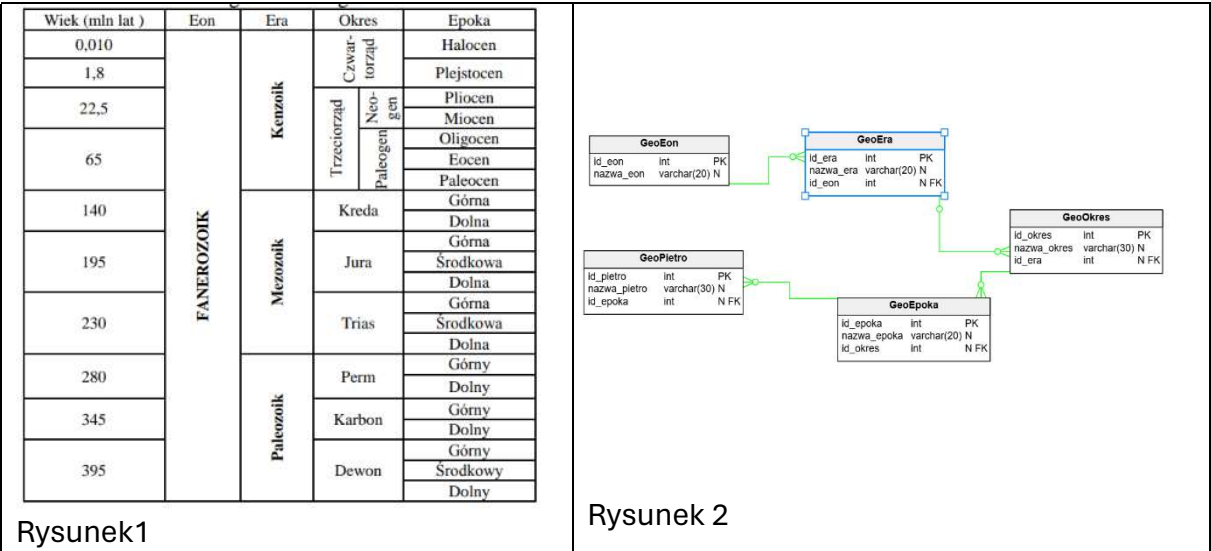
1.Cel

Celem ćwiczenia jest zbadanie wydajności różnych schematów baz danych, zarówno znormalizowanych, jak i zdenormalizowanych, w kontekście operacji złączeń i zagnieżdżonych zapytań. Przedstawiono ideę normalizacji i denormalizacji oraz ich wpływ na wydajność w systemach zarządzania bazami danych, takich jak SQL Server Management Studio i PostgreSQL.

2.Użyte tabele

Wykorzystano jako przykład tabelę geochronologiczną(Rysunek 1).Tabela ta przedstawia taksonomię dla czterech jednostek geochronologicznych – eonu, ery, okresu i epoki. Wymiar ostatniej jednostki – piętra został pominięty na zdjęciu ze względu na obszerność, ale został użyty do bazy danych.

Do analizy użyto znormalizowanej tabeli geochronologicznej(Rysunek 2), zdenormalizowanej(Rysunek 3) , tabeli Milion(Rysunek 4) oraz tabeli Dziesięć(Rysunek5)



d) Dodajemy klucze obce:

```
ALTER TABLE GeoEra ADD FOREIGN KEY (id_eon) REFERENCES GeoEon(id_eon);
ALTER TABLE GeoOkres ADD FOREIGN KEY (id_era) REFERENCES GeoEra(id_era);
ALTER TABLE GeoEpoka ADD FOREIGN KEY (id_okres) REFERENCES GeoOkres(id_okres);
ALTER TABLE GeoPietro ADD FOREIGN KEY (id_epoka) REFERENCES GeoEpoka(id_epoka);
```

Następnie wypełniamy tabele wartościami

3) tabela zdenormalizowana

Formę zdenormalizowaną tabeli geochronologicznej osiągnięto tworząc jedną tabelę ZdGeoTabela zawierającą wszystkie dane z powyższych tabel.

```
SELECT p.id_pietro, p.nazwa_pietro, ep.id_epoka,
       ep.nazwa_epoka, o.id_okres, o.nazwa_okres, er.id_era,
       er.nazwa_era, eo.id_eon, eo.nazwa_eon
INTO ZdGeoTabela
FROM GeoPietro p
JOIN GeoEpoka ep ON ep.id_epoka = p.id_epoka
JOIN GeoOkres o ON o.id_okres = ep.id_okres
JOIN GeoEra er ON er.id_era = o.id_era
JOIN GeoEon eo ON eo.id_eon = er.id_eon;
ALTER TABLE ZdGeoTabela ADD PRIMARY KEY (id_pietro);
```

