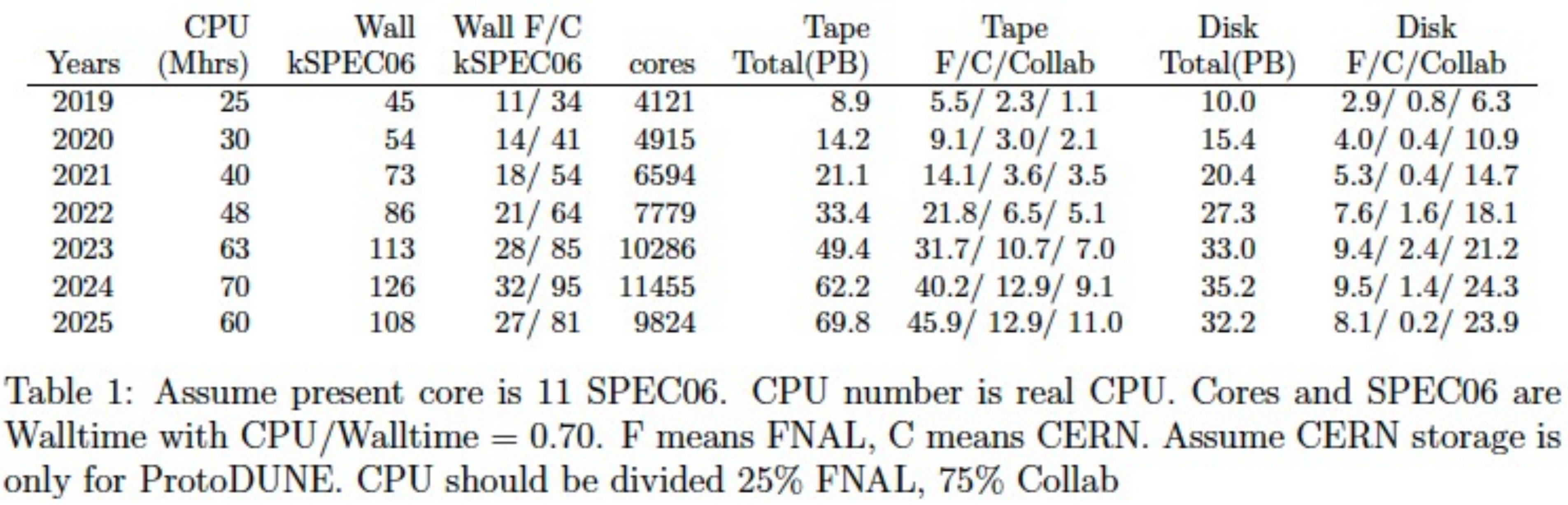
Computing Resource needs for 2022-23

H Schellman

January 26, 2022

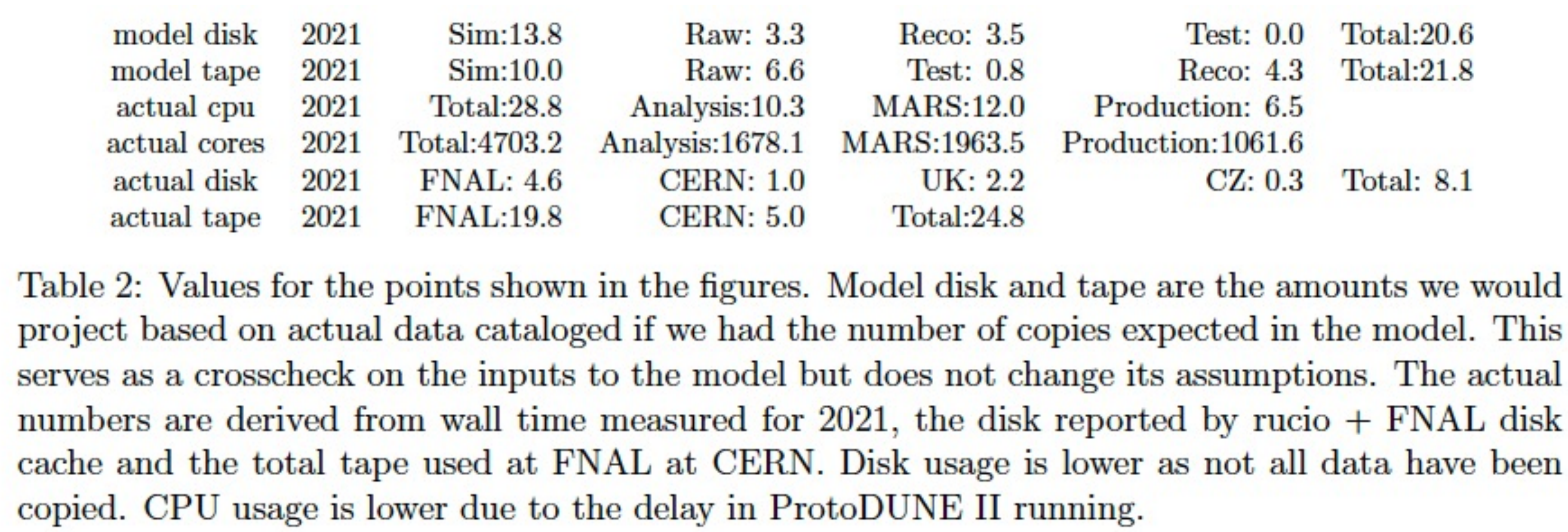
Requests from dune docdb-23419



Years	CPU (Mhrs)	Wall kSPEC06	Wall F/C kSPEC06	cores	Tape Total(PB)	Tape F/C/Collab	Disk Total(PB)	Disk F/C/Collab
2019	25	45	11/ 34	4121	8.9	5.5/ 2.3/ 1.1	10.0	2.9/ 0.8/ 6.3
2020	30	54	14/ 41	4915	14.2	9.1/ 3.0/ 2.1	15.4	4.0/ 0.4/ 10.9
2021	40	73	18/ 54	6594	21.1	14.1/ 3.6/ 3.5	20.4	5.3/ 0.4/ 14.7
2022	48	86	21/ 64	7779	33.4	21.8/ 6.5/ 5.1	27.3	7.6/ 1.6/ 18.1
2023	63	113	28/ 85	10286	49.4	31.7/ 10.7/ 7.0	33.0	9.4/ 2.4/ 21.2
2024	70	126	32/ 95	11455	62.2	40.2/ 12.9/ 9.1	35.2	9.5/ 1.4/ 24.3
2025	60	108	27/ 81	9824	69.8	45.9/ 12.9/ 11.0	32.2	8.1/ 0.2/ 23.9

Table 1: Assume present core is 11 SPEC06. CPU number is real CPU. Cores and SPEC06 are Walltime with CPU/Walltime = 0.70. F means FNAL, C means CERN. Assume CERN storage is only for ProtoDUNE. CPU should be divided 25% FNAL, 75% Collab

Actual #'s for 2021

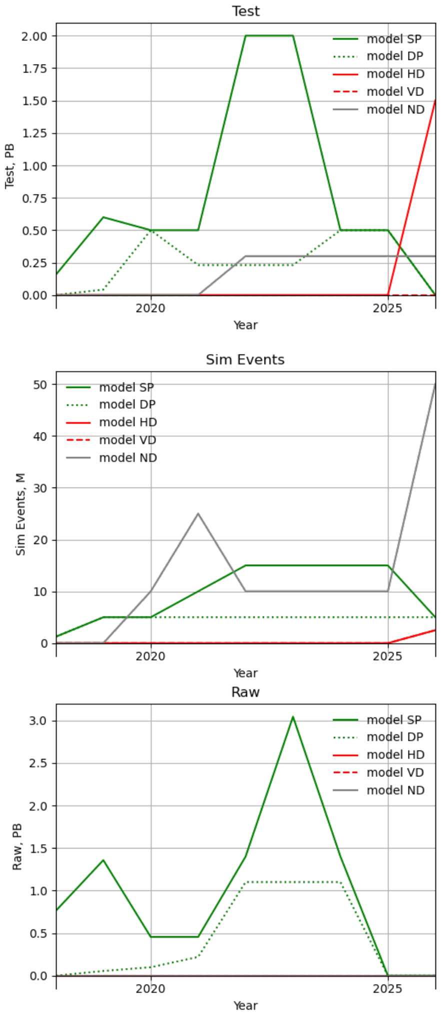


model disk	2021	Sim:13.8	Raw: 3.3	Reco: 3.5	Test: 0.0	Total:20.6
model tape	2021	Sim:10.0	Raw: 6.6	Test: 0.8	Reco: 4.3	Total:21.8
