DUNE Software and Computing News and Labor Estimates

Tom Junk

Dune Software and Computing Bi-Weekly

November 17, 2015

Computing news from Fermilab

- Facility outages can be found here:
  http://cdorg.fnal.gov/fop/outages.html

For the past few months the most reliable outage information has been sent to the experiment liasons via e-mail and we redistribute them to the dune-computing-news@fnal.gov mailing list.

We had set up dune-computing-news to be a subscriber to the gp-downtime mailing list but all mails were rejected as spam by listserv, so Eileen and I ended up just forwarding them.

Some outages are short, or occur when we are in meetings or away, so this isn't perfect. The web page is useful but hasn't been updated since September.

Security Updates

- SHA1 Certificate support is going away on Jan 16, 2016
- Shared account guidelines updated for experiments wanting to use the OPOS service (Offline Production Operations Service)
- That's us! We need to allow Fermilab SCD people to be able to use the dunepro account. We've asked for help in shepherding MC jobs. 35t data are coming.
- We've met with OPOS but MCC 5.0 is the first test of their services, and is coming in the next week(s)
- The maximum of 10 people in .k5login does not include OPOS people or support from SCD.
- Maximum simultaneous connections are not going to be enforced.
- But after 15 minutes of inactivity, SSH sessions will be logged off/terminated

New GPGrid going into production

- Would like to get rid of quotas
- New GPGrid will have fair share and priorities
- What is the maximum job length? User should be able to put in an estimate of job run time (clock time, not CPU) for purposes of scheduling.
  - some remote sites have short slots
  - Helps in draining partitionable slots.
  - GPGrid team is investigating returning logs for jobs that hit the time limit
  - Users should be able to specifiy maximum job length in the next release of jobsub. Would like simple mnemonics – short, medium, and long.
  - Maximum requestable – GPGrid would like to limit it to 2 days
- Default resources are 1 CPU and 2 GB of memory

New GPGrid going into production

- Grid managers would like to get rid of quotas and just use the Condor fairshare algorithm.
- Lots of flexibility in using fairshare – priority-based job scheduling.
- Would get rid of OPPORTUNISTIC and DEDICATED

- Partitionabile slots – some configured for more memory and/or more cores if need be. Dark Energy Survey wanted very large memory.

Data File Transfer from 35t → Tape

- Data are written to lbnedaq6.fnal.gov:/storage/data by *artdaq*.
- lbnedaq6.fnal.gov is on a private network – not visible to the Fermilab campus network. Must connect via lbne35t-gateway01.fnal.gov or lbne35t-gateway02.fnal.gov
- gateway01 is to be used for Run Control. Owners of the machine do not want "offline" software running on it (not even a debugger!)
- gateway02 is intended for use in data transfer. Has 7.3 TBytes of RAID (6 I believe..) disk.
- lbnedaq6 and lbnedaq7 have 9.8 T of disk space each.
- Data write rate is about 50 Mbytes/sec to the lbnedaq6 disk from the *artdaq*.

Data transfer for 35t

- Several options for data transfer:
  - Copy files from lbnedaq6(7) directly onto gateway02's disk (they're mounted); copy files from gateway02 to /pnfs/lbne/scratch/lbnepro/dropbox/data where they get transferred to tape. To have one script in control, it should run on gatweway02 and use ssh to initiate the first copy. Checksums!
  - On lbne35t-gateway02, run a scp process with the target in dCache. No local storage on gateway02's disks, so saves a disk write and read. (bare disk writes work at 80 Mbytes/sec). But scp takes CPU. Got maybe 60 Mbytes/sec doing this as a test. Can run several simultaneously – gateway02 has 16 cores. May want to compute checksums on both ends of this transfer.

Data Transfer for 35t

- Another transfer option

Compress files on gateway02 – copy first as in the first option, but run *art* with compression turned on (*artdaq* has it turned off). With data files taken so far, it's good for a 46% reduction in file size at ROOT compression level 1. On my desktop, a 9.6 GByte file took 5 min, 30 sec. time and 4 min. of CPU to do this. Higher compression levels hardly improve the compression factor but take a lot more CPU.

