

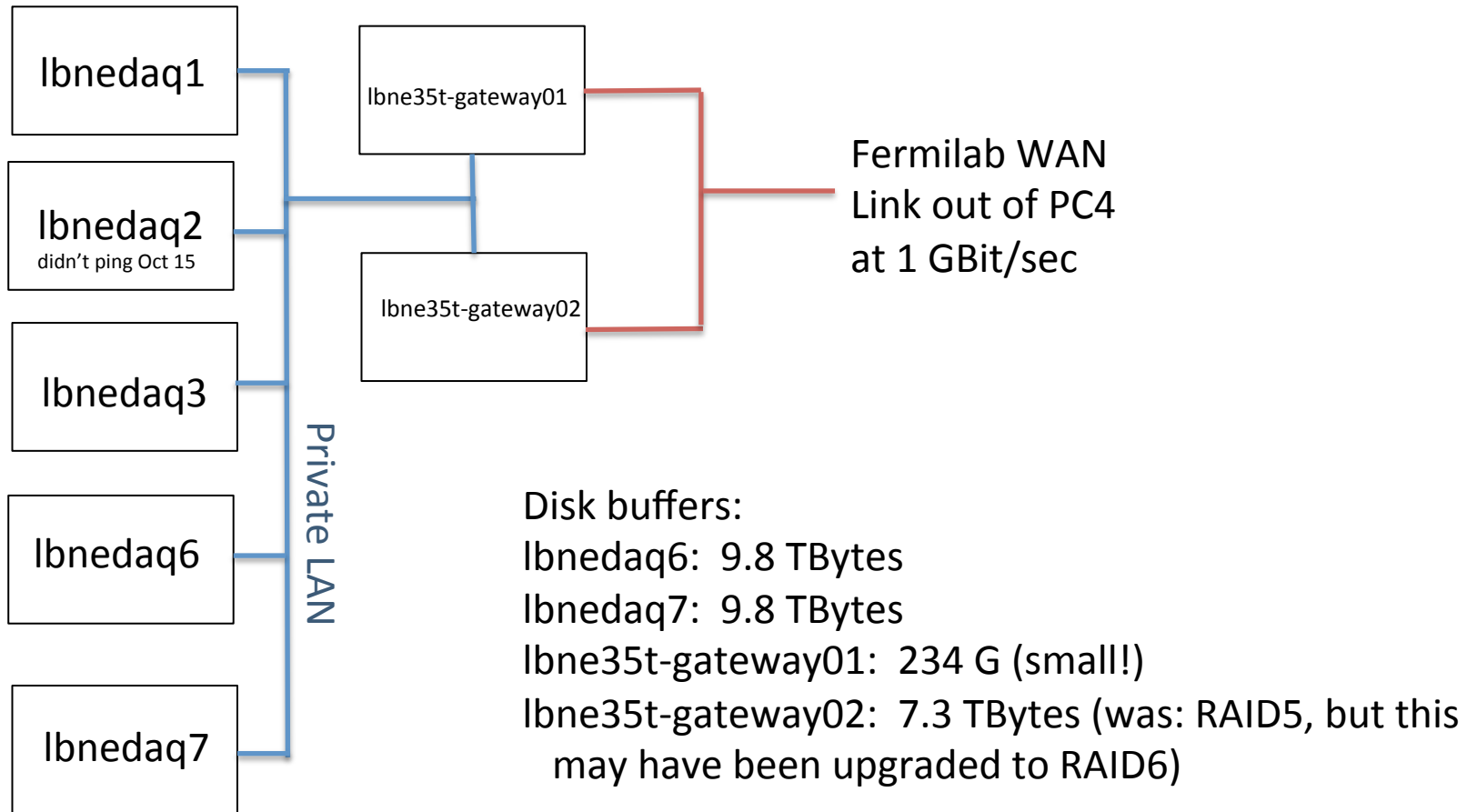
# 35-ton Data Handling

Tom Junk

35-ton Online/Offline Interface

Oct 16, 2015

## 35-ton Data Buffering and Transfer at PC4



## Data Transfer So Far

- Just a backup of test data files from lbne35t-gateway01 to enstore using SAM
- I've been copying them over to

`/pnfs/lbne/persistent/users/trj/`

in batches of one month's worth of files at a time.  
Files with names of this form:

`lbne_r003570_sr01_20151008T175941.root`

are copied (even if the run and subrun indexes are empty), but files with names of the form

`RootOutput-06a1-d032-a39e-fcf1.root`

have been omitted. Most of these are very small and look improperly closed.

## Data Transfer So Far

I followed Qizhong's instructions for extracting and uploading metadata, and placing files in the FTS dropboxes for storage in Enstore (Fermilab's tape system).

