DUNE Computing Overview

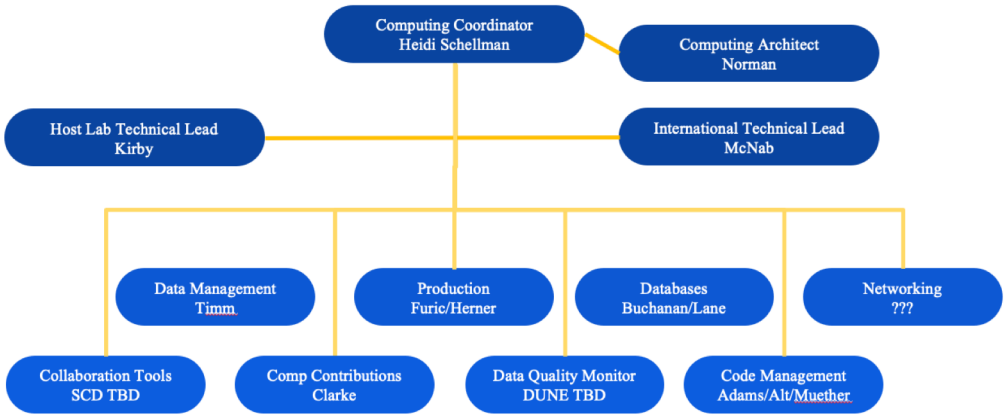
Michael Kirby, Heidi Schellman

October 15, 2019

DUNE Computing Outline

- Current Computing Campaigns
  - ProtoDUNE-SP Reprocessing
  - ProtoDUNE-DP Processing
- Progress on Major Systems
  - Data Management
  - Databases
- ***Summary of Event Data Model Workshop***
- ***Summary of Computing Model Workshop***
- Future Direction
  - Formation of Computing Contribution Board
  - Future plans towards a Computing CDR and TDR

DEEP UNDERGROUND NEUTRINO EXPERIMENT
COMPUTING CONSORTIUM



Fermilab

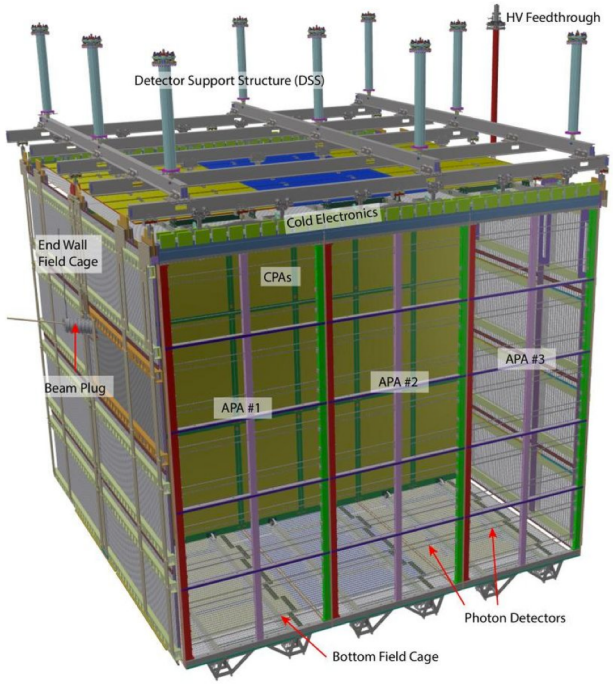
Success with ProtoDUNE 2018/2019

ProtoDUNE Projects @ CERN

- ProtoDUNE are two separate technology demonstrators being hosted by CERN
- Designed to prove viability of detector and readout technology
  - Have associated physics program in hadron response which is desired for DUNE energy scale and response calibration

- ProtoDUNE Single Phase (SP)
  - Ran with beam fall 2018
  - Continues to run with cosmic rays (Present)
  - First testbed for portions of Computing and Data Model

- ProtoDUNE Dual Phase (DP)
  - Ramping up for cosmic ray running (August-Present)
  - Different computing requirements driven by detector readout needs
  - Second testbed for refining Computing and Data Model



Fermilab

Information prepared by Ken Herner

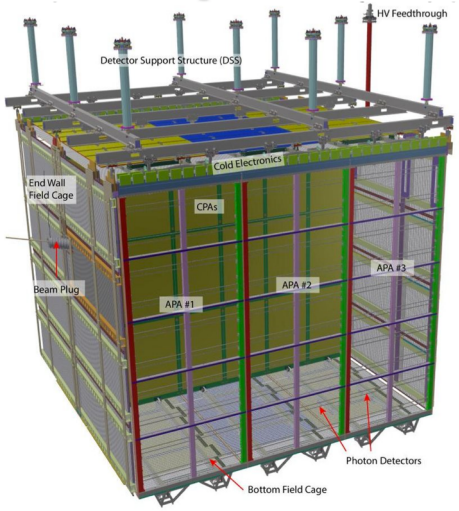
Success w/ ProtoDUNE Single Phase Reprocessing Campaign

- Reprocess ALL *good beam runs* with latest production release†
- Generated MC for 1, 2, 3, 6, 7 GeV beam conditions
- Large Scale campaign
  - Run across OSG/WLCG resources
  - 4 - 6 week to complete
    - ~8 M beam events , 7 min/evt, ~ 1M CPU hours for data
    - ~ 4 M MC events, 10 min/evt ~ 700k CPU hours for MC

- Identified Scale Behavior of DUNE/ProtoDUNE workflows
  - Several issues with file metadata + data catalog
    - learning experience and improving services
  - New *Tensorflow* algorithm exceeding CPU limits on worker nodes
    - Opportunity for the future improvements

- Major contribution from summer intern/student programs:
  Hakan Solak (U. Chicago)
  Robert Schiattarella (Napoli)

Data set	Data % Complete	MC % complete
Runs 5387, 5809, 5770, 5834, 5826, 5432, 5786, 5204	96.7%	--
0.3 GeV	95.0%	--
0.5 GeV	91.3%	--
1 GeV	95.9%	99.5%
2 GeV	97.7%	105%
3 GeV	85.9%	92.4%
6 GeV	91.5%	97.8%
7 GeV	87.8%	103%

