



POLITECHNIKA WARSZAWSKA
Wydział Elektroniki i Technik Informacyjnych
Instytut Informatyki

Rok akademicki 2013/2014

PRACA DYPLOMOWA MAGISTERSKA

Jakub Turek

[TYTUŁ]

Praca wykonana pod kierunkiem
dra inż. Jakuba Koperwasa

Ocena:

.....

*Podpis Przewodniczącego Komisji
Egzaminu Dyplomowego*

Kierunek: Informatyka
Specjalność: Inżynieria Systemów Informatycznych
Data urodzenia: 1990.01.09
Data rozpoczęcia studiów: 2013.02.20

Życiorys

Urodziłem się 9 stycznia 1990 roku w Łodzi. W 1997 roku rozpocząłem edukację w Szkole Podstawowej nr 7 w Łodzi. W latach 2003-2006 kontynuowałem naukę w Gimnazjum nr 42 im. Władysława Stanisława Reymonta w Łodzi. Od 2006 roku uczyłem się w Liceum Ogólnokształcącym nr 31 im. Ludwika Zamenhofs w Łodzi. W 2009 roku zdałem egzaminy maturalne i ukończyłem szkołę licealną z wyróżnieniem. W latach 2009-2013 studiowałem dziennie informatykę na Wydziale Elektroniki i Technik Informatycznych Politechniki Warszawskiej. Ukończyłem studia z wynikiem celującym i odebrałem tytuł zawodowy inżyniera. Obecnie kończę Pracę Dyplomową Magisterską pod kierownictwem Instytutu Informatyki. We wrześniu 2012 roku rozpocząłem pracę zawodową jako programista aplikacji do zarządzania procesami biznesowymi oraz aplikacji mobilnych w firmie Xentivo, gdzie pracuję do dziś. Moją pasją jest tworzenie aplikacji mobilnych oraz internetowych, które uruchamiane są w środowisku iOS.

.....

Podpis studenta

Egzamin dyplomowy:

Złożył egzamin dyplomowy w dniu:

z wynikiem:

Ogólny wynik studiów:

Dodatkowe uwagi i wnioski Komisji:

.....

Streszczenie

Celem Pracy Dyplomowej jest stworzenie interfejsu programowania aplikacji umożliwiającego efektywne rozwiązywanie problemów w modelowaniu dziedziny danych opartych o bazę danych Apache Cassandra.

Client-server Augmented Reality applications framework for Android system

Summary

The goal of this thesis is to create a framework supporting the development of multiuser applications for Android system. The framework should consist of a client-server bus, as well as several Augmented Reality components. An additional goal is to explore the possibility of adapting commonly used programming practises dedicated to large projects during Android applications development. All the goals have been fully accomplished - the output is the implementation of the client-server bus as well as AR components allowing to track the device's geographic coordinates and to render a stable three-dimensional graphics on the display of the device. In order to demonstrate the features of the framework, a sample multiplayer game has been created. The thesis includes a description of the created components' design process, as well as their functionality and structure.

Spis treści

1	Wstęp	4
2	Apache Cassandra	5
2.1	Model danych	6
2.2	Dystrybucja danych	6
2.3	Konsekwencje dla modelowania dziedziny	7
2.4	CQL	9
2.5	Modelowanie - wzorce i antywzorce	10
3	Mapowanie obiektowo-relacyjne	11
3.1	Mapowanie obiektowo-relacyjne dla Cassandra	11
3.2	Wnioski z badania wydajności	14
3.3	Mapowanie obiektowe dla Cassandra - koncepcja	16
4	Modelowanie obiektowe dla Cassandra	18
4.1	OMC - podstawowe pojęcia	19
4.2	Lista życzeń użytkownika - studium przypadku	20
4.3	Wsparcie dla wzorców modelowania	22
4.4	Przetwarzanie partiami	28
4.5	Wsparcie dla liczników	29

Rozdział 1

Wstęp

Tu będzie wstęp

Rozdział 2

Apache Cassandra

Apache Cassandra jest bazą danych NoSQL¹, która powstała w wyniku połączenia rozwiązań wykorzystywanych w Dynamo² oraz BigTable³. Cassandra początkowo była rozwijana dla potrzeb portalu społecznościowego Facebook. Baza danych powstała z myślą o rozwiązaniu problemu pełnotekstowego przeszukiwania skrzynek odbiorczych użytkowników, w których dziennie zapisywane były miliardy wiadomości. Głównym celem, do których dążyli twórcy Cassandra była możliwość wykorzystania jej do przechowywania ogromnych ilości danych w bardzo rozproszonym środowisku, gdzie awarie pojedynczych węzłów zdarzają się na porządku dziennym. W tych warunkach baza danych musi zapewniać szybki i niezawodny dostęp do danych. [2]

Apache Cassandra wykorzystywana jest w wielu serwisach na całym świecie. Najbardziej znaczące przykłady użycia produkcyjnego to eBay⁴, Instagram⁵ oraz GitHub⁶. Największa światowa instalacja Cassandra obejmuje około 15000 węzłów, na których przechowywane jest łącznie ponad 4 petabajty danych. [3]

W przeciwieństwie do relacyjnych baz danych, Apache Cassandra nie zapewnia wsparcia dla reguły ACID⁷. Zamiast tego zostały zrealizowane postulaty twierdzenia CAP: „we współdzielonym systemie plików można zachować maksymalnie dwie z trzech właściwości: spójności, dostępności oraz podatności na partycjonowanie”. [4] Apache Cassandra priorytetyzuje właściwości dostępności oraz partycjonowania. Spójność danych jest odwrotnie proporcjonalna i może być regulowana w zależności od czasu odpowiedzi. Wysoka spójność danych oznacza wolniejszą odpowiedź bazy.

¹NoSQL (ang. Not Only SQL) - podzbiór baz danych, które zapewniają inne sposoby modelowania dziedziny niż tradycyjny model oparty na tabelach i relacjach.

²Amazon DynamoDB - zdecentralizowana, wysoce skalowalna baza danych typu klucz-wartość.

³Google BigTable - rozproszony system bazodanowy, który dobrze skaluje się dla ogromnych ilości danych.

⁴eBay - największy portal z aukcjami internetowymi na świecie

⁵Instagram - portal pozwalający na umieszczanie fotografii.

⁶GitHub - usługa pozwalająca na przechowywanie i wersjonowanie kodu źródłowego aplikacji.

⁷ACID (ang. Atomic, Consistency, Isolation, Durability) - zasada atomowości, spójności, izolacji i trwałości, które gwarantują poprawne przetwarzanie transakcji w bazach danych.

2.1 Model danych

Model danych Apache Cassandra jest analogiczny do BigTable. [1] Można przedstawić go jako dwuwymiarową mapę trójek wartości:

`Map<RowKey, Map<ColumnKey, Triple<Value, Timestamp, TTL>>>`

gdzie `RowKey` to identyfikator wiersza, `ColumnKey` to identyfikator kolumny, `Value` to wartość komórki, `Timestamp` to czas aktualizacji komórki, a `TTL` to czas życia danej wartości. [5] Na rysunku 2.1 przedstawiona jest schematyczna ilustracja wiersza danych. Pogrubiona wartość w lewej komórce to klucz wiersza, natomiast wyróżnione wartości w pierwszym wierszu oznaczają klucze poszczególnych kolumn. Każda komórka składa się z trzech wartości: wartości, czasu życia oraz „odcisku czasu”.

	ABC	DEF	...	XYZ
123	test value	another test value	...	not a test value
456	20	∞	...	∞
	1291987837942000	1291980736812000	...	1291980736212000

Rysunek 2.1: Przykładowy wiersz modelu danych o identyfikatorze 123456. Wartość komórki (123456, DEF) to „another test value”.

2.2 Dystrybucja danych

Do dystrybucji danych wykorzystywana jest funkcja skrótu, która zachowuje kolejność elementów. Węzły są przedstawiane na w topologii pierścienia. Algorytm dystrybucji zostanie omówiony na przykładzie ze schematu 2.2:



Rysunek 2.2: Schematyczna ilustracja dystrybucji danych w bazie danych Apache Cassandra.

1. Każdemu z węzłów $\{A, B, C, D, E, F\}$ przypisywany jest token, który zawiera się w zakresie wartości przyjmowanych przez funkcję skrótu. Strategię wyboru tokenu można konfigurować. Przykładową strategią jest wybór losowy. W omawianym przykładzie węzłom zostały przypisane tokeny o wartościach $\{-16, -9, -3, 4, 9, 17\}$.
2. Użytkownik bazy danych przesyła żądanie do dowolnego węzła, który pełni funkcję koordynatora dla danej operacji. Koordynator nadzoruje wpisanie danych do odpowiednich węzłów. W omawianym przykładzie rolę koordynatora pełni węzeł E .
3. Każdy węzeł przechowuje dane, których funkcja skrótu zawiera się w przedziale $(token_{n-1}, token_n]$, gdzie n to numer kolejny węzła rosnący zgodnie z ruchem wskazówek zegara. W przykładzie węzeł C przechowuje wiersze o wartościach funkcji skrótu z przedziału $(-9, -3]$, natomiast węzeł D z przedziału $(-3, 4]$. Wartości funkcji obliczane są cyklicznie, stąd węzeł A przechowuje wiersze o skrócie z przedziału $(-\infty, -16] \cup (17, \infty)$. W przykładzie wiersz o kluczu z funkcją skrótu wartości -10 zostanie utrwalony na węźle B .
4. Dane replikowane są na n węzłach, gdzie n to wartość konfigurowalnego współczynnika replikacji. Poza węzłem macierzystym (wyznaczanym w punkcie 3 algorytmu) dane są replikowane na $n - 1$ kolejnych (zgodnie z ruchem wskazówek zegara) węzłach. W omawianym przykładzie dane zostaną zreplikowane na węzłach C i D .

