Muon g-2 FCRSG FY23

Adam Lyon (FNAL) for Muon g-2

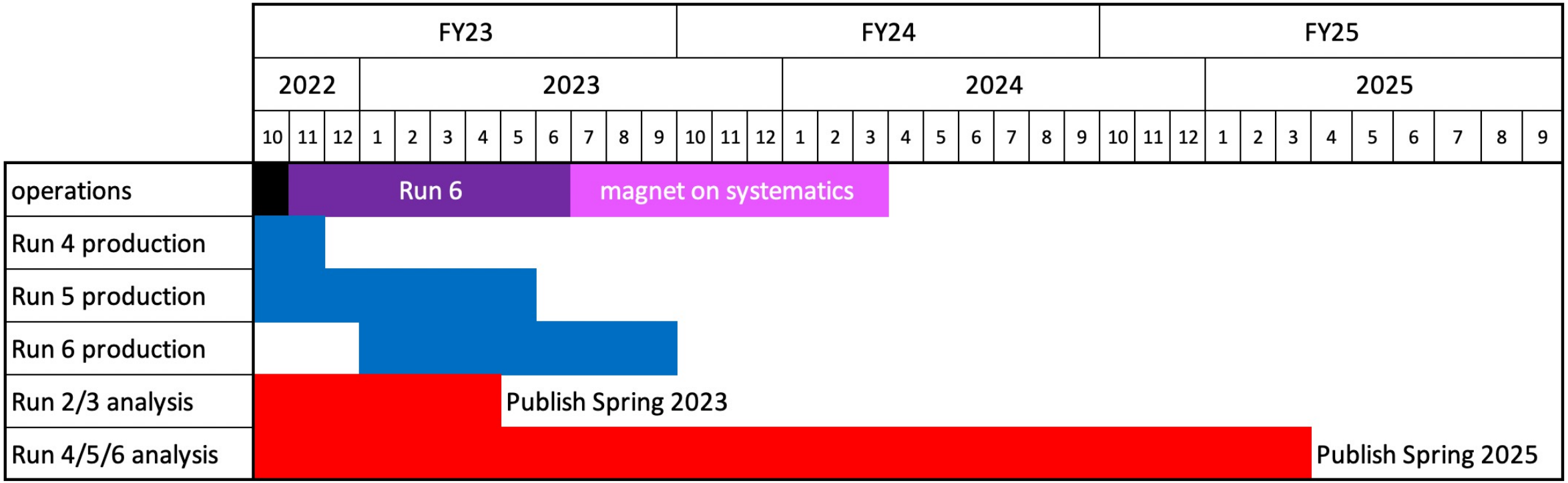
New information this year

- ~~5-year~~ 3-year vision of computing needs (see next slides)
  - Data taking (Run 6) ends this summer
  - Final publication scheduled for Spring of 2025
  - Need to formulate a data preservation plan and execute
  - We do not envision Muon g-2 needing computing resources after FY25 except for data preservation
    - May need to revisit this statement for the next FCRSG. There are follow-on papers that may take longer (e.g. EDM).
  - Note there are **ideas** for follow-on experiments that won't be Muon g-2. Though far from concrete.

- I/O Needs and Schedule (see next slides)
  - Production continues and we plan to complete by end of FY23
  - FY24, FY25 are analysis and simulation only computing

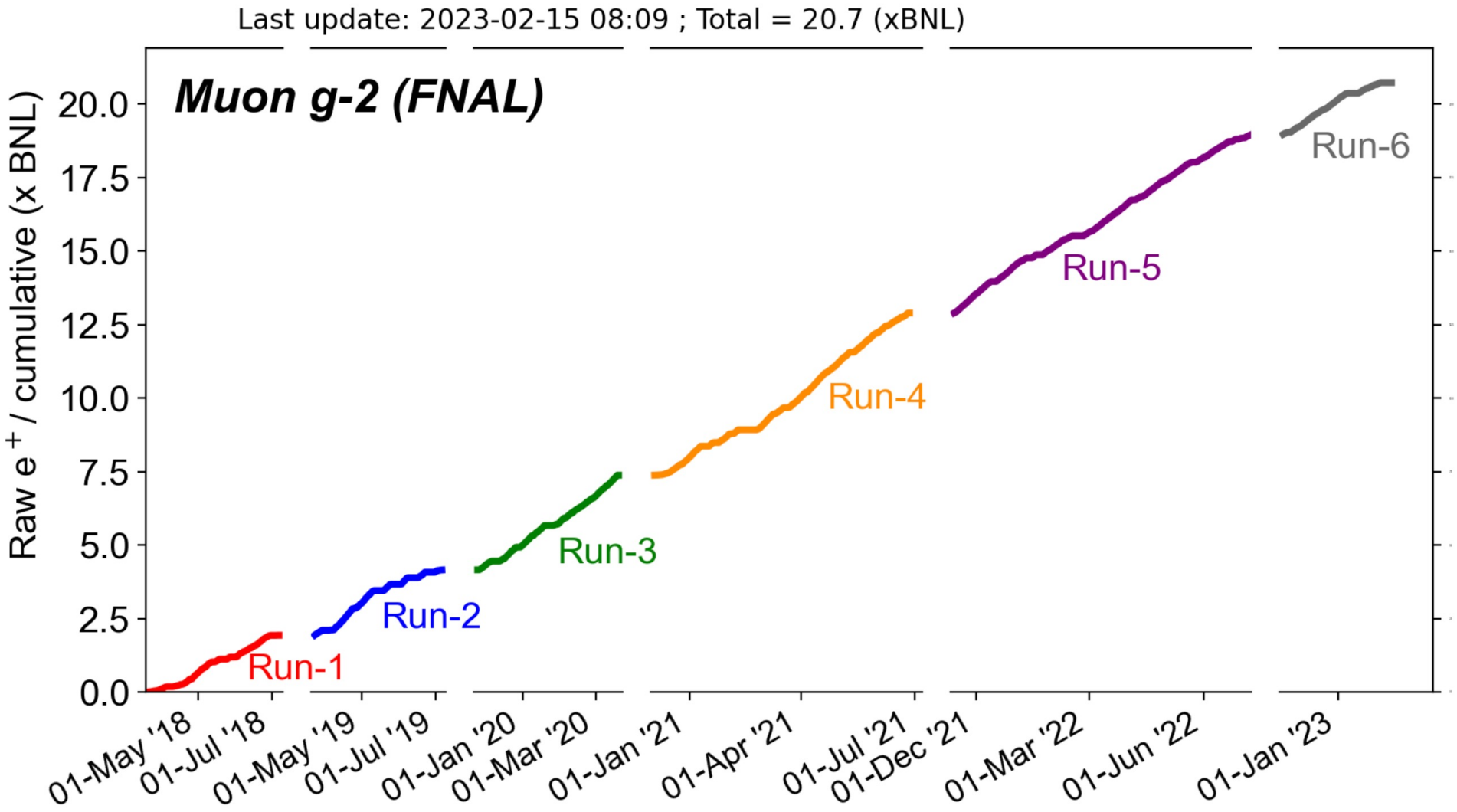
- GPUs – While we use GPUs extensively online in our DAQ, we have not found a use case where GPUs help offline
- EAF – We haven't used it much. I think newer analyzers will be interested. Will talk about this more once our next publication is complete (see last slide)

