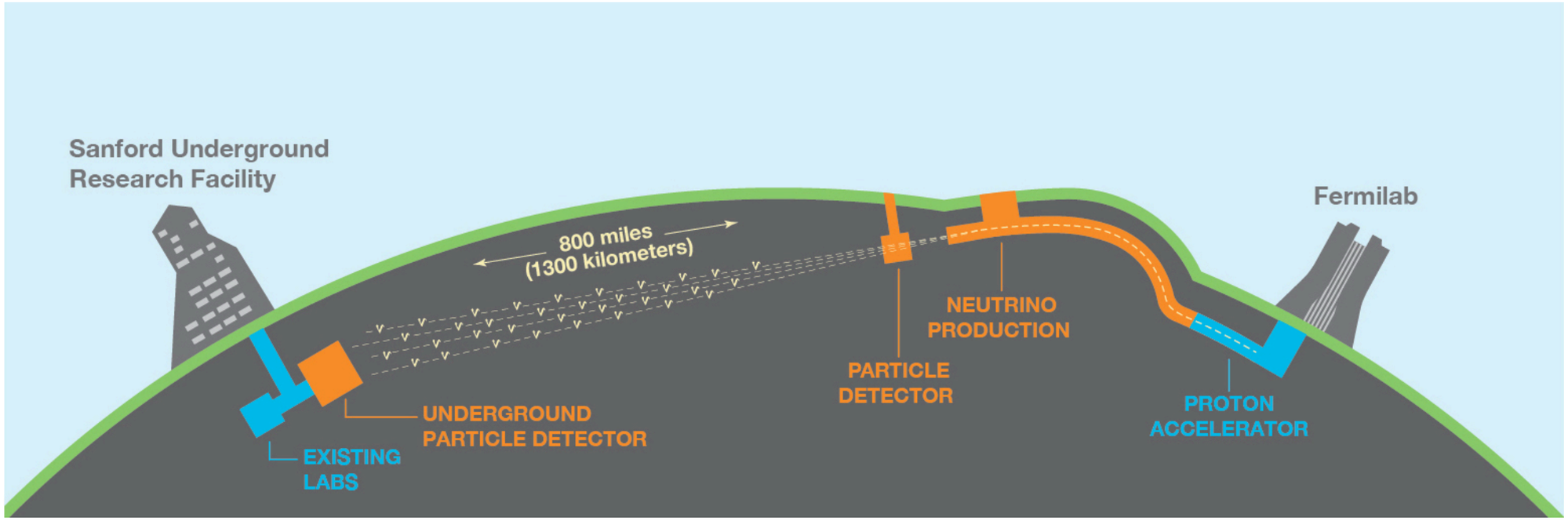
Status of DUNE Offline Computing

Mike Kirby on behalf of DUNE Collaboration

26TH International Conference on Computing in High Energy & Nuclear Physics

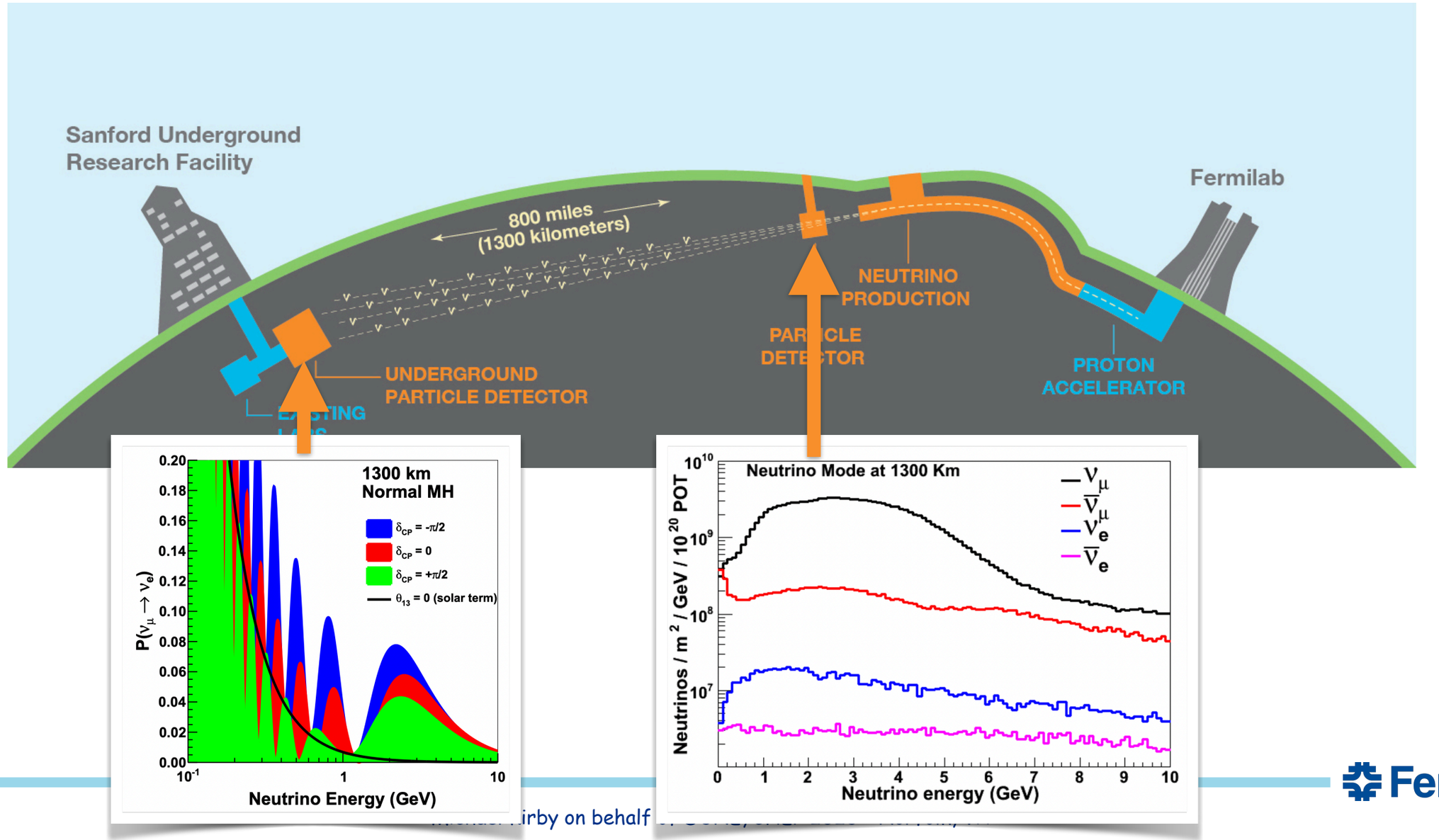
May 9, 2023 - Norfolk, VA

Quick reminder about DUNE



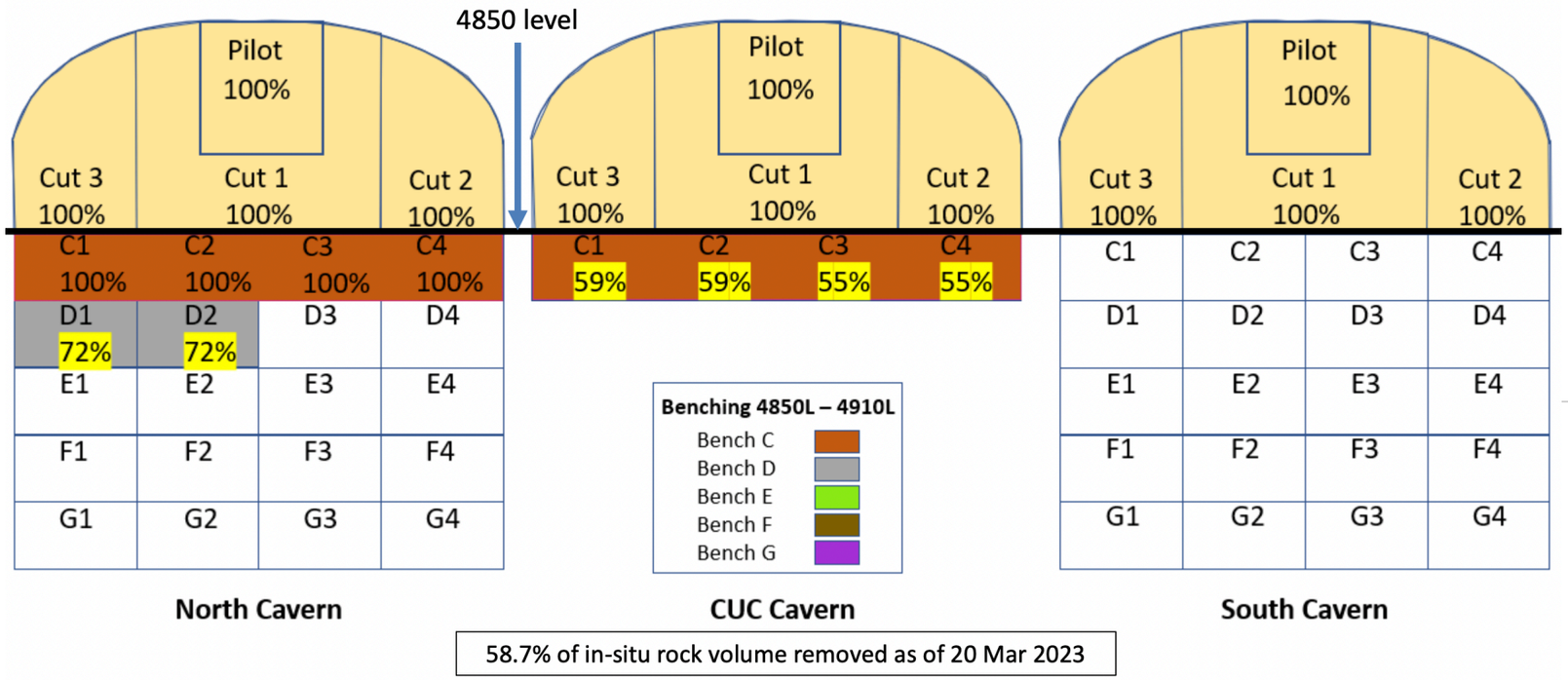
- neutrino experiment studying neutrino oscillation parameter (mass ordering, matter vs antimatter asymmetry, unitarity), proton decay, supernova neutrinos, and more.
- four very large LAr TPC (17 kT) at 4850 ft underground in Lead, SD (Homestake Mine)
- near detector onsite at Fermilab being designed (3 sub-detectors, two that move)
- two prototypes at CERN - (ProtoDUNE II Horizontal Drift - ProtoDUNE II Vertical Drift)

