FCRSG for DES and Rubin 2023

Brian Yanny, Liz Buckley-Geer

Feb 15, 2023

Summary plan for DES, Rubin/LSST going forward:

- DES will wind up required science results by about end CY 2024. As host institution for DES, DES has responsibility, supported by DOE, to keep ‘cosmology, Dark Energy science outputs’ available and accessible for a period (5 years, say), first for collaboration then for public.
- FNAL resource $ **are** currently available to host disk (say dcache, with tape copy) archive of DES data products with index support (limited database) and network access for, say transfer to NERSC for further analysis. Currently working to setup details of this arrangement. We understand miniboone has done a similar thing (on a smaller scale), and we can follow that model.
- Rubin does not formally use FNAL compute resources, but up to about 8 FTE Compute professionals and scientists are paid for on the project, some shared with FNAL, some 100% on Rubin. There have been informal arrangements between FNAL and Rubin (USDF based at SLAC): loan of 1.2PB dcache space until USDF gets tape archive up (few more weeks needed) and small test VM for Rucio devel.

Experiment Organization Chart for Offline Computing DES

Brian Yanny
DESDM Project Scientist

CS Liaison – Liz Buckley-Geer

Analysis

COSMOSIS
Marc Paterno
Jim Annis

Photo-z
Huan Lin

DELVE
Alex Drlica-Wagner

Data Access

Database
A ManningNCSA)
Chris Pond(NCSA)

Science Portal
Luiz Da Costa(Linea)
Adriano Pieres(Linea)

DESGW Production Operations
Ken Herner - Lead

Dillon Brout (Chicago)
Nora Sherman (Michigan)

DES Production Operations
Brian Yanny - Lead

NCSA
Robert Gruendl - Lead
Felipe Menanteau – Science Stack prep
Greg Daues – Stack build
Doug Friedel – Tape Backup

Reconstruction Software
Emmanuel Bertin(iFA Paris)
Erin Sheldon(BNL)
Gary Bernstein(Penn)

er(ANL)

Calibration

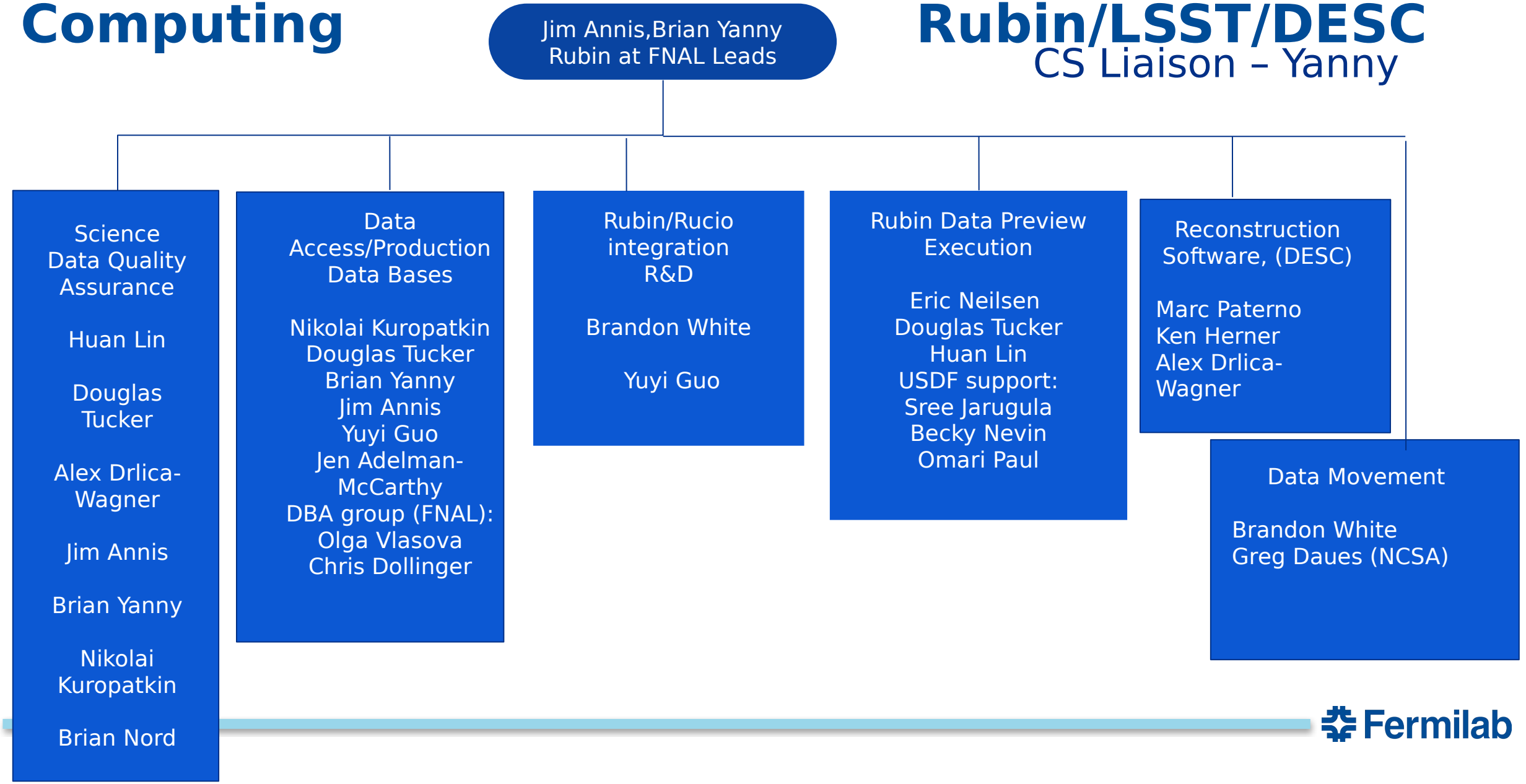
Douglas Tucker - Lead
Eli Rykoff (SLAC/Stanford)
David Burke(SLAC)