Muon g-2 Plan



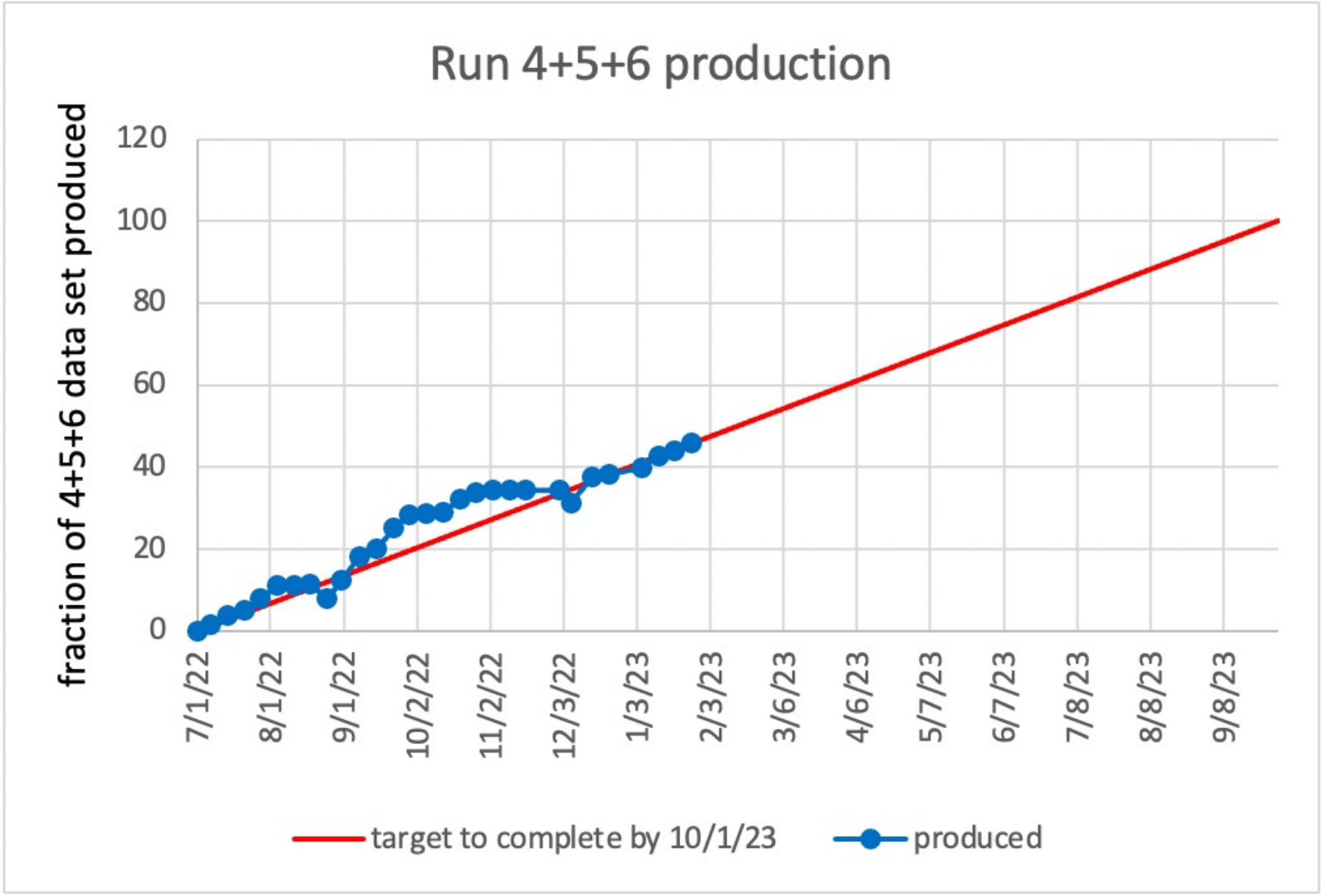
Fermilab

Running history



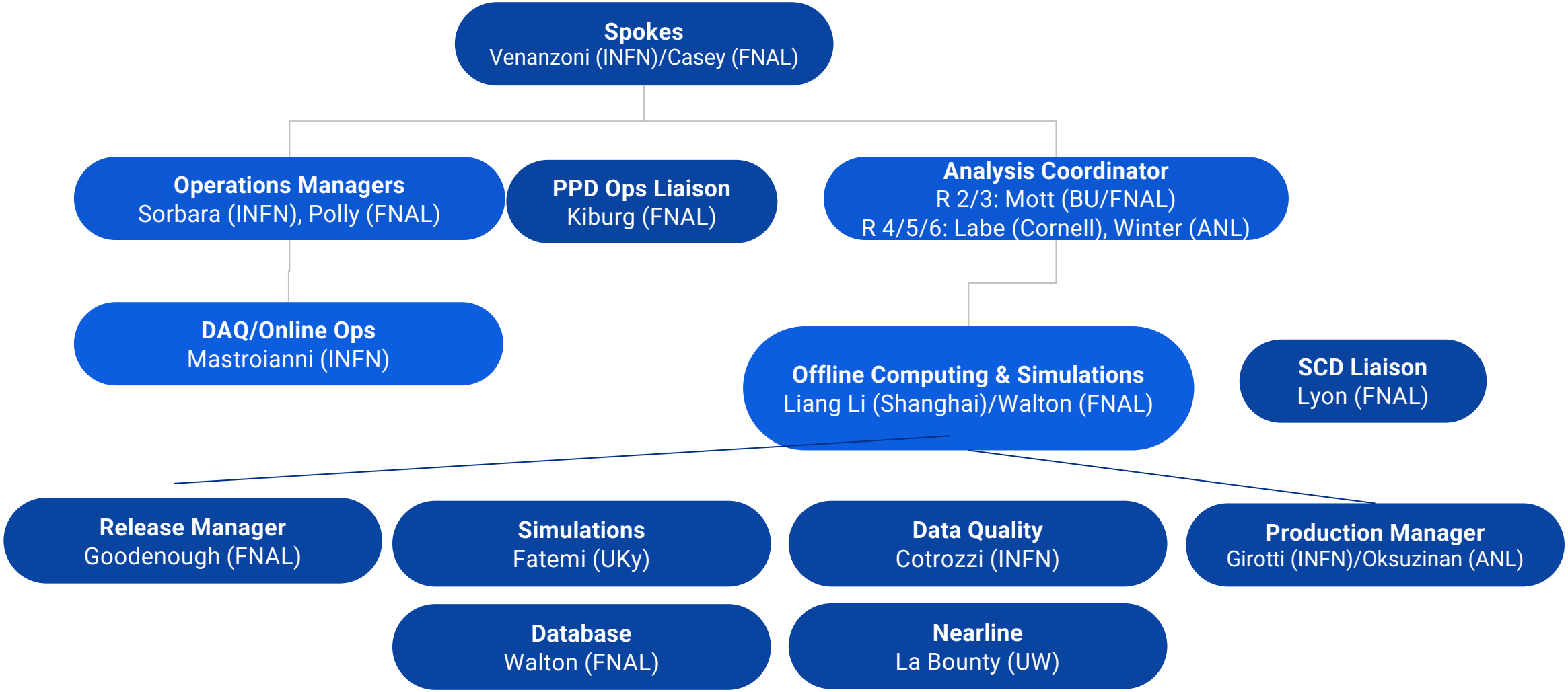
Fermilab

Production progress

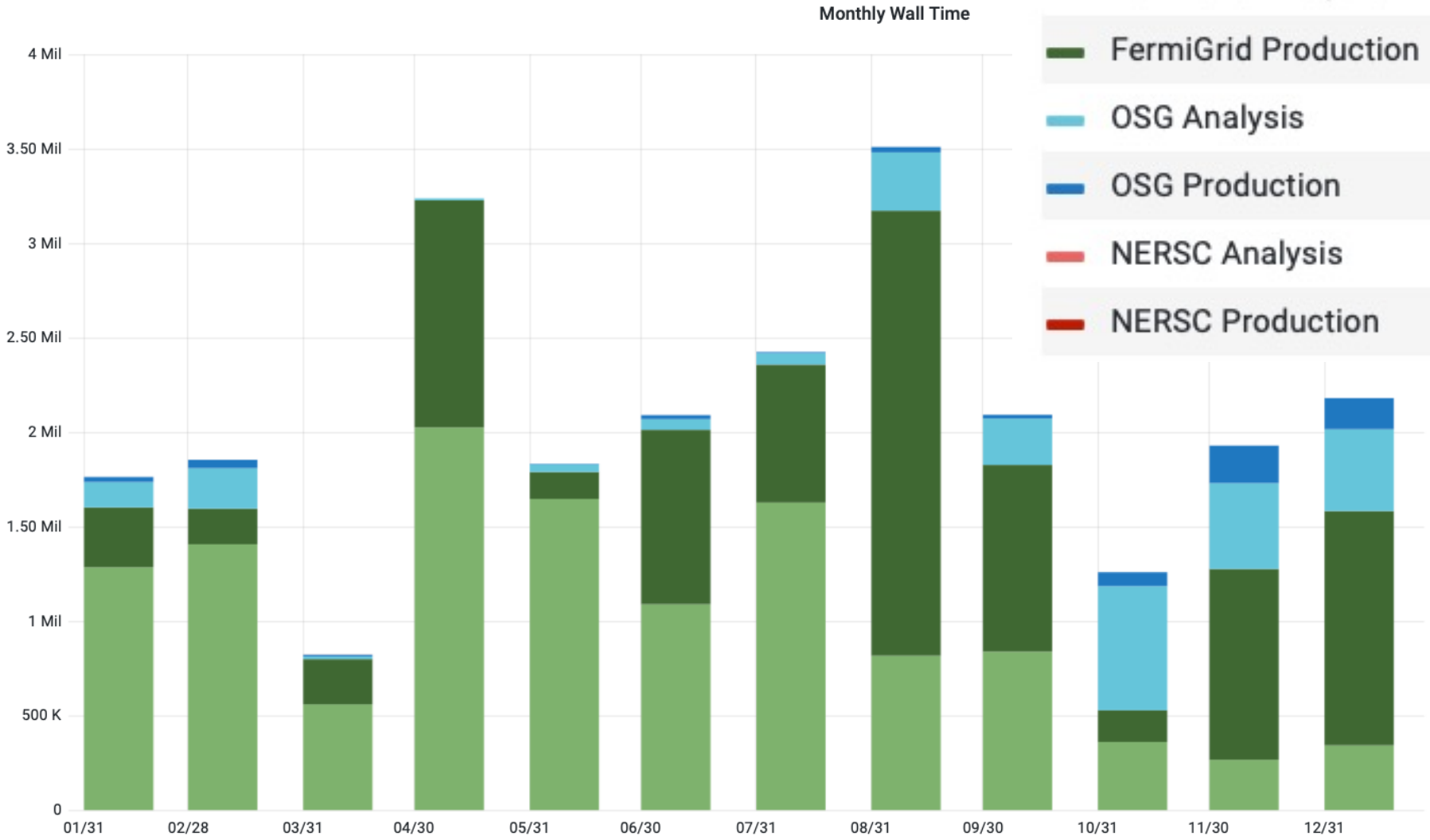


Fermilab

Experiment Organization Chart for Offline Computing



Experiment CPU Usage over the past year



	CY20	CY21	CY22
FermiGrid Analysis	23.2 Mil	17.2 Mil	13.1 Mil
FermiGrid Production	15.6 Mil	8.21 Mil	8.61 Mil
OSG Analysis	2.14 Mil	1.59 Mil	2.33 Mil
OSG Production	1.23 Mil	2.94 Mil	467 K
NERSC Analysis	0	0	0
NERSC Production	504 K	915 K	443

Fermilab

Recent production improvements

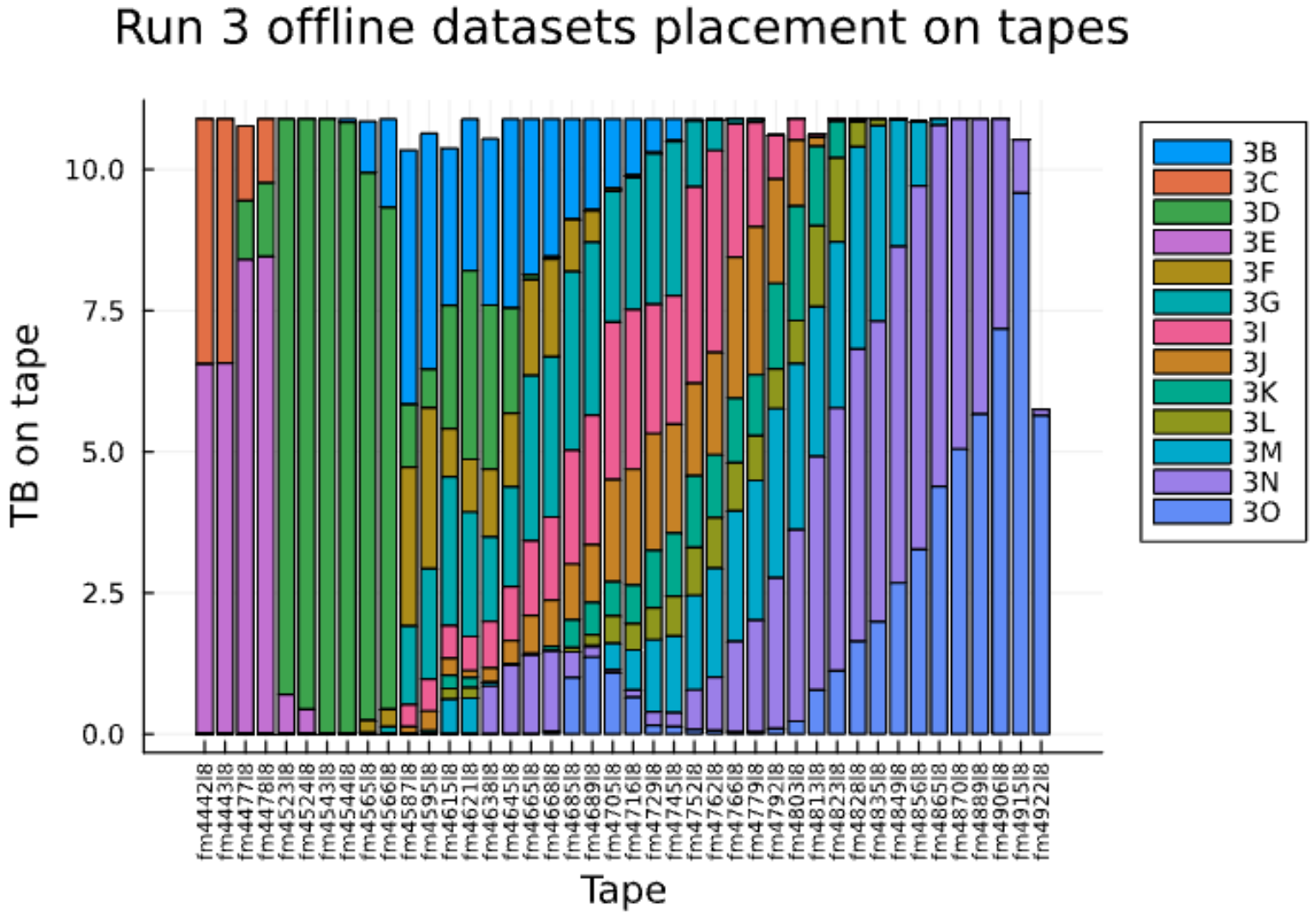
- Migration mode prestaging
  - Regular prestaging involves submitting individual files to dCache for restore
    - If the tape system is busy, this can be very inefficient
    - Enstore queue is **much** smaller than the dCache queue
    - Unable to group requests by tape. Leads to few files read per mount and many mounts per tape
  - In the tape crisis, Muon g-2 was using the majority of tape drives in the G2 robot
  - Migration mode prestaging devised to prestage large amounts of data with few resources
    - Prestage an **entire** tape at a time to a staging area
    - The data we could prestage with 30 drives we can now do with 6 !!
    - This works because we have been careful with file families – so needed data are all on sets of tapes
    - But this requires **manual** work from the dCache/Enstore team (and they've been super helpful!)

