Software development for the Rucio Scientific Data Management system

Lev Pambuk, Odesa National University of Technology

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

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.

Motivation

- The LHC experiments are, combined, managing nearly an **exabyte** of data. The planning modernization of the LHC will increase this volume to an unmanageable size. New mechanisms and techniques are **required** to manage **more efficiently** storage resources and deliver data to processing endpoints.

Goal of the project

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

Milestone 1: Introductory work for the first weeks

Getting used to the Rucio developer environment

- Correcting typos in docstrings for Python API of DIDClient
- Removing link to deprecated Google Form in Rucio WebUI
- Rearranging of imported statements in file

Milestone 2:
Client developments

[Week 3]

- Improving error reporting during upload when reusing a deleted DID
- Fixing Webdav Protocol stat in order to return data as specified
- Fixing listing order of replicas in Rucio client

Milestone 3:
Operations developments

[Week 4]

- Working on converting containers in Lifetime Model exceptions to datasets
- Allowing to declare suspicious replicas by RSE and LFN

Milestone 4:
Core developments

[Week 5-8]

- Adding parameters to daemons’ stop() function
- Modification of rse_settings dictionary datatype
- Adding size information to list_rules()
- Adding customisable replica sorter algorithm
- Fixing Heartbeats endpoints

Milestone 5:
Transfer developments

[Week 8-10]

- Adding option to specify OIDC token for communication with the storage element
- Development of dynamic bulk size submissioning to FTS

Timeline of the planned work

- 19 June - 3 September

Thank you for listening!