

Bazy danych
Patrycja Sala
WGGiOŚ, Geoinformatyka

Analiza wydajności złączeń i zagnieżdżeń dla schematów znormalizowanych i zdenormalizowanych

1. Wprowadzenie

Celem analizy było zbadanie wydajności złączeń oraz zagnieżdżeń dla schematów znormalizowanych i zdenormalizowanych. Bazowano na artykule: *Wydajność złączeń i zagnieżdżeń dla schematów znormalizowanych i zdenormalizowanych*, Autorzy: Łukasz Jajeśnica, Adam Piórkowski, Akademia Górniczo – Hutnicza, Katedra Geoinformatyki i Informatyki Stosowanej.

2. Konfiguracja sprzętowa i programowa

Testy przeprowadzono na dwóch systemach zarządzania bazami danych, zainstalowanych na dwóch różnych komputerach.

PostgreSQL 13.2

CPU: 2,3 GHz Dwurdzeniowy procesor Intel Core i5

RAM: 8 GB 2133 MHz LPDDR3

Dysk: Macintosh HD

S.O.: MacOS Catalina 10.15.5

MS SQL Server 18

CPU: Intel(R) Core(TM) i7-4700MQ CPU @ 2.40GHz

RAM: 16 GB DDR3 Samsung

Dysk: HDD 1000GB

S.O.: Windows 10

3. Zapytania testowe

W celu przeprowadzenia testów stworzono tabelę geochronologiczną w wersji zdenormalizowanej i znormalizowanej oraz tabelę wypełnioną liczbami od 0 do 999 999. Zapytania wykonywano zarówno bez, jak i z nałożonymi indeksami.

3.1. Zapytanie 1 (1 ZL)

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,68)=(GeoTabela.id_pietro));
```

3.2. Zapytanie 2 (2 ZL)

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,68)=GeoPietro.id_pietro) NATURAL JOIN GeoEpoka  
NATURAL JOIN GeoOkres NATURAL JOIN GeoEra NATURAL JOIN GeoEon;
```

3.3. Zapytanie 3 (3 ZG)

Złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci zdenormalizowanej, wykonywane poprzez zagnieżdżenie skorelowane:

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

3.4. Zapytanie 4 (4 ZG)

Złączenie syntetycznej tablicy miliona wyników z tabelą geochronologiczną w postaci znormalizowanej, wykonywane poprzez zagnieżdżenie skorelowane, a zapytanie wewnętrzne jest złączeniem tabel poszczególnych jednostek geochronologicznych:

```
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);
```

3. Wyniki testów

Każdy test przeprowadzono 10 razy, wyniki skrajne pominięto.

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

PostgreSQL							
<i>bez indeksów</i>				<i>z indeksami</i>			
1 ZL	2 ZL	3 ZG	4 ZG	1 ZL	2 ZL	3 ZG	4 ZG
209	421	8795	154	156	667	10926	176
166	603	8522	168	163	492	9335	189
184	563	8485	180	170	655	9794	189
150	602	8815	158	146	700	9692	150
213	451	8912	166	164	699	9179	266
188	641	8669	163	151	636	10244	174
233	451	8571	157	159	620	9031	152
175	560	8598	177	149	553	9007	259
177	554	8564	166	153	587	10938	194
148	599	8690	167	171	481	11286	251
MS SQL Server							
<i>bez indeksów</i>				<i>z indeksami</i>			
1 ZL	2 ZL	3 ZG	4 ZG	1 ZL	2 ZL	3 ZG	4 ZG
50	58	21025	56	35	42	21684	50
38	39	21346	44	27	28	21752	30
54	44	21260	39	34	30	22122	26
48	39	21451	38	37	28	21447	25
41	38	20988	41	28	33	21747	30
38	34	21407	40	33	28	21532	45
39	43	21119	38	32	31	21702	31
35	37	21641	40	28	26	21732	27
39	40	21615	39	27	30	21728	27
39	39	21538	44	28	30	21730	28

Tabela 2. Średnie oraz minimalne czasy wykonania zapytań 1 ZL, 2 ZL, 3 ZG i 4 ZG [ms]

PostgreSQL								
	1 ZL		2 ZL		3 ZG		4 ZG	
	MIN	ŚR	MIN	ŚR	MIN	ŚR	MIN	ŚR
Bez indeksów	148	184	421	545	8485	8662	154	166
Z indeksami	146	158	481	609	9007	9943	150	200

