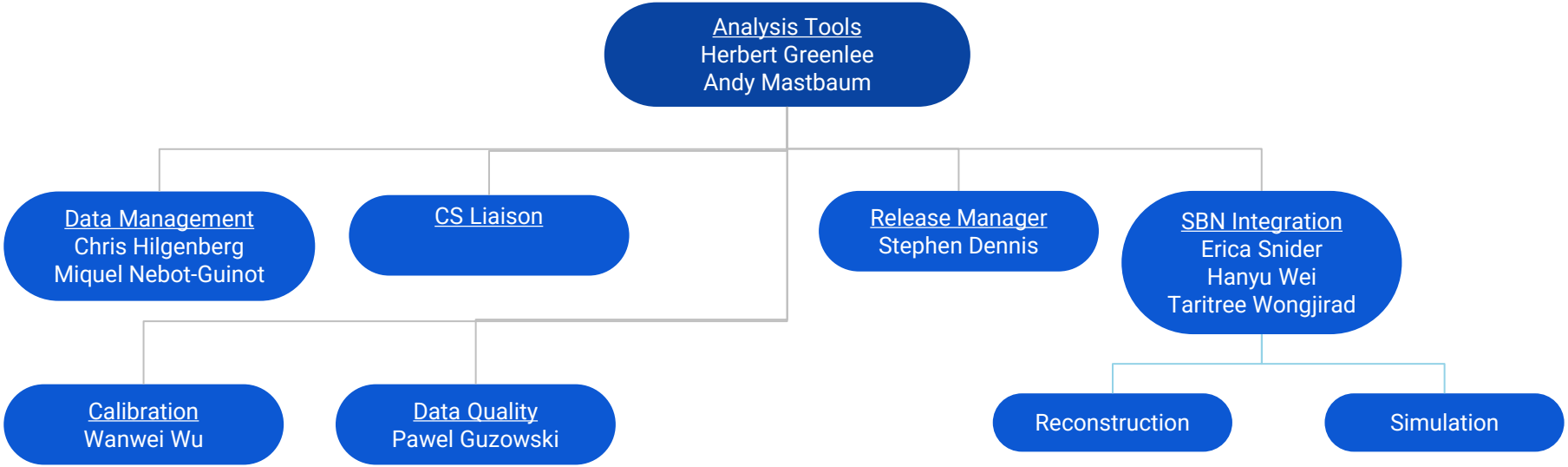
MicroBooNE FCRSG 2022

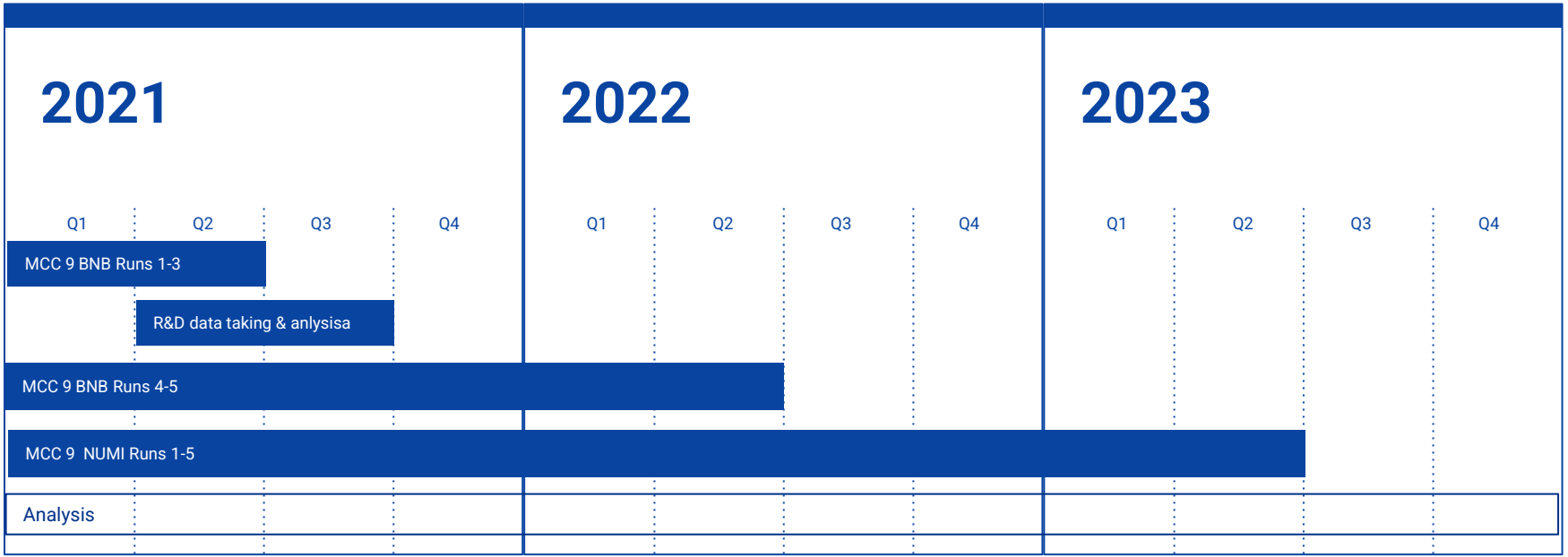
MicroBooNE Organization Chart for Offline Computing