Fermilab

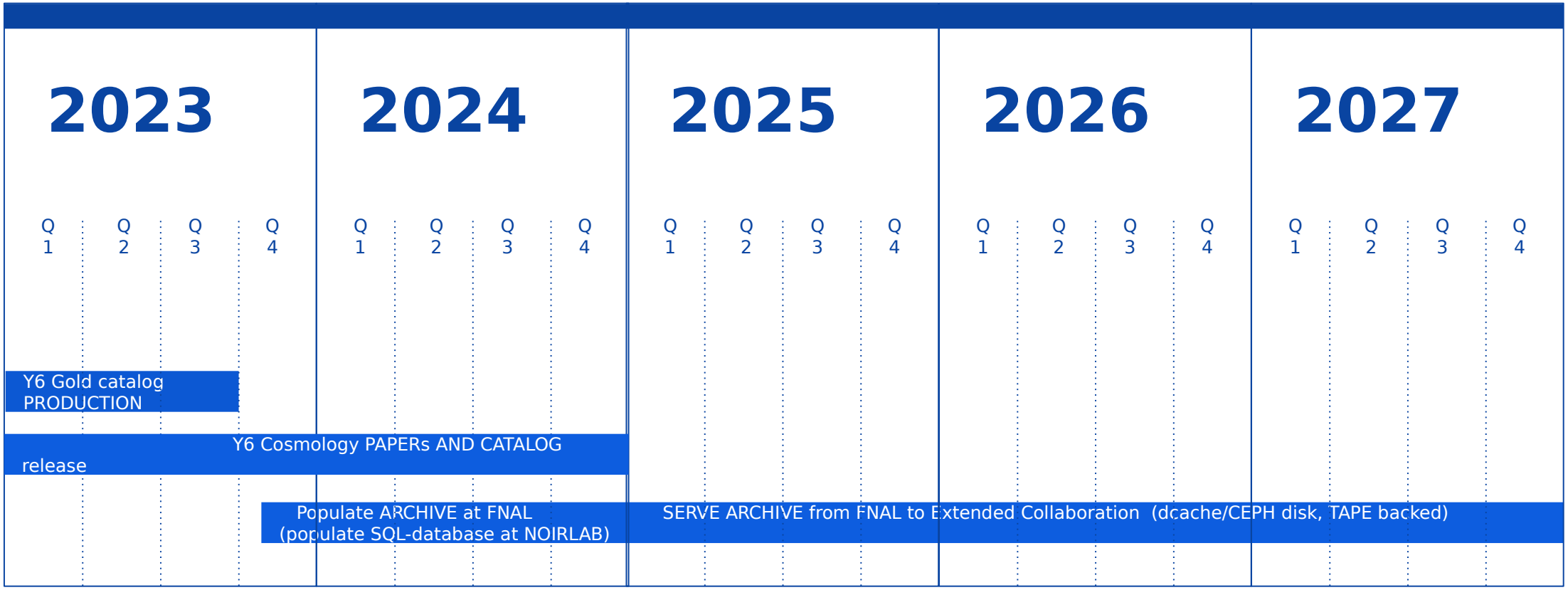
Experiment Organization Chart for Offline Computing

