Science and Technology Facilities Council

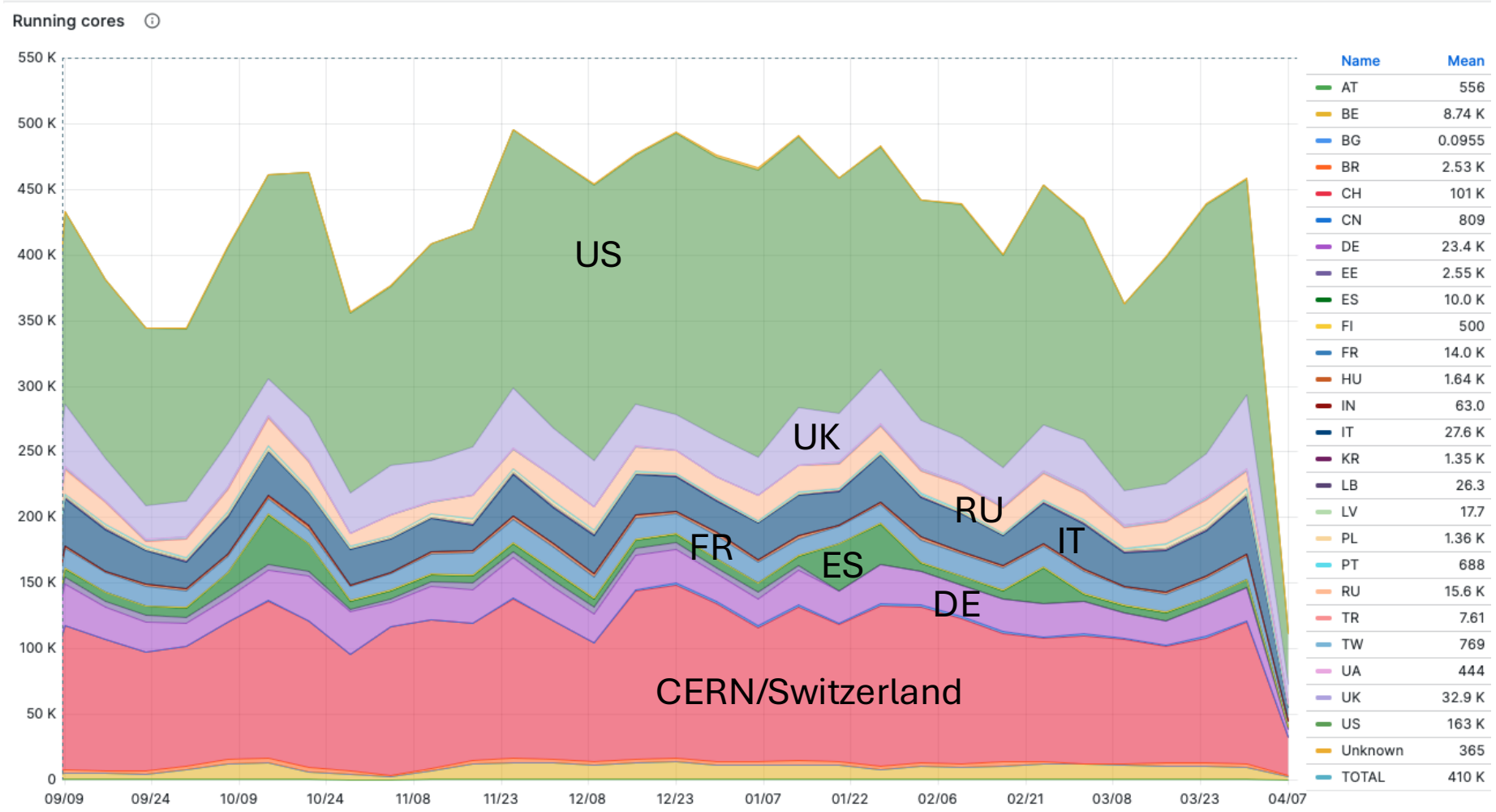
CMS

Katy Ellis, GridPP53, 10th April 2025

GridPP
UK Computing for Particle Physics

Global CMS ops

Compute by country

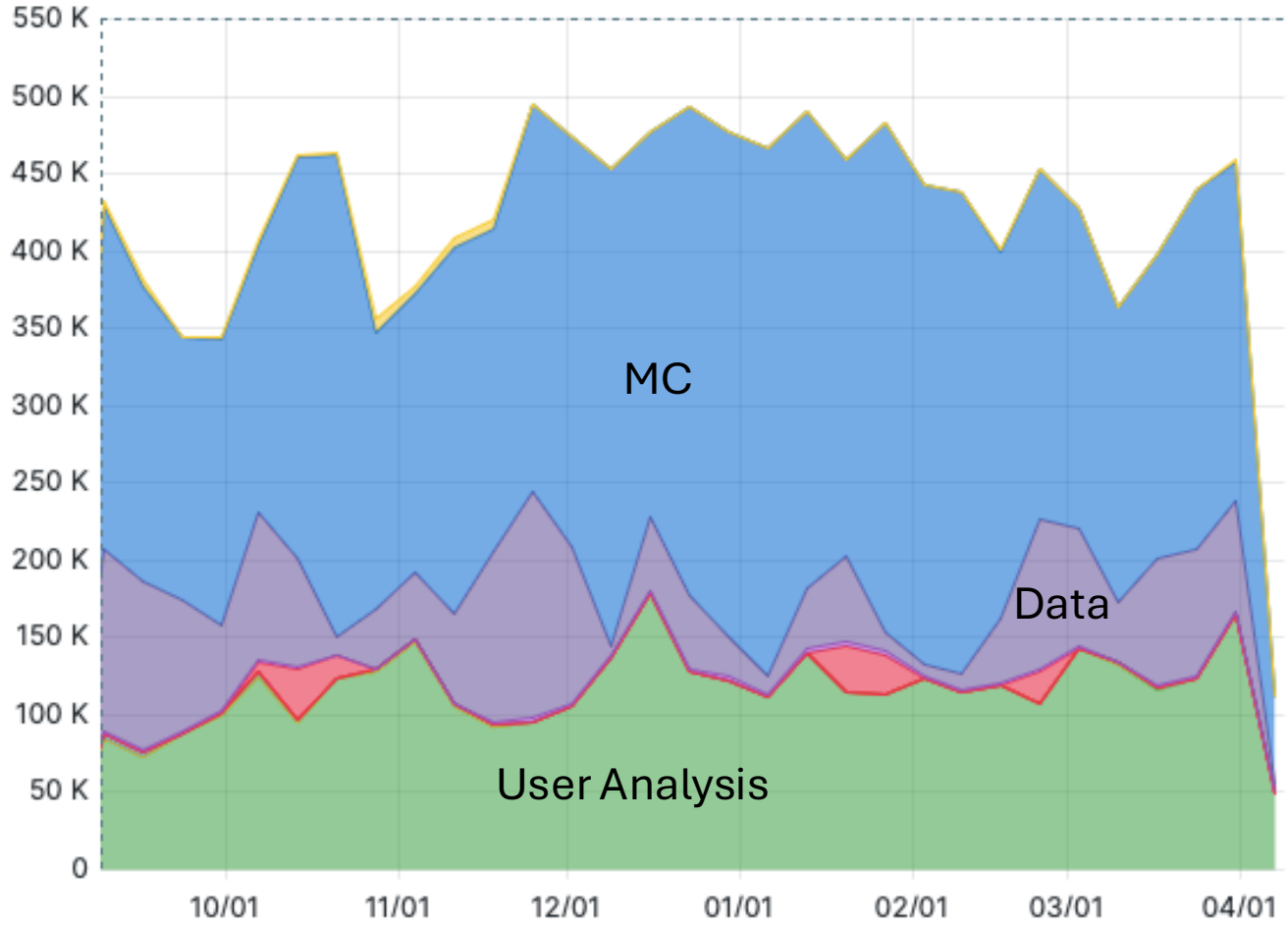


Job mix (all sites)

Total completed jobs

	total	percentage
Analysis	158838655	49.6%
Production	102613404	32.0%
Processing	23067913	7.2%
Merge	17229990	5.4%
Cleanup	9143135	2.9%

Running cores



Name	Mean
Analysis	113 K
Cleanup	68.6
Express	702
Harvesting	4.80
LogCollect	34.4
MN5_test_GEN	3.97 K
Merge	1.51 K
Processing	56.4 K
Production	233 K
Repack	1.37 K
TOTAL	410 K

