

WOJSKOWA AKADEMIA TECHNICZNA

Systemy baz danych

Sprawozdanie z zajęć laboratoryjnych

Rozproszone Bazy Danych

Wykonali: Mizera Joanna, Owczarek Monika

Grupa: I8E2S4

Data laboratorium: 21.05.2019

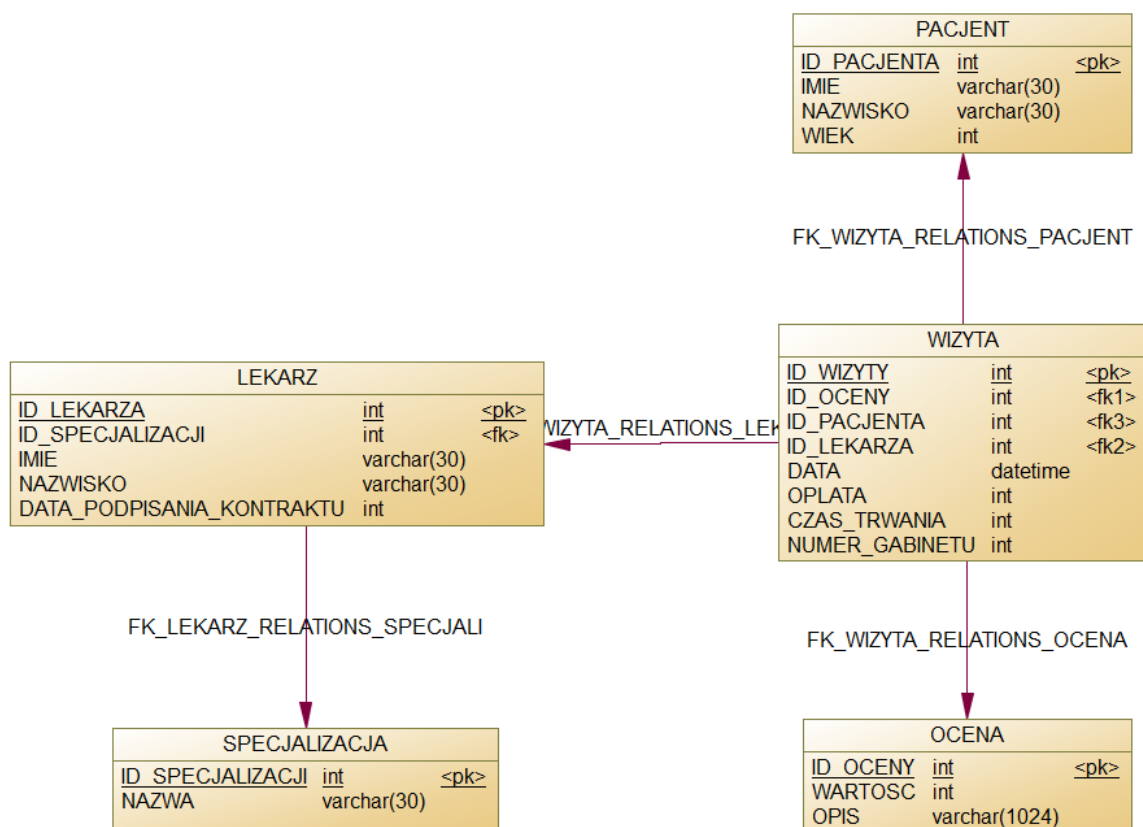
1. Opis modelu bazy relacyjnej

Baza danych przechowuje informacje dotyczące pacjentów, lekarzy, wizyt, specjalizacji lekarskich oraz wystawionych ocen na forum pewnej przychodni specjalistycznej. Pacjent może ocenić wizytę, w której uczestniczył, w skali od 1 do 5 gdzie:

- 1 – oznacza bardzo słaby poziom wizyty,
- 2 – oznacza słaby poziom wizyty,
- 3 – oznacza średni poziom wizyty,
- 4 – oznacza dobry poziom wizyty,
- 5 – oznacza bardzo dobry poziom wizyty.