2.3 Konsekwencje dla modelowania dziedziny

Wewnętrzna struktura i mechanizm dystrybucji danych wykorzystywany w Apache Cassandra zmienia podejście do modelowania dziedziny znane z relacyjnych baz danych. Zbudowanie efektywnego modelu danych Cassandra wymaga skupienia się w podobnym stopniu na zdefiniowaniu encji z modelowanego świata, jak również na analizie odwołań, które będą wykonywane do obiektów z tego świata. [6]

Założmy, że celem jest modelowanie danych dla sklepu internetowego. Zakupów dokonują użytkownicy, którzy mogą wstawić wiele przedmiotów z oferty sklepu na listę życzeń. W przypadku baz opartych o język SQL jest to klasyczny problem relacji typu wiele-do-wielu, do modelowania których wykorzystywana jest najczęściej tabela pośrednia.



Rysunek 2.3: Modelowanie listy życzeń w relacyjnej bazie danych.

Diagram prezentujący zamodelowaną relację dla listy życzeń jest przedstawiony na rysunku 2.3. W tabeli Użytkownik (*User*) przechowywane są imię, nazwisko oraz identyfikator. W tabeli Przedmiot (*Item*) znajduje się nazwa, cena, a także inne właściwości: opis, kategoria oraz waga. Tabela Lista życzeń (*Wishlist*) łączy ze sobą użytkownika i przedmiot poprzez wykorzystanie kluczy obcych.

Powyższy model jest wykorzystywany w widoku listy życzeń na profilu użytkownika. Na liście życzeń prezentowane są informacje o nazwie przedmiotu oraz jego cenie. Po kliknięciu nazwy użytkownik przenoszony jest do strony przedmiotu. Na listingu 2.4 zaprezentowano zapytanie, które wyświetla listę życzeń.

```
SELECT item.name, item.price
FROM Item item, Wishlist wishlist
WHERE wishlist.userId = 202;
```

Rysunek 2.4: Zapytanie, które pobiera wszystkie przedmioty na liście życzeń użytkownika o identyfikatorze 202.

Cassandra umożliwia utworzenie dokładnej repliki relacyjnego modelu danych. Zostało to przedstawione na rysunku 2.5.

User		name	surname	Wishlist		userId	itemId
	123	Janusz	Kowalski		51	123	579
		name	surname			userId	itemId
	124	Marcin	Nowak		52	124	232

Item		name	price	desc	category	weight
	232	Master Chef	20.34	Cooking recipes	BOOKS	0.2
		name	price	desc	category	weight
	579	Seat Hit	159.99	Wooden armchair	FURNITURE	10.8

Rysunek 2.5: Wynik błędnego przeniesienia relacyjnego modelu danych do Cassandra.

Taki model jest jednak niepoprawny. Nie umożliwia on filtrowania zawartości listy życzeń po identyfikatorze użytkownika. Wynika to z faktu, że pobranie odpowiednich wierszy listy życzeń wymaga znajomości ich identyfikatorów, podczas gdy widok dysponuje wyłącznie odniesieniem do użytkownika. Błąd ten można łatwo naprawić zastępując encję *Wishlist* encją *WishlistByUser*, co przedstawia diagram 2.6.

WishlistByUser		579
	123	-
		232
	124	-

Rysunek 2.6: Definicja encji listy życzeń umożliwiająca filtrowanie względem użytkownika.

Poprawiony model można poddać dalszej optymalizacji. Wyświetlenie listy życzeń użytkownika wymaga odwołania do encji *Item*, w której znajdują się informacje o nazwie i cenie przedmiotu. Ponieważ przedmioty mogą być rozmieszczone na różnych węzłach, silnik Cassandra nie może wykonać złączenia w sposób optymalny - zapytanie o każdą pozycję listy życzeń jest wykonywane osobno. W celu przyspieszenia wykonywania operacji należy wykonać denormalizację. Dołączając do encji *WishlistByUser* informacje o nazwie i cenie produktu można uniknąć wykonywania kosztownych złączeń. Pozostałe dane przedmiotu zostaną pobrane dopiero po przejściu na jego stronę. Efekt denormalizacji jest przedstawiony na diagramie 2.7.

WishlistByUser	123	579
		(„Master Chef”, 20.43)
	124	232
		(„Seat Hit”, 159.99)

Rysunek 2.7: Zdenormalizowana postać listy życzeń.

2.4 CQL

Efektywne modelowanie i obsługa danych w Apache Cassandra wymaga dobrej znajomości wewnętrznej struktury bazy danych. Dodatkowym utrudnieniem przy korzystaniu z początkowych wersji Cassandra była konieczność wykorzystania skomplikowanego interfejsu programistycznego opartego o wywoływanie zdalnych procedur Thrift⁸. Thrift jest platformą pozwalającą budować aplikacje przenośne między różnymi językami programowania. Dzięki temu rozwiązaniu baza danych dostępna była dla wszystkich platform. Niestety, skutkiem ubocznym było skomplikowanie interfejsu dostępowego.

Wraz z wydaniem 1.2 Apache Cassandra wprowadzony został nowy interfejs dostępu do tej bazy danych. Interfejs ten nosi nazwę CQL⁹ i jest językiem zapytań, którego składnia wzorowana jest na SQL. Poza podobieństwami składniowymi języki te nie mają cech wspólnych. Nie są wzajemnie zgodne. W chwili obecnej CQL jest rekomendowanym standardem komunikacji z Apache Cassandra. [7] Na listingu 2.8 przedstawiono zapytanie w języku CQL, które opisuje omawianą wcześniej encję *User*. Wynikiem wykonania tego zapytania jest prosty schemat modelu danych zaprezentowany na diagramie 2.9.

```
CREATE TABLE User (
    userId uuid PRIMARY KEY,
    name text,
    surname text);
```

Rysunek 2.8: Zapytanie CQL, które tworzy encję *User*.

⁸Dokumentacja interfejsu dostępna jest pod adresem <https://wiki.apache.org/cassandra/API10>.

⁹CQL (ang. Cassandra Query Language) - język zapytań Cassandra.

User		name	surname
	userId	null	null

Rysunek 2.9: Wynik wykonania zapytania 2.8.

2.5 Modelowanie - wzorce i antywzorce

Pomimo, że CQL znacząco upraszcza modelowanie w Cassandrze stworzenie efektywnego schematu danych nie jest zadaniem prostym. Aby ułatwić ten proces twórcy i użytkownicy Cassandry zaczęli opisywać wzorce i antywzorce modelowania oraz dostępu do danych. Pełnią one analogiczną rolę do wzorców i antywzorców projektowych znanych z programowania. Na ich opis składa się możliwie ogólna definicja problemu, a także poprawny (lub niepoprawny) sposób jego rozwiązania.

Przykładem antywzorca modelowania jest kolejka oparta na kolumnach. [8] Kolejka to struktura danych, w której ilości wykonywanych operacji wstawiania, usuwania i odczytu są do siebie zbliżone. W przypadku Cassandry usuwanie kolumn z wiersza nie jest wykonywane natychmiast po odebraniu żądania. Zamiast tego usunięta kolumna jest oznaczana specjalnym znacznikiem *tombstone* i fizycznie usuwana dopiero po upływie pewnego czasu. Takie działanie przyspiesza znacząco operację usuwania kosztem operacji odczytu. Kiedy w wierszu występuje wiele znaczników *tombstone* operacje filtrowania zakresu kolumn wykonywane są znacznie wolniej.

Przykładem wzorca dostępu do danych jest unikanie konfliktów synchronizacji poprzez uaktualnianie wyłącznie zmodyfikowanych wartości. [9] Encja *Item* z diagramu 2.5 ma 5 właściwości. Wyświetlenie ekranu aktualizacji przedmiotu wymaga pobrania zawartości całego wiersza z bazy danych. Wzorzec stanowi, że jeżeli na tym ekranie zostanie zmieniony wyłącznie opis to do Cassandry należy przesłać żądanie uaktualniające zawartość wyłącznie jednej komórki - *desc*. Pozostałe wartości z formularza mogą być już nieaktualne. Przesłanie kompletu informacji poskutkowałoby nadpisaniem aktualnych wartości. Postępowanie według wzorca pozwala wykorzystywać mechanizm rozwiązywania konfliktów wbudowany w Cassandrę.

Rozdział 3

Mapowanie obiektowo-relacyjne

Jedną z największych zalet relacyjnych baz danych w kontekście programowania obiektowego jest dostępność bibliotek, które umożliwiają proste mapowanie obiektowo-relacyjne¹. Wykorzystanie ORM znacząco upraszcza proces tworzenia oprogramowania wykorzystującego bazy danych:

- Umożliwia definiowanie modelu danych poprzez tworzenie klas reprezentujących obiekty wykorzystywane w aplikacji.
- Zarządzanie połączeniami, sesjami i transakcjami bazodanowymi jest znacząco uproszczone lub dokonywane automatycznie.
- Zwiększa bezpieczeństwo aplikacji. Mechanizmy ORM dostarczają narzędzi do ochrony przez atakami typu SQL Injection². [10]
- Umożliwia modyfikacje środowiska bazodanowego bez zmieniania kodu aplikacji. Mechanizmy ORM wspierają wiele silników bazodanowych, które mogą różnić się od siebie implementowanymi standardami języka SQL.

O popularności mechanizmów ORM świadczy fakt powstawania oficjalnych standardów mapowania obiektowo-relacyjnego włączanych do specyfikacji języków programowania. Przykładem takiego standardu jest JPA³ dla języka Java.

3.1 Mapowanie obiektowo-relacyjne dla Cassandra

Wraz z pojawieniem się języka CQL pojawiła się szansa wykorzystania istniejących implementacji mechanizmów ORM do przechowywania danych z użyciem silnika Cassandra. Motywacją do stworzenia takiego rozwiązania jest zwiększenie wydajności istniejących aplikacji bez modyfikacji ich kodu. Minimalna rekonfiguracja aplikacji i podłączenie jej

¹W dalszej części pracy używany będzie skrót ORM od angielskiego pojęcia object-relational mapping.

