

DUNE Computing Overview

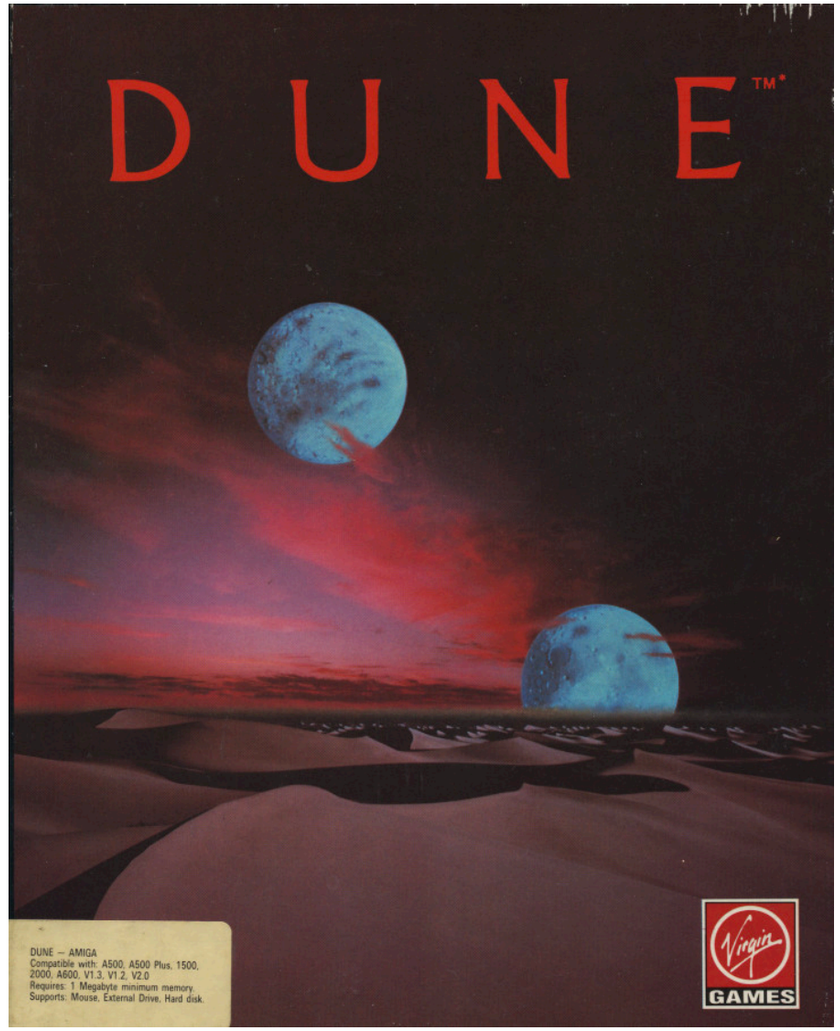
Michael Kirby, Fermilab/Scientific Computing Division

May 20, 2019

Outline for DUNE Computing

- DUNE Computing Status and Activities
  - Continued globalization of DUNE Computing
  - Production Processing
  - Data Management
  - Software and Tools Management
  - Database development
- Plans for the future
  - Dual Phase Data Challenge
  - Results of Analysis User Survey -
    Grid Utilization by analysis groups and survey results
  - Data Model Workshop
  - Computing Model Workshop
  - Utilizing HPC platforms for both analyzers
    and production

Earliest version of DUNE Computing



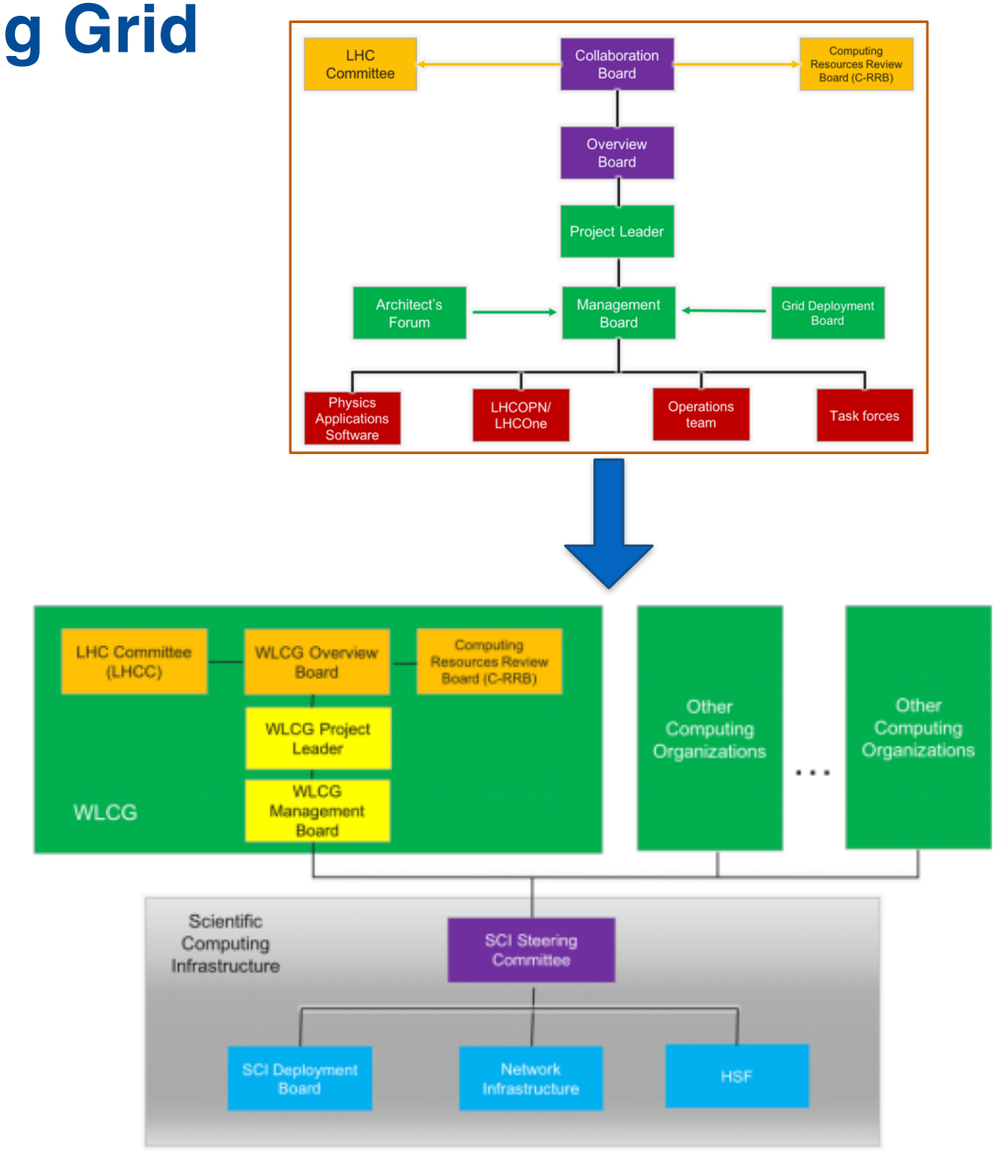
*no current plans to support the Amiga platform