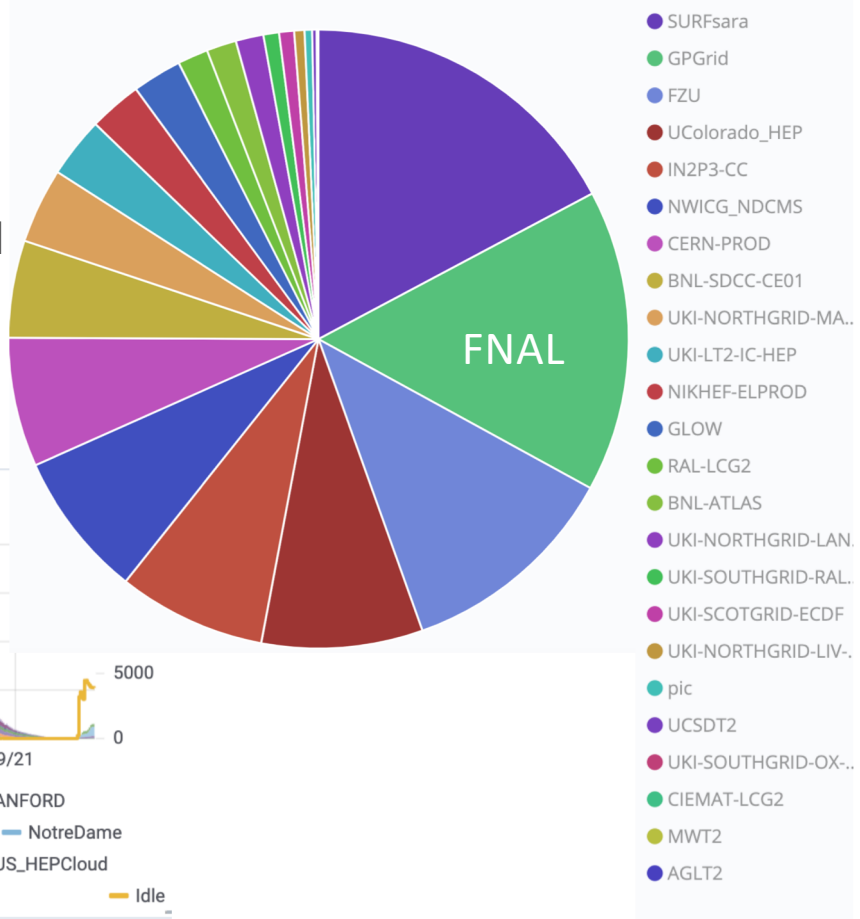


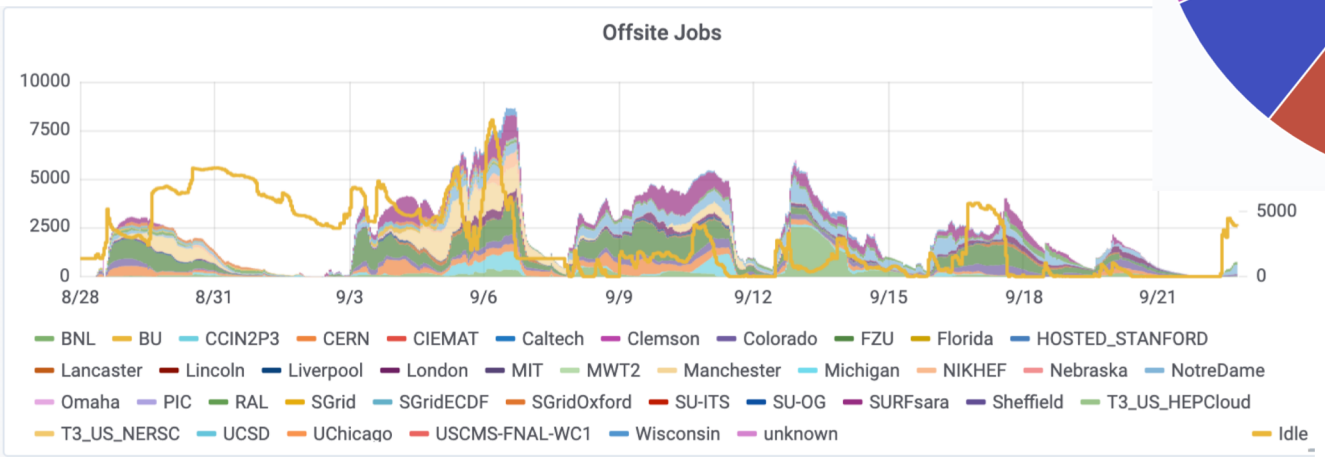
Fermilab

Information prepared by Ken Herner

Leveraging Distributed Resources w/ ProtoDUNE SP

- **Major contributions from over 25 institutions**
- NIKHEF, FZU, France, CERN, BNL, UK, Spain – account for international contribution > 75% of total
- Fermilab contribution significant but not largest in pool
- For ProtoDUNE-SP, offsite concurrent peak: ~8000 non-FNAL jobs



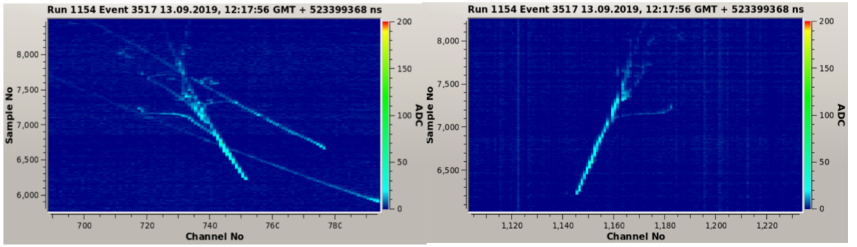


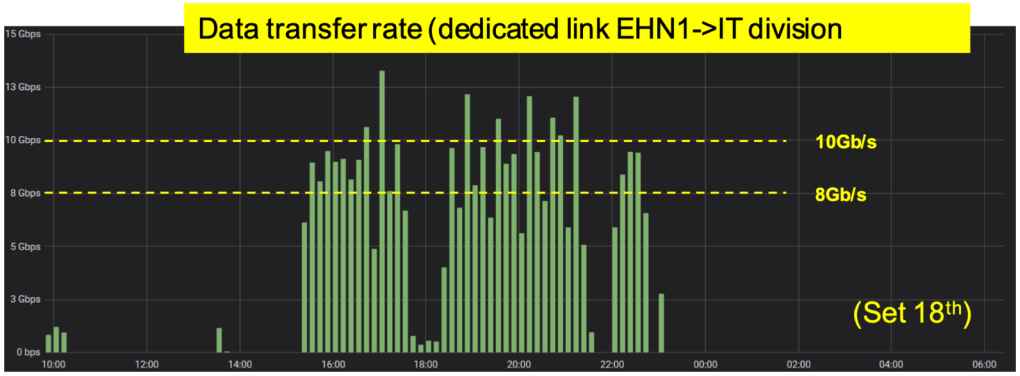
Fermilab

Information prepared by Ken Herner

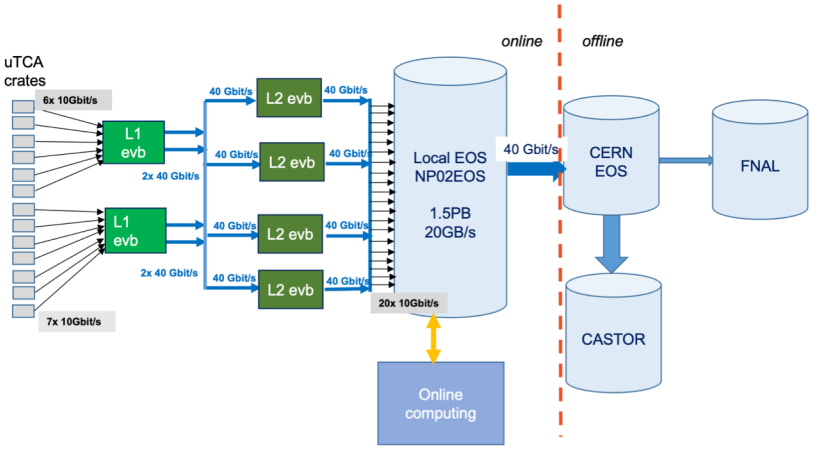
Success w/ ProtoDUNE Dual Phase: Data Archiving

