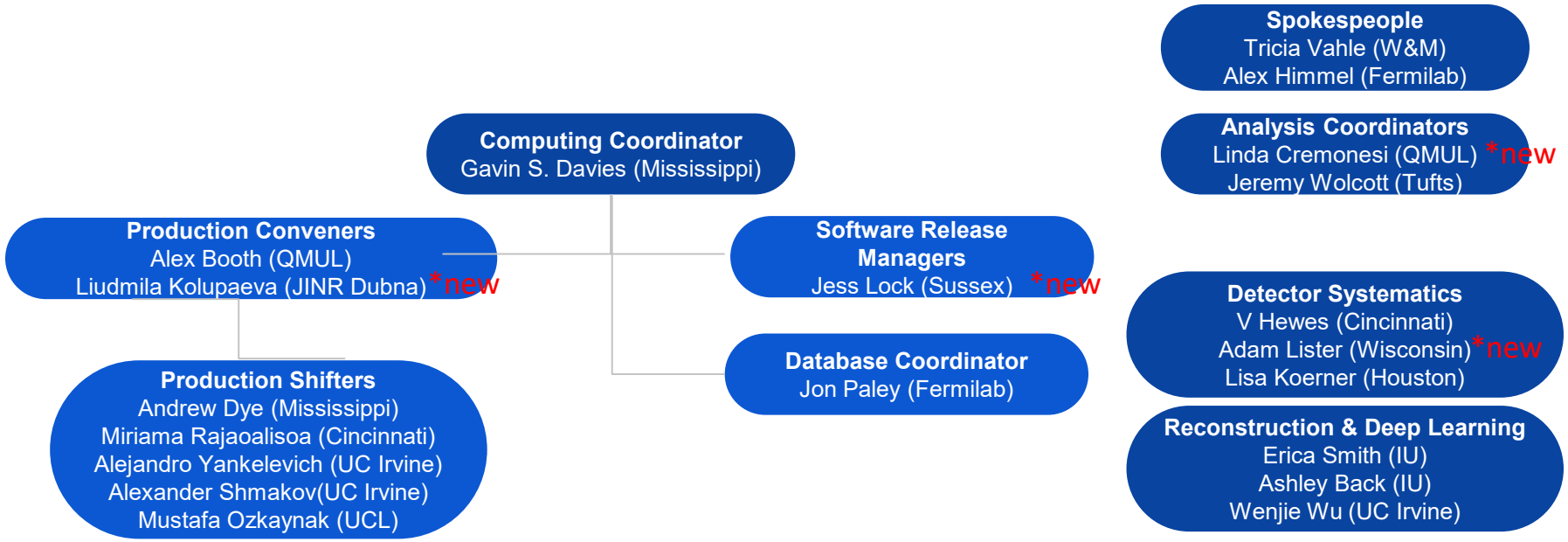
Fermilab | U.S. DEPARTMENT OF ENERGY | Office of Science

