Heidi Schellman for the Computing Consortium

DUNE COMPUTING UPDATE

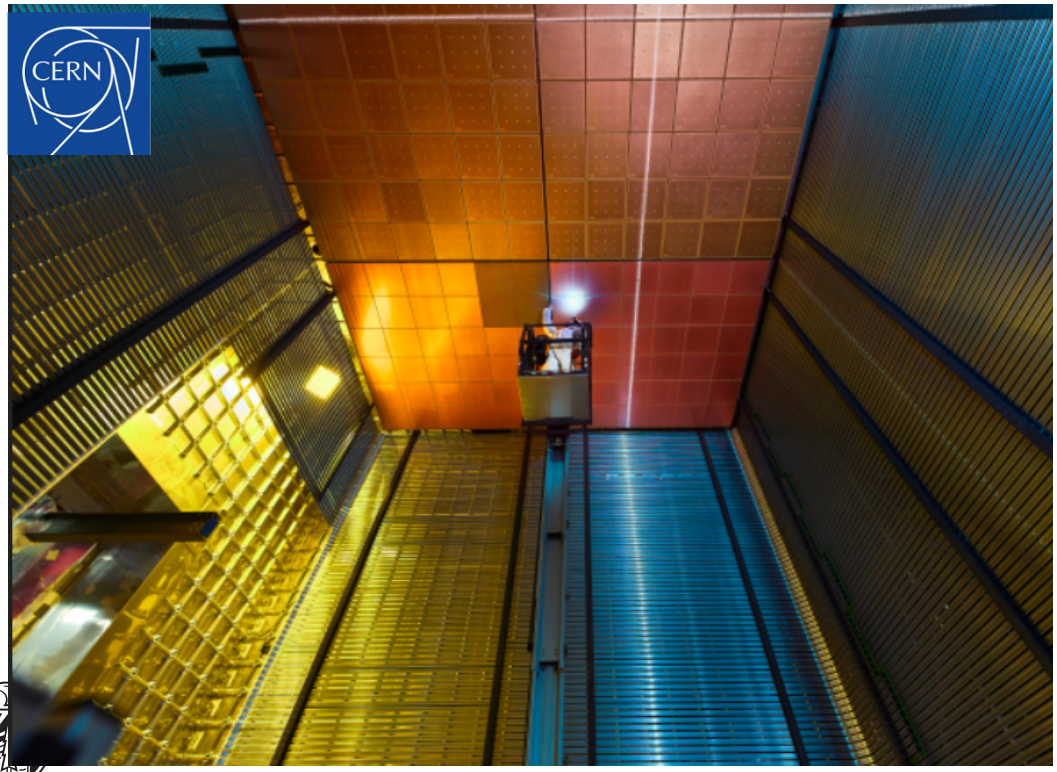
Oregon State

Series of workshops

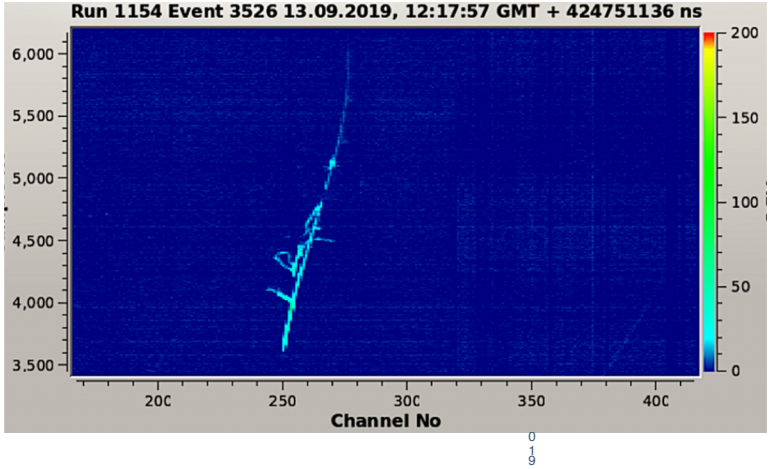
- October 18 – Kickoff workshop in Edinburgh
- DUNE collaboration meetings in January and May
- August 19 - Data Model workshop at BNL
  - Worked through interfaces with DAQ and databases
- September 9 - Computing Model workshop at FNAL – joint with GDB
  - Workflow use cases
  - Progress on uniform authentication
  - Draft Storage model
- December 19 – Database workshop at Colorado State
  - Hardware databases
  - Conditions database
- Formation of Frameworks task force (Andrew Norman and Paul Laycock)