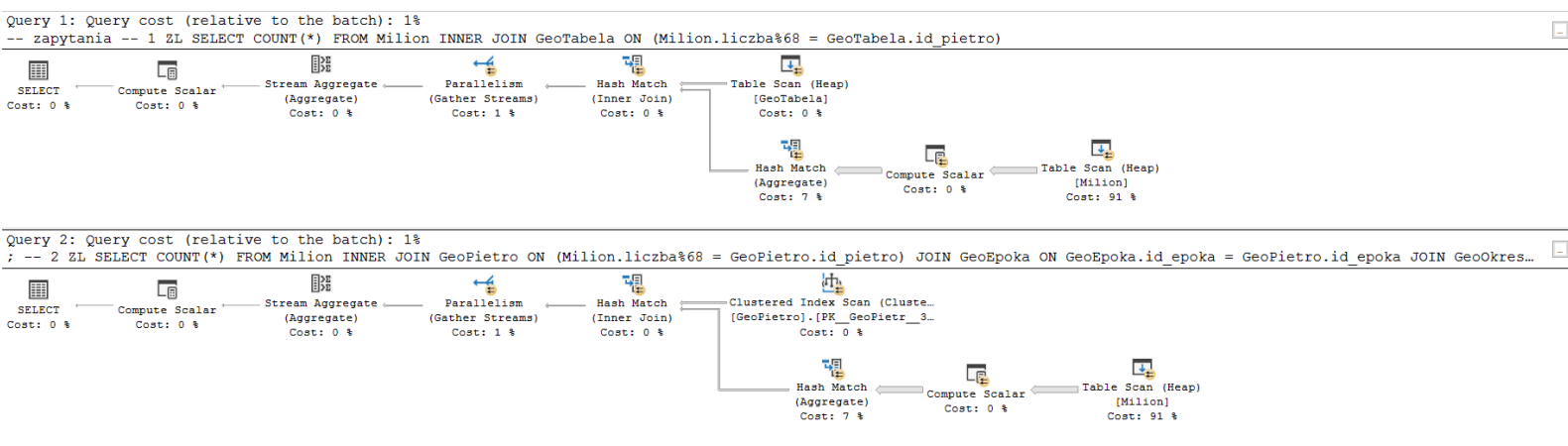
MS SQL Server								
	1 ZL		2 ZL		3 ZG		4 ZG	
	MIN	ŚR	MIN	ŚR	MIN	ŚR	MIN	ŚR
Bez indeksów	35	42	34	41	20988	21339	38	42
Z indeksami	27	31	26	31	21447	21718	25	32

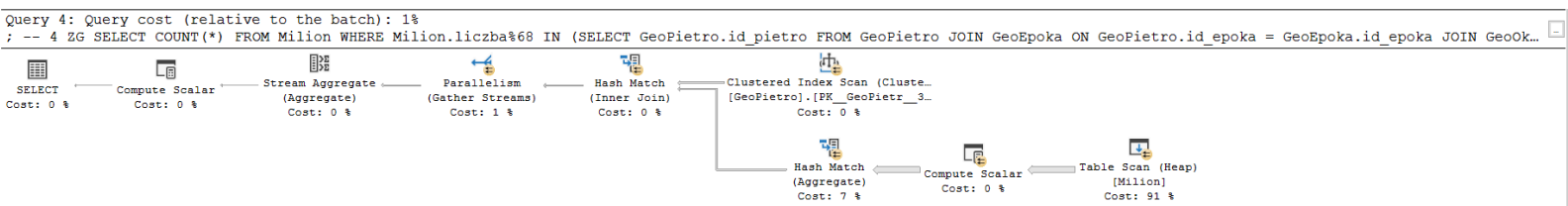
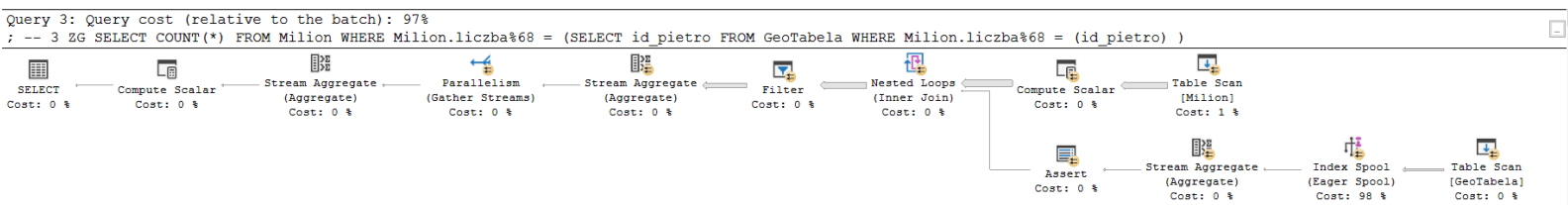
5. Wnioski

Postać zdenormalizowana jest wydajniejsza w większości przypadków. Jedyne przypadki, gdzie postać znormalizowana jest zdecydowanie szybsza to zagnieżdżenie skorelowane. Zapytanie zawierające zagnieżdżenie skorelowane z tabelą zdenormalizowaną jest dużo wolniejsze w wykonaniu niż pozostałe zapytania. Indeksacja w znaczący sposób nie poprawiła, ani nie pogorszyła czasu wykonywania zapytań w przypadku obu systemów zarządzania bazami danych.

6. Plany wykonania zapytań w MS SQL Server

6.1. Bez nałożonych indeksów





6.2. Z nałożonymi indeksami

