PIC
port d'informació
científica

RUCIO

Rucio JupyterLab Extension status

7th Rucio Community Workshop, 1st October 2024

Francesc Torradeflot, Enrique Garcia, Giovanni Guerrieri

Outline

- Introduction
- Overview
  - Features
  - Architecture
- Current status:
  - Latest developments
  - Future work
- Use cases
- Demo

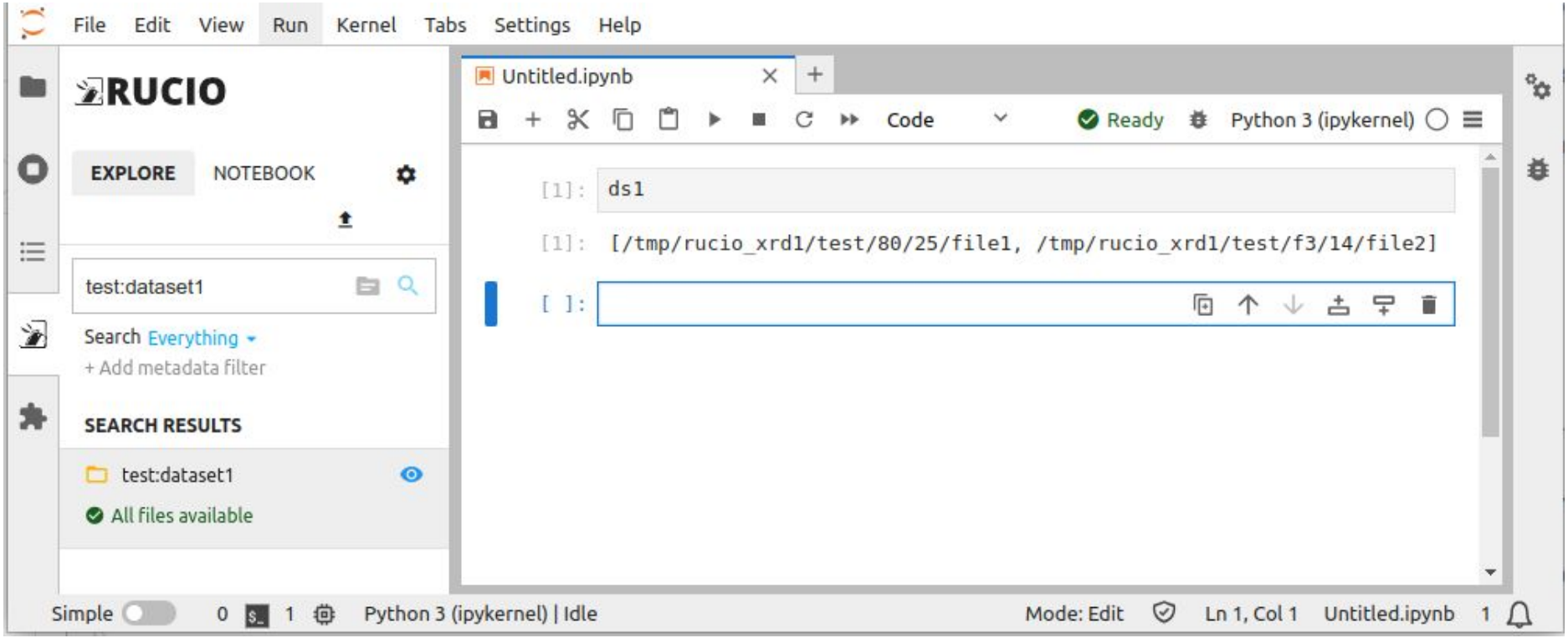
Introduction

The Rucio JupyterLab Extension was developed by Muhammad Aditya Hilmy in 2020 in the context of a Google Summer of Code internship. He is still (by far) the main contributor to the project.

It was adopted as an official Rucio component in 2024 and is now maintained by Francesc Torradeflot and Enrique Garcia.

Its main purpose was to provide the scientists an easier access to the data, by bridging the gap between the analysis tool (JupyterLab - SWAN) and the data lake (Rucio).