NOvA FCRSG FY23

Gavin S. Davies

Organization Chart for Offline Computing

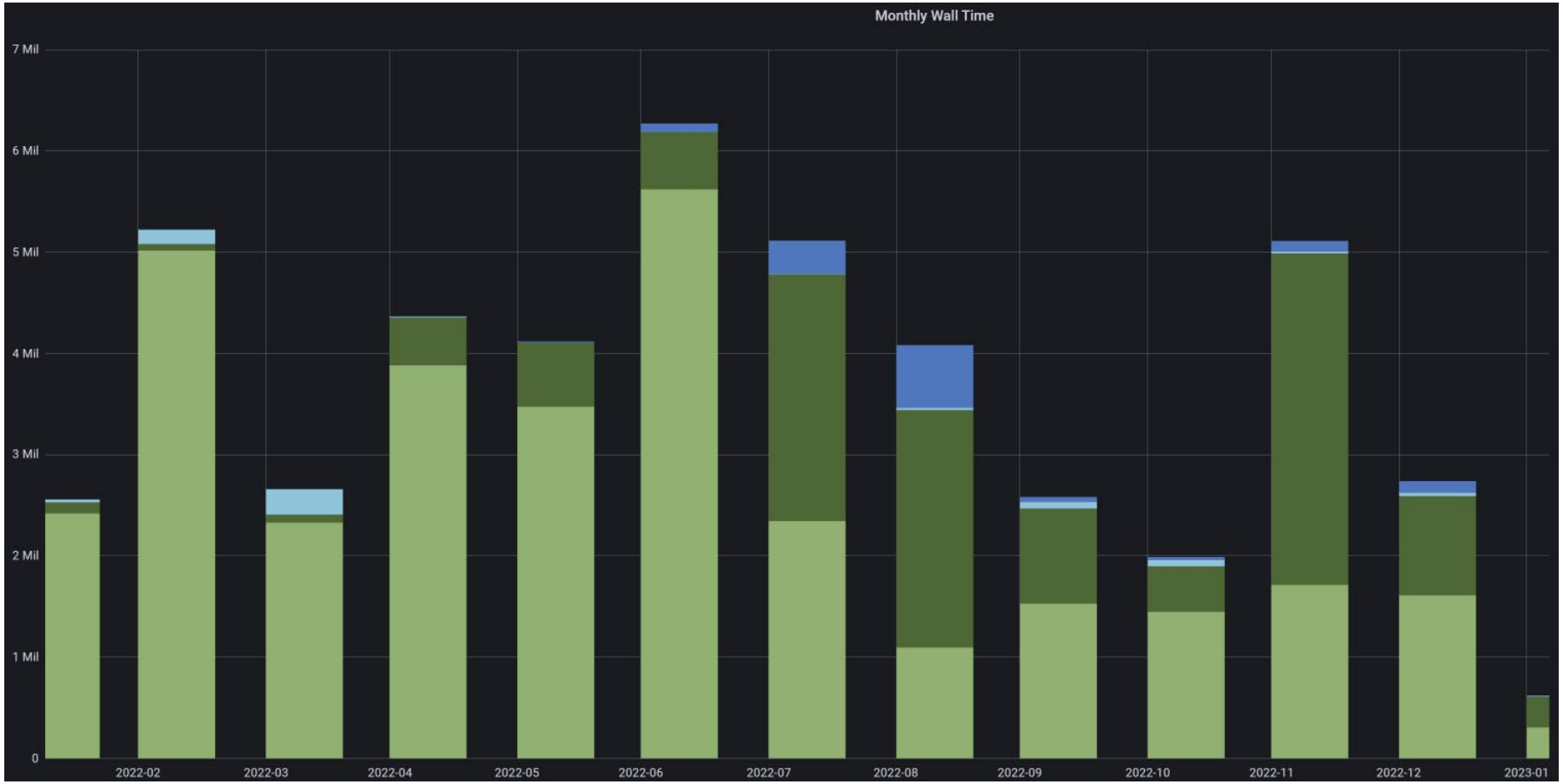


Computing Coordinator serves as CS Liaison (GD)

Production group consists of several “shifters” that handle submissions weekly
GENIE/GEANT/CRY development coordinated by Detector Systematics group

ML development coordinated within Reco & DL sub-group

Experiment CPU Usage over the past year



	avg	total
FermiGrid Analysis	2.52 Mil	32.8 Mil
FermiGrid Production	972 K	12.6 Mil
OSG Analysis	50.4 K	655 K
OSG Production	103 K	1.35 Mil
NERSC Analysis	0	0
NERSC Production	0	0

Fermilab

Why so little on the OSG?