Fermilab DUNE

Coordination with Worldwide LHC Computing Grid

- WLCG undergoing a transition in the organizational structure towards a Scientific Computing Infrastructure Steering Committee
- Approved middleware and interfaces (CREAM, ARC, HTCondorCE, dCache,EOS, SRM, etc)
- Tickets (GGUS), registry (GOCDB), and accounting (APEL) from EGI

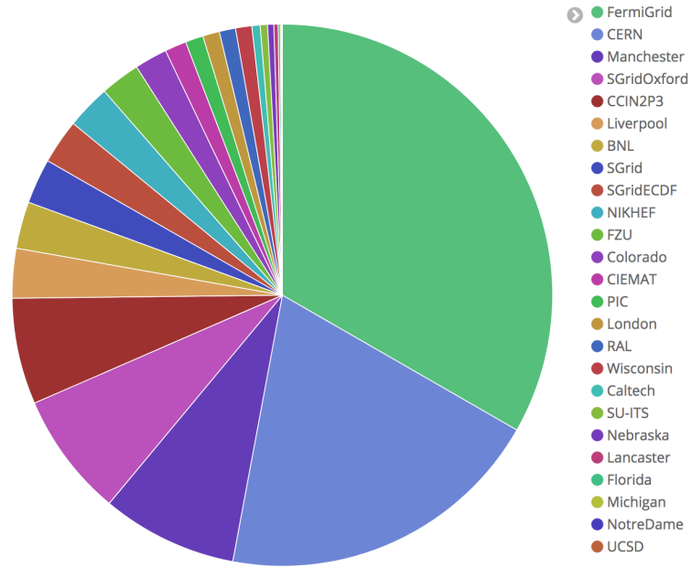
- DUNE is now an **observing member** of the WLCG Management Board - announced at the HSF/OSG/WLCG Workshop at JLab this spring
- will participate in Global Deployment Board (GBD) for infrastructure planning
- DUNE has already started to participate at an operational level through GGUS tickets - report and aid in identification of site specific issues (UK and French sites currently)
- Hoping to have a GDB meeting in the US later this year with focus on DUNE-relevant areas
- **Technical contributions** made with development and evaluation of services (not focused on getting everything into production operations)
  - Edinburgh + FNAL: development work on RUCIO for Data Management
  - RAL-PPD + Manchester: exploring DIRAC capabilities for Workload Management (FNAL kx509 certs and file catalog filed w/ metadata from SAM)
  - RAL-PPD: Running small MC simulation on WLCG and RAL SE integration



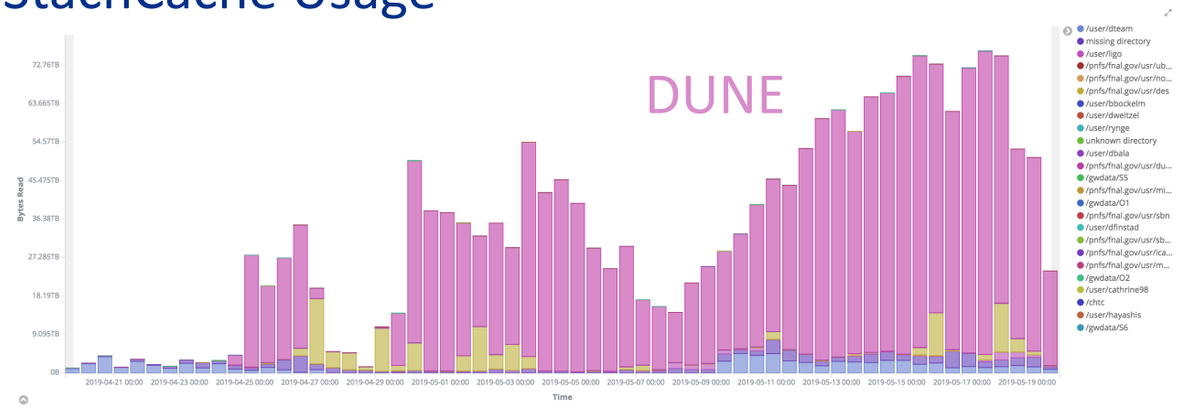
DUNE Production current status - Ken Herner, Ivan Furic

- Production group is tasked with executing the planned processing of data and MC samples as determined by the collaboration
- Campaigns completed since Jan Collab mtg:
  - protodune single phase 1 GeV MC
  - Partially completed single particle request
  - **7.5 ms readout-run reconstruction**
  - Ongoing: process cosmic runs, HV scans, readout time tests for SP data
- Modification in processing:
  - Reducing the number of events/job in DualPhase Sim
  - splitting up processing the 7.5 ms run into 20-event chunks
- DUNE is the **largest user of StashCache**
- Currently submission and tracking is based at FNAL using POMS/GlideInWMS/HTCondor - additional options being explored

