Centre de Calcul
de l'Institut National de Physique Nucléaire
et de Physique des Particules

Archival of Raw Images at FrDF

fabio hernandez

Data Facilities and Multisite Processing, Lyon, February 13th, 2025

CONTEXT

- Each observing night, about 2000 science + 500 calibration exposures will be recorded

  each exposure is composed of about 200 `.fits` *files, each one in the range 20 to 30 MB*

  the typical size of a exposure is 6 GB (compressed)

- Approximate sizes for raw dataset

  20 TB per night, 5 PB per year

- Approximate number of files for raw dataset

  500k per night, 150M per year, 1.5B files for the entire survey

- FrDF to archive on tape a full copy of raw data

PROBLEM

- We would like to avoid handling that many small files on tape

  a single file per exposure seems a better granularity for tapes

  2.5k files per night, 750k per year, 7.5M for the entire survey

- FrDF is sized to keep on disk only the raws it is assigned for processing

  since they all belong to the same spatial region, no specific need to move data between tape and disk for preparing a campaign

  that amounts to 40% of the size of the entire raw image dataset

PROBLEM (CONT.)

- Raw images are to be transferred from USDF either to `IN2P3_RAW_DISK` or `IN2P3_RAW_TAPE` (or both?)
- Since raw images are registered into Rucio as individual files (one per sensor) we cannot easily remove them from the RSE

we could implement some service to mediate interaction between dCache and HPSS

that service would perform aggregation of all the relevant sensor files into a single file per exposure

many details still to be sorted out, since this is not the way dCache works with tertiary storage

overall, this seems to add complexity and make the chain fragile

PROPOSAL FOR A WAY FORWARD

- All the sensor files of an exposure would be aggregated at USDF to create a zip file

  the per-exposure zip file would registered into Rucio

- All the per-exposure zip files would be replicated to `IN2P3_RAW_TAPE`

  this automatically copies the file to HPSS and removes the file from disk, but the file remains known to dCache

  a request to read that file automatically triggers a copy from HPSS to dCache disk

  this mechanism is well understood and in production since many years for LHC experiments

PROPOSAL (CONT.)

- Only the tracts assigned to FrDF would be transferred to `IN2P3_RAW_DISK`

  either one file per sensor or, better, one file per exposure if Butler can handle this

- Pros

  we handle the use case of archiving raws separately from the use case of processing raws: all remote facilities would be treated in a homogenous way

  we reduce the number of raw files to replicate and increase their unitary size, which is good for FTS

- Cons

  requires creation and storage of per-exposure file at USDF before registration into Rucio

  the total amount of raw data replicated from USDF would increase

ADDITIONAL INFORMATION

- Experimentation with zip files handling in Rucio done by G. Daues

  documented in DM-20003

  it is possible to add files to a zip file already registered into Rucio

  the constituent files are individually known to Rucio

Questions & Comments