DUNE ND File Storage and Archiving

Michael Kirby, Fermilab/Scientific Computing Division

Nov 6, 2019

What is meant by File Storage and Archiving?

- File is transferred to tape-backed dCache
  - means that the file will be given a location on dCache (e.g. /pnfs/dune/tape_backed/dunepro/mcc11/mc/)
  - file copied to an enstore tape volume for permanent storage (in general these are rarely deleted and it is a focused deliberate process)
  - file registered within a file catalog (metadata entered into a database)
- normally achieved using the Fermilab File Transfer Service (F-FTS)
  - utilizes a dropbox in /pnfs/dune/scratch
  - FTS scans multiple dropboxes frequently for new files
  - registers the files in file catalog
  - determines the dCache location from the file metadata
  - transfers file to dCache and deletes file from dropbox

🟢 **Fermilab**

What has to be done to utilize F-FTS?

- make sure that every file to be archived has a unique name
- Create JSON metadata file for every file place into the dropbox
- transfer file and JSON metadata file into the dropbox

- That's it.

🟦 Fermilab

How do you make a JSON metadata file?

- This is where it gets complicated….
- Two basic components
  - details of the file (name, file size, checksum, creation time, etc)
  - physics description of the contents of the file (mc/data, fd/nd, larsoft version, fhicl file, run_type, data_stream, etc)
- some of the metadata fields (not the values) needs to be defined ahead of time
- then some of the fields need to have values defined
- Bare minimum template is here: https://wiki.dunescience.org/wiki/Minimal_Metadata_template

🟦 **Fermilab**

Metadata minimum template

- file_name: is the file name - this needs to be unique within DUNE
- file_size: is the file size in bytes
- user: is the fnal unix user name of the person who is storing the file (need to generalize)
- content_status: can be used to flag bad files/data - not used unless it is REALLY bad
- file_type: is "detector" for data or "mc" for mc. "photon_detector" recently defined for np02
- file_format: describes the format, can be root, artroot, binary ...
- data_tier: is the stage in processing - raw, decoded, reconstructed, ... (should be lower case)
- application: is a 3-form with top level (art), specific application (reco, filter, sim...) and version
- event_count: number of events in the file
- data_stream: allows separation of different raw data types - map to different tape families for better access. test, cosmics, physics ....
- start_time: UTC time that file writing started (required for raw, optional for reconstruction)
- end_time: UTC time that file writing ended
- runs: list of 3-forms of [run, subrun, run_type] detector type tells 35T from far detector 1, from protodune sp in particular run_type is very important...
- parents: list of files that were processed to produce this file. Can be many->one

‡ Fermilab

More complete metadata for a file

- $> samweb get-metadata PDSPProd2_protoDUNE_sp_g4_p6GeV_35ms_sce_datadriven_23739153_66_99d5d2ce-58eb-4058-a412-ba724c6ca118.root
- File Name: PDSPProd2_protoDUNE_sp_g4_p6GeV_35ms_sce_datadriven_23739153_66_99d5d2ce-58eb-4058-a412-ba724c6ca118.root
- File Id: 35997352
- Create Date: 2019-09-10T20:02:00+00:00
- User: dunepro
- Update Date: 2019-09-10T20:20:57+00:00
- Update User: dunepro
- File Size: 1892720913
- Checksum: adler32:cda3401d
- Content Status: good
- File Type: mc
- File Format: artroot
- Group: dune
- Data Tier: simulated
- Application: art g4 v08_27_01
- Event Count: 10
- First Event: 661
- Last Event: 670
- Start Time: 2019-09-10T17:43:49+00:00

- End Time: 2019-09-10T18:10:03+00:00
- Data Stream: out1
- art.file_format_era: ART_2011a
- art.file_format_version: 13.0
- art.first_event: 661.0
- art.last_event: 670.0
- art.process_name: G4
- art.run_type: physics
- beam.momentum: 6.0
- beam.polarity: 1
- DUNE.campaign: PDSPProd2
- DUNE.production_status: test
- DUNE_MC.beam_energy: 6.0
- DUNE_MC.detector_type: protoDUNE SP
- DUNE_MC.electron_lifetime: 35ms
- DUNE_MC.liquid_flow: no
- DUNE_MC.space_charge: yes
- DUNE_MC.with_cosmics: 1
- MC.liquid_flow: no
- MC.space_charge: yes
- MC.with_cosmics: 1
- Runs: 23739153.0001 (physics)

🔷 **Fermilab**

Tasks

- Setup dCache dropbox
- update FTS to scan dropbox
- declare a file family for ND files
- setup metadata fields and decide upon standardization
- utilize metadata_extract.py to get "details of file"
- write script to expand to get physics information
- test F-FTS transfers

🎗️ Fermilab