Overview

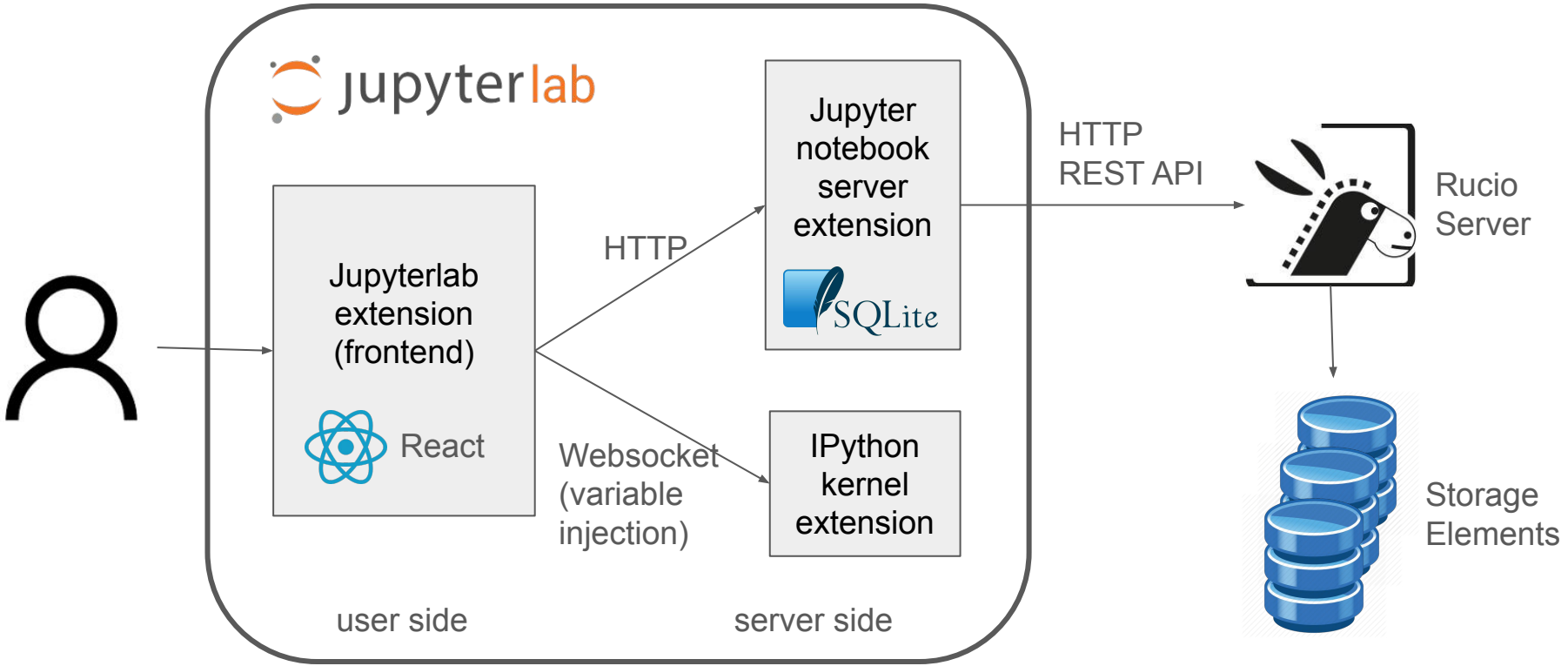

File Edit View Run Kernel Tabs Settings Help
RUCIO
EXPLORE
NOTEBOOK
test:dataset1
Search Everything
+ Add metadata filter
SEARCH RESULTS
test:dataset1
All files available
Untitled.ipynb
Code
Ready
Python 3 (ipykernel)
[1]: ds1
[1]: [/tmp/rucio_xrd1/test/80/25/file1, /tmp/rucio_xrd1/test/f3/14/file2]
[ ]:
Simple
Python 3 (ipykernel) | Idle
Mode: Edit
Ln 1, Col 1
Untitled.ipynb

Overview - Features

- Browse Rucio data from the Lab sidebar
- Replicate data with just one click
- Resolves file path automagically
- Inject path to notebook as a variable
- Supports three methods of authentication (currently)
  - Username & Password
  - X.509 User Certificate (or Proxy)
  - OIDC tokens (with limitations)
- Supports two modes of operation:
  - **Replica mode:** uses network-attached storage as a Rucio Storage Element (RSE), utilizes Rucio's file transfer capability.
  - **Download mode:** downloads data directly to the user's directory using Rucio clients.

