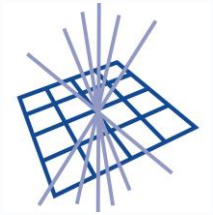
Science and Technology Facilities Council
Rutherford Appleton Laboratory



LHCb status

GRIDPP 53

alexrg

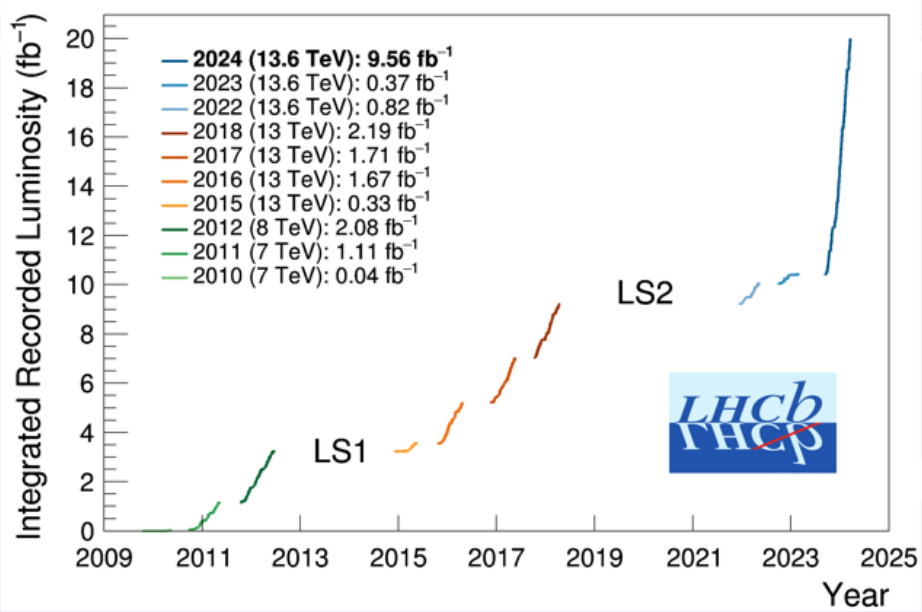
Contents

In this talk I'll try to cover the following topics:

1. LHCb 2024 Data Taking campaign.
2. RAL developments.
3. LHCb news.
4. LHCb problems.
5. LHCb resource usage by the UK sites.

2024 Data Taking campaign

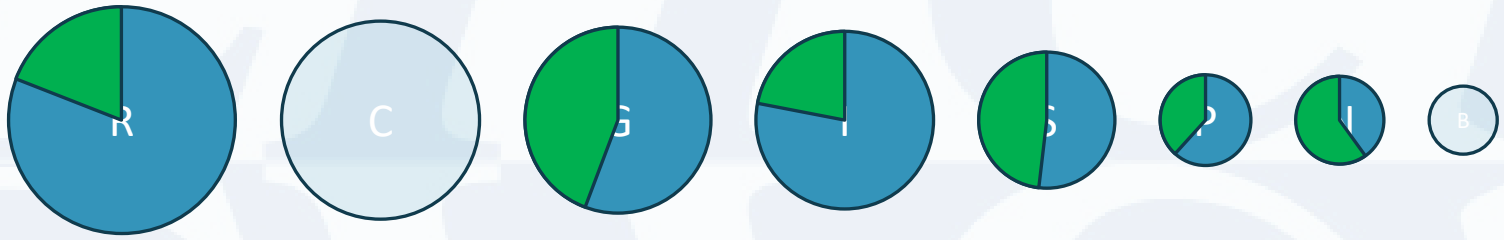
- The 2024 Data Taking campaign was very successful for LHCb
  - Recorded more luminosity than in all previous years combined
- That means that quite a lot of space was needed to accommodate the data
  - 4 week run extension increased the demand even more



2024 Data Taking campaign: storage

LHCb uses 8 Tier-1 sites. By September 2024, only a few of them had enough free space (disk) left..

Tape*



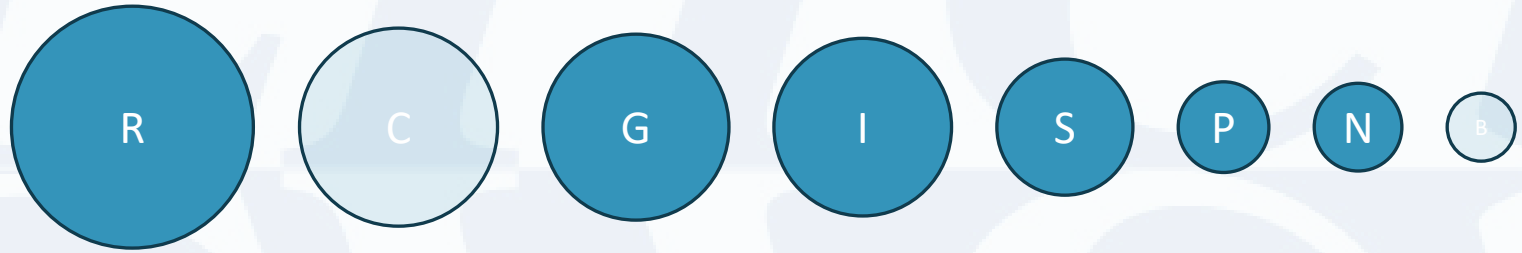
Disk*

*Diagram areas are proportional to the amount of pledged resources

2024 Data Taking campaign: storage

RAL Tier-1 saved the day by releasing the 2025/26 disk pledge in September and increasing tape allocation. Many thanks for this!

