MetaModelProposal

MetaModel – uniform data access across datastores

Proposal for Apache Incubator

Abstract

MetaModel is a data access framework, providing a common interface for exploration and querying of different types of datastores.

Proposal

MetaModel provides a uniform meta-model for exploring and querying the structure of datastores, covering but not limited to relational databases, various data file formats, NoSQL databases, Salesforce.com, SugarCRM and more. The scope of the project is to stay domain-agnostic, so the meta-model will be concerned with schemas, tables, columns, rows, relationships etc.

On top of this meta-model a rich querying API is provided which resembles SQL, but built using compiler-checked Java language constructs. For datastores that do not have a native SQL-compatible query engine, the MetaModel project also includes an abstract Java-based query engine implementation which individual datastore-modules can adapt to fit the concrete datastore.

Background

The MetaModel project was initially developed by eobject.dk to service the DataCleaner application (http://datacleaner.org). The main requirement was to perform data querying and modification operations on a wide range of quite different datastores. Furthermore a programmatic query model was needed in order to allow different components to influence the query plan.

In 2009, Human Inference acquired the eobjects projects including MetaModel. Since then MetaModel has been put to extensive use in the Human Inference products. The open source nature of the project was reinforced, leading to a significant growth in the community.

MetaModel has successfully been used in a number of other open source projects as well as mission critical commercial software from Human Inference. Currently MetaModel is hosted at http://metamodel.eobjects.org.

Rationale

Different types of datastores have different characteristics, which always lead to the interfaces for these being different from one another. Standards like JDBC and the SQL language attempt to standardize data access, but for some datastore types like flat files, spreadsheets, NoSQL databases and more, such standards are not even implementable.

Specialization in interfaces obviously has merit for optimized usage, but for integration tools, batch applications and or generic data modification tools, this myriad of specialized interfaces is a big pain. Furthermore, being able to query every datastore with a basic set of SQL-like features can be a great productivity boost for a wide range of applications.

Initial goals

MetaModel is already a stable project, so initial goals are more oriented towards an adaption to the Apache ecosystem than about functional changes.

We are constantly adding more datastore types to the portfolio, but the core modules have not had drastic changes for some time.

Our focus will be on making ties with other Apache projects (such as POI, Gora, HBase and CouchDB) and potentially renaming the 'MetaModel' project to something more rememberable. This includes comply with Apache Software Foundation license for third party dependencies.

Current status

Meritocracy

We intend to do everything we can to encourage a meritocracy in the development of MetaModel. Currently most important development and design decisions have been made at Human Inference, but with an open window for anyone to participate on mailing lists and discussion forums. We believe that the approach going forward should be more encouraging by sharing all the design ideas and discussions in the open, not only just the topics that have been "dragged" into the open by third parties. We believe that meritocracy will be further stimulated by granting the control of the project to an independent committee.

Community

The community around MetaModel already exists, but we believe it will grow substantially by becoming an Apache project. With MetaModel used in a wide range of both open and closed source application, both at Human Inference (HIquality MDM), it's open source projects DataCleaner, SassyReader and AnalyzerBeans and by other parties (such as the Quipo data warehouse automation project), we believe that the critical mass to sustain a community is there.

Core developers

MetaModel was founded by Kasper Sørensen in 2009. Later it was incorporated as a core library by Human Inference, meaning that more than 20 developers have been involved in its making in this commercial setting. Furthermore a smaller number of contributors have submitted patches for the library. Others have started building other interesting data-oriented libraries on top of MetaModel, for instance the 'vasc' open source project by Willem Cazander, which is an implementation of the Java Persistence API (JPA) for all the datastores supported by MetaModel.

Alignment

MetaModel already makes good usage of existing Apache projects such as POI, CouchDB and OpenOffice. Furthermore developers from the Apache Gora project have indicated a need for a project like MetaModel to solve specific uniform datastore access needs.

