



Contribution ID: 136

Type: **Lightning talk**

SWAN, Rucio, and Jupyter

Tuesday, 26 January 2021 09:15 (15 minutes)

The LHC experiments at CERN produce an enormous amount of scientific data. One of the main computing challenges is to make such data easily accessible by scientists and researchers. Technologies and services are being developed at CERN and at partner institutes to face this challenge, ultimately allowing to turn scientific data into knowledge.

SWAN (Service for Web-based ANalysis) is a platform allowing CERN users to perform interactive data analysis directly using a web browser. This service builds on top of the widely-adopted Jupyter Notebooks. It integrates storage, synchronisation, and sharing capabilities of CERNBox and the computational power of Spark/Hadoop clusters. Both scientists at CERN and at partner institutes are using SWAN on a daily basis to develop algorithms required to perform their data analysis. Full analyses can be performed using Notebooks as long as all the required data are available locally.

The Rucio data management system was principally developed by the ATLAS experiment to deal with Exabytes of data in a scalable, modular, and reliable way. Nowadays, Rucio has become the de-facto data management system in High Energy Physics and many other scientific communities such as astronomy, astrophysics, or environmental sciences are evaluating and adopting it.

In the Exabytes-scale era, the challenge to move large amounts of data in the local file system of a Notebook is faced on a daily basis by each individual scientist, causing duplication of effort and delaying the analysis results. The integration of Rucio in the Jupyter Notebook environment is a challenging but necessary R&D activity from which the worldwide scientific community would greatly benefit.

Starting from an idea at the previous CS3 conference, in less than a year a JupyterLab extension was developed and tested in the context of Google Summer of Code and the EU-funded project ESCAPE. This extension integrates Rucio functionalities inside the JupyterLab UI, to link experiment data into notebooks that require them, and to transparently make the data present in the ESCAPE DataLake available using Rucio.

Primary authors: HILMY, Muhammad Aditya; LASSNIG, Mario (CERN); BARISITS, Martin (CERN); Dr DI MARIA, Riccardo (CERN); CASTRO, Diogo (CERN); TEJEDOR SAAVEDRA, Enric (CERN); BOCCHI, Enrico (CERN)

Presenter: LASSNIG, Mario (CERN)

Session Classification: Novel Data Science Environments

Track Classification: Main session: User Voice: Novel Applications, Data Science Environments & Open Data