Tape*

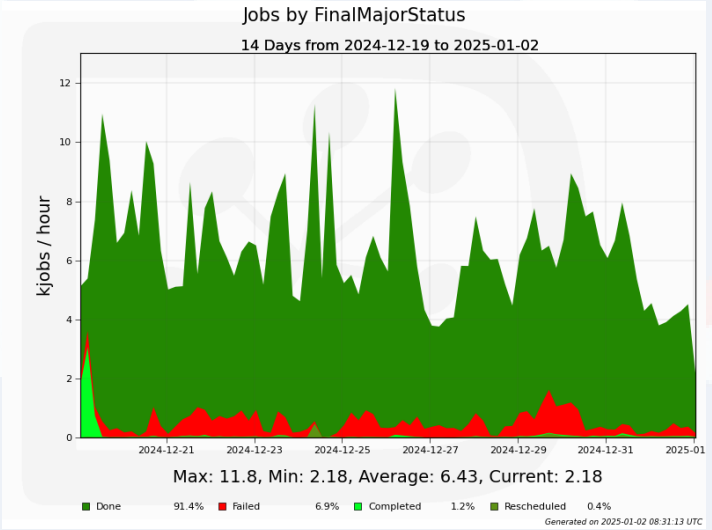


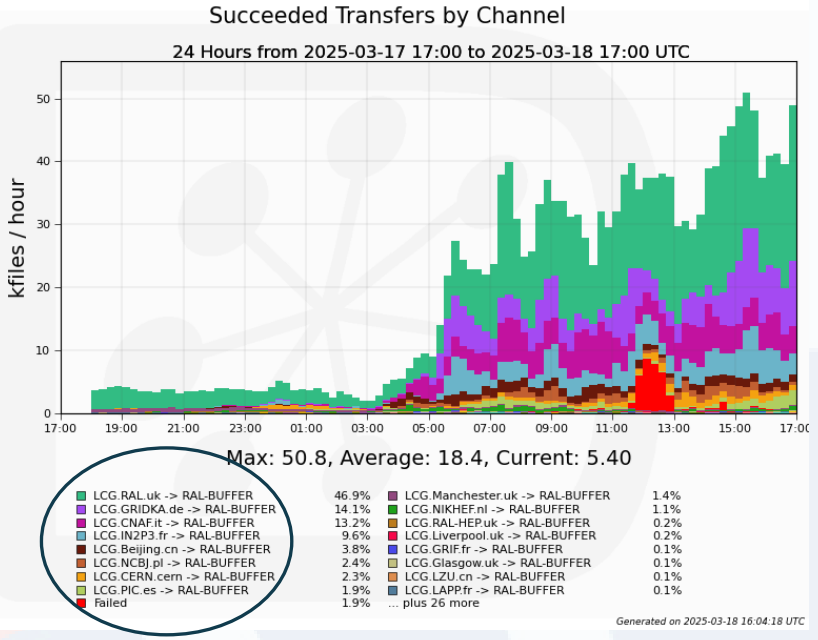
Disk*

*Diagram areas are proportional to the amount of pledged resources

2024 Data Taking campaign: reprocessing

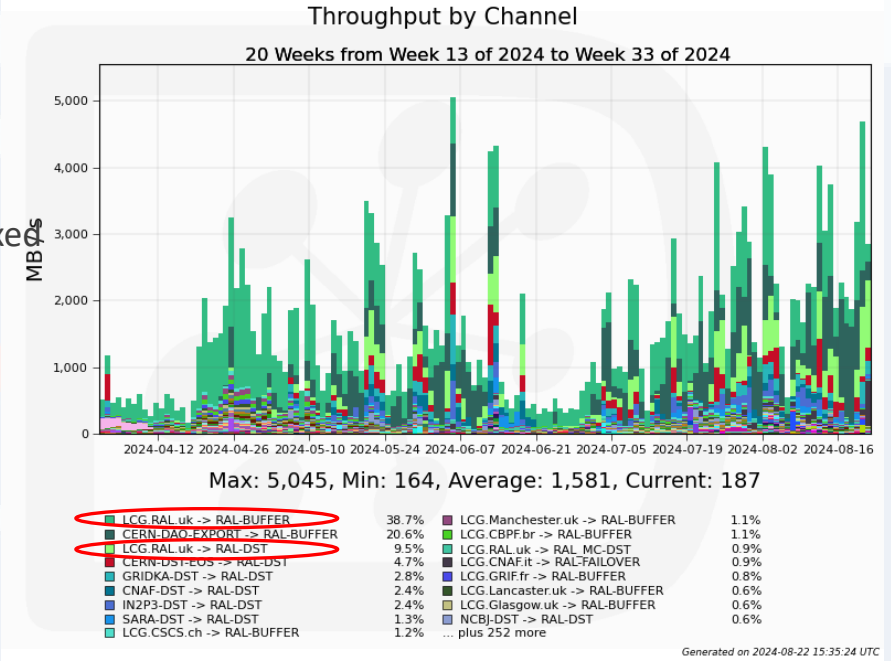
- A bit chaotic this year
  - EOS-pilot (test SE at CERN) was used for Turbo stream reprocessing
    - QOS of the SE is lower, it caused some problems
- Incident at RAL during the Christmas break (+second occurrence afterwards)
  - Pilots were failing due to nproc limit set by xrootd
    - Not RAL's fault
    - Can not be fully resolved
    - Fully mitigated at RAL



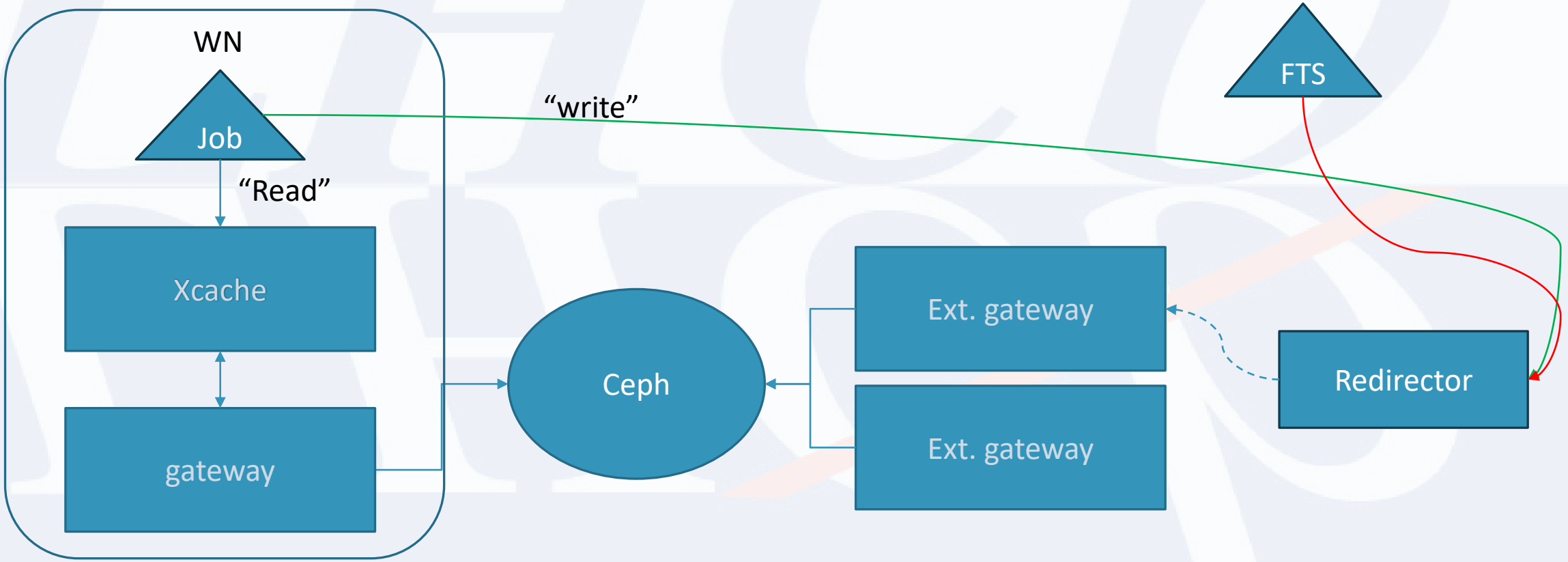


Xrootd development at RAL

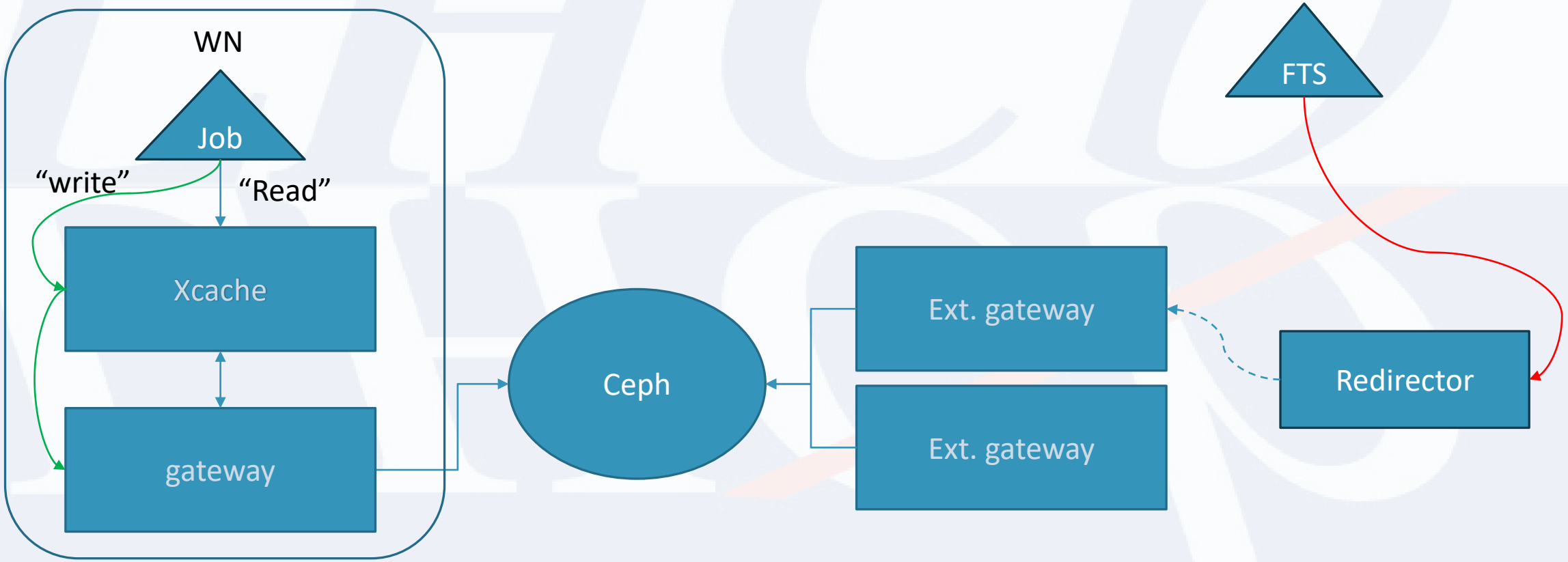
- The majority of LHCb uploads to ECHO come from the batch farm
- Each WN has its own gateway and xcache
- Since xcache is read-only, this local Xrootd infrastructure was used only for reads
- Not anymore!
  - LHCb jobs at RAL now upload their output data to ECHO via local gateways
    - Checksum requests (readback) are still redirected to the external ones, this is to be fixed
      - Easy fix, though