Content from Muhammad's Rucio-SWAN Integration Project slides

Overview - Architecture

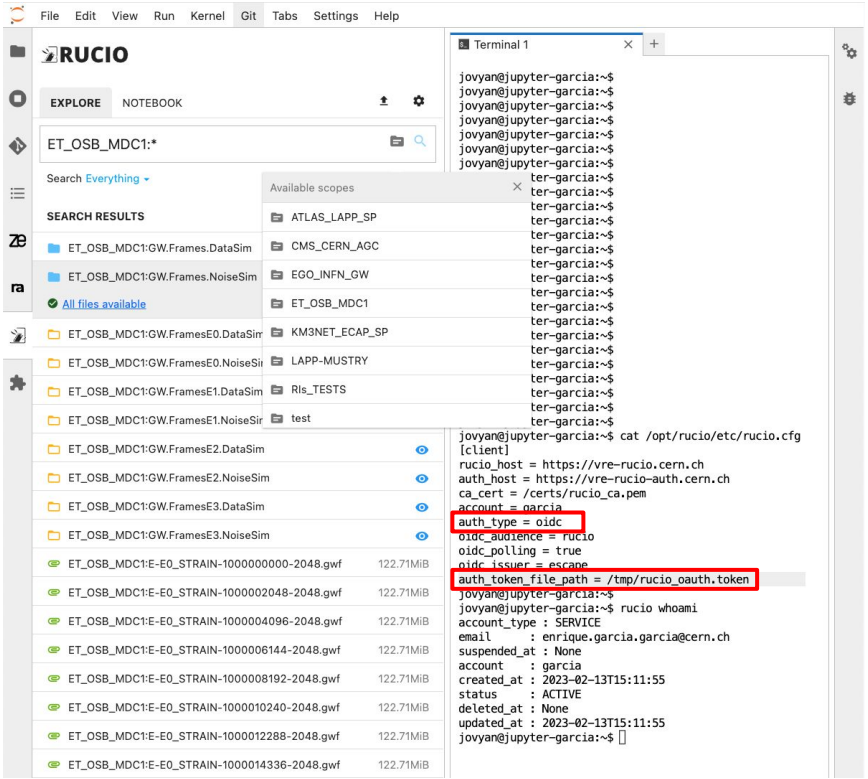


Current Status

- Latest developments
  - #27 Migration from JupyterLab v3 to v4 (Done)
  - #29 Creation of a test environment (relying on Rucio's test env) (Done)
  - #36 Apply filters in DID search. (credit to Georgy Skorobogatov) (In Review)
- Future work
  - #35 Include docker image publication in CI/CD
  - #39 Improve handling of errors in Rucio REST API calls

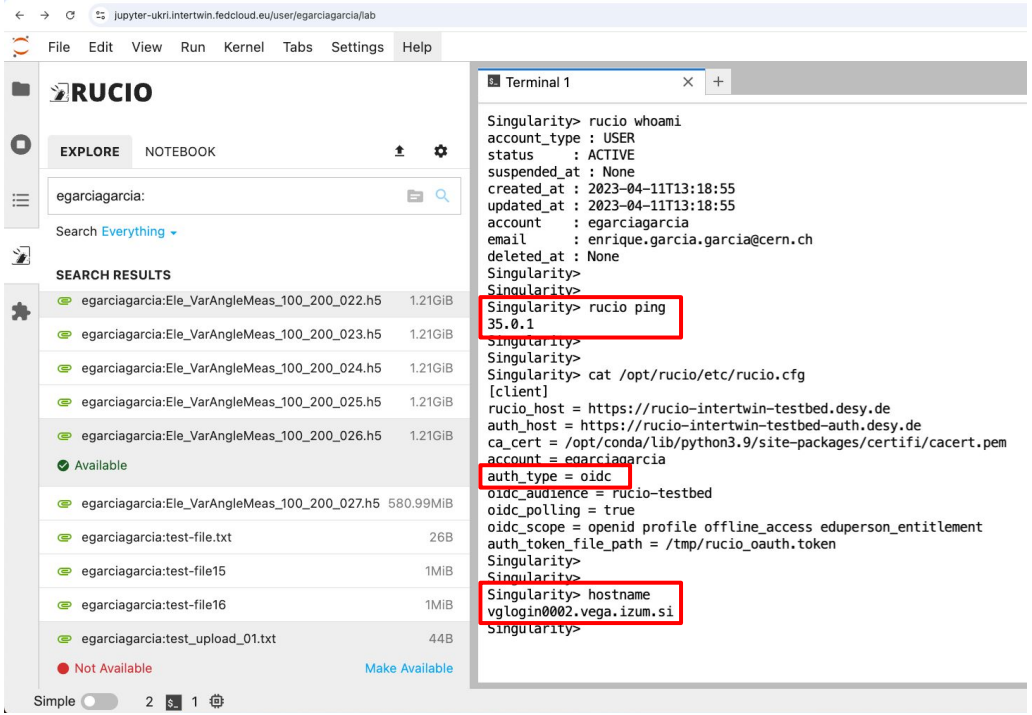
Use Cases: CERN VRE

- CERN VRE is an analysis facility developed within ESCAPE project
  - First environment hosting the extension since it's development.
- Extension connected to ESCAPE Rucio Instance (deployed @CERN)
  - Interacts with an EOS instance
    - Mounted in the jlab localhost
    - Configured as a RSE
  - Uses OIDC tokens as auth method
  - Tested with 1.28.0 → 1.30.0 releases
  - Release v34 raised some problems with OIDC authentication
    - X509 proxy works fine, though.
- Extension is currently being integrated in the CERN/SWAN framework.



Use Cases: interTwin

- EU project building a prototype of a Digital Twin Engine
  - Data infrastructure based on the ESCAPE Data Lake (RUCIO + FTS + Federated AAI)
  - Rucio instance deployed @ DESY
- Rucio extension available @ VEGA HPC Center
  - Uses OIDC as auth method
  - Interacts with a dCache instance
    - Volume mounted on VEGA
    - Configured also as a RSE



Thank you