Rubin/LSST/DESC

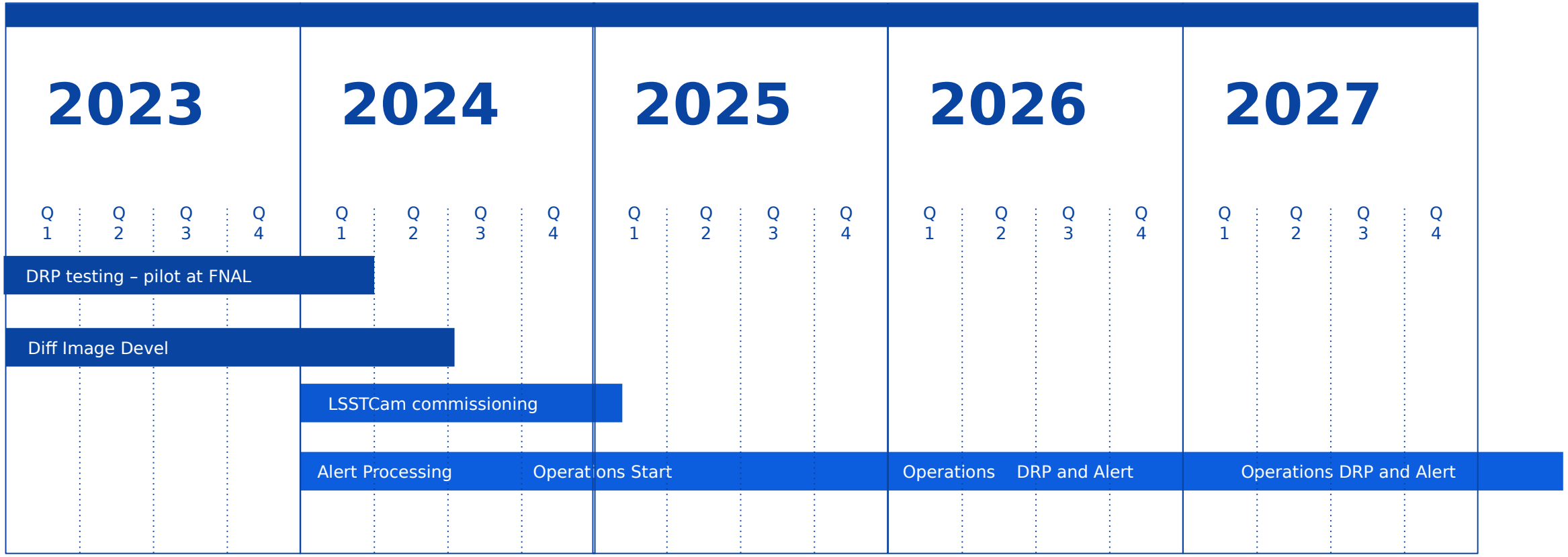
CS Liaison – Yanny



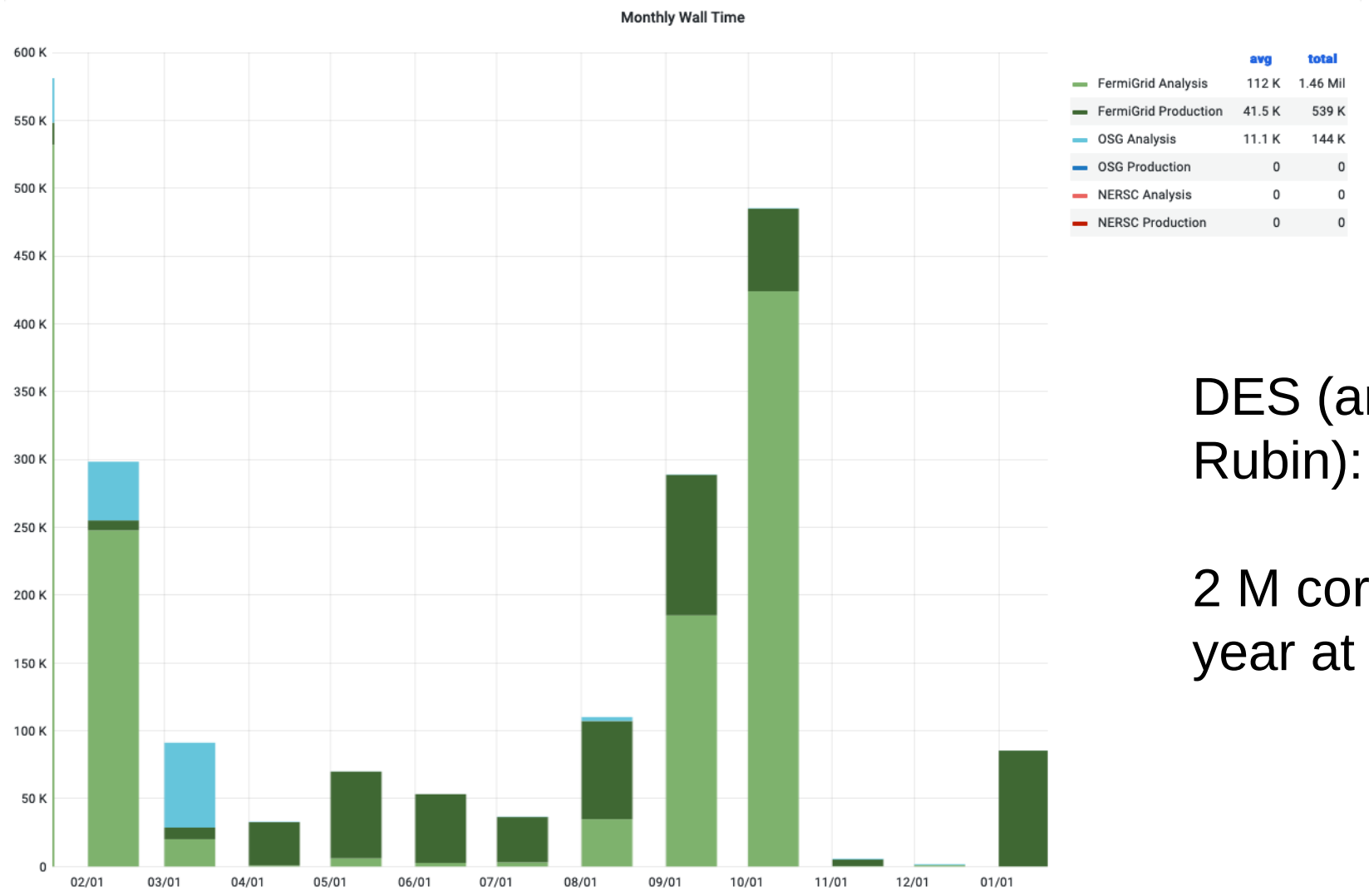
5 Year Plan – DES – compute at FNAL through 2024, then ‘archive access for 5 years’