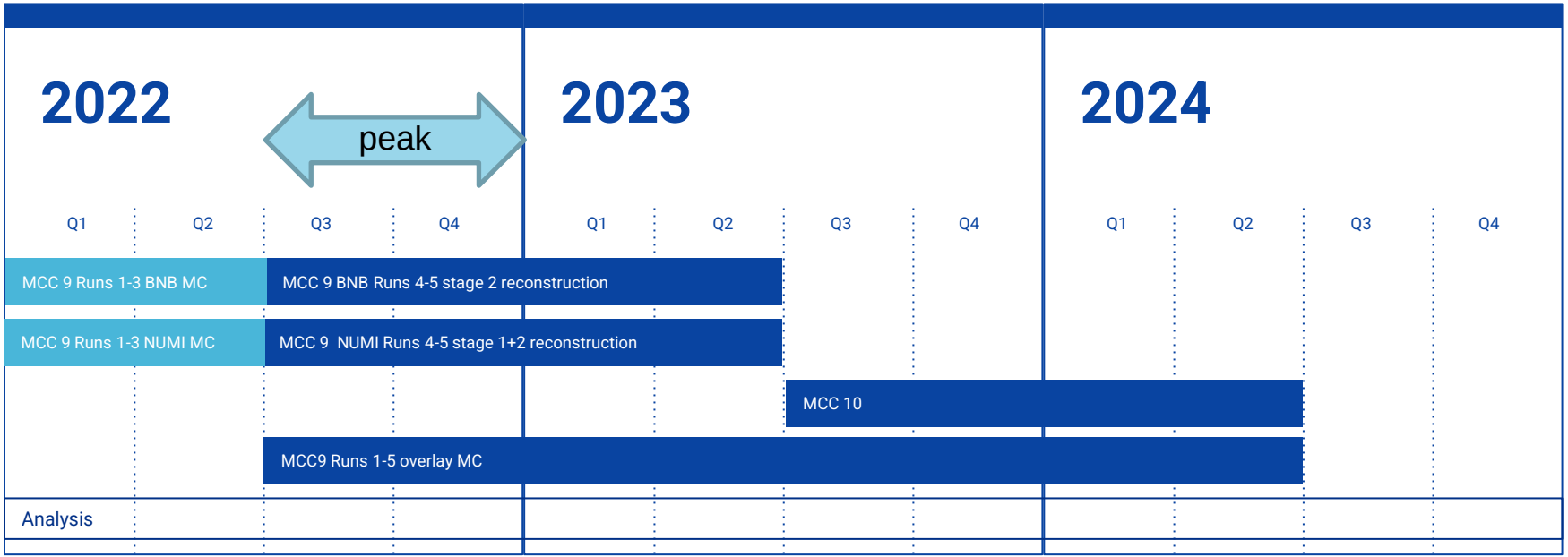
Overview of MicroBooNE Computing in 2022