- **Started August 2019**
- DAQ running at 10 - 30 Hz - 30 evts → 3 GB file
- 28 Aug – 13 Oct ~2M evts recorded ~ 65k files
- ~163 TB of raw data recorded from the DAQand transferred from np02eos to CERN EOSPUBLIC
- Cataloged and archived at CERN CASTOR and FNAL Enstore





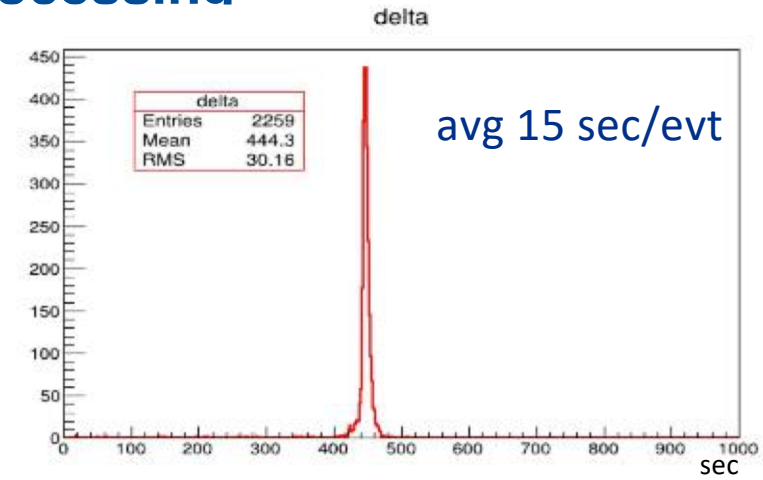
NP02 DAQ/network infrastructure

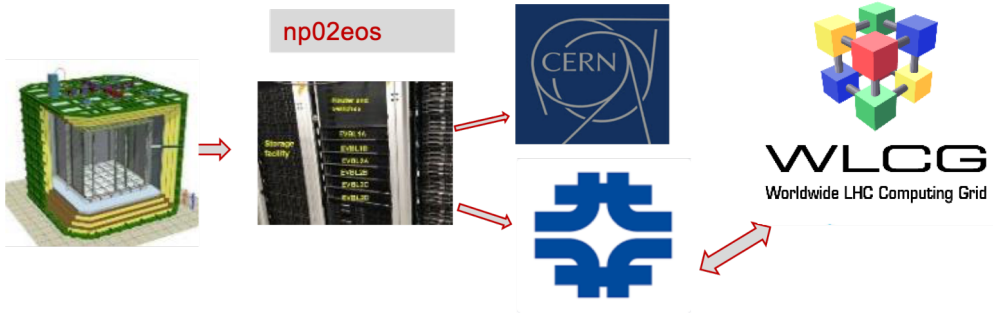


Information prepared by Elisabetta Pennacchio

ProtoDUNE Dual Phase online and offline Processing

- Fast online reconstruction being run at online processing facility - 30 servers, 450 cores, SLURM
- Optimization of reconstruction parameters and submission scripts have been performed
  - 7.5min processing time per raw datafile (15s/evt)
- First pass hit reconstruction almost ready for full production usage.





Notes:
- Offline processing through LArSoft foreseen
- Ingest module for raw data into LArSoft tested in July '19
  - Validation in progress
- Will launch systematic reconstruction campaign once validated
- Build upon the Data Challenge-3 processing experience

Fermilab
Information prepared by Elisabetta Pennacchio

Recent DUNE Data Ingest

September 2019

8.2 PiB

6.9 PiB

5 GB/s

-1 GB/s

- Recent: **1.17PB** ingested in September 2019
- NP02: 88TB cosmic, 20TB test, 1.6TB pedestal (not updated for Oct)
- NP04: 45TB cosmic, 14TB test
- PDSPProd MC: 234TB simulated (g4), 329TB detector-simulated(detsim), 284TB (reco) and 33 TB Pandora
- PDSPProd Reco: 184TB full-reco, 34 TB Pandora files
- **OVERALL DUNE data on tape 8.6PB**

Fermilab

Information prepared by Steve Timm

Design & Success other Infrastructure for ProtoDUNE and Early DUNE 2019-

Database Necessities & Software Integration

- DUNE has identified a CRITICAL need for early designs of databases to support:
- ProtoDUNE Physics and Technology Demonstration Programs
  - Calibrations, Detector Hardware, Channel Mapping, Data Quality etc….
    - Direct application to DUNE designs and needs
  - CERN Accelerator & Beam information etc…
    - No direct application to DUNE designs, but feed into parts of computing model

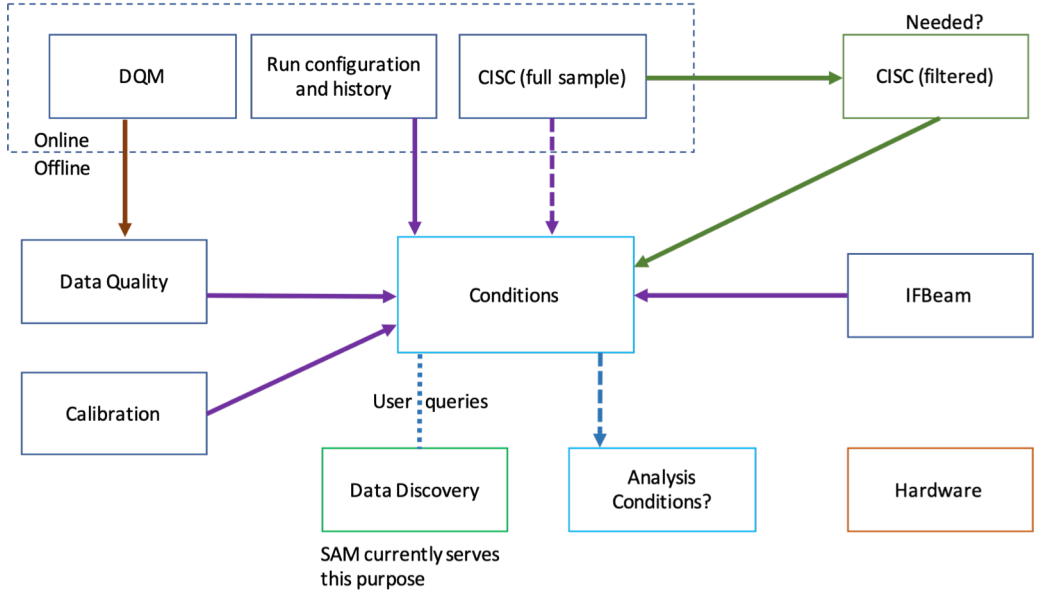
- DUNE/LBNF have needs separate needs for Logistics, Hardware, QA, etc…
  - Immediate need, but need to have a design that supports use of the information at later stages of the project and into the operations/physics phases.
  - Need to emphasize connections between databases

- **Early Need for consistent software stack to support both simulation and early commissioning of detectors.**

Information prepared by Norm Buchanan

Databases Design Proposal

- Active planning and design of Database Ecosystem for DUNE
  - Based on what is currently being done for ProtoDUNE
- Need to emphasize connections between databases
- Individual databases will “feed down” into centralized conditions database that will serve offline jobs.
- Current Implementation of conditions database utilizes unstructured RDB with blobs O(10 MB/run) of metadata coming from DAQ
- **Database workshop Planned @ Colorado State U. Dec 3-4**



