

Bazy danych NoSQL, ćwiczenie 3

temat: Baza Cassandra, podstawy pracy

autor: Artur Gramacki

1. Pobierz i zainstaluj oprogramowanie Cassandra:

- pobierz oprogramowanie ze strony
<http://www.apache.org/dyn/closer.lua/cassandra/3.11.4/apache-cassandra-3.11.4-bin.tar.gz>
- wypakuj zawartość pliku *tar.gz* do wybranego przez siebie katalogu, np *C:\cassandra*,
- pobierz i zainstaluj interpreter języka Python 2.7 (wymagany do pracy Cassandra), strona:
<http://portablepython.com/wiki/PortablePython2.7.6.1> i wypakuj do dowolnego katalogu.

2. Ustaw wymagane zmienne środowiskowe. Wymagane jest Java w wersji 8. Przykładowe ustawienie pokazujemy poniżej. Dopiero po ustawieniu tych zmiennych możliwe będzie uruchomienie w oknie poleceń serwera Cassandra.

```
set JAVA_HOME=c:\Program Files\Java\jre1.8.0_191
set PATH=c:\Programy\Portable_Python_2.7.6.1\AppData
set PATH=%PATH%;%SystemRoot%\system32\WindowsPowerShell\v1.0
```

3. W oddzielnym oknie wiersza poleceń również ustaw zmienne *JAVA_HOME* i *HOME*, a następnie uruchom narzędzie klienta *cqlsh* i połącz się z uruchomionym serwerem Cassandra. Po wykonaniu kilku ćwiczeń z wykorzystaniem programu klienta *cqlsh* możesz przejść do pracy z bardziej wygodnymi graficznymi narzędziami (np. *TablePlus* lub *RazorSQL*. To ostatnie oprogramowanie jest 30-dniowym trialem).

4. Następujące tutoriale mogą być pomocny w czasie wykonywania ćwiczenia:
<https://www.guru99.com/cassandra-data-types-expiration-tutorial.html>
oraz
<https://www.tutorialspoint.com/cassandra/index.htm>. Oczywiście różnych tutoriali w internecie jest „zatrzęsienie”, wybierz taki, który będzie Tobie najbardziej pasował.

5. Wykonaj następujące polecenia:

- utwórz nową przestrzeń kluczy o nazwie *filmoteka*,
- w przestrzeni w przestrzeni kluczy *filmoteka* utwórz rodzinę kolumn (tabelę) o nazwie *filmy*. Samodzielnie ustal sensowne kolumny. Pamiętaj o zdefiniowaniu klucza głównego. Wzoruj się na zbiorze danych, który był używany w instrukcji dotyczącej bazy MongoDB. Dane dostępne są tutaj w wersji kompletnej
<http://staff.uz.zgora.pl/agramack/files/BazyDanych/NoSQL/entities.7z>
a tutaj
<http://staff.uz.zgora.pl/agramack/files/BazyDanych/NoSQL/simple-entities.json>
w wersji okrojonej (wygodniejszej do testów początkowych). Do rodziny kolumn wstaw około 10 wierszy,
- w przestrzeni kluczy *filmoteka* utwórz rodzinę kolumn (tabelę) o nazwie *reżyserzy* zawierającą kolumny: *reżyser_ID* (klucz główny), *imie*, *nazwisko*, *rok urodzenia* (ew. kilka innych kolumn, według uznania). Do rodziny kolumn *reżyserzy* wstaw stosowne dane,
- wyświetl strukturę rodziny kolumn *filmoteka* (polecenie DESCRIBE),
- wyświetl raport w następującym układzie:

```
film_ID tytuł_filmu reżyser_ID reżyser_imie reżyser_nazwisko,
reżyser_rok_urodzenia
```

- wykonaj kilka przykładowych poleceń INSERT, UPDATE, DELETE. Używaj również klauzuli WHERE.
 - w bazie Cassandra zdefiniowano również polecenie UPSERT. Na odpowiednich przykładach wyjaśnij działanie tego polecenia. Zauważmy, że w klasycznym języku SQL w bazach relacyjnych takie polecenie nie występuje.
 - zademonstruj działanie funkcjonalności *Time to Live (ttl)*,
 - zdefiniuj dowolną tabelę, gdzie użyjesz możliwie jak największej ilości różnych wspieranych przez Cassandrę typów danych. Wpisz 3 przykładowe (i sensowne) wiersze (<http://cassandra.apache.org/doc/latest/cql/types.html>),
 - usuń przestrzeń kluczy *filmoteka*.
6. Wykonaj następujące ćwiczenia dotyczące tzw. *kolekcji*. Pamiętaj, że kolekcji używamy do przechowywania małych zbiorów danych, np. telefony czy maile jednej osoby. Dla zawartości, które mogą się w nieskończoność rozrastać, np. lista komentarzy użytkowników, kolekcje nie są dobrym rozwiązaniem. Kolekcja nie może przechowywać więcej niż 64kB danych. Próba przechowania w kolekcji więcej niż 64kB danych skończy się tym, że te dane zostaną utracone.
- utwórz nową przestrzeń kluczy o nazwie *kolekcje*,
 - Utwórz po kilka przykładowych kolekcji typu:
 - a) *set*,
 - b) *list*,
 - c) *map*.
7. Zdefiniuj własny typ danych i wstaw tam kilka przykładowych wierszy. Możesz wzorować się na przykładzie podanym na wykładzie.
8. Utwórz klaster składający się z 3 węzłów (np. 3 różne komputery w laboratorium). Możesz wzorować się np. na tutorialu <https://www.jamescoyle.net/how-to/2448-create-a-simple-cassandra-cluster-with-3-nodes>. Pamiętaj o tym, że zanim zaczniesz „spinać” komputery w klaster, to na żadnym węźle (instancji Cassandra) nie może być danych. Gdy jakieś dane są, to trzeba wyczyścić na każdym węźle katalog */data/system*. Na każdym węźle w określony sposób należy zmodyfikować pliki konfiguracyjne *cassandra.yaml* oraz *cassandra-rackdc.properties*. Należy też usunąć z każdego węzła plik *cassandra-topology.properties*, gdyż będzie na „przeszkadzał” (w sprawozdaniu wyjaśnij dlaczego).