Writable WN gateways

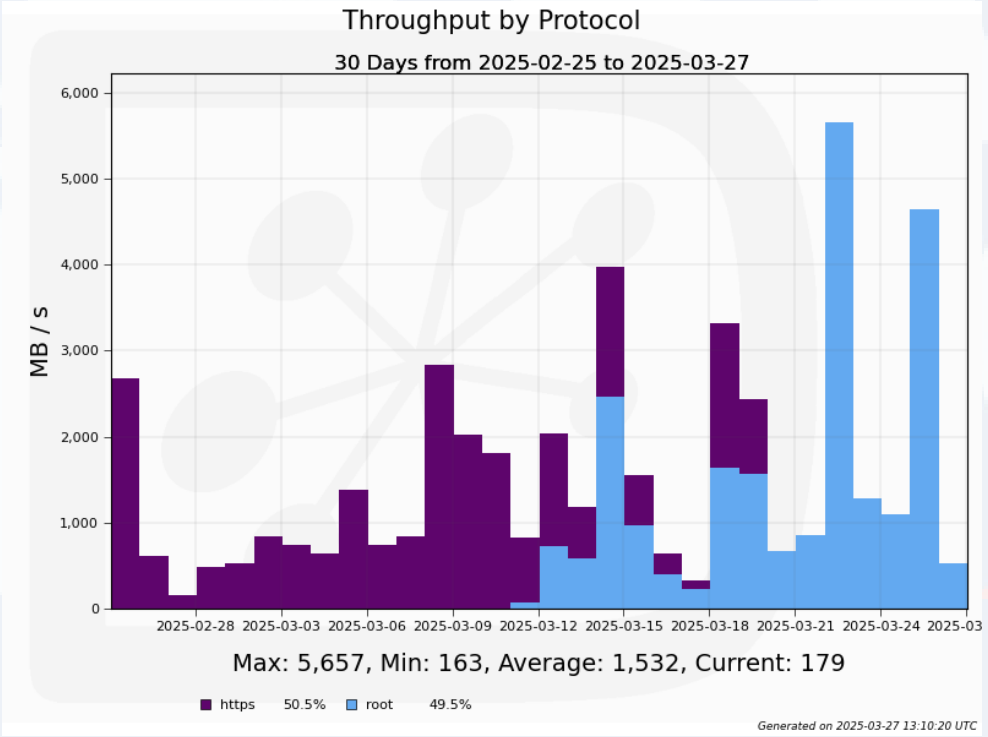


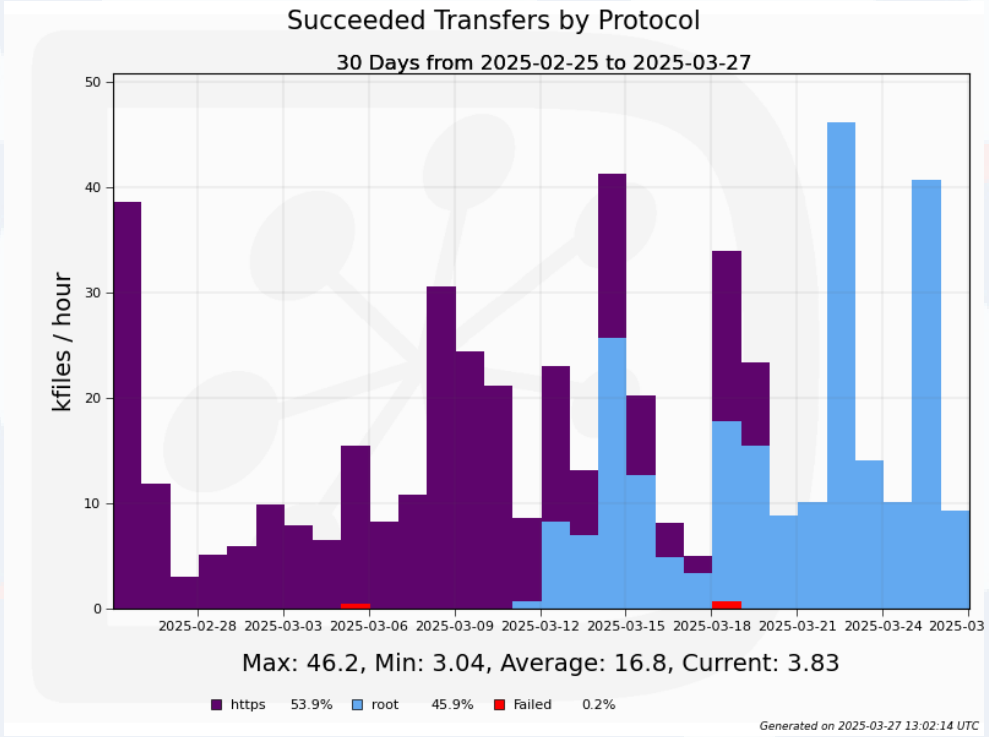
Writable WN gateways



Writable WN gateways

- The new setup reduces the impact of a gateway failure (once checksum [readback] redirection is deployed), removes load from external gateways and maybe more..





LHCb News: future plans

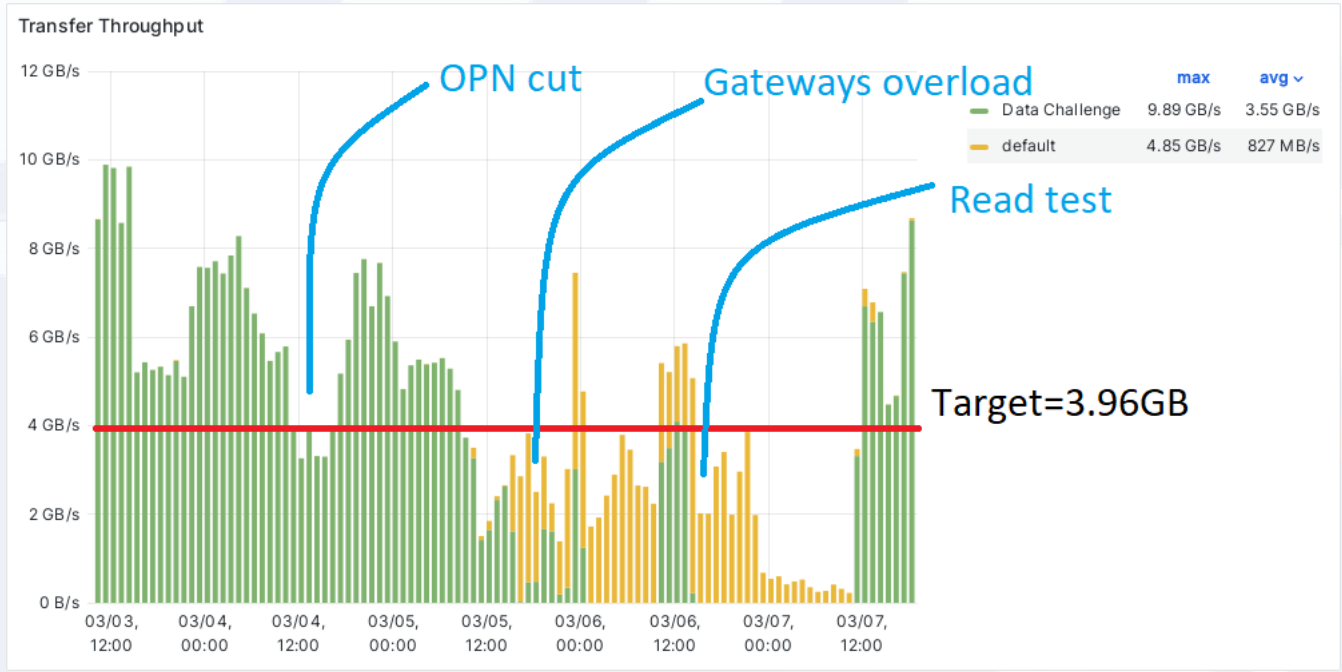
- 2025 Luminosity target for pp run is 12 fb^{-1}
  - 30% More than in 2024!
  - Data taking should tentatively start ~beginning of May
- 2026 is a data-taking year as well
  - Therefore some resources are needed as well
  - The exact numbers are not finalized yet