CY2019 Database Model Tasks

- collect use cases for all DBs
- define interfaces and document
- develop requirements (hardware DB available for review and feedback now)
- CY2020 - decide on implementation and prepare for ProtoDUNE II

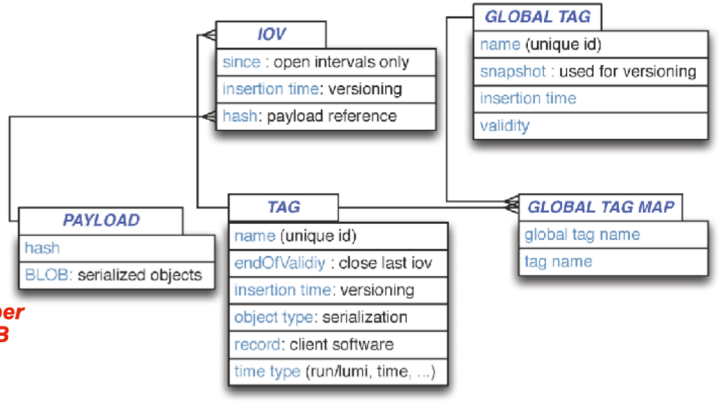
Fermilab

Information prepared by Norm Buchanan

Databases Design Proposal

- Alternate implementation being considered
- Based on Belle II Database Infrastructure
- Presented at Data Model workshop (Aug)
- Strong connection to HSF-based approach
- Paul Laycock leading this portion of investigation presented HSF-based approach used for Belle II
- Continued planning at DB workshop Dec 3-4

- Metadata Model: relational DB
  - Takes care of correlating versions of different calibrations for a coherent versioning system
- ***Payloads factorised from metadata***
- ***No need to define a schema per subsystem, payload is a BLOB***
- IOVs and payloads resolved independently:
  - ***Cache-friendly design***



P. Laycock - BNL

CY2019 Belle-II Investigation Database Model Tasks

- Collect use cases for all DBs
- Define and document proposed interfaces
- Develop requirements with impact statements
- Review current DB hardware availability and present feedback to experiments
- CY2020 Decision Point - decide on implementation path for ProtoDUNE II

Fermilab

Information prepared by Norm Buchanan

Software Management Report

- Ongoing release maintenance work: David Adams, Christoph Alt, Tom Junk, Tingjun Yang (Software Rel. Team)

- Effort is designed to:
  - **Keep up with rapid DUNE development**
  - Remain current/compatible w/ LArSoft releases (+leverage new features)
  - *Serve as gatekeepers w.r.t. decisions on software features to included in collaboration wide data processing, simulation and reprocessing activities*

Future Development Focus

- Python 2.x -> 3.x Transition
- New package management and deployment strategies (ala SPACK)
- Migration to public version control tools (note: LArSoft scheduled to migrate to github)
- Changes to architectural support models (i.e. shed support for OSX in favor of containerization strategies)
  - Will enable better use of future platforms and environments (e.g. HPC centers) while retaining capabilities (i.e. developers can still run on their laptops for dev/debug)
- Long term art strategy will require substantial changes/adaption
- Address large memory footprint for DUNE events and multi-threading needs for LAr data processing in context of data processing/reconstruction frameworks

DUNE International Computing Consortium Activities (circa 2019)

Computing Consortium Organizational Work

- Formal establishment in Winter 2019
  - Formation of structure and organization at winter collaboration meeting (CERN)
- Population of working groups in Spring/Summer 2019
- Generated Executive Summary for Computing (White Paper) [for TDR] Spring 2019
- Organized and Held Directed Workshops:
  - **DUNE DATA MODEL** : Location – Brookhaven National Lab (August 14-16)
  - **DUNE COMPUTING MODEL** : Location – Fermilab (Sept 9-10)
    [Held in conjunction with Grid Deployment Board meeting]
- Held Multiple Joint and Auxilary Sessions at Sept DUNE Collaboration Meeting
  - Including Joint meeting with Calibration Consortium (9/25)
- Upcoming Workshop:
  - **DUNE DATABASE MODEL** : Location - Colorado State (Dec. 3-4)

- In planning stages for CERN collaboration meeting in late January ‘20

Towards
The Data and Computing Models for DUNE

Workshops to develop DUNE Computing and Data Model

- **Goal of the Data Model Workshop (Brookhaven National Lab)**
  - map the data flow for the DUNE far detector required for primary physics analysis in terms of data tiers, volumes, lifetimes, and interplay with additional services
  - generate requirements from the data model for additional Software Research and Development projects in the upcoming years
  - estimate minimal level of resources (primarily storage and effort) necessary to meet these requirements

- **Goal of the Computing Model Workshop (FNAL)**
  - draft a model for computing infrastructure that matches the requirements determined from the Data Model Workshop
  - identify baseline and unique workflows for DUNE processing and develop requirements based upon those workflows
  - establish a model for computing centers and sites based upon a forward looking concepts and ideas - focus on the impact on resource requirements and cost

DUNE Data Model Workshop (BNL Aug 14-16, 2019)

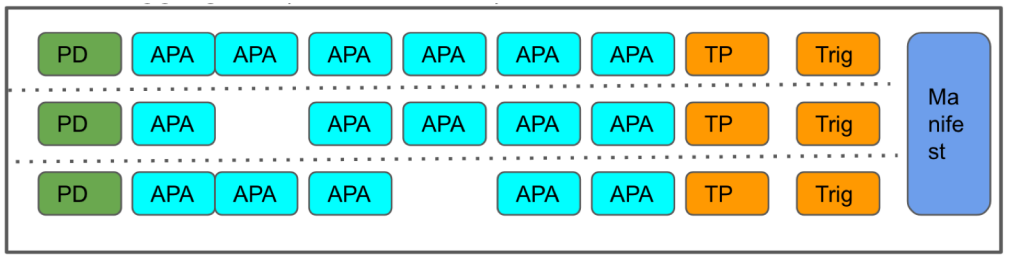
- ***Extensive Discussion with DAQ Consortium regarding Model and Needs***

Topics:

- Tightly coupling of data model to physics requirements
  - beam data
  - calibration streams
  - extended readout streams
- Database structures and schema
- Metadata service and data discovery integrated with Rucio (WLCG/HSF)
- Developed requirements for processing framework and identification of needed projects
- Data reduction studies, data tiers hierarchies and data volumes, and derived datasets and lifetimes

Source	Annual Data Volume	Assumptions
Beam interactions	27 TB	10 MeV threshold in coincidence with beam time, including cosmic coincidence; 5.4 ms readout
^{39}Ar, cosmics and atmospheric neutrinos	10 PB	5.4 ms readout
Radiological backgrounds	< 2 PB	< 1 per month fake rate for SNB trigger
Cold electronics calibration	200 TB	
Radioactive source calibration	100 TB	< 10 Hz source rate; single APA readout; 5.4 ms readout
Laser calibration	200 TB	10^6 total laser pulses; half the TPC channels illuminated per pulse; lossy compression (zero-suppression) on all channels
Random triggers	60 TB	45 per day