5 Year plan – Rubin – compute not local to FNAL, only limited test stands for small scale development



Experiment CPU Usage over the past year



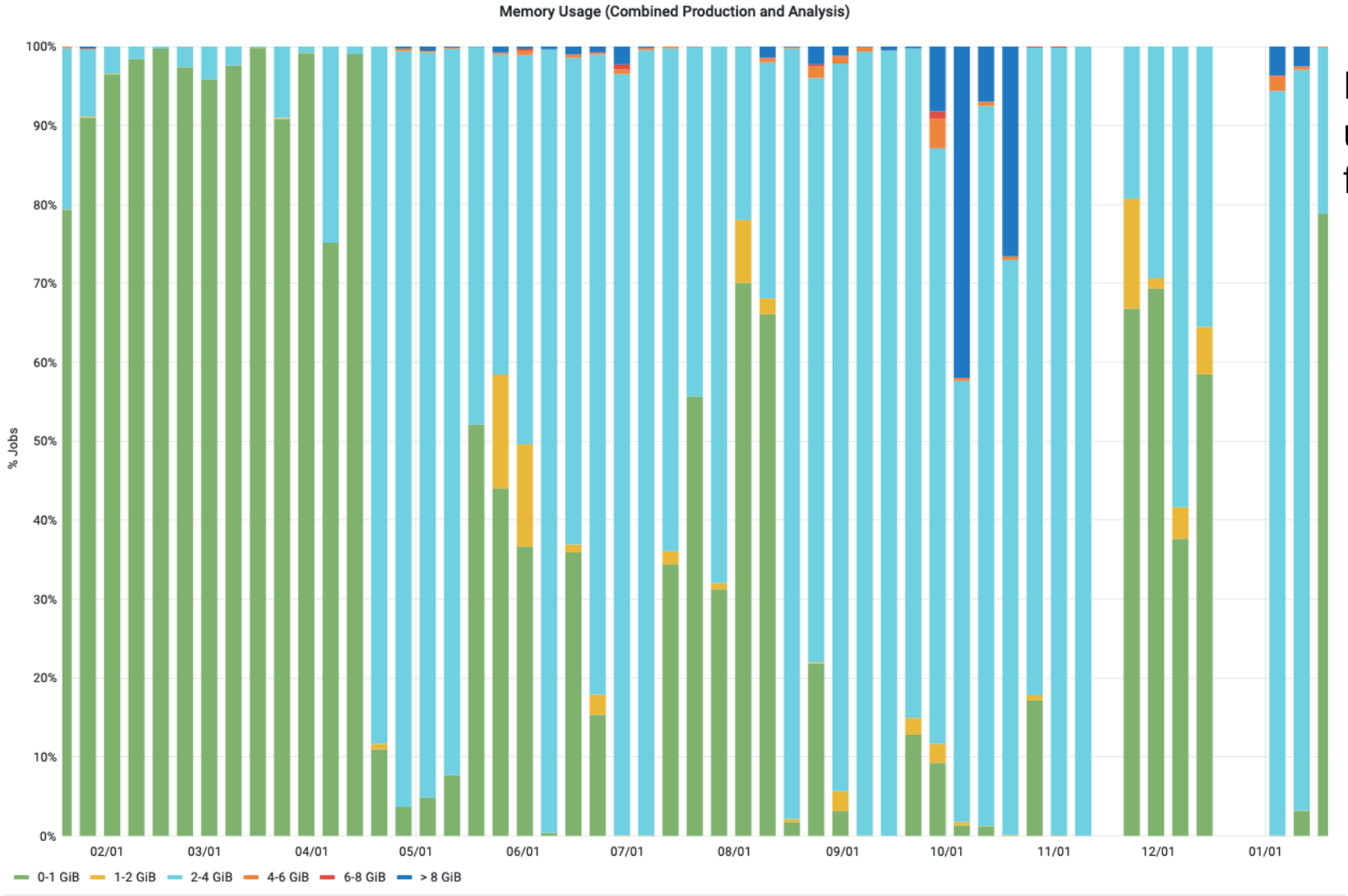
	avg	total
FermiGrid Analysis	112 K	1.46 Mil
FermiGrid Production	41.5 K	539 K
OSG Analysis	11.1 K	144 K
OSG Production	0	0
NERSC Analysis	0	0
NERSC Production	0	0

DES (and minimal Rubin):

2 M core-hours in past year at FNAL.

Fermilab

Memory footprint over the past year



Dedicated DES grid nodes utilized for large memory footprint (>8GB) jobs

DES (includes DESGW, DES-SN production)

Could request <= 2GB for vast majority of jobs.

