

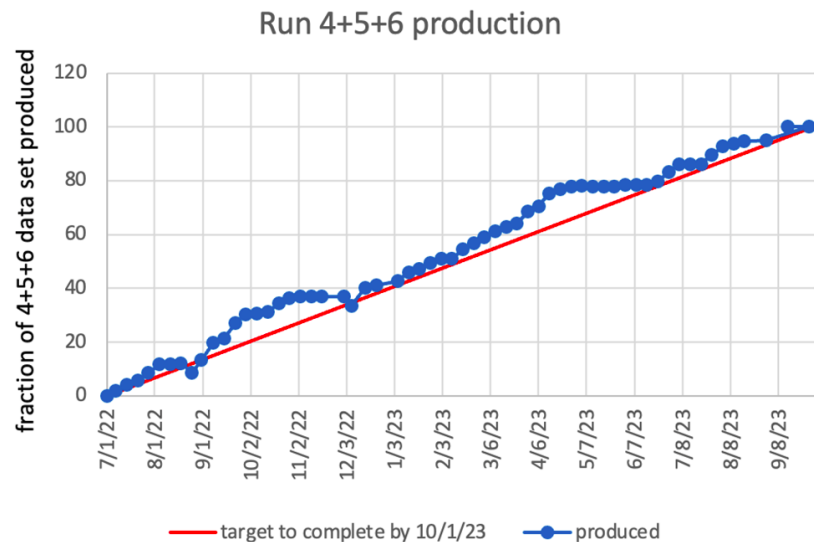
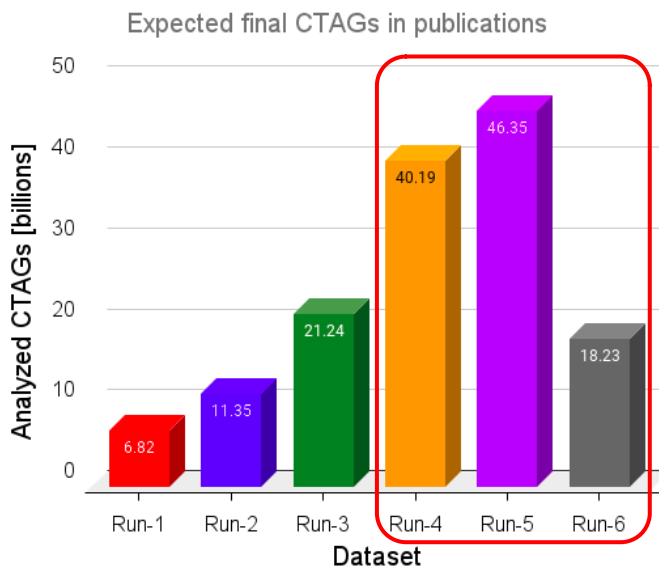
g-2 cache pool increase request

SPPM Meeting
26 Oct 2023

P. Girotti on behalf of the g-2 offline team

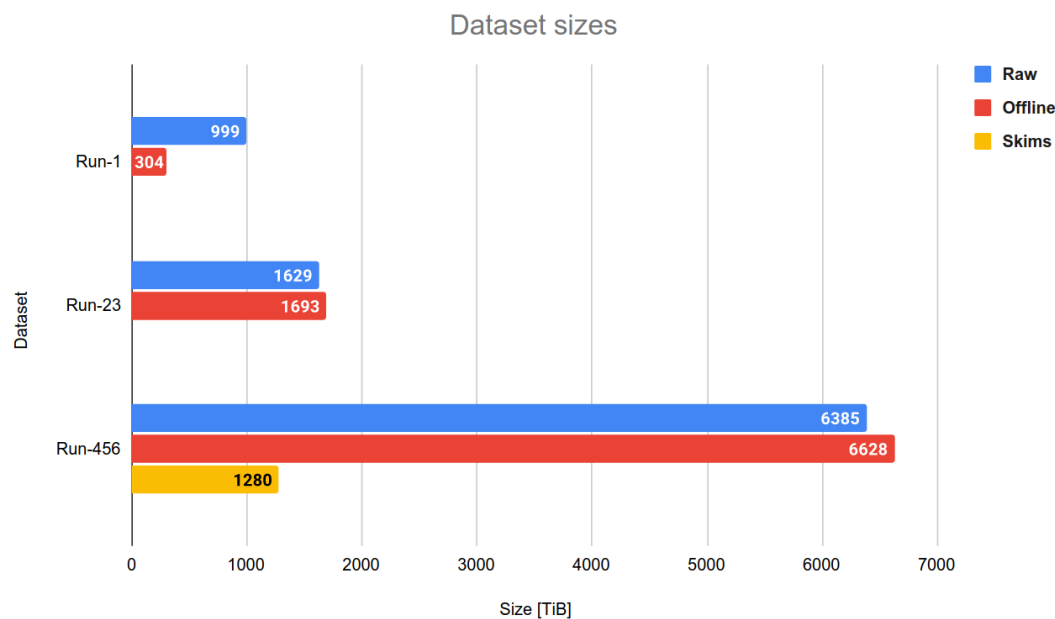
Muon g-2 status

- Muon g-2 Experiment finished collecting 6th and final run on July 9th, 2023
- Run-2/3 results published on Aug 10th
- Run-4/5/6 data now fully produced and ready to be analyzed
 - 75% of the experiment statistics
 - Run-4/5/6 publication expected in the first half of 2025



Our data

- Run-456 **offline** data is ~ 6.5 PiB \rightarrow Too large to be staged all at once on the public cache pool
- Our solution: create a **skimmed** version of the data by keeping only the minimal dataproducts that would still be useful for most of the analyses
 - Result: 1280 TiB ($\sim 5.2x$ reduction), 95% complete right now
- Still, quite large amount of data, and some full **offline** datasets are still needed for complete systematic studies



Storage options

- For Run-23 analysis, we kept the **offline** data (1700 TiB) staged and pinned on the public *readWritePools* for two years until publication
- In order to allow smooth production of the 6+ PiB of **raw** Run-456 data we requested and proficiently used a dedicated private disk cache (*GM2Pools*, now 1400 TiB in size)
- Run-456 production is finished, so we are planning to store the **skimmed** data (1300 TiB) on GM2Pools
 - Data always available to the 7 analysis teams with no more prestaging needed. Analysis work can proceed at the fastest pace possible, which is in line with lab management's priorities
- We desire to keep 3-4 full **offline** datasets staged too for complete systematic analyses (~700 TiB)
- To allow this we would like to request an expansion of the GM2Pools cache to 2100 TiB → 1300 + 700 + 100(buffer)

Summary

- g-2 is starting to analyze the massive (and final) Run-456 dataset
- Skims with size reduction of 5.2x implemented to keep data handling manageable
- Proposing to store all the analysis-relevant data to the GM2Pools disk cache, requesting expansion to 2100 TiB
- This would drastically reduce the g-2 footprint on the public cache
- In the next months, we will need to prestage all the Run-23456 offline data (8 PiB) on the public cache, in small chunks, each chunk pinned for less than a month, for other productions
- After Run456 analysis complete (exp. first half of 2025), we can give the entire GM2Pools cache pool back
 - We are willing to be flexible (if we get this additional disk space) to give some back before we complete the analysis if there is a site-wide disk crisis
- We don't need GM2WritePools anymore (~110 TiB) → could be converted to GM2Pools?