Optimized location of the Far Detector

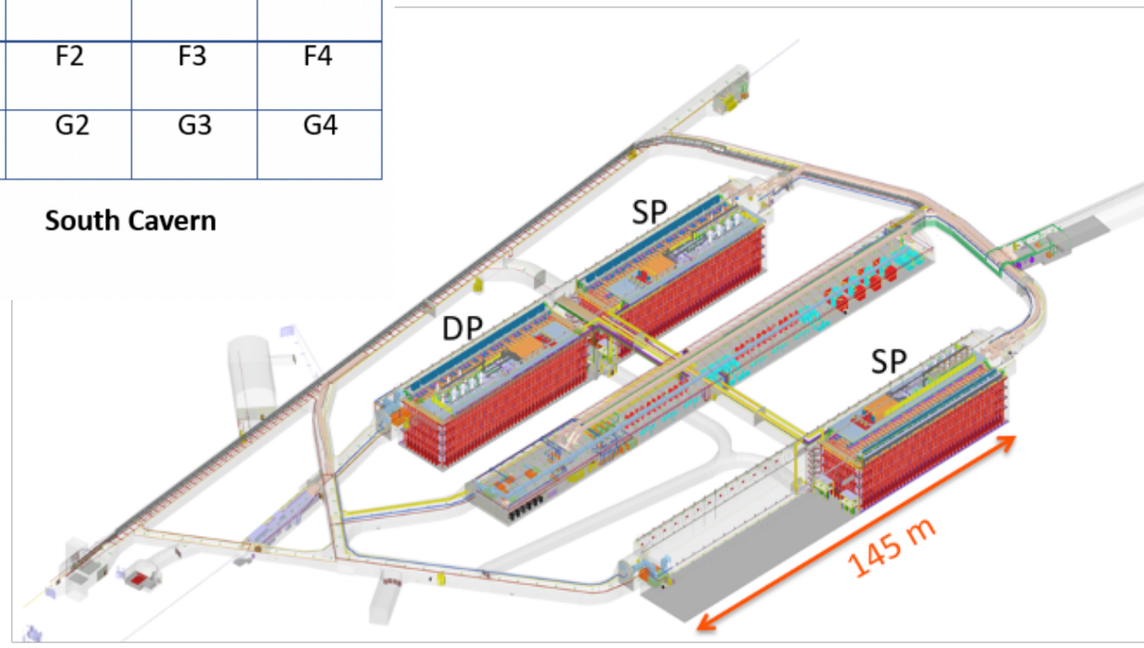


Excavation continues…

Slide: C. Mossey



58.7% of in-situ rock volume removed as of 20 Mar 2023



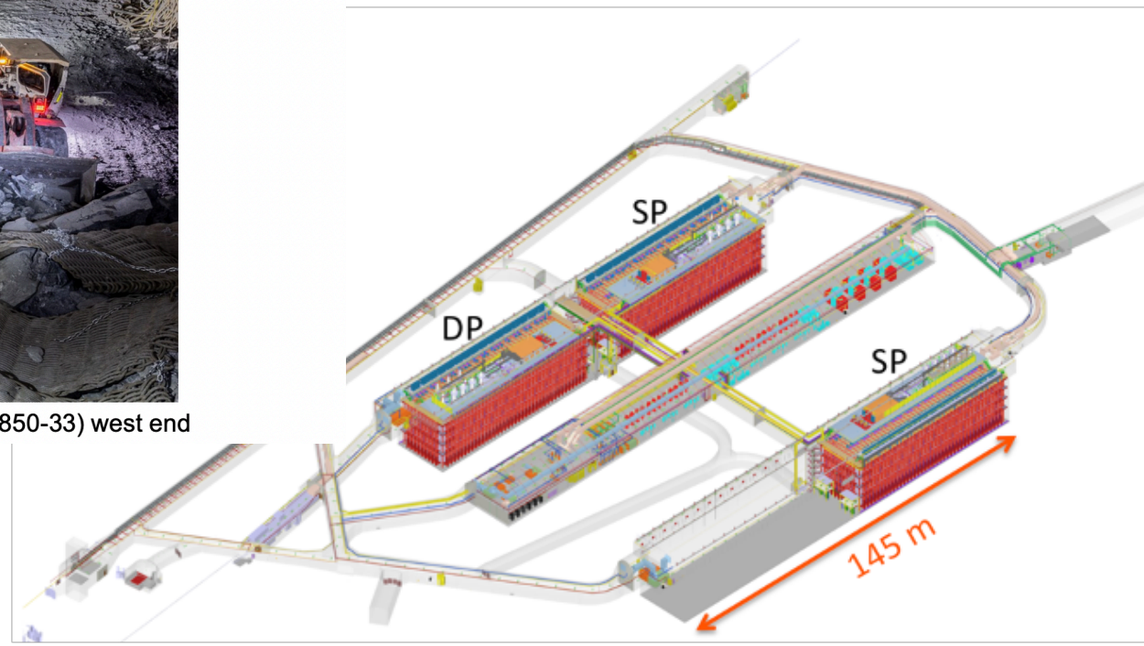
Excavation continues…

Slide: C. Mossey

North Detector Cavern – West End

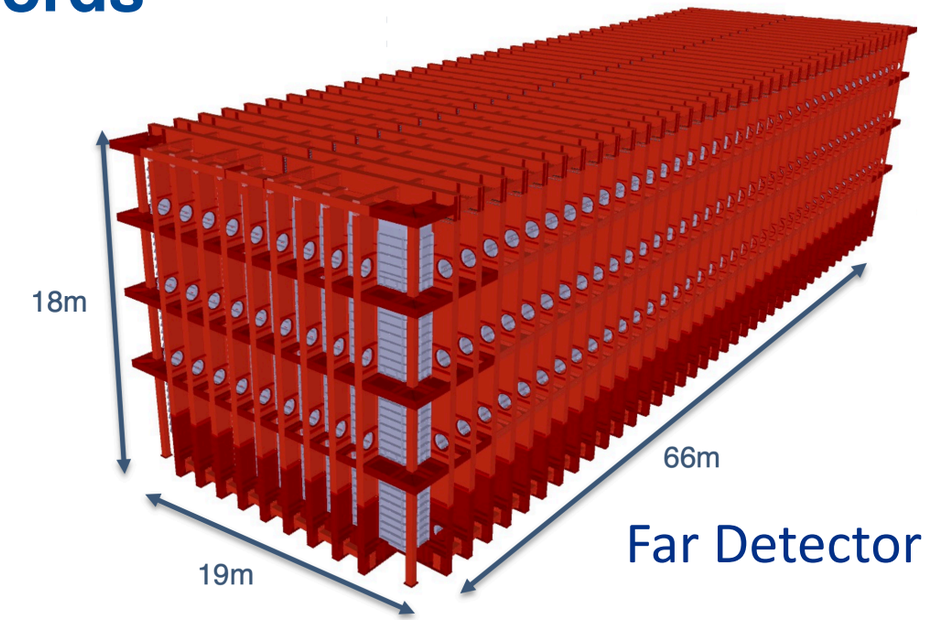
Photo by Matt Kapust, SDSTA; 19 Jan 2023

Drilling holes for blast charges for bench C (left) and removing muck (right) in North Detector Cavern (4850-33) west end



Far Detector Dataflow and Trigger Records

- beam coincidence events are extremely important, but of limited total volume
  - ~1 Hz beam rate
  - active online trigger in development
  - Region-of-Interest within module
  - online compression and zero-suppression being considered
- solar neutrino triggered events
- cosmic ray events and calibrations
- supernova readout events
  - ~140 TB in 100 seconds - one FD module
  - work w/ trigger primitives for immediate optical follow up
  - transfer out 4 hours and process in 4 hours for precision optical observations
- DUNE requirement - less than 30 PB/year total to permanent storage from all active FDs



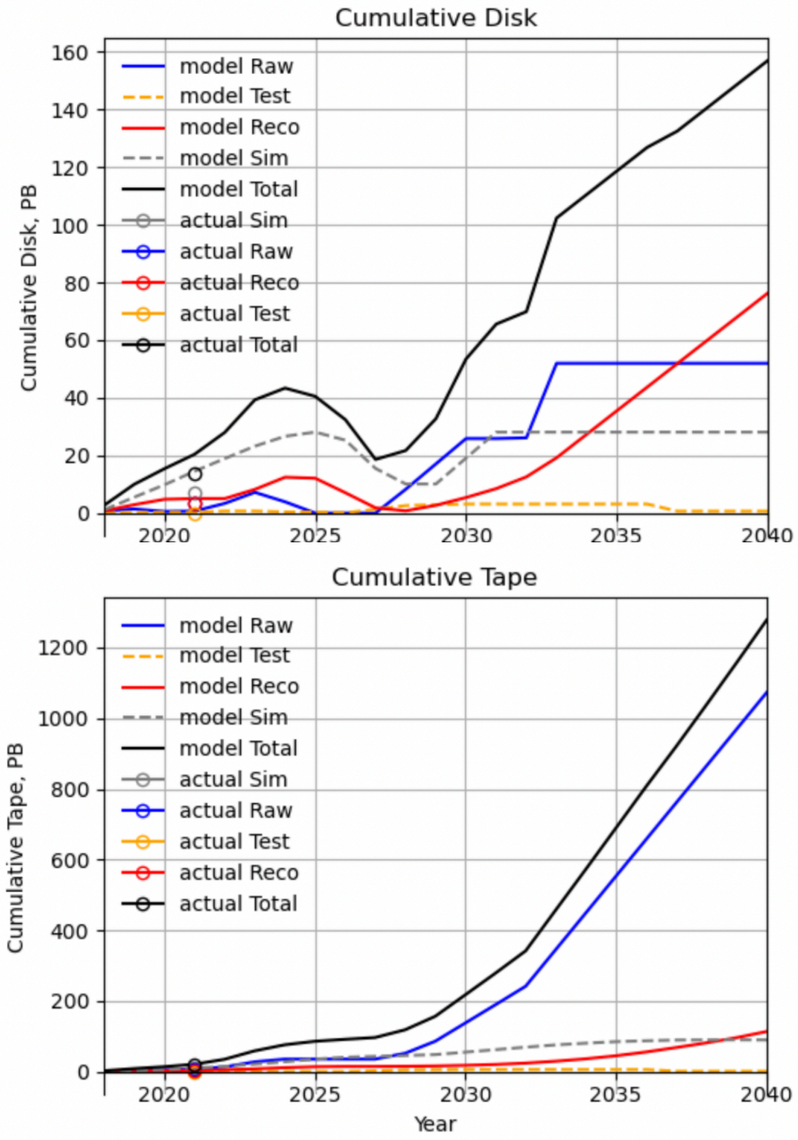
Far Detector Cryostat

Process	Rate/module	size/instance	size/module/year
Beam event	41/day	3.8 GB	30 TB/year
Cosmic rays	4,500/day	3.8 GB	6.2 PB/year
Supernova trigger	1/month	140 TB	1.7 PB/year
Solar neutrinos	10,000/year	$\leq$3.8 GB	35 TB/year
Calibrations	2/year	750 TB	1.5 PB/year
Total			9.4 PB/year

Recently published DUNE Computing CDR - https://arxiv.org/abs/2210.15665

Data Placement Strategy

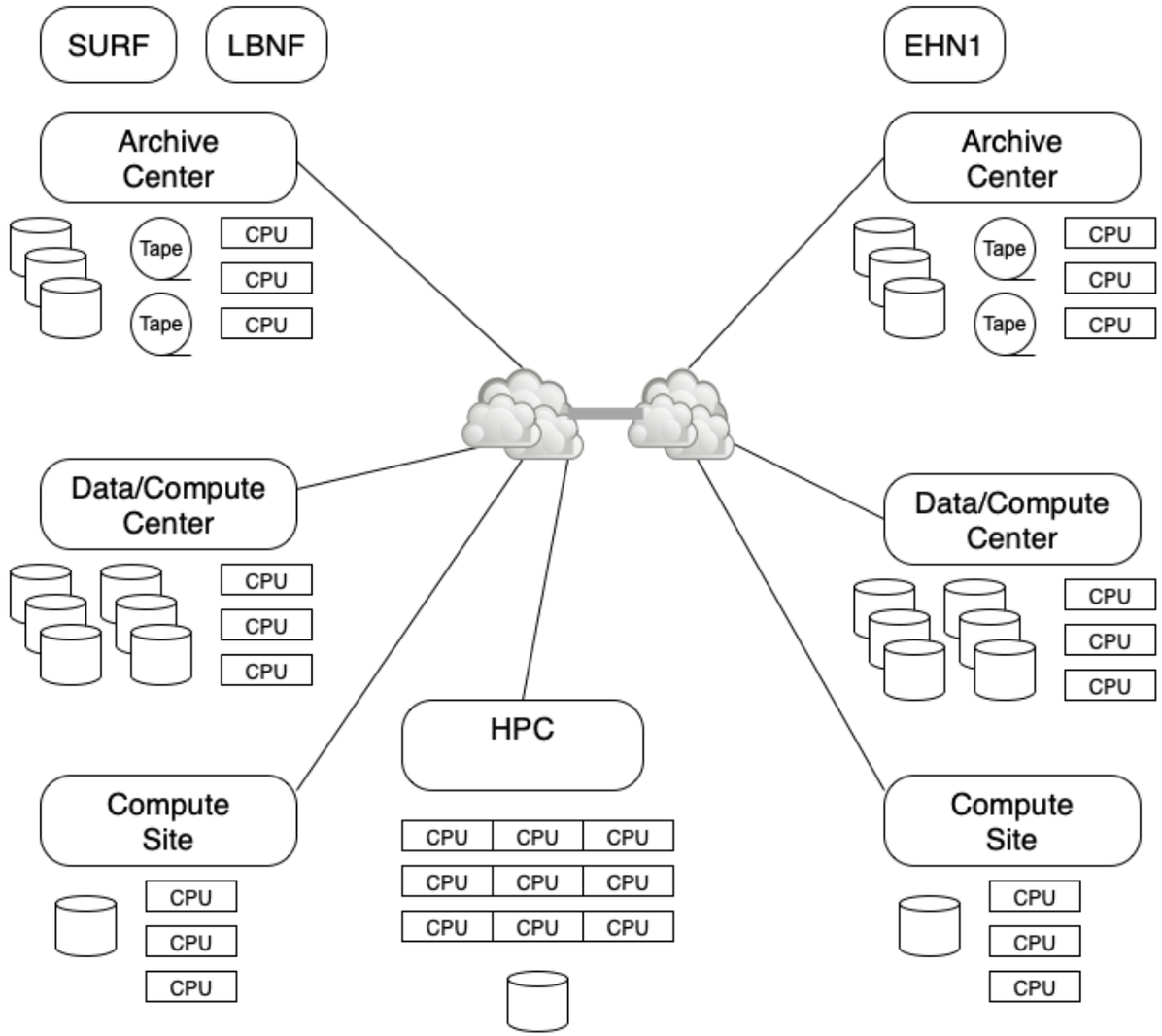
- accomplished with Rucio and FTS3 - Scale tests of new DUNE data pipeline - S. Timm
- 2 copies of raw data on tape
  - one copy on each side of an ocean
  - 6 months on disk
- 1 replica of reco/sim on tape
  - distribute across global Rucio SEs
  - annual reco pass over all data
  - annual sim campaign to match
  - production resident on disk for 2 years
- Assume 2 disk copies of reco and sim
  - impose shorter lifetimes on tests & sim stages
  - R&D exploring data tiers and formats
- DUNE HDF5 Experience - B. Chowdhury



