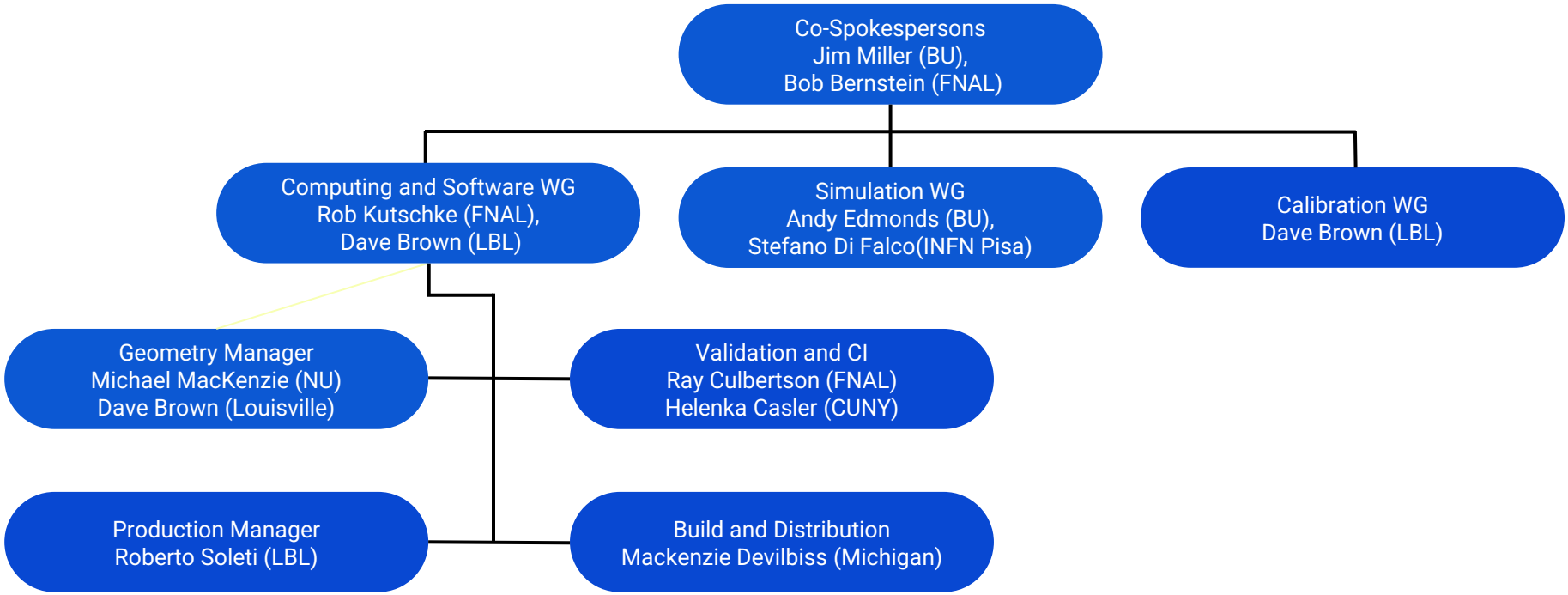
Mu2e FCRSG 2022

Mu2e-doc-42310-v3

Rob Kutschke, Fermilab SCD

June 1, 2022

Mu2e Organization Chart for Offline Computing

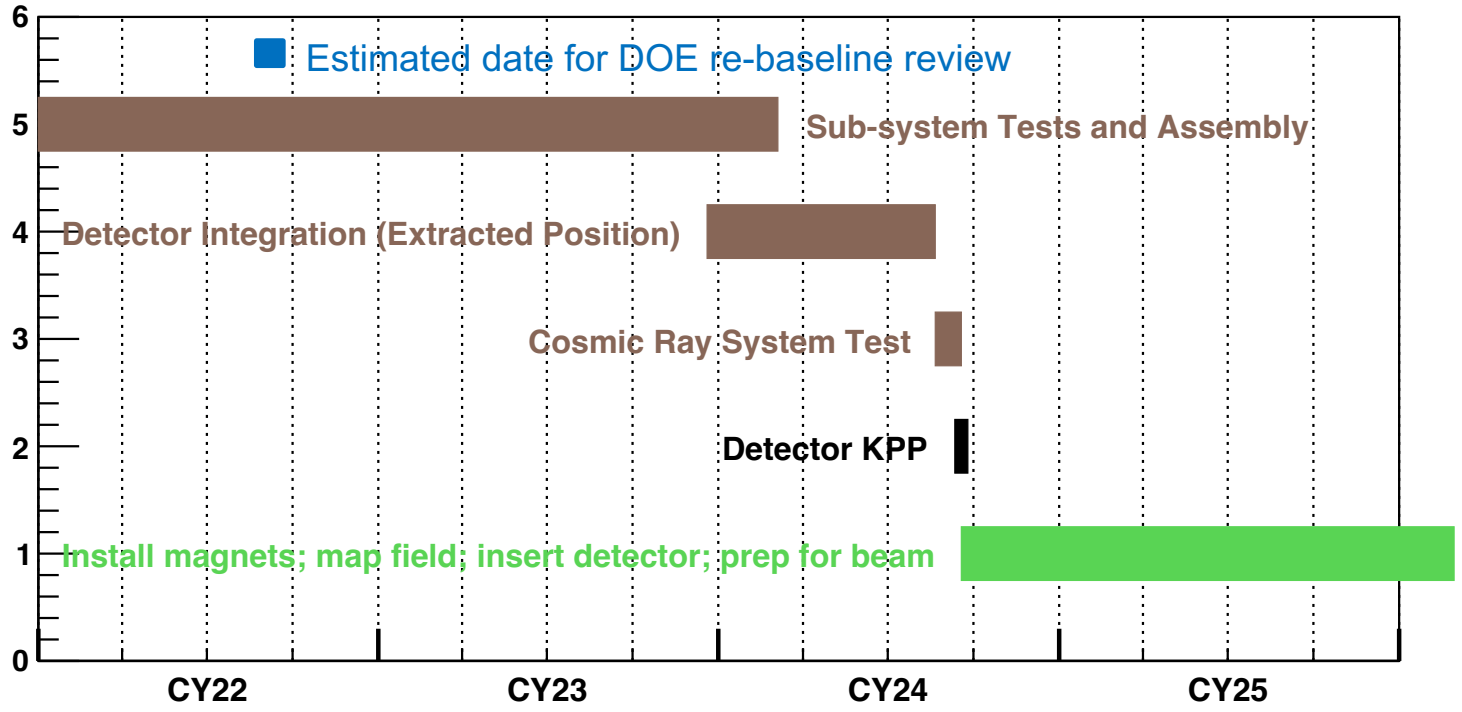


- Other roles filled transiently as needed
- Continue to lose people to higher priority on-project tasks (detector construction)
- Expect to staff up after detector KPP in late CY 2024
- Algorithm development is the responsibility of the subsystems; some people wear both hats.