²SQL Injection (ang. wstrzykiwanie SQL) - typ ataku polegający na wykorzystywaniu specjalnie spreparowanych wartości, które dołączone do szablonu zapytania SQL powodują szkodliwe działania niezgodne z intencją programisty.

³Java Persistence API.

pod klaster bazodanowy skutkowałyby zwiększeniem wydajności działania aplikacji. Ponadto wykorzystanie istniejących interfejsów mogłoby umożliwić obsługę różnych typów baz danych należących do ruchu NoSQL.

Przykładem projektu, który wykorzystuje mechanizm ORM do obsługi NoSQLowych baz danych jest Kundera.[11] Kundera zapewnia implementację mapowania obiektowego zgodną ze standardem JPA 2.0 między innymi dla Cassandra, HBase, MongoDB oraz Neo4j. Dodatkowo biblioteka wspiera obsługę wielu mechanizmów równocześnie.

3.1.1 Wydajność

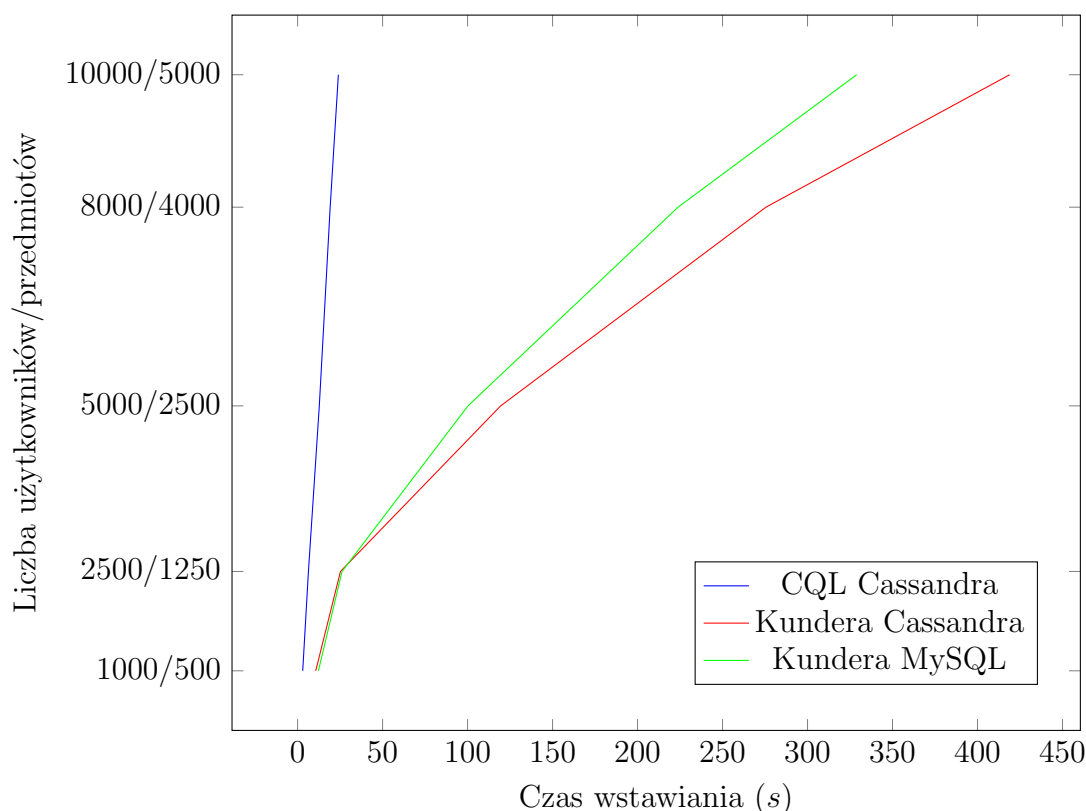
Wyniki pomiarów na stronie domowej projektu Kundera świadczą o tym, że biblioteka nie wprowadza znacznych narzutów wydajnościowych względem bezpośredniego wykorzystania interfejsu bazy danych. Ważniejsze jest jednak sprawdzenie jak duży narzut wprowadza dostosowanie relacyjnego modelu danych do Apache Cassandra. W tym celu Autor przeprowadził testy porównawcze masowego wstawiania i pobierania obiektów. Dla każdej z dwóch operacji zostały przeprowadzone trzy testy. Pierwszy test mierzy czas referencyjny. Jest wykonywany dla zdenormalizowanego modelu Cassandra przedstawionego na diagramie 2.7. Drugi test sprawdza czasy dla znormalizowanego modelu zaprezentowanego na diagramie 2.3, opisanego w JPA i uruchomiony na Cassandrze z wykorzystaniem biblioteki Kundera. Trzeci test wykorzystuje identyczny opis modelu jak drugi, jednakże wykonywany jest dla silnika MySQL. Wszystkie testy przeprowadzane były jednowątkowo. Czas opóźnień przesyłania danych jest pomijalny, gdyż testy były prowadzone na środowisku lokalnym.

W trakcie testów mierzony był całkowity czas wykonania danej operacji. Przyjęte zostały następujące założenia:

- Liczba użytkowników i przedmiotów są parametryzowane i skalowane liniowo.
- Liczba przedmiotów, które użytkownik może mieć na swojej liście życzeń zawiera się w przedziale $[0; 10]$.
- Odczytywana jest parametryzowalna liczba użytkowników.

Wyniki czasu wstawiania wielu rekordów zostały przedstawione na wykresie 3.1. Czasy pobierania wielu rekordów zostały zebrane na wykresie 3.2.

Wyniki zebrane na wykresie 3.1 pokazują, że wykorzystanie mechanizmów mapowania obiektowo-relacyjnego dla bazy danych Cassandra jest bardzo złym wyborem. Narzut związany z konwersją modelu danych do postaci relacyjnej jest tak duży, że w praktyce pojedynczy węzeł Cassandra osiąga gorsze wyniki zapisu niż baza danych MySQL. Zmiana silnika bazodanowego dla istniejącego kodu nie tylko nie poprawi wyników wydajnościowych, ale może wręcz spowodować spowolnienie działania aplikacji. Jedynym zyskiem z takiego rozwiązania będzie możliwość wykorzystania natywnego mechanizmu klastrowania węzłów Cassandra. Dzięki temu wyeliminowany zostanie pojedynczy punkt awarii systemu. Ten sam efekt można jednak uzyskać stosując rozwiązania dedykowane dla relacyjnych baz danych. Przykładami takich rozwiązań mogą być MySQL Cluster CGE dla bazy danych MySQL oraz Oracle RAC dla baz Oracle Database Enterprise Edition.



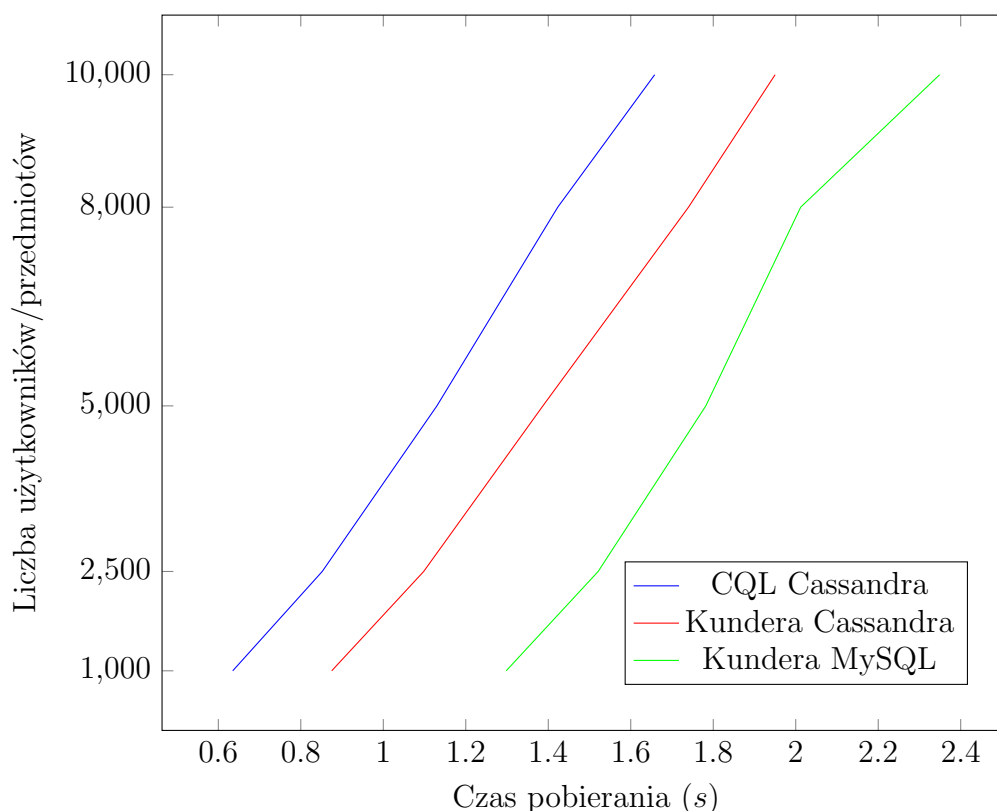
Rysunek 3.1: Porównanie czasu wstawiania wielu rekordów.

Potencjalnym zastosowaniem mapowania obiektowo-relacyjnego dla bazy danych Cassandra są środowiska integracyjne dla wielu aplikacji. Wykorzystując homogeniczny model danych można odwoływać się do różnych silników bazodanowych. W bibliotece Kundera istnieje takie rozwiązanie. Zostało nazwane Polyglot Persistence⁴. W praktyce dostosowanie modelu danych ORM do istniejących schematów baz danych jest problematyczne.