actual cpu	2021	Total:28.8	Analysis:10.3	MARS:12.0	Production: 6.5	
actual cores	2021	Total:4703.2	Analysis:1678.1	MARS:1963.5	Production:1061.6	
actual disk	2021	FNAL: 4.6	CERN: 1.0	UK: 2.2	CZ: 0.3	Total: 8.1
actual tape	2021	FNAL:19.8	CERN: 5.0	Total:24.8		

Table 2: Values for the points shown in the figures. Model disk and tape are the amounts we would project based on actual data cataloged if we had the number of copies expected in the model. This serves as a crosscheck on the inputs to the model but does not change its assumptions. The actual numbers are derived from wall time measured for 2021, the disk reported by rucio + FNAL disk cache and the total tape used at FNAL at CERN. Disk usage is lower as not all data have been copied. CPU usage is lower due to the delay in ProtoDUNE II running.

Resource estimates to 2026

- 2 copies of raw data on tape (+ 1 copy on disk if possible)
- 1 copy of "test" data stored for 6 months
- 1 copy of reco/sim on tape
  - Currently assume 1 reco pass over all data and 1 sim pass/year
  - Assume reco/sim resident on disk for 2 years
- Assume 2 disk copies of reco and sim
  - impose shorter lifetimes on tests and intermediate sim steps.