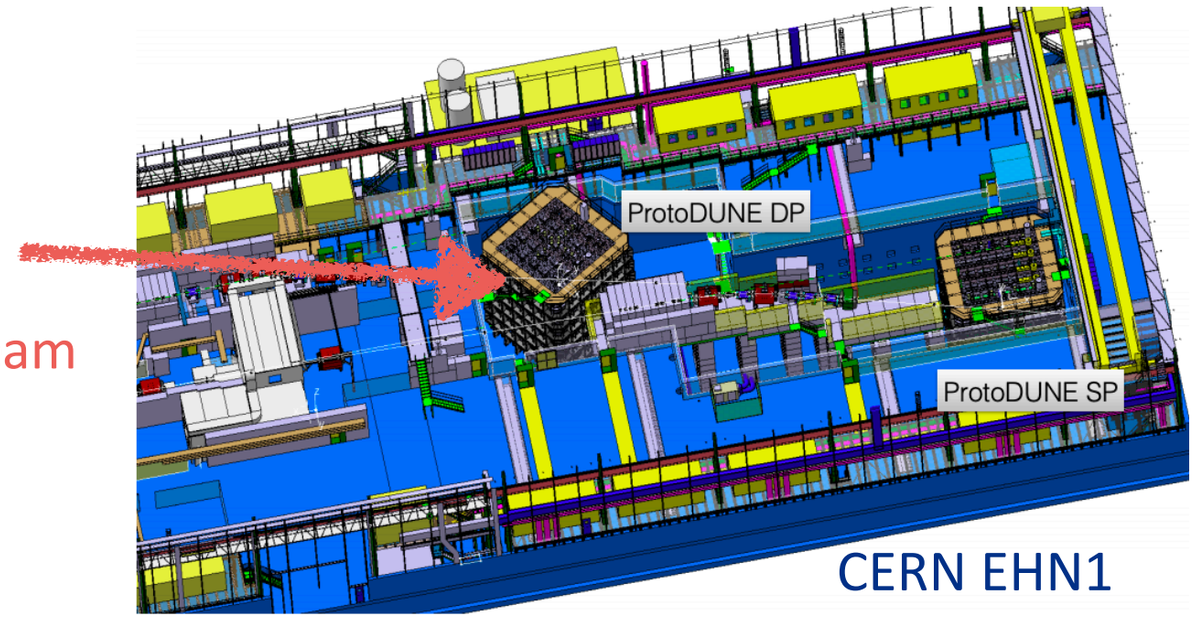
DUNE Computing CDR

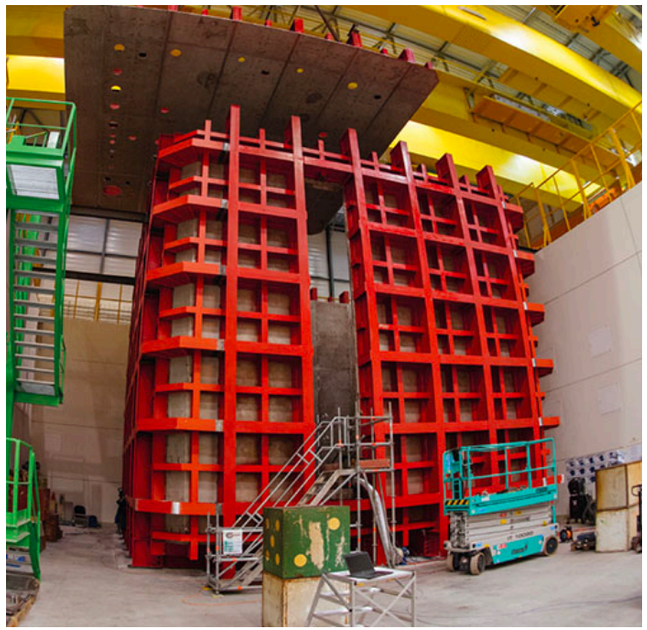
DUNE Computing Resource Model

- less “tiered” than current WLCG model —> flatter model proposed by HSF DOMA working group
  - take advantage of existing WLCG sites that can add DUNE access
  - require reasonable minimum size - storage elements
  - allow for CPU only sites with data streaming
- collaborating institutions (or groups of institutions) provide significant disk resources (~1PB chunks)
- plan to use common tools for most services
- participation in the HSF process important to provide and integrate new solutions



CERN and Neutrino Platform currently hosting ProtoDUNE

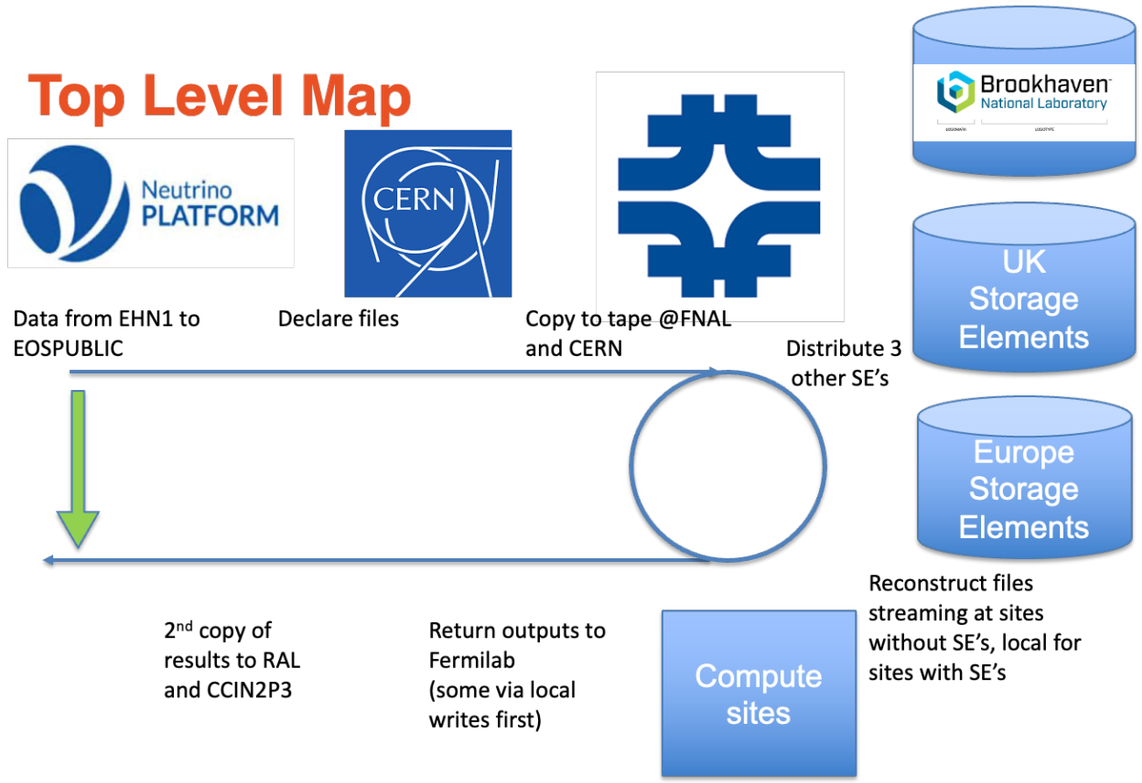




- ProtoDUNE Single Phase and Dual Phase
  - constructed and operated during the timeframe of 2018 - 2020
  - ProtoDUNE SP took 6 weeks of beam (~25 Hz)
  - invaluable information about performance, construction, and operations
- ProtoDUNE II currently under construction
  - Horizontal Drift and Vertical Drift
  - HD anticipate LAr filling 2023Q4
  - VD assembly underway for 2024 operations
  - Beam operations in 2023/2024
  - DUNE Database Development - L. Gerlach & A.P. Vizcaya Hernandez

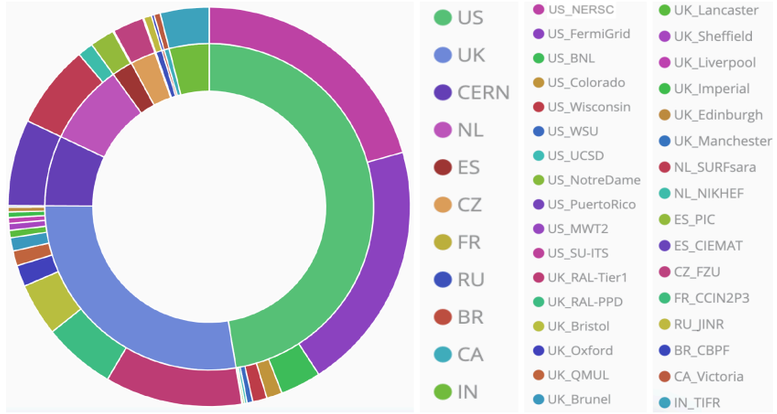
Summer 2022 Data Challenge 4 - ProtoDUNE

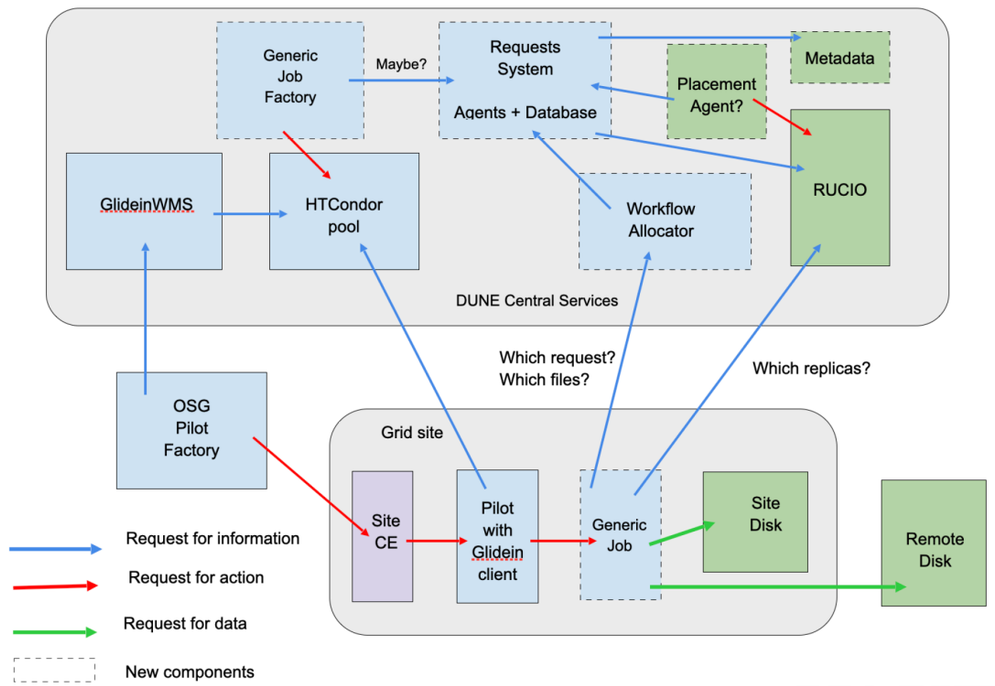
- Goals of the Data Challenge 4 - test all the services and procedures that will be used in the forthcoming beam runs of PD-HD and PD-VD
- Phase 1 - Data Pipeline
  - Goal - test data path EHN1->CERN->FNAL
  - transfer, declare, and replicate “raw data” at needed scale
  - 3.6 GBytes/s achieved across atlantic
- Phase 2 - Data Processing
  - Goal - sustain 5000 concurrent jobs for keep up processing
  - significant drop in CPU efficiency for jobs where large input data files not located “near” job
  - The Workflow System (now “justIN”)
  - The Data Dispatcher



Summer 2022 Data Challenge 4 - ProtoDUNE

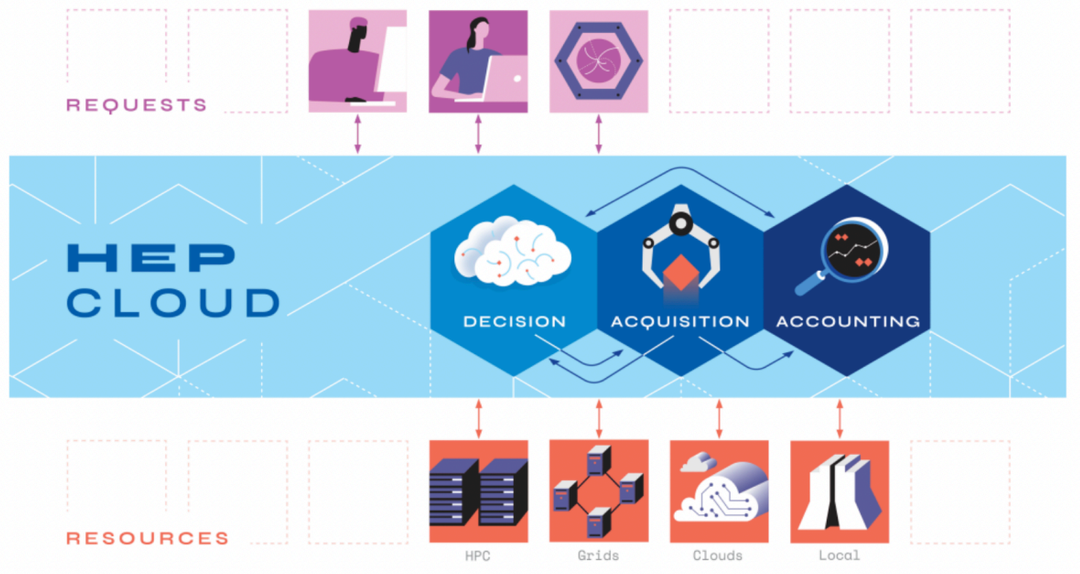
- Goals of the Data Challenge 4 - test all the services and procedures that will be used in the forthcoming beam runs of PD-HD and PD-VD
- Phase 1 - Data Pipeline
  - Goal - test data path EHN1->CERN->FNAL
  - transfer, declare, and replicate “raw data” at needed scale
  - 3.6 GBytes/s achieved across atlantic
- Phase 2 - Data Processing
  - Goal - sustain 5000 concurrent jobs for keep up processing
  - significant drop in CPU efficiency for jobs where large input data files not located “near” job
  - The Workflow System (now “justIN”)
  - The Data Dispatcher