Site usage
last 4 months
3.43 M Hours
1.98 M Sim
1.45 M data

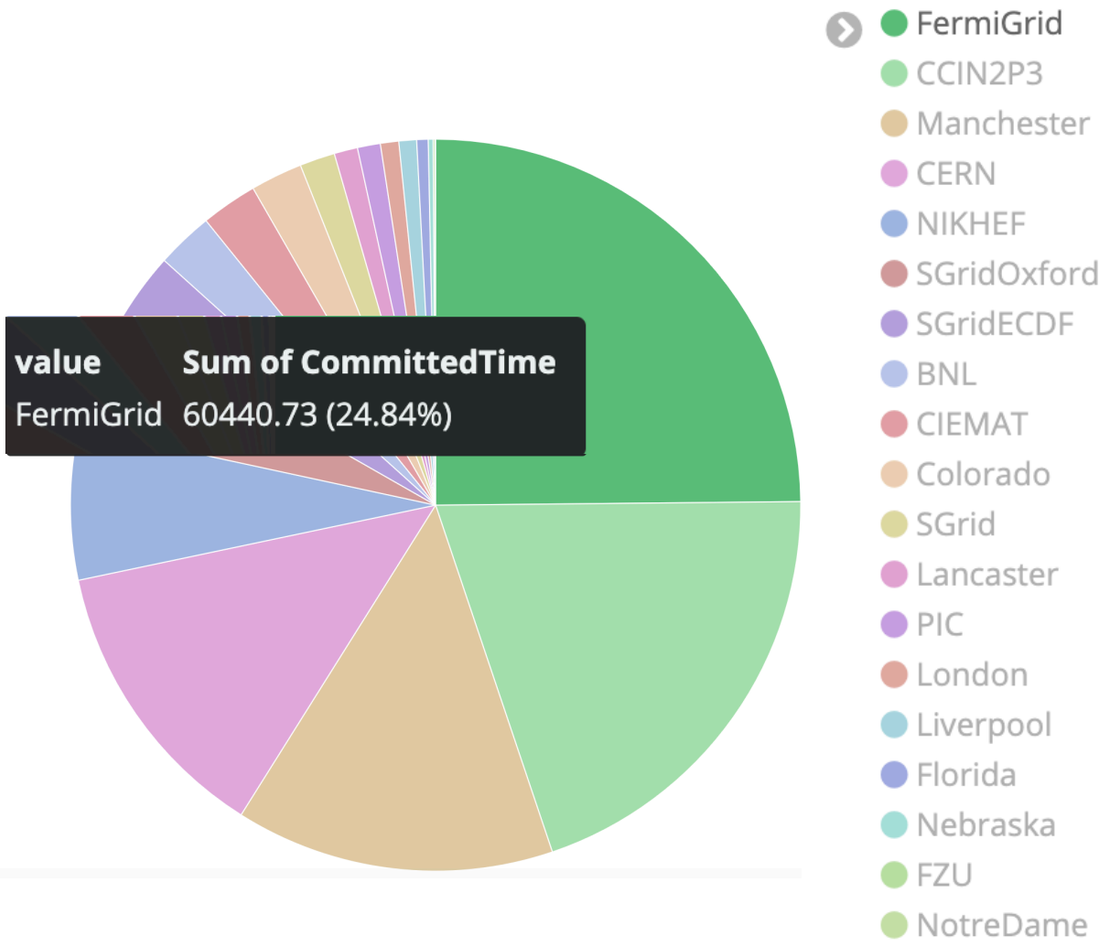


StachCache Usage



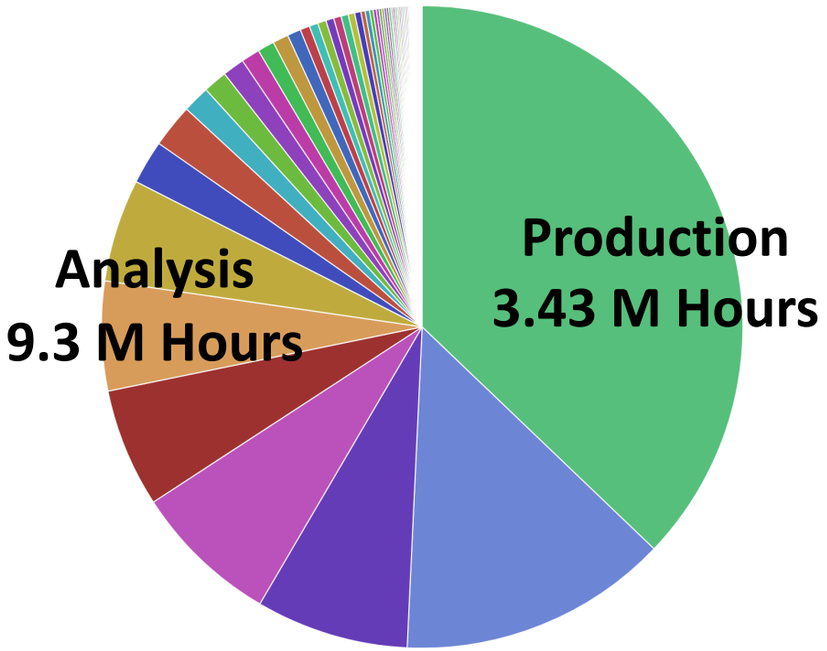
DUNE Production achieved an important milestone

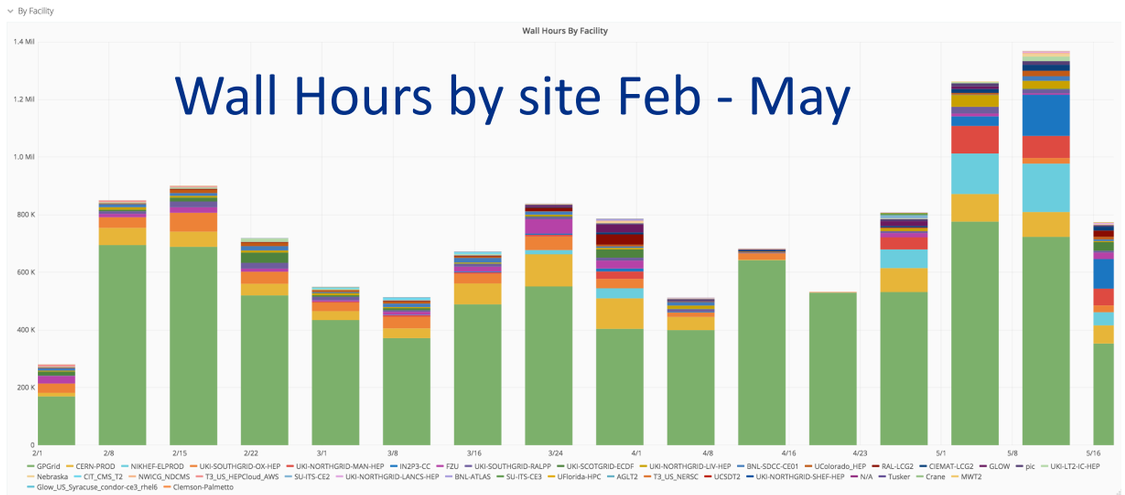
- End of last week with processing of SP and DP Simulation, 75% of DUNE Production grid jobs run on resources outside of FNAL
- Thanks to IN2P3, GridPP (Manchester, Oxford, Edinburgh, Lancaster, etc), CERN, NIKHEF, CIEMAT+PIC, Colorado, Florida, Nebraska, Prague, Notre Dame, and many others



DUNE Production Upcoming Goals

- Running completely in SLF7
- Upcoming Campaigns:
  - Finish Dual Phase Simulation production
  - "Mini production" to reprocess some ProtoDUNE Single Phase data with update dunetpc release
  - Continue cosmic processing
- ProtoDUNE Dual Phase data challenge and first cosmics processing from NP02 data
- Production group is looking for additional effort with shifters and testing of new features - great opportunity to help get your site involved and integrated

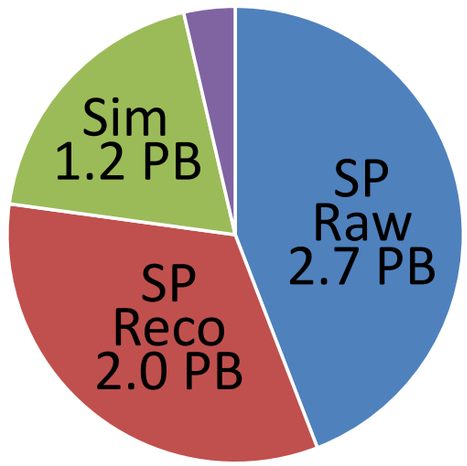




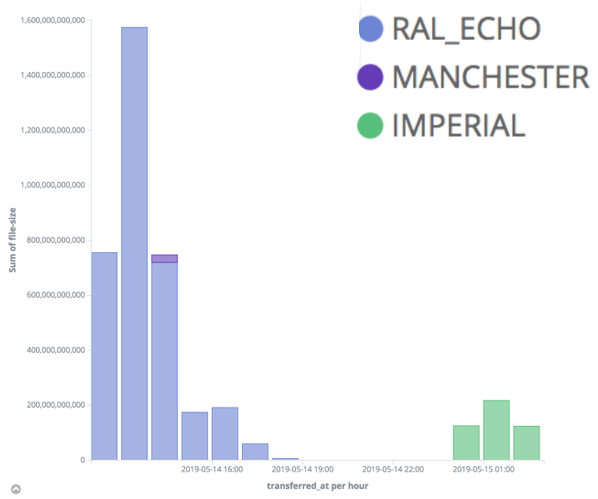
Data Management and Data Volumes - Steve Timm, Adam Aurisano

