RUCIO support DB test cases
Andrea Adelfio for WP4 (INFN Perugia)

Spoke 3 Technical Workshop, Trieste October 9 / 11, 2023

Finanziato
dall'Unione europea
NextGenerationEU

Ministero
dell'Università
e della Ricerca

Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

Technical Objectives, Methodologies and Solutions

Develop a new Archiving System that satisfies these requirements:

- **data interoperability with standard formats for data from observations, analyses, simulations, etc**
- **access policy and replication rules**
- **data integrity and validation**
- **data federation**
- **efficient data management based on metadata content**

DATA LAKE based on the RUCIO technology, a Data Management System used to organize large quantities of scientific data, capable of efficient data discovery and categorization using metadata.

Finanziato
dall'Unione europea
NextGenerationEU

Ministero
dell'Università
e della Ricerca

Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

Accomplished Work, Results

September 2023:
-**discovery of the RUCIO software through INFN RUCIO testbed and its documentation**
-**practice through available tutorials**
-**definition of use cases to obtain the standard formats to be manipulated**

October 2023:
-**test on data and metadata manipulation within the RUCIO command line interface**
-**discovery and practice of the RUCIO Python library**
-**test on data and metadata manipulation using RUCIO library for Python implementations**
-**look for the suitable use case with its standard format to begin the integration**

Next Steps and Expected Results (by next checkpoint: April 2024)

October 2023:
- define the perimeter of data formats to be manipulated in the data lake (FITS, HDF5, ROOT, VTK…)
- find the ideal data format to begin the <u>software prototyping</u> (e.g. FITS)

November 2023:
- start writing a code with basic functions to access the DB, manipulate and query on metadata

Dicember 2023 – April 2024:
- a Python script based on the RUCIO technology that is capable of managing metadata of collected data and can query storaged data based on the metadata stored in the database.