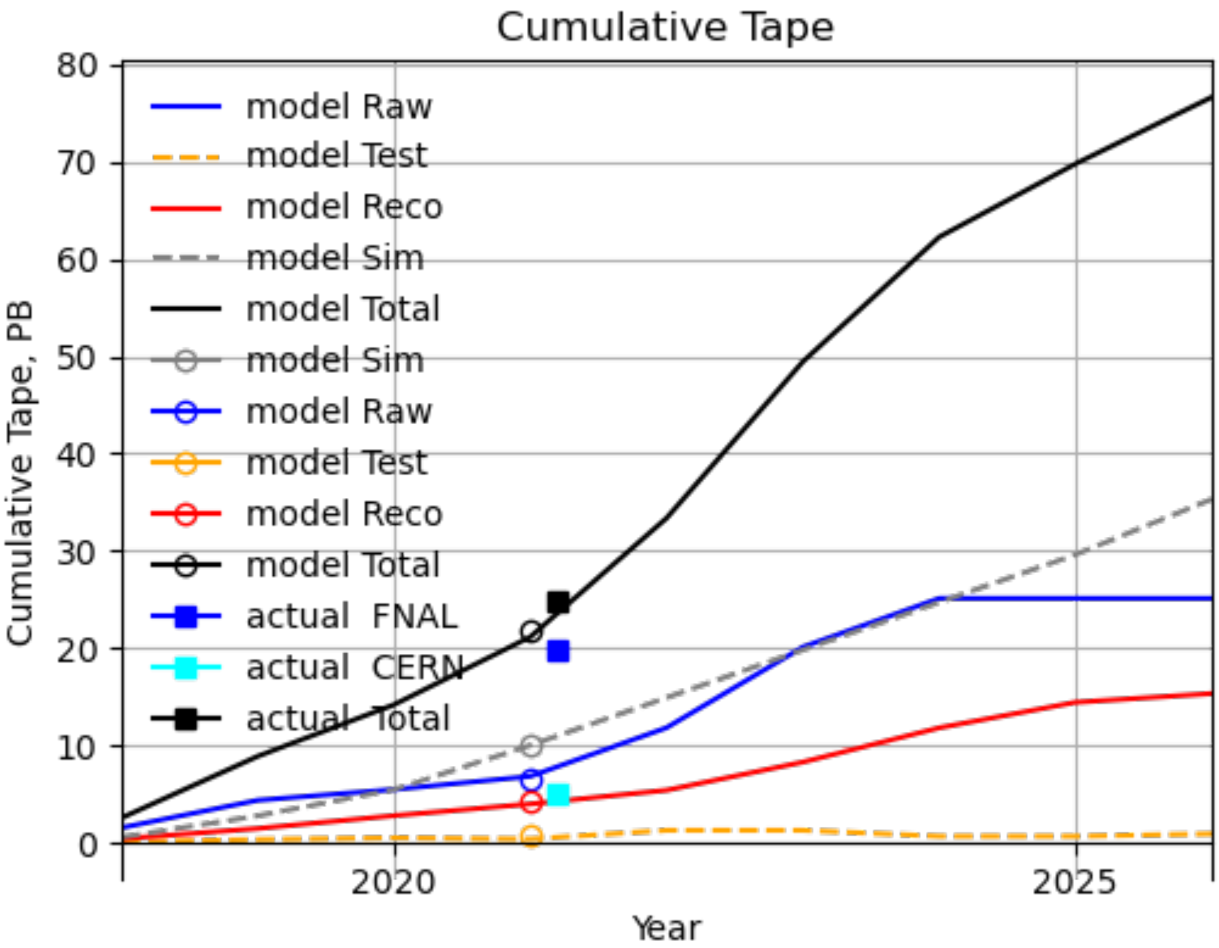
Assume ProtoDUNE runs in late 2022 and early 2023. 20 M physics events in 2022 and 43M in 2023.
+ commissioning

TAPE

Catalog sizes, PB
"Sim":10.0,
"Raw":3.3,
"Test":1.7,
"Reco":4.3
without duplicates

Actual size, PB
"FNAL":19.804,
"CERN":5.02

Catalog sizes count each file once
Model sizes apply replication rules
Actual includes multiple copies.