OSG Canary Job Summary

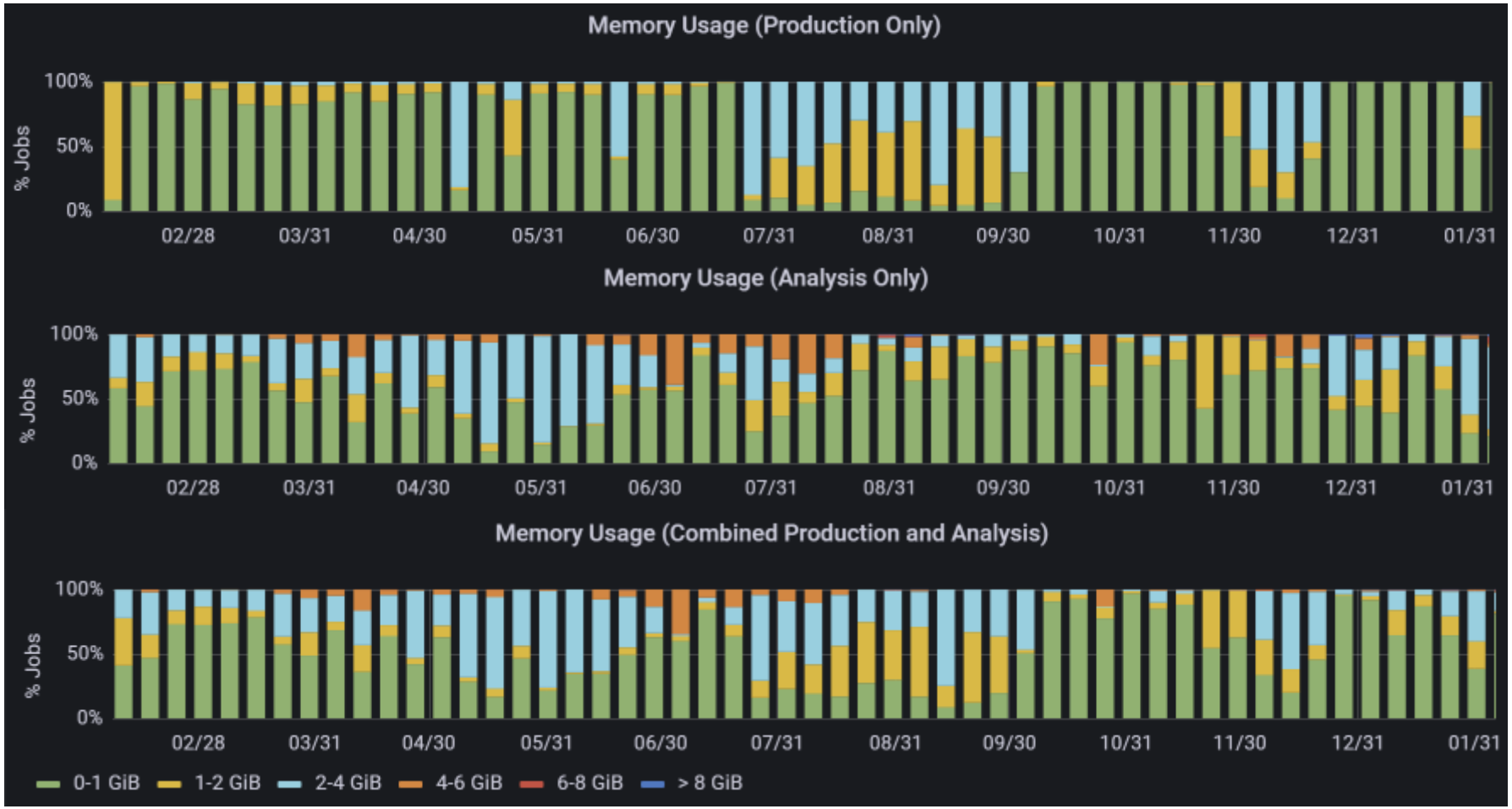
Website last updated: 2023-02-10 01:00:14

OSG Site	rockgen	g4	detsim	reco	CAFAna
BNL					
BU					
Caltech					
Clemson					
Colorado					
Cornell					
FNAL					
FZU					
HOSTED_STANFORD					
Hyak_CE					
JINR					
JINR_CLOUD					
JINR_Tier2					
SU-OG					
Lincoln					
Michigan					
MIT					
MWT2					
Nebraska					
NotreDame					
Omaha					
OSC					
SDSC-PRP					
SLATE_US_NMSU_DISCOVERY					
SMU					

We track OSG sites with canary jobs: https://nusoft.fnal.gov/nova/production/canary/summary.h

Most never run, no effort within the collaboration to track issues down site-by-site.