Fermilab

CPU and memory efficiency over the past year

DES+Rubin

CPU efficiency improved over FY22 plot
Memory Efficiency still low – perhaps over requesting GB/core

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

Goals	Where does the experiment need to contribute	Where does SCD need to contribute
DES: Set up network accessible DES archive for (extended) collaboration access. Enables continued analysis processing of DES archive with cpu either on Wilson Cluter, EAF, Fermigrid or remote sites (i.e. NERSC).	Spec hardware purchase. Design file layout, load archive, test connectivity. Inform users how to transfer efficiently.	Add additional dcache/CEPH storage, help test access mechanisms
DES: Enable AI/ML analysis of DES datset, with GPU assist.	Develop codes which make use of GPU capabilities.	Assist in porting codes from regular CPUs to GPUs and onto HPC platforms.
Rubin: Enable advanced analysis of early LSST data	Define advanced: image manipulation, and display, huge data volume (PB) accessibility to users.	Help suggest solutions.

Campaign Schedules

	2023	2024	2025	2026	2027
Processing campaigns (start month-end month if known). **Include when you expect to be prestaging**	DES BFD (moment calc): Feb-Mar DES Deepfield,Balrog,imsim: Mar-Sep	DES followup imsim: Jan-Jun Rubin testing (non-FNAL resources)	Rubin DRP (non-FNAL resources) Rubin Alerts (non-FNAL)	Rubin DRP (non-FNAL resources) Rubin Alerts (non-FNAL)	Rubin DRP (non-FNAL resources) Rubin Alerts (non-FNAL)
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.**	DES BFD: 0.5 M hrs, 10TB DES deepfield, Balrog,imsim (2M hrs, incl. 1M NERSC) 50 TB	DES followup imsim: 5M hr (mostly NERSC) 20TB	DES user follow up analysis campaigns (grab files from archive, Process on variety of compute sites, NERSC, Wilson, EAF) 5 M hrs, 50 TB	DES user follow up analysis campaigns (grab files from archive, Process on variety of compute sites, NERSC, Wilson, EAF) 5 M hr, 50 TB	DES user follow up analysis campaigns (grab files from archive, Process on variety of compute sites, NERSC, Wilson, EAF) 5 M hr, 50 TB
Conference or result targets (month if known)	DES infrastructure paper(s).	DES Y6 (final) cosmology result papers	Additional DES analysis papers	Serendipitous Discovery papers (DES and Rubin)	Serendipitous Discovery papers (DES and Rubin)

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)	6	5	5.5	6	5	4	2	2	2
Actual Used	4	3	2.1		N/A	N/A	N/A	N/A	N/A
Efficiency	54%	50%	42%	%	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)				20	20	20	20	10	10
Actual Used				15	N/A	N/A	N/A	N/A	N/A
Efficiency	%	%	%	% 75%	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)					10	10	10	10	10
Actual Used					N/A	N/A	N/A	N/A	N/A
Efficiency	%	%	%	%	N/A	N/A	N/A	N/A	N/A