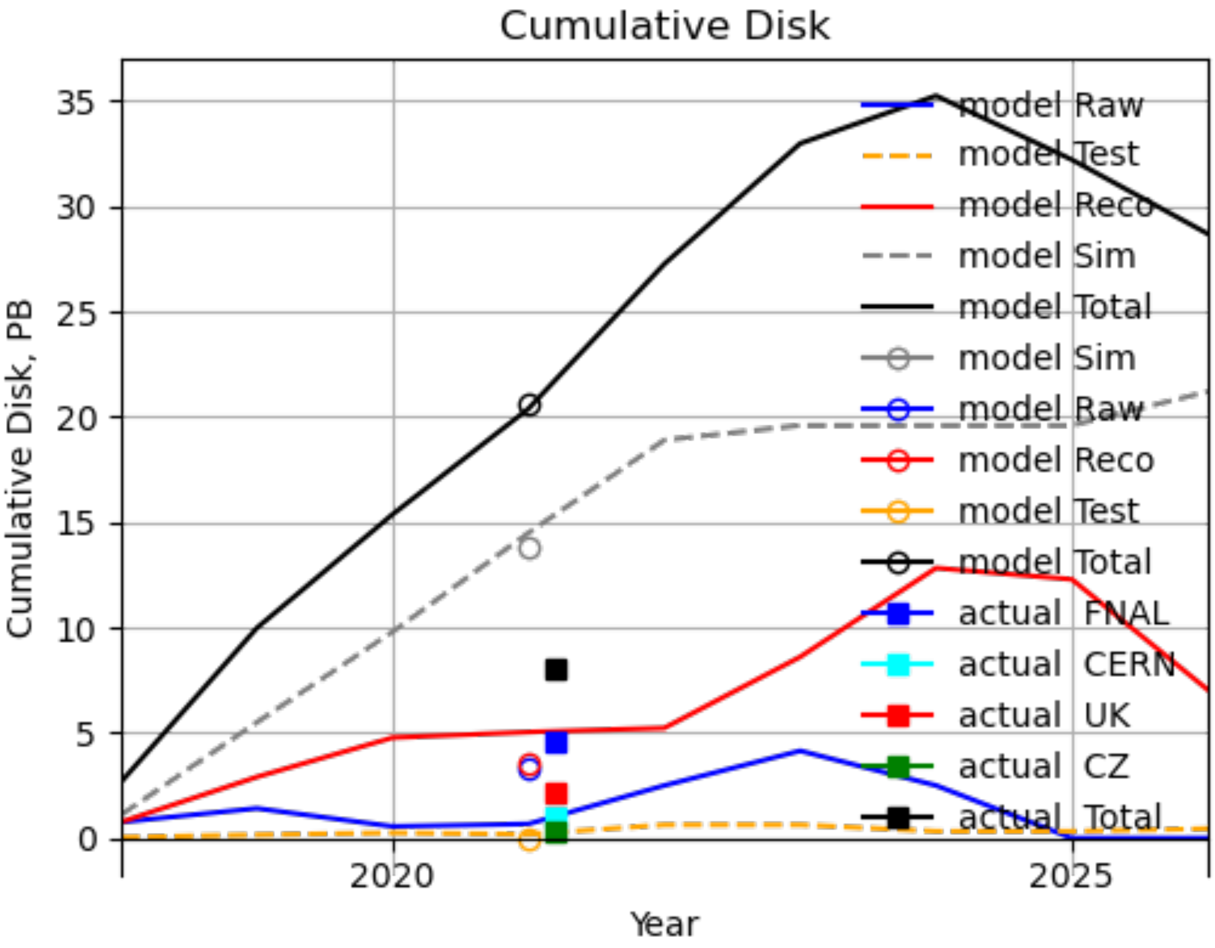
lines are projections used to make the request in July 2021
open circles are validation based on actual 2021 numbers

DISK

```
Catalog sizes, PB
"Sim":6.9 (last 2 years)
"Raw":3.3, (mainly cosmics)
"Reco":1.75, (last 2 years)
"Test":0.0
```

```
reported sizes from rucio, PB
"FNAL":4.6, (cache)
"CERN":0.975, (eos)
"UK":2.177,
"CZ":0.30
more being added to rucio
```

Catalog sizes count each file once
Model sizes apply replication rules
Actual includes multiple copies but is incomplete

