

Contribution ID: 440

Type: Poster

Prototyping a Bulk Data Management System for CTAO with Rucio

Bulk Data Management, including the long-term archiving of massive datasets, is critical for advancing highenergy gamma-ray astrophysics research by ensuring data accessibility and scientific reproducibility. Within the Cherenkov Telescope Array Observatory (CTAO), managing and preserving petabyte-scale data poses unique challenges. To address these challenges, we present our prototyping efforts for the Bulk Data Management System (BDMS), a key sub-system of CTAO's Data Processing and Preservation System (DPPS) designed for long-term preservation. BDMS leverages Rucio —the open-source data management system developed at CERN and follows the Open Archival Information Systems (OAIS) standard to manage the replication of data products between CTAO Data centers, ensure their long-term preservation, and provide an interface to ingest, query, and retrieve these data products.

We provide details on the BDMS architecture and its main functional blocks, namely: Ingest, Data Management (which includes data transfers, track preservation, and monitoring), Archival Storage, File Query and Access, and BDMS Administration. Additionally, we present a couple of use-cases focused on ingest, data management, and metadata handling.

Our prototyping contributions include containerized deployment using Helm charts and continuous integration tests on a Kubernetes (K8s) cluster provided by DESY Computing/Data center; metadata management by implementing a setup to extract and store metadata from raw and simulated data products, thereby enabling high-level dataset queries; and integration with DIRAC for workload management. Finally, we outline our future plans that include integrating Indigo IAM tokens into our prototyping efforts, and setting-up monitoring for BDMS storages and file transfers.

Collaboration(s)

Author: HASAN, Syed Anwar Ul

Co-authors: BILAND, Adrian (ETH Zurich); LI, Hancheng; LINHOFF, Maximilian (TU Dortmund | CTAO); LYARD, Etienne; Dr SAVCHENKO, Volodymyr (EPFL, Switzerland); WALTER, Roland (University of Geneva)

Presenter: HASAN, Syed Anwar Ul

Session Classification: PO-2

Track Classification: Gamma-Ray Astrophysics