- MicroBooNE took data for five years from 2015-2020 (known as runs 1-5).
  - MicroBooNE took detector R&D data in 2021.
  - MicroBooNE data taking is complete.
- MicroBooNE is the first large LArTPC experiment to do a complete physics analysis using automated reconstruction.
  - MicroBooNE has produced 49 publications (published and submitted) so far, with more on the way.
- MicroBooNE has been running data processing campaigns (data and MC) using software versions known as MCC9 since Dec., 2018.
  - In 2021, MicroBooNE released its first results on the MiniBooNE BNB Low Energy Excess (LEE), based on runs 1-3.

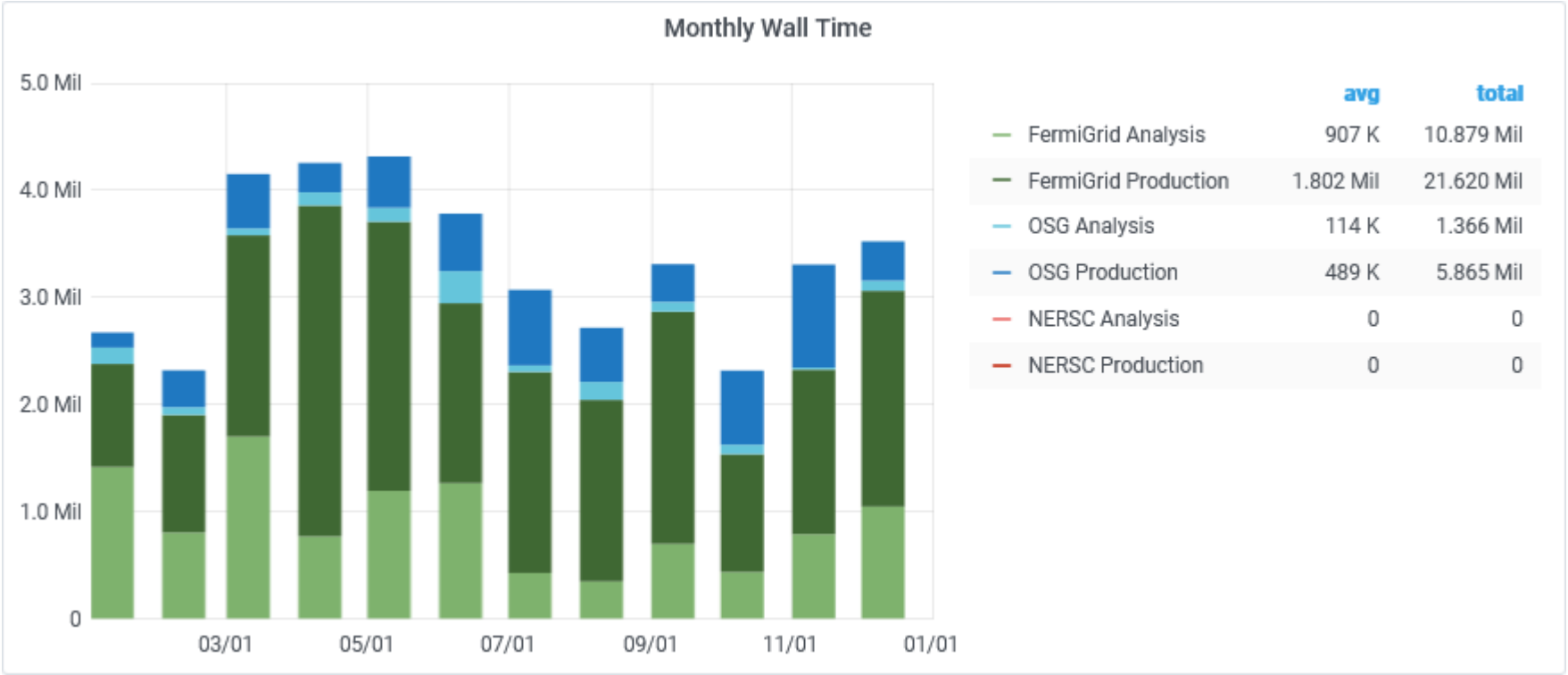
Important Dates to Remember (slide from FCRSG 2021)