Oregon State

ProtoDUNE Dual-Phase



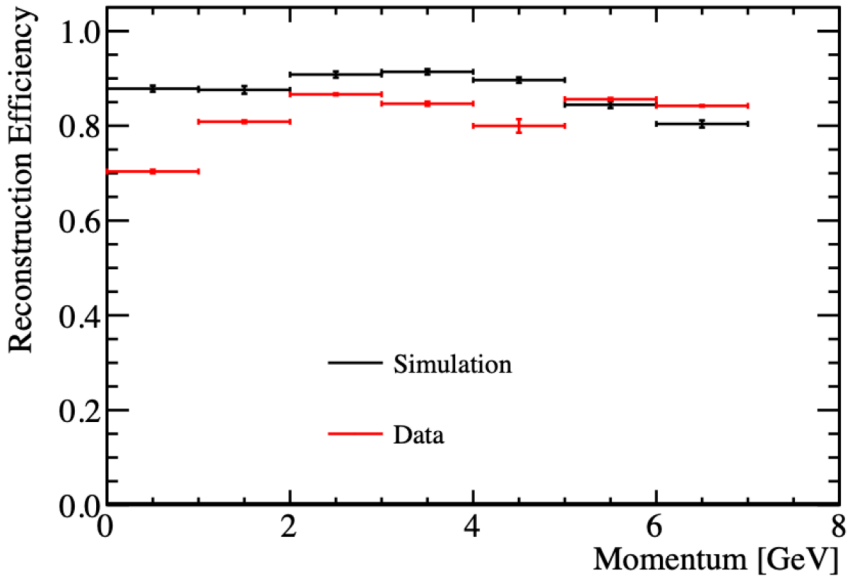
- Gas amplification raises S/N
- Data taking started late Aug 2019
- ~157 TB of raw data so far
- 110 MB/event with 2 CRP available
- No compression in first tests
- First reconstruction tests have been successful

