



# TOWARDS A HOLISTIC VIEW OF BATCH JOBS, DATA TRANSFERS, AND STORAGE

July 24, 2024

# THE CRISIS

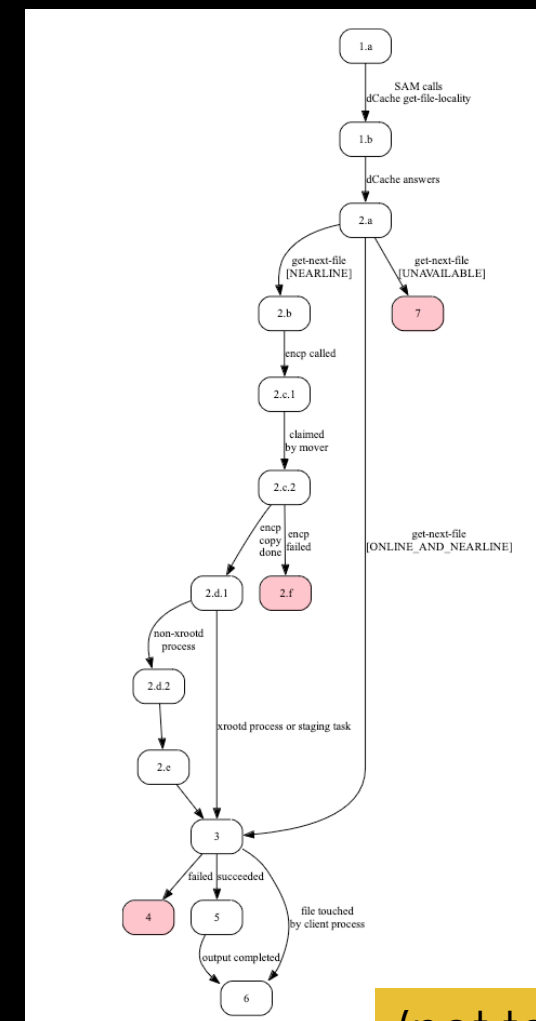
- In early October 2022 there were issues raised by some experiments about the performance of their data staging jobs.
- Urgent meetings were convened to determine the cause of the “tape crisis” and to devise a plan for fixing it.
  - “In the last two weeks, our ability to stage data off of tape is at a standstill ...”
  - “Problems starting in February ...”
  - “Unbearable since August ...”
- The urgent issue was resolved as quickly as possible.
- Less time-critical, but still important, were issues regarding monitoring:
  - Users have a hard time understanding what is going on between their jobs and the data handling and storage systems.
  - System managers have detailed views of individual subsystems, but it is hard to get a holistic view of the interactions between the subsystems.
  - It is hard to answer the question “Is the system ‘working’”.

# THE INVESTIGATION

- Marc Paterno and Saba Sehrish were tasked with talking to users, service providers, and management to understand how systems work and make suggestions for monitoring.
- They produced a very thorough report detailing how the batch, interactive, data handling (SAM), disk storage (dCache), and tape storage (Enstore) systems work together.
- They made a number of suggestions for both general monitoring and specific visualizations.
- [Data Handling Monitoring Report](#) (Sharepoint link)

# RECOMMENDATIONS

- Visibility into system states
  - SAM project and related batch submission states
  - SAM process and related batch job states
  - File locality and staging states. Diagram shows how complex this can be, *for a single file*.
  - Service states – disk and tape queues, movers, drives, pools
- Visualizations
  - File size histograms
  - File state dot plots
  - File-oriented performance data table
  - More...



(not to scale)

# THE RESPONSE

- That was January 2023. What have we done since?
  - No magic wand, and limited effort
  - All of this hinges on having the data and being able to correlate it.
  - We have lots of logs – SAM, Enstore (as of last week), IFDH, HTCCondor.
    - Legacy logs are unstructured, hard to parse
    - And then, we still have to correlate events across systems
      - Jobsub job id
      - PNFS id
      - SAM project id
      - POMS task id
    - Some we can find correlations for
  - Once we have the data and can correlate it, what about visualizations?
    - Online monitoring is different from offline data analysis. Different tools, different paradigms.
    - We do need to better cater to offline data analysis use cases.

# AND EVERYTHING IS CHANGING

- Enstore is on the way out
  - CTA is on the way in
  - We are working on a monitoring-first approach
  - Building off CERN experience
- DUNE and mu2e won't use SAM – instead they have "Shrek" (Metacat, Data Dispatcher, Rucio and friends)
  - A lot of work to be done there, monitoring (to the extent discussed here) was not included in the design.



# OUR PRIORITIES

- Work on collecting missing data, making it useful, and correlating it.
  - Work with service providers to integrate better first-class monitoring, especially tracing to help correlate events and track state.
  - Provide better interfaces for offline data analysis with familiar tools.
- 
- ... keeping the ship afloat and on course with a skeleton crew.