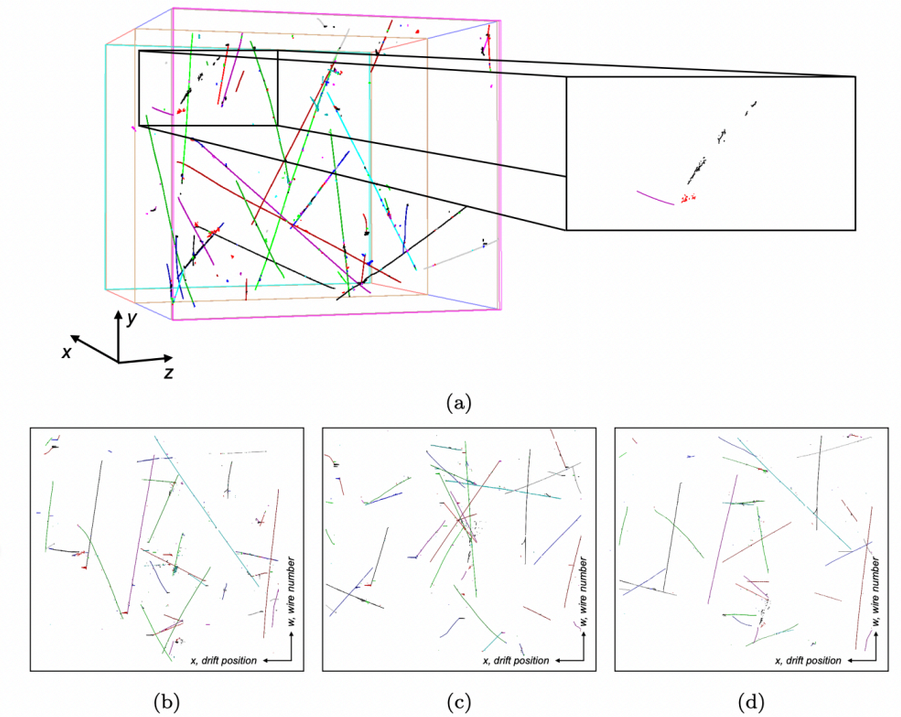
DUNE Software and Algorithmic Developments

- DUNE will take advantage of variety of architectures and algorithms
  - actively utilized resources at NERSC facilities for simulation and reconstruction of data
  - HEPCloud providing centralized submission point for production
- Neutrino interaction vertex-finding in a DUNE far-detector using Pandora deep-learning - A. Chappell
- Differentiable Simulation of the DUNE Near Detector Liquid Argon Time Projection Chamber - S. Gasiorowski
- Using parallel I/O libraries for managing HEP experimental data - A. Bashyal



DUNE Software and Algorithmic Developments

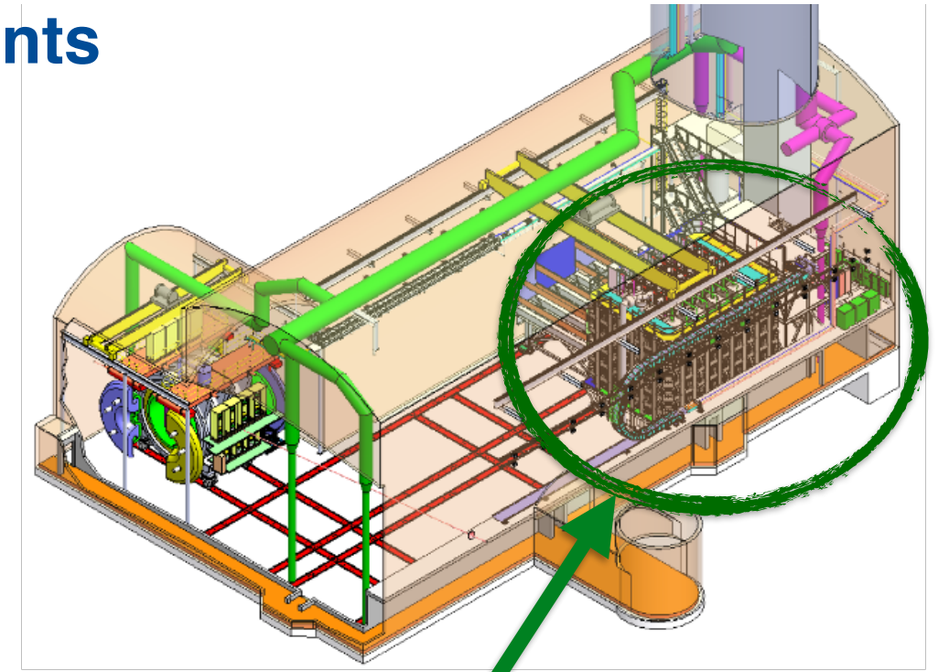
- DUNE will take advantage of variety of architectures and algorithms
  - actively utilized resources at NERSC facilities for simulation and reconstruction of data
  - HEPCloud providing centralized submission point for production
- Neutrino interaction vertex-finding in a DUNE far-detector using Pandora deep-learning - A. Chappell
- Differentiable Simulation of the DUNE Near Detector Liquid Argon Time Projection Chamber - S. Gasiorowski
- Using parallel I/O libraries for managing HEP experimental data - A. Bashyal



Pandora Reconstruction of ProtoDUNE Data
arXiv:2206.14521

DUNE Software and Algorithmic Developments

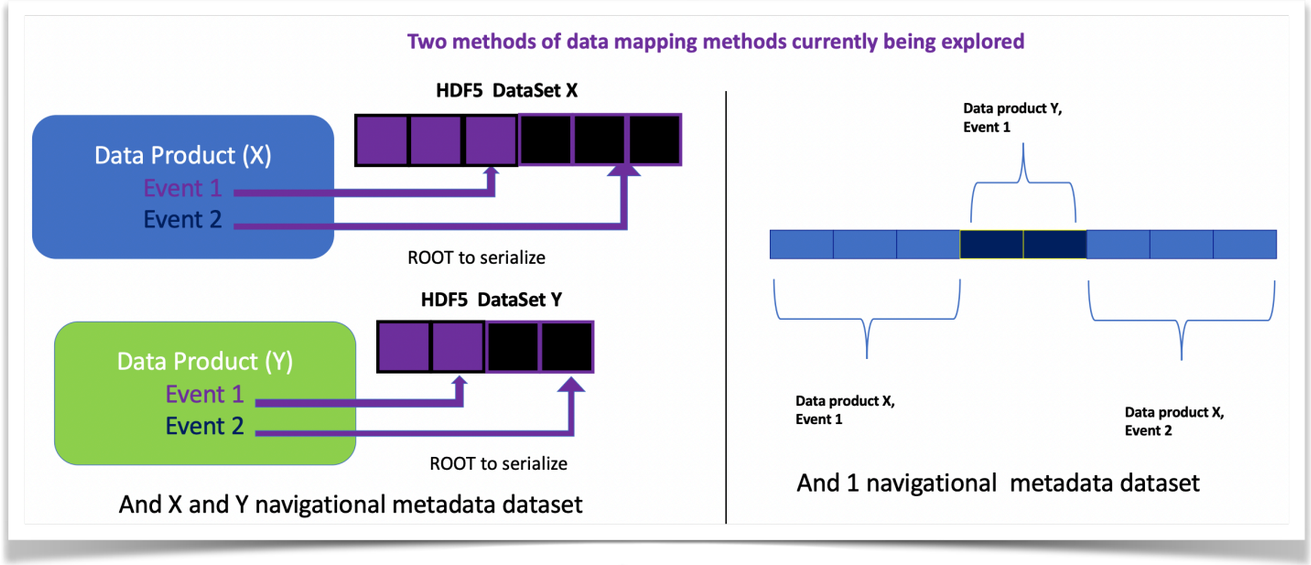
- DUNE will take advantage of variety of architectures and algorithms
  - actively utilized resources at NERSC facilities for simulation and reconstruction of data
  - HEPCloud providing centralized submission point for production
- Neutrino interaction vertex-finding in a DUNE far-detector using Pandora deep-learning - A. Chappell
- Differentiable Simulation of the DUNE Near Detector Liquid Argon Time Projection Chamber - S. Gasiorowski
- Using parallel I/O libraries for managing HEP experimental data - A. Bashyal



GPU based simulation of ND - LAr
2023 JINST 18 P04034

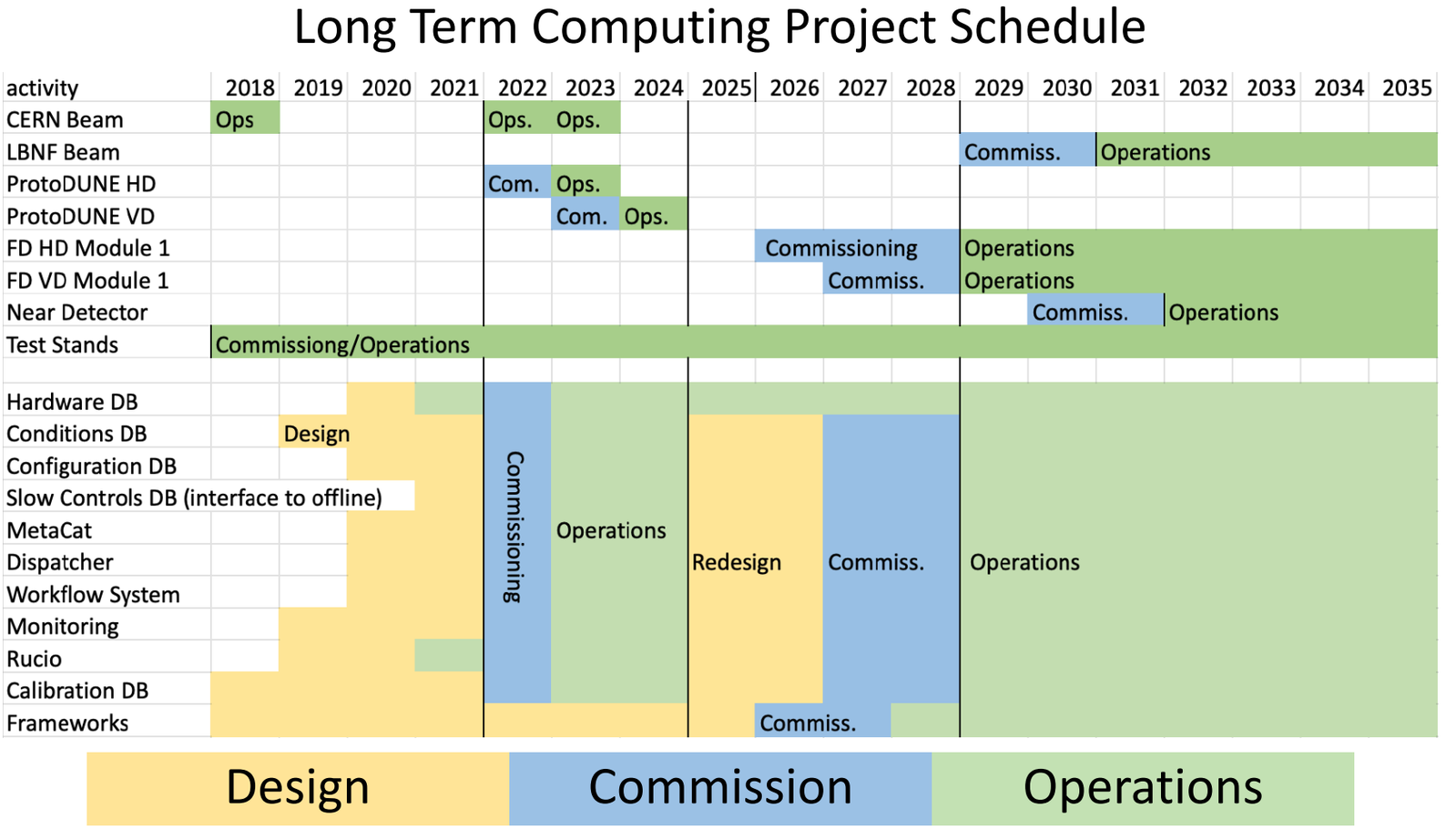
DUNE Software and Algorithmic Developments

