Contribution ID: **214**   Type: **Talk**

Archive Metadata for efficient data colocation on tape

Tuesday 22 October 2024 15:00 (18 minutes)

Due to the increasing volume of physics data being produced, the LHC experiments are making more active use of archival storage. Constraints on available disk storage have motivated the evolution towards the "data carousel" and similar models. Datasets on tape are recalled multiple times for reprocessing and analysis, and this trend is expected to accelerate during the Hi-Lumi era (LHC Run-4 and beyond).

Currently, storage endpoints are optimised for efficient archival, but it is becoming increasingly important to optimise for efficient retrieval. This problem has two dimensions. To reduce unnecessary tape mounts, the spread of each dataset - *the number of tapes containing files which will be recalled at the same time* - should be minimised. To reduce seek times, files from the same dataset should be physically colocated on the tape. The Archive Metadata specification is an agreed format for experiments to provide scheduling and colocation hints to storage endpoints to achieve these goals.

This contribution describes the motivation, the review process with the various stakeholders and the constraints that led to the Archive Metadata proposal. We present the implementation and deployment in the CERN Tape Archive and our preliminary experiences of consuming Archive Metadata at WLCG Tier-0.

Primary author: LEDUC, Julien (CERN)

Co-authors: Dr GUENTHER, Jaroslav (CERN);   AFONSO, Joao (CERN);   SKOVOLA, Konstantina;   DAVIS, Michael (CERN);   BUGEL, Niels Alexander;   OLIVER CORTES, Pablo (CERN);   BACHMANN, Richard (CERN);   ALFAGEME PEREZ, Sergio;   BAHYL, Vladimir (CERN)

Presenter: LEDUC, Julien (CERN)

Session Classification: Parallel (Track 1)

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