- Using Sequential Access via Metadata (SAM) to catalog and access files
- **Rucio instance** used to distribute files to secondary locations
  - discussion of **data retention policy in the Wed session - 2:00 pm**
- Three main storage facilities:
  - FNAL dCache/ENSTORE - raw data, reconstructed data, simulation, analysis ntuples
  - CERN CASTOR - raw data and reconstruction
  - CERN Public EOS - raw data and reconstruction
- Expanding soon to include IN2P3 and RAL
  - allocation for processing and tape storage of ProtoDUNE Dual Phase reconstructed data at IN2P3
  - **1PB disk and 500 TB** tape allocation at RAL for storage of data
    - place simulation reco files there
    - likely use as testbed for Rucio and DIRAC
- Additional disk only sites (Edinburgh, Imperial, Liverpool, Manchester, Prague, QMUL, RAL)
- Development on Rucio to allow for rules and policies that will meet DUNE requirements

DUNE Tape Usage To Date



Since Jan 1, 2019	Vol PB
Raw Data	0.910
Reco Data	0.478
Simulation	0.575



Software Management - David Adams, Christoph Alt, Tom Junk

- Weekly releases of dunetpc and duneutil, keeping up with *art* and larsoft.
  16 releases since late January
- New releases support
  - SL6
  - SL7
  - macOS 10.13 (High Sierra)
  - macOS 10.14 (Mojave).

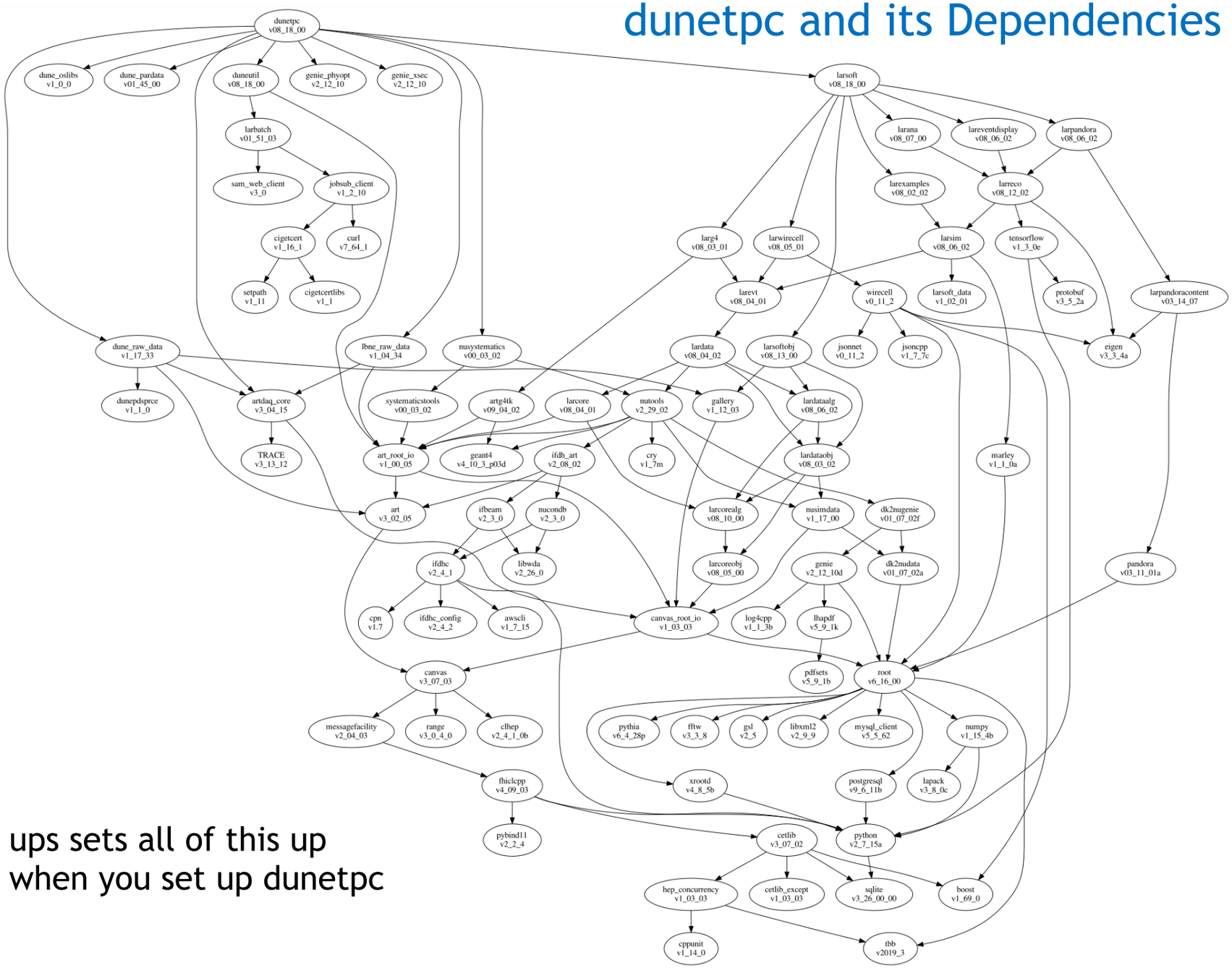
Debug and Prof builds supported

gcc and clang supported (macOS only has clang)

cvmfs and scisoft used to distribute built code

- macOS 10.12 (Sierra) is not supported for new releases.
- Offline releases of dune_raw_data and lbne_raw_data made for each *art* release used by LArSoft. dune_raw_data has a separate online git branch
- Old releases are have been cleaned out of CVMFS to minimize usage - thanks Gavin Davies

dunetpc and its Dependencies



ups sets all of this up
when you set up dunetpc

Highlights of work Since Jan 2019

- dunetpc v08.x builds with *art* V3, which introduces multithreading capabilities.
  - Event-level threading - many threads per job, still one thread per event
  - may not drop memory footprint enough
  - **multiple threads for a single event** is a request for development
  - still need to **reduce memory usage** and try using the *art* thread tools.
- work on improving the debugging tools - forge_tools v19_0_3 installed
- Clean up old releases of dunetpc that did not set up anymore - ease CVMFS maintenance and performance
- dune_oslibs library shims: SL7 flavors simplified.
- Photon visibility libraries now distributed via StashCache: /cvmfs/dune.osgstorage.org

Upcoming Software Management Projects

These projects depend on how *art* and LArSoft evolve…

- Speed up builds and developer experience
- Remove old, dead code in dunetpc Split dunetpc into multiple git repos
- CMS is able to check out just part of a git repository
  - potential migration to Spack and SpackDev which uses new approach
- Move to a pull-request model
  - LArSoft proposes to move in this direction
  - Uses **GitHub** instead of Redmine
  - Needs reviewers and coordinators to handle pull requests.
  - **DUNE may need to contribute reviewers to LArSoft** (even if dunetpc isn't moved to pull-request)

