Opis rozwiązania - BSON

Projekt – RSO MongoDB

Autorzy:

* Tomasz Adamiec
* Piotr Cebulski
* Marek Kowalski
* Mateusz Rosiewicz
* Paweł Sokołowski
* Marcin Wnuk

Warszawa, 2013

Wstęp

* 1. Czym jest BSON

BSON (skrót od Binary JSON) jest binarnym formatem stosowanym do przechowywania dokumentów. Każdy dokument składa się z zestawu elementów typu nazwa-wartość. Gdzie wartość może należeć do wielu różnych typów (takich jak int, double, string) w tym typów złożonych takich jak dokument w dokumencie czy tablica elementów.

* 1. Zastosowanie BSON

W naszym projekcie BSON stosowany jest do przesyłania, zapisywania i przeszukiwania zawartości bazy danych. Pole "\_id", które każdy BSON zawiera, stosowane jest jako klucz główny ze względu na swoją unikalność.

Opis rozwiązania

* 1. Ogólny zarys

Rozwiązanie modułu BSON składa się z pięciu klas:

* BSON - klasa abstrakcyjna zawierająca statyczne metody służące do parsowania tablicy bytów zawierającej BSON na BSONDocument
* BSONDocument - sparsowany dokument BSON, zawierający obiekty BSONElement różnych typów
* BSONElement - pojedyncza para nazwa - wartość
* BSONtype - wyliczenie możliwych typów wartości zawartych w BSON
* ObjectID - pole ObjectID dokumentu BSON, zawiera 4 numeryczne pola: czas utworzenia, id maszyny, id procesu, licznik

Najważniejszymi elementem rozwiązania jest klasa BSON. Zawarte w niej metody parseBSON i getBSON stanowią podstawę działania całego pakietu.

* 1. Parsowanie BSON

Parsowanie BSON (czyli przejście z tablicy bajtów na obiekt klasy BSONDocument) wykonywane jest przez metodę parseBSON. Najpierw odczytywane jest pierwsze pole dokumentu BSON, czyli jego długość.

Następnie dla każdego elementu (pary nazwa-wartość) zawartego w dokumencie proces wygląda tak samo. W pierwszej kolejności odczytywany jest jego typ i nazwa. Parsowanie wartości przechowywanej w elemencie jest zależne od typu i jest wykonywane przez metodę parse klasy BSON.

* 1. Tworzenie BSON

Tworzenie BSON (czyli tablicy bajtów z obiektu klasy BSONDocument) jest procesem odwrotnym do jego parsowania. W pierwszej kolejności tworzona jest lista, w której będą przechowywane tablice bajtów odpowiadające poszczególnym elementom zawartym w dokumencie.

Kiedy wszystkie elementy są już przetworzone na tablice bajtów (zajmuje się tym metoda getByte) obliczana jest łączna długość wszystkich tablic. W ostatnim kroku wszystkie tablice łączone są w jedną całość, do której dodawana jest na początku łączna długość i na końcu bajt zerowy.