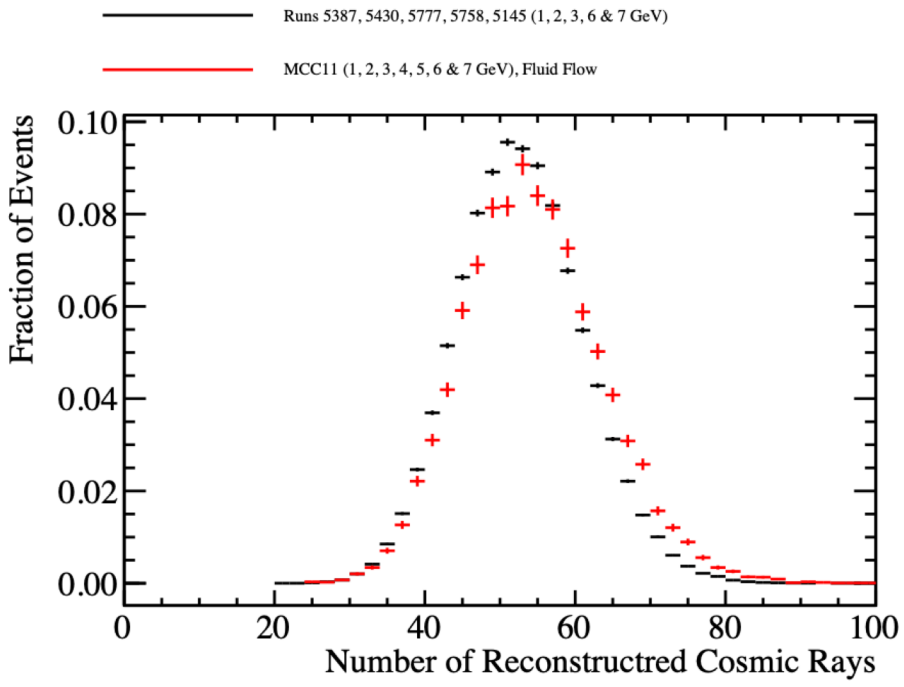


Oregon State

SP - Reconstruction Quality

1.7 PB of raw cosmic events in 2019
554 TB of beam and cosmics reconstructed with 2-D deconvolution – PASS2

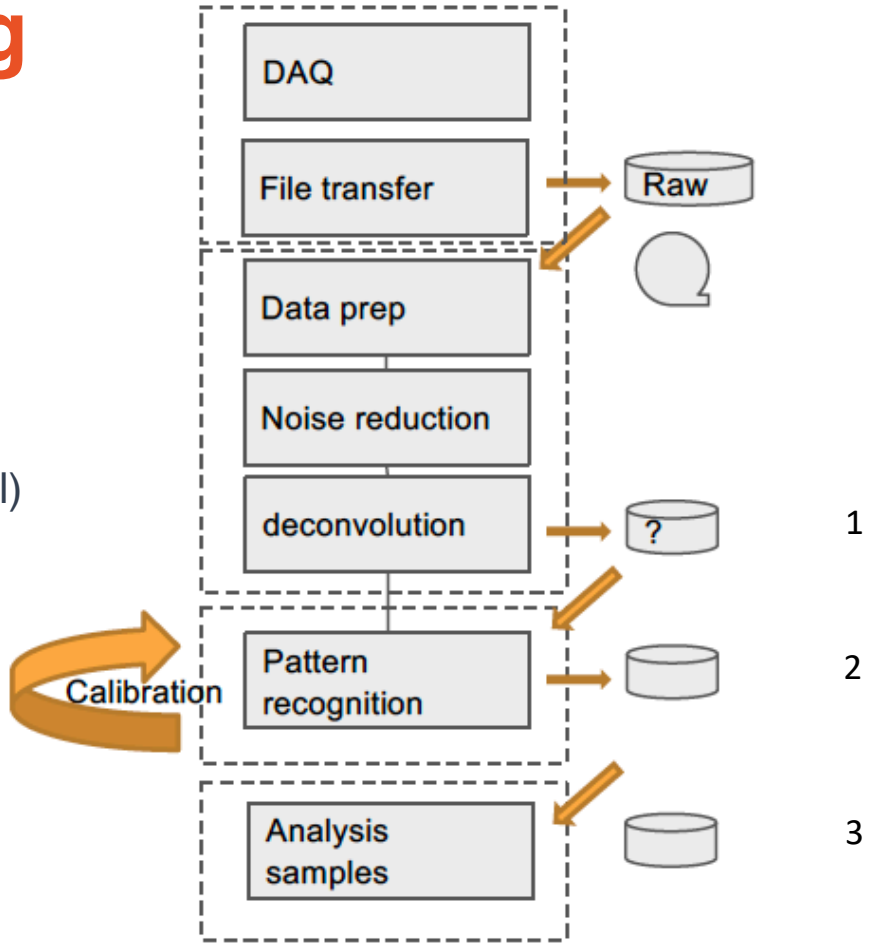




Oregon State

LAr TPC data processing

- hit finding and deconvolution
  - **x5 (ProtoDUNE)**
  - **x100 (Far Detector)** data reduction
  - Takes ~30 sec/APA
  - Do it 1-2 times over expt. lifetime
- Pattern recognition (Tensorflow, Pandora, WireCell)
  - Some data expansion
  - Takes ~30-50 sec/APA now
  - Do it ? times over expt.
- Analysis sample creation and use
  - multiple[2] iterations
  - Chaos (users) and/or order (HPC)