Migration Mode Prestaging

We have a script that queries the SAM database to determine the tapes for a set of datasets.

Because we read the whole tape, we may read parts of datasets we do not need.

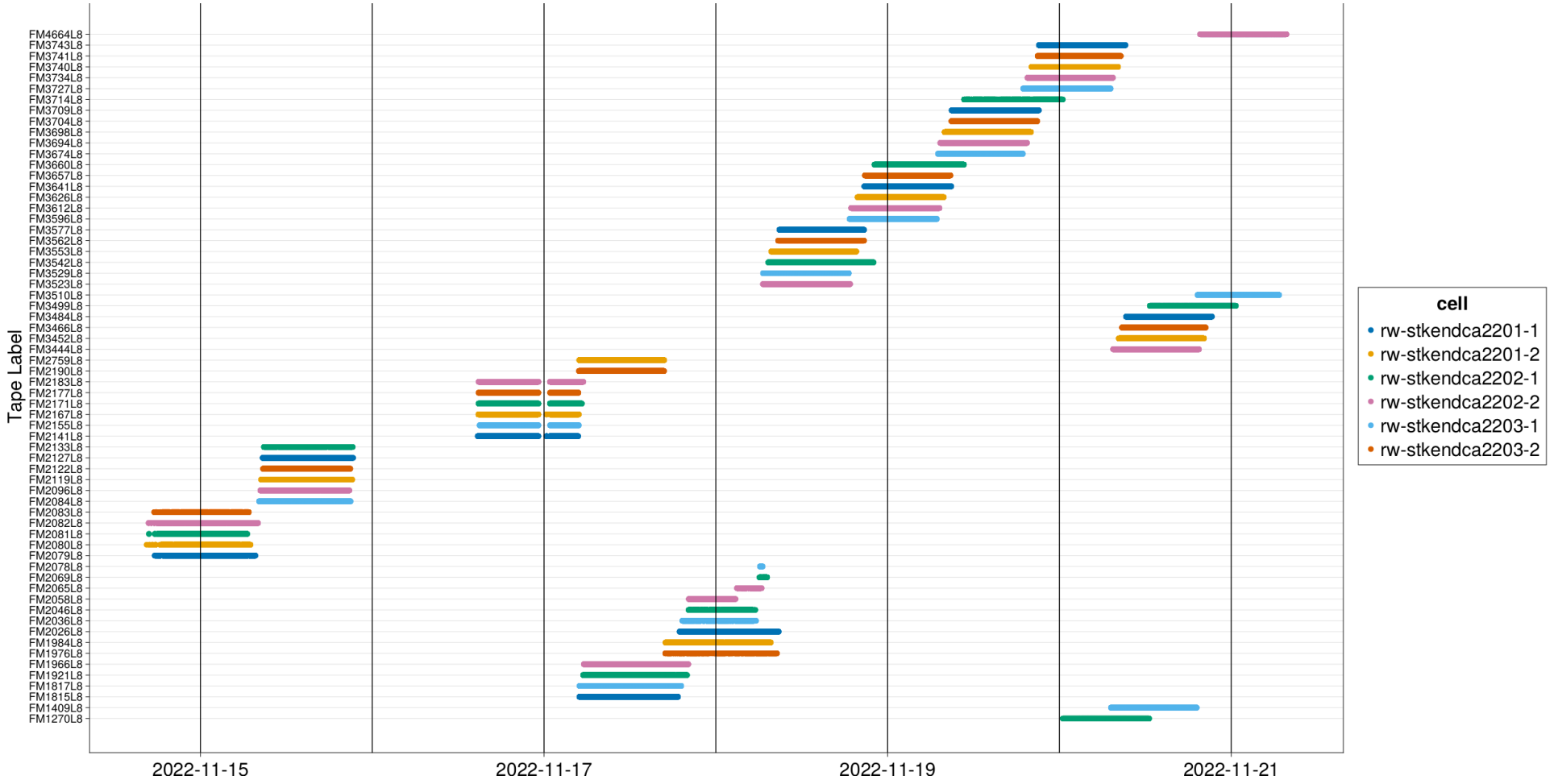
But it's likely we'll need those data later, and the script accounts for this.