Każdy lekarz pracujący w przychodni specjalistycznej posiada określoną specjalizację lekarską np. kardiolog, internista, laryngolog. W bazie ukazane zostały tylko te tabele, które są wykorzystane do wykonania polecenia laboratoryjnego.

2. Model bazy relacyjnej



3. Zapytanie

3.1. Opis zapytania

Zapytanie zwracające imię i nazwisko lekarza, nazwę jego specjalizacji, oraz liczbę przeprowadzonych przez niego wizyt z pacjentami poniżej 80 roku życia, których ocena ma wartość większą niż średnia wszystkich ocen.

3.2. Kod zapytania SQL

```
select sel1.DANE, sel1.NAZWA, count(*) as LICZBA_WIZYT from
(select
concat (l.IMIE, ' ', l.NAZWISKO) as DANE,
s.NAZWA as NAZWA,
o.WARTOSC as OCENA,
w.DATA as DATA_WIZYT
from
wizyta w
inner join OCENA o on w.ID_OCENY = o.ID_OCENY
inner join LEKARZ l on w.ID_LEKARZA = l.ID_LEKARZA
inner join PACJENT p on w.ID_PACJENTA = p.ID_PACJENTA
inner join SPECJALIZACJA s on l.ID_SPECJALIZACJI = s.ID_SPECJALIZACJI
where
p.WIEK < 80 and
o.WARTOSC >
(select
avg(o1.WARTOSC)
from
WIZYTA w1
inner join OCENA o1 on w1.ID_OCENY = o1.ID_OCENY)) as sel1
group by DANE, NAZWA
order by "LICZBA_WIZYT" desc;
```

3.3. Kod zapytania SQL z widokiem

```
create view
dbo.SREDNIA_VIEW WITH SCHEMABINDING
as
select avg(o.WARTOSC) AS SREDNIA
from dbo.WIZYTA w
inner join dbo.OCENA o on w.ID_OCENY = o.ID_OCENY;

select sel1."DANE", sel1."NAZWA", count(*) as "LICZBA_WIZYT" from
(select concat (l.IMIE, ' ', l.NAZWISKO) as DANE, s.NAZWA as NAZWA, o.WARTOSC as OCENA,
w.DATA as DATA_WIZYT
from wizyta w
inner join OCENA o on w.ID_OCENY = o.ID_OCENY
inner join LEKARZ l on w.ID_LEKARZA = l.ID_LEKARZA
inner join PACJENT p on w.ID_PACJENTA = p.ID_PACJENTA
inner join SPECJALIZACJA s on l.ID_SPECJALIZACJI = s.ID_SPECJALIZACJI

where o.WARTOSC > (select SREDNIA from SREDNIA_VIEW) and p.WIEK < 80
)as sel1
group by "DANE", "NAZWA"
order by "LICZBA_WIZYT" desc;
```

3.4. Kod zapytania w MongoDB

```
db.WIZYTA.aggregate([
  {"$match": {
    "$and": [{
      "pacjent.wiek": {
        "$lt": 80
      }
    }
  ]}, {
    "ocena.wartosc": {
      "$gt": db.WIZYTA.aggregate([
        {"$group": {
          "_id": null,
          "srednia": {
            "$avg": "$ocena.wartosc"
          }
        }
      ]
    ).map(function(u) { return u.srednia }))[0]
    }
  ]
}, {
  "$group": {
    "_id": {
      "Imie": "$lekarz.imie",
      "Nazwisko": "$lekarz.nazwisko",
      "Specjalizacja": "$lekarz.specjalizacja.nazwa"
    },
    "Wizyty": {
      "$sum": 1
    }
  }
}, {
  "$sort": {
    "Wizyty": -1
  }
}, {
  "$limit": 100
}])
```

