



Contribution ID: 112

Type: **Innovation Grants**

Extending Rucio to support external metadata catalog

In this talk, I will present the progress made in deploying and customizing Rucio as a scalable, metadata-integrated prototype Data Lake as a solution for Data Management (DM) of the Interoperable Data Lake (IDL) project. Rucio is an open-source DM software designed for large-scale scientific experiments such as those in high-energy physics (HEP). Initially developed by CERN for the ATLAS experiment, it is a robust solution for managing any type of data across geographically distributed storage systems.

The talk will focus on a custom did-metadata plugin developed to enable communication with an external database, AyraDB, to handle project-specific metadata with a predefined structure. This communication was achieved using an API client developed by CherryData, supporting both ingestion and query operations.

Finally, we developed a custom Rucio client to enable relevant features for the IDL project, not allowed by the default client, such as a combined file upload and metadata assignment.

An overview of the future steps will be presented, particularly focusing on the integration of a JupyterHub system into the Kubernetes cluster that seamlessly communicates with the Data Lake.

Giorno preferito

12 Dicembre Mattina

Primary author: PACIOSELLI, Luca (Istituto Nazionale di Fisica Nucleare)

Presenter: PACIOSELLI, Luca (Istituto Nazionale di Fisica Nucleare)