LHCb News: UK DC

- Exceeded DC24 target rate
  - Had not been achieved during DC24
- Used new inject tool, similar to ATLAS and CMS
  - Might be useful for future challenges, especially for smaller sites
- OPN cut reduced the throughput, but it was still very close to the target
- Gateway overload reduced the performance as well
- “Read” test (ECHO->CERN) failed completely
  - Due to problems with EOS-PILOT SE at CERN

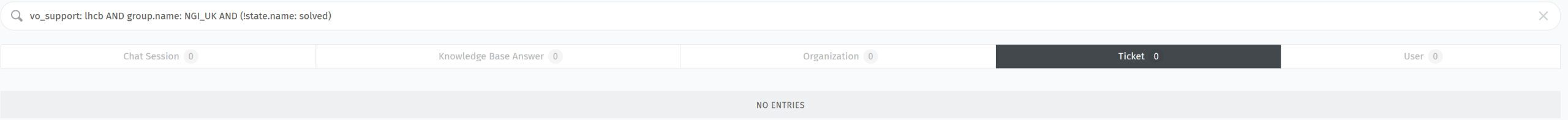
LHCb News: UK DC

- UK DC in one picture



Problems

At the time of writing, there are no opened LHCb tickets against UK sites!



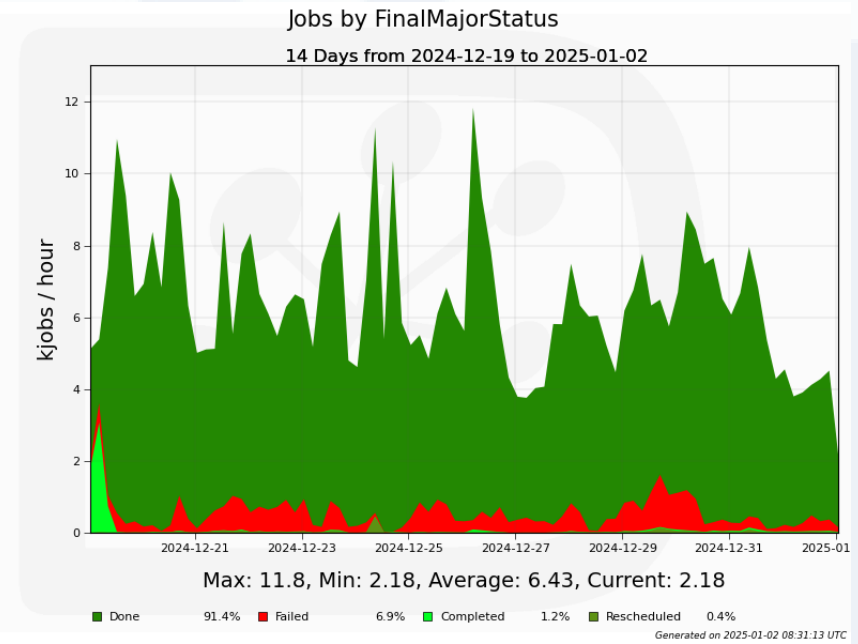
At RAL, apart from a few minor incidents, there was one serious issue for LHCb – the problem with LHCb pilots due to xrootd nproc limit. Not RAL's fault though.

Problems: Xrootd nproc issue

- At Liverpool jobs were failing due to failure in thread creation (GGUS ticket)
  - LHCb jobs were setting nproc limit to 30k, not clear where exactly
- During Christmas break we noticed the same problem at RAL
  - 30k limit excess on “big” (384 cores) WNs
    - The default 2M limit was overridden, unfortunately..
- Turned out that the limit is set by xrootd client whenever it is imported 😱
  - Limit is 30k, and hardcoded, so no way to change it 😭
  - Very annoying, e.g. simple code like this triggers limit set

```
import gfal2
ctx = gfal2.creat_context()
# Limit is set if gfal-plugin-xrootd is installed
```

  - Limit is “global”, and threads are considered as processes as well
    - So, if the user is running a lot of threads, after (e.g.) gfal context creation it may be problematic to create additional threads/processes
    - That’s why jobs/pilots were failing
  - Github issue, fix should be in 5.8



Problems: Xrootd nproc issue (backup)

- Can not wait for 5.8, however
  - Some LHCb software (e.g. for run[12] data processing) is frozen, so there xrootd version can not be changed
    - Only affects streaming jobs, so we are not affected yet, but might see this problem again in the future
- LHCb put a mitigation in place, something like

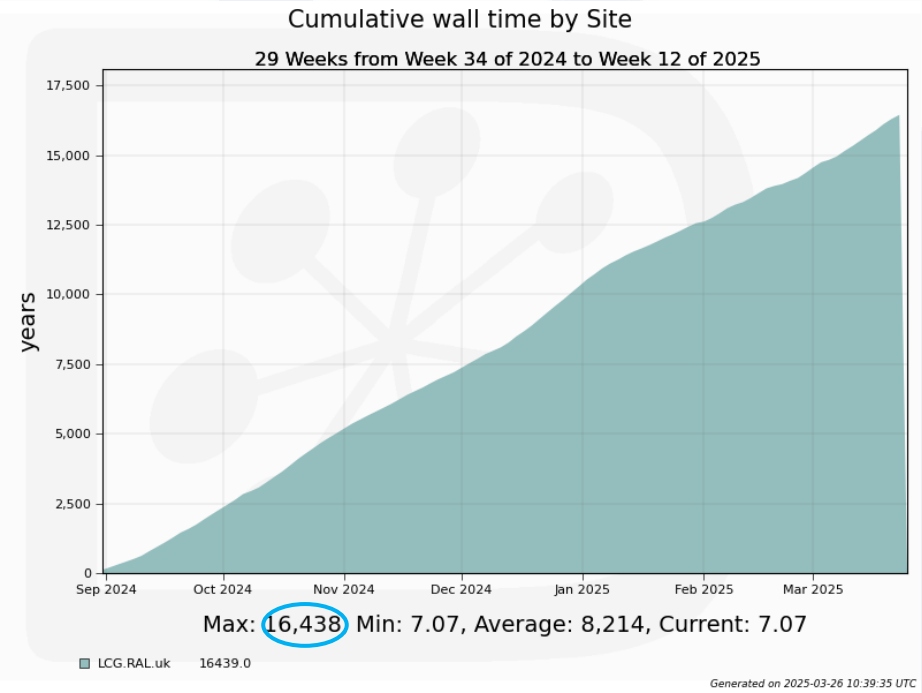
```
limit = getrlimit(NPROC)
ctx = gfal2.creat_context()
if getrlimit(NPROC)!= limit:
    setrlimit(limit)
```

- RAL mitigation is in place as well – disable `setrlimit` syscall inside the job containers
  - Not sure why, but xrootd client did not crash after this somehow!