Wykres 3.1 demonstruje jak ogromną przewagę szybkości przy zapisie posiada poprawnie zaprojektowany model danych Apache Cassandra. Przypuszczalnie osiągnięty czas mógłby być jeszcze lepszy, gdyż ograniczenie wydajności zapisu następowało prawdopodobnie po stronie klienta testowego. Zgodnie ze specyfikacją Apache Cassandra jest w stanie obsługiwać bez opóźnień znacznie więcej jednoczesnych żądań zapisu.

Czasy pobierania listy życzeń użytkownika przedstawione na wykresie 3.2 nie różnią się znacząco od siebie. Przewaga szybkości pobierania danych z wykorzystaniem CQL ponownie wynika z zastosowania lepszego modelu danych. Denormalizacja pól encji *Item* pozwala pominąć pobieranie dla każdego użytkownika dodatkowego wiersza z bazy danych.

⁴Polyglot Persistence (dosłowne tłumaczenie to „zapis poliglotyczny”) - opis mechanizmu znajduje się na stronie <https://github.com/impetus-opensource/Kundera/wiki/Polyglot-Persistence>.



Rysunek 3.2: Porównanie czasu pobierania wielu rekordów.

3.2 Wnioski z badania wydajności

Na podstawie wyników przeprowadzonych badań można wnioskować, że mechanizmy mapowania obiektowo-relacyjnego nie nadają się do wykorzystania przez bazę danych Apache Cassandra. Wynika to z faktu, że mechanizmy te wspomagają normalizację i zachowywanie relacji między encjami, natomiast efektywne modelowanie dziedziny danych w Cassandra dąży do denormalizacji i maksymalnego uproszczenia zależności pomiędzy obiektami.

Drastyczna różnica wydajności pomiędzy szybkością zapisu danych w modelach zoptymalizowanym i niezoptymalizowanym świadczy o tym, że do efektywnego modelowania dziedziny w Apache Cassandra niezbędna jest wiedza na temat sposobu przechowywania danych przez tę bazę i podejmowanie świadomych wyborów. Dostarczona przez ORMy mechanizmy mogą natomiast wyłącznie spowolnić korzystanie z Cassandra.

Podczas badania wydajności Autor natrafił również na problemy z przenośnością kodu źródłowego pomiędzy różnymi systemami bazodanowymi. Wykonanie operacji wstawiania partii danych w bazie relacyjnej oraz Cassandra wymagało modyfikacji kodu źródłowego aplikacji testowej. W relacyjnej bazie danych nie można było ukończyć testu wstawiania i pobierania danych wykorzystując jedynie elementy zdefiniowane w JPA. Pomimo wielokrotnej zmiany konfiguracji w pewnym momencie testu dochodziło do całkowitego zablokowania puli połączeń. Dopiero przejście do transakcji bezstanowej zdefiniowanej w Hibernate ORM⁵ umożliwiło zakończenie testu.

⁵Hibernate ORM - środowisko aplikacyjne implementujące standard JPA dla baz relacyjnych. Zdefi-

Opisane problemy z przenośnością kodu mogą być następstwem fundamentalnych różnic pomiędzy relacyjnymi systemami bazodanowymi a Apache Cassandra. Przykładowo transakcja, czyli jedno z podstawowych pojęć związanych z zapytaniami w języku SQL, nie istnieje w systemie Cassandra. Dostosowanie implementacji do interfejsu mapowania obiektowo-relacyjnego wymaga sztucznego blokowania zapytań po stronie klienta, co prowadzi do wystąpienia niepotrzebnych opóźnień i nadmiernego wykorzystania mechanizmów synchronizacji wielowątkowej. Z drugiej strony brak lub nieodpowiednie wykorzystanie transakcji w przypadku bazy relacyjnej prowadzi do błędów zapytań.

Wyniki badań nie świadczą o niemożliwości implementacji mapowania obiektowego dla Cassandra. Dla optymalnej wydajności należy jednak zastosować rozwiązane dedykowane:

- Definiowanie modelu powinno być skoncentrowane wokół wewnętrznej reprezentacji danych w Cassandrze. W stosunku do mapowania obiektowo-relacyjnego kluczowe są aspekty obniżenia istotności relacji i wsparcie dla operacji denormalizacji z poziomu interfejsu.
- Mapowanie powinno w prosty i transparentny sposób udostępniać realizację poprawnych wzorców modelowania. Przykładowo lista wartości powinna móc być zamodelowana przynajmniej na dwa sposoby: jako kolumna typu list lub jako lista kolumn o nazwach zawierających wartości listy. Mapowanie obiektowe dla Cassandra powinno umożliwiać wymienienie fizycznej struktury danych bez zmiany wykorzystującego ją kodu źródłowego.
- Mapowanie obiektowe dla Cassandra powinno być proste. Język CQL, w porównaniu do SQL, ma prostą składnię i nie udostępnia wielu operacji. Ponadto skomplikowany mechanizm mapowania mógłby niekorzystnie kontrastować z wydajną bazą danych. Przykładem nieczytelnego mapowania z dużą liczbą opcji konfiguracyjnych może być pole encji opisane w JPA przedstawione na listingu 3.3.

```
@ManyToMany
@JoinTable(name = "wishlist",
           joinColumns = {
               @JoinColumn(name = "userId",
                           referencedColumnName = "userId") },
           inverseJoinColumns = {
               { @JoinColumn(name = "itemId",
                           referencedColumnName = "itemId") },
               foreignKey = @ForeignKey(name = "userId_fk"),
               inverseForeignKey = @ForeignKey(name = "itemId_fk"))
private Set<Item> wishlistItems = new HashSet<Item>();
```

Rysunek 3.3: Przykład nieczytelnego mapowania obiektowo-relacyjnego.

niowana w ramach tego środowiska transakcja bezstanowa nie jest częścią interfejsu JPA, a więc nie może być wspierana przez Kunderę.

3.3 Mapowanie obiektowe dla Cassandra - koncepcja

W większości rozwiązań ORM encja modelu danych jest opisywana jako definicja klasy, której obiekty reprezentować będą instancję tej encji. Rozważmy listing 3.4, który przedstawia hipotetyczne dostosowanie standardu JPA do denormalizacji dla modelu 2.7.

```
@Table(name = "wishlist")
class Wishlist {
    @Id(type = IdType.PARTITION_KEY)
    private String userId;
    @Denormalize(clustering_keys = { "itemId" },
                fields = { "name", "price" })
    private Item item;
}
```

Rysunek 3.4: Hipotetyczny przykład dostosowania definicji JPA do denormalizacji.

W powyższym przykładzie denormalizacja jest modelowana jako zdegenerowany przypadek relacji jeden-do-jednego. Przypisanie zmiennej *item* do obiektu klasy *Wishlist* i utrwalenie tego obiektu w bazie danych spowoduje uzupełnienie wartości kolumn *itemId*, *name* oraz *price* w tabeli *wishlist*. Po pobraniu obiektu klasy *Wishlist* z bazy danych przechowywany reprezentant klasy *Item* posiada tylko część informacji. Takie rozwiązanie posiada istotną wadę. Podczas korzystania z mechanizmu nie można jednoznacznie stwierdzić czy brak uzupełnienia pola *desc* w obiekcie klasy *Item* jest spowodowany tym, że opisu faktycznie brakuje, czy też operacje dokonywane są na niekompletnym (zdenormalizowanym) obiekcie. Rozwiązaniem pozbawionym tej wady jest denormalizacja jawna przedstawiona na listingu 3.5.

```
@Table(name = "wishlist")
class Wishlist {
    @Id(type = IdType.PARTITION_KEY)
    private String userId;
    @Id(type = IdType.CLUSTERING_KEY)
    private String itemId;
    @Column(name = "name")
    private String name;
    @Column(name = "price")
    private String price;
}
```

Rysunek 3.5: Hipotetyczny przykład zastosowania denormalizacji jawnej w JPA.

Denormalizacja jawna posiada jednak inne wady. Przede wszystkim model nie podaje, że kolumny *itemId*, *name* oraz *price* powinny pochodzić od instancji klasy *Item*. Ponadto użytkownik mechanizmu musi pamiętać o ręcznym wypełnieniu pól, co jest z kolei narażone na błędy. Eleganckie połączenie tych dwóch opcji nie jest możliwe w języku Java.

Ze względu na mechanizm metaklas w Pythonie stworzono eleganckie i proste implementacje ORM dla relacyjnych baz danych. Dwa najpopularniejsze mechanizmy to Django

ORM⁶ oraz SQLAlchemy⁷. Wykorzystując metaklasy osiągalne jest stworzenie mapowania o interfejsie wymienionych mechanizmów, który rozwiązuje problem zasygnalizowany w przypadku języka Java. Ilustruje to listing 3.6.

```
class Wishlist(Model):
    userId = TextField(partition_key=True)
    item = DenormalizedField(relates=Item,
                             clustering_keys=['itemId'],
                             fields=['name', 'price'])

w = Wishlist()
w.item = Item(id='1', name='n', price='10.0')
id = w.item_itemId      # id equals '1'
name = w.item_name      # name equals 'n'
price = w.item_price     # price equals '10.0'
i = w.item               # exception raised
```

Rysunek 3.6: Hipotetyczny przykład denormalizacji w mapowaniu w języku Python.

Dzięki metaklasom można dokonać następujących modyfikacji modelu *Wishlist*:

- Dla każdej instancji *DenormalizedField* automatycznie dodać do klasy pola, które będą przechowywać zdenormalizowane wartości. W przykładzie definicja klasy została rozszerzona o pola *item_itemId*, *item_name*, *item_price*.
- Każdą instancję *DenormalizedField* zamienić na właściwość tylko do zapisu, która automatycznie ustawia wartości zdenormalizowanych pól.