Current status

- Processing chain exists and works for both protoDUNEs
  - Data stored on **tape** at FNAL and CERN, staged to dCache in 3-8 8GB files
  - Use **xrootd** to stream data to jobs
  - Processing for both SP and DP takes ~**500 sec/event** (80 sec/APA)
    - Signal processing is < 2 GB of memory
    - Pattern recognition is 2-3 GB
  - Copy 2 GB output back as a single transfer.
  - SP TensorFlow pattern recognition likes to grab extra CPU's (fun discussion)
- Note: ProtoDUNE-SP data **rates** at 25 Hz are equivalent to the 30 PB/year expected for the full DUNE detector. (Just for 6 weeks instead of 10 years)
- ProtoDUNE-DP
  - Data transfer and storage chain operational since August – up to 2GB/s transfer to FNAL/IN2P3
  - Reconstruction has started

Oregon State

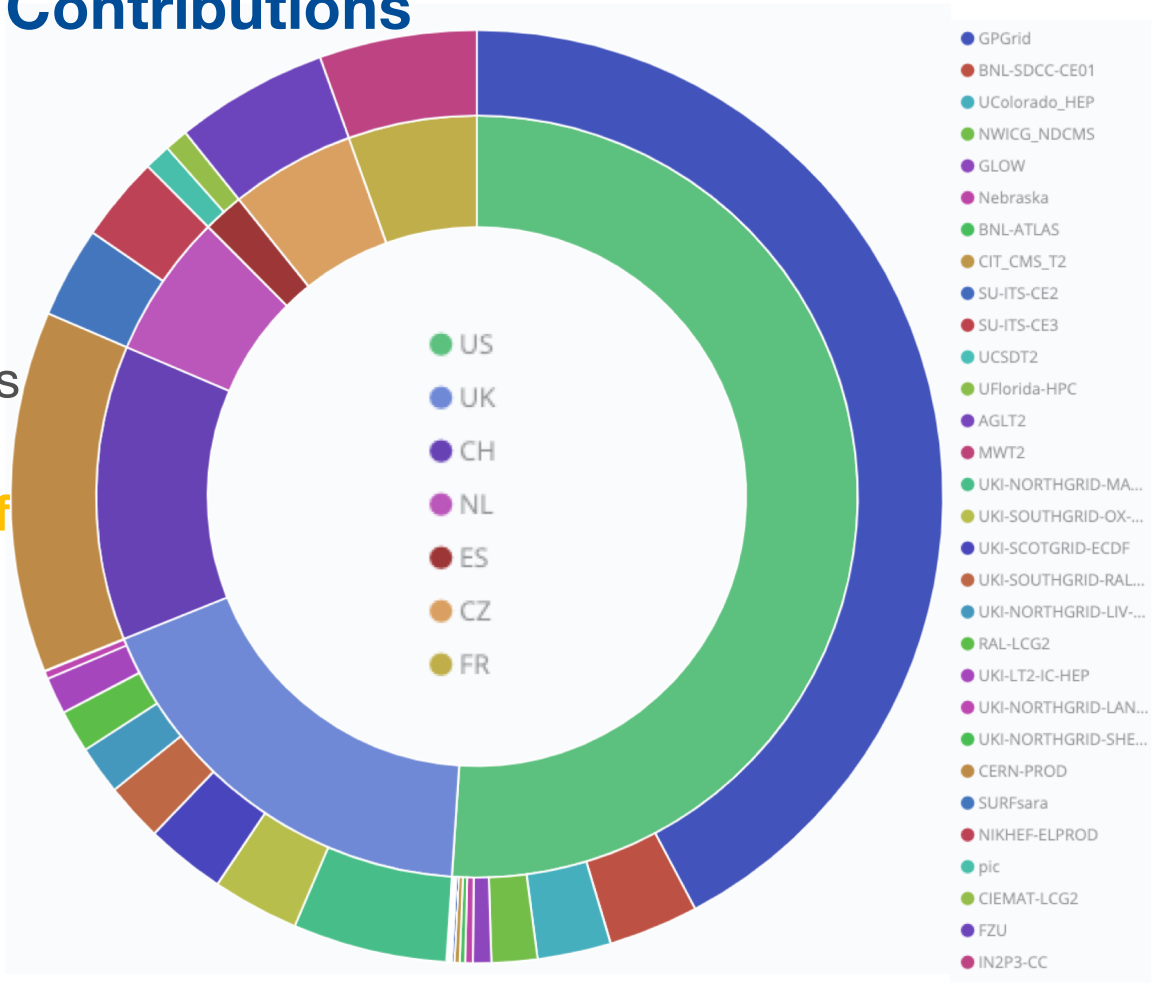
International CPU Contributions

PDUNE-SP data took 6 weeks to collect

Reprocessing passes are generally 4-6 weeks on ~8000 cores

In 2019 so far, **49% of production wall hours are from outside USA**

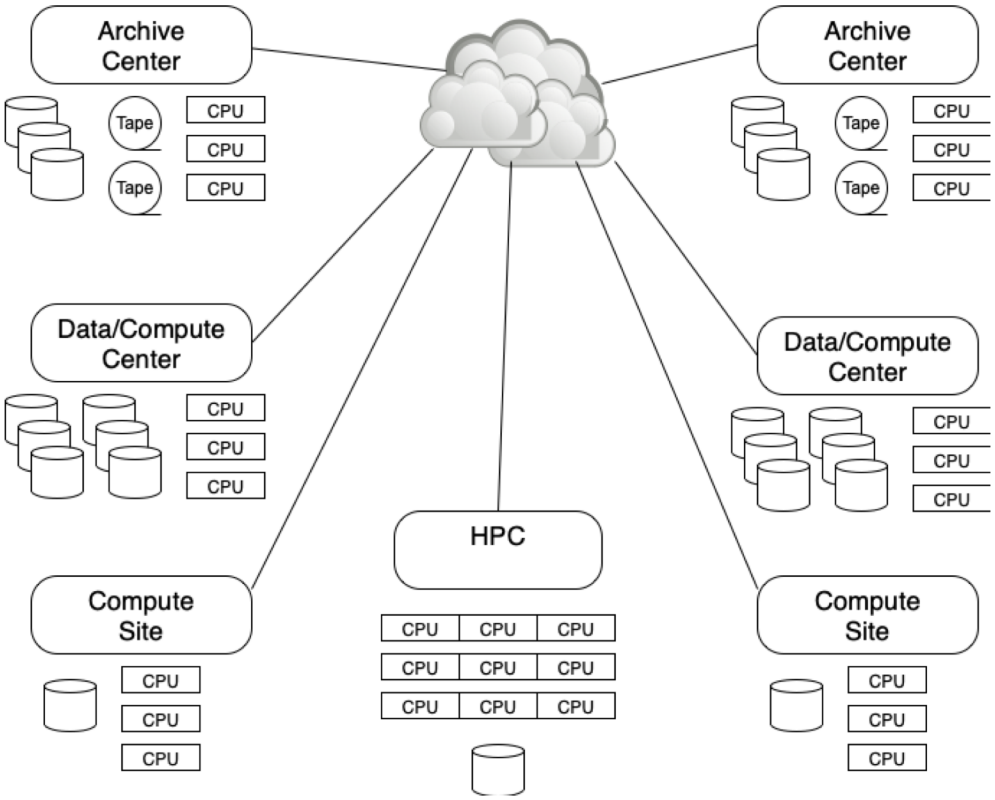
Actively working to add more sites and countries



Ken Herner's talk

Draft Distributed computing model

- Less “tiered” than current WLCG model → **DOMA**
- Collaborating institutions (or groups of institutions) provide significant disk resources (~1PB chunks)
- **Rucio** places multiple copies of datasets
- **We likely can use common tools:**
  - **But need our own contribution system**
  - **And may have different requirements for dataset definition and tracking**



Future compute model draft

- Tape archive for raw data at two places (FNAL + external to US)
  - Second site may be a consortium rather than a single site
  - Currently FNAL/CERN
- Important derived data sets are disk resident and controlled by Rucio
  - Minimum of two copies
- Data tracked through a replacement for the existing SAM system
- Data delivered either by local copies or through streaming (xrootd)
- Jobs submission systems being evaluated
  - FNAL POMS/GlidenWMS submissions work
  - Demonstration at RAL of running a DIRAC job using SAM for data delivery