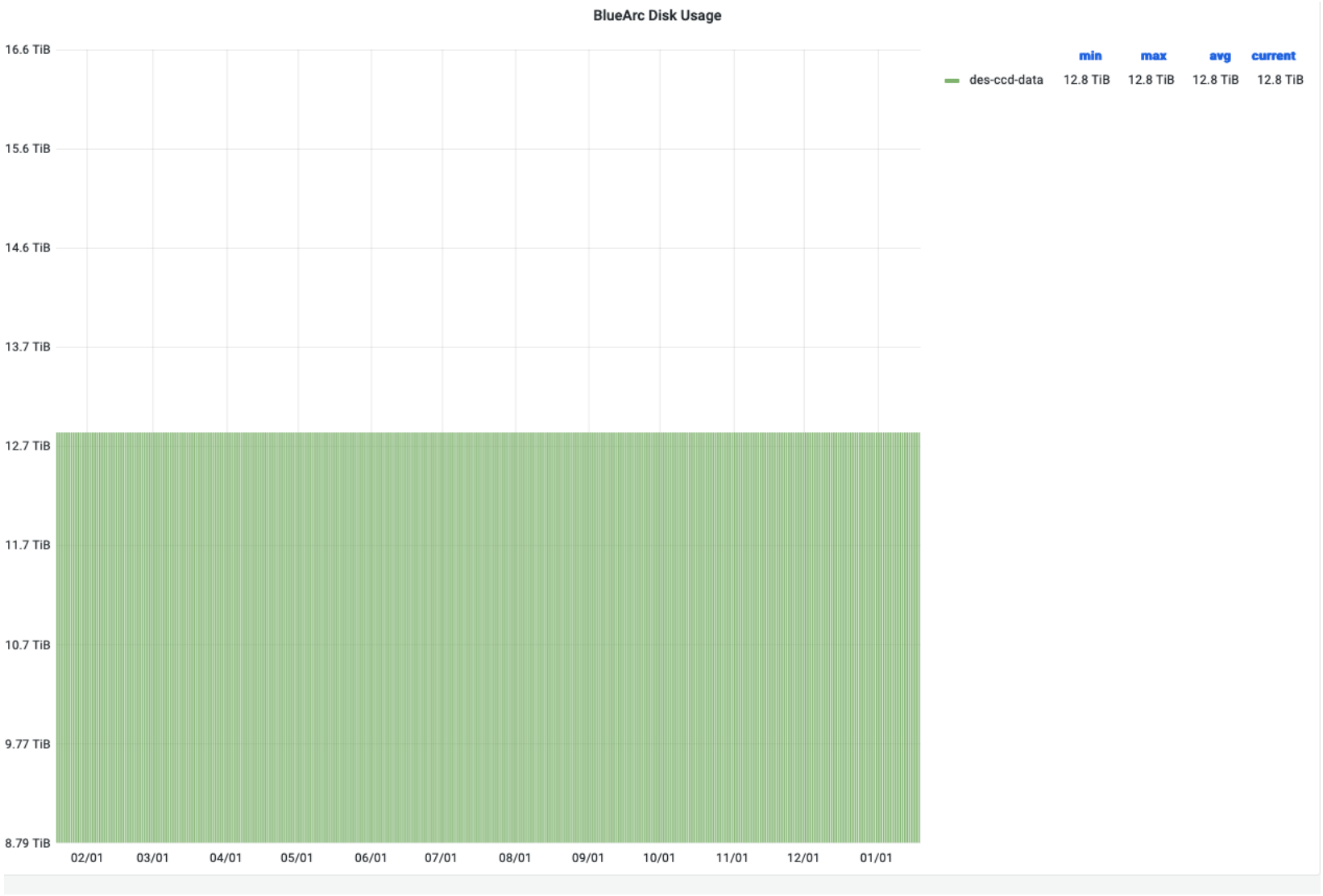
Looking for five-year projections this cycle

CPU – GPU 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)						10	20	20	20
Actual Used					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. Any particular GPU type(s) needed? FNAL has a budgeted amount for GPUs this year. What would be the best configuration for your work?