- DUNE will take advantage of variety of architectures and algorithms
  - actively utilized resources at NERSC facilities for simulation and reconstruction of data
  - HEPCloud providing centralized submission point for production
- Neutrino interaction vertex-finding in a DUNE far-detector using Pandora deep-learning - A. Chappell
- Differentiable Simulation of the DUNE Near Detector Liquid Argon Time Projection Chamber - S. Gasiorowski
- Using parallel I/O libraries for managing HEP experimental data - A. Bashyal



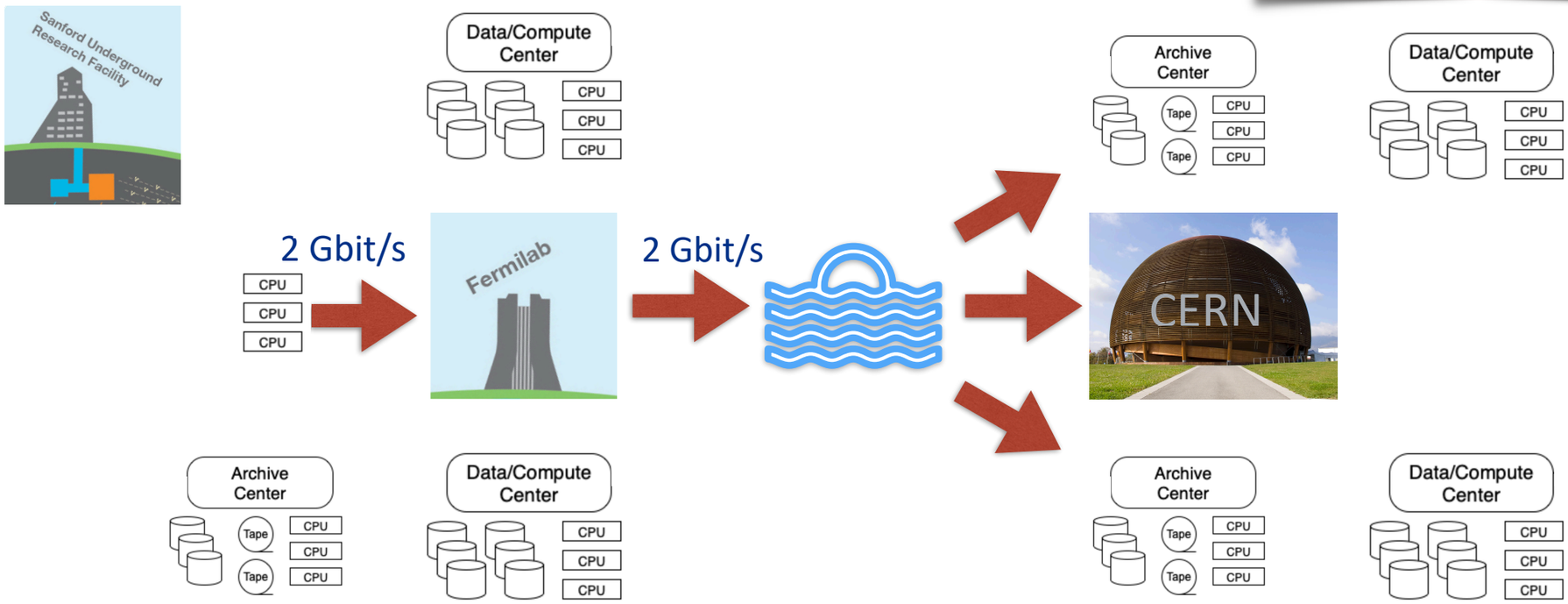
tentative DUNE future timeline

- **Fall 2023** - operations of ProtoDUNE HD
- **Spring 2024** - operations of ProtoDUNE VD
- **2023 - 2024** - DUNE computing operations at scale with PD II data
- **FD HD Module 1**
  - 2027 construction
  - 2028 commissioning
  - 2029 physics
  - FD VD Module - 1 year offset
- **2025-2027** - use this time for development addressing unique DUNE Challenges



DUNE Involvement in WLCG Data Challenge 24

"FD" Raw Data
archival storage



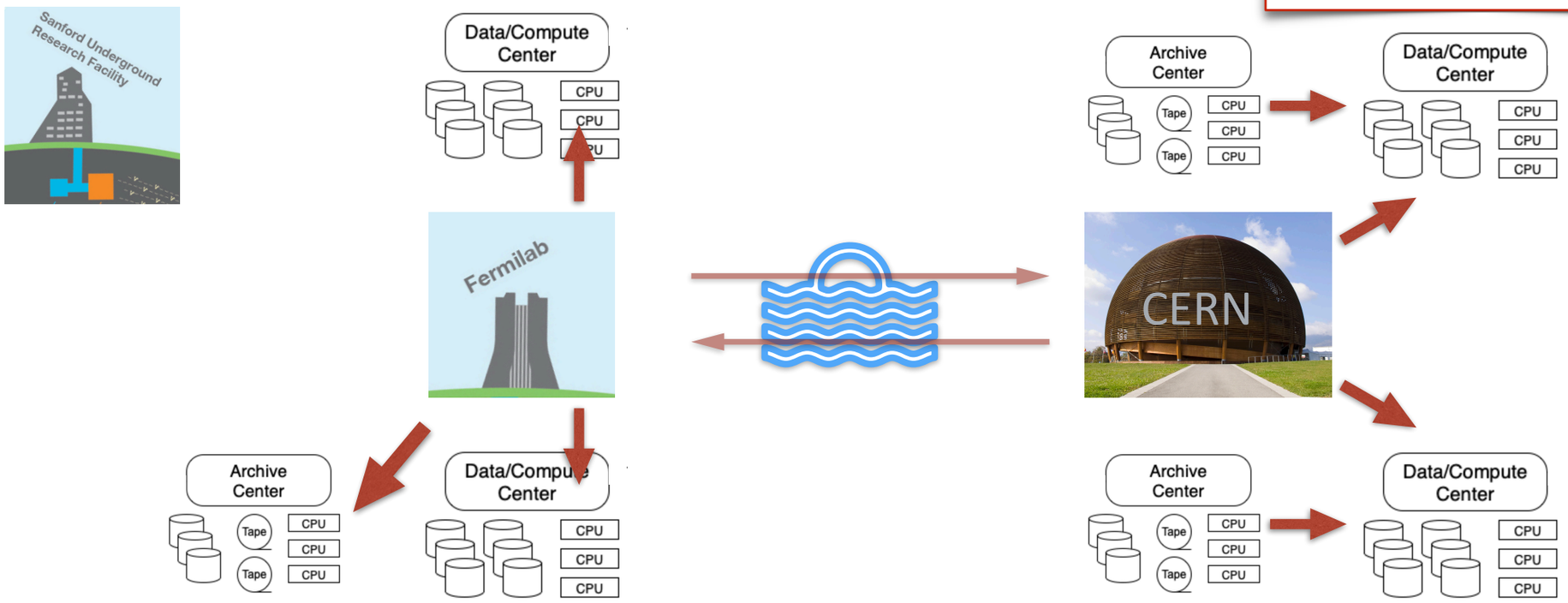
not to scale, not a technical design, it's just a cartoon

- Simulate the archival of 25% of the raw data rate from the Far Detector
  - translates to 2 Gbit/s from SURF to FNAL
  - replicate that "FD" raw data to archival storage facilities around the world
  - replicate the "FD" raw data to disk storage elements around the world for prompt access from compute elements
- Both job submission and RSE to RSE w/ token authentication/authorization

DUNE Involvement in WLCG Data Challenge 24

"FD" Raw Data

raw processing

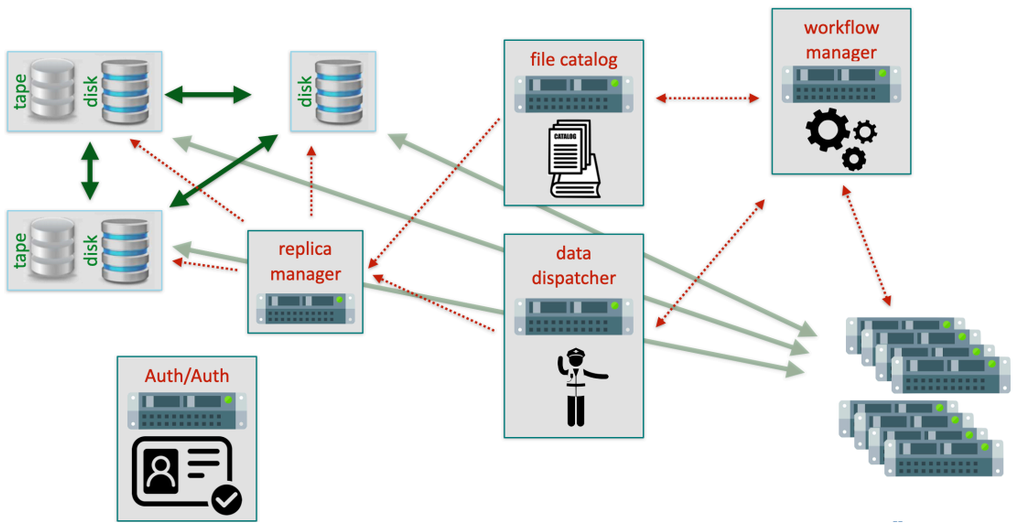


not to scale, not a technical design it's just a cartoon

- Maintain continuous processing workload at distributed sites commensurate with 25% "FD" raw data rate
  - utilize compute elements across the WLCG and OSG
  - match the locality of jobs with locality of data at nearby RSEs
- Both job submission and RSE to RSE w/ token authentication/authorization

Summary

- DUNE successfully utilizing WLCG resources
- Looking forward to addressing challenges
  - more complete Rucio integration
  - developing new workflows and workflow management - including access to HPC
  - integrate GPU software and hardware for processing - data prep especially
- looking forward to improved understanding that will come from involvement in DC24 and ProtoDUNE II operations
- exploring ideas for analysis centers
- This research is based upon work supported by the US Department of Energy, Office of Science, Office of High Energy Physics.