Estimated DAQ Volume from a FD Single Phase Module



Potential file based design of raw data model

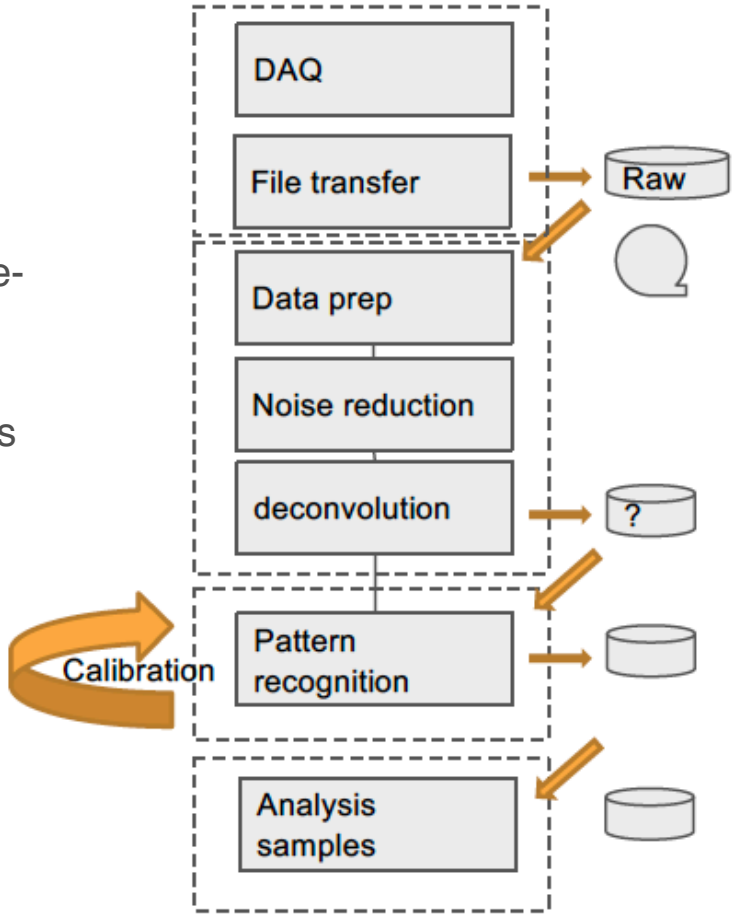
Fermilab

FD Beam Data Processing

- Main production processing of LBNF beam **trigger records** for reconstruction of far detector interactions targeting physics measurements.
- Multiple stages suitable for different architectures and different live-cycles
  - 1) hit finding and deconvolution – 1-2 iterations
  - 2) Pattern recognition (Tensorflow, Pandora) – multiple iterations
  - 3) Analysis sample creation and use – multiple[2] iterations
- *Identified significant opportunities for data reduction*

Data reduction estimates	MB/evt	MB/evt	
	Rawdigit	Wirehits	reduction
ProtoDUNE data	53.2	4.6	11.6
ProtoDUNE MC	46.1	11.9	3.9
FD nue	2073.6	11.6	179.1
FD bkg	2073.6	7.1	291.1

TPC Only Data Volumes



Fermilab

FD Beam Data Processing

ACTIVE PROCESS OF ENUMERATING AND MODELING KNOWN DUNE/NEUTRINO WORKFLOWS

- Definition of the input and output size for each of those tasks:
  - Task 1 Input: Raw data (5K beam, 10K solar, 1.5M cosmic) - 9.8 PB/module/year
  - Task 1 Output: Hits + ROI - 50 TB/module/year
  - Task 2 Input: Hits + ROI - 50 TB/module/year
  - Task 2 Output: Evt classification + physics quantities - 65 TB/module/yr
- Data lifetime of the output for each stage:
  - Raw data - infinite lifetime
  - Task 1 Output - 5 year lifetime (reprocess hit generation once)
  - Task 2 Output – 1-2 year lifetime (reprocess event class twice/yr)

Far Detector								
Data Type	**Amount/yr**	**Copies**	**Num of versions**	**Total/yr**	**Disk**	**Tape**	**Lifetime on disk**	**Note**
Raw	**10 PB**	**2**	**1**	**20 PB**		**100%**	**Infinite**	
Hits	**50 TB**	**3**	**2**	**300 TB**	**100%**	**100%**	**5 year**	
Reco	**65 TB**	**3**	**2**	**390 TB**	**100%**	**0%**	**2 year**	
MC	**100 TB**	**3**	**2**	**600 TB**				**10:1 Ratio with raw?**
Raw like	**1 PB**	**1**	**1**	**1 PB**	**100%**		**2 year**	**~10% Raw for ML Training**

These are 0th-pass estimates without photon detector

Fermilab

Offline Workflow Descriptions

- Have completed (or are performing) similar studies for other workflows:
  - Calibrations (laser, pns, etc)
  - Simulations (including near detector)
  - Background estimation (data driven)
  - Analysis processing/Selection
  - Parameter estimation

- *Use the summary of these workflows to define resource needs for computing model*

- **Expect we will be able to test these models of workflows with Run II of ProtoDUNE SP and ProtoDUNE DP**

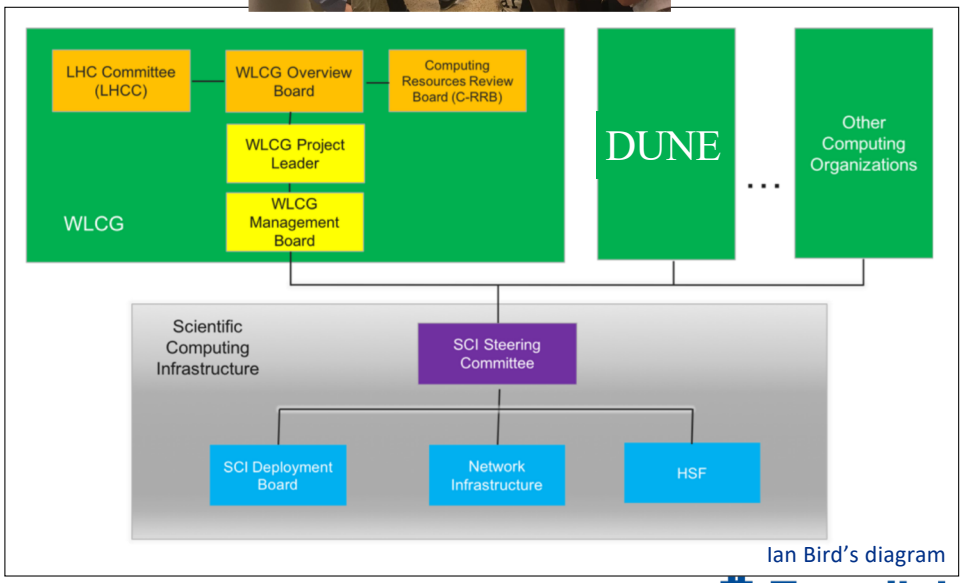
Far Detector								
Data Type	**Amount/yr**	**Copies**	**Num of versions**	**Total/yr**	**Disk**	**Tape**	**Lifetime on disk**	**Note**
Raw	**10 PB**	**2**	**1**	**20 PB**		**100%**	**Infinite**	
Hits	**50 TB**	**3**	**2**	**300 TB**	**100%**	**100%**	**5 year**	
Reco	**65 TB**	**3**	**2**	**390 TB**	**100%**	**0%**	**2 year**	
MC	**100 TB**	**3**	**2**	**600 TB**				**10:1 Ratio with raw?**
Raw like	**1 PB**	**1**	**1**	**1 PB**	**100%**		**2 year**	**~10% Raw for ML Training**