3.2 PostgreSQL

Wszystko wykonujemy analogicznie jak w MySQL

4. Testy wydajności

4.1 SQL Server Management Studio

Etap 1-obejmuje zapytania bez nałożonych indeksów na kolumny danych (jedynymi indeksowanymi danymi były dane w kolumnach będących kluczami głównymi poszczególnych tabel,)

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:

```
--1
SET STATISTICS TIME ON SELECT COUNT(*)
FROM Milion
INNER JOIN ZdGeoTabela
ON (Milion.liczba % 68 = ZdGeoTabela.id_pietro);
SET STATISTICS TIME OFF
```

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

```
--2
SET STATISTICS TIME ON
SELECT COUNT(*) FROM Milion
INNER JOIN GeoPietro ON (Milion.liczba % 68=GeoPietro.id_pietro)
INNER JOIN GeoEpoka ON GeoPietro.id_epoka=GeoEpoka.id_epoka
INNER JOIN GeoOkres ON GeoEpoka.id_okres= GeoOkres.id_okres
INNER JOIN GeoEra ON GeoEra.id_era=GeoOkres.id_era
INNER JOIN GeoEon ON GeoEon.id_eon=GeoEra.id_eon
SET STATISTICS TIME OFF
```

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

```
--3
SET STATISTICS TIME ON
SELECT COUNT(*)
FROM Milion
WHERE Milion.liczba % 68 =
(SELECT id_pietro FROM ZdGeoTabela WHERE Milion.liczba % 68 = id_pietro);
SET STATISTICS TIME OFF
```

Zapytanie 4 (4 ZG), którego celem jest 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:

```
--4
SET STATISTICS TIME ON
SELECT COUNT(*)
FROM Milion
WHERE Milion.liczba % 68 IN
(SELECT GeoPietro.id_pietro FROM GeoPietro
INNER JOIN GeoEpoka ON GeoPietro.id_epoka = GeoEpoka.id_epoka
INNER JOIN GeoOkres ON GeoEpoka.id_okres = GeoOkres.id_okres
INNER JOIN GeoEra ON GeoOkres.id_era = GeoEra.id_era
INNER JOIN GeoEon ON GeoEra.id_eon = GeoEon.id_eon);
SET STATISTICS TIME OFF
```

Etap 2- nałożenie indeksów na wszystkie kolumny biorące udział w złączeniu

```
CREATE INDEX idxEon ON GeoEon(id_eon);
CREATE INDEX idxEra ON GeoEra(id_era, id_eon);
CREATE INDEX idxOkres ON GeoOkres(id_okres, id_era);
CREATE INDEX idxEpoka ON GeoEpoka(id_epoka, id_okres);
CREATE INDEX idxPietro ON GeoPietro(id_pietro, id_epoka);
CREATE INDEX idxLiczba ON Milion(liczba);
CREATE INDEX idxZdGeoTabela ON ZdGeoTabela(id_pietro, id_epoka, id_era, id_okres, id_eon);
```

Następnie analogicznie przeprowadzono zapytania.

