

1. Cel

Celem niniejszej pracy jest przetestowanie oprogramowania służącego do mapowania dokumentowej bazy danych na bazę relacyjną, w celu jej graficznej analizy. Zbadany zostanie sposób działania komponentów niezbędnych do przeprowadzenia tego zadania, a przede wszystkim interakcje między nimi.

2. Analizowane produkty

Do przechowywania danych została użyta dokumentowa baza danych MongoDB. Ten nierelacyjny system zarządzania bazą danych zapewnia łatwą skalowalność i jest rozwiązaniem coraz bardziej popularnym. Dane w MongoDB składowane są w wielu dokumentach, zapisanych w formacie JSON. Format ten jest uniwersalny i pozwala na stosowanie niejednorodnej struktury danych; rekordy o tym samym typie danych mogą posiadać różne pola o dowolnej strukturze. Taki format składowania danych może być ciężki do interpretacji przez programy służące do ich wizualizacji. Najbardziej powszechnym i zarazem prostym w interpretacji formatem składowania danych jest model relacyjny, występujący np. w bazach danych typu SQL. BI Connector to narzędzie rozwijane przez Mongo, pozwalające na mapowanie dokumentowej bazy danych do bazy relacyjnej typu MySQL. Oprogramowanie to tworzy tymczasowy schemat relacyjnej bazy danych na podstawie bazy dokumentowej, który po połączeniu z bazą może być użyty przez programy służące do wizualizacji danych. Jako program służący do wizualizacji danych został użyty Tableau. Jest to desktopowa aplikacja, pozwalająca na graficzną wizualizację danych w sposób zdefiniowany przez użytkownika, co znacznie ułatwia ich zrozumienie. Umożliwia również odkrycie nowych zależności pomiędzy przechowywanymi danymi. Wszystko to w połączeniu z intuicyjnym interfejsem tworzy bardzo wygodne narzędzie dla Business Intelligence.

3. BI Connector - działanie.

W procesie mapowania danych dokumentowych na relacyjne, wykonywanym za pomocą BI connectora, można wydzielić dwa etapy. Pierwszym z nich jest stworzenie modelu relacyjnej bazy danych. Odpowiednie polecenie, odwołując się do wybranej bazy danych Mongo, tworzy plik schematu bazy danych o rozszerzeniu drdl. W pliku tym znajduje się nazwa bazy danych oraz wszystkie zawarte w niej kolekcje zmapowane na tabele. Każde pole obiektu w bazie Mongo jest mapowane na odpowiednią kolumnę danej tabeli. Wszystkie typy danych zostają zamienione na ich SQL'owe odpowiedniki (np. MongoType: string na SQLType: varchar). Dla właściwości obiektów zawierających tablice danych, zostają utworzone odpowiednio nazwane tabele, powiązane kluczem obcym z tabelą główną. Gwarantuje to niestratne mapowanie, przy zachowaniu spójności typów danych wewnątrz Mongo. Plik schematu jest w pełni edytowalny, jednak ingerencja użytkownika w jego strukturę może zaburzyć funkcjonowanie SQL'owej reprezentacji bazy dokumentowej. Polecenie to nie musi odwoływać się do całej bazy danych, dodając flagę -c możliwe jest mapowanie pojedynczej kolekcji. Użytkownik może także definiować nowe tabele, będące przefiltrowanymi widokami innych tabel za pomocą funkcji pipeline. Cała funkcjonalność opisana jest przejrzysto na stronie producenta. Drugi etap polega na uruchomieniu silnika, który na bieżąco rzutuje dane z Mongo, przez schemat *.drtl na bazę typu MySQL. Pozwala to na bieżące śledzenie zawartości bazy danych Mongo, przy zachowaniu pierwotnie zdefiniowanej struktury danych. Silnik ten pozwala na odwoływanie się do bazy danych typu Mongo w sposób analogiczny do odwoływania się do bazy typu MySQL. Umożliwione zostaje zdefiniowanie własnych

połączeń między tabelami, selekcja, filtrowanie krotek. Nie jest możliwa edycja, jak i usuwanie danych - narzędzie to umożliwia tylko projekcję danych.

