18 October 2024

Report LHCb

Lucio Anderlini
INFN Firenze

Matteo Barbetti
CNAF

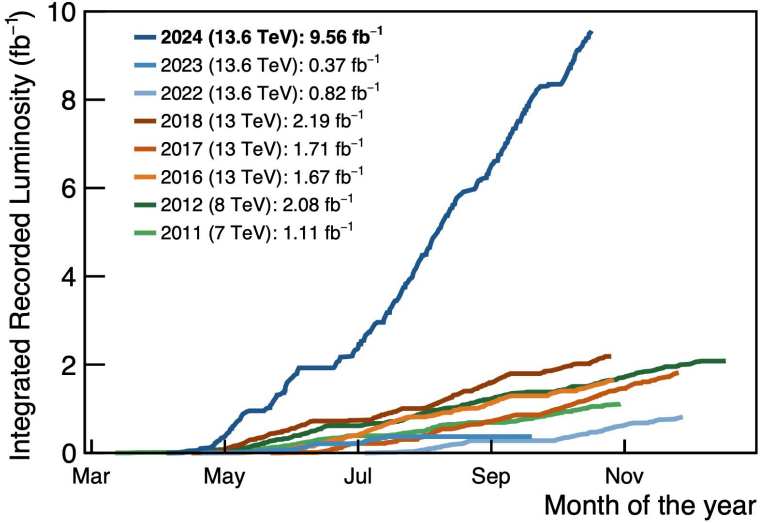
LHCb completed the *pp* run for this year

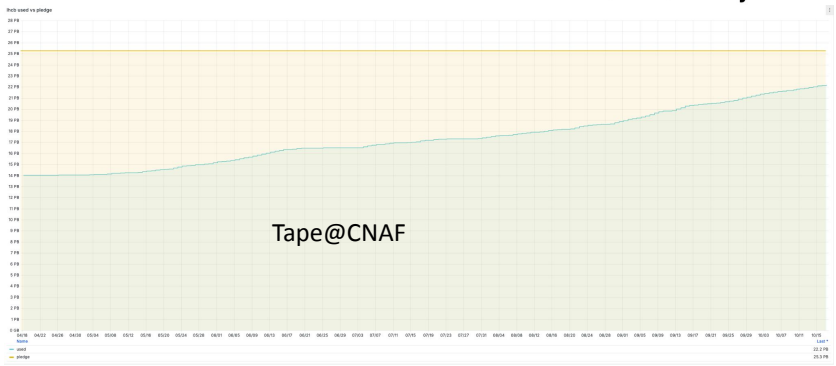
After a *slow start*, LHCb has continuously streamed data to the Tape at CNAF since May 2024.

This process encountered some **limitations** due to the well-known issues with OceanDisk (GGUS:167716), which **particularly affects** the LHCb's dataflow.

Among the solutions implemented/investigated:

- preparation of a **tape buffer** hosted on hardware not affected by OceanDisk issues
- creation of a **hot-storage region** based on NVMe to store data processed by Sprucing/Merge jobs
- limiting the number of jobs with **high I/O rate** (e.g., WGprod/Sprucing/Merge)
- limiting the number of concurrent **FTS transfers**



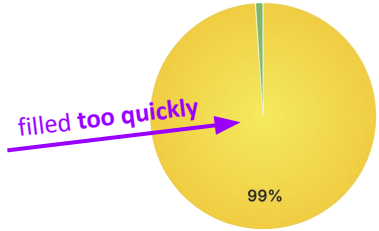


Hot-storage region

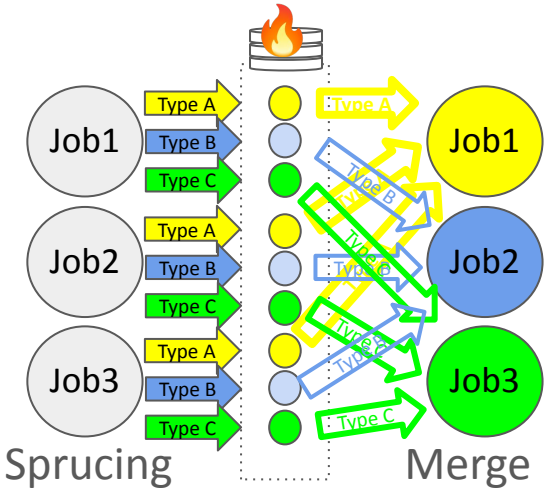
The scope of the ***hot*-region** is to (temporary) store files produced by Sprucing jobs which, processed by Merge jobs, are combined into a single merged file, then deleting all the input files used.