Database Progress - Norm Buchanan, Jonathan Paley

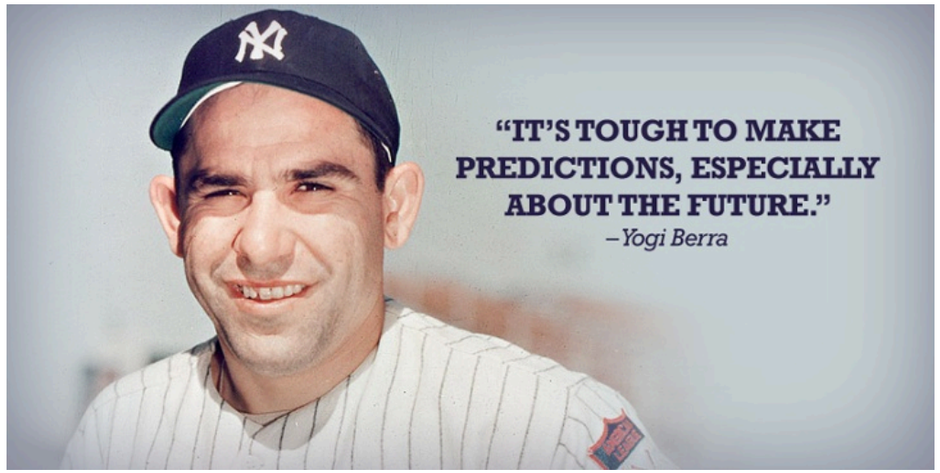
- Successful replication of the Beam Instrumentation Database from CERN to FNAL and into the IFBeam structure
  - Allows access to information needed for offline processing
  - **API/interface completed** and ready for incorporation into software modules
- Hardware Database
  - photon detector schema is complete and population has begun
  - Anode Plane Assembly schema is complete and has a few entries (Flor Blaszczyk)
  - work still needed to **integrate between the hardware database and other infrastructure**
- ProtoDUNE Single Phase DAQ configuration is stored at CERN in MongoDB -> successful replication into a UConDB at FNAL
- RunDB information has now been integrated into the SAM file queries so that datasets built based upon RunDB information are now possible
- Work is needed to improve upon integration of operations databases into production - **meeting planned for tomorrow**

Plans for the future

- Plans for the future
  - Federated Identity
  - Dual Phase Data Challenge
  - Results of Analysis User Survey - Grid Utilization by analysis groups and survey results
  - Data Model Workshop
  - Computing Model Workshop
  - Utilizing HPC platforms for both analyzers and production

Plans for the future

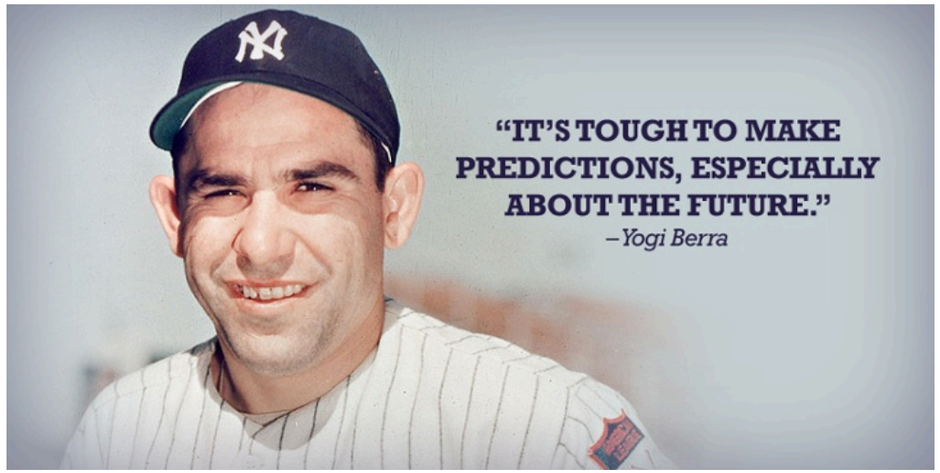
- Plans for the future
  - Federated Identity
  - Dual Phase Data Challenge
  - Results of Analysis User Survey - Grid Utilization by analysis groups and survey results
  - Data Model Workshop
  - Computing Model Workshop
  - Utilizing HPC platforms for both analyzers and production

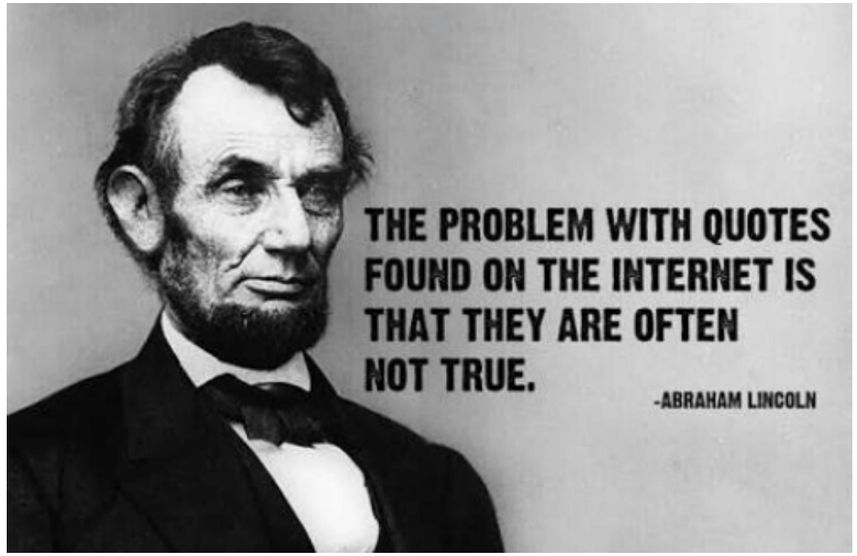


Fermilab DUNE

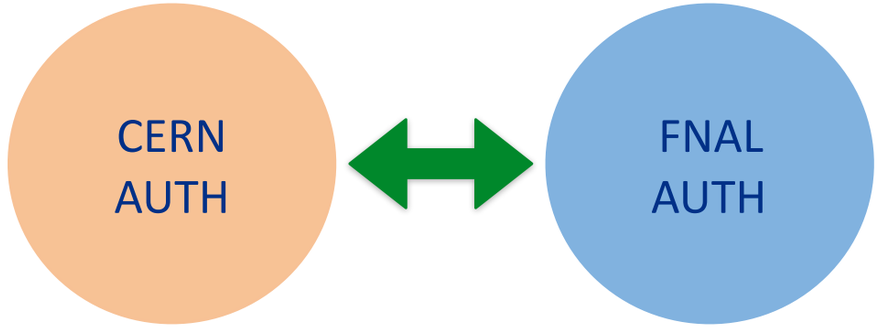
Plans for the future

- Plans for the future
  - Federated Identity
  - Dual Phase Data Challenge
  - Results of Analysis User Survey - Grid Utilization by analysis groups and survey results
  - Data Model Workshop
  - Computing Model Workshop
  - Utilizing HPC platforms for both analyzers and production





Federated Identity Management



Collaborative Services

- FNAL and CERN have been working on Federated Identity support whereby membership in one organization grants access to other organization(s)
- CERN has implemented FNAL SSO for authentication on CERN indico and working on edms
- FNAL last week started the process of having DocDB as the first use case
- Goal is to have many “browser-based” services utilizing federated identities
  - Team Centers - engineering docs, etc
  - Service Now - support tickets and report problems to FNAL
  - Redmine - DUNE/LArSoft/FIFE wiki pages and tutorials
  - Indico - easier use of indico with site specific to meeting
  - other requested services?