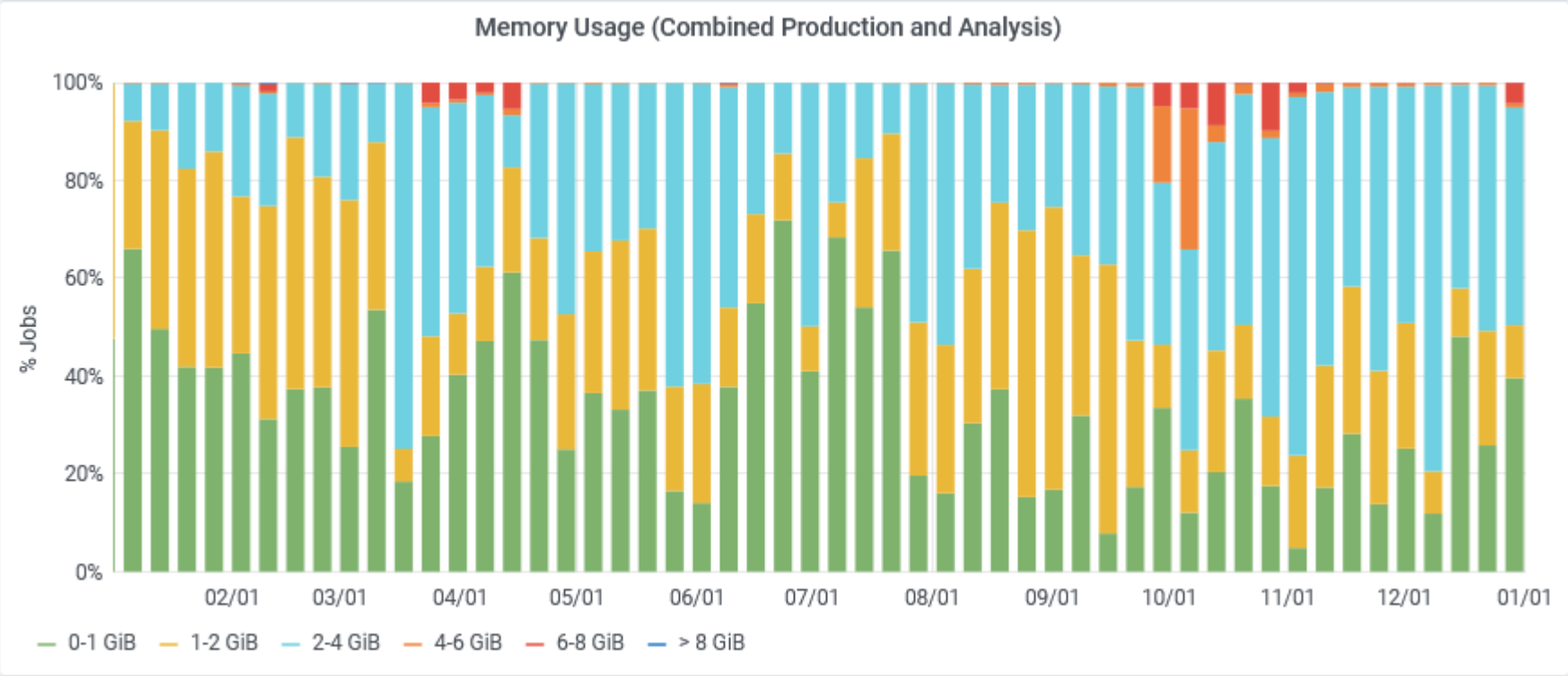
Important Dates to Remember



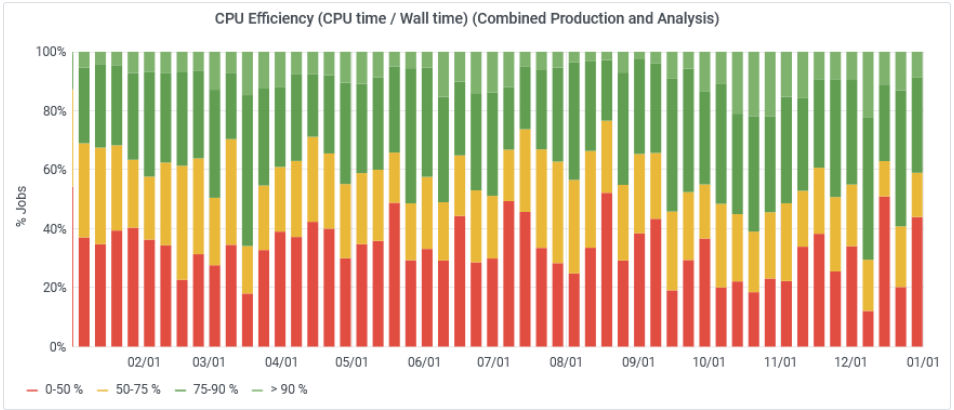
CPU - Experiment Usage Over the Last Year