4. Test dla 100 000 rekordów

3.1. Zapytanie bez użycia widoku

Query executed successfully. DELLMI00 (14.0 RTM) | DELLMI00\mi00j (54) | RBD | 00:00:00 | 634 rows

| DANE | NAZWA | LICZBA_WIZYT |
|----------------------|------------|--------------|
| Mery Blaszkiewicz | Onkolog | 219 |
| Isabella Bythway | Laryngolog | 203 |
| Stesha Isenor | Internista | 202 |
| Colleen Gonnely | Laryngolog | 195 |
| Celestina L'Episcopi | Onkolog | 185 |
| Marlowe Tailby | Onkolog | 182 |
| Fairleigh Lammiman | Laryngolog | 175 |
| Barnabe Madgett | Internista | 170 |
| Ware Chastel | Ginekolog | 168 |
| Erasmus Joire | Kardiolog | 161 |

Properties: Current connection parameters, Aggregate Status, Connection, Connection Details.

Wynik: 446 ms

3.2. Zapytanie z użyciem widoku

Query executed successfully. DELLMI00 (14.0 RTM) | DELLMI00\mi00j (55) | RBD | 00:00:00 | 634 rows

| DANE | NAZWA | LICZBA_WIZYT |
|----------------------|------------|--------------|
| Mery Blaszkiewicz | Onkolog | 219 |
| Isabella Bythway | Laryngolog | 203 |
| Stesha Isenor | Internista | 202 |
| Colleen Gonnely | Laryngolog | 195 |
| Celestina L'Episcopi | Onkolog | 185 |
| Marlowe Tailby | Onkolog | 182 |
| Fairleigh Lammiman | Laryngolog | 175 |
| Barnabe Madgett | Internista | 170 |
| Ware Chastel | Ginekolog | 168 |
| Erasmus Joire | Kardiolog | 161 |

Properties: Current connection parameters, Aggregate Status, Connection, Connection Details.

Wynik: 336 ms

3.3. Zapytanie z wykorzystaniem cache MS SQL Server

Query executed successfully. DELLMI00 (14.0 RTM) | DELLMI00\mi00j (54) | RBD | 00:00:00 | 634 rows

| | DANE | NAZWA | LICZBA_WIZYT |
|----|----------------------|------------|--------------|
| 1 | Mery Blaszkiewicz | Onkolog | 219 |
| 2 | Isabella Bythway | Laryngolog | 203 |
| 3 | Stesha Isenor | Internista | 202 |
| 4 | Colleen Gonnely | Laryngolog | 195 |
| 5 | Celestina L'Episcopi | Onkolog | 185 |
| 6 | Marlowe Talby | Onkolog | 182 |
| 7 | Fairleigh Lammiman | Laryngolog | 175 |
| 8 | Barnabe Madgett | Internista | 170 |
| 9 | Ware Chastel | Ginekolog | 168 |
| 10 | Erasmus Joire | Kardiolog | 161 |

Properties

Current connection parameters

Aggregate Status

Elapsed time: 00:00:00.1952755

Connection Details

Connection name: DELLMI00 (DELLMI00)

Connection details: Azure Active Dir, Connection elapsed: 00:00:00.1952755, Connection encrypted: Not encrypted, Connection finished: 05.06.2019 14:19:00, Connection row: 634, Connection started: 05.06.2019 14:19:00, Connection status: Open, Display name: DELLMI00, Login name: DELLMI00\mi00j, Server name: DELLMI00, Server version: 14.0.1000, Session Tracing: SPID 54

Wynik: 195 ms

3.4. Zapytanie w MongoDB

```
},
"responseLength" : 10192,
"protocol" : "op_msg",
"millis" : 217,
"planSummary" : "COLLSCAN",
"ts" : ISODate("2019-06-06T18:36:40.991Z"),
"client" : "127.0.0.1",
"appName" : "MongoDB Shell",
"allUsers" : [ ],
"user" : ""
```

