



Contribution ID: 4

Type: **not specified**

MetaCat - Metadata Catalog for Rucio-based Data Management Systems (Remote)

Thursday 10 November 2022 16:15 (15 minutes)

Metadata management is one of three major areas of scientific data management along with replica management and workflow management. Metadata is the information describing the data stored in a data item such as a file or an object. It includes the data item provenance, recording conditions, format and other attributes. MetaCat is a metadata management database designed and developed for High Energy Physics experiments. As Rucio is becoming a popular product to be used as the replica management component, MetaCat was designed to be conceptually compatible with Rucio and to be able to work with Rucio as the replica management component.

Main objectives of MetaCat are:

- Provide a flexible mechanism to store and manage file dataset metadata of arbitrary complexity
- Provide a mechanism to retrieve the metadata for a file or a dataset
- Efficiently query the metadata database for files or datasets matching user defined criteria expressed in terms of the metadata
- Provide a transparent mechanism to access external metadata sources to logically incorporate the external metadata into queries without copying them

One of the MetaCat features is MQL - metadata query language developed specifically for this application. The article will discuss the architecture, functionality and features of MetaCat as well as the current status of the project.

Author: MANDRICHENKO, Igor (Fermi National Accelerator Lab. (US))

Presenter: MANDRICHENKO, Igor (Fermi National Accelerator Lab. (US))

Session Classification: Technology Sessions