Porównanie bazy relacyjnej z bazą dokumentową MongoDB

# Wstęp

Realizowany Projekt ma na celu wskazanie różnić pomiędzy bazą relacyjną SQL, a bazą dokumentową. W tym celu została utworzona aplikacja testująca wydajność baz dla funkcji CRUD(od ang. create, read, update and delete) dla 10, 100, 1000, 10000 operacji.

|  |  |  |
| --- | --- | --- |
| **Działanie** | **Instrukcja SQL** | **MongoDB** |
| Create | INSERT | INSERT |
| Read (Retrieve) | SELECT | FIND |
| Update | UPDATE | UPDATE |
| Delete (Destroy) | DELETE | DELETE |

Aplikacja została napisana w języku C# przy wykorzystaniu frameworka ASP.NET 4.  
Do realizacji połączenia z bazą danych MongoDB wykorzystano oficjalny driver dostępny na oficjalnej stronie MongoDB.

# Terminologia i pojęcia

Wiele pojęci używanych w relacyjnych bazach danych SQL ma swoje odobiwedniki w MongoDB. Poniższa tabela przedstawia kilka z nich.

|  |  |  |
| --- | --- | --- |
| **MySQL** | **Instrukcja SQL** |  |
| Table(tabela) | Collection(kolekcja) |  |
| Row(rekord) | Document(dokument) |  |
| Column(kolumna) | Field(pole) |  |
| Joins(złączenia) | Embedded documents(dokumenty osadzone), linking |  |
|  |  |  |

# Charakterystyka MongoDB

Monogo jest otwartym, nierelacyjnym systemem zarządzania bazą danych udostępnionym na otwartej licencji GNU AGPL. Charakteryzuje się dużą skalowalnością, wydajnością oraz brakiem ściśle zdefiniowanej struktury obsługiwanych baz danych. Zamiast tego dane składowane są jako dokumenty w stylu JSON, co umożliwia aplikacjom bardziej naturalne ich przetwarzanie, przy zachowaniu możliwości tworzenia hierarchii oraz indeksowania. Doskonale sprawdza się w rozwiązaniach internetowych.

Zamiast przechowywać dane w serii rekordów i tabel, jak ma to miejsce w SQL, MongoDB zawiera kolekcje i dokumenty. Kolekcję można porównać do tabeli, jednak dokument różni się od rekordu. Dokumenty ze względu na to, że nie mają ściśle określonej struktury mogę przechowywać znacznie więcej informacji.

Załóżmy, że chcemy przechowywać informację o użytkownikach i ich rolacha.

W SQL będzie to wyglądać tak.

|  |  |
| --- | --- |
| **USERS** | |
| **user\_id** | **email** |
| 51 | xxx@onet.pl |
| 52 | yyy@onet.pl |
| **ROLES** | |
| **role\_id** | **name** |
| 1 | Admin |
| 2 | Moderator |
| **USER\_ROLES** | |
| **role\_id** | **user\_id** |
| 1 | 51 |
| 2 | 51 |
| 2 | 52 |

Jak widać ustalenie ról wymaga przeprowadzenia operacji złączeń, gdy w przypadku MongoDB sprowadza się do utworzenia dokumentu o następującej postaci:

1. {
2. "id" : ObjectId("5678cdef19875672av1456avasceg"),
3. "email" : xxx@onet.pl,
4. "roles" : ["admin", "operations"]
5. }

Podejście to może rodzić pewne wątpliwości co do na przykład nadmiarowości danych, albo możliwości aktualizacji nazwy roli np. z admin na administrator. W SQL głównym celem jest przygotowanie struktury danych w taki sposób, by uniknąć redundancji danych, więc modyfikacji danych sprowadzałaby się tylko do zmiany nazwy w rekordzie tabeli **ROLES**.

Jeżeli chodzi o MongoDB głównym celem jest ułatwienie użycia danych w aplikacji. W takim przypadku zmiana nazwy roli będzie wymagała znacznie większej ilości operacji. Z założenia jednak operacja aktualizacji danych przy tym podejściu ma być rzadką operacją.