4. Testy dla 250 000 rekordów

4.1. Zapytanie bez użycia widoku

Query executed successfully. DELLMI00 (14.0 RTM) | DELLMI00\mi00j (54) | RBD | 00:00:00 | 634 rows

| DANE | NAZWA | LICZBA_WIZYT |
|------|----------------------|----------------|
| 1 | Isabella Bythway | Laryngolog 545 |
| 2 | Mery Blaszkiewicz | Onkolog 505 |
| 3 | Stesha Isenor | Internista 503 |
| 4 | Marlowe Tailby | Onkolog 500 |
| 5 | Celestina L'Episcopi | Onkolog 461 |
| 6 | Colleen Gonnely | Laryngolog 449 |
| 7 | Barnabe Madgett | Internista 428 |
| 8 | Fairleigh Lammiman | Laryngolog 421 |
| 9 | Zondra McLurg | Laryngolog 421 |
| 10 | Jacinda Feldbau | Laryngolog 409 |

Properties: Current connection parameters, Aggregate Status (Elapsed time: 00:00:00.8272261), Connection (Connection name: DELLMI00 (DELLMI00\mi00j)), Connection Details (Azure Active Dir, Connection elapsed time: 00:00:00.8272261, Connection encrypted: Not encrypted, Connection finished: 05.06.2019 14:13:00, Connection rows: 634, Connection started: 05.06.2019 14:13:00, Connection status: Open, Display name: DELLMI00, Login name: DELLMI00\mi00j, Server name: DELLMI00, Server version: 14.0.1000, Session Tracing: SPID 54).

Wynik: 827 ms

4.2. Zapytanie z użyciem widoku

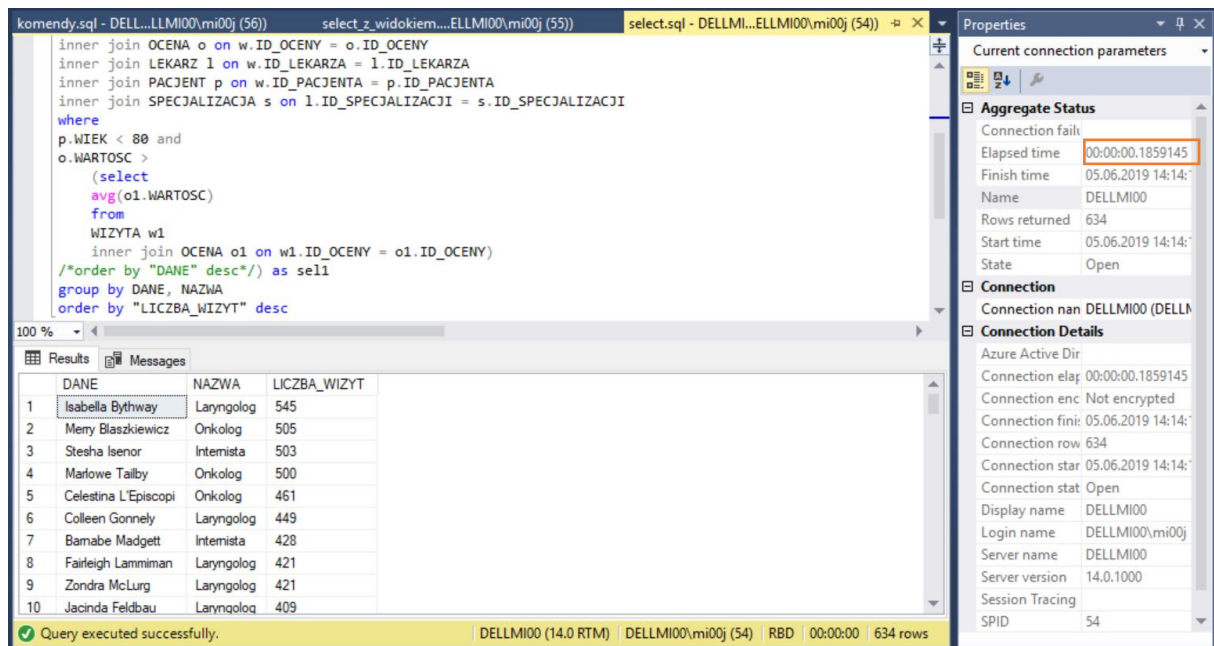
Query executed successfully. DELLMI00 (14.0 RTM) | DELLMI00\mi00j (55) | RBD | 00:00:00 | 634 rows