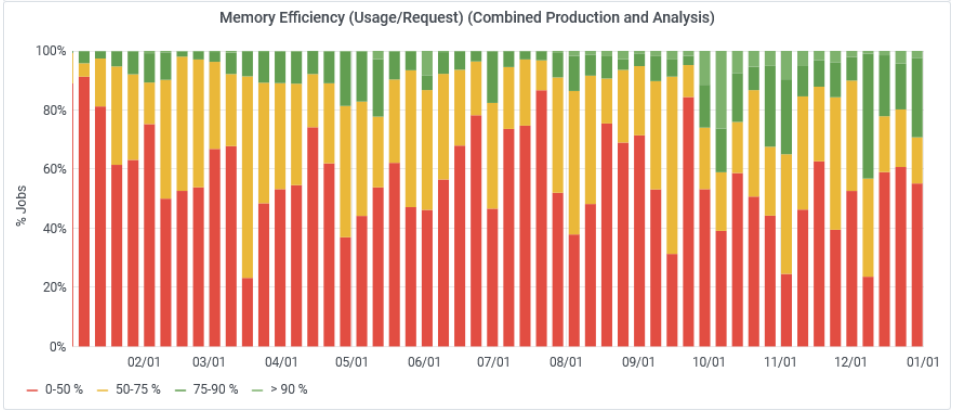


Memory Footprint Over the Last Year



CPU and Memory Efficiency Over the Last Year





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

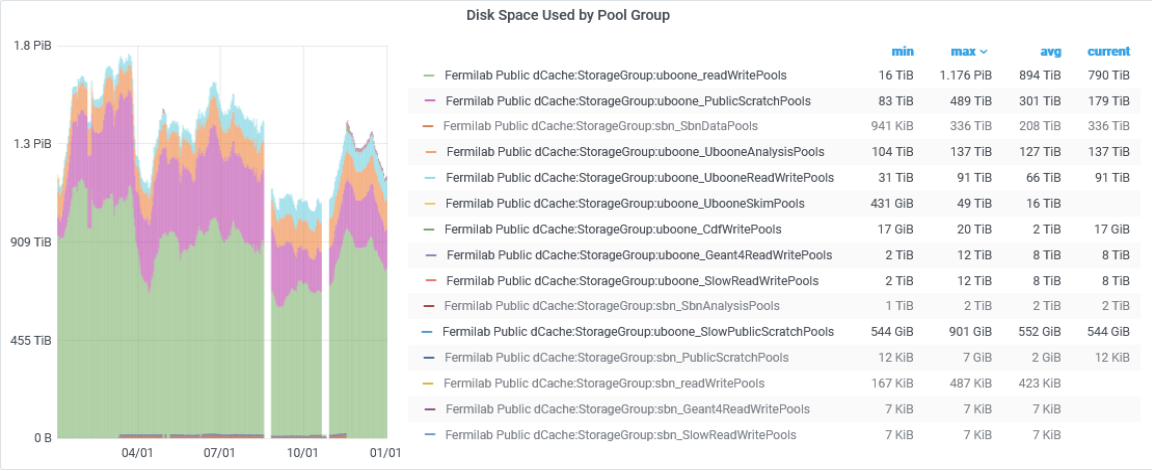
	2018	2019	2020	2021	2022	2023	2024
Requested	7	20	54	43	30	30	15
Actual Used	11.5	29.3	40.6	32.5	15.1 thru 5/10	N/A	N/A
Efficiency	53.9%	63.8%	56.7%	63.9%	63.8% YTD	N/A	N/A

CPU Adaptations Going Forward

How can experiment use OSG/HPC/HEPCloud going forward

- MicroBooNE uses opportunistic OSG slots for all workflows.
  - Opportunistic OSG accounts for about 20% of batch computing resources.
- MicroBooNE has special priority on Bern OSG site.
  - 1000 4 GB slots / 600 TB local storage element.
  - Used for run 5 data reconstruction in 2020 & 2021.
  - Will be used for run 4/5 BNB stage 2 reconstruction (about to start).
- MicroBooNE has an allocation of 12.8M hours at ANL/Theta HPC facility.
  - R&D on Theta workflows in NUMI group.
  - Workflow has been demonstrated with limited number of jobs. Challenge is to scale. I/O limited.

Disk: dCache Usage and Predictions (in TB)



MicroBooNE dedicated dCache pools

- UbooneAnalylsisPools (persistent) – 151 TB
- UbooneReadWritePools (write) – 100 TB (can be returned)
- UbooneSkimPools was returned in 1/2021
- Dedicated data pool 1.5 PB requested at Mar. 3, 2022 SPPM (see following slide).

	Analysis (Persistent)	Other Dedicated (Write)
Current	151 TB (actual)	100 TB (actual)
2022	151 TB	0
2023	151 TB	0
2024	151 TB	0

Dedicated Data Pool Request

- MicroBooNE requested a dedicated persistent data pool of 1.5 PB.
  - Orginally requested in an SPPM on Mar. 3, 2022.
- Anticipated uses are as follows.
  - Only to be used for production (not accessible to analyzers).
  - Prestage all run 1-5 beam-off data (914 TB = 524 TB runs 1-3 + 390 TB runs 4-5).
    - These data are read repeatedly for generating overlay MC.
  - Prestage NUMI run 4-5 raw data (502 TB).
    - These data can be deleted once they have been reconstructed.
  - Temporary storage for newly generated data and MC.
    - Retain data that will be used for further processing.
      - Branching workflows (detector systematics).
    - Front end for FTS (prevent FTS backlogs).

