

Continue work on MongoDB support for Slick

Abstract

Slick is a database query and access library for Scala. It allows you to work with data almost as if you were using Scala collections. For now Slick supports top relational databases such as MySQL, Postgres, MSSQL Server, and Oracle. Unfortunately there is no support for MongoDB probably most popular NoSql database. On GSOC 2014 there was a project that prepared base for MongoDB Driver in Slick. That contribution includes connection handling, plain JSON queries and lifted embedding queries for non-nested objects. Previous implementation is based on relational model in Slick and Casbah official MongoDB driver for Scala. My proposal is to update and improve previous implementation and make missing piece, hence support for nested documents, arrays and DBRefs which are essential in mongo databases and in practice most of mongo documents use it. To do that it is necessary to make some changes in Slick internals. My idea is to extend Table, TableQuery classes and make it more flexible. Also projection needs changes hence, there is a need to add new features to Type and Shape classes. Moreover, I would like to provide Slick syntax for querying nested objects.

Benefits

MongoDB is one of the most popular NoSql databases. Support for MongoDB will increase range of potential use cases for Slick and would make that platform more flexible.

About mongo

MongoDB is a document-oriented database hence, there is no separation between data and schema. Each document has dynamic schema and other documents, arrays and references can be nested into it. Documents are stored in BSON format, which is an extension of JSON format.

Related Work

There is a prototype of mongo driver on <https://github.com/dvinokurov/>. Implementation includes connection handling, plain JSON queries and lifted embedding queries for non-nested objects. That project is based on Casbah official Scala driver for MongoDB. I would like to continue that project. I already rebased that branch with changes from current master branch. You can find it on my [Github](#) .

Plan

Improve current implementation

First part of my work on GSOC would be improving current implementation. There are almost thirty TODO-s left in mongo driver code. It would be necessary to do tasks such as:

- add missing methods for instance in MongoBackend in trait DatabaseFactoryDef there is only one method to create connection with database(by URL) and there is a need for implementing method such as forDriver, forConfig etc. That's only one example
- remove unused deprecated or code
- add missing documentation
- add new dependencies if it is needed

Since last year session handling in slick changed to Action based API. It is necessary to implement new features added to RelationalProfile, hence RelationalActionComponent and BasicActionComponent. MongoBackend needs implementation new features from DatabaseComponent.

Implementing nested objects

1.Data structures

The hardest part of project is to add support for:

- DBRefs - references to documents in different collections
- Embedded Objects - MongoDBObject that contains another MongoDBObject
- Embedded Arrays (Scala List/Set) arrays nested within other arrays

Currently, mapping in slick is done by extension of abstract class Table. In such a structure we can define columns which are wrappers for Scala values. That concept is working very well with relational model but is not enough for document-oriented model because of existence more complex and flexible types I mentioned above. Columns in table can only store primitive values like Int or String. My proposition is to create new data structure Document (naming can be changed) which would be extension of Table class. That structure would let you mix columns with other Documents, arrays and DBRefs. In such a way we could define hierarchical schema of document in particular collection. I see a need for adding new features to projection hence, ProvenShape returned by projection should take nested tuples of values as parameters. Following, we need create DocumentQuery class that would be extension of TableQuery (that class would representing document in database and add some operations that can be performed on Documents). Not sure but probably it would be good to make something like DocumentNode that would be Node representing Document in database. Probably it is necessary to make some changes in Rep types. It would be good to provide new profile for instance NoSQLProfile. Such an abstraction could be used by other NoSQL databases. Here I created some prototype how it would look like in simple hello world program [[link](#)]. I believe that concept of DBRef is similar to foreign key and I would try to reuse code of slick to make support to it. Making schema of document could be similar to that in [Mongoose](#).

2.Querying documents

Querying mongo documents from Slick would be still based on Casbah official Scala driver for mongo. Based on what is already done I would like to provide support in MongoInsertInvokerComponent for inserting values to nested document and MongoQuery

for querying such a document. Important in that part is to implement Slick syntax for querying documents.

Schedule

13 April – 30 May Coursera Principles of Reactive Programming

27 April – 25 May - establish communication channels with mentor, setting up environment (install necessary tools) and agree on working methodology.

I believe that top priority in this project would be working with Type system of slick to provide flexibility. That part of the project must be finished completely.

1-3 week (25 May-14 June)

- implementing action based API in MongoDB implementation
- continue working on improving current implementation add missing comments

4-8week (22 June-19 July)

- implement new classes hence, Document, DocumentQuery, NewShape
- write unit tests for new classes

8-10week (19 July - 2 August)

- implementation of querying and inserting nested documents in mongo

11-13 week (3august-23august)

- writing tests and documentation
- provide sample code

Working methodology

I would like to stay in touch with mentor before and during GSOC. Good idea is to establish channels of communications like email or Skype (or other tools used by mentor). I noticed that some of participants of GSOC write on mailing list weekly results and I would like to do that as well(mentor would know exactly where I am in a project and what problems I had). If it possible I would like to have frequent code reviews.

About me

My name is Adam Kozuch and I'm a student of Informatics on University of Gdansk(Poland). Currently I'm on first year of master studies. My time zone is UTC+01:00. My email address is adam.kozuch@gmail.com. My Github account <https://github.com/adamkozuch>.

Professional Experience

Alan Systems 2013.07-2013.08 Software Developer Intern - Working with c# and Entity Framework

Polskapresse 2014.12-2015.02 Software Developer Intern - Working with PHP and MySQL (strong accent on working with database)

Skills

-good knowledge of c#

-PHP and MySQL

-MSSQL Server and ORM(Entity Framework)

-MongoDB – NoSql course most of it was about mongo. On course I was doing things like imports, queries or MapReduce. I was working with pure mongo and Java Driver. You can see effects on my [Github](#) (visuals are in polish because I did exercises in polish and recently translated it into English, original repository you can find on my [BitBucket](#))

-Scala – learning Scala for couple of months and have strong understanding of courser Functional Programming Principles in Scala. To be sure that I can handle a project I decided to do a coursera [Principles of Reactive Programming](#) (13 April-30 May)

-JVM – basic understanding (currently taking course in it)

-sbt – basic understanding

-git – intermediate - I was working with it on my internships and during my open source contributions

-slick- my knowledge of slick include understanding hello-slick-3.0 Activator, internals of MongoDB driver implementation and basic knowledge of Slick type system

Open Source

My open source experience includes couple of contributions to Scala libraries such as cassovary, scalding and summingbird. You can see it on my Github.

English

I have knowledge of written and spoken English. I'm not fluent but I'm able to talk in real time (we can arrange Skype talk if you want to know more about me.

Other commitments

During GSOC, I am not going to have any jobs or vacations. I treat that project very seriously and if all works out I would like to make that project part of my master's thesis I have a plan to work 13 weeks. In the first 2 weeks I would like to work 30 hours a week (finishing coursera and classes at university). Another eleven weeks I would like to work at least 45 hours a week.

If you have any question just contact me via email or Skype. If you are not sure that I can handle a project just give me an issue that I can prove my skills. If you found some mistakes in my proposition or you think I should do things differently or should plan my time differently I'm open to suggestions. I'm aware that this project is tough and demands hard work and creativity. I have both things.