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, support top databases such as MySql, Postgres, MSSQL Server, and Oracle. Unfortunately there is no support for Mongo DB probably most popular NoSql database. On GSOC 2014 project prepared base for MongoDB[1]. That contribution include connection handling, plain JSON queries and lifted embedding queries for non-nested objects. Previous implementation is based on relational model and Casbah driver. My proposal is to update and improve previous implementation and make missing piece, hence support for nested documents, arrays and references to other documents which are essential in mongo databases and in practice most of mongo collections 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 need changes hence, there is a need to add new features to Type and Shape classes.

Benefits

MongoDB is one of the most popular NoSql databases.[1] Support for MongoDB would increase range of potential use cases for Slick and would made that platform more flexible.

About mongo

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

Related Work

There is a implementation on https://github.com/dvinokurov/. Implementation include include 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 implementation with current changes to slick. You can find it on my Github https://github.com/adamkozuch/slick/tree/tmp/mongodb.

Plan

Improve current implementation

First part of my work on GSOC would be improving current implementation. There is almost thirty TODO-s left in mongo driver code. I 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. Thats only one example.
- remove unused deprecated or code
-add missing documentation
-add new dependencies if it is needed

Since last year couple of thing changed in slick session handling by Action based API. It is necessary to implementing new features added to RelationalProfile, hence RelationalActionComponent and BasicActionComponent. MongoBackend need 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 - MongoDBObjects that contains another MongoDBObjects
-Embedded Arrays (Scala List/Set) arrays nested within other arrays

Currently, mapping in slick is conducted by extension of class by Table abstract class. 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 which would be extension of Table class. That structure would let you mixing columns with other Documents, arrays and Dbrefs.

To do that it is necessary to provide new kind of Shape class that would project schema of Document. Write about type-safety. Following, we need create DocumentQuery class that would be extension of TableQuery (that class would representing document in database). I see a need for extension of ProvenShape class that it would be possible to make a projection on data type such as document(idea is that you can write nested tuples in projection). Not sure but probably it would be good to make something like DocumentNode that would representing Document in database. Probably it is necessary to make some changes in Rep types.

2.Implementing it in mongo
TODO

Schedule

13 April – 30 May Coursera Principles of Reactive Programming
27 April – 25 May - establish communication canals with mentor, setting up environment (install necessary tools) and agree on working methodology.
1-4 week (25 May-21 June may)

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

5-8week (22 June-19 July)
-working with lifted API – implement new types hence, Document, DocumentQuery, NewShape,

8-10week (19 July - 2 August)
-working with mongo integrate new features into it. Add new features from Casbah.

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 canals 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 frequently code review.

About me

My name is Adam Kożuch and I'm a student of Informatics on University of Gdansk(Poland). Currently I'm on first year of master degree. 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 Working with c# and Entity Framework
Polskapresse 2014.12-2015.02 Working with PHP and MySql(strong accent on working with database –projecting and integrating into project)

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 https://github.com/adamkozuch/MongoExcercises/tree/master (visuals are in polish because I did excersises in polish and recently translatet it into English, original repository you can find on my BitBucket https://bitbucket.org/AdamKozuch)
-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)(https://www.coursera.org/course/reactive).
-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 include couple of contributions in Scala libriaries 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 and jobs or vacations. I treat that project very serious and if all work out I would like to make that project part of my master thesis I would like to work 13 weeks. Fist two week 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 to do that I can prove my skills.