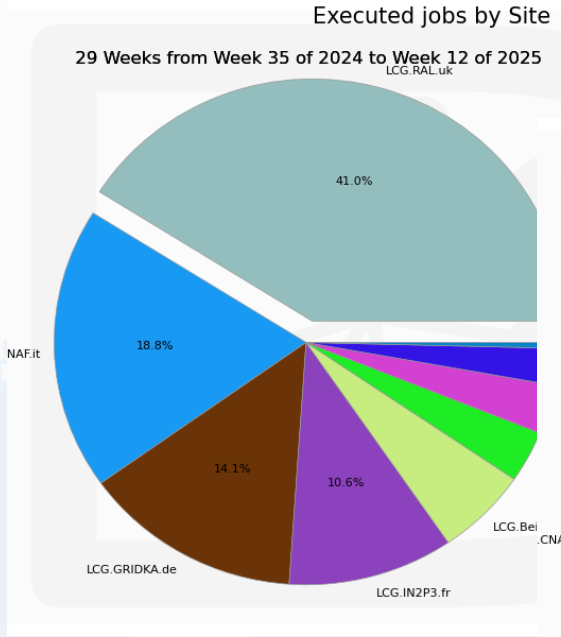
LHCb Jobs at RAL T1

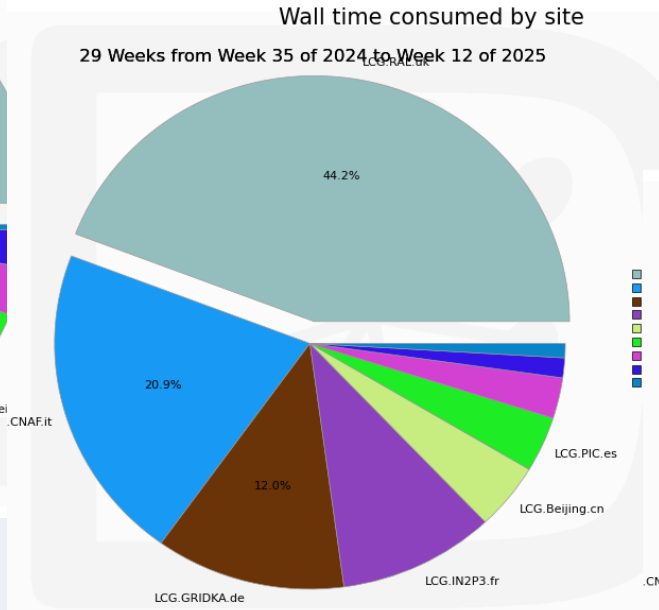
Nowadays it's a little tricky to calculate average HS23 consumption…

- DIRAC code for calculating normalization factor is being updated, so can not be trusted
- If we use static normalization factor from RAL, we have:
  - 16438*12.7/0.56 = 372790 HS23
  - Blue = Wallclock years; purple = Normalization factor, green = = years in the reporting period
  - Pledge is 180 kHS23, pledge for FY 2025/2026 is 228 kHS23

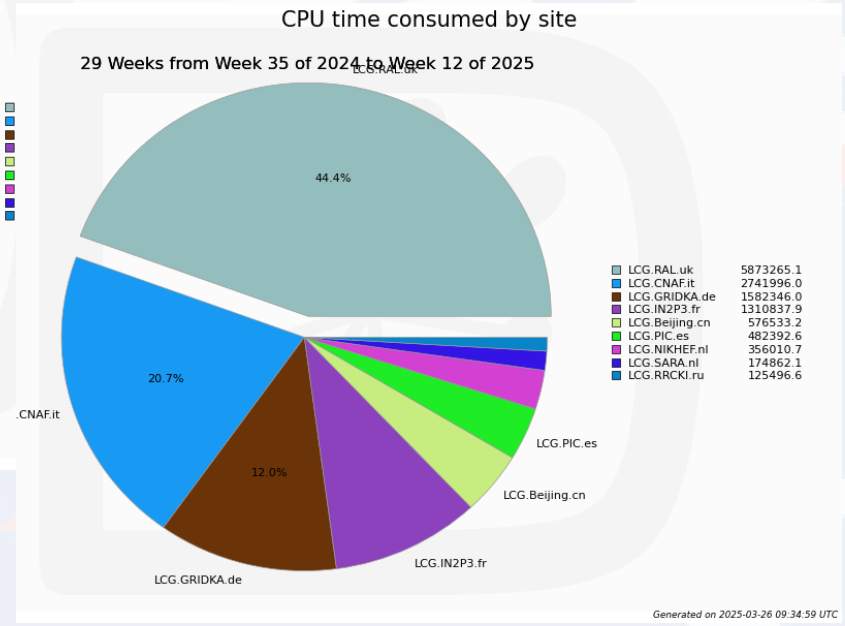


Comparison

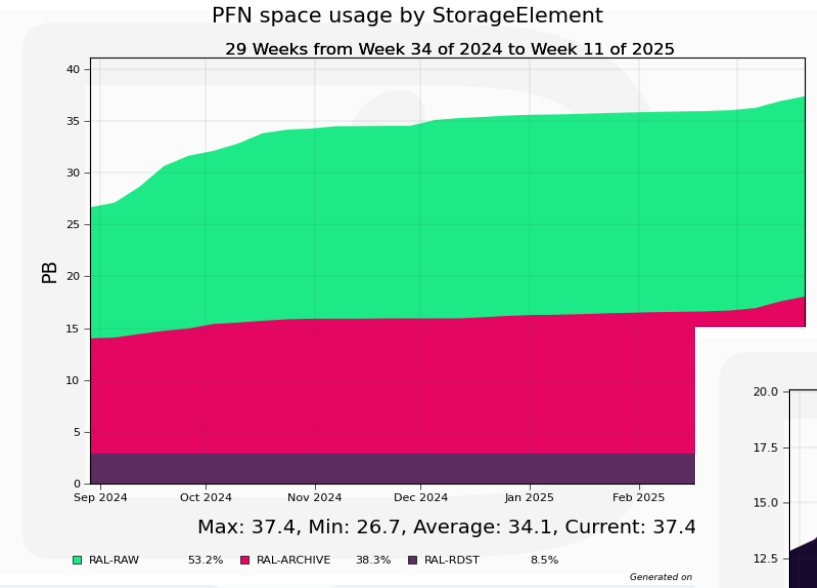


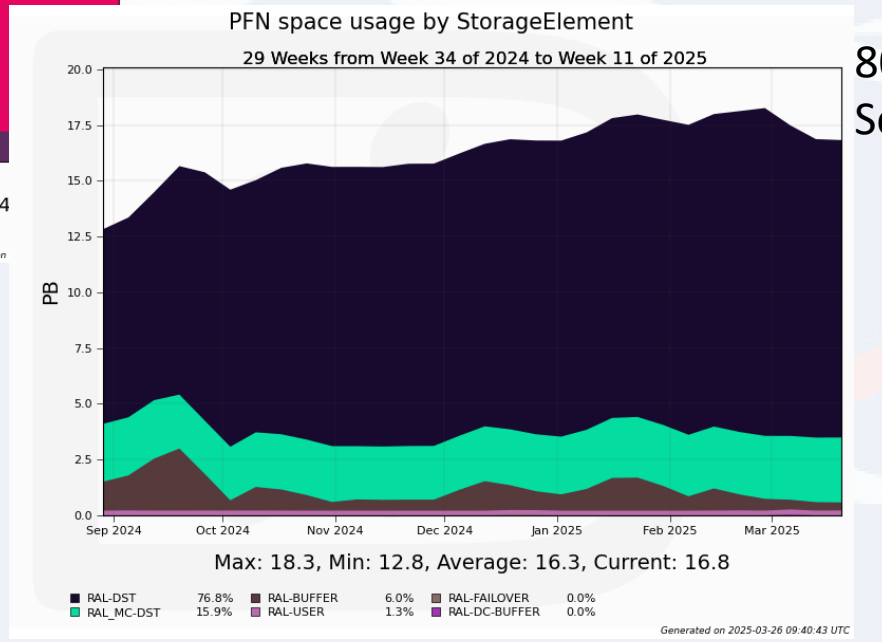


RAL has provided the biggest amount of CPU resources among all T1 sites, in terms of executed jobs (first plot), walltime (second), and cputime (third). Even more than the “usual” 30% share.



