Data Management for extreme scale computing

Gillardo Frédéric
gillardo@lapp.in2p3.fr

11/04/2019
CTA Data Workflow

- Canary Island: Online Process
- Chile: Online Process
- eXtreme Data Cloud
- Calib
- Reconst
- DL3 / High level data
- Scientist Interface
**Use Case Goals**

**Objectives:**
- Archive PBs of data using an ingest method.
- Index files using metadata contained in the header files
- Query files using metadata parameters

**XDC Services Requirements:**
- **OneData:**
  - Onedata
    - Metadata management
    - QoS (policy definition)
User Stories

As an “Archive Manager”, I can configure the system to extract metadata during the preprocessing operation.

Related with :

As an “Archive Manager”, I can define rules based on metadata to archive files on tape only or on disc & on tape, at ingest time.
As an “Archive Manager”, I can define rules based on metadata to duplicate archive files on tape on low latency storage to be soon quickly retrieved.
As an “Archive Manager”, I can define rules based on metadata to create replicas on a specific data center, at ingest time.
As an “Archive Manager”, I can define rules based on metadata to prevent deletion of files
As an “Archive Manager”, I can delete files based on metadata values

As a “Archive User”, I can make a query based on metadata parameters to get a list of logical files names
OneData installation at LAPP and CCIN2P3
Conclusion

What has been demoed:
• Metadata index creation using REST API
• Deploy the « preprocessing’s service » using docker compose
• Ingest a HDF5 files
• Query File ID using HDF5 header
• Retrieve file using CDMI interface and File ID