| DANE | NAZWA | LICZBA_WIZYT |
|------|----------------------|----------------|
| 1 | Isabella Bythway | Laryngolog 545 |
| 2 | Mery Blaszkiewicz | Onkolog 505 |
| 3 | Stesha Isenor | Internista 503 |
| 4 | Marlowe Tailby | Onkolog 500 |
| 5 | Celestina L'Episcopi | Onkolog 461 |
| 6 | Colleen Gonnely | Laryngolog 449 |
| 7 | Barnabe Madgett | Internista 428 |
| 8 | Fairleigh Lammiman | Laryngolog 421 |
| 9 | Zondra McLurg | Laryngolog 421 |
| 10 | Jacinda Feldbau | Laryngolog 409 |

Properties: Current connection parameters, Aggregate Status (Elapsed time: 00:00:00.7996085), Connection (Connection name: DELLMI00 (DELLMI00\mi00j)), Connection Details (Azure Active Dir, Connection elapsed time: 00:00:00.7996085, Connection encrypted: Not encrypted, Connection finished: 05.06.2019 14:30:00, Connection rows: 634, Connection started: 05.06.2019 14:30:00, Connection status: Open, Display name: DELLMI00, Login name: DELLMI00\mi00j, Server name: DELLMI00, Server version: 14.0.1000, Session Tracing: SPID 55).

Wynik: 799 ms

4.3. Zapytanie z wykorzystaniem cache MS SQL Server



The screenshot displays the SQL Server Enterprise Manager interface. The main window shows a query execution plan for a query named 'select_z_widokiem...'. The query is a complex join with multiple tables and a subquery. The results table shows 10 rows of data. The Properties pane on the right shows connection details and aggregate status.

Query Text:

```
inner join OCENA o on w.ID_OCENY = o.ID_OCENY
inner join LEKARZ l on w.ID_LEKARZA = l.ID_LEKARZA
inner join PACJENT p on w.ID_PACJENTA = p.ID_PACJENTA
inner join SPECJALIZACJA s on l.ID_SPECJALIZACJI = s.ID_SPECJALIZACJI
where
p.WIEK < 80 and
o.WARTOSC >
(select
avg(o1.WARTOSC)
from
WIZYTA w1
inner join OCENA o1 on w1.ID_OCENY = o1.ID_OCENY)
/*order by "DANE" desc*/ as sell
group by DANE, NAZWA
order by "LICZBA_WIZYT" desc
```

Results Table:

| | DANE | NAZWA | LICZBA_WIZYT |
|----|----------------------|------------|--------------|
| 1 | Isabella Bythway | Laryngolog | 545 |
| 2 | Mery Blaszkiewicz | Onkolog | 505 |
| 3 | Stesha Isenor | Internista | 503 |
| 4 | Marlowe Talby | Onkolog | 500 |
| 5 | Celestina L'Episcopi | Onkolog | 461 |
| 6 | Colleen Gonnely | Laryngolog | 449 |
| 7 | Barnabe Madgett | Internista | 428 |
| 8 | Fairleigh Lamminan | Laryngolog | 421 |
| 9 | Zondra McLurg | Laryngolog | 421 |
| 10 | Jacinda Feldbau | Laryngolog | 409 |