Highlights of the Past Year

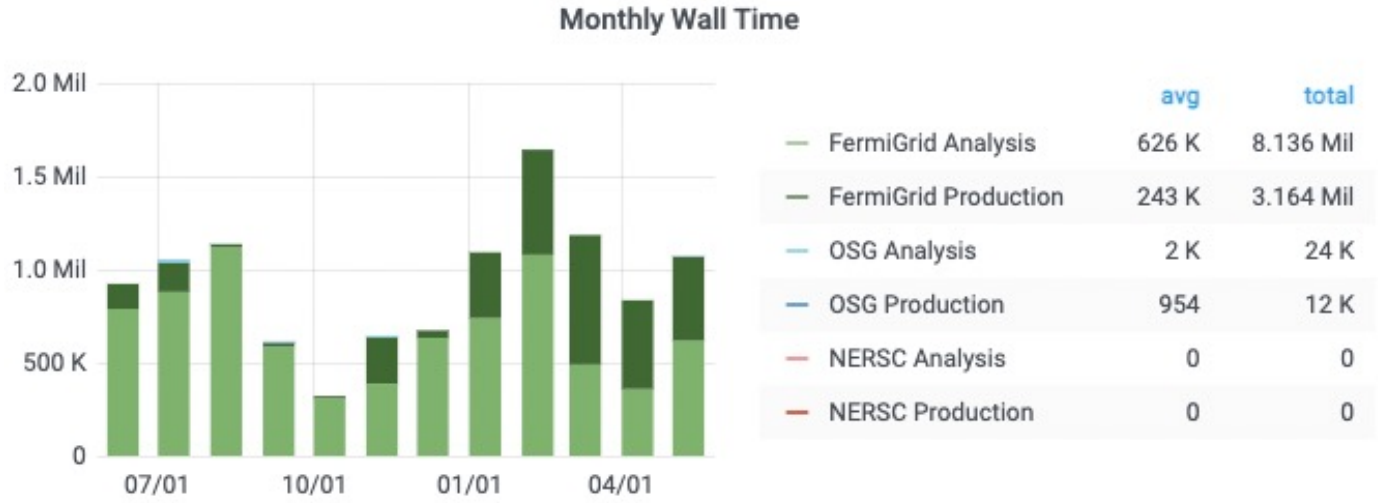
- Sensitivity Update (SU2020) paper is nearing the end of internal review
  - Final use of samples that we generated starting in 2015
- Main body of MDC2020 is complete
  - Miscalibrated, misaligned datasets for use in algorithm development
    - 3 variants: perfect, best, startup
  - Still to come:
    - Additional signal-like background sources
    - More detailed and realistic models of miscalibration and mis-alignment
- Support for Vertical Slice Tests and test stand measurements
  - First real data in our Conditions DB
- Routine use of Conditions DB in both sim and reco.
- Ongoing simulations to understand backgrounds from cosmic rays.
- Simulations to support Mu2e-II work to be shown at Snowmass

Preliminary Mu2e Revised Schedule



- Peak usage driven by simulation;
- Hard to predict timing – we run when code is ready.
- Low volume test stand and cosmic ray data throughout this period.

CPU Wall Hours: Mu2e Usage Over the Last Year

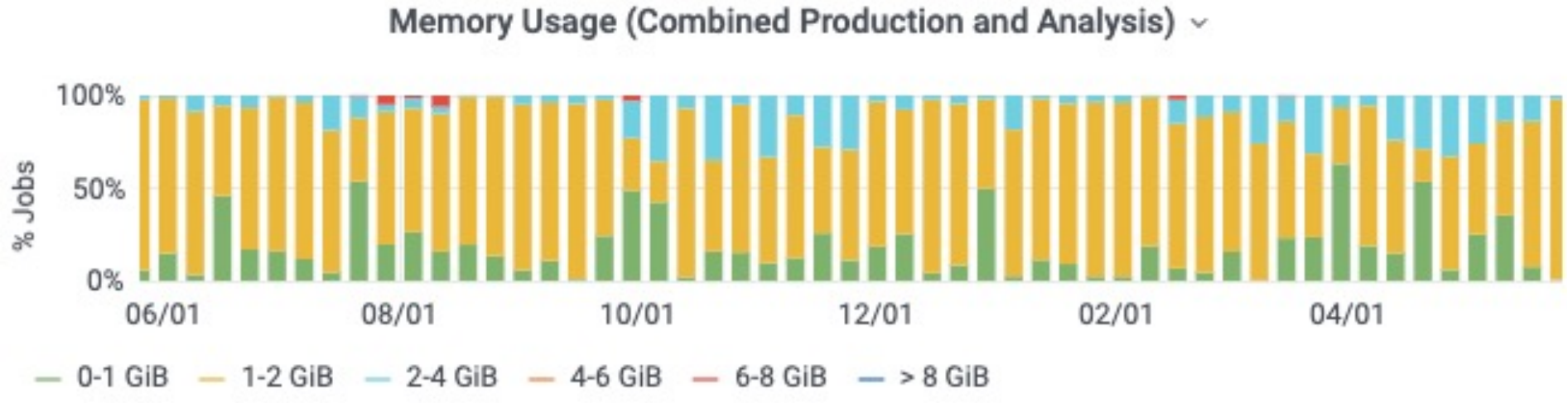


	avg	total
FermiGrid Analysis	626 K	8.136 Mil
FermiGrid Production	243 K	3.164 Mil
OSG Analysis	2 K	24 K
OSG Production	954	12 K
NERSC Analysis	0	0
NERSC Production	0	0

- Have run on OSG in the past; stopped this year due to frequent job failures.
- Plus 12.2M KNL-core hours on THETA/BeBop at ANL (G4MT); managed using ANL tools
- Mu2e G4MT workflow being used to develop HEPCloud access to Theta.
- Only G4 code and some generators are MT ready.
- Most simulations have elastic deadlines; some exceptions.