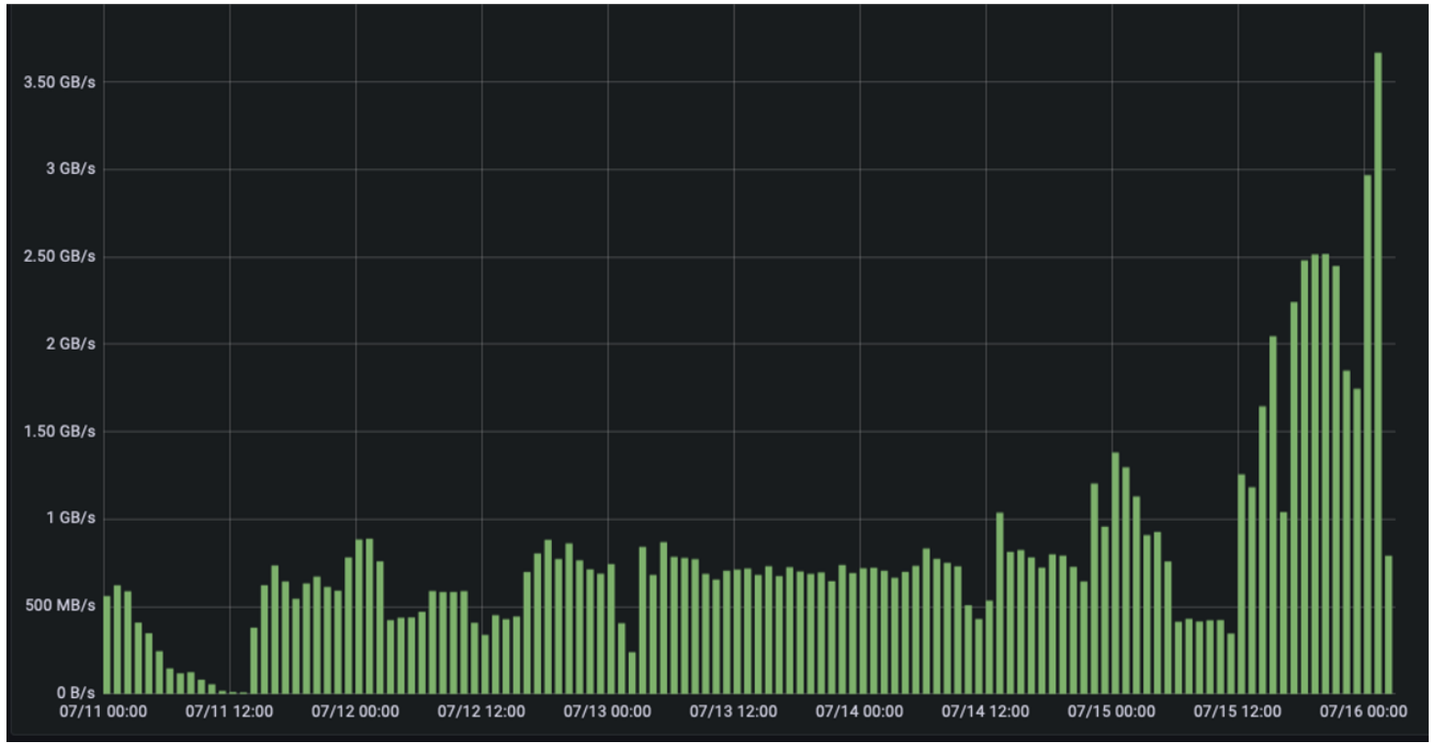


Thank you to the organizing committee

Backup

Data Challenge 4 - Phase 1 - Data Pipeline for ProtoDUNE

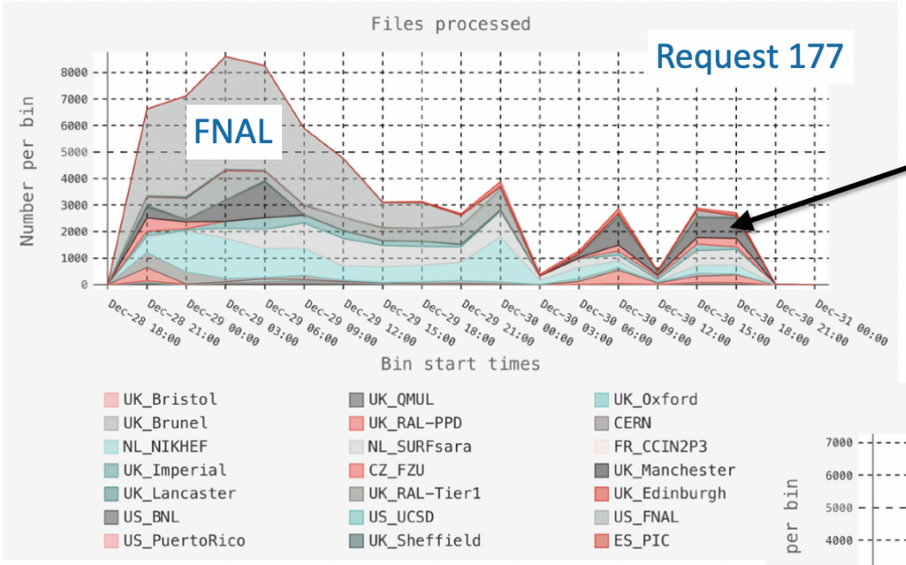
- Ran 5 days, July 11-15, Midnight UTC to Midnight UTC
- Generated ~500TB of data total
- Near end of the challenge reached a peak of 3.6GB/s=28.8Gbit/s
- utilized FTS-3 for initial ingest, Rucio for replication, https and 3rd party transfers
- helped to identified issues that have to be fixed before beam run begins



DC-4 Phase 2 - Data Processing

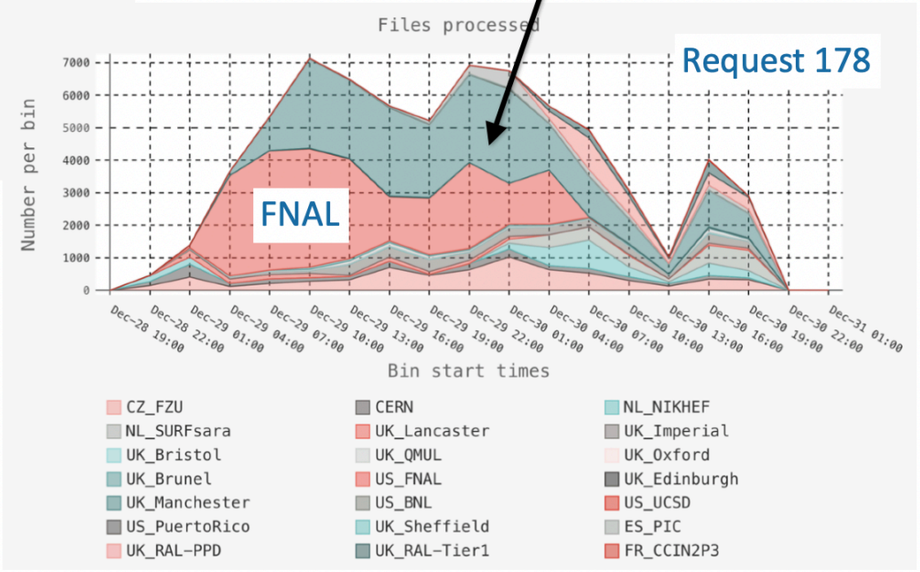
- Successfully shown ability to easily sustain 5000 jobs
- more importantly, automatic matching between the job locality and data locality
- looking for improved efficiency in the “flatter” computing model

(Charts from WFS Dashboard)



Not much Manchester

Lots of Manchester!



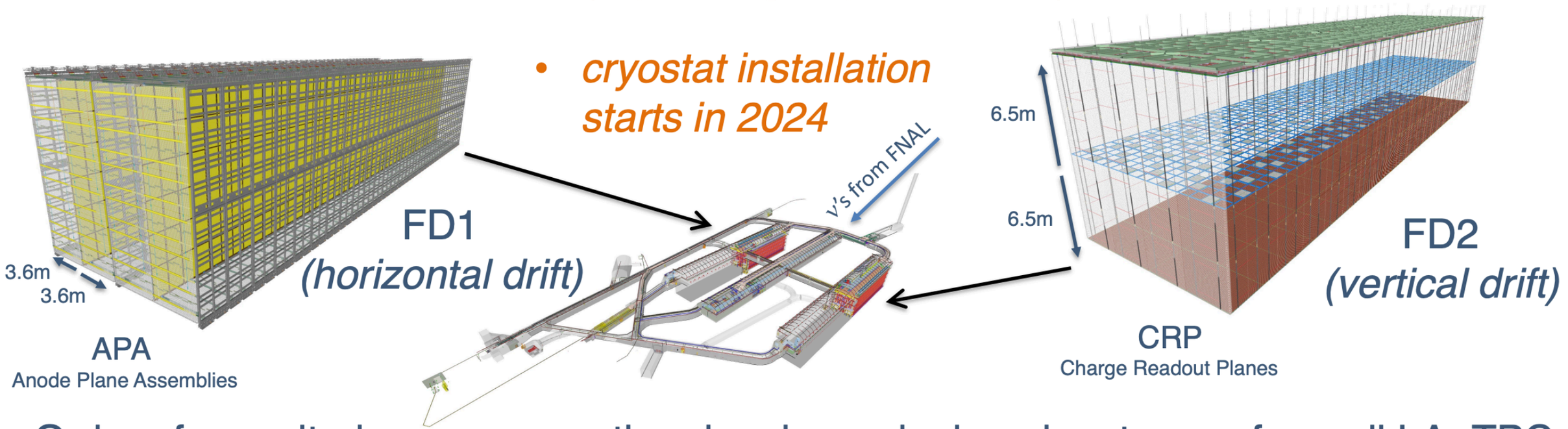
Allocations follow the data to suitable sites, using replica locations originally from Rucio

Central Utility Cavern

DUNE Far Detectors (Phase I)

Sam Zeller's talk
@ FNAL P5 Town Hall

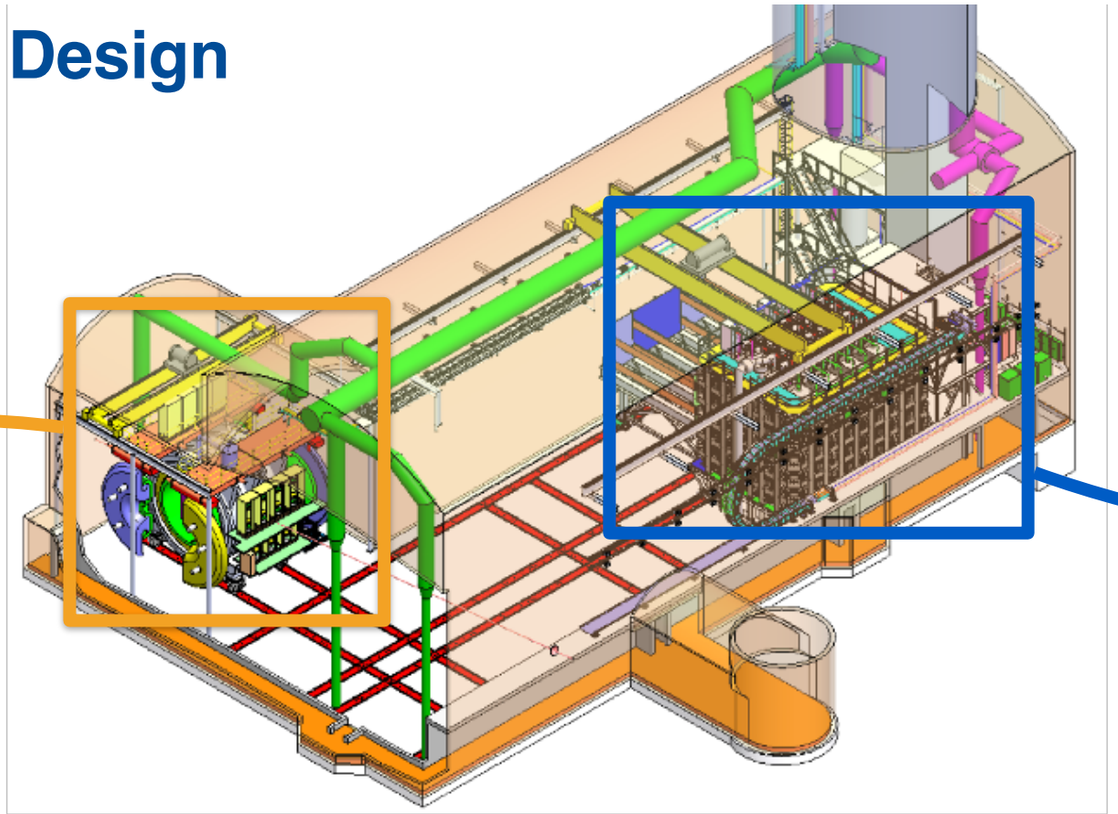
- Phase 1 will include caverns for 4 detector modules in South Dakota and 2 far detector modules, each 17 kton of LAr, the largest LAr TPCs ever constructed.
  - *FD1: horizontal drift (ala ICARUS, MicroBooNE)*
  - *FD2: vertical drift (capitalizing on protoDUNEs)*



- Order of magnitude more mass than has been deployed up to now from all LAr TPCs

DUNE Near Detector Design

on-axis neutrino detector for beam monitor



ND that is functionally identical to FD - but moveable

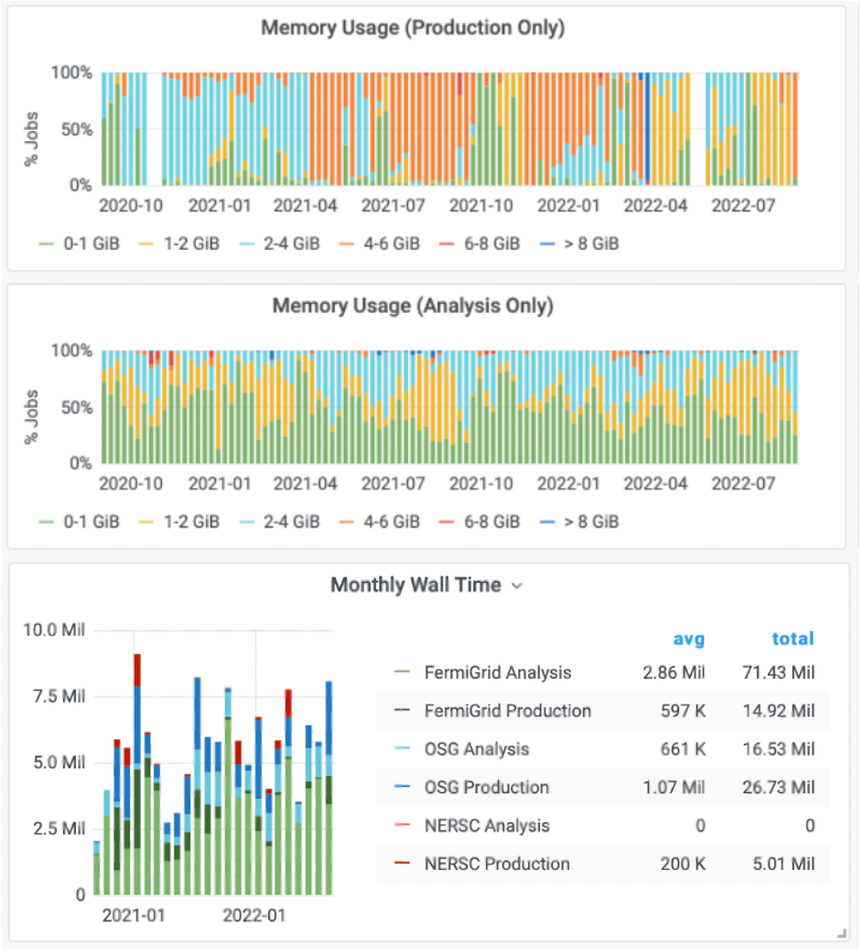
- Near Detector - three sub-detectors serving different purposes
- ND-LArTPC: Highly segmented Liquid Argon Time Projection Chamber
- TMS - Muon Spectrometer (Phase 1) / ND-GArTPC - Gaseous Ar TPC (Phase 2)
- SAND - scintillator-based tracking and active argon target for on-axis beam monitoring
- DUNE-PRISM: Movement of LAr + TMS/GArTPC transverse to the beam, sampling beam profile, energy spectrum, and ν interactions to reduce systematic uncertainties