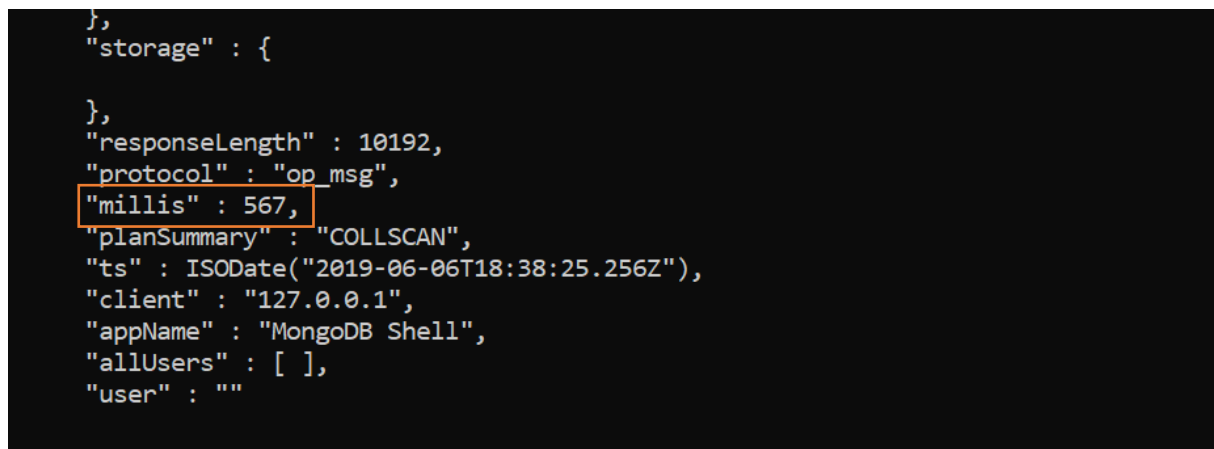
Properties Pane:

- Aggregate Status:** Connection failed, Elapsed time: 00:00:00.1859145, Finish time: 05.06.2019 14:14:14, Name: DELLMI00, Rows returned: 634, Start time: 05.06.2019 14:14:14, State: Open.
- Connection:** Connection name: DELLMI00 (DELLMI00\mi00j (54)).
- Connection Details:** Azure Active Dir, Connection elapsed: 00:00:00.1859145, Connection encrypted: Not encrypted, Connection finished: 05.06.2019 14:14:14, Connection row: 634, Connection started: 05.06.2019 14:14:14, Connection state: Open, Display name: DELLMI00, Login name: DELLMI00\mi00j, Server name: DELLMI00, Server version: 14.0.1000, Session Tracing: SPID 54.

Status Bar: Query executed successfully. DELLMI00 (14.0 RTM) | DELLMI00\mi00j (54) | RBD | 00:00:00 | 634 rows

Wynik: 185 ms

4.4. Zapytanie w MongoDB



The screenshot displays the MongoDB Shell output for a query. The result is a JSON object with various fields, including 'storage', 'responseLength', 'protocol', 'millis', 'planSummary', 'ts', 'client', 'appName', 'allUsers', and 'user'.

```
{
  "storage" : {
  },
  "responseLength" : 10192,
  "protocol" : "op_msg",
  "millis" : 567,
  "planSummary" : "COLLSCAN",
  "ts" : ISODate("2019-06-06T18:38:25.256Z"),
  "client" : "127.0.0.1",
  "appName" : "MongoDB Shell",
  "allUsers" : [ ],
  "user" : ""
}
```