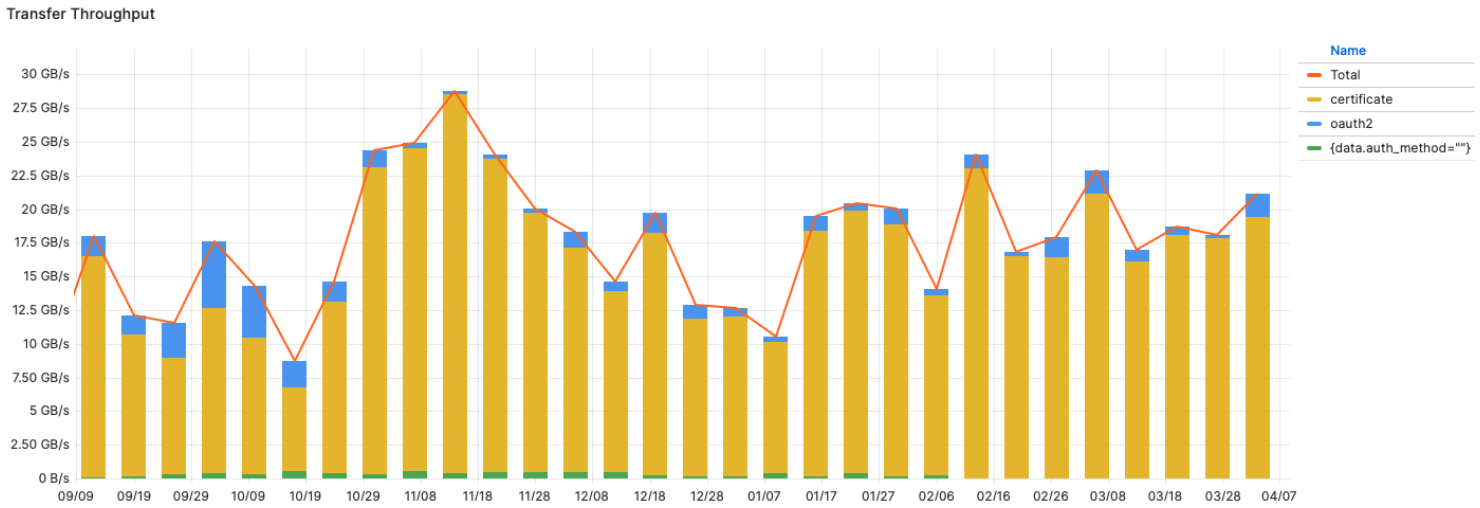
Most Analysis jobs are single core; most central production jobs are multicore

Comments on (low) job efficiency

- Some jobs are intrinsically low efficiency
  - High I/O, low CPU usage
- Sometimes mistakes are made over the number of cores
- Sometimes data is read (via AAA) from further away than necessary
- Recently, a more subtle issue found with input data:
  - Investigated by Core Software (CMSSW) experts
  - (Quote David Lange)
    It depends on event processing rate, basket sizes of the data being read and number of cores (and XRootD server behavior)