Can run many of these in parallel. Input file is on gateway02's disk, output file is in the dCache dropbox, to save another disk write. May lose ability to check the checksum on both sides.

Will need to extract metadata after the transfer to dCache dropbox as the file size changes.

Current file transfer 35t

- Current 35t file transfer script:
  - Looks for data files on lbnedaq6. Skips runs with no run and subrun number, skips RootOutput*.files which haven't been closed yet or not yet properly closed.
  - Computes checksum on lbnedaq6
  - scp's the file to gateway02
  - computes the checksum there. If failure or mismatch of checksums, try again up to 10 times
  - Extract metadata – run number, # events (uses root to open file), file size. Gets start and end times, detector configuration, run mode, and name out of online database (Thanks Jon and Erik!)
  - Uploads metadata to SAM
  - Copies files to dropbox in dCache. (no checksum check yet. This step is a bit risky as dCache goes down a lot – while the checksum is not too likely to fail, the entire copy probably will. Does the copy hang or just quit with an error code? Not sure..)
  - Does not delete files – just mv's them to a transferred_files directory, along with JSON metadata files.
  - Keeps a list of transferred files in a text file. To do: automate a web page out of this.

35t File transfer

- run mode, confuguration name, and the text string for the data name (Currently “Test”) are gotten from the online database, queried on run number.

- Some of these queries fail, and metadata cannot be extracted for them. To investigate, but many runs at the moment are short and serve their purpose online, to see if things work.

- SAM station has an expired certificate (expired on Nov. 14, just as I was uploading files). Result of OSG Grid certificates being distributed without a complete DN – missing OU=Services field. – Files not being uploaded to Enstore. SCD (Mike Diesburg) says they will get a new certificate installed ASAP.