<experiment>

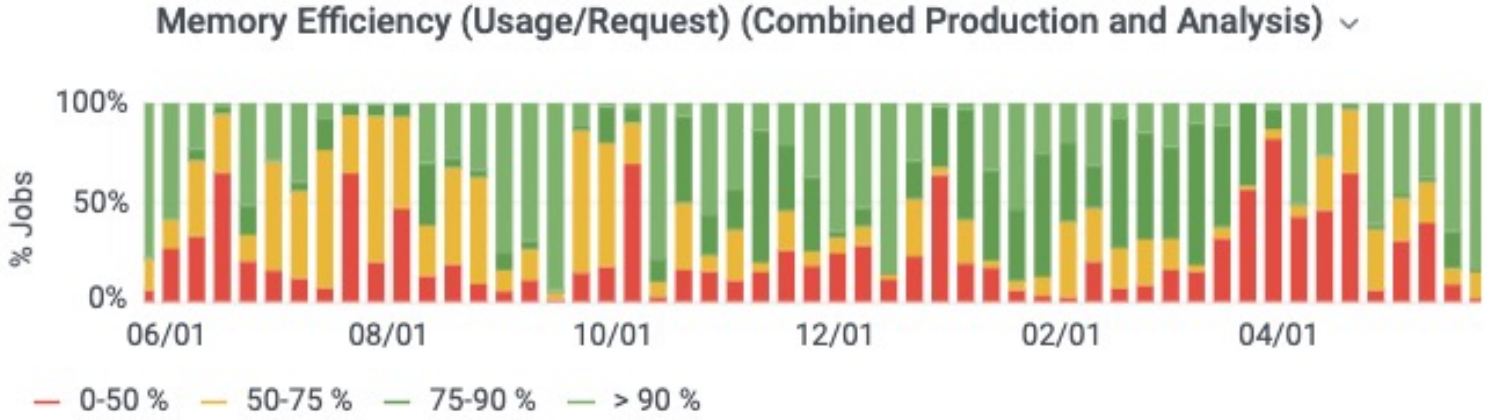
Memory Footprint Over the Last Year



- Routinely run G4 jobs with 2 slots and 2 threads in under 3 GB.
- Trigger and Reco jobs < 2 GB
- Pile-up modelling jobs; a few years ago >> 2GB; now <~2 GB
- Redesigned data products and workflows
- Cosmic ray jobs have a tail to very high memory use for jobs with high energy particles

Mu2e CPU and Memory Efficiency Over the Last Year

Small improvement from last year.
Some red are pileup jobs – I/O bound.



Big improvement from last year.
Will continue user education.

<experiment>

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

	2018	2019	2020	2021	2022	2023	2024
Requested	11	15	11	11	11	15	15
Actual Used	12.1	11	10.9	10.5	4.8 YTD	N/A	
Efficiency	65%	86%	90	81%	78% YTD	N/A	

<experiment>

CPU Adaptations Going Forward

Last summer+fall: low availability of OSG cycles plus high job failure rate. Stopped using it.

No Mu2e funding for commercial cloud resources

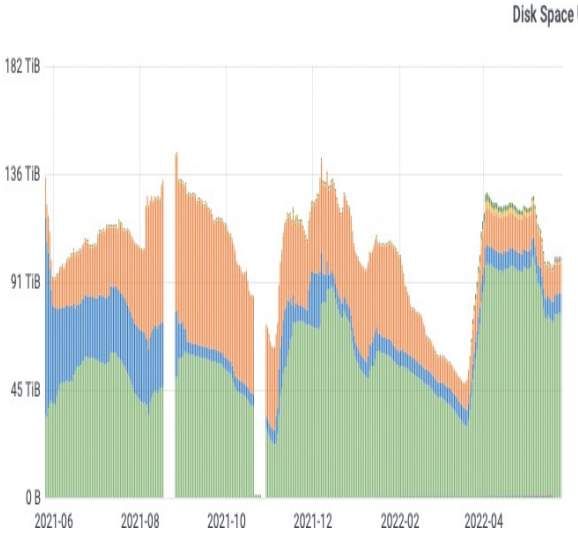
Director's HPC allocations at ANL (14M hours in 2022) for Mu2e collaborators who are ANL employees. Most often used for CRV studies (ANL institutional responsibility).