4. Konfiguracja

Po pobraniu wszystkich pakietów instalacyjnych (Mongo, Tableau, BI connector), należy podążać zgodnie z wytycznymi podanymi na stronach producentów. Po pomyślnej instalacji zajmiemy się konfiguracją MongoDB. Zakładamy, że MongoDB zostało zainstalowane w folderze „C:\mongodb”. MongoDB potrzebuje folderu, w którym będą przechowywane wszystkie dane, dlatego będąc w folderze, w którym zainstalowane zostało Mongo, należy otworzyć konsolę oraz zatwierdzić komendę „md \data\db”. To polecenie utworzy potrzebny folder. Zawartość pobranego pakietu BI Connectora należy wypakować do folderu „C:\mongodb\bin”.

Uruchomienie MongoDB odbywa się poprzez zatwierdzenie w konsoli polecenia „C:\mongodb\bin\mongod.exe”. Tak zostaje uruchomiony główny proces bazy danych MongoDB, i jeśli wszystko przebiegło prawidłowo, na konsoli powinna pojawić się wiadomość „oczekiwanie na połączenia”. Należy pozostawić tę konsolę otwartą.

Następnie należy połączyć się z bazą danych jako klient. W nowej konsoli należy zatwierdzić polecenie „C:\mongodb\bin\mongo.exe”. W ten sposób zostaje utworzony proces klienta połączonego z bazą danych, który pozwala na dowolną interakcję z nią.

Następnym krokiem jest uruchomienie BI Connectora. W folderze bin należy wykonać polecenie „mongodrdl --host HOST -d dbname -o schema.drdl”, gdzie HOST to nazwa hosta, którą można sprawdzić w logach serwerach, a dbname to nazwa baza danych używanej w Mongo. Polecenie to tworzy schemat relacji na podstawie zawartości bazy danych MongoDB. Jest to fizyczny plik o rozszerzeniu .drdl, do którego następnie odwołuje się proces „pośrednika”. Aby ów proces uruchomić należy zatwierdzić polecenie „mongosqld.exe --schema schema.drdl --mongo-uri HOST”, gdzie schema.drdl to utworzony wcześniej plik ze schematem, a HOST to nazwa hosta. Po wykonaniu tych czynności możliwe jest analizowanie danych z bazy Mongo np. przez oprogramowanie Tableau.

Po uruchomieniu Tableau Desktop należy kliknąć Connect to a Server „MySQL”, w polu tekstowym podać IP serwera (domyślnie adres 127.0.0.1), które jest wyświetlone na konsoli po wykonaniu polecenia mongosqld oraz port (domyślnie 3307). Tak skonfigurowane środowisko pozwala na dalszą pracę z danymi za pomocą Tableau. Dane dostępne w Tableau są SQL'ową reprezentacją danych z bazy Mongo. Wszelkie modyfikacje w bazie Mongo skutkują natychmiastową reakcją BI Connectora, i co za tym idzie, zmianą wizualną wewnątrz Tableau.

5. Ocena mechanizmów, przeprowadzone testy.

Przeprowadzone zostały testy sprawdzające zachowanie BI Connectora, manipulując formatem i zawartością danych w MongoDB.

Pierwszą ważną uwagą jest fakt, że BI Connector pozwala na modyfikację i podgląd na żywo tylko zmian nie wpływających na strukturę bazy SQL. To znaczy, możliwa jest edycja wartości pól obiektów w bazie Mongo, a nie dodawanie nowych pól lub edycja już istniejących. Wynika to z faktu, że plik definiujący schemat bazy danych tworzony jest jednorazowo, przed uruchomieniem procesu pośredniczącego w komunikacji SQL - Mongo. Aby schemat danych widoczny wewnątrz Tableau był aktualny (po np. dodaniu nowej właściwości do obiektu), należy powtórzyć pierwszą część polecenia omawianego w pkt. 4 (tworzenie pliku schematu bazy *.drdl), a następnie na nowo uruchomić proces pośrednika.