Fermilab

Migration Mode Prestaging

- Whole tape reads at avg 245 MB/s
- Takes ~12.5 hrs to read an entire tape
- With 6 drives we can prestage 130 TB/day
- Max with regular prestaging was ~90 TB/day with 30 drives



Other improvements

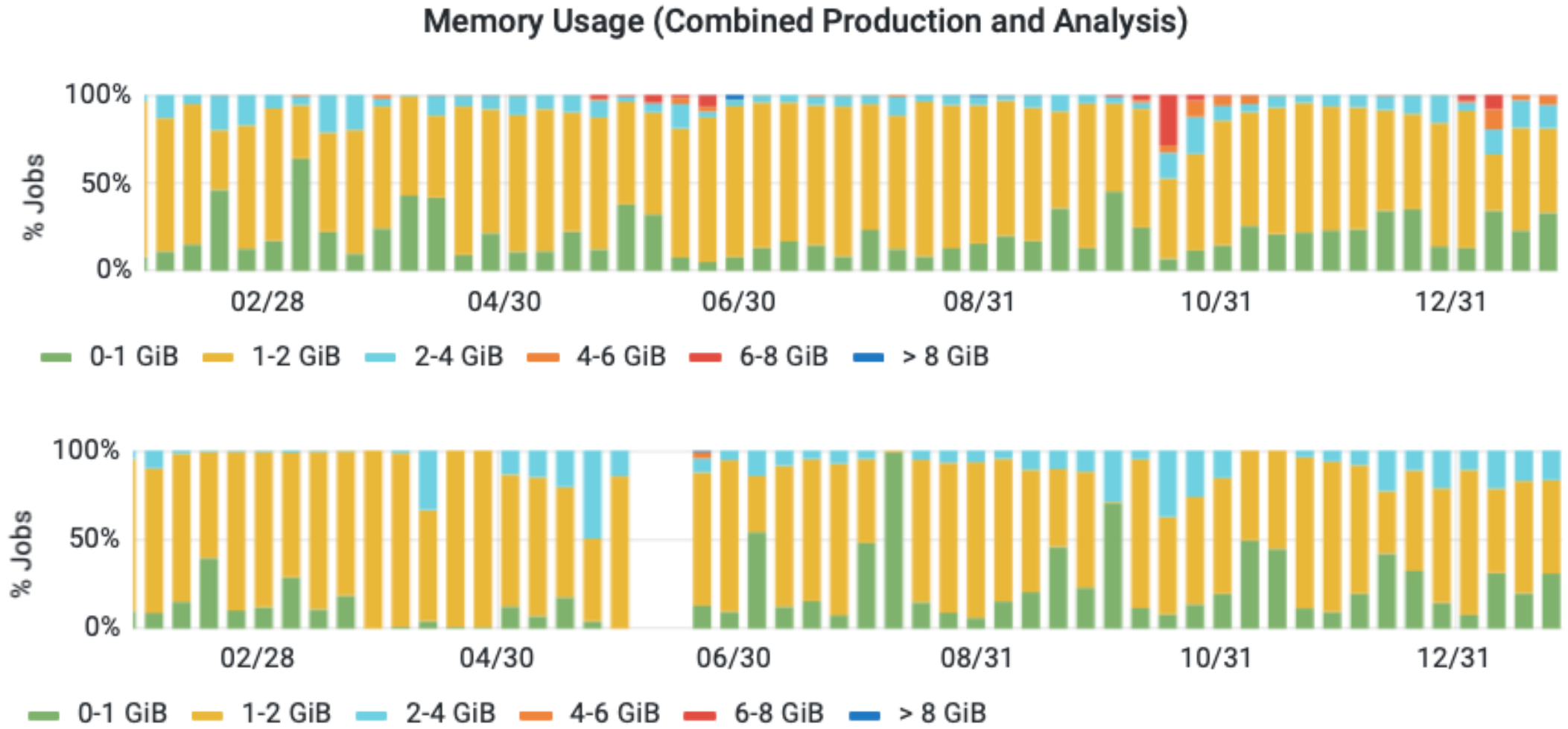
- We are currently taking Run 6 data
- We are running pre-production on this data soon after it is written to disk
  - In the past, we did not start pre-production until after the run completed
  - Doing “in-situ” pre-production allows us to skip the pre-staging step
  - We will need to pre-stage for full production

- Our Read Pool was increased to 1.5 PB to allow us to pre-stage more as we run production

- Our Write Pool was increased from 53 TB to 109 TB to increase the lifetime of Run 6 raw data to give us more time for in-situ pre-production.