Production Plans for 2022

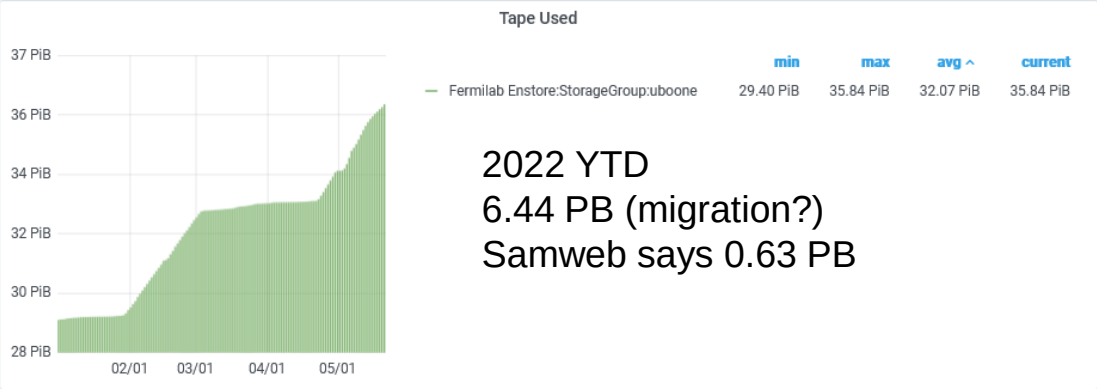
- In 2021, 59% of newly produced data was overlay MC (as opposed to real data).
- In 2022 so far, the figure is 87% of newly produced data is overlay MC.
- We expect this pattern to continue in the future. Most new data storage will be overlay MC.
  - The main bottleneck for these kinds of production is reading beam-off raw data from tape.
- The remaining real data production campaigns are as follows.
  - Runs 4/5 stage 2 reconstruction of BNB data.
    - Input data for this campaign is already staged to Bern.
  - Runs 4/5 stage 1+2 reconstruction of NUMI data.
    - Starting now.

Dedicated Data Pool Possible Scope Reductions

- If the full 1.5 PB can not be granted, the first use of the data pool that we would eliminate is prestaging NUMI raw data ahead of reconstruction.
  - This would save about 500 TB, reducing the needed size to 1.0 PB.
- If 1.0 PB is still too large, we would prioritize prestaging runs 4-5 beam-off data ahead of runs 1-3 beam-off data.
  - We estimate that about 70% of future overlay MC production will be for runs 4-5.
  - With 500 PB we could keep all of runs 4-5 beam off data.
- Summary
  - We think we can manage our remaining beam-on data production with existing resources (including Bern).
  - Having beam-off data (especially runs 4-5) on disk will speed up our remaining overlay MC production and take pressure off the tape system.

Tape - Usage and Predictions (in PB)