Age of files in NAS

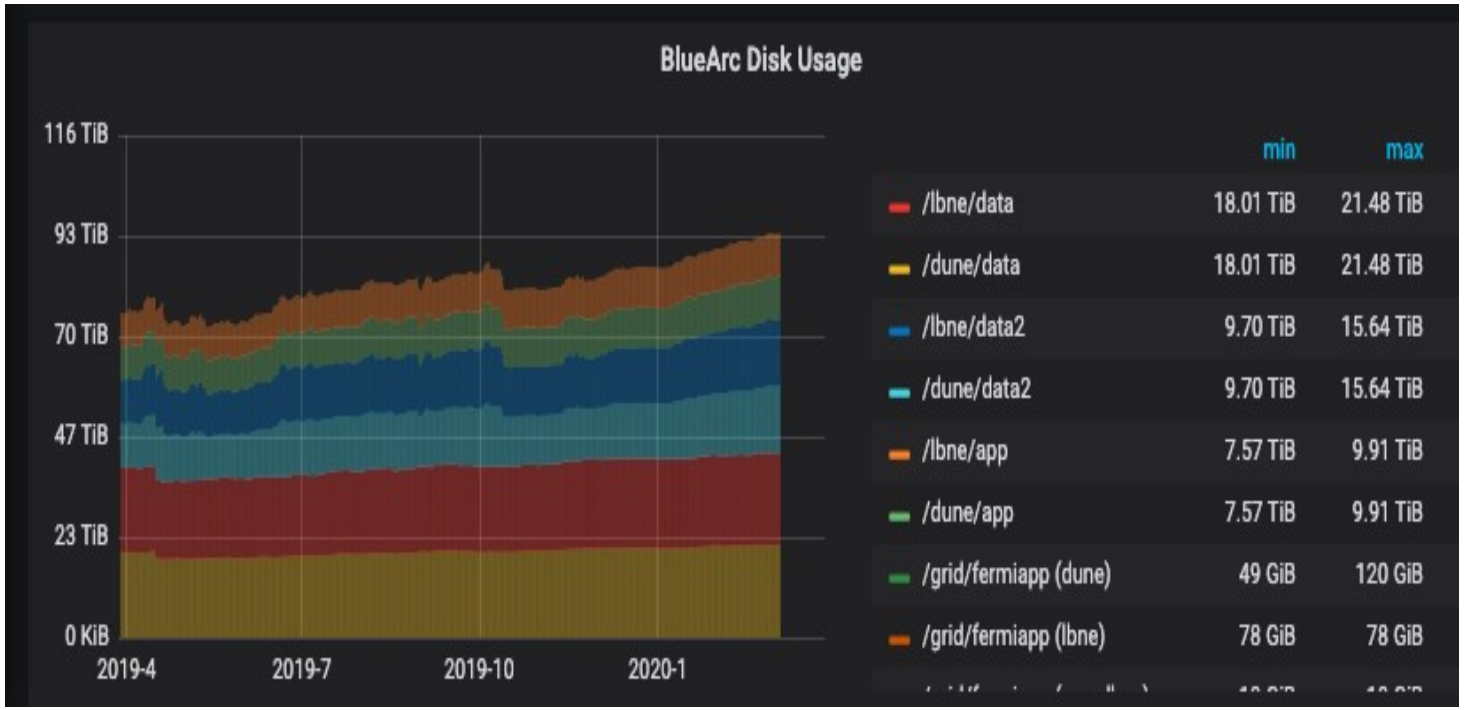


Mostly legacy data from DECam. No expansion planned.

Legacy DES disks (800TB) could be migrated to CEPH potentially.

Fermilab

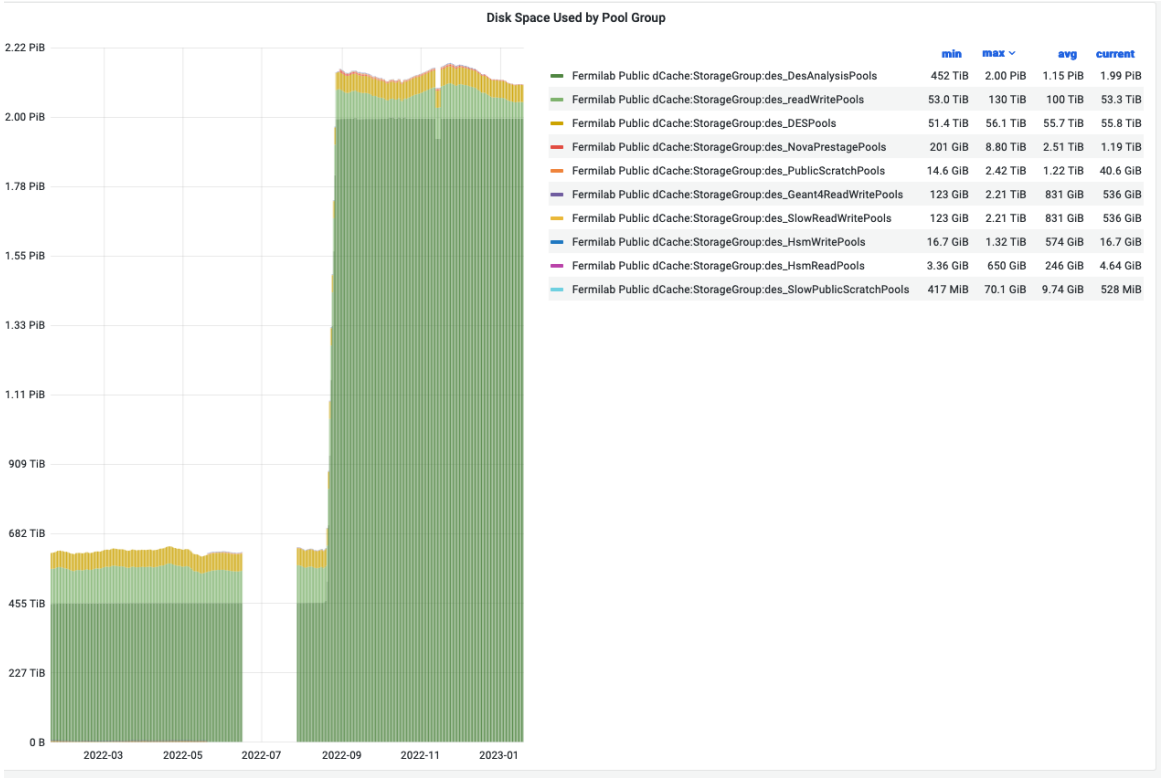
NAS Usage and Projections



	App	Data
2022		
2023		
2024		
2025		
2026		
2027		

Justify how much needs to be migrated. If you can't explain what it is or you're going to do with it, strongly consider abandoning it.
Current usage:
https://fifemon.fnal.gov/monitor/d/r6UDhH-iz/sppm-sc-pmt-prep?orgId=1