Known risks

Orphaned products

The contributors and the contributing organization (Human Inference) have a very strong dependence on MetaModel already and will continue to have that for a long time. The continued need for this vendor to support new types of datastores and maintain existing functionality will ensure that MetaModel is not in the risk of being orphaned.

Inexperience with Open Source

MetaModel is already open source, and has been so for many years. Main contributors of the project have also contributed to other open source projects such as DataCleaner and Apache Mahout. The openness of Apache is arguably more extensive, but we are only encouraged and delighted to be handling the project in a more open manner.

Homogenous Developers

Frequent committers are currently located in Denmark, The Netherlands and India. They are used to working in a distributed environment.

Reliance on Salaried Developers

Initial committers for MetaModel will depends on salaried based developers to contribute to this project, but given the dependence on MetaModel from both commercial and open source projects, the project would continue without issue if no salaried developers contributed to the project.

The goal is build diverse community to contribute back to MetaModel project.

Relationship with Other Apache Products

MetaModel depends on several Apache products including: commons-lang, commons-io, commons-codec, http-components, POI, CouchDB, OpenOffice and XMLBeans.

Furthermore MetaModel is built by Apache Maven.

An Excessive Fascination with the Apache Brand

The ASF has a strong brand, and that brand is in itself very attractive.

We are interested in joining the ASF in order to increase our contacts and visibility in the open source world. Furthermore, we have been enthusiastic users of Apache Software Foundation projects, and would feel honored by getting the opportunity to join and contribute back to the community.

Documentation

Information on MetaModel can be found at: http://metamodel.eobjects.org

Initial source

MetaModel has been developed since 2009 and have undergone a couple of major changes (indicated by the 2.x and 3.x versions).

The code is used in mission critical systems and is considered very stable and high performing.

The source includes a fork of the xBaseJ project's code, which will be removed upon incubation. This code was originally GPL licensed, but granted with a special license to MetaModel to be forked and relicensed using the current LPGL license of MetaModel.

Removal of the xBaseJ code will effectively mean that the Apache variant of MetaModel will not have support for dBase database files. We imagine that the dBase module could live on as a separate pluggable module under the LGPL license, outside of Apache.

External dependencies

The dependencies all have Apache compatible licenses. These include BSD and MIT licensed dependencies.

Required resources

Mailing lists

- metamodel-private (with moderated subscription)
- metamodel-dev
- metamodel-commits

Source Repository

A subversion (http://svn.apache.org/repos/asf/incubator/metamodel/trunk/) or git (https://git-wip-us.apache.org/repos/asf/incubator-metamodel.git) repository is needed.

Currently MetaModel's code is hosted at http://eobjects.org/svn/MetaModel but will be moved to an Apache repository.

Issue tracking

JIRA MetaModel (METAMODEL)

Other resources

We would like to have wiki page located at: http://wiki.apache.org/MetaModel

In later development phase a set of database servers (specifically MongoDB, CouchDB, MySQL, PostgreSQL, MS SQL Server (Express), Firebird) should be made available for integration testing. Currently this is done internally at Human Inference.

Initial committers

Kasper Sørensen (i.am.kasper.sorensen [at] gmail.com), Project Founder, works at Human Inference

Ankit Kumar (ak.ankitkumar [at] gmail.com), works at Human Inference

Sameer Arora (sameer11sep [at] gmail.com)

Henry Saputra (hsaputra [at] apache.org)

Juan José van der Linden (delostilos [at] gmail.com), works for Quipu

Arvind Prabhakar (arvind at apache dot org)

Matt Franklin (mfranklin at apache dot org)

Noah Slater (nslater at apache dot org)

Sponsors

Champion

Henry Saputra (hsaputra at apache dot org)

Nominated mentors

Henry Saputra (hsaputra at apache dot org)

Arvind Prabhakar (arvind at apache dot org)

Matt Franklin (mfranklin at apache dot org)

Noah Slater (nslater at apache dot org)

Sponsoring entity

The Apache Incubator.