- We have about 2.5 TBytes of test data in the dropbox waiting to be put on tape

- Completed files not showing up in “completed” directory

35t File transfer News

- Monitoring of FTS server

http://lbnesamgpvm01.fnal.gov:8787/fts/status

Linked on

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

and

https://cdcvs.fnal.gov/redmine/projects/dune/wiki/Computing_How-To_Documentation

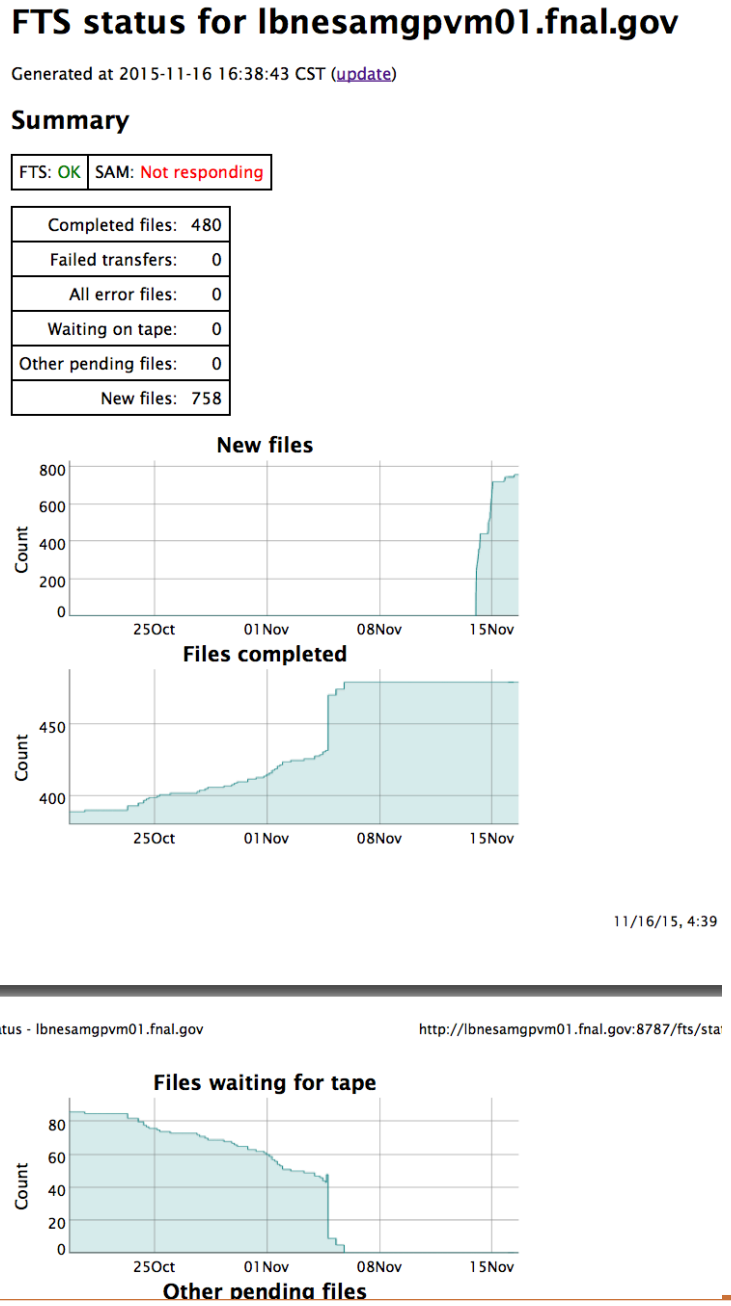
FTS and SAM Monitoring

http://lbnesamgpvm01.fnal.gov:8787/fts/status

These graphs (and more) will be available on FIFEMON

http://fifemon.fnal.gov

when the FTS version on lbnesamgpvm01 is upgraded to v 5.0.0



Metadata Issue

lbne.data.detector_type appears to get truncated to the value “35t” when the file is uploaded to SAM.

Wanted it to be of the form

35t:rce00:rce01:rce02:ssp00

This is how the metadata is uploaded initially to SAM and the way SAM returns it on queries before the file makes it into Enstore.

Asking for this to be fixed. Might have to put the detector configuration string in the name or the run mode. Would like to query on the active detector components.

Software Versions Coming

- A reminder – GEANT 4.10.0 and ROOT 6 are coming to LArSoft!
- At the *art* stakeholders' meeting and the LArSoft coordination meeting, DUNE has expressed neither reservations for moving to the new versions, nor is eager to make the schedule move faster.
- We are interested in keeping up with the latest versions at this point, and for the foreseeable future.
- ROOT6 takes more memory, GEANT 4.10.0
- At some point we would like to freeze G4 version (root should be "lossless"). A coordination issue: What if some LArSoft stakeholders insist on a newer version of G4, while others have frozen for physics reasons.

Software and Computing Effort Estimates for LBNC

S&C and Computing Model Labor Estimates Through 2019 (First pass guess!)

Group	Task	Prio	Start Date	Dur	FTE
Computing Model	Document the DUNE Computing Strategy	H	Q1 2016	48	0.3
S&C	Data Transfer and Cataloging, protoDUNE	H	Q1 2017	36	0.5
S&C	Liaison to Fermilab SCD, incl. annual resource request	H	Q1 2016	48	0.5
S&C	Grid VO Admin	H	Q1 2016	48	0.2
S&C	Interactive Accounts and Disk Admin	H	Q1 2016	48	0.1
S&C	Database Development	H	Q1 2016	48	0.3
S&C	Liaison to LArSoft	H	Q1 2016	48	0.3
S&C	Liaison to the art team	H	Q1 2016	48	0.1
S&C	Software support	H	Q1 2016	48	0.2
S&C	Collaborative Tools	H	Q1 2016	48	0.5
S&C	DAQ Interface	H	Q1 2016	48	0.5
S&C	Investigating Leadership Computing Facilities	M	Q1 2016	48	0.1

	2016				2017				2018				2019			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Computing/Software WG *Resp: Junk, Farbin*																
Software and Computing group	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3
Computing model	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.1					0.2	0.2	0.2	0.2

ProtoDUNE S&C Labor Estimates

Computing – the sim/reco is separate – T. Kutter proposed of order 22 FTE's needed for sim/reco at the Sep. Collab meeting Estimates from Maxim