Grid Services

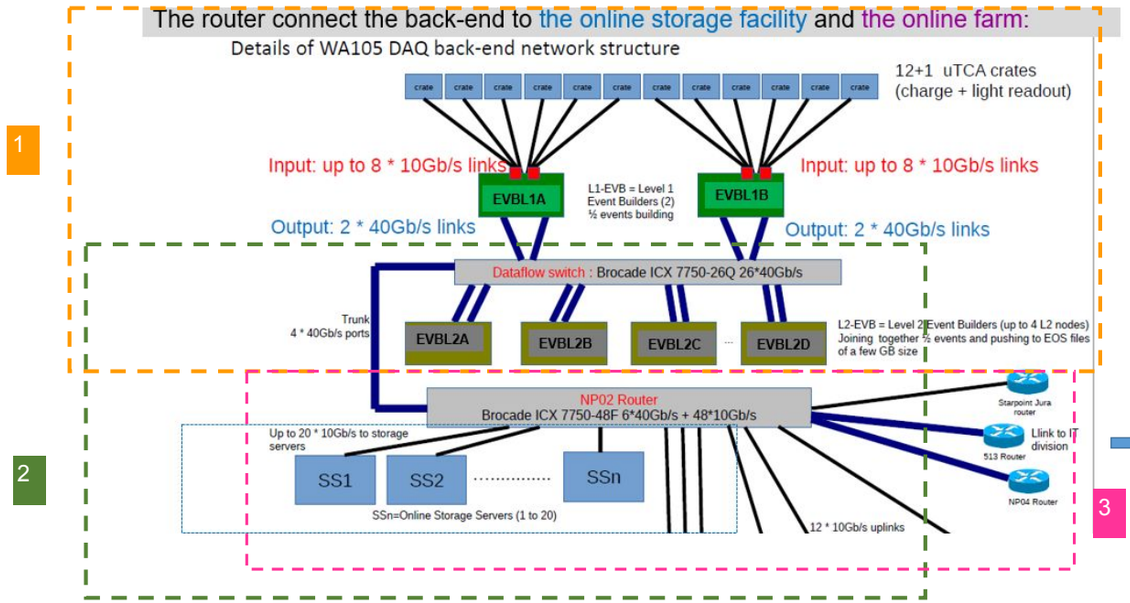
- DUNE is not an WLCG member, but will benefit from work moving to federated tokens (no more certs)
- allow access to site resources, storage elements, data management, etc
- currently 3 token solutions are being examined by the community
- FNAL Security team is actively participating with WLCG and has offered to represent DUNE with the authentication working group
  - will interface with DUNE security rep to consult with and discuss the needs of DUNE going forward
- Interactive computing access has not been decided

REMINDER: Multi-Factor Authentication at FNAL will soon require an RSA token for using the VPN - https://computing.fnal.gov/mfa

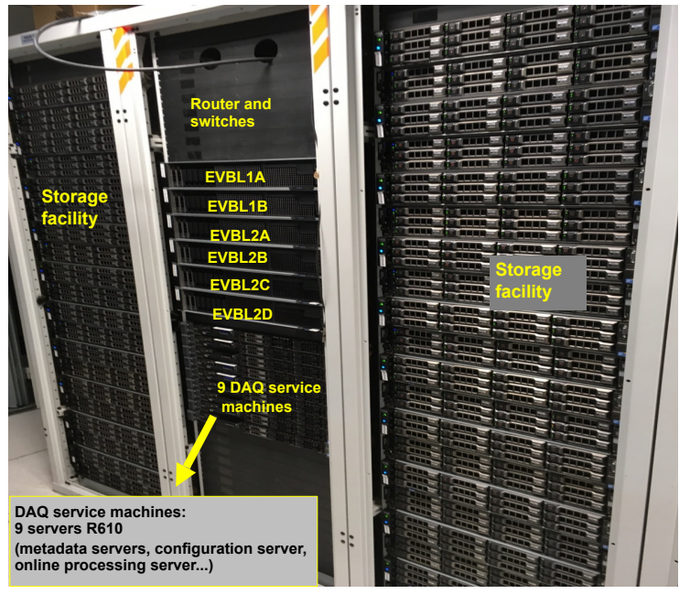
Fermilab DUNE

ProtoDUNE Dual Phase Data Challenge 3 plans - Elisabetta Pennacchio

- The first (extremely useful) joint data challenge (DC2) was performed in April 2018.
  - The DAQ back-end system was already operational
  - continuing to perform “data challenge” exercises for any modification or addition of new elements
  - currently operational workflows are not anticipated to change with DC3
  - general verification services provided by CERN IT division (link to IT and central CERN EOS) and for the data transfer to FNAL



EHN1 DAQ room



4 main data-flow blocks in the NP02 DAQ system:

1. From the uTCA to the L1 Event builders and L2 Event builders (from bits to datafiles….)
2. From the L2 event builders to NP02 EOS instance
3. From NP02 EOS (1 PB) instance to central CERN EOS
4. From the central CERN EOS to CASTOR and Fermilab

ProtoDUNE Dual Phase Data Challenge 3 Scope

1. JSON metafile generation - repetition of what has been done for DC2, structure of the metadata files has slightly changed. Finalization expected shortly, and some new files have already been generated

2. File transfer from CERN EOS to CASTOR - not tested last year, no major surprises/difficulties are expected, however should perform a complete test with heavy data volume conditions

3. File transfer from CERN EOS to FNAL - already tested last year, ensure service continues at same level

4. At Dune Core Computing Meeting (May 13th) new proposal to include NP02 raw data processing to the scope of the data challenge. This is doable, and it will be done in collaboration with DUNE data management and production team.

Raw data processing drives the organization of the data challenge.



Fermilab
DUNE

ProtoDUNE Dual Phase DC3 planning

- protoDUNE DP cosmic MC data (MC11) will be used as input data
  Data files have to be moved to CERN EOS *before* the data challenge .
The total volume has to be decided.