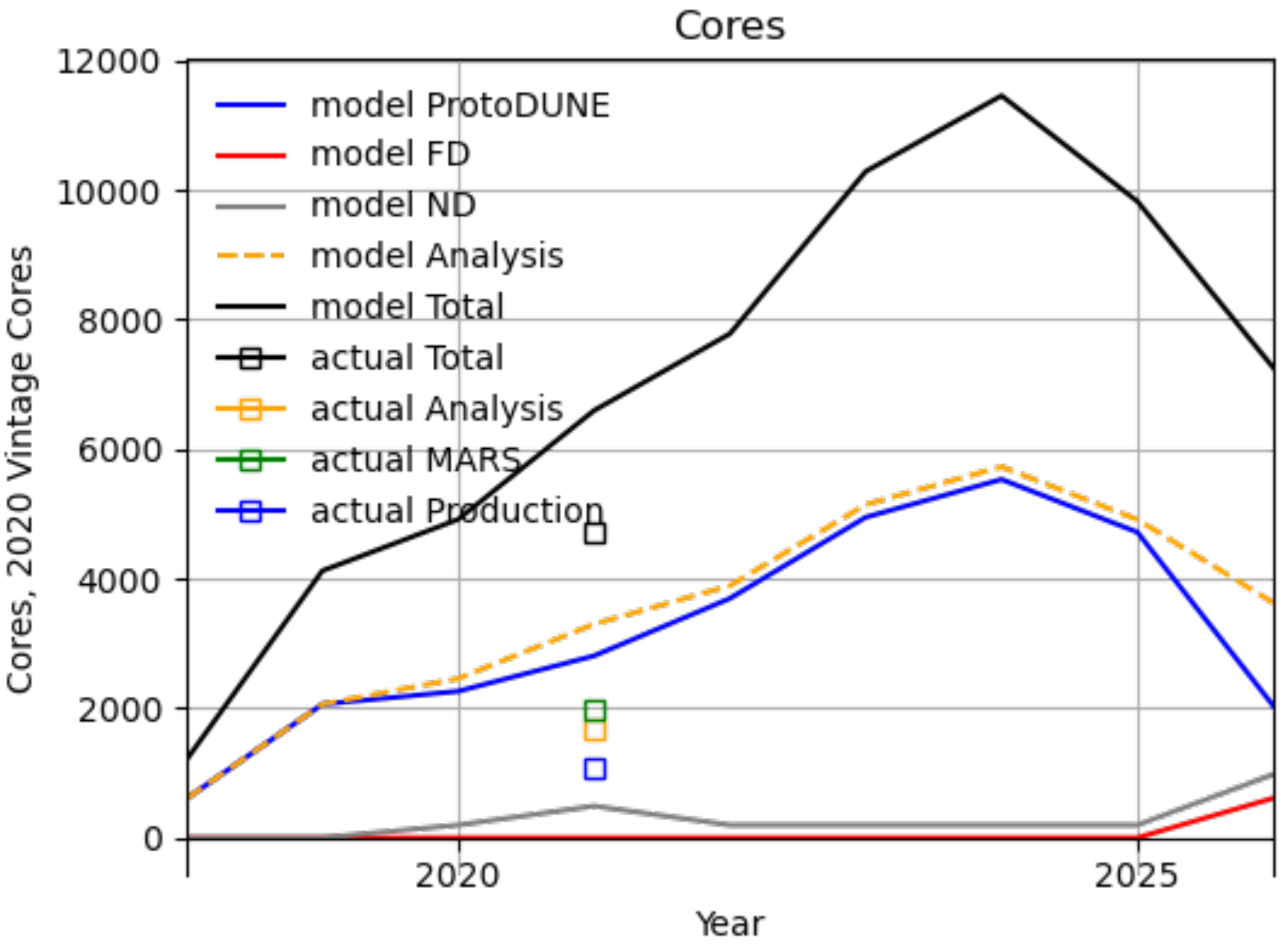


lines are projections used to make the request in July 2021
open circles are validation based on actual 2021 numbers

CORES

2021 measured
Analysis = 1679 cores
MARS = 1963 cores
Sim = 1062 cores
Total = 4703 cores