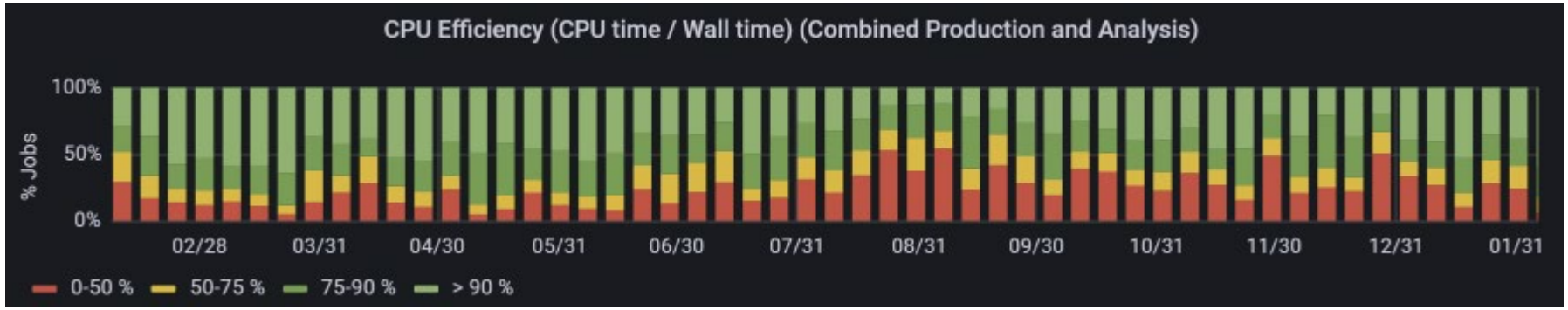
Memory footprint over the past year



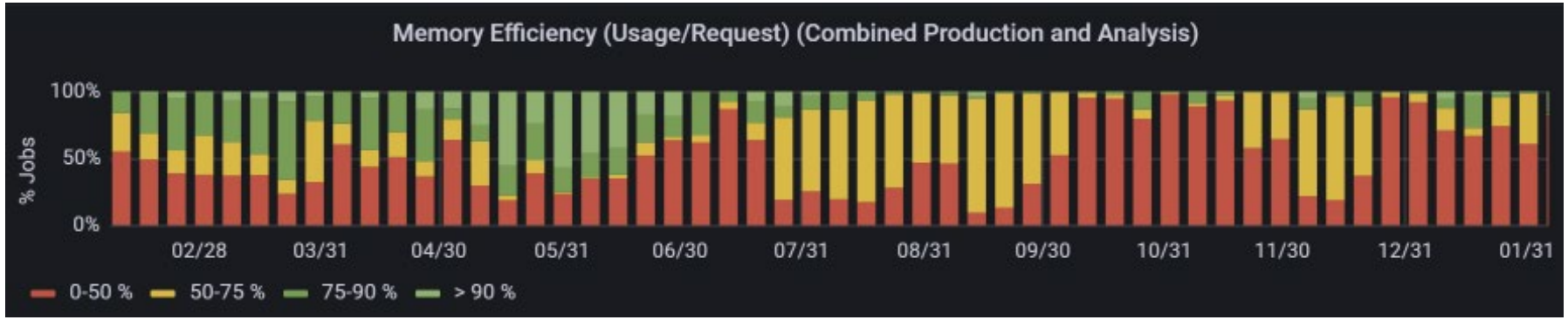
Broadly good memory footprint due to extensive efforts to reduce and keep below 2 GB

Fermilab

CPU and memory efficiency over the past year



Overall very efficient CPU usage



Note periods of “poor” memory efficiency correspond to periods with many analysis jobs using < 1GB of memory, not excessive request for large memory slots.

Fermilab

What do you want to achieve in computing over the next 5 years?

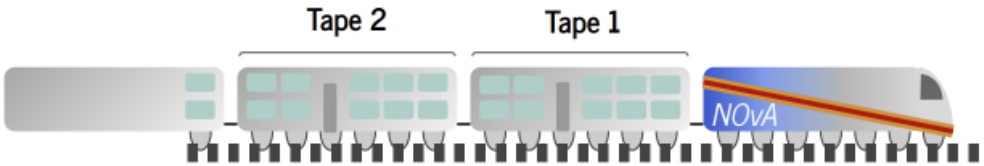
None of these are desires, but rather requirements to stay up-to-date
Very low on required expertise for some of these.

SCD-NOvA workshop was incredibly fruitful

Goals	Where does the experiment need to contribute	Where does SCD need to contribute
SL7 -> AlmaLinux upgrade Linux and DAQ support	Port of experiment specific code, building, and testing	May need help from experts Possibly need waiver or container for online (DAQ) systems
Software packaging and Distribution Migrate to using Spack	Port of experiment specific setup code (SRT), testing, and deployment	Documentation and support, providing Spack builds of standard packages
Authentication	From certificates to tokens	Support for FTS, job submission, SAM access in grid jobs

Become consumers of Spack products, unlikely to switch to *creating* Spack packages unless substantial support is available.

NOvA Freight Train



- Brand new workflow started in March 2022 where we process files tape by tape based on advice from SCD experts.
- Motivation: we have many files from many trigger streams scattered over many tapes and prestaging is a major production bottle-neck
  - Visit each tape once only.
  - Prestage and process all files from that tape before moving onto the next one
  - Most processing happening on offsite HPCs (ALCF, and soon NERSC)
- Two projects: raw train and artdaq train

Raw train

- Prestage tapes
- Copy raw files to persistent

→ Num. of tapes: 351
→ Num. of files: ~1.1 million
→ Total size: 1.34 TB

Artdaq train

- Prestage tapes
- Send NuMI and cosmic files to ALCF for doing cosmic filtering on GPUs
- Process exotics files in FermiGrid
- Reco planned in the future

→ Num. of tapes: 5003 (14 raw trains)
→ Num. of files: ~6 million (6 raw trains)
→ Total size: 2.63 PB (**2000 raw trains**)

This has been the major focus of production development in the last year.