DUNE Computing Model Workshop (FNAL Sep 9-12, 2019)

- Actively working with WLCG and other stakeholders to evolve HEP computing model
- Dune is observer on WLCG Management Board
- DUNE representative on Grid Deployment Board (GDB) - many members of the GDB who are on both DUNE and LHC experiment
- FNAL hosted Sept 2019 GDB meeting, Authorization Workshop, and FIM4R Workshop
- Proposed evolution would have DUNE as a member of Scientific Computing Infrastructure Steering Committee

FIM4R **Federated Identity Management**
For Research

new experience
Tornado Warning

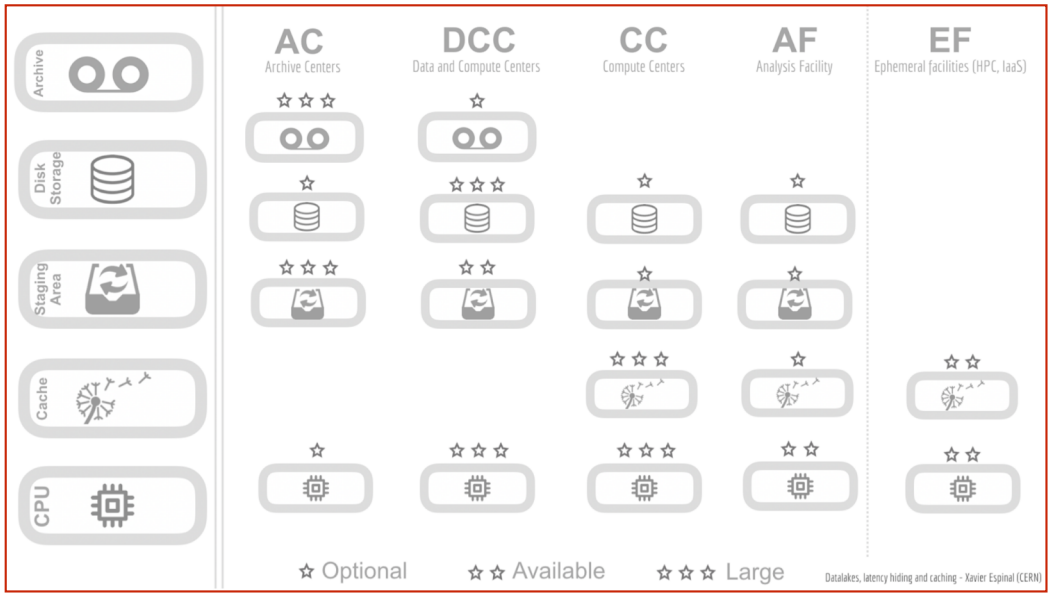


Ian Bird's diagram

Fermilab

DUNE Computing Model Workshop (FNAL Sep 9-12, 2019)

- Develop computing requirements based upon workflows matching data model
- Map those requirements onto a new model for CPU, storage, services, and infrastructure
- Exploring R&D for new storage models and support levels for centers and sites
  - based upon the **HSF DOMA** working group proposal
  - sites can be federated across institutions or countries
  - recruit additional resources by de-emphasis of 24x7 support level at DDC, CC, and AF
  - replication of data (and services to support fall over) may show lower costs



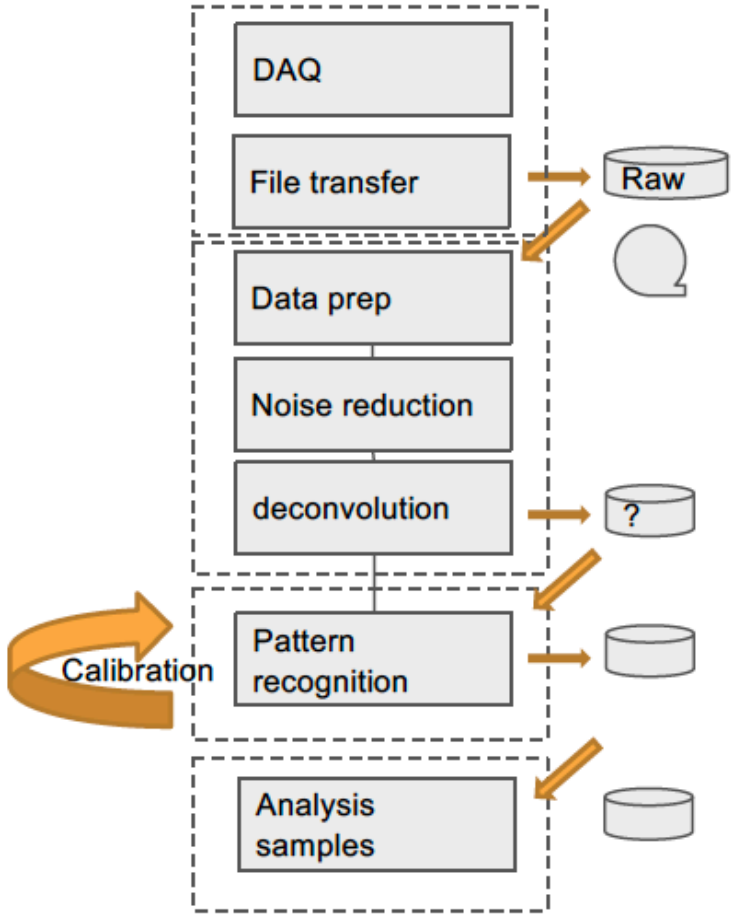
Data Access in DOMA, HSF/OSG/WLCG Joint Workshop J-LAB Newport News, VA 19-23 March 2019