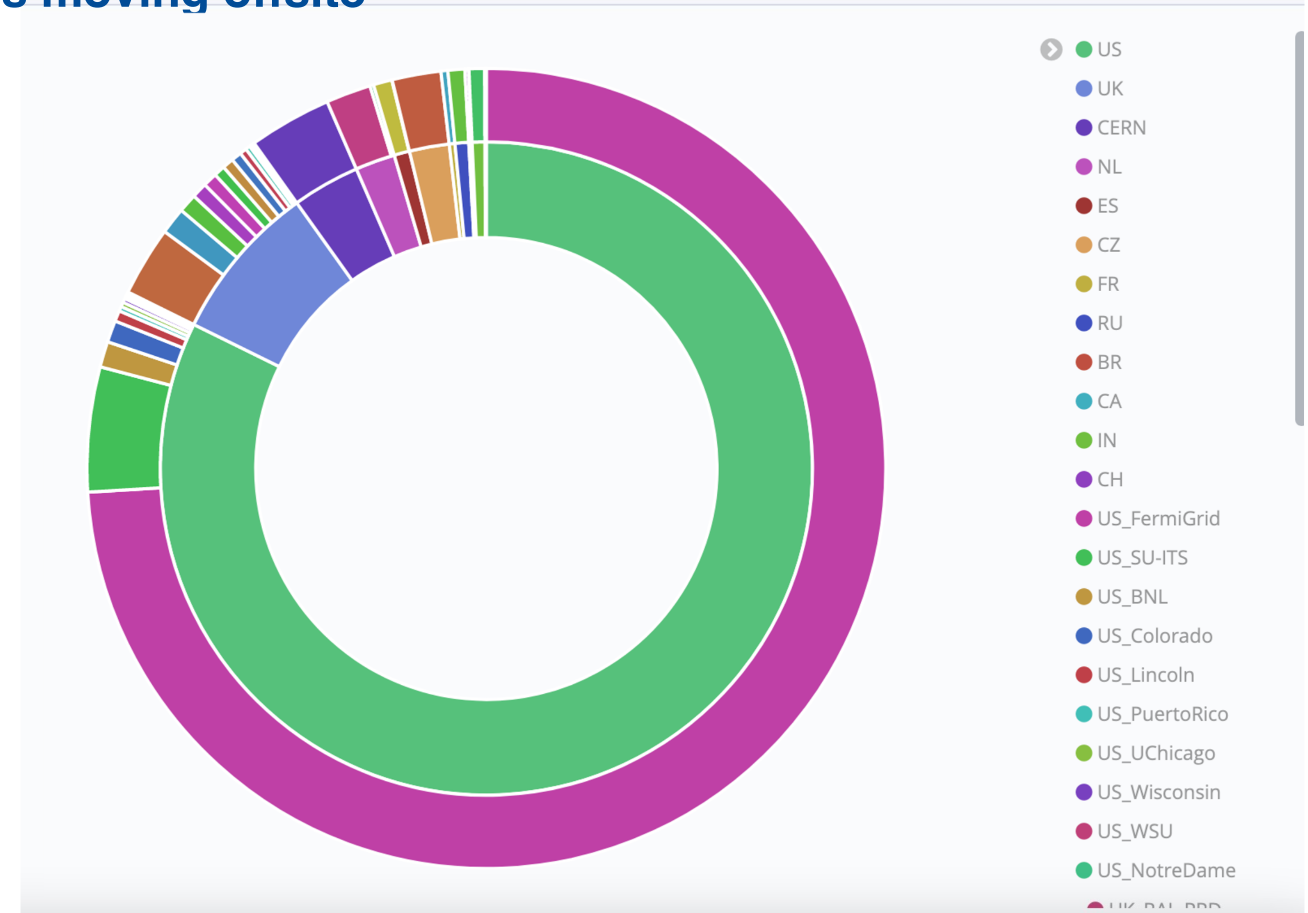
core ~ 11 SPEC06



lines are projections used to make the request in July 2021
squares are actual 2021 numbers

Where we are now: Analysis is moving offsite

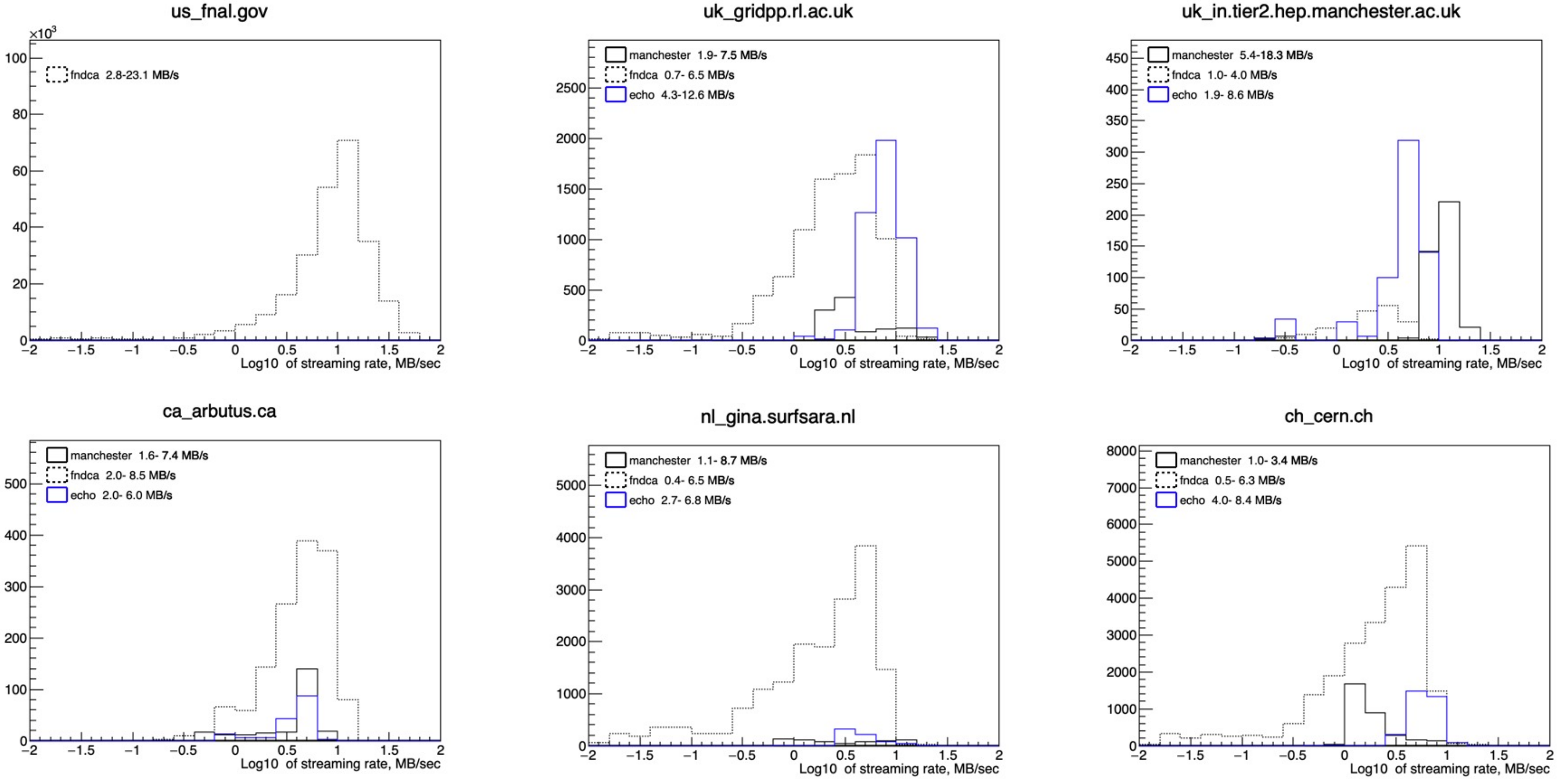
- This shows August 2021core-hrs contributions to analysis
- US sites are contributing as are many sites worldwide