Fermilab

Campaign Schedules

	2023	2024	2025	2026	2027
Processing campaigns (start month-end month if known). **Include when you expect to be prestaging**	“freight-train” FD NuMI and cosmic data (last Nu24 prod computing) Prep inputs for prod6 ~September, miniprod6 general validation, training inputs	miniprod6 part 2, Q1 validate trained networks Prod6 Campaign, ~12 month campaign similar to 2019/20 campaign for Nu26	Continued prod6	FD Data top-up for prod6 Potential (ND-only?) reco-only campaign	no estimate
Storage + CPU estimates (call out any special resource needs if known, e.g. HPC or GPU). **Include amount(s) to be prestaged and file families, in addition to space needed for new outputs.**	Large prestaging for FT Processing largely at ALCF & NERSC ~5k tapes to prestage 3 PB of data, ~1/3 through Rotating storage of 500 TB	On-site running ramps up Prod6 likely larger than 2019-20 campaign Need more storage 500 TB is always at 99% full, v. careful management			
Conference or result targets (month if known)	Working towards Neutrino 2024, theses	Neutrino 2024 (based on ‘22-‘23 computing)		Neutrino 2026 (based ‘24-’25 computing)	

CPU @ Fermilab Prediction Going Forward and Accuracy of Your Predictions [units of Million (1 CPU, 2GB) wall hours per CY]

	2019	2020	2021	2022	2023	2024	2025	2026	2027
Requested (could have multiple values for different MWC combinations)	31.60	29	33	35	30 analysis/sim + 10 production = 40	30 analysis/sim + 25 production = 55	30 analysis/sim + 10 production = 40	30 analysis/sim + 15 production = 45	30 analysis/sim + 10 production = 40
Actual Used	25.94	40.5	41.4	44.7	3 to date	N/A	N/A	N/A	N/A
Efficiency	77%	67%	74%	80%	N/A	N/A	N/A	N/A	N/A

FY 22 : 33 M is analysis, 12 M is production

CPU – non-FNAL HTC Resources Going Forward and Accuracy of Your Predictions [units of Million (1 CPU, 2GB) wall hours per CY]

	2019	2020	2021	2022	2023	2024	2025	2026	2027
Requested (could have multiple values for different MWC combinations)	N/A	N/A	N/A	N/A	2	2	2	2	2
Actual Used	??	??	2.2	2	N/A	N/A	N/A	N/A	N/A
Efficiency	%	%	%	%	N/A	N/A	N/A	N/A	N/A

Historically have run more on the OSG, no longer have the effort within the collaboration to make sites work one-by-one.
https://nusoft.fnal.gov/nova/production/canary/summary.html

CPU – HPC Resources Going Forward and Accuracy of Your Predictions [units of Million (1 CPU, 2GB) wall hours per CY]

TO DO

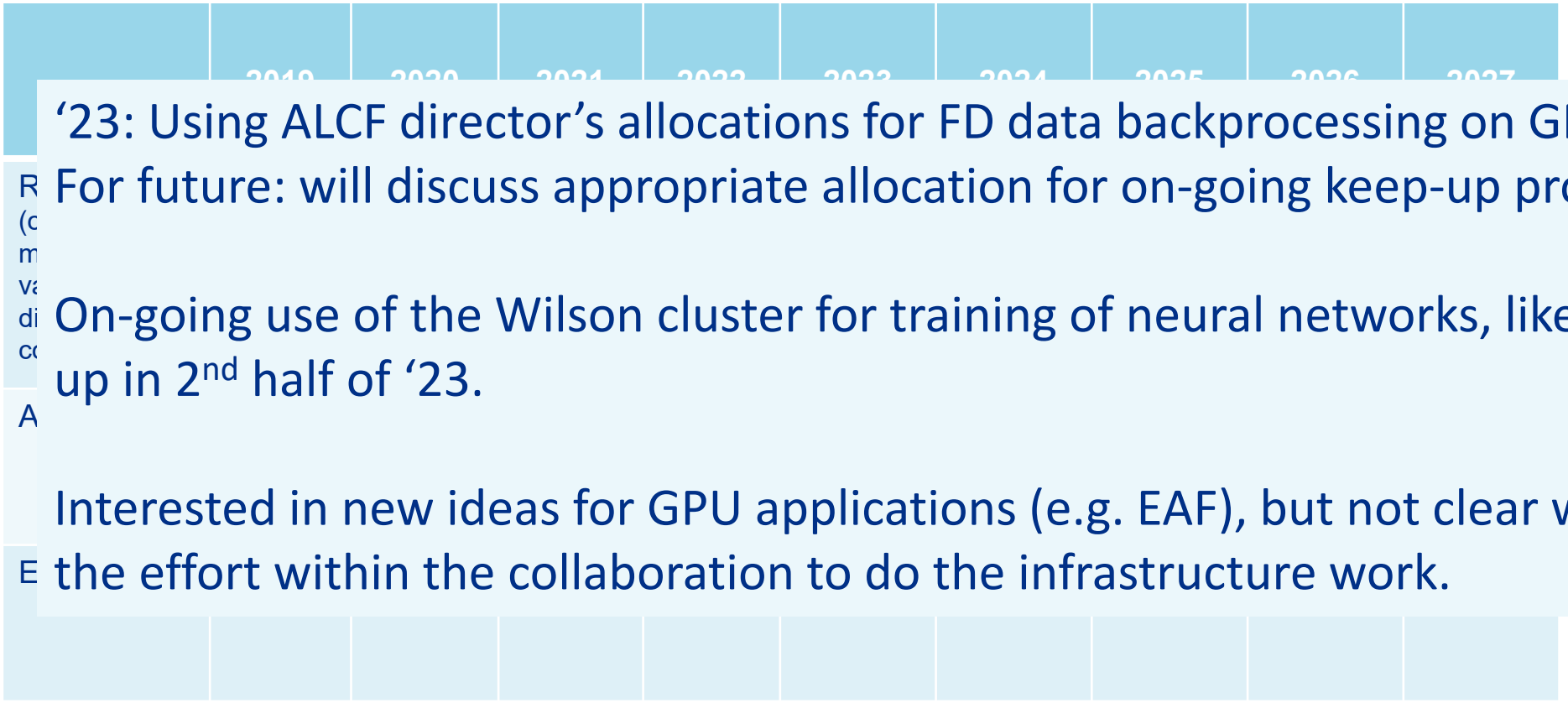
Now using HEPCloud/NERSC in tandem with ALCF for freight-train.