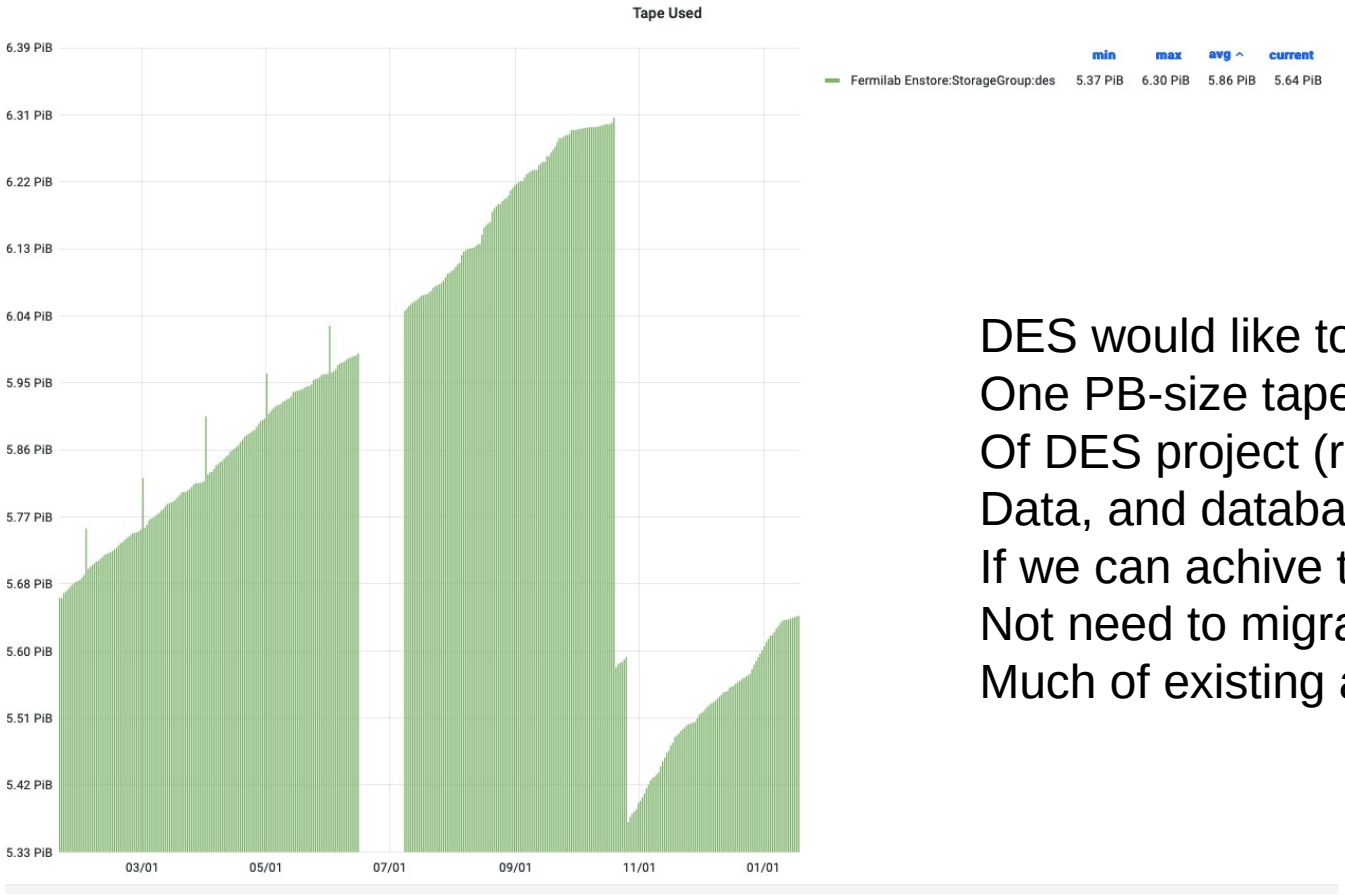
dCache Usage and Predictions (in TB)



Total dedicated r/w (tape backed): 0 TB
Total persistent: 538TB
Total dedicated other (staging pools, etc.): 0 TB

	Analysis (Persistent)	Other Dedicated (Write)
Current	X TB (actual)	Y TB (actual)
2022	538	
2023	538	
2024	538	
2025	538	
2026	538	
2027	538	

Tape usage and predictions (in PB)



DES would like to create
One PB-size tape archive
Of DES project (raw, processed
Data, and database backups).
If we can achive this, than would
Not need to migrate forward
Much of existing archive.

	Total Added By End of Year
At end 2021	5.67PB (actual)
2022	5.59PB (recycled unused DB backups)
2023	
2024	+1PB DES Archive bkup
2025	
2026	
2027	

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

Data Lifetimes

DES: DES plans to create a 1PB dcache disk archive, backed by a 1PB tape archive.
Dcache disk archive could be CEPH or ‘regular dcache’. Plan is for this disk archive to live for 5 years (through FY28) and serve extended collaboration.
Archive contains raw images, processed images, object catalogs, ‘value added catalogs’, ‘dark energy Cosmology catalogs’.

At end of 5 year disk archive lifetime, disk archive could be deprecated, working on the assumption that Rubin/LSST has superceded or absorbed DES.

1 PB tape backup to live longer.

Analysis Facility Use

- Both DES and Rubin collaborators plan to use the EAF.
- An essential ‘requirement’ is the ability to access large DES (and eventually Rubin)
- datasets (or subsets of them) in an juptyer-notebook like environment efficiently.
- For instance, if the data is in dcache, or on CEPH POSIX or Object-Store S3 style disk, then mechanisms could be used to access large volumes sequentially, or subsets of data could be selected via a SQL-style data query.
- Also important would be the ability of users to ‘share’ access to analysis results
- (like a group-id mechanism)
- A few 10s of TB of ‘local, persistent storage’ for each user should be available.
- GPUs eventually will be important for AI/ML type analyses, but we are not ready to use them yet.