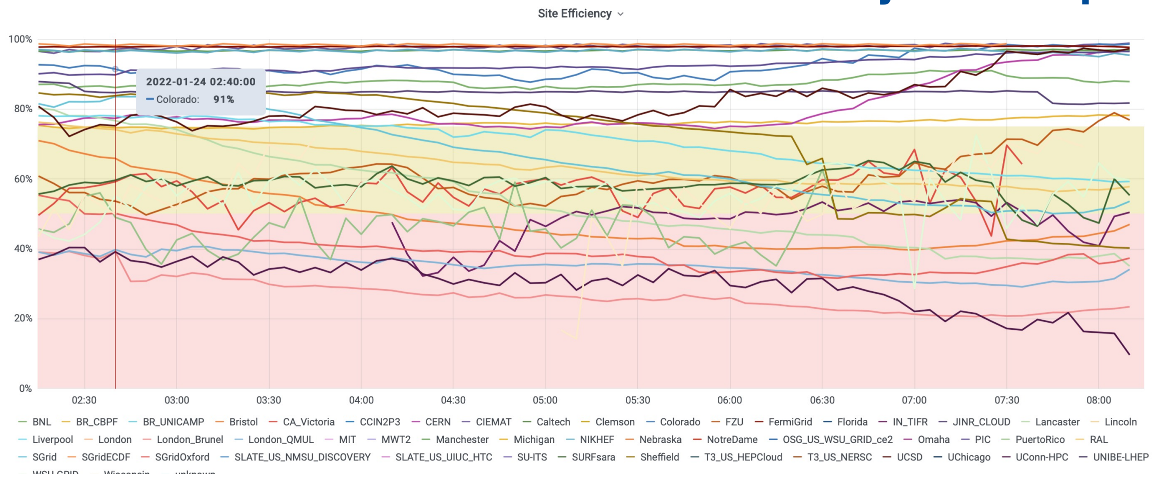
Why two copies on disk

- FNAL+CERN does not have enough disk to keep one copy of all useful samples live at all times. Sim +reco for last 2 years is close to 9PB and we also need to stage raw data for reprocessing. Prestaging has been a major hurdle for the analysis efforts!
- Analysis jobs are IO bound. The closer the disk the better.
- We are moving files to the available rucio controlled space in Europe and India. Partial raw data and beam data/sim samples for ProtoDUNE.
- SAM now has simple routing enabled to use local data when available and use data from Europe when in Europe. Should raise efficiency significantly. Same expected for India once enabled.   Proposed data access/workflow systems will make this even more efficient.

Log 10 of Streaming rates for different site disk combinations (pdspana)
No difference between sites for sim or reco (~0.1 MB/s)