Defining the Computing Challenges

Offline reconstruction experience and impact on workflow and resources

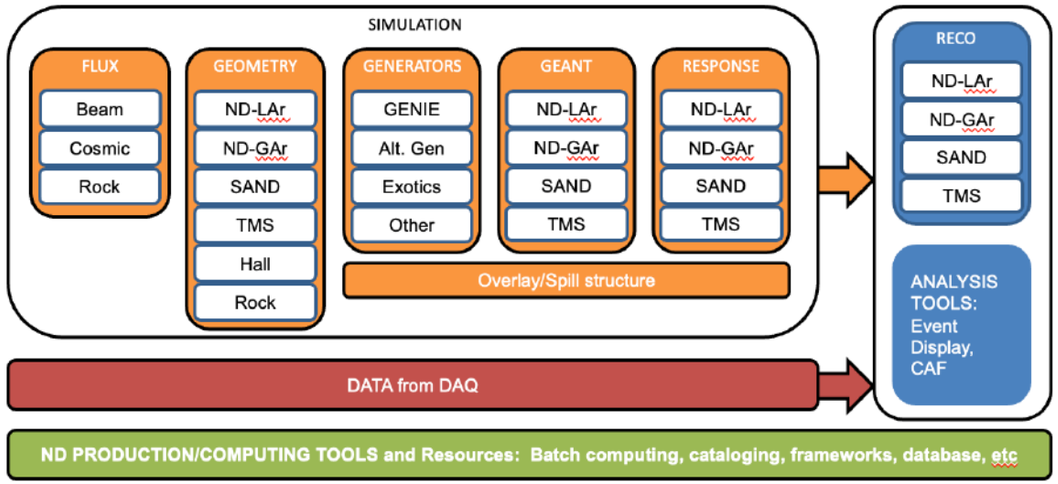
- ProtoDUNE offline reconstruction
  - signal processing memory “pinch point”
  - adapt workflow & software to fit site constraints
  - adjust to future expectation of localized readout
  - process time extended trigger records
- Near Detector simulation and reconstruction
  - strong algorithm development and important achievements helping to define ND design
  - need for integration into a framework
  - reco simulation samples for resource estimates
  - advances in ML/AI show need for an improved standard development environment

Last 2 years Grid job stats



Offline reconstruction experience and impact on workflow and resources

- ProtoDUNE offline reconstruction
  - signal processing memory “pinch point”
  - adapt workflow & software to fit site constraints
  - adjust to future expectation of localized readout
  - process time extended trigger records
- Near Detector simulation and reconstruction
  - strong algorithm development and important achievements helping to define ND design
  - need for integration into a framework
  - reco simulation samples for resource estimates
  - advances in ML/AI show need for an improved standard development environment

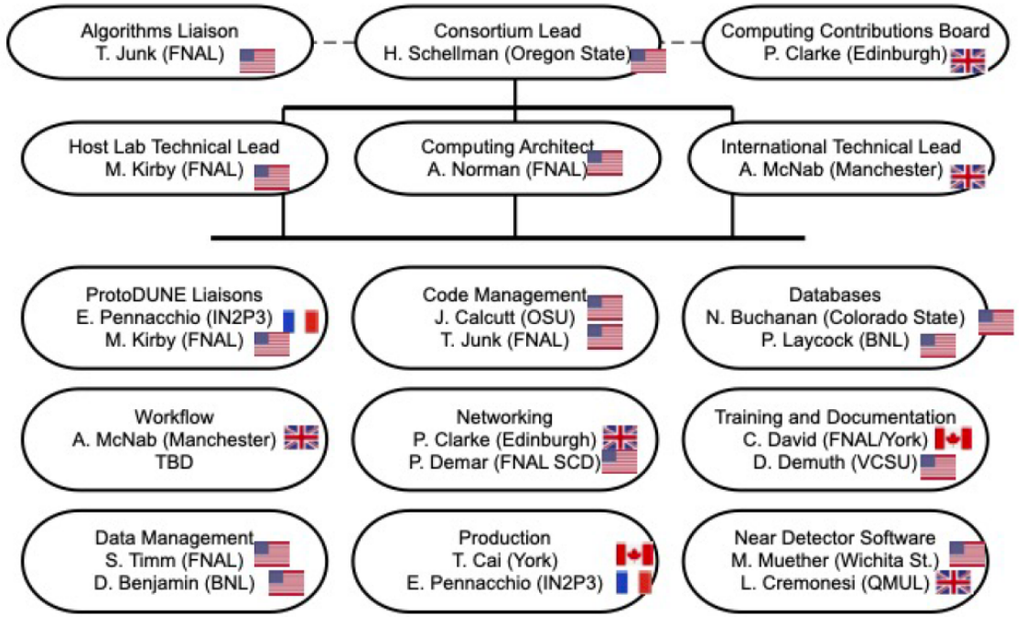


GArSoft run time stats

Module Label	time/event (sec)
RootInput(read)	0.00
init:EventInit	1.31275e-05
hit:CompressedHitFinder	0.00488308
tpcclusterpass1:TPCHitCluster	0.0091922
vechit:tpcvechitfinder2	0.0103787
patrec:tpcpatrec2	0.0130245
trackpass1:tpctrackfit2	0.014211
vertexpass1:vertexfinder1	0.000851081
tpccluster:tpccathodestitch	0.0269436
track:tpctrackfit2	0.0135842
vertex:vertexfinder1	6.19847e-05
veefinder1:veefinder1	9.96417e-05

Overview of the Computing Consortium

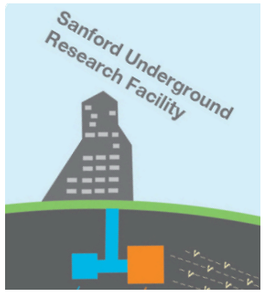
- organization of the working groups and management
- define source of resources
  - CCB - computing resources
  - division of effort for development
- layout interfaces with other consortia



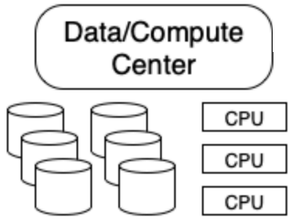
Consortium	Sites	• Collaboration
• Project management • Core software development • Core operations • Interfaces to DUNE consortia • Interfaces to other projects • Training • User support	• Storage • CPU • Networking	• Algorithms • Operations help • User support • Production group • Validation • Calibrations • New ideas

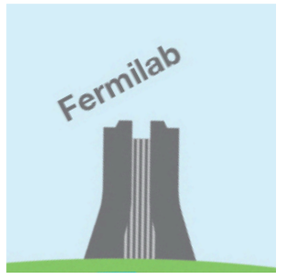
Cartoon version of data movement

FD Raw Data
archival storage

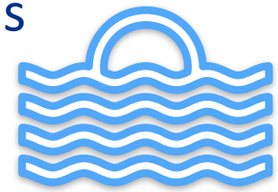


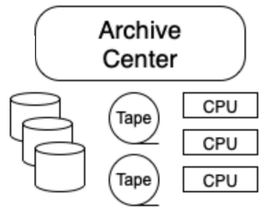
8 Gbit/s

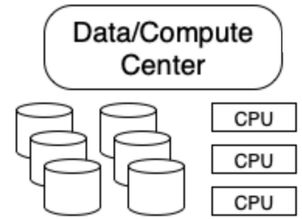


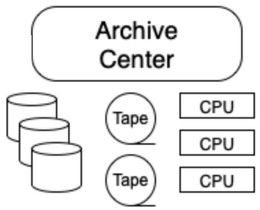


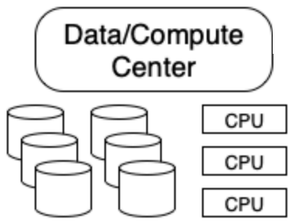
8 Gbit/s

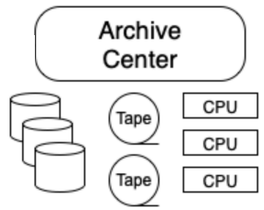


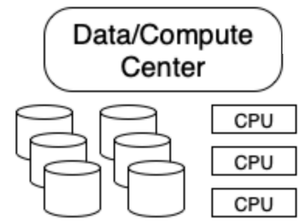










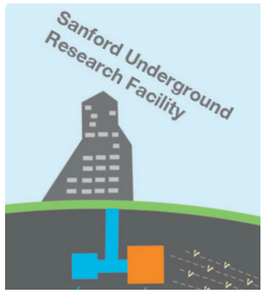


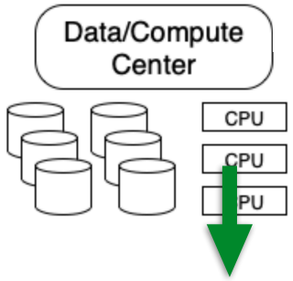
not to scale, not a technical design
it's just a cartoon

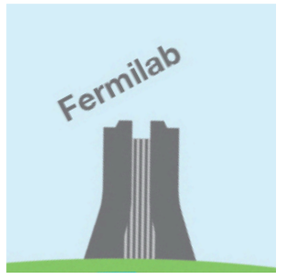
Use multiple archive sites
not in North America

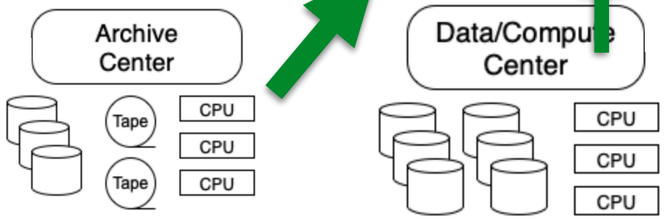
Cartoon version of data movement

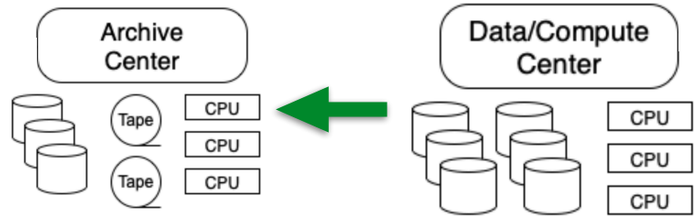
FD Data
prod output

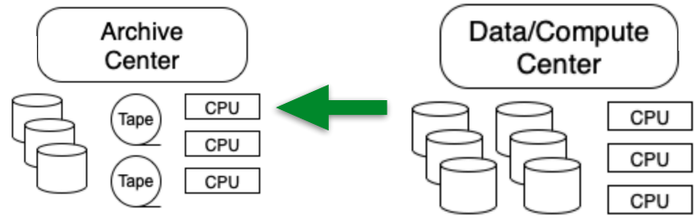








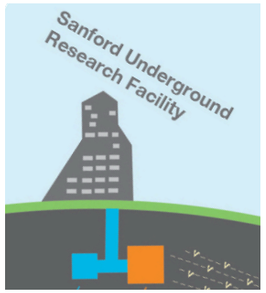


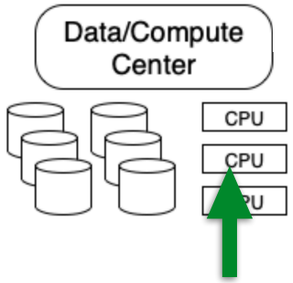


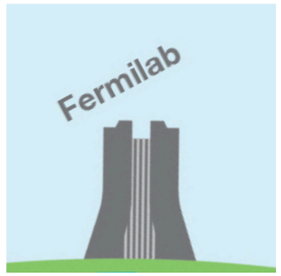
not to scale, not a technical design
it's just a cartoon