Usunięcie lub zmiana nazwy pola w kolekcji Mongo powoduje błędne odwołania pliku *.drdl do pól obiektów.

```
db.mondial.cities.update({}, {$unset: {province:1}}, false, true);
```

```
db.mondial.cities.updateMany( {}, { $rename: { "population": "new_population" } } )
```

Sort fields Data source order Show aliases Show hidden fields 1000

mondial.cities _id	mondial.cities country	mondial.cities name	mondial.cities population	mondial.cities province	mondial.cities test_field
GR-Anatolikos Maked...	GR	Kavala	58 790,00	Anatolikos Makedonia...	null
GR-Anatolikos Maked...	GR	Komotini	null	Anatolikos Makedonia...	null
GR-Attikis-Athina	GR	Athina	664 046,00	Attikis	null
GR-Attikis-Peiraias	GR	Peiraias	163 688,00	Attikis	null
GR-Attikis-Peristeri	GR	Peristeri	139 981,00	Attikis	null
GR-Attikis-Acharnes	GR	Acharnes	106 943,00	Attikis	null
GR-Dytikis Elladas-Pa...	GR	Patra	213 984,00	Dytikis Elladas	null

mondial.cities _id	mondial.cities country	mondial.cities name	mondial.cities population	mondial.cities province	mondial.cities test_field
GR-Anatolikos Maked...	GR	Kavala	null	null	null
GR-Anatolikos Maked...	GR	Komotini	null	null	null
GR-Attikis-Athina	GR	Athina	null	null	null
GR-Attikis-Peiraias	GR	Peiraias	null	null	null
GR-Attikis-Peristeri	GR	Peristeri	null	null	null
GR-Attikis-Acharnes	GR	Acharnes	null	null	null
GR-Dytikis Elladas-Pa...	GR	Patra	null	null	null

Dodanie nowego pola do pojedynczego obiektu kolekcji, spowoduje dodanie tego pola z wartością *null* do każdego innego rekordu wewnątrz kolekcji widocznej przez Tableau. Zmiana ta jest widoczna dopiero po utworzeniu nowego pliku *.drdl.

```
db.mondial.countries.insert({
  name: "test_country3",
  code: "TST3",
  capital: "test_capital_3",
  area: 52103210,
  funny_facts: "this country is cool!"
})
```

Abc mondial.countries _id	# mondial.countries area	Abc mondial.countries capital	Abc mondial.countries code	Abc mondial.countries name
5956a48efd05c3706...	52 103 210,00	test_capital	TST	test_country
5956ae10fd05c3706...	52 103 210,00	test_capital_new	TST2	test_country2
581cb5a519ec2deb4...	17 075 200,00	R-Moscow-Moskva	R	Russia
581cb5a519ec2deb4...	9 976 140,00	CDN-Ontario-Ottawa	CDN	Canada

Abc mondial.countries _id	# mondial.countries area	Abc mondial.countries capital	Abc mondial.countries code	Abc mondial.countries funny_facts	Abc mondial.countries name
595a067a83a7a1d87...	52 103 210,00	test_capital_3	TST3	this country is cool!	test_country3
5956a48efd05c3706...	52 103 210,00	test_capital	TST	this country is cool!	test_country
5956ae10fd05c3706...	52 103 210,00	test_capital_new	TST2	this country is cool!	test_country2
581cb5a519ec2deb4...	17 075 200,00	R-Moscow-Moskva	R	<i>null</i>	Russia
581cb5a519ec2deb4...	9 976 140,00	CDN-Ontario-Ottawa	CDN	<i>null</i>	Canada