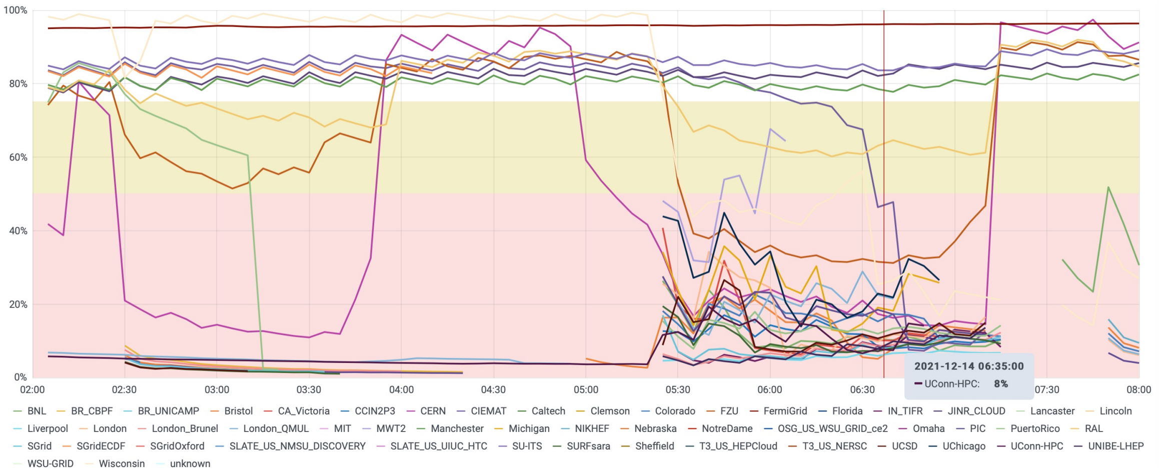


Recent site efficiencies. Mix of analysis and production



1/24/22 After sam optimization

Analysis jobs are still IO limited to ~ 50% efficiency



12/13/21 before sam optimization

Disk and CPU request

	Cores requested	Cores pledged	Cores used	Disk request PB	Disk pledged PB	Disk used PB
2021	6594	11536	4700	20.4	13.0	9.0
2022	7780	~10095	---	27.3	19.0	

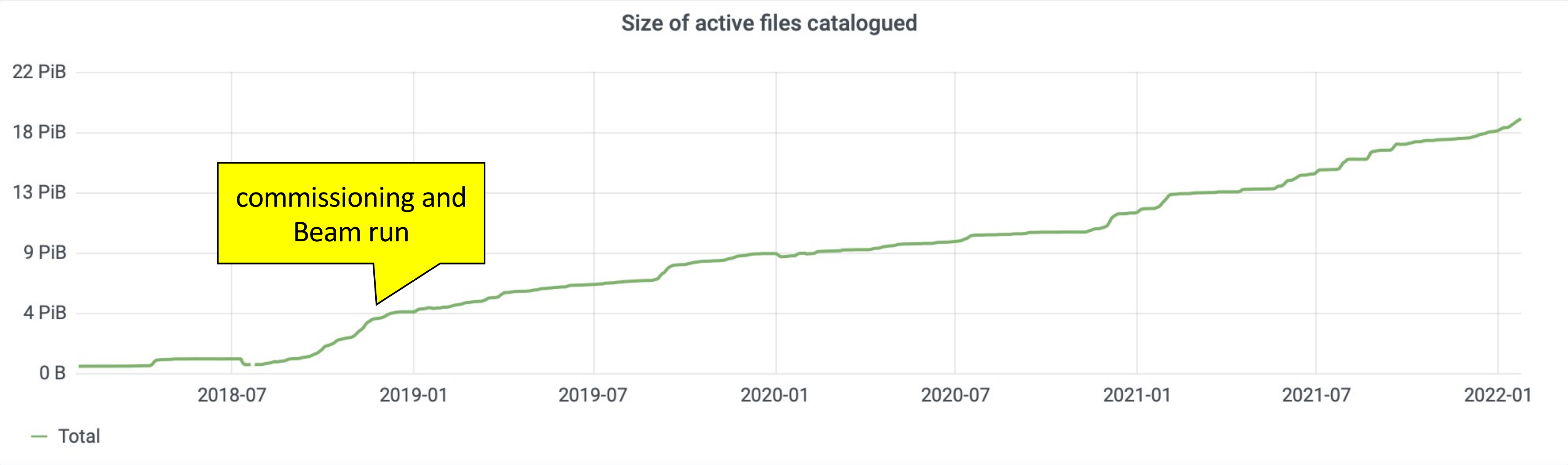
2021 request was based on assumption that PD II started in late 2021.

Disk was also underutilized in 2021 due to rucio startup. Rucio is now being brought up across sites and will fill pledged disk over next month. More will be needed to accommodate live disk copies of all sim/reco.

Disk/CPU increase in request from 2021 to 2022 driven by needs of the protoDUNE runs ProtoDUNE-II runs (VD coldbox, HD coldbox, ProtoDUNE-HD/VD commissioning, beam, cosmics)

Why so much more disk and CPU needed in 2022-2023?

- Need to move more samples to collaboration disk for fast access
- Expect to take significant ProtoDUNE RunII data before April 23
  - In 2018 we logged around 3.5 PB of raw and reconstructed data over a few months.
  - this is built into the model and explains the increase of 7 PB from 2021-22 to 2022-23



Matching pledges to needs

Requests for 2022-2023 are significantly larger due to the PD II running.

Here is how we can prioritize to use pledged resources effectively.

- Disk - pledge < full need
  - Prioritize protoDUNE 2 raw and reconstructed data for disk
  - Must reduce # of available copies and/or disk lifetimes for samples.
  - Some samples will need to be staged from tape.
- Cores - pledges > request but we can still optimize
  - Prioritize data and calibration samples
  - Rely more on opportunistic resources
  - Continue optimization of analysis job placement to raise efficiency