Storage usage



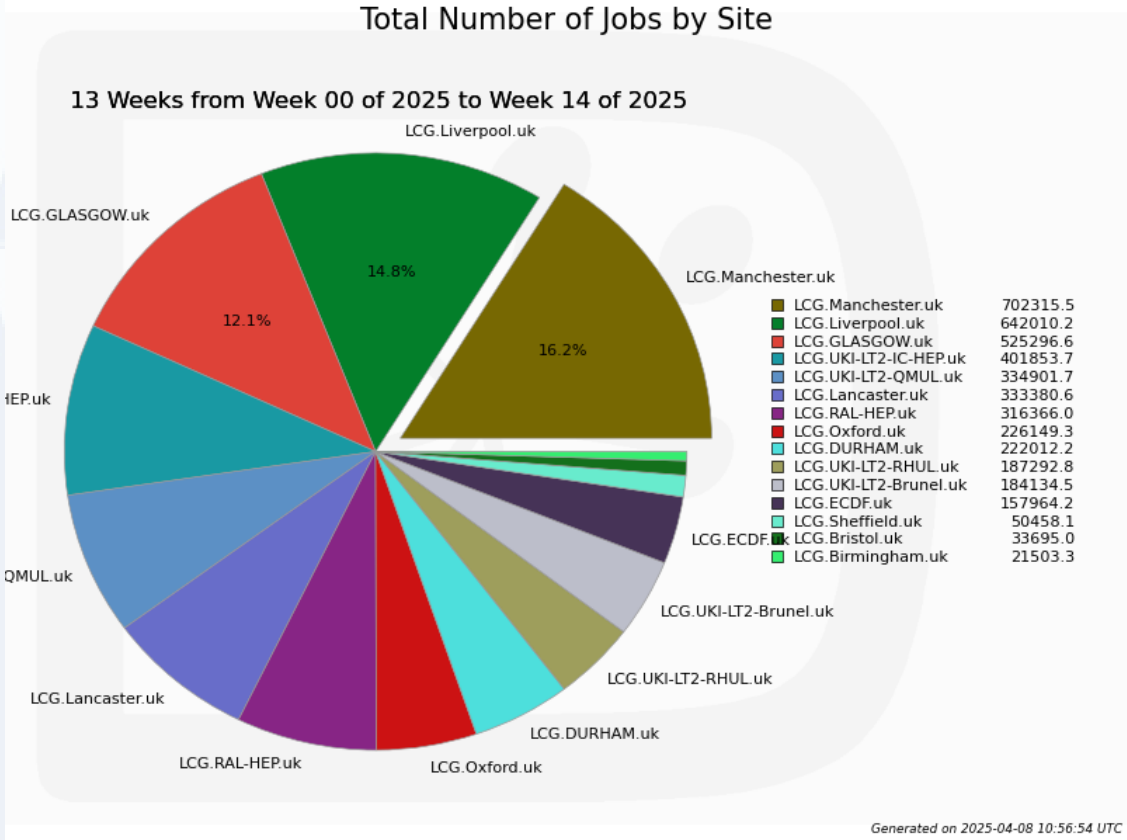


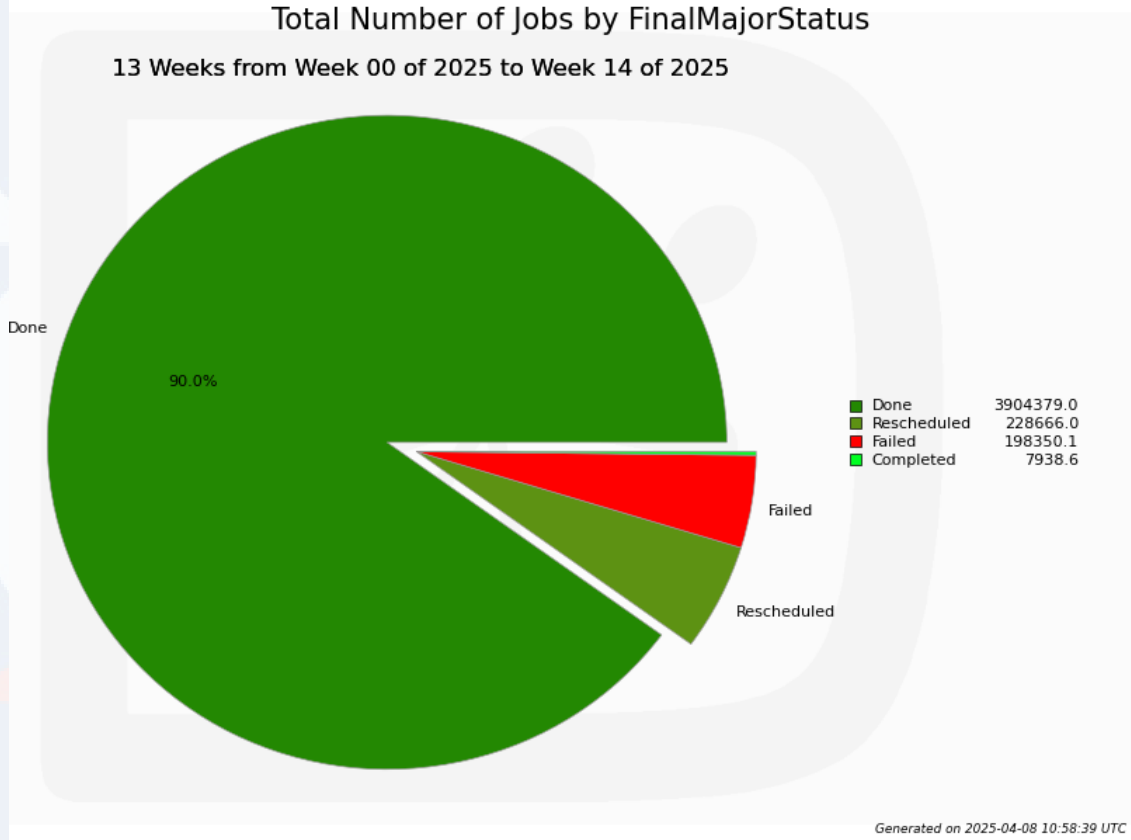
- Tape usage increased significantly (37.4PB). That is:
  - More than “initial” 2023/2024 allocation (33PB)
  - 95% of the additional allocation (39PB)

80% of the 2024/2025 allocation (released in September 2024) is used

Tier-2 statistics

After DIRAC downtime data only since Jan, sorry…





Conclusion

- GridPP saved the LHCb DataTaking campaign in 2024, many thanks for this!
- Smooth operations in the past 6 months.
- New developments and improvements are happening.

LHCb News: Dirac X (backup)

- The migration to Dirac X is ongoing
- There was a deep downtime for database upgrade/service installation last week
  - Full drain of the Grid
  - Some services are now delivered by Dirac X!
    - E.g. job sandbox management

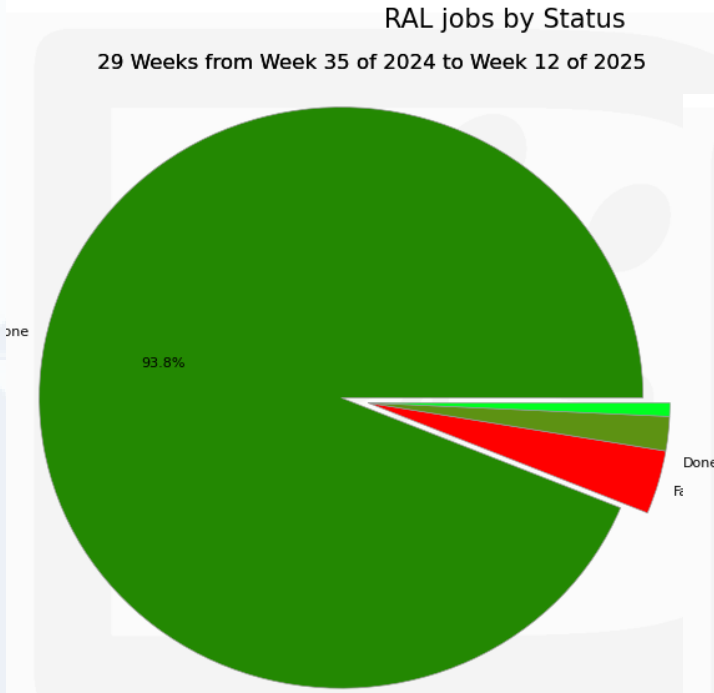
LHCb News: HLT storage (backup)