Dzięki takim modyfikacjom możliwe jest ustawianie wartości zdenormalizowanych pól poprzez podawanie całej instancji obiektu *Item*, jak w listingu 3.4. Jednocześnie pobieranie wartości z obiektu możliwe jest tylko poprzez odwołania do zdenormalizowanych pól. Nie da się pomylić ze sobą zdenormalizowanych i pełnych instancji klasy *Item*.

⁶Django ORM - część platformy Django, która odpowiada za operacje bazodanowe. Dokumentacja jest dostępna pod adresem <https://docs.djangoproject.com/en/dev/topics/db/>.

⁷Strona domowa projektu jest dostępna pod adresem <http://www.sqlalchemy.org>.

Rozdział 4

Modelowanie obiektowe dla Cassandra

Na podstawie wyników wydajnościowych przedstawionych w sekcji 3.2 Autor zaproponował odmienną koncepcję mechanizmu mapowania obiektowego dla bazy danych Cassandra. Aby uwypuklić różnice pomiędzy przedstawioną propozycją i tradycyjnymi mechanizmami ORM Autor postuluje stosowanie odmiennego nazewnictwa. Omówiona w sekcji 3.3 koncepcja, która będzie rozwijana w dalszej części pracy, nazywana będzie **modelowaniem obiektowym**:

- Mechanizmy mapowania obiektowo-relacyjnego nie wymuszają na użytkowniku kolejności projektowania komponentów. Osiągalne są zarówno podejście *code first*¹, jak i *database first*². W proponowanym przez Autora modelowaniu obiektowym, ze względu na występowanie struktur wysokiego poziomu odpowiadających wzorcom projektowym, podejście *database first* jest możliwe tylko teoretycznie lub w bardzo ograniczonym zakresie. Znacznie bardziej efektywne jest podejście *code first*. Posiada ono wsparcie implementacyjne - generację schematu - i jest rekomendowane przez Autora.
- W mechanizmach mapowania obiektowo-relacyjnego mapowanie odpowiada najczęściej typom danych, natomiast w obrębie danego typu jest ono jednoznaczne. Przykładowo identyfikator konwertowany jest na liczbowy klucz główny, lista przechowywana jest jako relacja, a ciąg znaków przekształcany jest na typ *VARCHAR*. W przypadku modelowania obiektowego dla Cassandra wybór wyznaczać będzie natomiast sposób przechowywania danej wartości. W zależności od narzuconego sposobu modelowania listy będzie mogła zostać ona zrzutowana na wartość kolumny o typie *list*, nazwy kolumn w wierszu lub osobną tabelę.
- Mapowanie obiektowo-relacyjne dostarcza przede wszystkim niskopoziomowe odpowiedzi typów bazodanowych. Wyjątkiem jest relacja wiele-do-wielu. Modelowanie

¹Code first (ang. dosłownie „najpierw kod źródłowy”) - w kontekście ORM oznacza to modelowanie dziedziny za pomocą kodu źródłowego, z którego następnie generowany jest schemat bazy danych.

²Database first (ang. dosłownie „najpierw baza danych”) - w kontekście ORM oznacza to modelowanie dziedziny jako schemat bazodanowy, do którego następnie pisany jest mapujący kod źródłowy.

obiektyw dla Cassandra skupia się zarówno na polach niskiego poziomu (przykładowo kolumna o wartości tekstowej), jak i na reużywalności komponentów wysokopoziomowych odpowiadających wzorcom projektowym.

- Częścią modelowania obiektyw dla Cassandra są opcjonalne mechanizmy mapowania obiekto-relacyjnego. Należą do nich generacja schematu oraz migracje pomiędzy wersjami modelu.

W dalszej części pracy omówiona zostanie konkretna implementacja modelowania obiektyw dostarczona wykonana przez Autora. Pełna nazwa tego projektu to **Object Modeling for Cassandra**. W dalszej części pracy używana będzie nazwa skrócona - *OMC*.

4.1 OMC - podstawowe pojęcia

Środowisko aplikacyjne OMC definiuje trzy podstawowe pojęcia: model, pole oraz silnik.

4.1.1 Model

Jest to klasa, która opisuje atomowy względem kodu źródłowego obiekt modelujący bazę danych. Obiekt ten, w zależności od definicji, może być mapowany na jedną lub wiele tabel. Jedynym wymaganiem, które musi spełnić klasa należąca do modelu jest dziedziczenie po klasie bazowej `Model`. Ciało modelu stanowią pola.

4.1.2 Pole

Pole jest to zmienna klasy definiującej model. Opisuje atomową wartość z punktu widzenia kodu źródłowego. W zależności od typu pole może być mapowane na jedną kolumnę, wiele kolumn lub autonomiczną tabelę w bazie danych. Aby pole klasy mogło zostać zakwalifikowane jako pole modelu, w definicji klasy należy mu przypisać obiekt dziedziczący po klasie bazowej `Field`.

4.1.3 Silnik

Silnik jest to obiekt, który odpowiada za nawiązywanie połączenia z bazą danych. Po utworzeniu silnika wystarczy przypisać go do całego modelu. Wszystkie operacje bazodanowe będą domyślnie wykonywane przez przypisany silnik.

Utworzenie silnika wymaga podania adresu URL opisującego połączenie, nazwanego z języka angielskiego *connection string*. Strukturę tego adresu przedstawia listing 4.1.

```
cassandra://{ip}:{port}/{keyspace}?rf={rf}&strategy={strategy}
```

Rysunek 4.1: Struktura adresu URL opisującego połączenie z bazą danych Cassandra.

Wszystkie elementy ujęte w znaki `{ }` w adresie 4.1 należy zamienić na wartości według poniższego objaśnienia:

ip adres IP węzła koordynatora, z którym będzie komunikować się kod źródłowy,

port port IP koordynatora (podanie portu jest opcjonalne; wartość domyślna to 9042),

keyspace przestrzeń nazw Cassandra, w której operuje OCM,

rf współczynnik replikacji³ podawany przy tworzeniu przestrzeni nazw, jeżeli ta nie istnieje (parametr jest opcjonalny),

strategy nazwa strategii replikacji podawanej przy tworzeniu przestrzeni nazw, jeśli ta nie istnieje (parametr jest opcjonalny).

Do utworzenia silnika służy metoda `Engine.create_engine(connection_string)`. Aby przypisać silnik do modelu należy użyć metody `Model.bind(engine)`. Przykład wywołania przypisującego modelowi silnik dla serwera lokalnego i przestrzeni nazw `test` został zaprezentowany na listingu 4.2.

```
engine = Engine.create_engine('cassandra://127.0.0.1/test')
Model.bind(engine)
```

Rysunek 4.2: Przykład wywołania tworzącego silnik i przypisującego go do modelu.

4.2 Lista życzeń użytkownika - studium przypadku

Encję użytkownika przy pomocy OMC można opisać modelem `User` zaprezentowanym na listingu 4.3. Model posiada trzy pola: identyfikator (`id`), imię (`name`) oraz nazwisko (`surname`).

```
class User(Model):
    id = UuidField(type=TimeUuid, auto_generate=True, partition_key=True)
    name = TextField(selectable=True)
    surname = TextField(selectable=True)
```

Rysunek 4.3: Tabela użytkownika opisana przy pomocy OMC.

Identyfikator jest opisany polem typu `UuidField`. Ze względu na zadeklarowany parametr `type=TimeUuid` pole zostanie zmapowane w Cassandrze jako dana o typie `timeuuid`. Mechanizm OMC posiada wbudowaną walidację dla identyfikatorów. Przypisanie polu `id` instancji modelu `User` błędnej wartości spowoduje zgłoszenie wyjątku `AttributeError`. Parametr `auto_generate=True` oznacza, że jeżeli identyfikator nie zostanie ustawiony manualnie, to zostanie wygenerowany automatycznie podczas mapowania modelu na bazę danych. Ustawienie `partition_key=True` oznacza natomiast, że pole będzie częścią klucza wiersza tabeli.

Imię i nazwisko opisane są polem typu `TextField`. Jest ono mapowane na bazodanowy typ `text`. Deklaracja `selectable=True` oznacza, że na polu zostanie utworzony indeks

³Współczynnik replikacji (ang. replication factor) - określa na ile węzłów kopiowany jest każdy wstawiony wiersz.

drugiego stopnia. Dzięki temu możliwe jest przeszukiwanie tabeli po wartości tego pola. Na listingu 4.4 przedstawiono odwołanie do OCM, dzięki któremu można wyszukać wszystkich użytkowników o imieniu Jakub i nazwisko Turek.

```
User.objects().find(name='Jakub', surname='Turek')
```

Rysunek 4.4: Przykład wyszukiwania użytkownika po imieniu i nazwisku.

Opis encji przedmiotu w OMC został zaprezentowany na listingu 4.5.

```
class Item(Model):
    id = UuidField(type=TimeUuid, auto_generate=True, partition_key=True)
    name = TextField(searchable=True, length=255)
    price = DecimalField(total_positions=8, decimal_positions=2)
    desc = TextField()
    category = DictionaryField(entries=['BOOKS', 'CLOTHES', 'FURNITURE'])
    weight = DecimalField(total_positions=5, decimal_positions=1)
```

Rysunek 4.5: Tabela przedmiotu opisana przy pomocy OMC.

W stosunku do poprzedniej definicji zostały użyte dwa nowe typy pól: `NumberField` oraz `DictionaryField`. Pierwsze z nich służy do przechowywania wartości liczb dziesiętnych. Cassandra nie posiada odpowiednika znanego z relacyjnych baz danych stałoprzecinkowego typu `NUMBER`. O umieszczanie poprawnie zaokrąglonych wartości musi zadbać autor oprogramowania korzystającego z Apache Cassandra. Aby to ułatwić, OMC dostarcza programowe wsparcie dla operacji stałoprzecinkowych. Pole `DecimalField` posiada dwa parametry. Opcja `total_positions` opisuje maksymalną liczbę cyfr dziesiętnych, z których może składać się liczba, a `decimal_positions` definiuje ile z tych cyfr należy do części ułamkowej pola. `DecimalField` w przezroczysty dla programisty sposób wspiera przypisywanie i wszystkie operacje arytmetyczne dla typów `int`, `long` oraz `float`.

