QueryArrow: Semantically Unified Query and Update of Heterogeneous Data Stores

Hao Xu
xuhao@renci.org

Ben Keller Antoine de Torcy Jason Coposky

University of North Carolina at Chapel Hill

June 14, 2017
Acknowledgement

This research is partially supported by

- The National Science Foundation under Grant Number OCI 0940841 "DataNet Federation Consortium" (Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation) and

- iRODS Consortium
Resources

- QueryArrow: https://github.com/xu-hao/QueryArrow fs branch
- QueryArrow Database Plugin:
  http://github.com/xu-hao/temporary-irods-mod qa3 branch
Motivtiing Applications

Big Metadata Challenge:
- Aggregation: integrating metadata from multiple data stores
- Policies: for example, metadata access control
- Unify query and update
- Semantically specification
- Heterogeneity: NoSQL, Indexing
- Improve Testing, Modularity, Robustness
QueryArrow is a semantically unified query and update system for heterogeneous metadata store

- **QueryArrow Service (QAS):** Register Databases and Execution of QAL
- **QueryArrow Language (QAL):** Configuration, QL/DML
- **QueryArrow Plugins (QAP):** Mappings between QAL and Databases
- **iRODS QueryArrow Database Plugin**
<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum QAP</td>
<td>aggregation</td>
</tr>
<tr>
<td>Translation QAP</td>
<td>policy support</td>
</tr>
<tr>
<td>Cache QAP</td>
<td>caching</td>
</tr>
<tr>
<td>Mutable Map QAP</td>
<td>in-memory mutable map</td>
</tr>
<tr>
<td>Immutable Map QAP</td>
<td>in-memory immutable map</td>
</tr>
<tr>
<td>ElasticSearch QAP</td>
<td>interfacing with ElasticSearch</td>
</tr>
<tr>
<td>Neo4j QAP</td>
<td>interfacing with Neo4j</td>
</tr>
<tr>
<td>PostgreSQL QAP</td>
<td>interfacing with Postgres</td>
</tr>
<tr>
<td>SQLite3 QAP</td>
<td>interfacing with SQLite3</td>
</tr>
<tr>
<td>CockroachDB QAP</td>
<td>interfacing with CockroachDB</td>
</tr>
<tr>
<td>FileSystem QAP</td>
<td>file system</td>
</tr>
</tbody>
</table>
**QAP Composition**

Higher-order plugin

- **Cache QAP**
- **Translation QAP**
- **Sum QAP**

- QAP 1
  - Data store 1
- QAP 2
  - Data store 2
- QAP 3
  - Data store 3
QAL Syntax

t ::= p | v \hspace{1cm} \text{terms}
a ::= P(t_1, \ldots, t_n) \hspace{1cm} \text{atom}
c ::= a | \text{insert } a | \text{delete } a | g c | b
| 1 | 0 | c \oplus c | c \otimes c \hspace{1cm} \text{command}

pt ::= \text{input? output? key? ty} \hspace{1cm} \text{parameter type}
R ::= \text{rewrite } a c | \text{rewrite insert } a c
| \text{rewrite delete } a c
| \text{predicate } P(pt_1, \ldots, pt_n)
| \text{import } \ldots
| \text{export } \ldots \hspace{1cm} \text{configuration}
"store-and-heap" Monad
https://github.com/xu-hao/CertifiedQueryArrow
QueryArrow Service

- Local
- Unix domain socket
- TCP
- HTTP
QueryArrow Database Plugin

- Plugs into iRODS, requires no changes to iRODS core except for bug fixes
- Talks to QueryArrow Service
- Implements 84 db plugin functions
- Backward compatibility with GenQuery
- Tested on 4.2.x:
  - 4.2.0: PostgreSQL QAP passed python tests suite except 2 tests using ssl plugin on Ubuntu 16.04/Centos 7
  - 4.2.1: PostgreSQL QAP passed python tests suite except 2 tests using ssl plugin on Ubuntu 16.04/Centos 7 is running
Run iRODS install script through QueryArrow
- Only Postgresql
- Combined Postgresql and Neo4j