5. Testy dla 500 000 rekordów

5.1. Zapytanie bez użycia widoku

Query executed successfully. DELLMI00 (14.0 RTM) | DELLMI00\mi00j (54) | RBD | 00:00:01 | 634 rows

| DANE | NAZWA | LICZBA_WIZYT |
|----------------------|------------|--------------|
| Isabella Bythway | Laryngolog | 1109 |
| Marlowe Talby | Onkolog | 1029 |
| Stesha Isenor | Internista | 1017 |
| Merry Blaszkiewicz | Onkolog | 999 |
| Celestina L'Episcopi | Onkolog | 929 |
| Colleen Gonnely | Laryngolog | 885 |
| Fairleigh Lammiman | Laryngolog | 860 |
| Jacinda Feldbau | Laryngolog | 854 |
| Zondra McLurg | Laryngolog | 845 |
| Barnabe Madgett | Internista | 821 |

Wynik: 1,218 s

5.2. Zapytanie z użyciem widoku

Query executed successfully. DELLMI00 (14.0 RTM) | DELLMI00\mi00j (55) | RBD | 00:00:00 | 634 rows

| DANE | NAZWA | LICZBA_WIZYT |
|----------------------|------------|--------------|
| Isabella Bythway | Laryngolog | 1109 |
| Marlowe Talby | Onkolog | 1029 |
| Stesha Isenor | Internista | 1017 |
| Merry Blaszkiewicz | Onkolog | 999 |
| Celestina L'Episcopi | Onkolog | 929 |
| Colleen Gonnely | Laryngolog | 885 |
| Fairleigh Lammiman | Laryngolog | 860 |
| Jacinda Feldbau | Laryngolog | 854 |
| Zondra McLurg | Laryngolog | 845 |
| Barnabe Madgett | Internista | 821 |

Wynik: 921 ms

5.3. Zapytanie z wykorzystaniem cache MS SQL Server

Query executed successfully. DELLMI00 (14.0 RTM) | DELLMI00\mi00j (54) | RBD | 00:00:00 | 634 rows

| | DANE | NAZWA | LICZBA_WIZYT |
|----|----------------------|------------|--------------|
| 1 | Isabella Bythway | Laryngolog | 1109 |
| 2 | Marlowe Talby | Onkolog | 1029 |
| 3 | Stesha Isenor | Internista | 1017 |
| 4 | Merry Blaszkiewicz | Onkolog | 999 |
| 5 | Celestina L'Episcopi | Onkolog | 929 |
| 6 | Colleen Gonnely | Laryngolog | 885 |
| 7 | Fairleigh Lammiman | Laryngolog | 860 |
| 8 | Jacinda Feldbau | Laryngolog | 854 |
| 9 | Zondra McLurg | Laryngolog | 845 |
| 10 | Bamabe Madgett | Internista | 821 |

Wynik: 267 ms

5.4. Zapytanie w MongoDB

```
}
},
"storage" : {
},
"responseLength" : 10192,
"protocol" : "op_msg",
"millis" : 885,
"planSummary" : "COLLSCAN",
"ts" : ISODate("2019-06-06T18:39:56.484Z"),
"client" : "127.0.0.1",
"appName" : "MongoDB Shell",
"allUsers" : [ ],
"user" : ""
}
```