<http://indico.fnal.gov/getFile.py/access?contribId=0&resId=0&materialId=slides&confId=10130>

Qizhong's scripts now are in duneutil

I made wiki pages, batched by month, listing each file and the metadata used.

<https://cdcvs.fnal.gov/redmine/projects/35ton/wiki/LBNE35tVerticalSliceTestDataFiles>

Up to the end of September – October files not transferred yet.

## Example Metadata JSON file

```
{
  "file_name": "lbne_r003290_sr01_20150930T214134.root",
  "file_size": 31248466,
  "event_count": 1950,
  "last_event": 1950,
  "runs": [ [ 3290, "test" ] ],
  "first_event": 1,
  "file_type": "test-data",
  "file_format": "root",
  "data_tier": "raw",
  "group": "lbne",
  "application": {
    "family": "art",
    "name": "daqag",
    "version": "v00_00_01"
  },
  "start_time": "2015-02-05T20:19:59",
  "end_time": "2015-02-05T20:20:04",
  "lbne_data.name": "35t_testdata_201509",
  "lbne_data.detector_type": "35t",
  "lbne_data.run_mode": "immediate triggered mode"
}
```

Generated automatically using Qizhong's scripts, which need to run root to get some of these items.

Currently filled in by hand

Any field or combination thereof can be used to query SAM to get a URL to the file for opening in art.

## Setting up samweb and an example samweb query

sam\_web\_client is setup automatically when you setup dunetpc.

Or you can just

```
$ source /grid/fermiapp/products/dune/setup_dune.sh
$ setup sam_web_client
$ kx509
```

from a fresh login. The kx509 step gets a KCA certificate using your Kerberos ticket to validate your identity.

Getting the September 2015 list of files

```
samweb -e lbne list-files "lbne_data.name 35t_testdata_201509"
```

## More example samweb queries

[https://cdcvs.fnal.gov/redmine/projects/sam-main/wiki/Sam\\_web\\_client\\_Command\\_Reference](https://cdcvs.fnal.gov/redmine/projects/sam-main/wiki/Sam_web_client_Command_Reference)

```
$ samweb -e lbne get-metadata lbne_r003216_sr01_20150928T175438.root
```

n.b. filenames must be unique in an experiment's namespace. Presumably the metadata upload will fail if a file already exists with a particular name. We're still using lbne for now. I asked Qizhong if we needed a DUNE SAM station and for now the LBNE one may just have to do.

```
$ samweb -e lbne get-file-access-url lbne_r003216_sr01_20150928T175438.root
```

The string you get from this can be used as the input filename to art.

project.py has tools for making projects out of collections of files.

## Moving forwards – Better data handling and cataloging

- Local buffers on lbne35t-gateway02 and lbne35t-gateway01 can be used to store data before transfer
- Instead of scp or sftp, the /lbne/data, /lbne/data2, and /pnfs/lbne filesystems have been mounted on lbne35t-gateway02. Files can simply be copied.
- /grid/fermiapp is mounted on lbne35t-gateway02 so we can run the metadata extractor on the gateway node
  - needs root
  - needs sam\_web\_client
- May want FTS dropbox on lbne35t-gateway02, but we already have three dropboxes:
  - /lbne/data/dunepro/dropbox/data
  - /lbne/data2/dunepro/dropbox/data
  - /pnfs/lbne/scratch/lbnepro/dropbox/dataThe last of these was created so we wouldn't have to worry about quotas on BlueArc



## What we have, what we need

We have:

- cron script that gets a Kerberos ticket and transfers files with scp and checks the checksums
- Appropriate Kerberos keytab so that lbnetdaq tickets can be obtained by a cronjob.
- Tested procedure for uploading files to tape and SAM
- We have scripts from T2K and NOvA (Giles and Jarek)

What we need

- File transfer script from lbnetdaq6 or lbnetdaq7 to lbne35t-gateway02. What authentication is needed? Or is gateway2's disk just mounted.
- Check integrity of copy before file is deleted.
- Check for list of files waiting to transfer – bottlenecks, network outages.

## What we need

- Script to extract the last three metadata items from database

data.name

test, calibration, commissioning, physics  
detector configuration

35t:rce00:rce01:rce02:ssp00:ssp01 ...

run mode – immediate triggered mode, burst mode, etc.

Request has gone to Jon Paley for a db query mechanism.

The value of data.name has to be entered by the shift crew or run coordinator, and the others presumably exist in the data file.

Can run config\_dumper on the file to extract some information, but these also live in the database.

## What we need

- File catalog on the web

-- this can be automated out of samweb commands for files that are registered on tape.

-- tool to list files waiting to be transferred and web interface— so experts or operators can see if data are waiting behind bottlenecks.

## Comment on Data Analysis

- Some run modes – burst mode, scope mode – may not be appropriate to analyze with LArSoft. ADC dumper modules and standalone programs (pedestal and noise finding) may be most appropriate.