Memory usage 2022

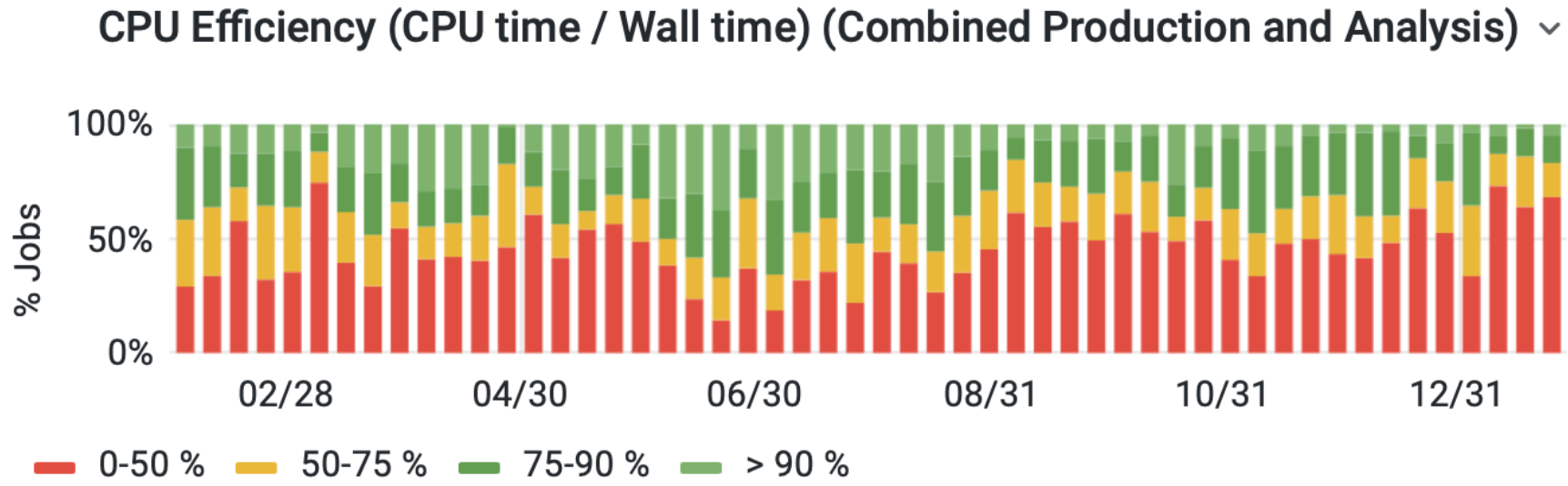
Improved compared to 2021

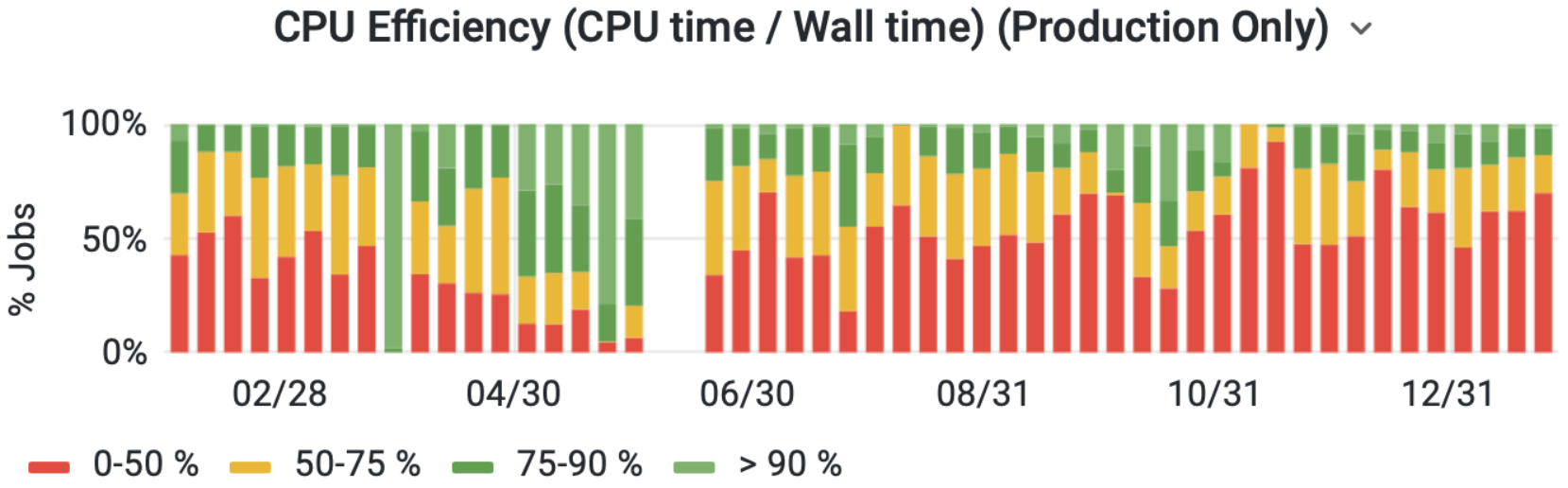


Memory is always a challenge. We add functionality to our code that increases memory, and then have to find savings elsewhere. Though most jobs run within 2 GB.

Fermilab

CPU Efficiency





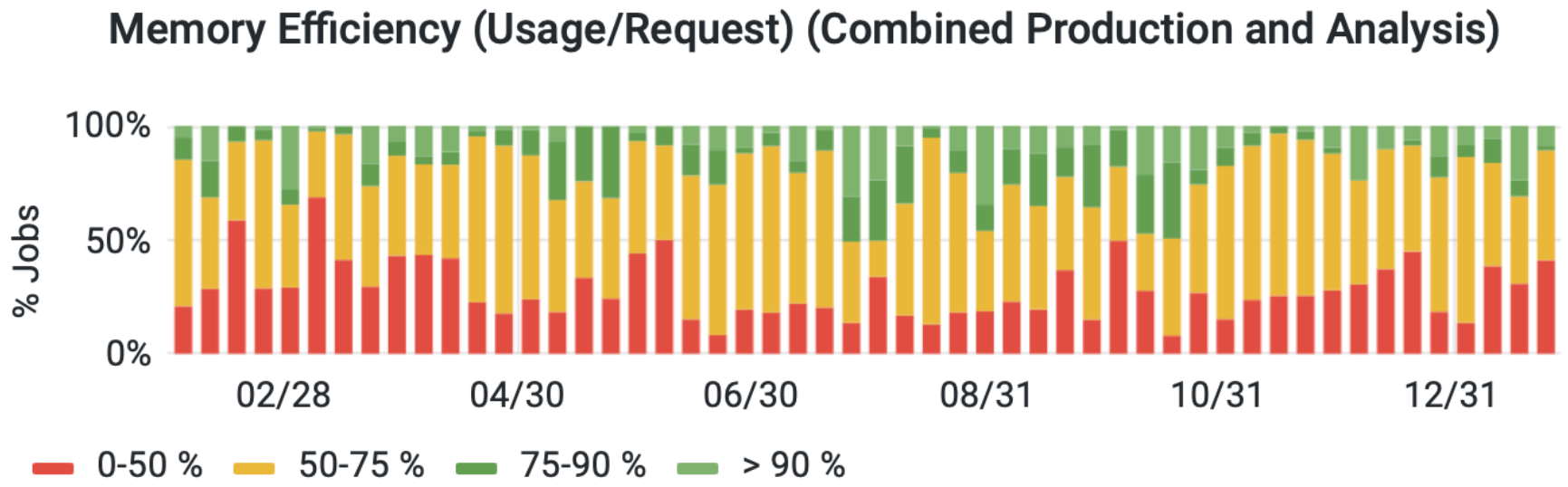
A plot weighted by length of job would look much better.

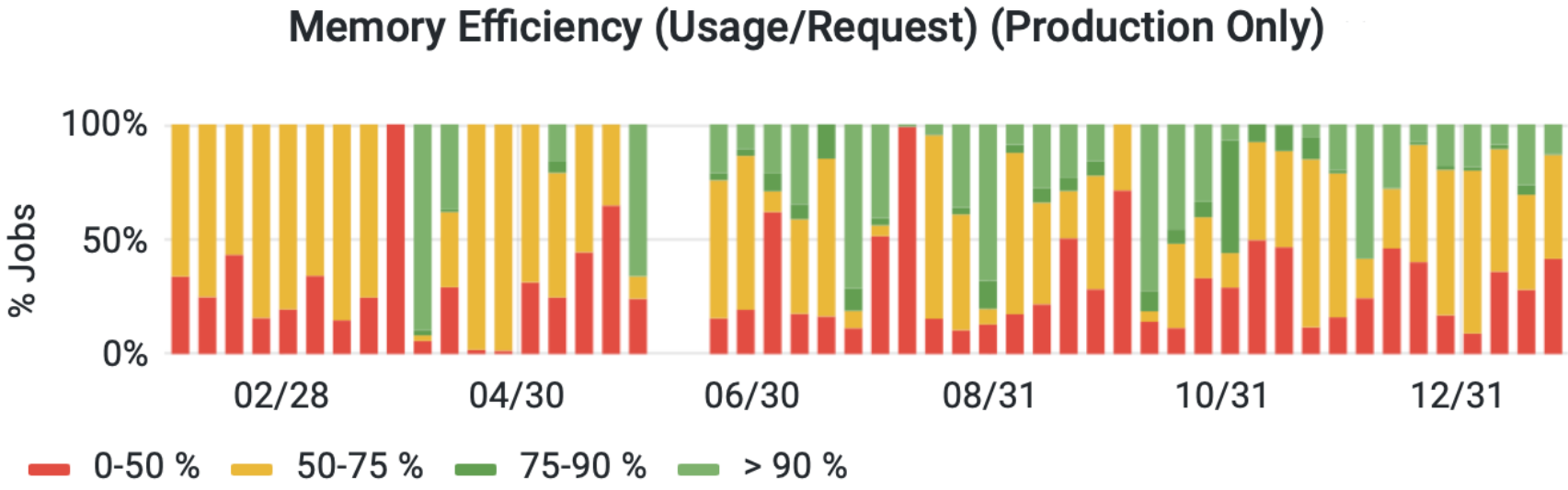
Long production jobs (e.g. 24 hrs) are very efficient.