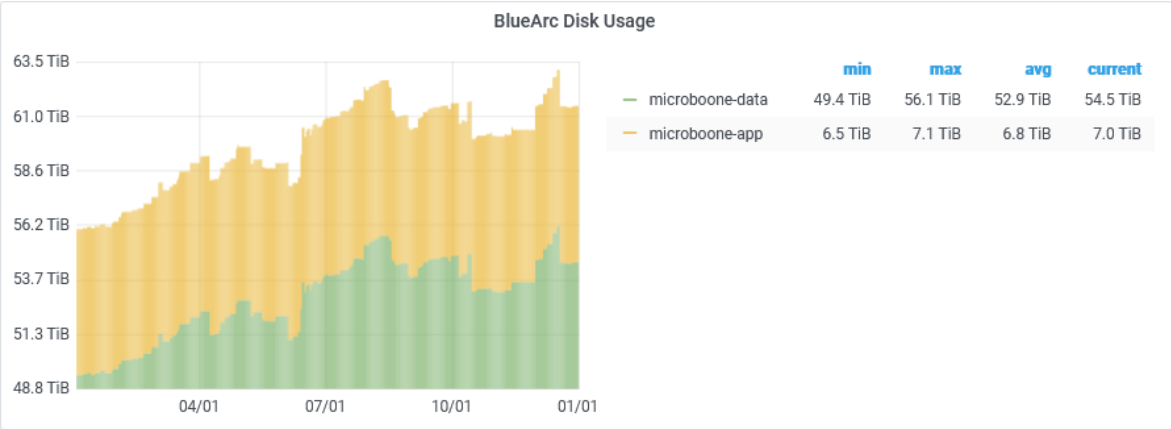
2021
2.76 PB



2022 YTD
6.44 PB (migration?)
Samweb says 0.63 PB

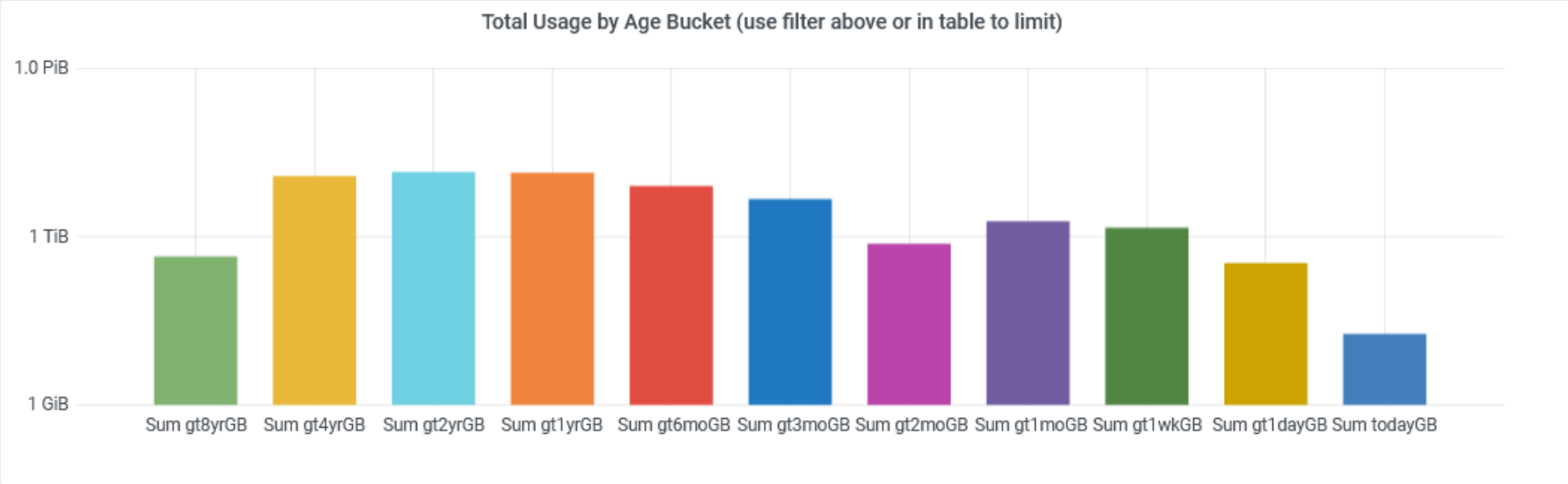
	Total Added By End of Year
At end 2021	+3.5 PB (pred) +2.76 PB (actual) 29.38 PB (total)
2022	+0.63 (actual YTD) +3.5 PB (req)
2023	+3.5 PB (req)
2024	

Disk: NAS Usage and Predictions (in TB Units)



	App	Data
2022	8 TB	60 TB
2023	8 TB	60 TB
2024	8 TB	60 TB

Age of files in NAS



Data Lifetimes

Please describe your plans, if any, to delete any datasets

- We have never specified a data lifetime at creation.
- We have had several data deletion campaigns in the past.
  - We have already deleted all MCC7 and earlier reconstructed data and MC.
  - We are internally discussing a deleteion plan for MCC8.
    - Deleting all MCC8 MC and reconstructed data would save about 7 PB (out of about 30 PB).
    - We could also consider not migrating MCC8 data.

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

- Finish reconstructing runs 1-5 using MCC9 software.
  - Stage 2 reconstruction of BNB data for runs 4-5.
  - Stage 1+2 reconstruction of NUMI data for runs 4-5.
- Physics analysis and results.
  - Complete flagship BNB LEE analysis for full dataset (runs 1-5).
  - Full oscillation analyses + many other physics analyses.
- Analyze and publish detector R&D data taken in 2021.
- Develop reconstruction and simulation beyond MCC9 (MCC10).
- Possible targeted reprocessing using a future MCC10.
- Use experience gained by running MicroBooNE to contribute to the SBN program.