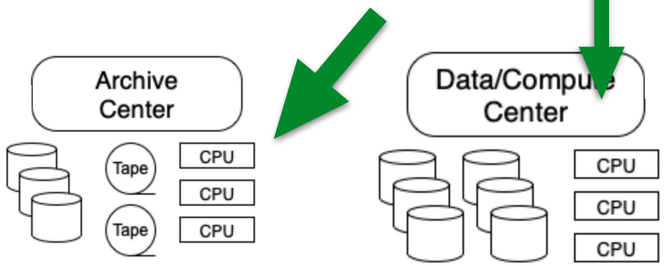
Cartoon version of data movement

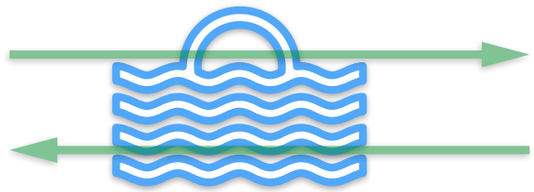
FD Data
analysis processing

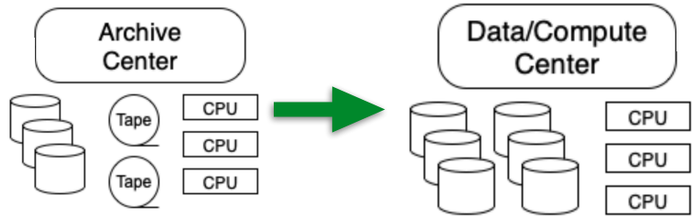


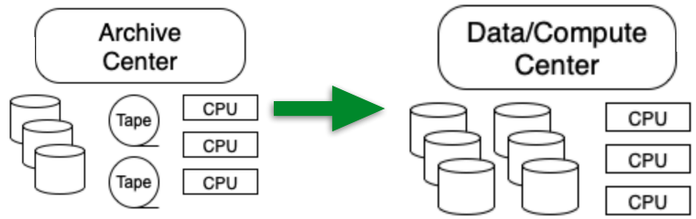








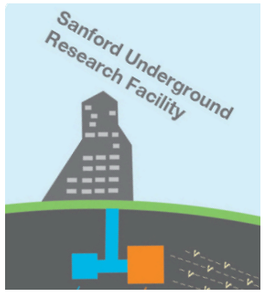




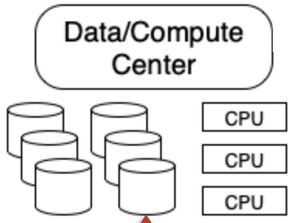
not to scale, not a technical design
it's just a cartoon

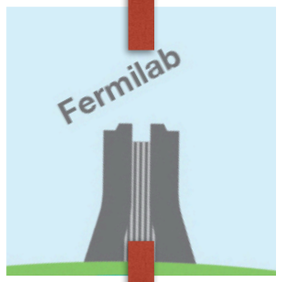
Cartoon version of data movement

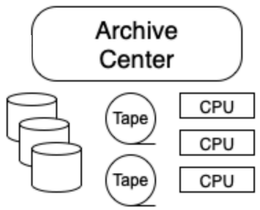
SuperNova Raw Data
rapid transfer & processing

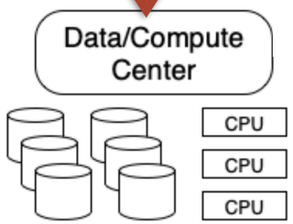


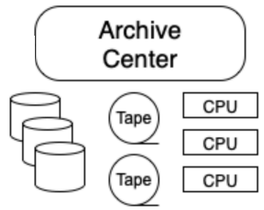
100 Gb/s

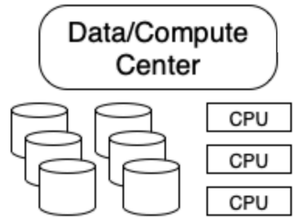


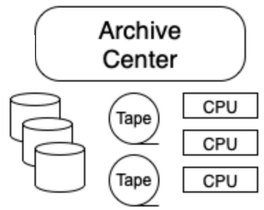


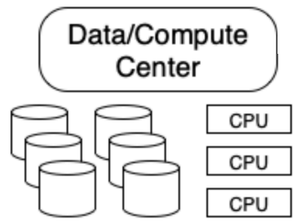










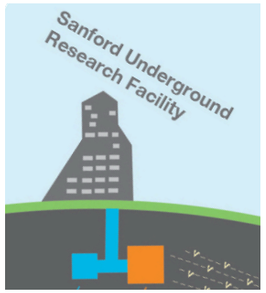


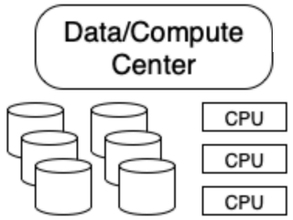
not to scale, not a technical design
it's just a cartoon

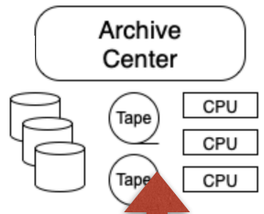
Not anticipated to be part of DC24

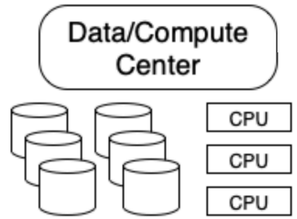
Cartoon version of data movement

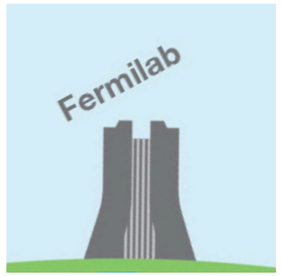
ProtoDUNE Raw Data
archival storage



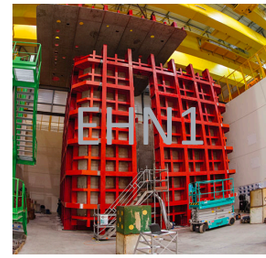


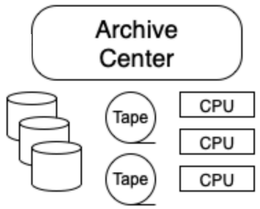


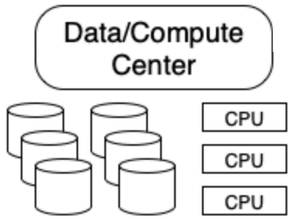


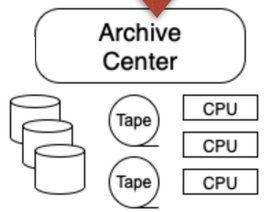


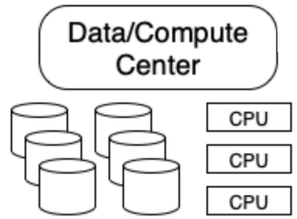
32 Gbits/s









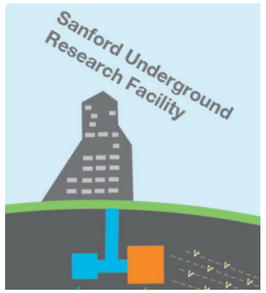


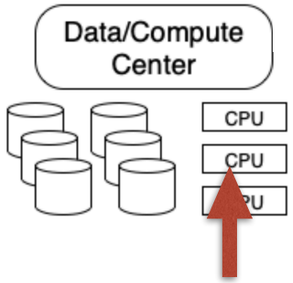
not to scale, not a technical design
it's just a cartoon

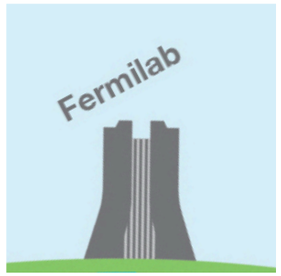
Cartoon version of data movement

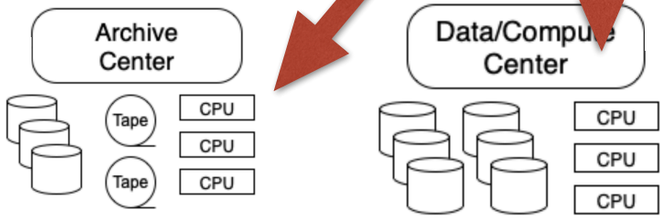
ProtoDUNE Raw Data

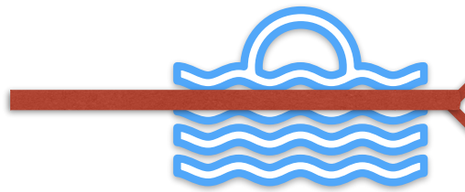
raw data processing

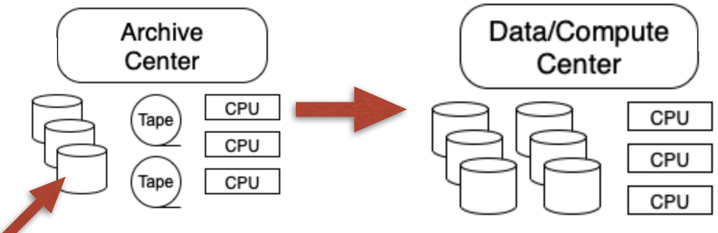


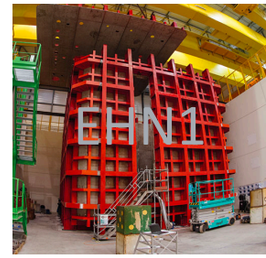


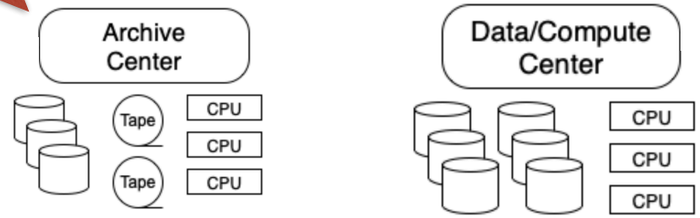










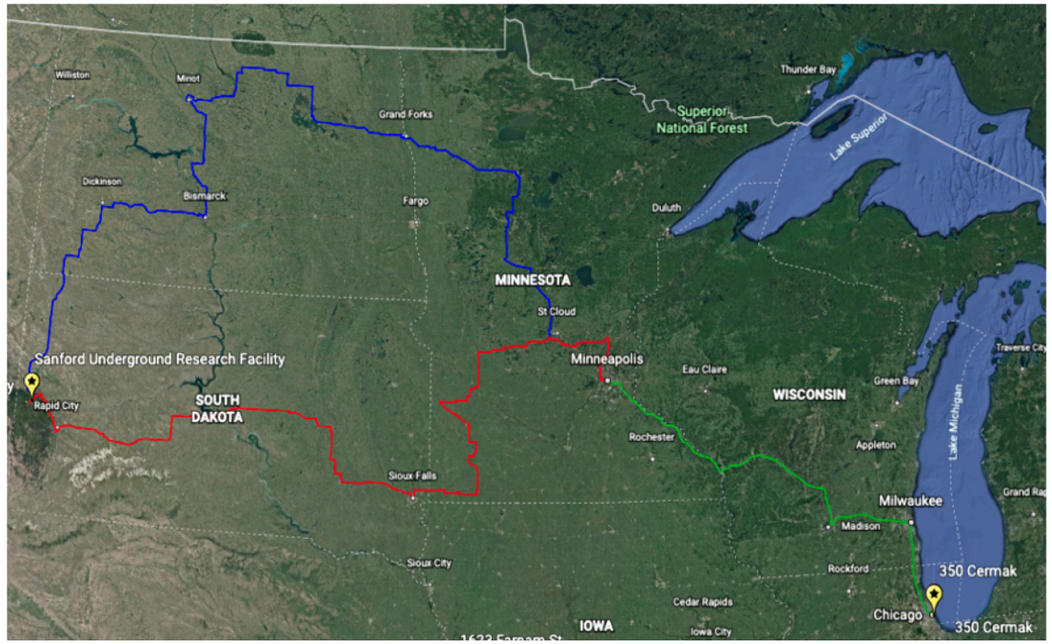


not to scale, not a technical design
it's just a cartoon

Listing the Unique DUNE Requirements

Networking for DUNE

- important requirements for operations
  - potential natural disasters combine to make multiple paths important
- interface between online and offline resources
- unique challenges with SNB data transfer and opportunity for fast processing
- wide area network:
  - DUNE is very well served by national science networks
  - ESnet, Geant, Janet, Surfnet, Renater….
- DUNE relies on these and thanks them.
  - fully engaged and participate in the global LHCOPN/LHCONE meetings
  - NRENS are ~1 TB backbones w/ 100 GB to many sites
  - discussed with ESnet potential/need for virtualised networks - not foreseen at this time



DUNE FD WAN Bandwidth Timeline Projections:

Date	*Stage of the experiment*	*Primary Path*	*Secondary Path*	*Tertiary Path*
Now	Cavern excavation	10GE	< 1GE via SURF	none
2025	Detector construction	10GE	< 1GE via SURF	none
2027	Computing/DAQ deployment	100GE	10GE	< 1GE via SURF
2028	Cryo deployment completed	100GE	10Gb/s+	10GE
2029	Start of science	100GE	10Gb/s+	10GE

vLAN service provided by REED/GPN (shared)
Dedicated circuit Ross Dry Bldg. to Chicago
Dedicated circuit Yates Complex to Denver (10GE or 100GE)

Offline Computing Frameworks

- thanks to **Frameworks Requirements Working Group**
- document available on docdb at DocDB 24423
- review by the HEP Software Foundation (HSF)
  - recommend Production and Analysis frameworks may be different
  - develop unified framework for all offline prod
  - plan interface with Workflow Management
- handle very large events on complex architectures
  - partial region and subsetting of trigger records
  - temporal stitching of trigger records
- threadsafe + i/o management
- work plans for development of dynamic time-window processing
  - FNAL LDRD funded for a prototype framework
  - DUNE coordinating closely with FNAL effort

Diagram from Kyle Knoepful FNAL LDRD