Short production jobs are i/o bound

Fermilab

Memory Efficiency





Fermilab

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

Goals	Where does the experiment need to contribute	Where does CSAID need to contribute
Migrate away from SL7	Builds and testing	May need a waiver or container to complete Run 6 production (up to end of FY23) May need help from experts
100B event simulation at NERSC	Manage and run	We'll definitely need help from the HEPCloud team
Complete the Muon g-2 analyses	Computing expertise within the experiment will likely decrease.	CSAI may need to provide more computing assistance.
Data Preservation	Need to figure this out	Need to figure this out

Campaign Schedules

	2023	2024	2025	2026	2027
Processing campaigns (start month-end month if known). **Include when you expect to be prestaging**	Runs 5 and 6 processing Now (Jan 30) through September Prestaging throughout	Analysis and Simulation Jobs Throughout 1 month 1B event simulation campaign for Fermigrid/OSG 100B event simulation campaign for NERSC (we may start this in 2023 – we need to better plan this campaign)	Analysis and small simulation Jobs Throughout	NONE	NONE
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.**	Prestaging 1.5 PB Run 5 left + ~2 PB Run 6 FermiGrid CPU 6-7M hours Will store ~ 3 PB for output	1B Simulation campaign: ~500K CPU hours (will allow for OSG too). ~ 50 TB of storage 100B event simulation at NERSC needs to be better understood.			
Conference or result targets (month if known)	Production Goal for FY23				

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)	21 all platforms	33 all platforms	45 all platforms	60 all platforms	15 analysis/sim + 20 production = 35 FermiGrid	15 analysis/sim FermiGrid	10 analysis/sim FermiGrid		
Actual Used	29	39	25	22	N/A	N/A	N/A	N/A	N/A
Efficiency		70%	70%	70%	N/A	N/A	N/A	N/A	N/A

Looking for five-year projections this cycle

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)	21 all platforms	33 all platforms	45 all platforms	60 all platforms	3 production + analysis	3 analysis/sim	2 analysis		
Actual Used	??	3.3	4.5	2.7	N/A	N/A	N/A	N/A	N/A
Efficiency	%	%	%	%	N/A	N/A	N/A	N/A	N/A

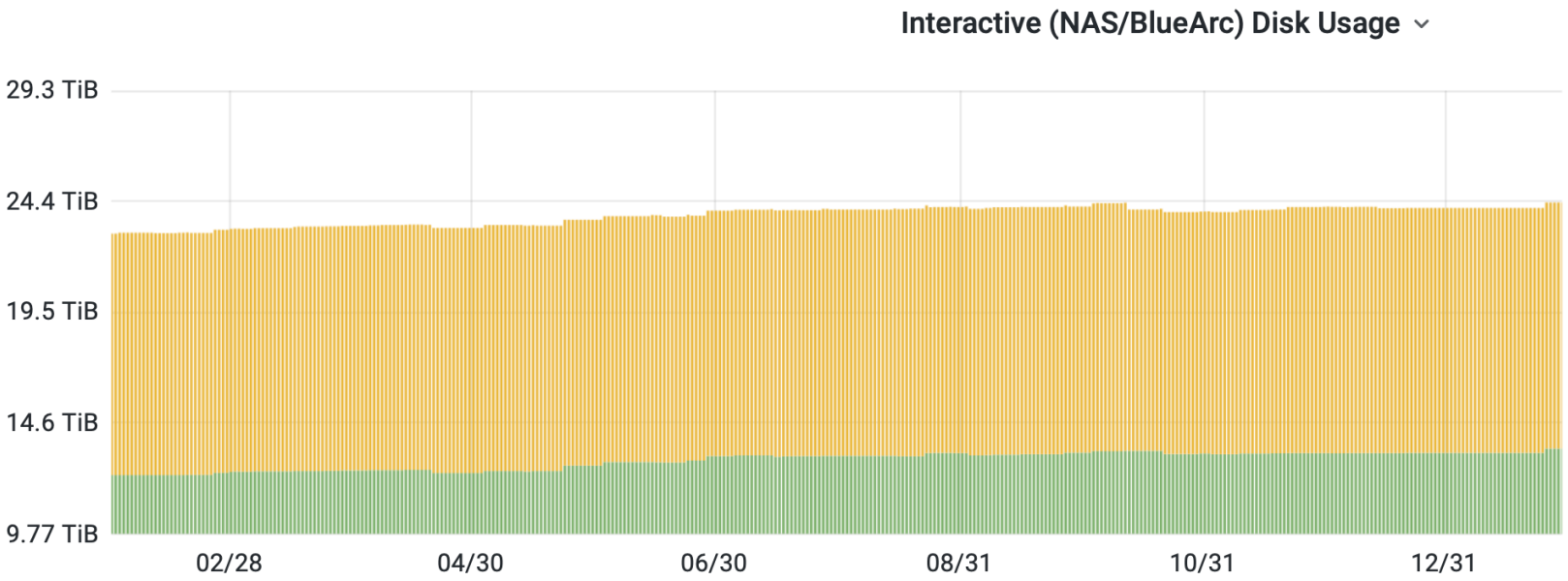
Looking for five-year projections this cycle

CPU – HPC 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)					100B event simulation campaign needs to be understood	100B event simulation campaign needs to be understood			
Actual Used		0.5	0.9	0.5	N/A	N/A	N/A	N/A	N/A
Efficiency	%	%	%	%	N/A	N/A	N/A	N/A	N/A