Development work to use some of FNAL allocation at NERSC via HepCloud (1.5 Mhours)

Can request NSERC allocations through LBL collaborators but have not done so.

Do not expect to use GPUs or other heterogenous resources (no people to do the work).

Disk: dCache Usage and Predictions (in TB)



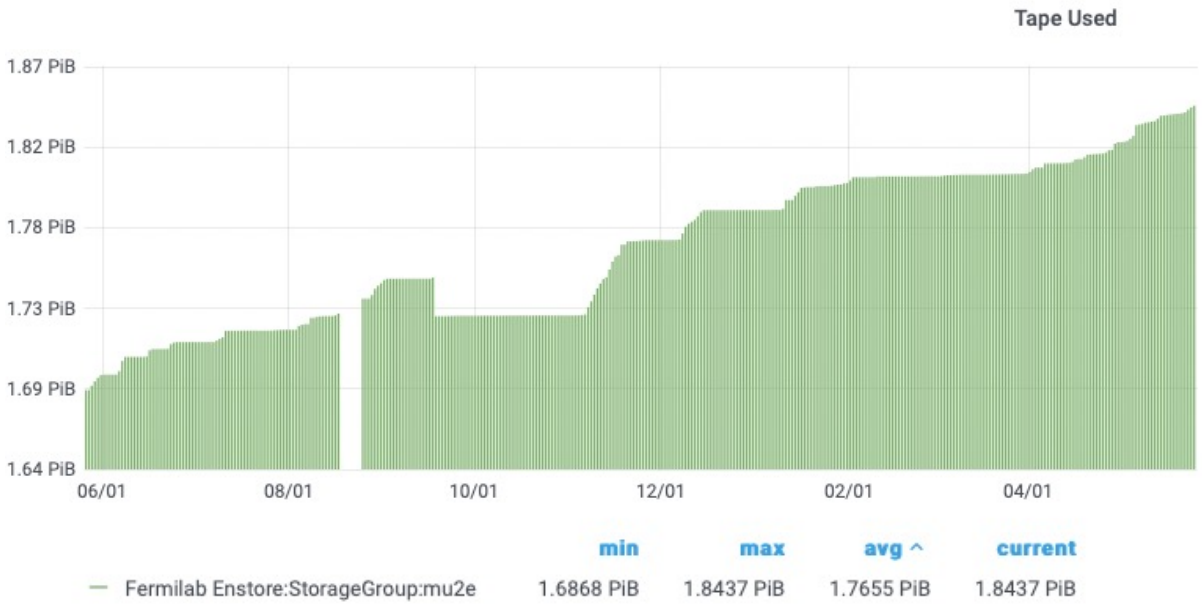
	min	max	avg	current
Fermilab Public dCache:StorageGroup:mu2e_readWritePools	1.2 TiB	101.7 TiB	59.3 TiB	77.1 TiB
Fermilab Public dCache:StorageGroup:mu2e_Mu2eAnalysisPools	5.0 TiB	83.1 TiB	13.7 TiB	8.5 TiB
Fermilab Public dCache:StorageGroup:mu2e_PublicScratchPools	9.8 TiB	65.9 TiB	30.5 TiB	12.6 TiB
Fermilab Public dCache:StorageGroup:mu2e_Geant4ReadWritePools	93 GiB	3.5 TiB	746 GiB	914 GiB
Fermilab Public dCache:StorageGroup:mu2e_SlowReadWritePools	93 GiB	3.5 TiB	756 GiB	914 GiB
Fermilab Public dCache:StorageGroup:mu2e_CdfWritePools	1.7 TiB	1.7 TiB	1.7 TiB	
Fermilab Public dCache:StorageGroup:mu2e_NovaPrestagePools	181 GiB	609 GiB	442 GiB	181 GiB
Fermilab Public dCache:StorageGroup:mu2e_HsmWritePools	2 GiB	470 GiB	156 GiB	470 GiB
Fermilab Public dCache:StorageGroup:mu2e_SlowPublicScratchPools	165 GiB	259 GiB	167 GiB	259 GiB
Fermilab Public dCache:StorageGroup:mu2e_HsmReadPools	1 GiB	246 GiB	104 GiB	236 GiB

