



Contribution ID: 389

Type: Talk

Distributed Data Management with Rucio for the Einstein Telescope

Wednesday 23 October 2024 14:24 (18 minutes)

Modern physics experiments are often led by large collaborations including scientists and institutions from different parts of the world. To cope with the ever increasing computing and storage demands, computing resources are nowadays offered as part of a distributed infrastructure. Einstein Telescope (ET) is a future third-generation interferometer for gravitational wave (GW) detection, and is currently in the process of defining a computing model to sustain ET physics goals. A critical challenge for present and future experiments is an efficient and reliable data distribution and access system. Rucio is a framework for data management, access and distribution. It was originally developed by the ATLAS experiment and has been adopted by several collaborations within the high energy physics domain (CMS, Belle II, Dune) and outside (ESCAPE, SKA, CTA). In the GW community Rucio is used by the second-generation interferometers LIGO and Virgo, and is currently being evaluated for ET. ET will observe a volume of the Universe about one thousand times larger than LIGO and Virgo, and this will reflect on a larger data acquisition rate. In this contribution, we briefly describe Rucio usage in current GW experiments, and outline the on-going R&D activities for integration of Rucio within the ET computing infrastructure, which include the setup of an ET Data Lake based on Rucio for future Mock Data Challenges. We discuss the customization of Rucio features for the GW community: in particular we describe the implementation of RucioFS, a POSIX-like filesystem view to provide the user with a more familiar structure of the Rucio data catalogue, and the integration of the ET Data Lake with mock Data Lakes belonging to other experiments within the astrophysics and GW communities. This is a critical feature for astronomers and GW data analysts since they often require access to open data from other experiments for sky localisation and multi-messenger analysis.

Primary authors: TANASIJCZUK, Andres Jorge (Universite Catholique de Louvain (UCL) (BE)); LEGGER, Federica (Universita e INFN Torino (IT)); Mr FRONZE', Gabriele Gaetano (INFN Torino (IT) and LIGO-Virgo-Kagra Collaboration (US/IT/JP)); LAVEZZI, Lia; ET AL.

Presenter: LAVEZZI, Lia

Session Classification: Parallel (Track 1)

Track Classification: Track 1 - Data and Metadata Organization, Management and Access