Software development for the Rucio Scientific Data Management system

Lev Pambuk / Лев Памбук
Location: Odesa, Ukraine
E-mail: levpambuk@gmail.com
Phone: +38 063 129 40 73
Education: Odesa National University of Technology

Project context

Rucio is a software framework that provides functionality to organize, manage, and access large volumes of scientific data using customisable policies. The data can be spread across globally distributed locations and across heterogeneous data centers, uniting different storage and network technologies as a single federated entity. Rucio offers advanced features such as distributed data recovery or adaptive replication, and is highly scalable, modular, and extensible. Rucio has been originally developed to meet the requirements of the high-energy physics experiment ATLAS, and is continuously extended to support LHC experiments and other diverse scientific communities.

With this summer fellow project we aim to bring critical developments to the common parts of Rucio, shared by many communities and experiments.

Mentors: Martin Barisits (CERN EP), Mario Lassnig (CERN EP)

Milestones and work plan

Milestone 0: Set up [Week 1]

- Getting acquainted with the development team and engineering practices
- Reading documentation
- Setting up the development environment

Milestone 1: Introductory work [Week 2]

- Typos in docstrings for Python API of DIDClient
  https://github.com/rucio/rucio/issues/6225

- Remove Link to Deprecated Google Form in Rucio WebUI
  https://github.com/rucio/rucio/issues/6179

- Import Statements Should Be at Top of File
  https://github.com/rucio/rucio/issues/6090
Milestone 2: Client developments [Week 3]

*Improve error reporting during upload when reusing a deleted DID*
https://github.com/rucio/rucio/issues/6091

*Webdav Protocol stat does not return data as specified*
https://github.com/rucio/rucio/issues/5977

*Rucio client lists replicas in wrong order*
https://github.com/rucio/rucio/issues/6141

Milestone 3: Operations developments [Week 4]

*Convert containers in Lifetime Model exceptions to datasets*
https://github.com/rucio/rucio/issues/6147

*Allow to declare suspicious replicas by RSE and LFN*
https://github.com/rucio/rucio/issues/5906

Milestone 4: Core developments [Week 5-8]

*Add parameters to daemons’ stop() function*
https://github.com/rucio/rucio/issues/6075

*rse_settings dictionary datatype*
https://github.com/rucio/rucio/issues/5972

*Add size information to list_rules()*
https://github.com/rucio/rucio/issues/5978

*Customisable replica sorter algorithm*
https://github.com/rucio/rucio/issues/6114

*Heartbeats endpoints don’t work*
https://github.com/rucio/rucio/issues/6180

Milestone 5: Transfer developments [Week 8-10]

*Add option to specify OIDC token for communication with the storage element*
https://github.com/rucio/rucio/issues/6128

*Dynamic submission bulk size to FTS*
https://github.com/rucio/rucio/issues/6228

Milestone 6: Conclusions [Week 11]

*Summary report and slides*
Present work at IRIS-HEP and CERN seminars