Pole `DictionaryField` dostarcza wsparcia dla słowników jednokrotnego wyboru. Parametr `entries` pozwala na wyspecyfikowanie listy dozwolonych wartości pola. Podanie wartości spoza listy powoduje zgłoszenie wyjątku `AttributeError`. Typ pola jest automatycznie wnioskowany z wartości parametru `entries`. Dla listy elementów typu `str` pole zostanie zmapowane na wartość typu `text`. W przypadku, gdy typ nie może być wywnioskowany na podstawie listy, należy podać go jawnie za pomocą parametru `type`.

Opis encji listy życzeń przedstawia listing 4.6.

```
class Wishlist(Model):
    userId = UuidField(type=TimeUuid, partition_key=True)
    item = DenormalizedField(related=Item,
                            clustering_keys=['itemId'],
                            fields=['name', 'price'])
```

Rysunek 4.6: Tabela listy życzeń opisana przy pomocy OMC.

Pole `DenormalizedField` spełnia postulaty przedstawione w sekcji 3.3. Istotny jest fakt, że na podstawie parametru `clustering_keys=['itemId']` pole `itemId` staje się częścią złożonego klucza głównego tabeli.

4.3 Wsparcie dla wzorców modelowania

OMC został zaprojektowany w taki sposób, aby dostarczać mechanizmów wspierających wzorce projektowe. W tej sekcji Autor opisuje wszystkie wzorce, które były analizowane, modelowane, testowane i optymalizowane w trakcie prac implementacyjnych nad OMC.

4.3.1 Szereg zdarzeń

Apache Cassandra została zaprojektowana z myślą o wspieraniu szeregów zdarzeń. Szereg zdarzeń to uporządkowane względem czasu wystąpienia wpisy do bazy danych, które opcjonalnie mogą posiadać z góry określony czas życia. Typowym przykładem szeregu zdarzeń są wartości odczytów z urządzeń pomiarowych. Wartości te zwyczajowo są pobierane cyklicznie, a do wykorzystania w czasie rzeczywistym przydatne są tylko najnowsze odczyty. Doskonały opis modelowania szeregów zdarzeń przedstawiony jest we wpisie Patricka McFadina na portalu **Planet Cassandra**. [12] Artykuł analizuje przykład stacji meteorologicznych. Przedstawia trzy sposoby zapisu szeregu zdarzeń w Apache Cassandra.

Odczyty dla tego samego urządzenia w pojedynczym wierszu Najprostszą możliwością jest zebranie wszystkich próbek dla danego urządzenia w jednym wierszu tabeli. Kolejne wartości zapisywane są ze śladem czasu w nazwie kolumny. Listing 4.7 przedstawia wzorec zapisany jako model OMC.

```
class TimeSeriesPatternOne(Model):
    weatherstation_id = TextField(partitioning_key=True)
    event_time = TimestampField(clustering_key=True, auto_on_create=True)
    temperature = TextField()
```

Rysunek 4.7: Modelowanie szeregu zdarzeń. Zapis kolejnych odczytów jako kolumny.

Zapis śladu czasu w nazwie kolumny można uzyskać poprzez dołączenie pola do złożonego klucza głównego tabeli jako *clustering key*⁴. Dołączenie to dokonuje się poprzez ustawienie flagi `clustering_key` na `True`.

Udogodnieniem dla programisty korzystającego z OMC jest możliwość automatycznej generacji śladu czasu dla chwili obecnej. Ślad czasu może być wygenerowany w trakcie tworzenia obiektu (`auto_on_create=True`) lub w trakcie wstawiania obiektu do bazy danych (`auto_on_save=True`).

Grupowanie odczytów dla urządzenia i fragmentu daty W przypadku gdy odczytów jest zbyt dużo by mieściły się w jednym wierszu sugerowanym rozwiązaniem jest dodatkowe partycjonowanie wierszy po komponencie daty: miesiącu, dniu, itd. Autor wzorca sugeruje utworzenie w modelu dodatkowego pola, do którego wpisywany jest komponent daty i dołączenie tego pola do klucza partycjonowania. Takie modelowanie dziedziny może prowadzić do potencjalnych błędów. Ta sama dana - ślad czasu - przechowywana jest w dwóch polach, które różnią się formatowaniem.

⁴Clustering key (ang. dosłownie „klucz klastrowania”) - część złożonego klucza głównego tabeli, która określa w jakiej kolejności grupowane są wpisy względem klucza partycjonowania (identyfikatora wiersza).

OMC posiada mechanizm partycjonowania dat po komponencie. W modelu znajduje się tylko jedna dana - faktyczny ślad czasu. OMC na jej podstawie automatycznie tworzy i formatuje klucz partycjonowania. Dzięki temu programista nie musi dbać o ręczne wyznaczanie komponentu daty. Listing 4.8 przedstawia opis modelu z partycjonowaniem względem dnia.

```
class TimeSeriesPatternTwo(Model):
    weatherstation_id = TextField(partitioning_key=True)
    event_time = TimestampField(clustering_key=True,
                                auto_on_create=True,
                                partitioning_by_day=True)
```

Rysunek 4.8: Modelowanie szeregu zdarzeń. Partycjonowanie zdarzeń po komponencie daty.

Pole `event_time` modelu `TimeSeriesPatternTwo` jest kluczem klastrowania. Dodatkowo posiada ono ustawioną flagę `partitioning_by_day=True`, która spełnia trzy zadania:

- Przy tworzeniu schematu tabeli dodaje dodatkową kolumnę `event_time_day`⁵ o typie tekstowym. Pole to jest automatycznie włączane do klucza partycjonowania.
- Przy zapisywaniu obiektu do bazy danych automatycznie wypełnia wartość dodatkowego pola o ślad czasu obcięty do z dokładnością do dnia.
- Przy wyszukiwaniu obiektów po dacie automatycznie filtruje dane zarówno po wartości pola `event_time_day`, jak i `event_time`.

Dostępne jest kilka rodzajów komponentów daty, po których można dokonać partycjonowania. Różnią się one ostatnim członem nazwy parametru. Wybór obejmuje partycjonowanie po latach (`_year`), miesiącach (`_month`), dniach (`_day`), godzinach (`_hour`), minutach (`_minute`) oraz sekundach (`_second`).

Odczyty z ograniczoną pamięcią Korzystając z wbudowanego mechanizmu Cassandra możliwe jest tworzenie automatycznie przedawniających się wartości. Wstawiając pozycję do bazy danych można ustawić czas życia, po którym zostanie ona usunięta. Przedawnianie się wartości może być wykorzystywane w połączeniu z odwrotną kolejnością sortowania wpisów (od najnowszego do najstarszego). Dzięki temu bardzo łatwo jest budować widoki, które prezentują kilka najnowszych informacji.

OMC zapewnia dwa rodzaje wsparcia dla czasu życia wartości. Czas życia może być ustawiany dla indywidualnych obiektów. Przy zapisywaniu obiektu należy na jego instancji wywołać metodę `save(ttl=20)`. Parametr `ttl` definiuje czas życia wartości w sekundach. Czas życia można również ustawić globalnie dla wszystkich instancji modelu. Wykorzystuje się do tego pole `__ttl__` klasy. Zaprezentowano to na listingu 4.9.

Dla każdego pola, które jest częścią klucza klastrowania można ustawić odwrotną kolejność sortowania elementów. Do tego celu należy ustawić flagę `descending_clustering` na `True`.

⁵Nazwa jest tworzona z nazwy pola oraz nazwy komponentu daty.


```
class TimeSeriesPatternThree(Model):
    __ttl__ = 20
    weatherstation_id = TextField(partitioning_key=True)
    event_time = TimestampField(clustering_key=True,
                                descending_clustering=True
                                auto_on_create=True)

    temperature = TextField()
```

Rysunek 4.9: Modelowanie szeregu zdarzeń. Przedawniające się odczyty jako kolumny z sortowaniem od najnowszego.

4.3.2 Wsparcie dla kolejek

W sekcji 2.5 został przytoczony antywzorzec kolejki. Problem został opisany przez Alekseya Yeschenko na blogu firmy DataStax. [8] Wynika on ze sposobu obsługi operacji usuwania przez silnik Apache Cassandra. Usuwanie nie jest realizowane natychmiast. Zamiast tego kasowana kolumna oznaczana jest specjalnym znacznikiem *tombstone* i rezyduje w pamięci aż do momentu, gdy nastąpi właściwe wyrzucanie. Przy dużej liczbie usunięć kolumn z wiersza może zdarzyć się, że zapytanie filtrujące z limitem będzie działało znacznie wolniej niż dla wiersza z identyczną ilością kolumn, ale bez poprzedzających usunięć. Wynika to z faktu, że stwierdzenie, że dana wartość jest „martwa” wymaga pobrania kolumny i zdeserializowania jej wartości. Jeżeli po deserializacji wykryty zostanie znacznik *tombstone*, wtedy pobierana jest kolejna wartość. Proces powtarzany jest aż do skutku.

Kolejka może być zamodelowana efektywniej w przypadku, gdy możliwe jest wskazanie miejsca, w którym rozpoczynają się nieusunięte kolumny. Przykładem może być kolejka typu FIFO⁶. Wyszukując kolejny element w kolejce można nałożyć warunek, że czas wstawienia elementu do kolejki jest późniejszy niż czas ostatnio usuniętego elementu. OMC dostarcza wsparcia dla takich kolejek, co zostało przedstawione na listingu 4.10.