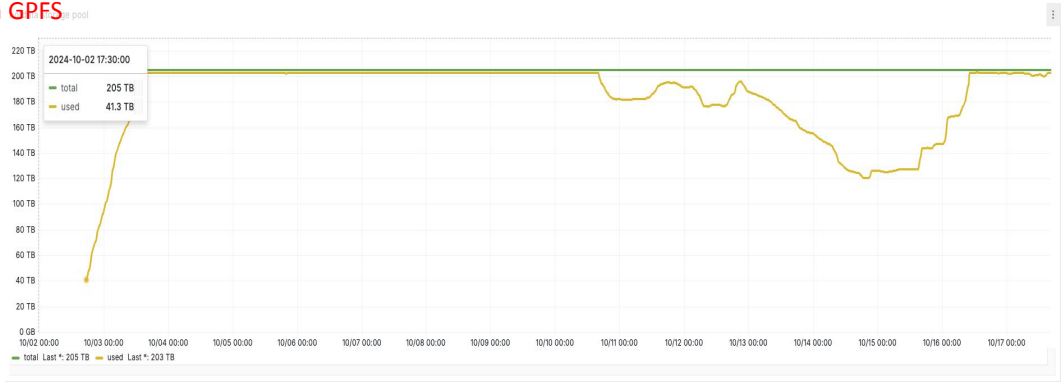
To pursue this goal, GPFS has been configured with a dedicated **placement policy**:

- `/gpfs_lhcb/disk/lhcb/**/*.dst` → NVMe (but this is the whole disk)
- `/gpfs_lhcb/disk/lhcb/buffer/**/*.dst` → NVMe (correct path)



`buffer/` isn't a fileset hence **no policy** can be set in GPFS





Proposed intervention to make LHCb-Buffer a fileset

LHCb is now taking data for the pA program:
the upcoming month should be relaxed on computing resources

Then we will start the end-of-year reprocessing (extremely IO-intensive).

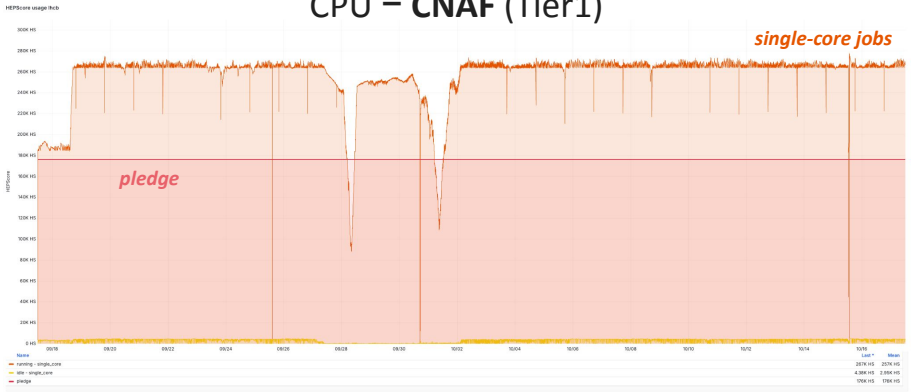
I propose:

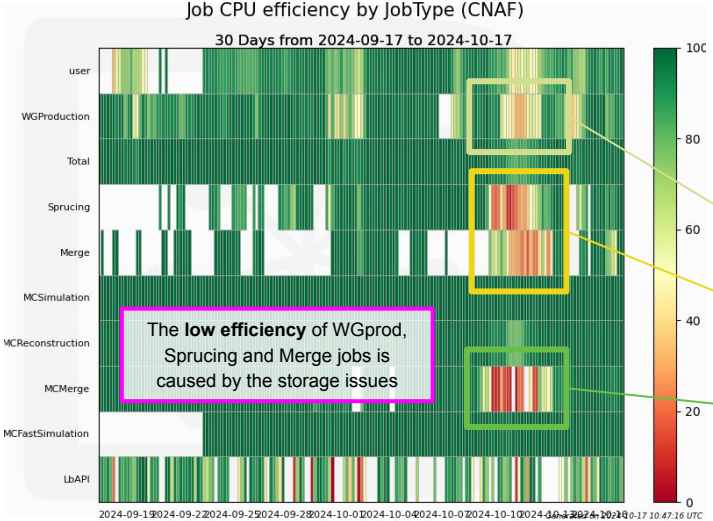
1. LHCb uses next week to clean-up the buffer (should become less than 100 TB)
2. let's schedule one-day storage downtime on w/c October
3. let's ban access to CNAF storage for jobs in the farm and stop FTS transfers
4. let's move the hot storage to its FileSet and implement the placement policy for *.dst files
5. we estimate we will need 1 PB of buffer for end-of-year activities, can we use more non-OD storage?

CPU usage

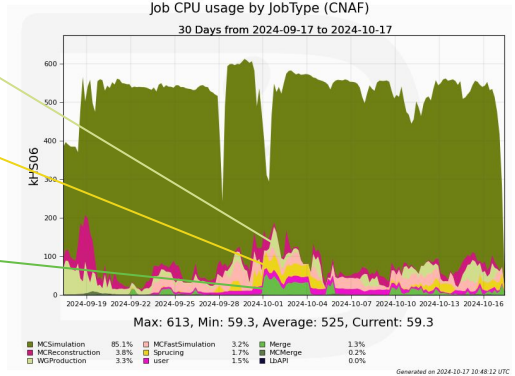
Leonardo nodes have ensured a **significant overpledge** during the last month, even if most of the resources are used for simulation production due to the well-known storage issues

CPU – **CNAF** (Tier1)