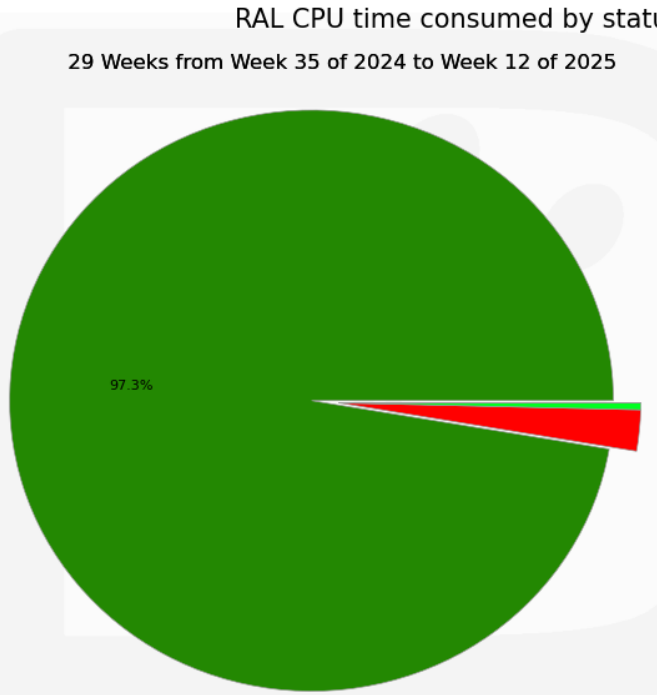
- There is a new storage being set up at the HLT farm for LHCb
- The interesting point is that it is ceph(fs)-based storage with xrootd in front of it
  - So very similar to the setup used at many UK sites!
    - E.g. Glasgow and Manchester

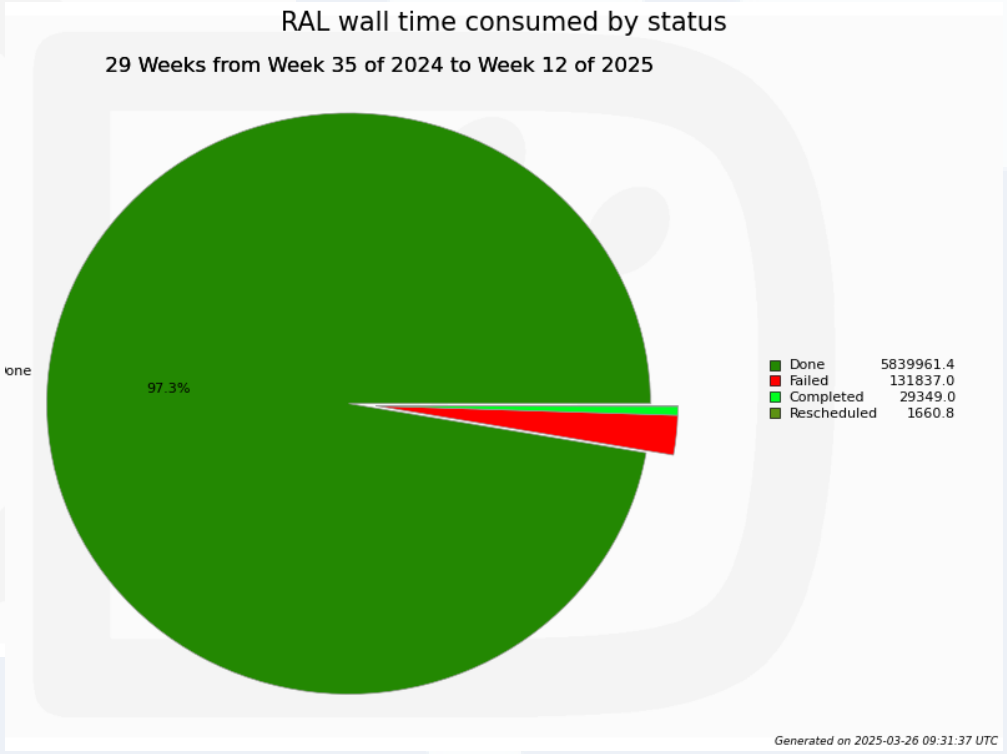
LHCb News: tokens (backup)

- Tokens are successfully used in production for HT-Condor CEs
- For ARC CEs we are waiting for ARC7
- For storage we are somewhere close to the beginning
  - ETF tests are progressing
    - May have first preprod instance working soonish
      - Delayed due to some issues
  - Production usage is more complicated..

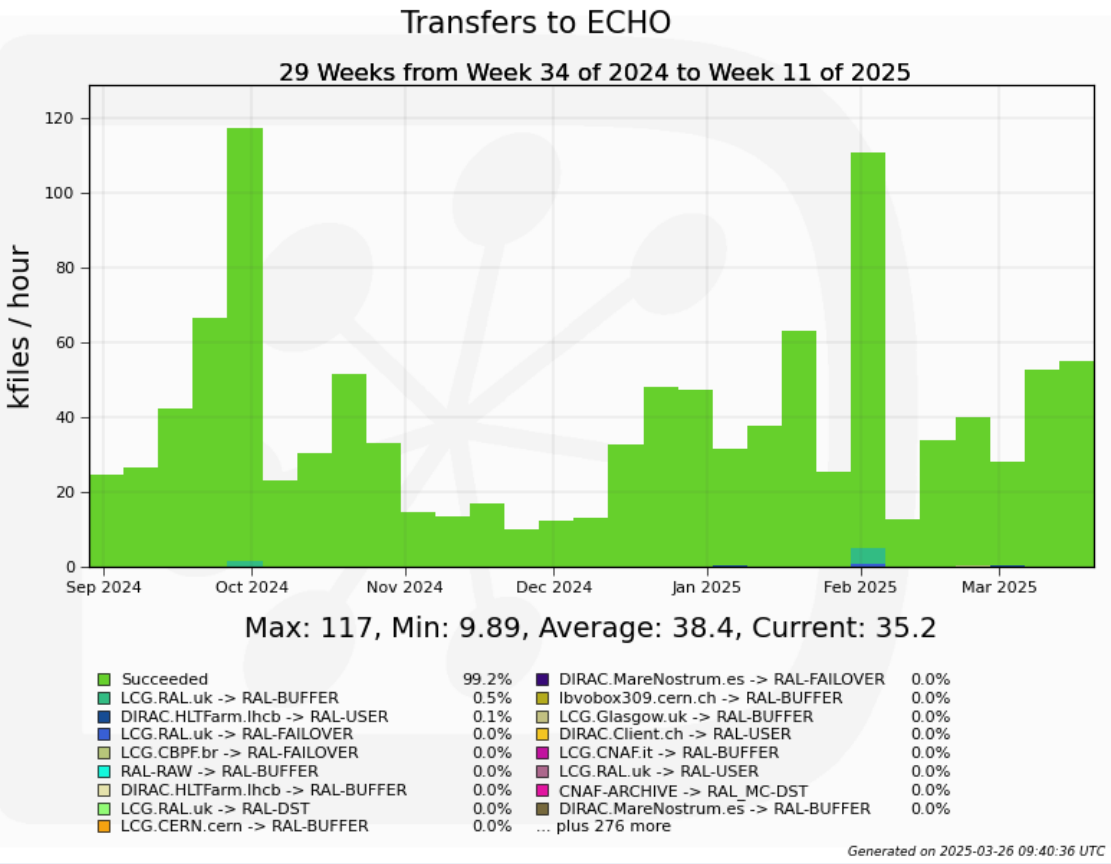
RAL T1 Job success rate (backup)