Parametr `__track_deletes__` umożliwia śledzenie operacji usuwania. Przyjmuje on parę łańcuchów znaków: nazwę pola, którego wartość jest śledzona podczas usuwania oraz warunek zapisywania śledzonej wartości. Na chwilę obecną obsługiwane jest śledzenie pól typu `UuidField` oraz `TimestampField`, a także dwa warunki śledzenia: `ASC` oraz `DESC`. Warunek *ascending* oznacza, że śledzona wartość jest aktualizowana tylko w przypadku, gdy jest mniejsza od usuwanej obecnie wartości. Warunek *descending* działa odwrotnie - aktualizuje śledzoną wartość tylko w przypadku, gdy jest większa od usuwanej obecnie.

Ustawienie parametru `__track_deletes__` powoduje, że śledzona wartość jest automatycznie dodawana do wszystkich zapytań wyszukiwujących obiekty. Oznacza to, że ostatnie wywołanie w listingu 4.10 będzie miało ustawione trzy warunki filtrowania:

- Nazwa elementu (`name`) jest równa `'3'` - wynika z `find(name='3')`.
- Wybierz maksymalnie jeden element (`LIMIT 1`) - wynika z zakresu `[:1]`.
- Czas dodania do kolejki (`enqueued_at`) większy od `second_element.enqueued_at` - automatycznie dodawane przez OMC.

⁶FIFO (First In, First Out) - rodzaj kolejki, w którym elementy obsługiwane są według kolejności wstawienia.

```

class FIFOQueue(Model):
    __track_deletes__ = ('enqueued_at', 'ASC')
    name = TextField(partitioning_key=True)
    enqueued_at = UuidField(type=TimeUuid,
                           auto_generate=True,
                           clustering_key=True)

    payload = DataField()

    # create three queue elements
    first_element = FIFOQueue(name = '1', payload = 'firstPayload')
    second_element = FIFOQueue(name = '2', payload = 'secondPayload')
    third_element = FIFOQueue(name = '3', payload = 'thirdPayload')
    # save elements into database
    first_element.save()
    second_element.save()
    third_element.save()
    # delete first two elements from database
    first_element.delete()
    second_element.delete()
    # get element with name '3'
    FIFOQueue.objects().find(name='3')[:1]

```

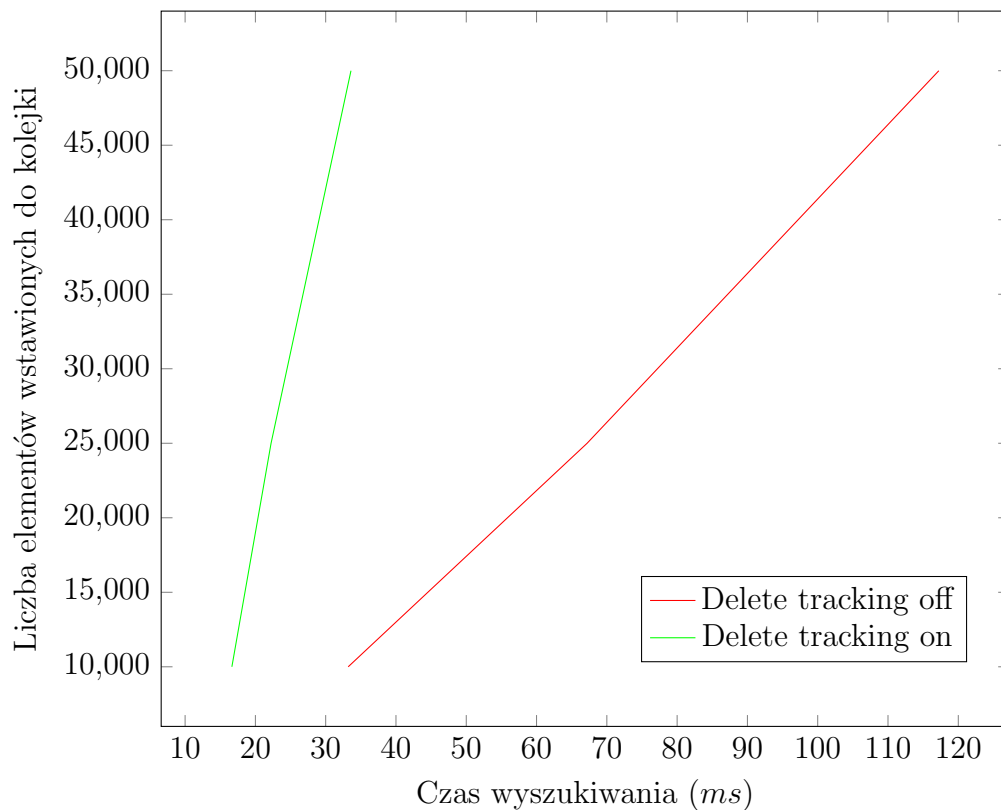
Rysunek 4.10: Modelowanie kolejki typu FIFO w OMC.

Wykorzystanie opcji `__track_deletes__` w sposób znaczący wpływa na wydajność kolejki FIFO. Na wykresie 4.11 przedstawiono porównanie szybkości działania wyszukiwania pierwszego nieusuniętego elementu kolejki przy włączonym oraz wyłączonym śledzeniu. Pomiar wydajności został przeprowadzony na lokalnej konfiguracji klastra Apache Cassandra. Czas mierzony był przez aplikację wykorzystującą mechanizm OMC. Aplikacja ta wstawiała do kolejki y przedmiotów, po czym usuwała pierwsze $y - 1$ przedmiotów z tej kolejki. Mierzony był czas wyszukiwania pierwszego elementu należącego do kolejki (z ograniczeniem LIMIT 1, podobnie jak przy wywołaniu metody `find` w listingu 4.10).

Pomimo znaczącej poprawy czasu dostępu do danych przy wykorzystaniu parametru `__track_deletes__` można zauważyć, że modelowanie nie jest optymalne. W przypadku braku usunięć danych z tabeli czas dostępu do pojedynczej danej powinien być w przybliżeniu stały. Z wykresu wynika jednak, że wyszukiwanie wydłuża się wraz ze wzrostem liczby elementów w kolejce. Pozwala to wyciągnąć wniosek, że efektywne modelowanie w Cassandrze powinno całkowicie eliminować lub radykalnie ograniczać ilość wykonywanych operacji usunięć.

4.3.3 Selektywna aktualizacja

Apache Cassandra jest przystosowana do obsługi wielu równoległych źródeł danych. Ze względu na brak transakcyjności jednym ze sposobów na uniknięcie problemów synchronizacji jest zalecana przez Dave’a Gardniera selektywna aktualizacja. [9] Edytując zawartość danego wiersza należy przysyłać do bazy danych wyłącznie wartości, które rzeczywiście uległy modyfikacji. Inaczej istnieje prawdopodobieństwo niecelowego nadpisania zmian



Rysunek 4.11: Porównanie czasu wstawiania wielu rekordów.

wprowadzonych równolegle przez inne źródło danych. OMC domyślnie korzysta z mechanizmu selektywnej aktualizacji, co przedstawiono na listingu 4.12.

```
class SelectiveUpdate(Model):
    name = TextField(partitioning_key=True)
    description = TextField(length=1024)
    priority = IntegerField()
    # find object with name 'su_test'
    su_test = SelectiveUpdate.objects().find(name='su_test').get()
    # update priority
    su_test.priority = 1
    # update object
    su_test.save()
```

Rysunek 4.12: Przykład selektywnej aktualizacji.

W pierwszej kolejności metoda `save()` sprawdza, którą z dwóch operacji, aktualizacji czy wstawienia obiektu, należy wykonać. Przy wykonywaniu aktualizacji do bazy danych przesyłana jest tylko wartość zmienionych pól, w danym przykładzie tylko priorytetu. Jeżeli pomiędzy pobraniem obiektu, a aktualizacją priorytetu i zapisaniem obiektu w bazie danych inny węzeł dokona zmiany opisu obiektu, opis ten nie zostanie ponownie nadpisany poprzednią wartością.

Mechanizm aktualizacji selektywnej można wyłączyć. Można dokonać tego wywołując metodę `save()` z parametrem `selective_update=False`. Alternatywnie selektywna aktualizacja może zostać wyłączona dla całego modelu poprzez zadeklarowanie pola `__selective_update__ = False`.

4.3.4 Indeks wartości unikalnych

Aby umożliwić wyszukiwanie po wartościach kolumny, która nie należy do klucza głównego tabeli należy użyć indeksu drugiego rzędu (*secondary index*). Według dokumentacji Apache Cassandra [13] indeksy drugiego stopnia powinny być używane na kolumnach, które silnie grupują wiersze encji nadrzędnej. Wynika to z wewnętrznej implementacji indeksów drugiego stopnia. Załóżmy, że w systemie istnieje tabela przechowująca dane użytkowników, w której zapisane są wiersze zaprezentowane na diagramie 4.13.

	name	surname	email	city
jturek	Jakub	Turek	J.Turek@stud.elka.pw.edu.pl	Warsaw
	name	surname	email	city
manisero	Michał	Aniserowicz	M.Aniserowicz@stud.elka.pw.edu.pl	Warsaw

Rysunek 4.13: Przykładowe wartości w tabeli użytkownicy.

Przykładowy wygląd danych w tabeli, która realizuje indeks drugiego stopnia dla miasta (kolumna `city`), został zaprezentowany na diagramie 4.14.

	jturek	manisero
Warsaw	null	null

Rysunek 4.14: Przykładowa tabela, która mogłaby realizować indeks drugiego stopnia dla kolumny `city` z diagramu 4.13.

Wadą rozwiązania przedstawionego na diagramie 4.14 jest brak skalowalności. Przy dodawaniu użytkowników wiersz rozszerza się o kolejne kolumny, które przechowywane są na tym samym węźle. Zajmuje to dużo miejsca na jednej partycji. Dodatkowo wszystkie zapytania o dany indeks zawsze odwołują się do jednego węzła, co skutkuje nierównomiernym rozłożeniem obciążenia. Ze względu na wymienione wady indeksy drugiego stopnia są realizowane lokalnie. Każdy węzeł zawiera informacje o wszystkich indeksowanych kolumnach, ale wyłącznie o kluczach wierszy, które znajdują się na danym węźle. Oznacza to, że jeżeli użytkownicy **jturek** oraz **manisero** znajdują się na różnych węzłach, również indeks jest równomiernie dystrybuowany pomiędzy te dwa węzły.