- Operations during DC:

1. data files are moved to CASTOR (new) and FNAL (already tested in 2018)

2. Data will then be processed at CCIN2P3 - 2019 30M HS06 CPU hours are available, and 2PB of tape space. MC jobs at CCIN2P3 write output to local SE but not tested at scale

- To organize this “raw data processing” in the best conditions, some discussions are now going on with Ken

Data processing in CCIN2P3 with output writing in the local SE is one of the main goal of the DC3, and it will allow the complete integration of CCIN2P3 in DUNE computing

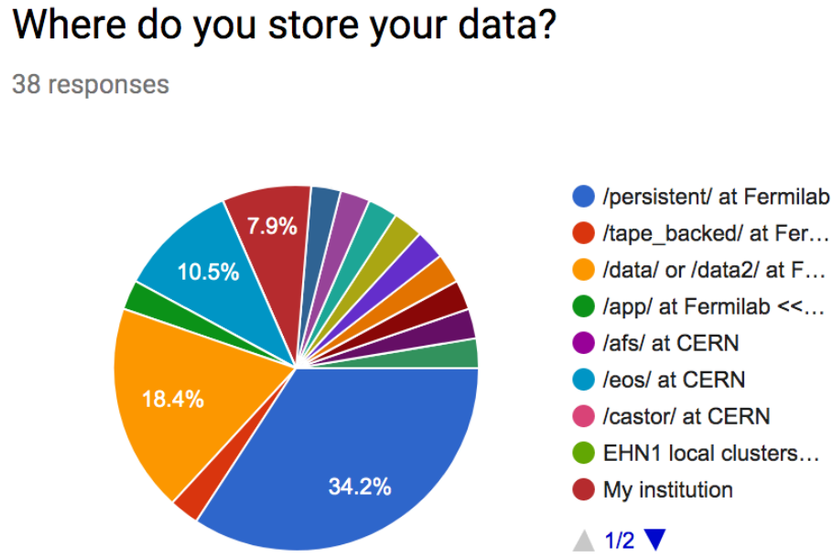
- Thanks to Steve and Ken, for all the discussion we had and we will have in the next weeks….



Fermilab
DUNE

User Analysis Survey

- Need to improve the user analysis experience
- Lots to learn about computing, and limited effort to teach and document all of it
- https://forms.gle/7TW9KbwPJdEfaHmp7
- 39 responses - 30 institutions
- large distribution of storage elements used - **developed a storage element requirement doc**
- job submission to sites
  - 34 users - Fermigrid & OSG
  - 10 users - CERN submission
  - 9 users - local institution clusters
- 30 users find data via SAM
- 75% coverage using xrootd



- support for **SciPy, NumPy, TensorFlow,** etc (wider data science tools)
- Beginners guide to computing & LArSoft
- Dummy Analysis
- **Computing Users Meeting scheduled for Tuesday at 4:00 pm**

Upcoming workshops for Computing Consortium

- Data Model Workshop Aug 14-16 tentatively at BNL
  - important to understand the data structure and logical flow from DAQ to analysis
  - Data Model Working group put together draft document
  - input from DAQ, Calibrations, physics groups need to understand requirements and latency for processing
- Computing Model Workshop Sep 9-11 tentatively at FNAL
  - trying to coordinate with the WLCG Grid Deployment Board to have focus on DUNE needs
  - computing model is important in order to identify projects and infrastructure that needs effort
  - understand how contributions from labs and institutions will be accounted
  - plan to complete computing model 2019, build out during 2020, measure success and identify weaknesses during the ProtoDUNE beam operations of 2021

Moving towards High Performance Computing

- Department of Energy is focusing resources on building out Leadership Class Facilities
- expectation is that these facilities will supply significant portion of the computing resources needed for science to be done
- New architectures - CPU, Accelerators (GPU FPGA, coprocessors, etc), interconnected memory
- Adapting algorithms and workflows to match these resources and developing new techniques
- proposed work to develop data prep code (calibration, 2D deconvolution, ROI & Hit finding in multithreaded/vectorized algorithms)
- active work now on running final fits on NERSC



PRESS RELEASE | ARGONNE NATIONAL LABORATORY

U.S. Department of Energy and Intel to deliver first exascale supercomputer

BY U.S. DEPARTMENT OF ENERGY | MARCH 18, 2019

Targeted for 2021 delivery, the Argonne National Laboratory supercomputer will enable high-performance computing and artificial intelligence at exascale.

News

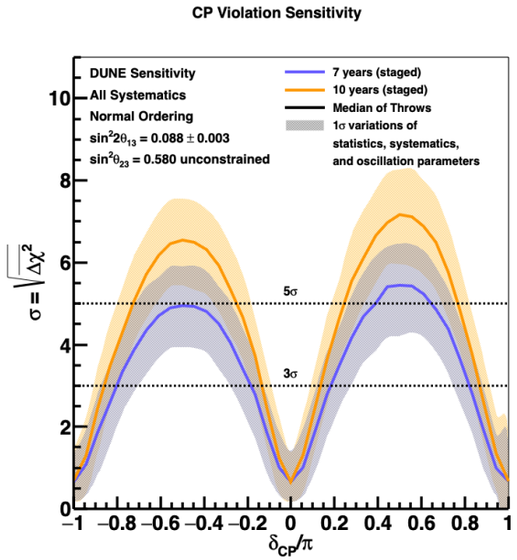
Media Contacts

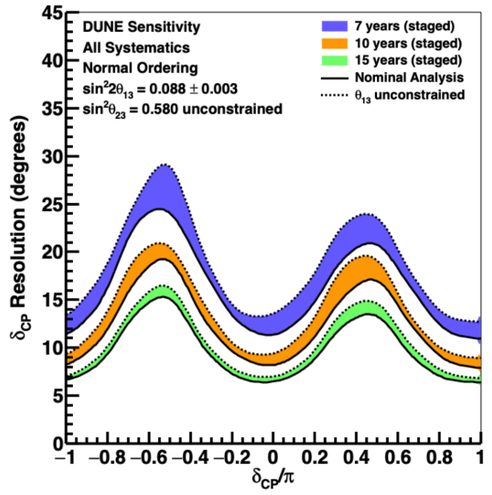
Press Releases

Fermilab DUNE

Final Fit for Long Baseline Oscillations at NERSC Cori - Callum Wilkinson, Ken Herner

- Sensitivity plots for TDR total 140k jobs
  - CPV/MH/octant sensitivity curve (each 40k jobs for 3 exposures (1 hierarchy)
  - Staging studies (5.5k for 16 exposures)
  - 2D Asimov plots (10k jobs for 3 exposures)
  - Staging and Systematic variation (5k jobs)
  - most jobs 12 CPU hours - 1.5 MB memory
- utilize small input files and MINUIT fitting procedure
- ideal for running at High Performance Computing
- next processing targeted for running on Cori at NERSC
  - Thanks to Pengfei Ding for container build scripts
- decrease the turnaround time **4-6 weeks** -> **2 days**
- Similar exploration of HPC facilities should be encouraged and incorporated into the DUNE Computing Model





Conclusions

- DUNE activity and influence within the greater HEP distributed computing community continues to grow
- successful **operation of production processing** and integration of diverse sites around the world
- data storage - active **development of community solutions** and utilization of numerous storage facilities around the world
- active evolution of the software and analysis environment to emphasize lower barrier to contributions, multithreading and lower barrier to contribution
- User analysis input will help guide interactive computing and grid tools - several tutorials this week for beginners and advanced analyzers
- **Data and Computing Model workshops planned for August (BNL) and September (FNAL)**
- **Contributions are always welcome** - contact the Computing Consortium to learn how to get involved

Tuesday, 21 May 2019

14:00 - 15:30 Computing: Model Discussion
Convener: Heidi Schellman (Oregon state)
Location: Apperance Room 13SE

16:00 - 17:30 Computing: User Meeting Joint with physics and SP/DP groups
Conveners: Heidi Schellman (Oregon state), Dr. Michael Kirby (FNAL)
Location: Hornet's Nest (WH8X)

Wednesday, 22 May 2019

14:00 - 15:30 Computing: Short and Long Term DUNE computing model
Convener: Heidi Schellman (Oregon state)
Location: Computing

Thursday, 23 May 2019

09:00 - 12:30 Computing: Tutorial
Convener: Heidi Schellman (Oregon state)
Location: One West (1W)

09:00 - 10:30 Near Detector LArTPC Analysis Tutorial
Convener: Chris Marshall
Location: Disappearance Room (13SW)

16:00 - 17:30 Computing: Single and Dual Phase ProtoDUNE - possible joi
Conveners: Heidi Schellman (Oregon state), Dr. Michael Kirby (FNAL)

Fermilab DUNE

ProtoDUNE Dual Phase Frontend - Backend testing

1. From the uTCA to the L1 Event builders and L2 Event builders (from bits to datafiles….)

We are testing this part since a few weeks: we are now carefully checking the results to understand the system and find out its time stability

2. From L2 event builders to NP02 EOS instance

This had been tested during several campaigns from December 2018 to February 2019, under different conditions (EOS configuration, optimization in the number of parallel transfers…).