Near term storage and CPU assumptions

- Raw data based on protoDUNE per-APA
  - SP / 2.5 compression
  - DP / 10 compression
- Reconstructed data assumes raw hits are not stored. Ends up being about 1/10 size of uncompressed raw data.
  - Both DP and SP take ~80 sec/APA of which ~1/3 is raw hit processing and the rest is pattern recognition.
- Assume 2 copies of raw data on tape
- Assume 2 copies of reconstructed data on disk – kept for 3 years
- ProtoDUNE simulation takes ~ 40 min/event for outputs are ~200 MB/event

Oregon State

Storage estimates docdb-17086

		2018	2019	2020	2021	2022	1st FD module
		asbuilt					compressed
SP	Events, M	10.9	19.4	6.5	23.5	30.0	2.4
	Raw data, TB	751	1,197	448	1,624	2,072	5,944
	Reco data, TB	1,501	395	134	487	621	59
per pass	CPU, MH	2	3	1	4	5	12

DP	Events, M	0.0	1.0	21.6	61.8	64.4	0.0
	Raw data, TB	0	114	238	1,359	1,416	0
	Reco data, TB	0	137	238	1,359	1,416	0
per pass	CPU, MH	0	0	4	10	11	0
---							
yearly							
total	Events, M	10.9	20.4	28.1	85.3	94.3	2.4
2x	Raw data, TB	1,501	2,621	1,371	5,965	6,975	11,888
2 passes	Reco data, TB	3,003	1,063	744	3,692	4,075	119
copies	reco data, TB	6,005	2,127	1,488	7,383	8,149	238
	CPU, MH x2	4	7	9	28	31	12
	total storage	7,507	4,748	2,859	13,348	15,124	12,126

Oregon State

Simulation and Cumulative storage

	sim events, M	2.5	10.0	10.0	10.0	10.0	10.0
	sim size	500	2000	2000	2000	2000	2000
	CPU - MHrs	1.875	7.5	7.5	7.5	7.5	7.5

Cumulative		**2018**	**2019**	**2020**	**2021**	**2022**	**1st FD module**
Raw *2	TAPE	1,501	4,123	5,493	11,458	18,433	30,321
Reco 2 versions for 2 years		3,003	4,066	1,807	4,436	7,766	4,193
Reco *2 copies	DISK	6,005	8,132	3,615	8,871	15,532	8,387
sim for 2 years		500	2,500	4,500	6,500	8,500	10,500
sim*2 on disk		1,000	5,000	9,000	13,000	17,000	21,000
total disk		7,005	13,132	12,615	21,871	32,532	29,387

Oregon State

CPU needs

RECONSTRUCTION

- ProtoDUNE events are more complex than our long term data.
  - ~**500** sec to reconstruct 75 MB compressed – 7 sec/MB
  - For FD, signal processing will dominate at about 3 sec/MB
  - < 30 PB/year of FD data translates to **~100 M CPU-hr/year**
  - That's **~ 12K cores** to keep up with data. But no downtimes to catch up.
- Near detector is unknown but likely smaller.