	’23: Plan on NERSC primarily for production processing of FD data ‘24: Plan on using NERSC primarily for Feldman-Cousins for Nu24								
Requested (could have multiple values for different MWC combinations)	Future plans less certain								
Actual Used					N/A	N/A	N/A	N/A	N/A
Efficiency	%	%	%	%	N/A	N/A	N/A	N/A	N/A

Fermilab

CPU – GPU Resources Going Forward and Accuracy of Your Predictions [units of Million (1 CPU, 2GB) wall hours per CY]

TO DO



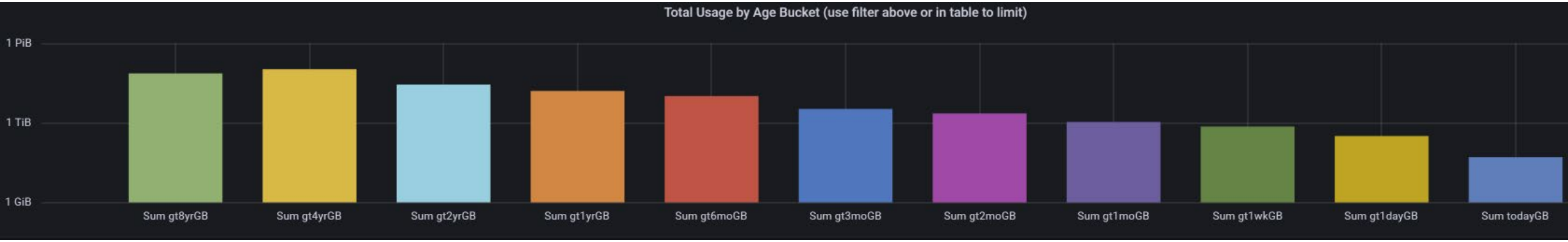
‘23: Using ALCF director’s allocations for FD data backprocessing on GPUs

For future: will discuss appropriate allocation for on-going keep-up processing

On-going use of the Wilson cluster for training of neural networks, likely to ramp up in 2nd half of ‘23.