1+2 : We have also been running the two parts together during the last weeks (typical single uninterrupted run duration ~ several hours)

3. From the NP02 EOS instance to central CERN EOS

We plan to use CERN FTS. We tested it last year in July, all scripts for the final configuration are ready and being commissioned.

All these parts are under systematic tests now and these tests will continue during all the next month. We do not think that we need to put them in the scope of DC3, since they are mainly related to NP02 online, and we need to test and exercise them in a constant way for other reasons (stability and stress tests of the front-end).

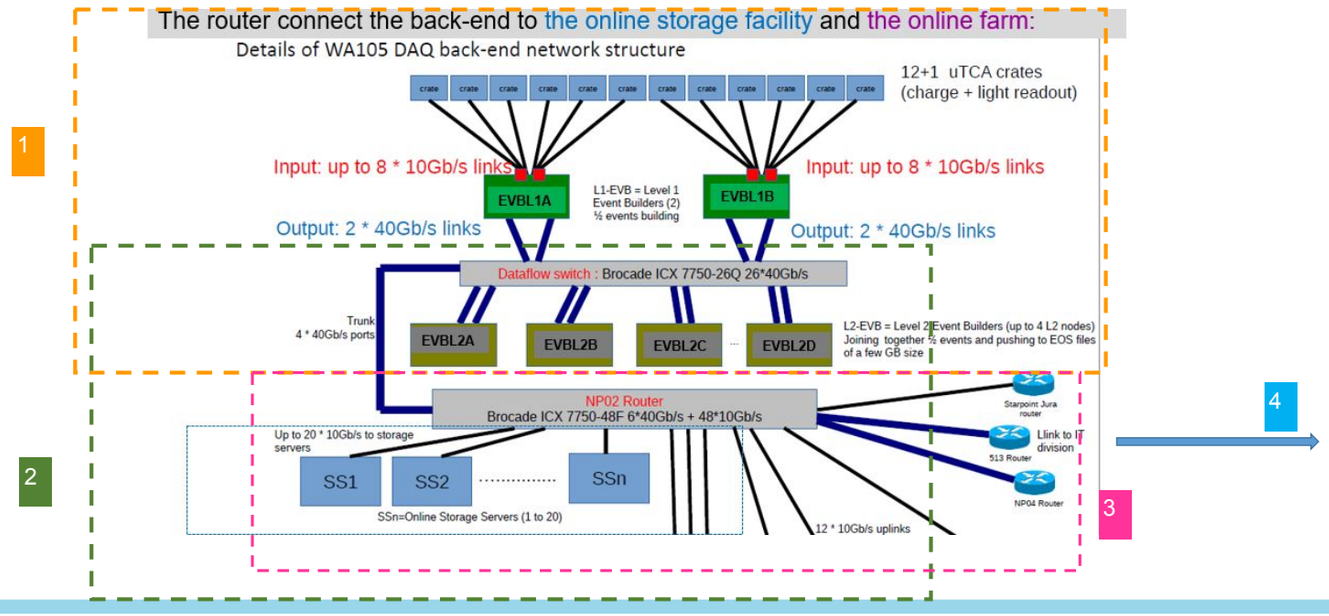
4. From central CERN EOS to CASTOR/FNAL

This is the main scope of the new data-challenge DC3

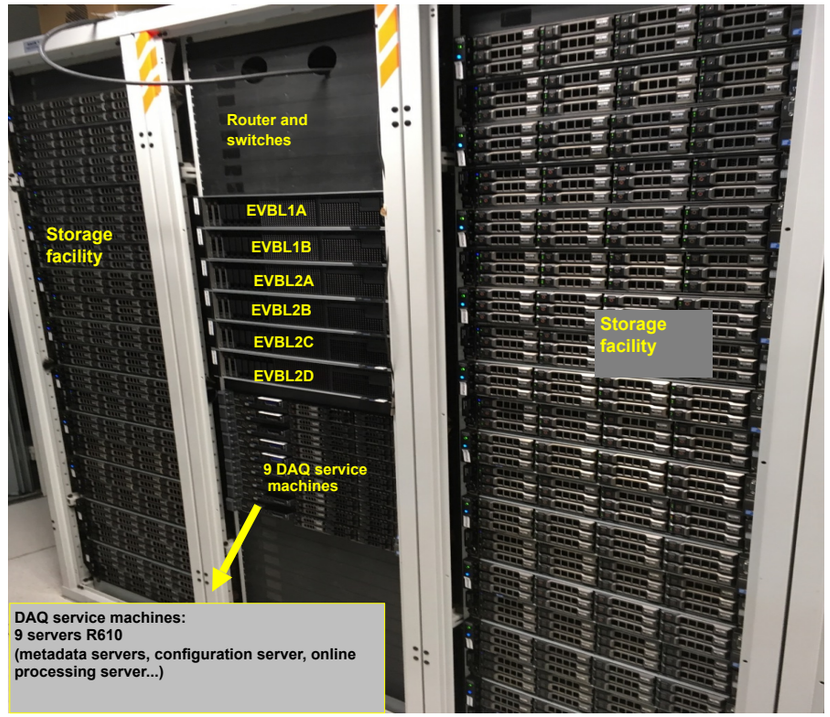
→ We would like to check that on the side of the central CERN EOS there are no problems in transferring the data to CASTOR + FNAL and that we reproduce the expected performance

ProtoDUNE Dual Phase Data Challenge 3 plans - Elisabetta Pennacchio

- Dual Phase DAQ event builder and nearline storage
  - front-end system: 12 uTCA crates, connected to back end system by 12 Ethernet optical links operating at 10 Gbit/s
  - back-end system:→2 L1 event builders, each assembling one half event
  - →4 L2 event builders: merge event halves into 3 GB that are then moved to the local EOS instance
  - →local EOS instance (np02eos) : 1PB capacity with RAID, 20 storage servers



EHN1 DAQ room



4 main data-flow blocks in the NP02 DAQ system:

1. From the uTCA to the L1 Event builders and L2 Event builders (from bits to datafiles….)
2. From the L2 event builders to NP02 EOS instance
3. From NP02 EOS instance to central CERN EOS
4. From the central CERN EOS to CASTOR and Fermilab