Group	Task	Prio	Start Date	Dur	FTE
protoDUNE S&C	DAQ Emulator to enable buffer integration	H	Q1 2016	2	1
protoDUNE S&C	Prototyping of the buffer farm ("subfarm")	H	Q1 2016	3	2
protoDUNE S&C	EOS interface and "consensus buffer flush"	H	Q2 2016	2	3
protoDUNE S&C	EOS-CASTOR interface (bidirectional)	H	Q2 2016	3	1
protoDUNE S&C	Prototyping express-streams to run at CERN, out of EOS	H	Q2 2016	2	2
protoDUNE S&C	Provisioning/installing DUNE software at CERN	H	Q3 2016	2	2
protoDUNE S&C	Mock data challenge (simulated express streams at CERN)	H	Q3 2016	2	1
protoDUNE S&C	Tech. evaluation and downselect for CERN-US data link	H	Q3 2016	2	1
protoDUNE S&C	Metadata generation and SAM interface	H	Q4 2016	3	1
protoDUNE S&C	Procurement, configuration and testing of bandwidth out c	H	Q4 2016	2	1
protoDUNE S&C	Scalability test of data transport CERN-US	H	Q1 2017	1	1
protoDUNE S&C	Data flow monitoring system	H	Q1 2017	2	2
protoDUNE S&C	Dress rehearsal of complete chain of data transmission (inc	H	Q2 2017	1	3
protoDUNE S&C	Installation and configuration of the HW buffer (subfarm)	H	Q2 2017	3	2
protoDUNE S&C	protoDUNE production system	H	Q3 2017	4	3
protoDUNE S&C	Integration with actual DAQ	H	Q3 2017	3	2
protoDUNE S&C	Data challenge with actual components in place (DAQ, subf	H	Q3 2017	4	3
protoDUNE S&C	XRootD data storage federation	H	Q4 2017	3	1
protoDUNE S&C	Commissioning	H	Q4 2017	3	3

		2016								2017								2018								2019							
		Q1		Q2		Q3		Q4		Q1		Q2		Q3		Q4		Q1		Q2		Q3		Q4		Q1		Q2		Q3		Q4	
		S	E	S	E	S	E	S	E	S	E	S	E	S	E	S	E	S	E	S	E	S	E	S	E	S	E	S	E	S	E	S	E
CERN prototypes WG	*Resp: Kutter,tbd*																														
protoDUNE coordination with WA105 computing		3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2

35t and FD reco

Group	Task	Prio	Start Date	Dur	FTE
35t sim/reco	TPC Paper S/N, charge yield purity, gaps, p, pizero	H	Q1 2016	18	2
35t sim/reco	PD performance, light yield	H	Q1 2016	18	1
FD Reco	nu_e CC identification	H	Q4 2015	6	1
FD Reco	track/shower reconstruction	H	Q4 2015	6	1
FD Reco	Various PID	M	Q2 2016	12	3
FD Reco	Validation of reconstruction	M	Q4 2015	12	1
FD Reco	Improvement upon existing reconstruction chain	M	Q1 2017	24	6
FD Reco	Wire Cell Pattern recognition	H	Q4 2015	6	0.5
FD Reco	Wire Cell Larsoft Integration	M	Q1 2016	6	0.3
FD Reco	Web-based Human Directed Pattern Validation	H	Q1 2016	12	0.5
FD Reco	FD Design Validation	H	Q4 2015	12	0.5

May need to put WA105 and dual-phase in a separate category

	2016				2017				2018				2019			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Computing/Software WG *Resp: Junk, Farbin*																
Software and Computing group	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3
Computing model	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.1					0.2	0.2	0.2	0.2
FD sim/reco	8	8	8	8	9	9	9	9	10	10	10	10	10	10	10	10
35T reco/analysis	3	3	3	3	3	2	1									
Beam simulation																
protoDUNE Sim/Reco/Ana																
Photon Det Sim/Reco	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Alex Himmel's task list for photon detector sim/reco is quite detailed. 2 FTE's

Analysis Framework HighLAND

- From Anselmo Villanueva (Valencia)
- A layer on top of Analysis Trees. Used on T2K ND work.
- To invite. Probably stakeholders are in the physics and sim/reco groups.
- We can supply support for versioning, code repositories, wikis, etc.