Lokalne indeksowanie wymusza odwołania do wszystkich węzłów dla każdego wyszukiwania. Zakładając, że indeksowanych wierszy (zarówno wszystkich, jak i tych z wyniku zapytania) jest dużo więcej niż węzłów, nadal znacząco przyspiesza to wyszukiwanie. Sytuacja zmienia się, gdy indeksowana jest wartość unikalna. Wtedy wbudowany w Cassandrę mechanizm wymusza wykonanie wielu zapytań zamiast jednego.

Indeks wartości unikalnych można zamodelować samodzielnie w prosty sposób. [14] Wystarczy utworzyć nową tabelę, w której unikalna wartość jest identyfikatorem. W wier-

szu dla danej unikalnej wartości zapisany jest odpowiedni identyfikator. Przykład takiego indeksu dla adresu e-mail (kolumna `email`) użytkowników przedstawiono na diagramie 4.15.

UserEmailIndex	J.Turek@stud.elka.pw.edu.pl	jturek
		null
	M.Aniserowicz@stud.elka.pw.edu.pl	manisero
		null

Rysunek 4.15: Przykład tabeli przechowującej indeks dla unikalnych wartości (adres e-mail użytkownika).

OMC wspiera automatyczne tworzenie indeksów dla wartości unikalnych. Na listingu 4.16 przedstawiono sposób utworzenia indeksu dla wartości unikalnych dla pola `email`.

```
class User(Model):
    id = UuidField(type=TimeUuid, auto_generate=True, partition_key=True)
    name = TextField()
    surname = TextField()
    email = TextField(searchable_unique=True)
    city = TextField(searchable=True)
```

Rysunek 4.16: Modelowanie tabeli użytkownicy z wykorzystaniem indeksu dla wartości unikalnych.

Flaga `searchable_unique=True` powoduje, że dla danego pola tworzony jest indeks unikalny. Indeks jest w całości zarządzany przez OMC. Podczas wyszukiwania metodą `User.objects().find(email='J.Turek@stud.elka.pw.edu.pl')` mechanizm wyszukiwania identyfikator użytkownika w stworzonym przez siebie indeksie. OMC aktualizuje indeksy podczas wykonywania operacji `save()` oraz `delete()`.

Zarówno indeks wartości unikatowych, jak również domyślny indeks drugiego stopnia (tworzony przy użyciu flagi `searchable=True`) nazywane są przy użyciu tej samej konwencji. Pierwszą składową jest nazwa encji, drugą - nazwa pola, a trzecia to słowo „index”. Dla przykładu 4.16 indeks wartości unikatowych zostanie umieszczony w tabeli o nazwie `user_email_index`.

4.4 Przetwarzanie partiami

Apache Cassandra zapewnia wsparcie dla operacji atomowych. [15] Wykorzystując partie⁷ można wykonać zestaw operacji, z których (cytując dokumentację) „jeżeli jedna się powiedzie, to wykonają się wszystkie”. Nie jest to jednak odpowiednik transakcji znanych z relacyjnych baz danych. Apache Cassandra nie zapewnia integralności wykonywanych w partii zapytań: dane mogą być modyfikowane równolegle przez inne zapytania. Przykład atomowych operacji został zaprezentowany na diagramie 4.17. Jest to zestaw dwóch zapytań, które usuwają z bazy danych użytkownika i właściwy mu indeks dla pola e-mail.

⁷Partia (ang. batch) - zgrupowany zestaw zapytań INSERT, UPDATE lub DELETE.

```
BEGIN BATCH
DELETE FROM user WHERE userId = 'jturek'
DELETE FROM user_mail_index WHERE mail = 'J.Turek@stud.elka.pw.edu.pl'
APPLY BATCH;
```

Rysunek 4.17: Przykład operacji na partii, która usuwa z bazy danych użytkownika *jturek* i jego indeks dla pola email.

Silnik wykorzystywany w OCM zapewnia wsparcie dla przetwarzania partiami. Potrafi kolejkować wszystkie zapytania DML⁸, a następnie wykonywać zapytania z kolejki jako operacja atomowa. Przykład wykorzystania operacji na partiach w OCM zaprezentowano na listingu 4.18.

```
# starts batch
Model.begin_batch()
# create new Item
item = Item(name='item', description='desc', price=12.5)
item.save()      # save is queued
wishlist = Wishlist(userId='jturek', itemId=item.id)
wishlist.save()   # save is queued
# executes item.save() and wishlist.save()
Model.apply_batch()
```

Rysunek 4.18: Przykład wykonania operacji na partii, która wstawia do bazy nowy przedmiot i dodaje go do listy życzeń użytkownika.

Wywołanie metody `Model.begin_batch()` powoduje rozpoczęcie kolejkowania zapytań DML. Operacja wstawienia nowego przedmiotu (`item.save()`) nie wykonuje się od razu, ale jest umieszczana w kolejce. Do kolejki trafia również operacja aktualizacji listy życzeń użytkownika (`wishlist.save()`). Wywołanie metody `Model.apply_batch()` powoduje, że partia dwóch zgrupowanych zapytań wykonuje się.

4.5 Wsparcie dla liczników

Apache Cassandra udostępnia wsparcie dla specjalnego typu danych - liczników. [16] Liczniki to kolumny specjalnego typu `counter`, na których można wykonywać tylko dwie operacje: inkrementacja lub dekrementacja. Poza specjalnym typem zmienia się również sposób interakcji z tabelą. Nie można wstawiać do niej wartości, a jedynie aktualizować liczniki dla różnych kombinacji pozostałych pól.

OMC wspiera modelowanie liczników. Przykład przedstawiono na listingu 4.19.

Obsługa modelu posiadającego licznik różni się od wykorzystania standardowego modelu. Przede wszystkim wartości wszystkich pól są ustawione w trybie `read only`, czyli nie jest możliwe wykonanie operacji `sweden.country = 'Germany'`. Dzięki temu wszystkie operacje wykonywane na liczniku są zawsze odpowiednio interpretowane, gdyż nie

⁸Data Modification Language (ang. język modyfikacji danych) - podzbiór zapytań obejmujących wstawianie (INSERT), aktualizację (UPDATE) oraz usuwanie (DELETE) danych.

```
class TripCounter(Model):
    country = TextField(partitioning_key=True)
    visits = CounterField()
    sweden = TripCounter(country='Sweden')
    sweden.visits.increment(1)
    sweden.save()
    poland = TripCounter(country='Poland')
    poland.visits.increment(5)
    poland.visits.decrement(2)
    poland.save()
    s_visits = TripCounter.find(country='Sweden').get().visits      # 1
    p_visits = TripCounter.find(country='Poland').get().visits      # 3
```

Rysunek 4.19: Przykład użycia liczników w OCM.

może się zmienić ich kontekst. W przeciwnym razie trudno byłoby ocenić intencje programisty, który najpierw inkrementuje licznik o 2, a następnie zmienia wartość kraju dla obiektu ze zwiększonym licznikiem. Mogło być to działanie celowe (operacje wykonane w nieintuicyjnej kolejności) lub zwykły błąd.

Pole typu `CounterField` udostępnia dwie operacje: inkrementację (`increment()`) oraz dekrementację (`decrement()`). Metody wywołane bez podania parametru standardowo zwiększają/zmniejszają wartość licznika o 1. Obie funkcje są łączne. Oznacza to, że tak jak w przykładzie można dokonywać wielu operacji na jednym liczniku, które przed uaktualnieniem obiektu w bazie danych są sumowane.

Po wykryciu pola licznikowego w definicji modelu OCM automatycznie zmienia sposób obsługi obiektu. Nowo powstały obiekt jest traktowany tak, jakby został wcześniej pobrany z bazy danych - operacja `save()` dokonuje aktualizacji, a nie wstawienia.

Bibliografia

- [1] <http://static.googleusercontent.com/media/research.google.com/pl//archive/bigtable-osdi06.pdf>
- [2] <http://www.datastax.com/documentation/articles/cassandra/cassandrathenandnow.html>
- [3] <http://cassandra.apache.org>
- [4] Towards Robust Distributed Systems, Dr. Eric A. Brewer, PODC Keynote
- [5] The data model is dead, long live the data model, Patrick McFadin
- [6] <http://www.ebaytechblog.com/2012/07/16/cassandra-data-modeling-best-practices-part-1/>
- [7] <https://wiki.apache.org/cassandra/API>
- [8] <http://www.datastax.com/dev/blog/cassandra-anti-patterns-queues-and-queue-like-datasets>
- [9] <http://www.slideshare.net/davegardnerisme/cassandra-concepts-patterns-and-antipatterns>
- [10] <http://doctrine-orm.readthedocs.org/en/latest/reference/security.html>
- [11] <https://github.com/impetus-opensource/Kundera>
- [12] <http://planetcassandra.org/blog/post/getting-started-with-time-series-data-modeling/>
- [13] <http://www.datastax.com/docs/1.1/ddl/indexes>
- [14] <http://www.wentnet.com/blog/?p=77>
- [15] http://www.datastax.com/documentation/cql/3.0/cql/cql_reference/batch_r.html
- [16] http://www.datastax.com/documentation/cql/3.1/cql/cql_using/use_counter_t.html

OŚWIADCZENIE

Oświadczam, że Pracę Dyplomową pod tytułem “[TYTUŁ]”, którą kierował dr inż. Jakub Koperwas, wykonałem samodzielnie, co poświadczam własnoręcznym podpisem.

.....

Jakub Turek