6. Repozytorium

<https://github.com/mi00j/rbd-przychodnia.git>

7. Filmik

<https://youtu.be/Y0ADtGQHCU>

8. Podsumowanie

8.1. Uwagi

- Baza MS SQL Server po uruchomieniu zapytania zaczyna przechowywać cache. Gdy zapytanie jest uruchamiane po raz pierwszy w SQL Server, jest ono kompilowane i generowany jest plan dla zapytania. Każde zapytanie wymaga planu zapytania, zanim zostanie faktycznie wykonane. Jest on przechowywany w pamięci podręcznej planu kwerend SQL Server. W ten sposób po ponownym uruchomieniu danego zapytania program SQL Server nie musi tworzyć kolejnego planu. Używa raczej tego zbuforowanego, który poprawi wydajność bazy danych. Czas trwania planu kwerendy w pamięci podręcznej zależy od tego, jak często zapytanie jest wykonywane. Plany, które są używane częściej, pozostają w pamięci podręcznej na dłuższy czas i odwrotnie. W związku z tym czas przetwarzania zapytania nie odzwierciedla rzeczywistego czasu, który musi zostać poświęcony na jego wykonanie. Ręczne usunięcie cache również nie przynosiło oczekiwanych rezultatów.
Ponadto wyniki dla 100k, 250k i 500k z wykorzystaniem cache nie różnią się od siebie znacząco. W każdym z przypadków wartości przyjmowały zakres od około 120 ms do 280 ms.
- Przetwarzanie zapytań w bazie MS SQL Server zależy od bardzo wielu czynników. Wyniki różniły się od siebie znacząco w przypadku pracy laptopa na baterii (dłuższy czas) i pracy podczas ładowania (krótszy czas). Wpływ na wyniki ma również liczba otwartych programów i aplikacji.
- Liczba rekordów dla poszczególnych testów była tak naprawdę większa o 1000 rekordów w tabeli LEKARZ, 1000 rekordów w tabeli PACJENT, 5 rekordów w tabeli OCENA, 5 rekordów w tabeli SPECJALIZACJA. Samych rekordów w tabeli WIZYTA było odpowiednio 100 000, 250 000, 500 000.

8.2. Napotkane problemy i ich rozwiązania

Największym problemem było napisanie takiego zapytania dla bazy NoSQL, aby działało w sposób identyczny jak zapytanie SQL. Problem ten wynikał na pewno z faktu, że nie korzystaliśmy nigdy z baz danych typu MongoDB. Niestety utworzone zapytanie nie dało optymalnego wyniku. Przyczyną jest na pewno słaba znajomość składni oraz możliwości bazy. Wpływ na otrzymany czas ma również struktura dokumentów umieszczonych w naszej kolekcji zawierająca zagnieżdżone pola.

Problemem było również zaimportowanie plików csv do MongoDB. Baza traktowała każdą wartość jako string przez co nie było możliwości wykonania potrzebnych operacji tj. obliczenia średniej ocen. Rozwiązanie problemu nie było najłatwiejsze. W celu dodania danych do bazy utworzony został generator pliku json.

8.3. Pomysły

Baza MongoDB chociaż posiada oczywisty potencjał nie radzi sobie w wielu aspektach przede wszystkim związanych z importem danych. Dobrym pomysłem byłoby dopracowanie tych szczegółów aby praca z bazą NoSQL była wydajniejsza i przyjemniejsza.

8.4. Ocena środowisk

- **Microsoft SQL Server Management Studio 2017**

SQL Server Management Studio to darmowe zintegrowane środowisko do zarządzania dowolną infrastrukturą SQL. W tym przypadku jest to SQL Server. Zapewnia narzędzia do konfigurowania, monitorowania i administrowania bazami danych. Środowiska używa się aby mieć możliwość wdrożenia, monitorowania i aktualizowania komponentów warstwy danych używanych przez aplikacje, a także tworzenia zapytań i skryptów.

W przypadku realizacji tego zadania, środowisko to nie przysparzało żadnych problemów. Import danych z pliku csv przebiegł pomyślnie, dzięki czemu możliwe było szybkie przedstawienie wyników zapytań dla bardzo dużej liczby rekordów.

- **MongoDB**

MongoDB jest nierelacyjnym systemem zarządzania bazami danych. Charakteryzuje się dużą skalowalnością wydajnością i brakiem konkretnej struktury obsługiwanych baz danych. Dla użytkownika, który nigdy nie miał styczności z tego rodzaju bazą, obsługa jej nie należy do najprostszych. Dokumentacja webowa, chociaż dosyć obszerna i poruszająca wiele różnych tematów, nie odpowiedziała na większość pytań, które pojawiły się przy korzystaniu ze środowiska w ramach realizacji tego zadania. Przyzwyczajenie do relacyjnych baz danych sprawiło, że z trudnością przychodziło przestawienie się na działanie bazy MongoDB.