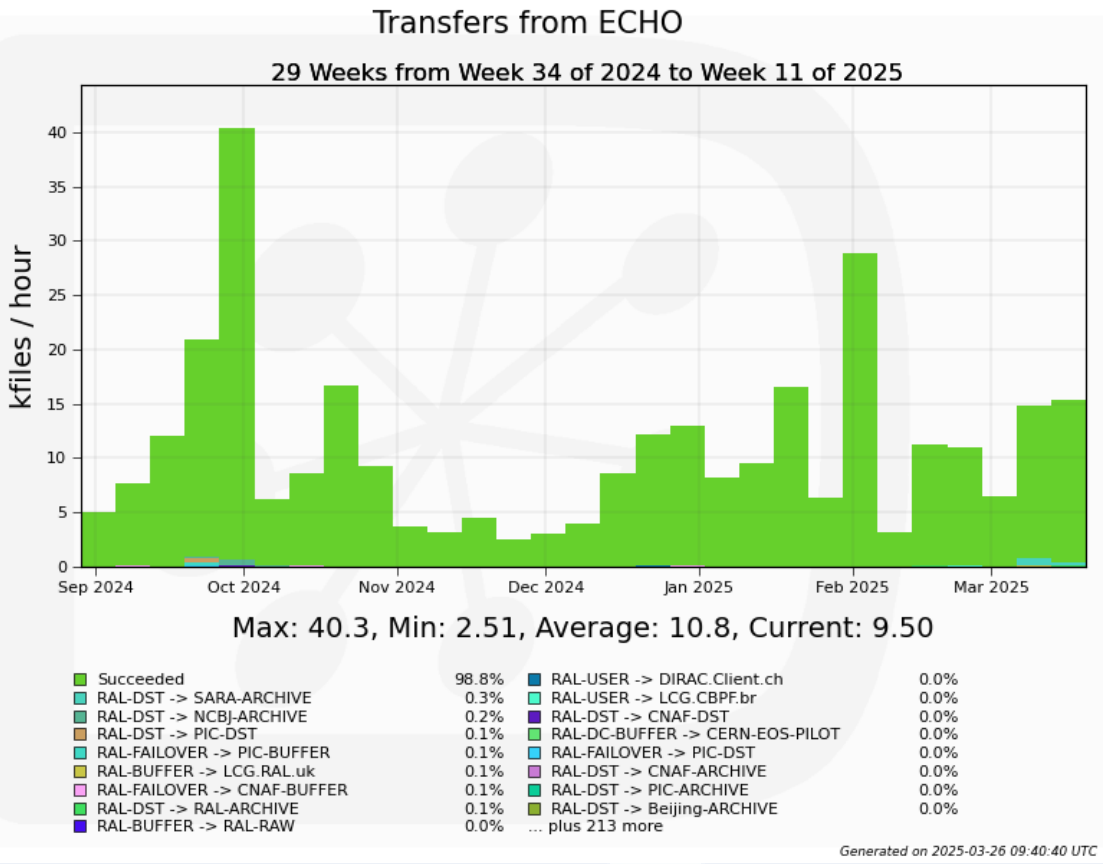
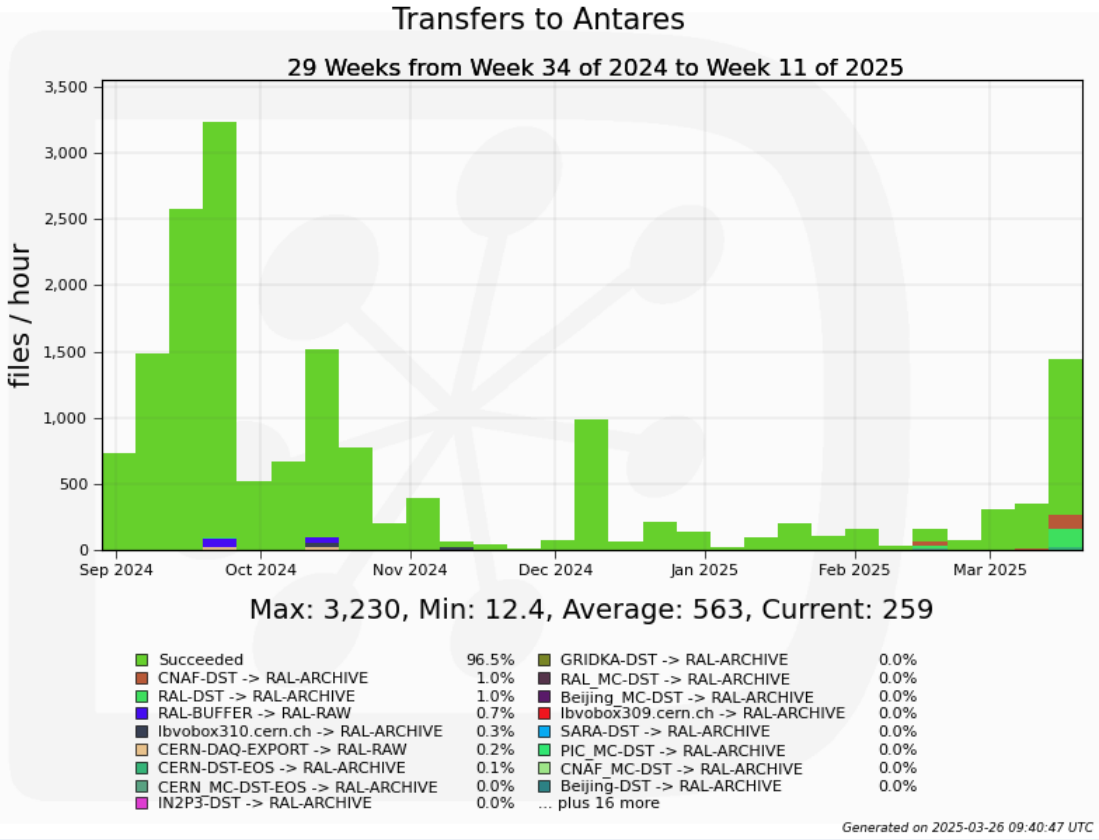
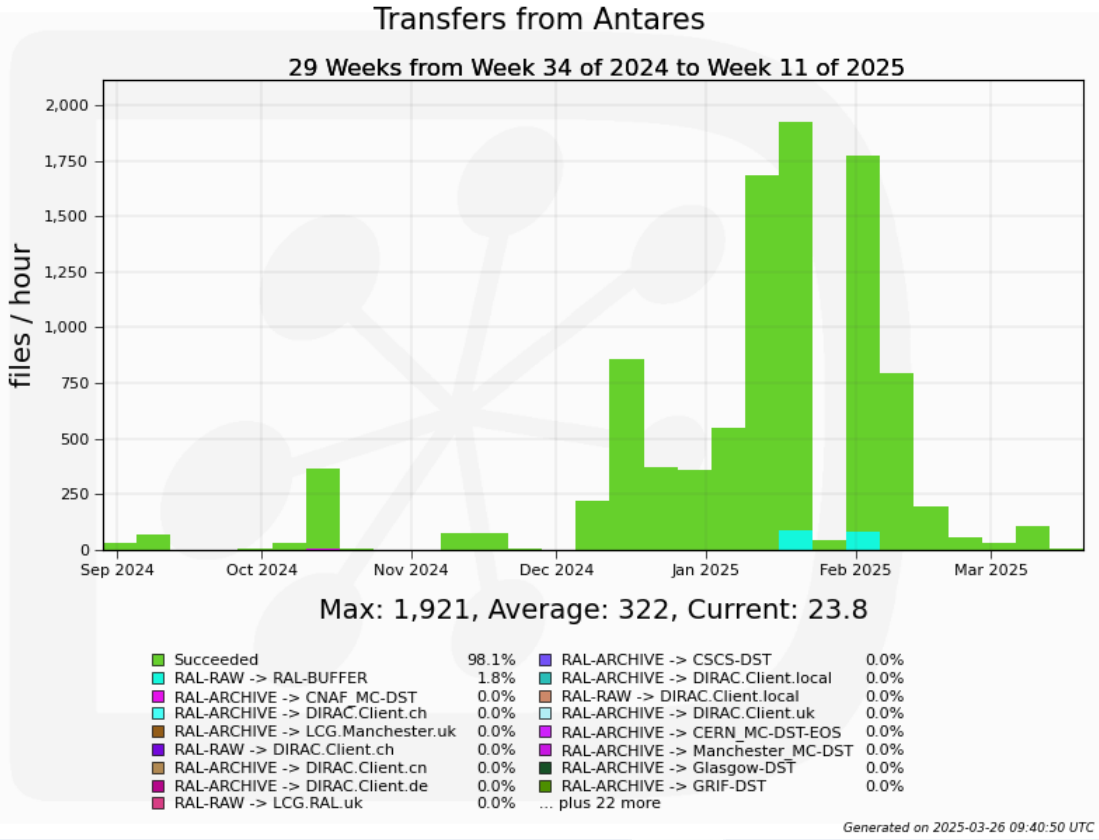






Transfers to ECHO (backup)





Transfers to Antares (backup)