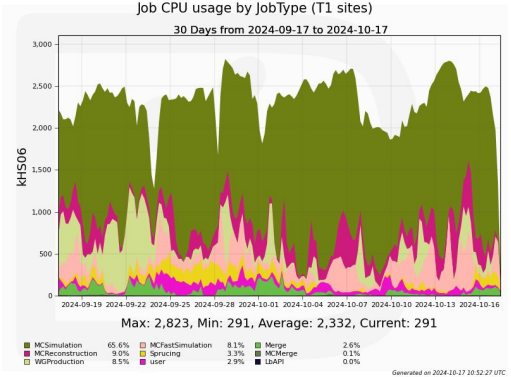




from DIRAC: **CNAF** (Tier1 + Tier2)

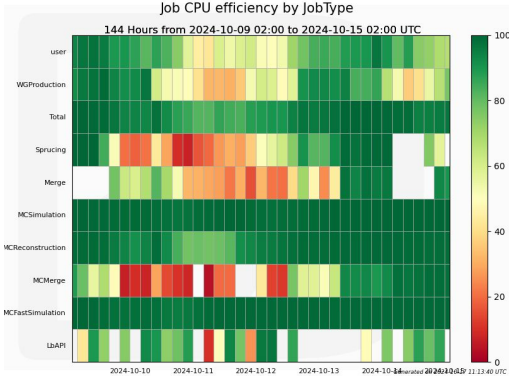


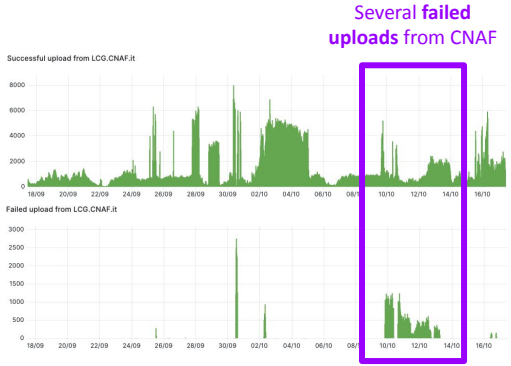
from DIRAC: **all** Tier1 sites

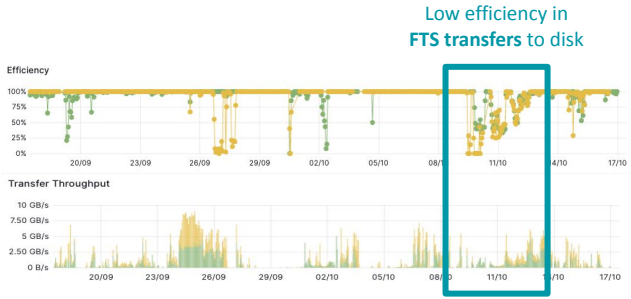


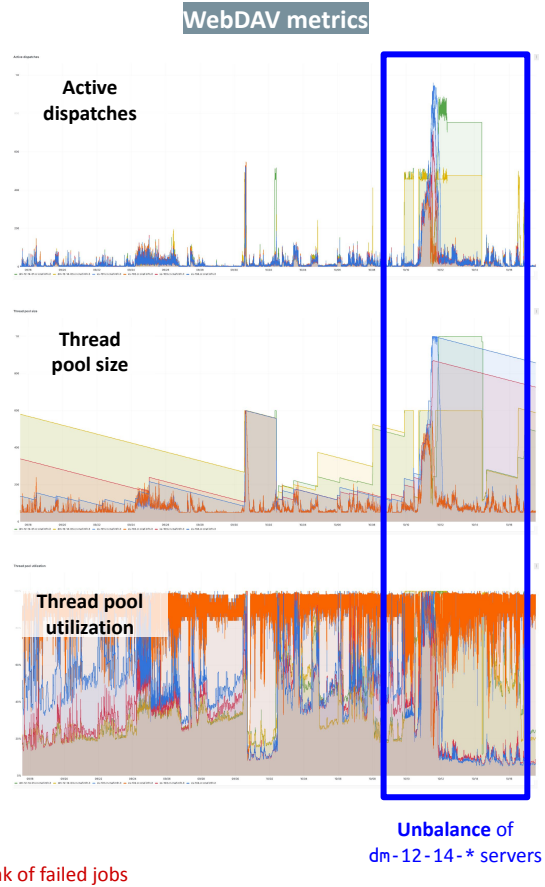
10th October event

Several data transfers failed (GGUS:168542)









History of corrupted files

- **11-13 Sept 2024** Tier-1 experienced some **network issues** (GOCDB:35910) due to a hardware problem with a core switch Arista at Technopole
- **25 Sept 2024** A network intervention was needed to restore the nominal conditions after the problem raised on September 11th → **the intervention caused network instability**
- **1 Oct 2024** **Checksum inconsistency** between source file and file at CNAF found by Chris (ELOG:39126)
- **2 Oct 2024** Comparing source file with the one at CNAF, it seems that some **corruption actions** occurred at CNAF at certain point of the data transfer, **probably due to the network issues**
- **16 Oct 2024** The storage team performed an **intense research campaign** checking more than 80k files on disk and 3.5k files on tape and identified 7 corrupted files (GGUS:168495)