Po edycji wartości pola jakiegokolwiek obiektu, o ile jest ona prawidłowa; tzn. nowa wartość jest zgodna z typem danych tego pola, możliwy jest podgląd aktualnej tabeli po wymuszeniu aktualizacji w Tableau przyciskiem "Update Now". Dane, których pola posiadają wartości pól niezgodne z ich typem, nie są widoczne w tabeli. Uniemożliwiają również dalszą obróbkę danych wewnątrz Tableau (np. funkcję join).

```
db.mondial.countries.update (
{
  name: "test_country4",
},
{
  name: "test_country4",
  code: "TST4",
  capital: "test_capital_4",
  funny_facts: 1234,
  area: "duza"
}
)
```

Po takiej modyfikacji pola "area" państwa "test_country4" rekord przestaje być widoczny w tabeli mondial.countries wewnątrz Tableau.

Dodanie wartości pola w formacie tablicy skutkuje stworzeniem nowej tabeli, zawierającej klucz obcy tabeli głównej i nowe zdefiniowane właściwości.

```
db.mondial.countries.update (
{
  name: "test_country3",
},
{
  name: "test_country3",
  code: "TST3",
  capital: "test_capital_3",
  area: 52103210,
  funny_table: [
    "this country was invented by me",
    "for the test",
    "over"
  ]
}
)
```

Tabela mondial.countries_funny_table wygląda następująco:

Abc mondial.countries_funny_table _id	Abc mondial.countries_funny_table funny_table	# mondial.countries_funny_ta... funny_table_idx
595a067a83a7a1d87...	this country was invented by me	0
595a067a83a7a1d87...	for the test	1
595a067a83a7a1d87...	over	2

Łączenie tabel możliwe jest wewnątrz interfejsu Tableau. Definiowanie warunku połączeniowego jest zadaniem użytkownika (możliwe jest połączenie za pomocą tych samych wartości pól/zdefiniowanej funkcji zależności). Program potrafi też połączyć tabelę automatycznie, jeżeli obie posiadają to samo pole z odpowiednimi wartościami (np. "_id").

Dodanie nowego pola, którego wartość jest obiektem, skutkuje dodaniem wielu kolumn nazwanych obiekt.właściwość do rekordów wewnątrz tabeli. Rekordy nie zawierające obiektu w bazie Mongo, posiadają standardowo wartość *null*.

```

db.mondial.countries.update (
{
  name: "test_country4",
},
{
  name: "test_country4",
  code: "TST4",
  capital: "test_capital_4",
  funny_facts: "people eat only potatoes",
  area: 52103210,
  president: {
    name: "Magda K.",
    height: 186,
    hobbies: [
      "swimming",
      "diving"
    ]
  }
}
)

```

# mondial.countries area	Abc mondial.countries name	# mondial.countries president.height	Abc mondial.countries president.name
52 103 210,00	test_country4	186,000	Magda K.
52 103 210,00	test_country	<i>null</i>	<i>null</i>
52 103 210,00	test_country2	<i>null</i>	<i>null</i>
52 103 210,00	test_country3	<i>null</i>	<i>null</i>

Dla pola hobbies powstała tabela mondial.countries_president_hobbies, zawierająca _id country jako klucz obcy.

6. Podsumowanie

BI Connector spełnia swoją funkcjonalność jako narzędzie umożliwiające szybką translację bazy dokumentowej Mongo na bazę relacyjną MySQL. Dwufazowość przetwarzania danych przez oprogramowanie uniemożliwia natychmiastowy podgląd modyfikacji ingerujących w strukturę bazy danych. Zrozumienie podstaw działania BI Connectora znacznie ułatwia późniejszą pracę z nim.

Różnice w składowaniu danych pomiędzy wymienionymi schematami są na tyle duże, że nie jest możliwe idealne odzwierciedlenie struktur danych, wraz z ich modyfikacją na żywo. Rozwiązanie tego problemu zaproponowane przez BI Connector powinno zadowolić większość użytkowników.

Konrad Matusiak 117193,
Szymon Kasperski 117196