4.2 PostgreSQL

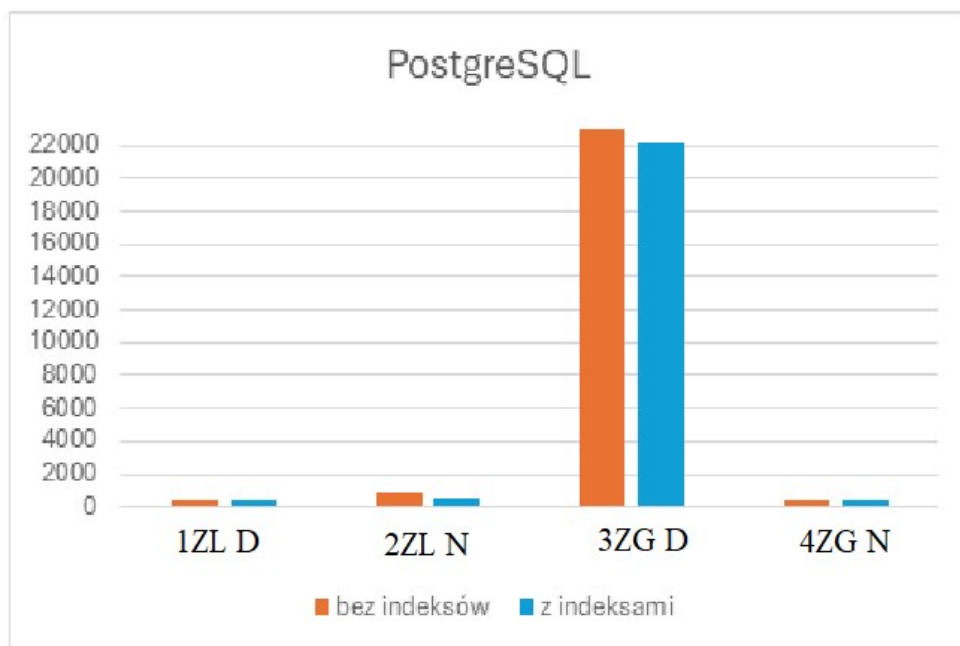
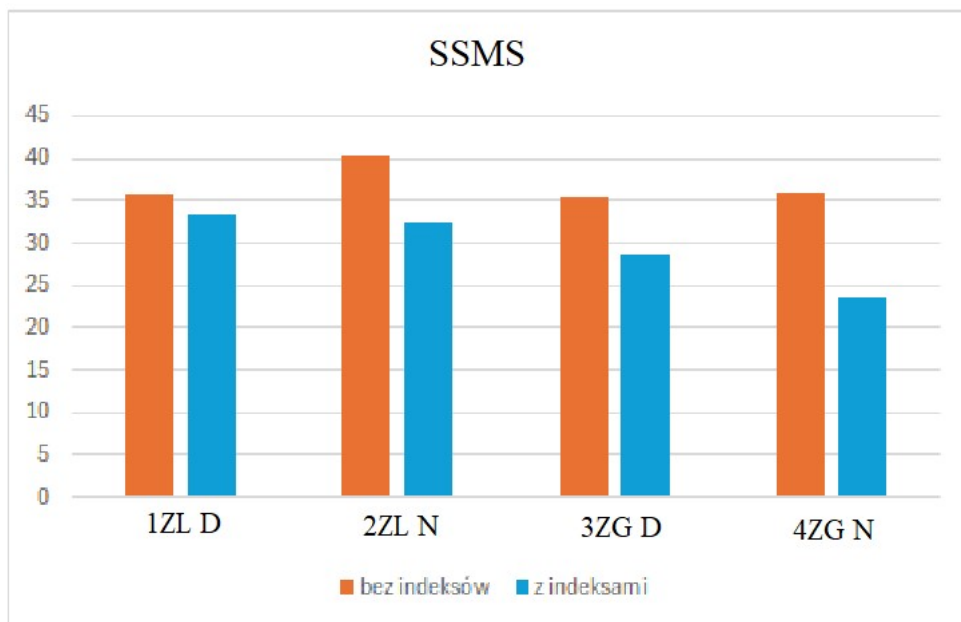
Wszystko wykonujemy analogicznie jak w SSMS

5.Wyniki

Testy przeprowadzono kilkakrotnie, a następnie policzono średnią.

	1ZL	1ZL	2ZL	2ZL	3ZG	3ZG	4ZG	4ZG
Bez indeksów	min	śr	min	śr	min	śr	min	śr
SSMS	34	35,6	38	40,2	33	35,4	32	36
PostgreSQL	316	356	701	768,2	21845	23423,6	325	400
Z indeksami								
SSMS	31	33,4	28	32,4	26	28,6	21	23,6
PostgreSQL	320	383,8	518	569	21535	22251,8	278	339

Wykresy z danymi dla SSMS i PostgreSQL, oś x przedstawia poszczególne zapytania, oś y czas wykonania zapytania w [ms].



5.Wnioski

- Postać zdenormalizowana jest na ogół wydajniejsza, ale w PostgreSQL czas wykonywania zagnieżdżenia skorelowanego był dużo wolniejszy dla wersji zdenormalizowanej zarówno z indeksami jak i bez

- Dodanie indeksów niemal we wszystkich przypadkach przyspieszyło czas wykonywania
- Wszystkie zapytania szybciej wykonały się w SQL Management Studio
- W PostgreSQL zapytanie 3 wykonywało się o wiele dłużej niż pozostałe
- W SQL Server Management Studio dodanie indeksów szczególnie przyspieszyło wykonanie zapytania 2 i 4
- W SQL Server Management Studio czas CPU był znacznie większy niż elapsed

6. Konfiguracja sprzętowa

- Procesor: Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz, 2112 MHz, Rdzenie: 4, Procesory logiczne: 8
- Windows 10 Home
- 64-bitowy system operacyjny, procesor x64
- SQL Server Management Studio 19
- pgAdmin 4

7. Bibliografia

- „Wydajność złączeń i zagnieźdzeń dla schematów znormalizowanych i zdenormalizowanych”, Łukasz Jajeńnica, Adam Piórkowski
- http://stareaneksy.pwn.pl/historia_ziemi/przyklady/?pokaz=tabela