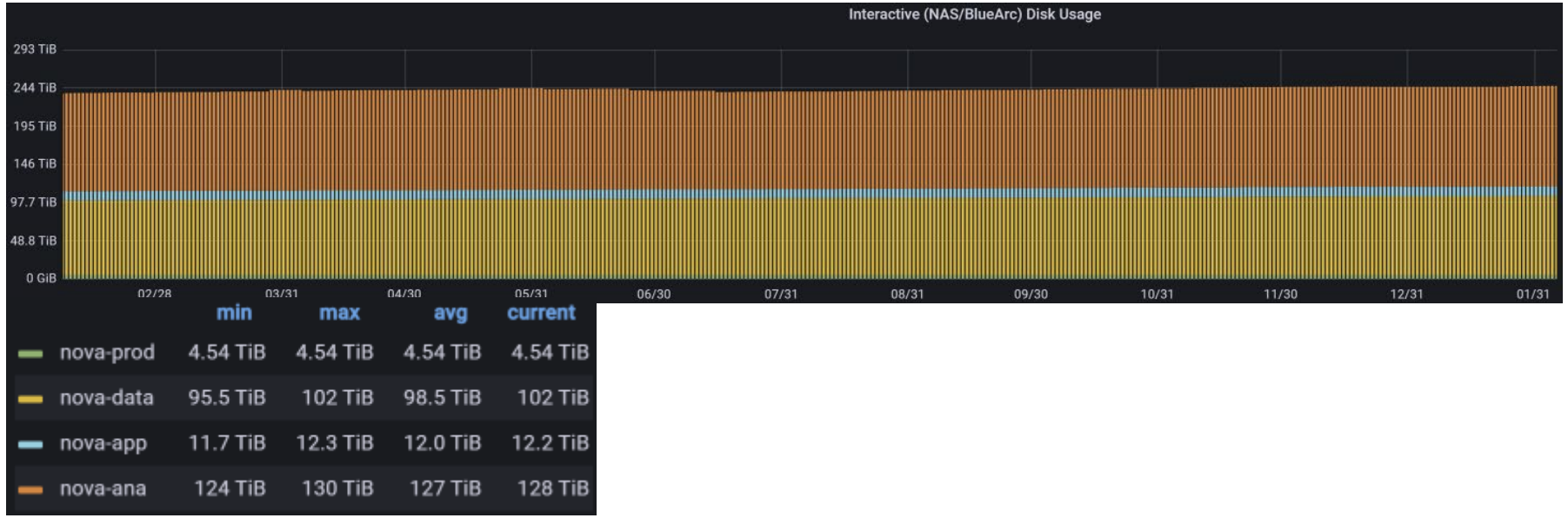
Interested in new ideas for GPU applications (e.g. EAF), but not clear we have the effort within the collaboration to do the infrastructure work.

Age of files in NAS



Fermilab

NAS Usage and Projections



No increase requested. Request to keep what we have.

	App	Data
2022	14 TiB	105 TiB
2023	14 TiB	105 TiB
2024	14 TiB	105 TiB
2025	14 TiB	105 TiB
2026	14 TiB	105 TiB
2027	14 TiB	105 TiB

Fermilab

dCache Usage and Predictions (in TB)

	min	max	avg	current
Fermilab Public dCache:StorageGroup:nova_readWritePools	468 TiB	2.08 PiB	1.02 PiB	684 TiB
Fermilab Public dCache:StorageGroup:nova_NovaAnalysisPools	369 TiB	527 TiB	433 TiB	501 TiB
Fermilab Public dCache:StorageGroup:nova_CdfWritePools	2.50 MiB	362 TiB	27.4 TiB	362 TiB
Fermilab Public dCache:StorageGroup:nova_NovaPrestagePools	5.04 TiB	263 TiB	152 TiB	139 TiB
Fermilab Public dCache:StorageGroup:nova_PublicScratchPools	7.67 TiB	259 TiB	39.4 TiB	249 TiB
Fermilab Public dCache:StorageGroup:nova_NovaWritePools	37.9 TiB	56.3 TiB	44.7 TiB	56.1 TiB
Fermilab Public dCache:StorageGroup:nova_Geant4ReadWritePools	1.86 TiB	21.6 TiB	9.32 TiB	2.60 TiB
Fermilab Public dCache:StorageGroup:nova_SlowReadWritePools	1.86 TiB	21.6 TiB	9.32 TiB	2.60 TiB
Fermilab Public dCache:StorageGroup:nova_NovaStagePools	3.56 TiB	20.4 TiB	3.93 TiB	3.56 TiB
Fermilab Public dCache:StorageGroup:nova_HsmWritePools	757 GiB	13.5 TiB	5.72 TiB	2.64 TiB
Fermilab Public dCache:StorageGroup:nova_HsmReadPools	146 GiB	10.4 TiB	3.56 TiB	686 GiB
Fermilab Public dCache:StorageGroup:nova_SlowPublicScratchPools	41.3 GiB	8.07 TiB	764 GiB	7.75 TiB

TO DO

	Analysis (Persistent)	Other Dedicated (Write)
Current	545 TB (actual)	Y TB (actual)
2023	Any increase!	
2024		