Goal: Minimize cost while maximizing capability through redundancy

Fermilab

Strawman Computing Model and Data Flow

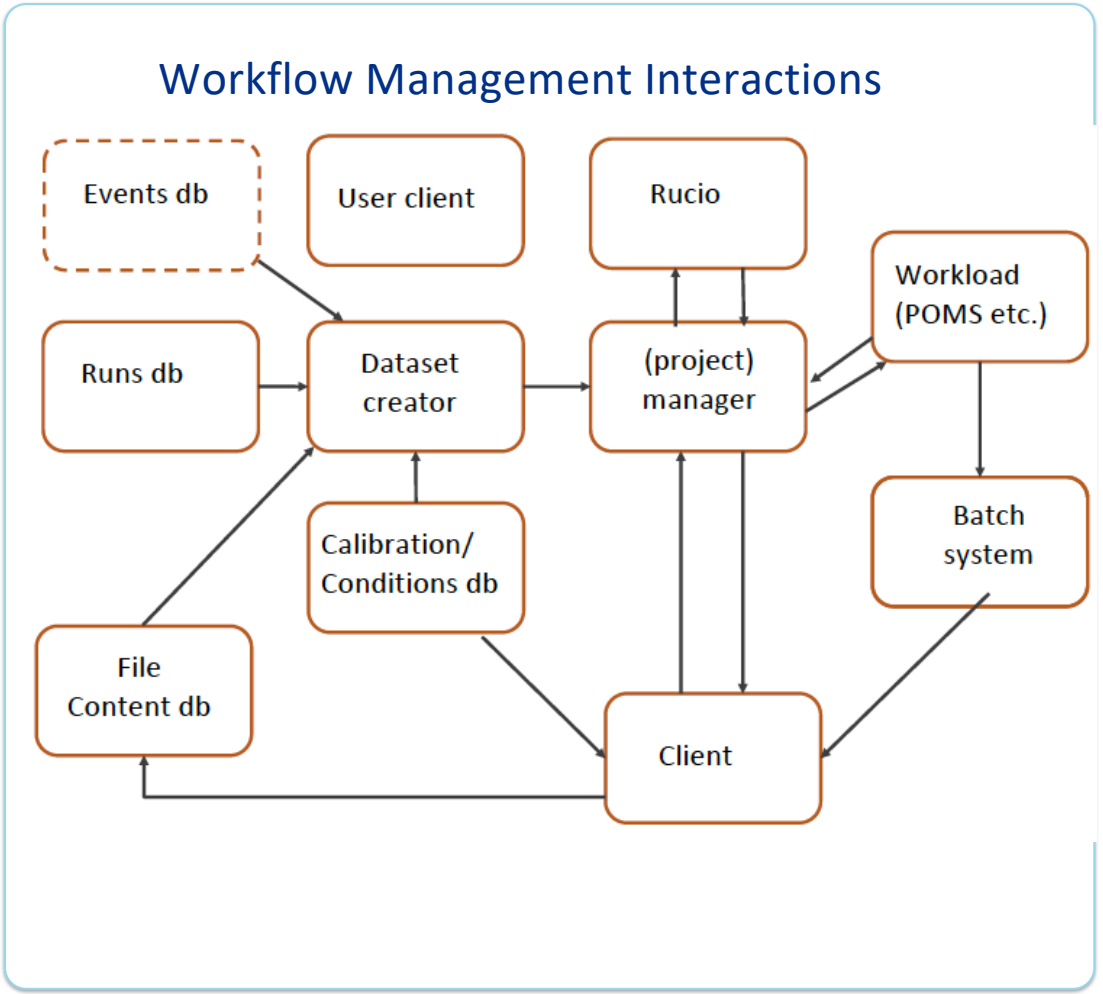
- Discuss the resource needs of production and analysis workflows for the DUNE offline
  - Description of the workflow mostly in terms of storage, bandwidth, and processing time
  - Also need to use these workflows to define the networking needs
  - Lifetime of output datasets become important
- Use the workflows to provide input into the computing model and develop resource requests for overall DUNE Computing



Fermilab

Basic model assumptions

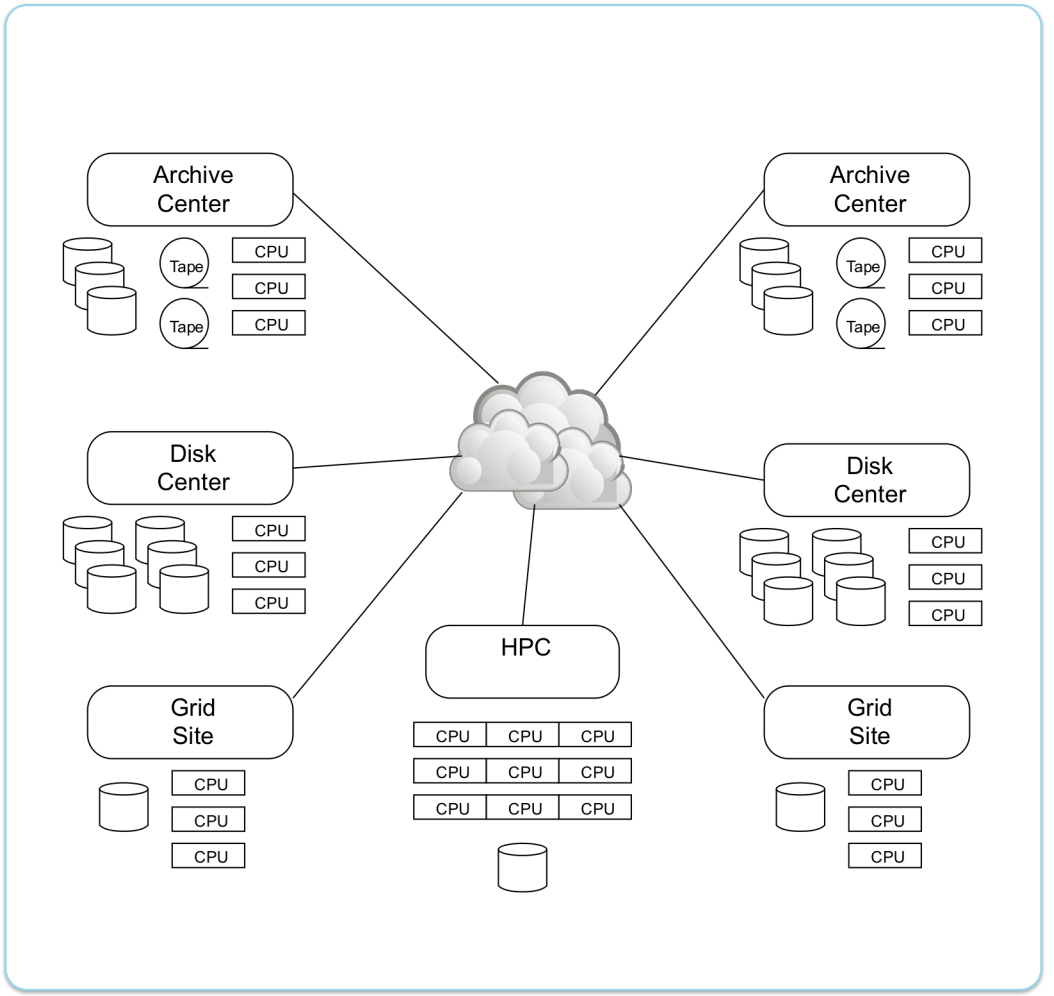
- Adapt standard grid and workflow tools to match data and compute resources
  - Current: POMS (GlidenWMS based)
  - Exploration w/ DIRAC
- Develop replacement for SAM components providing data metadata cataloging and dataset ident.
- Rucio controls file placement and replication across location
- Develop replacement for SAM “Project” manager that tracks real-time file delivery and status (bookkeeping)