Looking for five-year projections this cycle

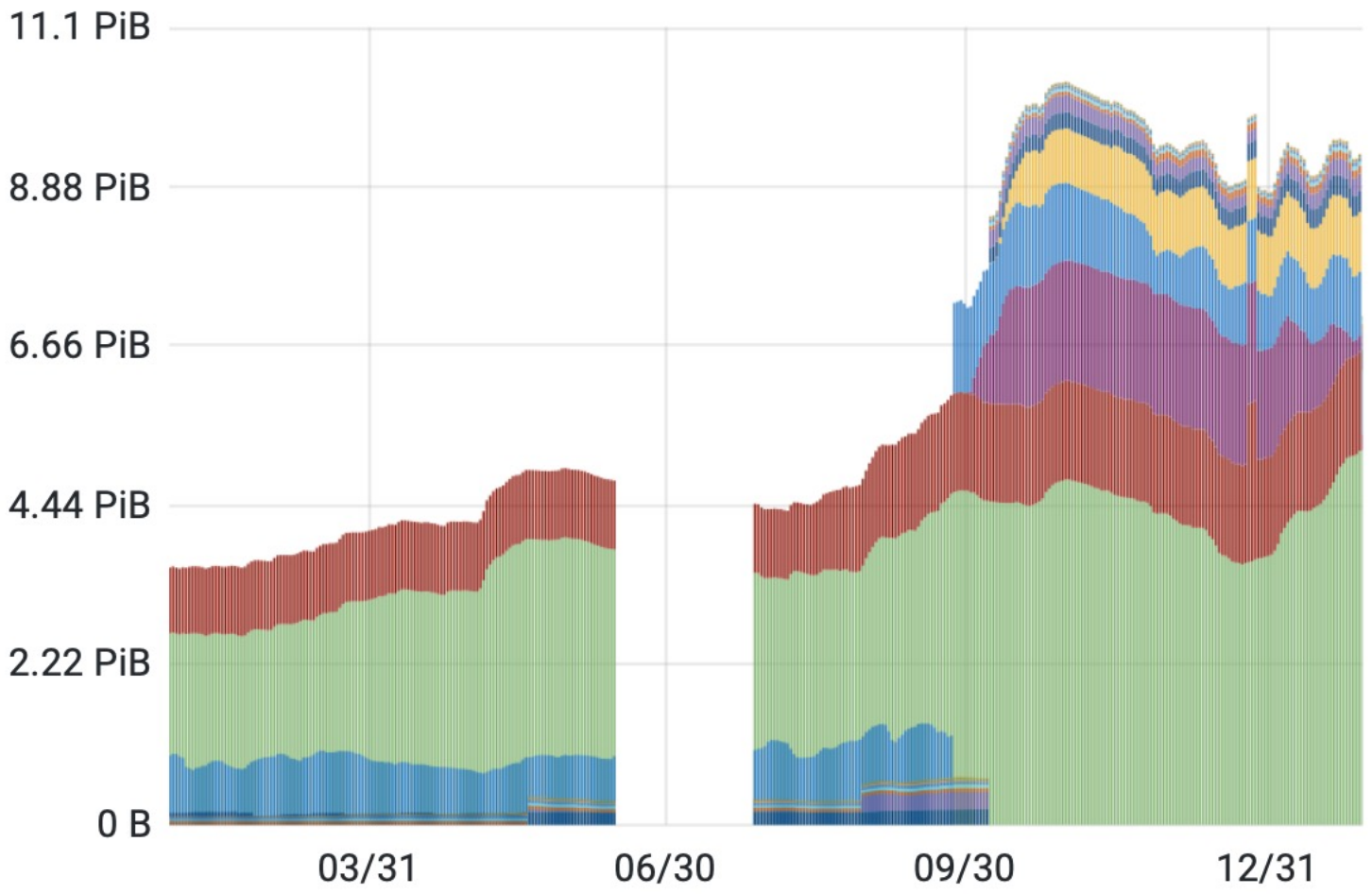
NAS Usage and Projections



	App	Data
2022	10.8	13.5
2023	12	15
2024	15	20
2025	15	20
2026	15	20
2027	0	0

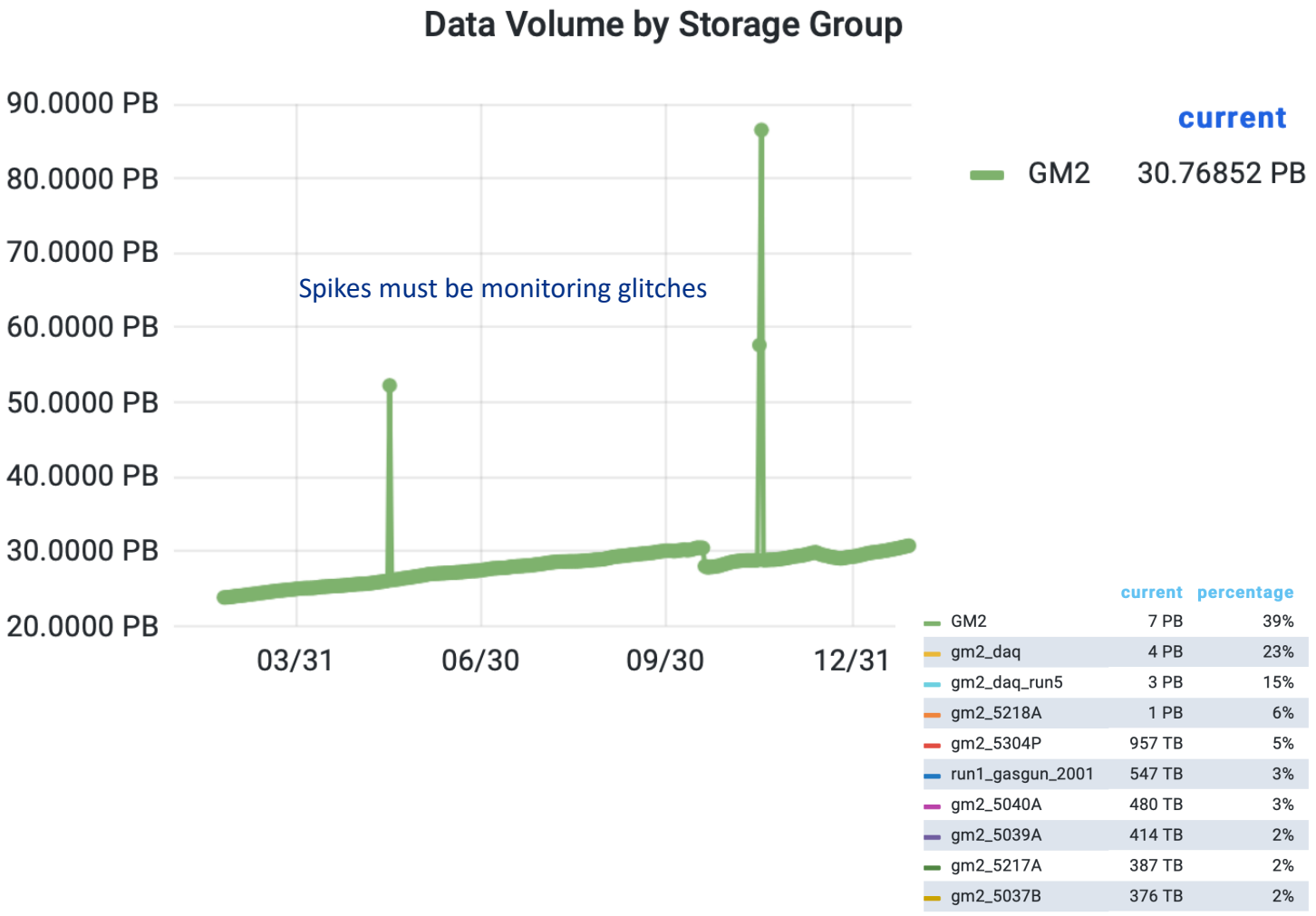
App is important for building analysis code and for launching jobs.
Data is important to supplement persistent dCache space

dCache Usage and Predictions (in TB)



	Analysis (Persistent)	Write Pool	Read Pool	Stage Pool
2022	128	53	940	850
Current	377	109	1400	850
2023	600	109	1400	850
2024	800	0	700	850
2025	800	0	700	850
2026	800	0	0	0
2027	0	0	0	0

Tape usage and predictions (in TB)



	current	percentage
GM2	7 PB	39%
gm2_daq	4 PB	23%
gm2_daq_run5	3 PB	15%
gm2_5218A	1 PB	6%
gm2_5304P	957 TB	5%
run1_gasgun_2001	547 TB	3%
gm2_5040A	480 TB	3%
gm2_5039A	414 TB	2%
gm2_5217A	387 TB	2%
gm2_5037B	376 TB	2%

	Total Added By End of Year
At end 2021	21.31 PB
2022	+7.5 PB requested +9.4 PB actual
2023	Run 6 Raw = 2 PB Run 5 Production = 3 PB Run 6 Production = 1.8 PB Sim/Analysis = 1 PB Total = +7.8 PB
2024	Sim/Analysis = +1.5 PB
2025	Sim/Analysis = +0.5 PB
2026	0 PB
2027	0 PB

Fermilab

Data Lifetimes

Every production/simulation version has its own file family

We can and have recycled obsolete production data

Analysis Facility

We haven’t used the EAF yet (not a good time to revamp analyses before Run 2/3 publication). Once the publication is done, people may be interested. The Run 4/5/6 dataset is large, even when skimmed and tuple-ized. Probably ~ 1 PB. Likely will only need CPUs