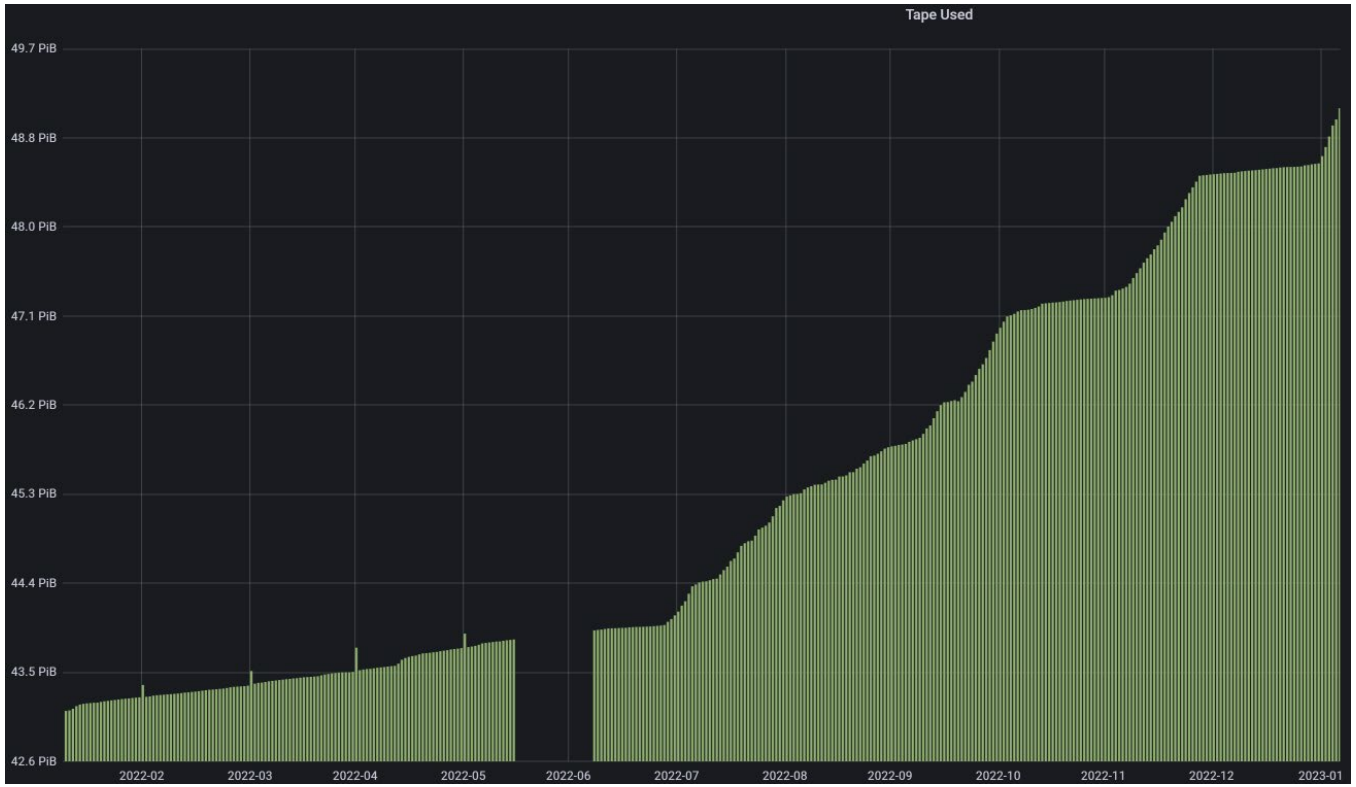
Persistent space is a *very* valuable tool now that prestaging is fraught.

Stays very full and must carefully managed, taking time of collaboration experts not doing other development instead.

Any increase in available persistent space would be put to use holding active analysis files, ART files for development of new algorithms, and buffering prestaging projects like Freight Train.

Fermilab

Tape usage and predictions (in TB)



	Total Added By End of Year
At end 2021	+3 PB (actual)
2022	+5 PB
2023	+4 PB
2024	+3 PB
2025	+3 PB
2026	+2 PB
2027	+2 PB

How are your file families structured? Should you revisit them?
See https://fifemon.fnal.gov/monitor/d/BSnVdWDnk/tape-data for details

Add word on file families

Data Lifetimes

We have never created a data set with a designated lifetime.

- We do not have plans to delete datasets; requires considerable care and effort to ensure that crucial, irreproducible data is not lost.
- There are datasets that we would not copy forward to new media (several Pb).
- We have explored what we could leave behind / delete.
- There is roughly 1.8 Pb of data from design and prototyping stages of experiment

- Ultimately, without provided tools, deleting datasets is not practical

Analysis Facility Use

- Prepare a slide answering these questions:
- Are you planning to use the Elastic Analysis Facility? If so, roughly how many concurrent users would you expect?
- Do the I/O needs of this work differ from traditional types? If so, how? What storage would need to be available to make effective use of the facility? Bulk storage, fast storage (e.g. NVMe)? Keep in mind that faster I/O = smaller capacity in general (about a factor of 10 less NVMe than HDD for the same money)
- What kind of hardware do you require? Currently has CPUs and (smaller number of) GPUs
- More details available at the August 2022 EAF Stakeholders Meeting