



Contribution ID: 294

Type: **Talk**

A Tape RSE for Extremely Large Data Collection Backups

Tuesday 22 October 2024 14:42 (18 minutes)

The Vera Rubin Observatory is a very ambitious project. Using the world's largest ground-based telescope, it will take two panoramic sweeps of the visible sky every three nights using a 3.2 Giga-pixel camera. The observation products will generate 15 PB of new data each year for 10 years. Accounting for reprocessing and related data products the total amount of critical data will reach several hundred PB. Because the camera consists of 201 CCD panels, the majority of the data products will consist of relatively small files in the low megabyte range, impacting data transfer performance. Yet, all of this data needs to be backed up in offline storage and still be easily retrievable not only for groups of files but also for individual files. This paper describes how SLAC is building a Rucio-centric specialized Tape Remote Storage Element (TRSE) that automatically creates a copy of a Rucio dataset as a single indexed file avoiding transferring many small files. This not only allows high-speed transfer of the data to tape for backup and dataset restoral, but also simple retrieval of individual dataset members in order to restore lost files. We describe the design and implementation of the TRSE and how it relates to current data management practices. We also present performance characteristics that make backups of extremely large scale data collections practical.

Primary author: HANUSHEVSKY, Andrew Bohdan (SLAC National Accelerator Laboratory (US))

Co-authors: CHE, Guangwei (SLAC National Accelerator Laboratory); NAKATA, Lance (SLAC National Accelerator Laboratory); YANG, Wei (SLAC National Accelerator Laboratory (US))

Presenter: HANUSHEVSKY, Andrew Bohdan (SLAC National Accelerator Laboratory (US))

Session Classification: Parallel (Track 1)

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