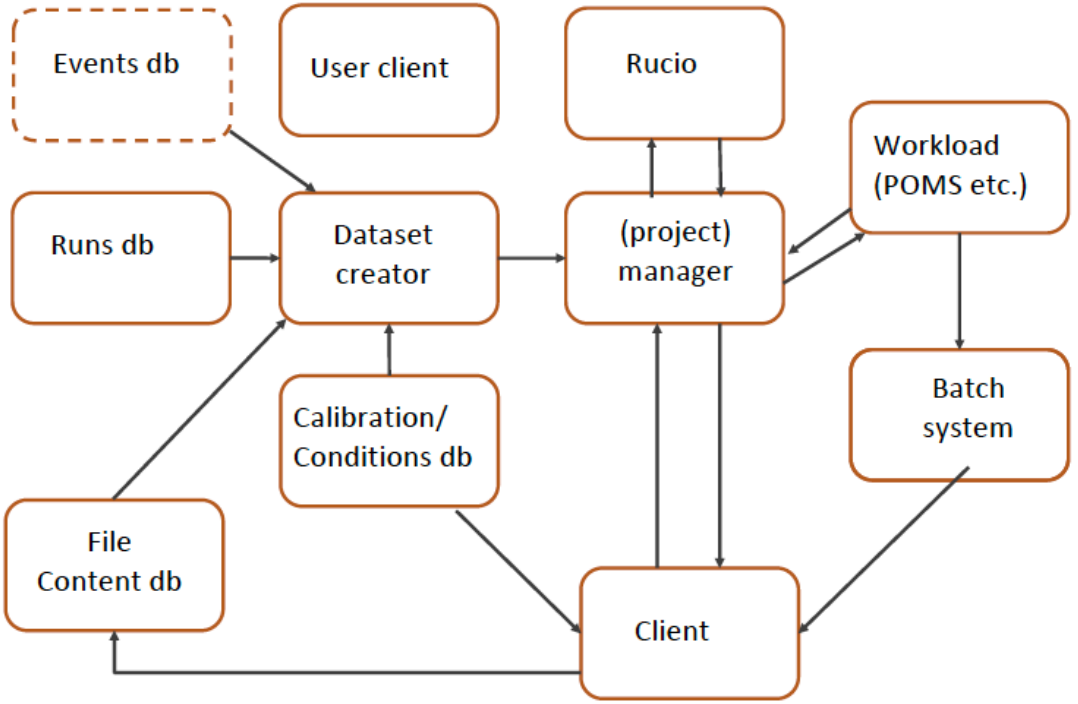
ANALYSIS (Here be Dragons)

- NOvA/DUNE experience is that data analysis/parameter estimation can be very large
  - ~ 50 MHrs at NERSC for NOvA fits

Oregon State

Data tracking project

- FNAL neutrino experiments use an updated version of the SAM* file database from D0/CDF
  - Needs a remodel (gut renovation?)
- Develop replacement for SAM components that describe data
  - Beam/detector config
  - Processing provenance
  - Normalization
- Use Rucio for file placement and location



Long term development personnel needs (based on LHCb experience) ~ 20 FTE

- Distributed Computing Development and Maintenance - 5.0 FTE
- Software and Computing Infrastructure Development and Maintenance - 6.0 FTE
- Central Services Manager and Operators - 1.5 FTE
- Computing Shift Leaders - 1.5 FTE
- User Support - 1.0 FTE
- Application Managers and Librarians - 2.0 FTE
- Database Optimization and Maintenance - 0.5 FTE
- Distributed Data Manager - 0.5 FTE
- Distributed Workload Manager - 0.5 FTE
- Distributed Production Manager - 0.5 FTE
- Overall Coordination 2.0 FTE

Oregon State

Unknowns for the future

- $$$ - many of those people are computing specialists that need to be paid
- Near detector:
  - Rate ~ 1 Hz, technology not yet decided – But software integration underway
  - Occupancies will be similar to ProtoDUNE at 1 Hz → O(1) PB/year?
- Processor technologies
  - HPC's
  - Less memory/more memory?
  - GPU's? << signal processing may love these!
- Storage technologies
  - Tape
  - Spinning disk
  - SSD
  - Something else?

Oregon State

US consortium

- 4 national labs (ANL, BNL, FNAL and LBNL)
- 5 US universities – Colorado State, Minnesota, Oregon State, Wichita State and William and Mary
- 3 thrusts
  - Databases – Hardware and integrated design
  - Data model and algorithms for large scale computation on diverse hardware
  - Coding standards, build tools and training to ensure ability of collaborators to contribute to robust algorithms for reconstruction and analysis
- White paper this month.
- Formal proposal going in early in 2020.

We stand on the shoulders of giants

- **Art framework, Larsoft, Pandora and WireCell**
  - NOvA
  - ArgoNeut
  - MicroBooNE
- **Models and simulation**
  - GEANT4 and Fluka
  - GENIE, Neut, GiBUU, NuWro, …
- **Beam models**
  - G4numi -> g4lbnf
  - ppfx
- **Infrastructure**
  - Jobsub/POMS
  - WLCG and OSG
  - Enstore, dCache
  - uCondb and ifbeam
  - SAM catalog
  - Elisa logbook
  - Rucio
  - Authentication systems
- **OSG/WLCG/HSF for new ideas!**

Oregon State

Pete explains the growth of contributions