For Reference:
Total r/w (tape backed): 6264 TB
Total scratch: 2333 TB
Total persistent: 2576 TB
Total other: 2131 TB
<experiment>

	Analysis (Persistent)	Other Dedicated
Current	148 TB (actual)	0
2022	148	*20? (Read)
2023	148	20 (Write) 20 (Write)
2024	148	20 (Read) 60 (Write)

* Request is in preparation

Tape - Usage and Predictions (in PB)

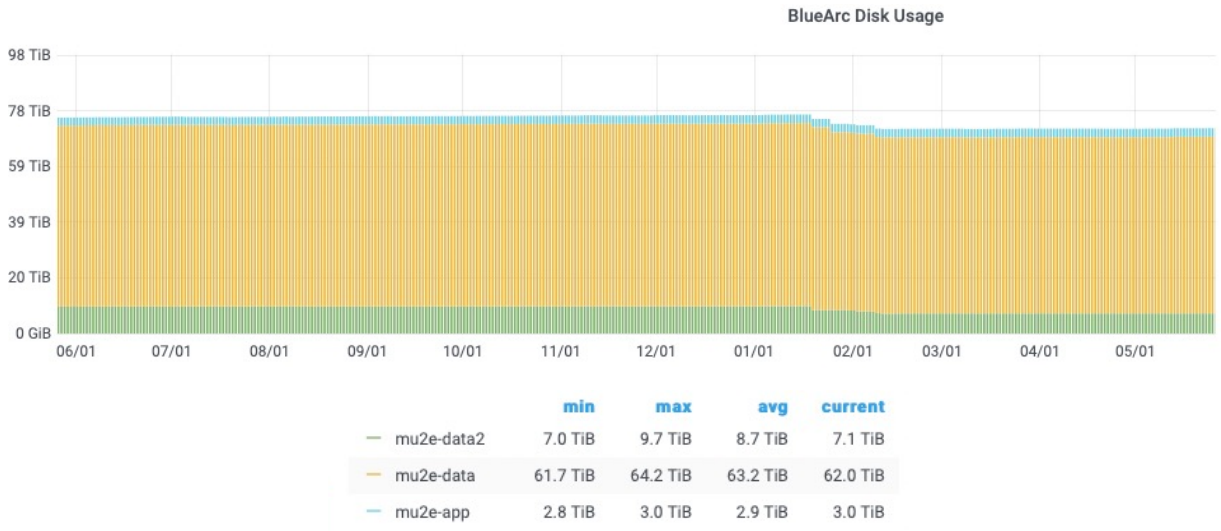


- Total above includes > 1 PB on T10K not migrated to LT08.

	Total Added By End of Year
At end 2021	0.5 PB
2022	0.04 PB YTD 0.2 PB Estimated
2023	0.3 PB
2024	0.4 PB

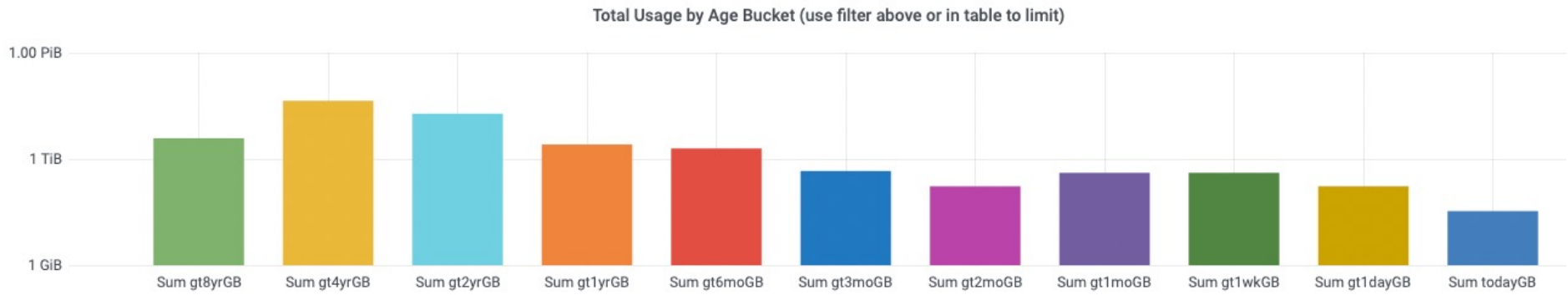
<experiment>

Disk: NAS Usage and Predictions (in TB Units)