Dane w MongoDB przechowywane są w postaci BSON(Binary JSON), który nieco różni się od JSON. Jedną z różnic jest wartość **ObjectId**, która jest typem używanym w MongoDB do unikatowej identyfikacji użytkowników(odpowiednik klucza podstawowego w SQL). Wartość ObjectId nie tylko jednoznacznieidentyfikujedokument ale na jej podstawie jesteśmy wstanie określić kiedy dokument został wygenerowany.

# Zalety formatu BSON

Zalety formatu BSON w porównaniu do tabel i rekordów:

* Naturalny sposób mapowania na obiekty w językach programowania
* Osadzone dokumenty i tablice eliminują konieczność stosowania złączeń
* Brak schematu danych przekładający się na ułatwienie polimorfizmu

# MongoDB skalowanie poziome i automatyczny sharding

MongoDB obsługuję skalowanie poziome , które pozwala na dodanie kolejnych maszyn do klastra i tym samym jego skalowanie. Jest to przeciwieństwo skalowania pionowego, które sprowadza się do zwiększania mocy obliczeniowej poszczególnych komputerów.

W przypadku baz SQL są ograniczone możliwości zwiększenia wydajności bazy. W grę wchodzą: analiza i optymalizacja zapytań, zwiększenie bufora, przeniesienie danych do innych źródeł, zmniejszenie poziomu użycia serwera podczas zapytań. Wszystkie te rozwiązania wymagają dużych nakładów czasowych związanych z pracą programistów.

Jedną z możliwości zwiększenia wydajności relacyjnej bazy danych jest zastosowanie tzw. Sharding sytemu, co oznacza podział danych między wiele baz danych. W przypadku relacyjnych baz danych jest to trudne przedsięwzięcie, dodatkowo wiążę się z możliwością wykonywania zapytań dotyczących całego zbioru danych. W przypadku MongoDB problem z shardingiem nie występuję, jednak istnieją pewne ograniczenia z nim związane.

# Realizacja połączenia w aplikacji.

1. var client = **new** MongoClient(**new** MongoClientSettings {
2. Server = **new** MongoServerAddress("localhost", 3979)
3. });
4. IMongoDatabase \_database = client.GetDatabase("dvproject");

# Przykładowy test.

1. **private** **float**[] MongoCreateTest(IMongoCollection<Crime> collection, **int**[] insertNumbers)
2. {
3. System.Diagnostics.Debug.WriteLine("\*\*\*\*\*\*\*\*\*\*CreateTest\*\*\*\*\*\*\*\*\*\*");
4. **float**[] results = **new** **float**[insertNumbers.Length];
5. **for** (**int** i = 0; i < insertNumbers.Length; i++)
6. {
7. var watch = System.Diagnostics.Stopwatch.StartNew();
8. mongoInsertOperation(collection, insertNumbers[i]);
9. watch.Stop();
10. var elapsedMs = watch.ElapsedMilliseconds;
11. results[i] = elapsedMs;
12. System.Diagnostics.Debug.WriteLine(insertNumbers[i] + "\t\t" + elapsedMs);
13. }
15. **return** results;
16. }
17. **private** **void** mongoInsertOperation(IMongoCollection<Crime> collection, **int** insertNumber)
18. {
19. **for** (var i = 0; i < insertNumber; i++)
20. {
21. collection.InsertOne(**new** Crime()
22. {
23. Id = ObjectId.GenerateNewId(),
24. DateRptd = "01/01/2016",
25. DrNo = 1,
26. DateOcc = "01/01/2016",
27. TimeOcc = 1,
28. Area = 1,
29. AreaName = "Cracov",
30. CrmCd = 1,
31. Rd = 1,
32. CrmCdDesc = "1",
33. Status = "fail",
34. StatusDesc = "fail",
35. Location = "Cracov",
36. CrossStreeet = "Blabla",
37. Location1 = "Cracov1"
38. });
39. }
40. }

# 8. Kod JavaSript odpowiedzialny za rysowanie wykresów.