

Provenance Management in Swift

Luiz Gadelha

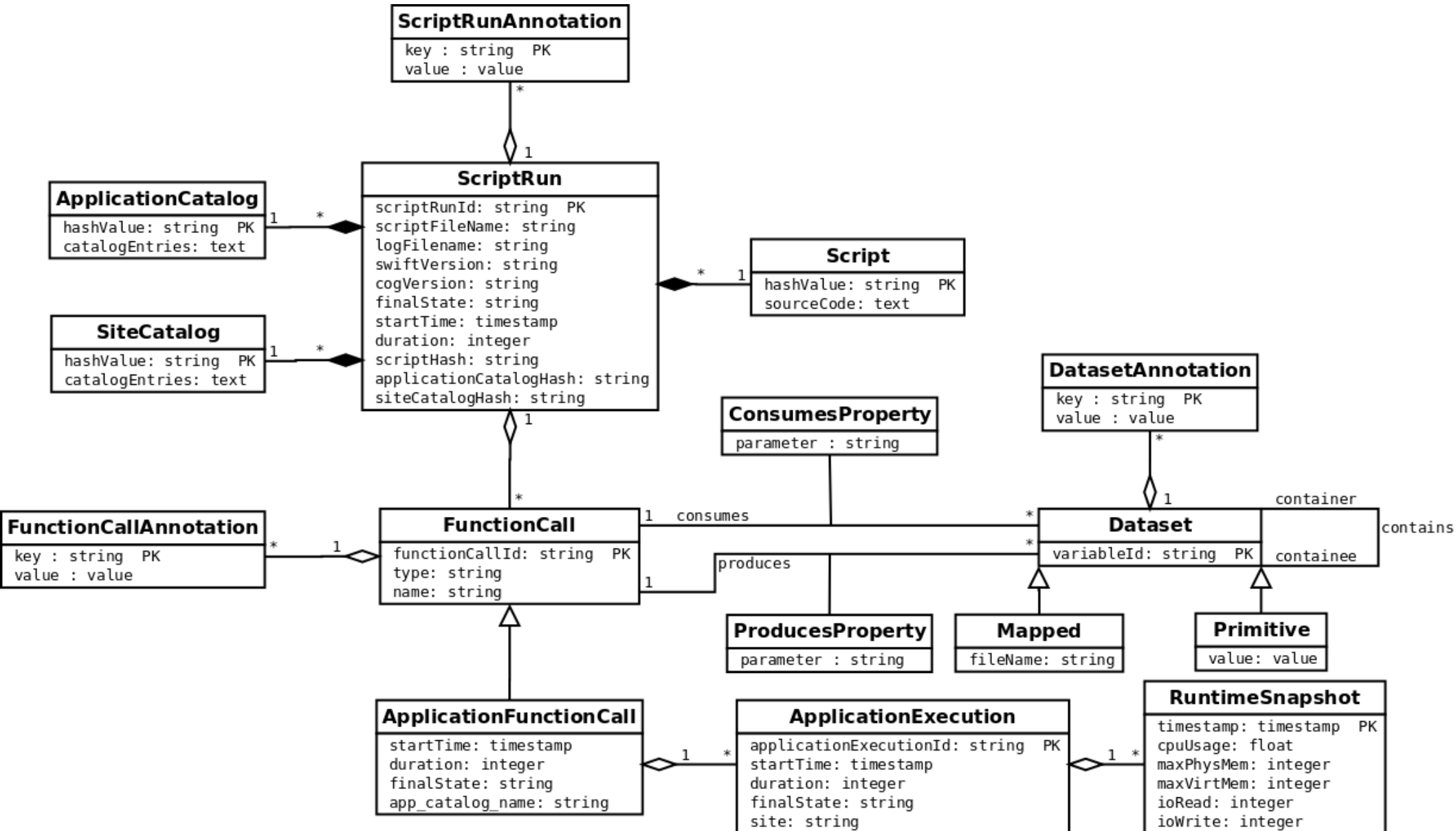
LNCC - National Laboratory for Scientific Computing
Petrópolis, Rio de Janeiro
Brazil

lgadelha@lncc.br

Provenance Tools

- Tools available:
 - Scripts for creating the provenance database
 - Scripts for extracting provenance from log files
 - SQL functions and stored procedures
 - SPQL query language implementation (ANTLR)
- Available at Swift SVN repository:
 - <https://svn.ci.uchicago.edu/svn/vdl2/provenancedb>
- Tutorial:
 - <http://www.Incc.br/~lgadelha/README.html>

Provenance Database Schema



Provenance Database Creation

- Database creation (requires PostgreSQL):

```
$ createdb provdb
```

```
$ psql -f prov-init.sql provdb
```

```
$ psql -f pql-functions.sql provdb
```

Provenance Tools Configuration

- Stored at etc/provenance.config
- Configure log files directory:

```
export LOGREPO=~/.swift-logs/
```

- Configure database connection:

```
export SQLCMD="psql -h db.ci.uchicago.edu -U provdb provdb"
```

Enabling Provenance in Swift

- Should be changed in swift.properties:
 provenance.log=true
 wrapperlog.always.transfer=true
- Should be changed in log4j.properties:
 log4j.logger.swift=DEBUG
 log4j.logger.org.griphyn.vdl.karajan.lib=DEBUG

Importing Provenance to DB

- Make sure provenancedb and log-processing (from Swift) directories are on execution path
- To import provenance information to DB run:
\$ swift-prov-import-all-logs
- To reinitialize DB and import:
\$ swift-prov-import-all-logs rebuild

Discussion

- Good default level of detail
- Integration with Swift
 - Moving scripts to Swifts trunk SVN repository
 - Integration with David's swiftlog
 - Embedded DB and online generation
- Integration with other tools and catalogues:
 - Galaxy
 - Globus Catalog
- Persistent dataset ids. across runs