	App	Data
2022	**3**	**80**
2023	**3**	**80**
2024	**3**	**80**

Age of files in NAS



- We are aware that this is an issue
- I started herding cats in January but ran out of energy.
- We will clean up before asking for significantly more space.

Data Lifetimes

- In 2020/21 Mu2ed retired ~1.1 PB of tape (~82% of total used at that time). These were not migrated to LT08.
- Simulated datasets: retire when successor is established; estimate 3-4 years.
- Will develop a plan for experimental data prior to start of data taking.
  Strawman plan:
  - Raw data keep until N years after last paper; N to be negotiated.
  - Derived data sets keep current generation plus previous 2.

What Do You Want to Achieve in Computing Over Next Three Years

Goals	Where does the experiment need to contribute	Where does SCD need to contribute
Big picture: Be ready to take and process cosmic ray data in 2024 and beam data in 2026.	See below for the components of this goal.	
Commission DAQ, Trigger, slow control	Trigger algorithms; detector control system; online DQM; configuration of the overall system.	Ryan Rivera's group works on everything except trigger algorithms.
Develop data-logging workflow	Requirements, joint effort on design, implementation	Consulting; add features to SCD tools as needed
Upgrade data processing workflows	Requirements, joint effort on design, implementation	We may request new features in POMS and related tools

- About the transition from SAM to to RUCIO / Metacat / Data Dispatcher
  - We have been advised, in the above work, to encapsulate the SAM dependence using tools supplied by SCD so that the backend of the tools can can swapped with minimal changes on the user facing side.
  - We will need support to understand how to do this.

What Do You Want to Achieve in Computing Over Next Three Years

Goals	Where does the experiment need to contribute	Where does SCD need to contribute
EVE7 Event display	Requirements, design, evaluate options, implementation	If SCD is offering to support event displays we are interested.
Offline DQM System	Requirements, design, implementation.	Will need some db support.
Calibration/Alignment Algorithms	Most of the work	See below.

- We will continue to want ongoing consulting from SCD on design issues, best practices, better use of profiling tools, code reviews etc on both code development and workflow development.

In all of the above, we will understand what other FNAL experiments do.

Anything else?

- There remain ongoing dCache reliability issues; we would like them addressed.
- We have been using CMS BOT for GitHub ←> Jenkins communication
  - It has proven extremely valuable
  - Support status is incidental consulting
  - We would like it, or it's successor to have official support.
- We been using home brew solutions to access ANL HPC
  - We look forward to ANL HPC being onboarded to HEPCloud
  - Our workflow is being used for development.
- Important postdocs will soon be moving on; we do not anticipate there will be an overlap with successors for skills/knowledge transfer.

<experiment>

Updates to the Mu2e Computing Model - 1

- Was presented at FCRSG 20201: https://indico.fnal.gov/event/47845/
- What's new?
  - Dates for steps in resource usage have been pushed back by about 18 months
  - Work started on data logging and Offline DQM – still early days.
  - DAQ buffer disks now in-hand
    - Will measure performance and identify any bottlenecks or resource limitations.
  - Last year we said that data from the Trk+Cal would be written to one file but data from the CRV would be written to a separate file.
    - Why: concerns about latency and pipeline depths in TDAQ
    - Would require a processing step to put them together in one file.
      - Cascading requirements for handling of corner cases
    - Now we expect the data from all 3 system to be written to a single file
    - The two-file scenario remains as a risk.

Updates to the Mu2e Computing Model - 2

- Will use the output of MDC2020 to update estimates for:
  - Trigger rejection
  - Trigger pre-scale factors for calibration and monitoring triggers
  - Event sizes on-spill and off-spill
  - CPU time needed for calibration and reconstruction
  - Needed size of persistent dCache
  - Many other topics out of the scope of this talk.