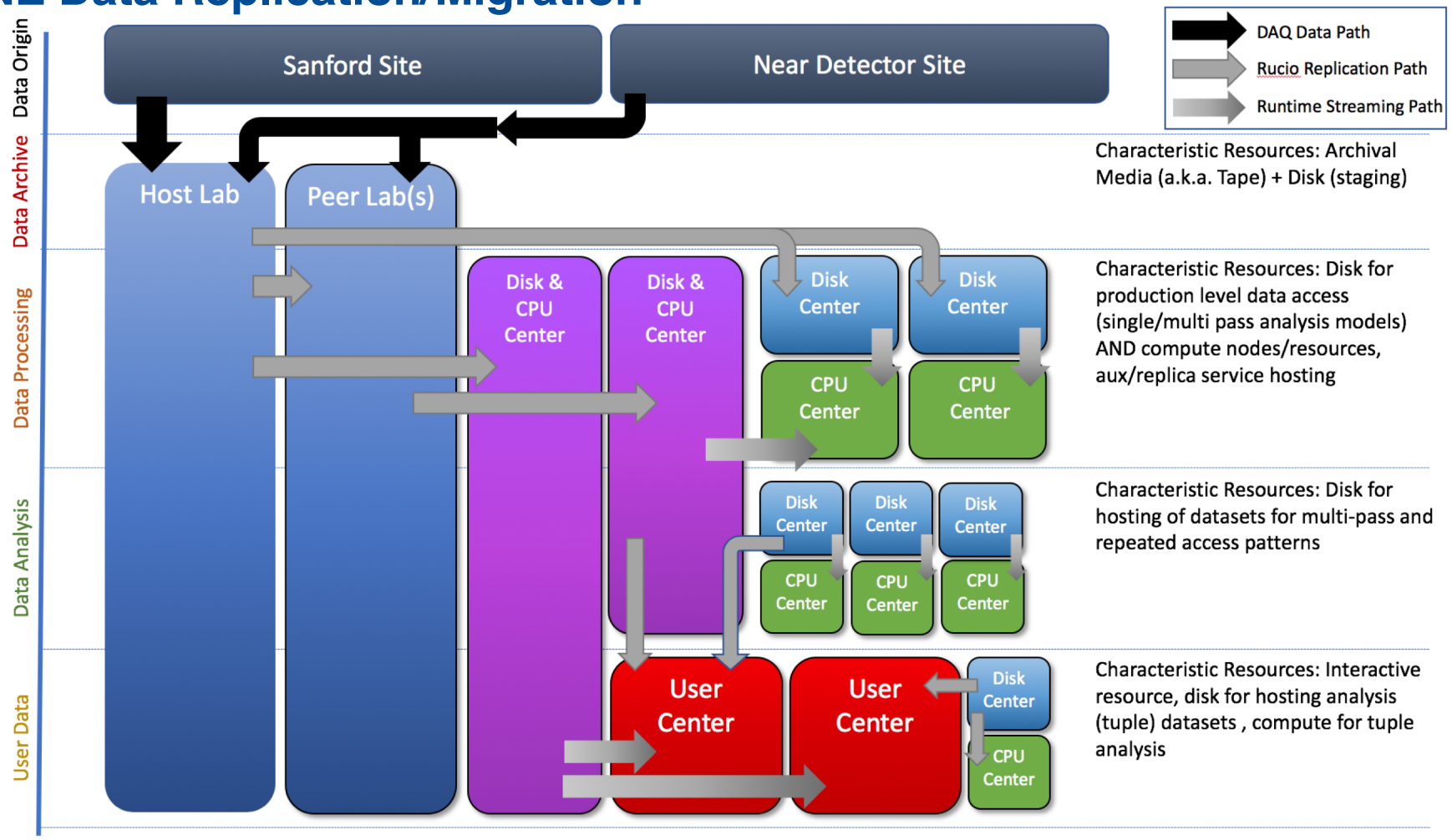
Distributed model

- Collaborating institutions provide significant disk resources (1PB chunks)
- Rucio places multiple copies of datasets
- Site matching takes into account CPU characteristics and data location
- Sites pull from availability regions

Is SAM-style late binding viable alternative to data pre-placement in a high-speed WAN network world?



DUNE Data Replication/Migration



Proposal for a "Computing Contribution Board"

- Caveat: discussion document has been presented to the Spokespeople and circulation to countries/IB reps has been happening the last two weeks - not finalized
- Organize and oversee computing contributions by DUNE partners - can be federated countries, etc (e.g. BeNeLux as an “economic” analogy)
- Defines resources as:
  - CPU resources for simulation, reconstruction and analysis of DUNE data
  - Storage resources dedicated to DUNE data
  - Professional personnel for development and maintenance of core computing infrastructure
  - Dedicated personnel to manage and operate DUNE computing
- physical computing resource provision being a collective responsibility shared between collaborating countries subject to some form of “fair share” expectation
- DUNE does not seek to formally oblige partners/nations to contribute at this stage. For now, the matter is left as a point of good citizenship

Computing Contribution Board Membership

- Peter Clarke has agree to lead the CCB and lead drafting of governance discussion document
- each DUNE partner would have a representative on the CCB
  - members are national liaisons in a position able to interact effectively with their respective funding agencies
- Computing Consortium would annually establish a computing resource request based upon the computing model and present to CCB
- Start of what could be a long process to negotiate with federations (institutions, countries, etc) about how to account for resources
- A technical, but important step in this direction, will change operation to use DUNE VO (outside the Fermi VO umbrella) shortly
- Allow for WLCG EGI accounting through APEL, and therefore matching of pledges to deliver of resources

Institution	Country
CBFP	Brazil
Unicamp	Brazil
CERN	CERN
FZU	Czech Republic
CCIN2P3	France
Indian groups	India
KISTI	Korea
Nikhef	Netherlands
Bern	Switzerland
CIEMAT	Spain
Edinburgh	UK
GridPP	UK
Manchester	UK
RAL/STFC	UK
Argonne	USA
Berkeley	USA
BNL	USA
Colorado State	USA
CU Boulder	USA
Fermilab	USA
Florida	USA
LBNL	USA
MInnesota	USA
Northern Illinois Univ.	USA
Notre Dame	USA
Oregon State University	USA
SLAC	USA
Texas, Austin	USA

Coming soon, QMUL, UK, York, CA

Computing Resources Review & Scrutiny

Note:
Transitioning (Fall '19) from FNAL Scientific Portfolio Management (SCPMT) process to a new Computing Resource board to review requests in a scientific context (CRSG Board)

- We need external input to ensure that our decisions make sense in a global context and in the context of available resources across the science programs.
  - Currently we rely on:
    - Long Baseline Neutrino Committee (LBNC)
    - International Computing Advisory Committee (ICAC)

The Long Baseline Neutrino Committee

International Computing Advisory Committee

Fermilab

Summary

- Demonstrated Success w/ ProtoDUNE-SP and ProtoDUNE-DP active processing campaigns
  - successfully processing data and efficiently utilizing international computing resources
  - reprocessing of ProtonDUNE-SP Run I data almost complete
  - successful data management and online DQM of ProtoDUNE-DP - offline to follow shortly
- Successful Event Data and Computing Model workshops in Aug and Sep
  - anticipate full development of the computing model by end of FY2020 (US)
  - Computing CDR (and TDR?) delivery anticipated on similar timeline
  - deployment of computing model ideally established for ProtoDUNE SP & DP beam operations following Long Shutdown 2 at CERN (late 2021 and 2022)
- Actively engaged in WLCG and HSF to help define the future HEP computing landscape
- Proposed a model for computing contributions and establishment of Computing Contribution Board to match requests to international resource contributions
- Proposed Fermilab CRSG to review annual requests for computing resources from the Event Data and Computing Models

Concerns

- User Support for a large collaboration
  (sustainability w/ limited consortium personnel)

- Upcoming draft documents requiring expert outside review
  (and how to find/fund this work)

- Planning activities and enabling participation
  (i.e. funding participation from Univ and Lab people to take part in workshops)

- Technology demonstrators and resources for R&D