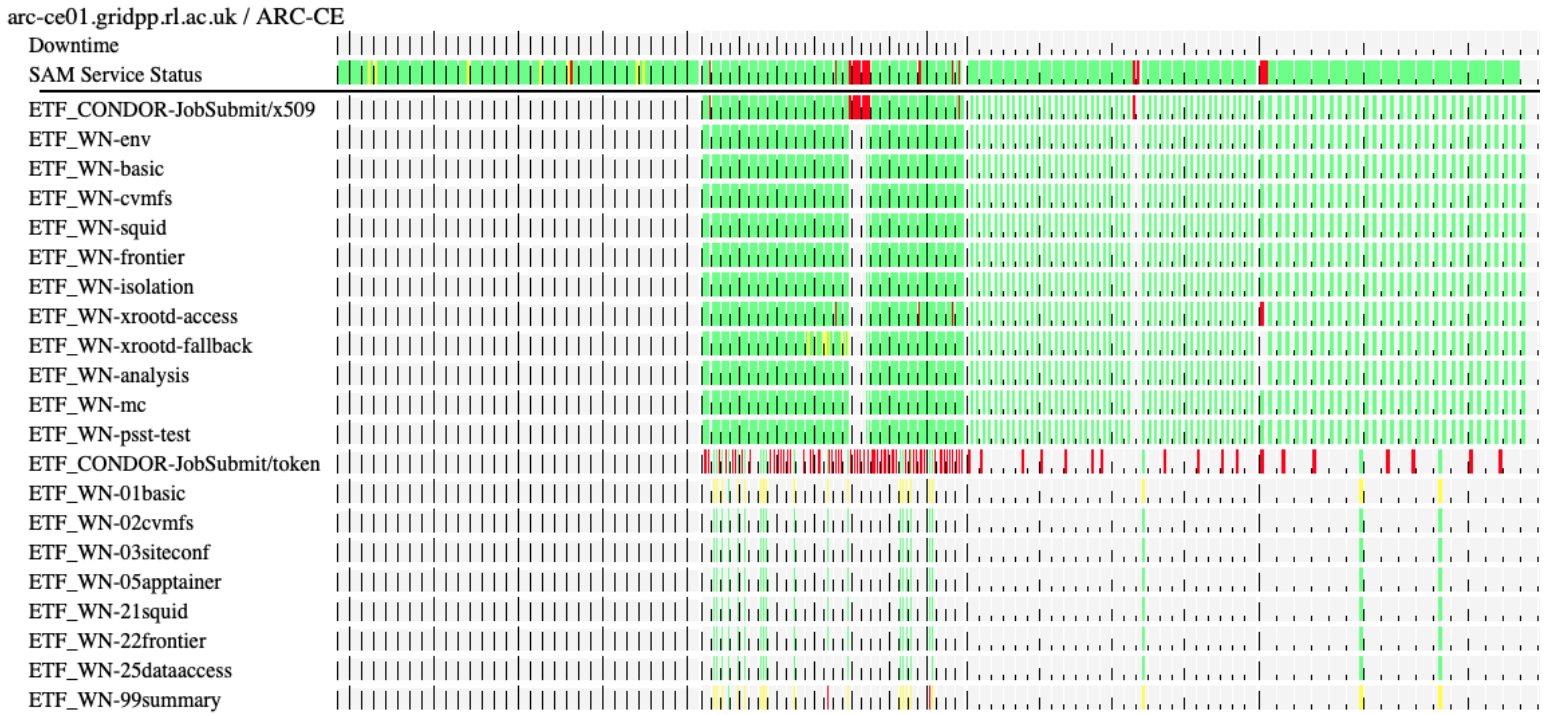
Token status

- Transfers:
  - Many sites are enabled, so many transfers should be token authenticated?
    - In theory yes, but in practice no!
    - Many transfers this year were to/from tape (not supported)
    - Anything multihop (not supported)
    - In FTS, ALL possible sources must be token-enabled, as well as the destination
    - Some issues with CERN-EOS



Token status

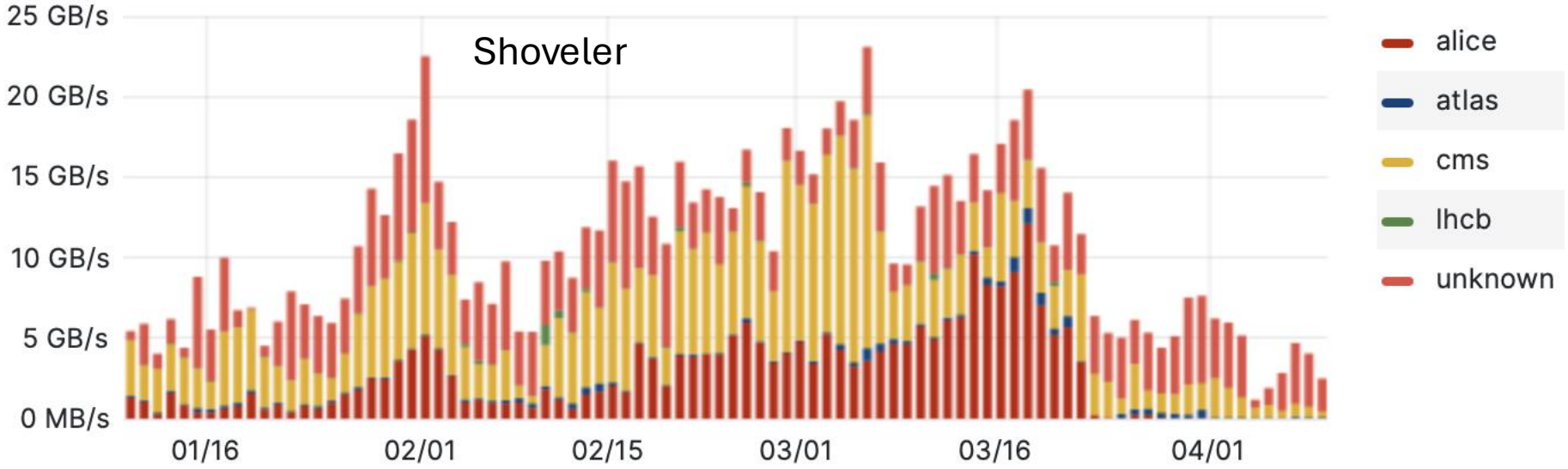
- Compute:
  - Not much to update (some internal work is done; user-facing work is not)
  - Many sites using HT-Condor CE passing SAM tests
  - Continuing issues with ARC-CE SAM tests – release of ARC-7 is good news



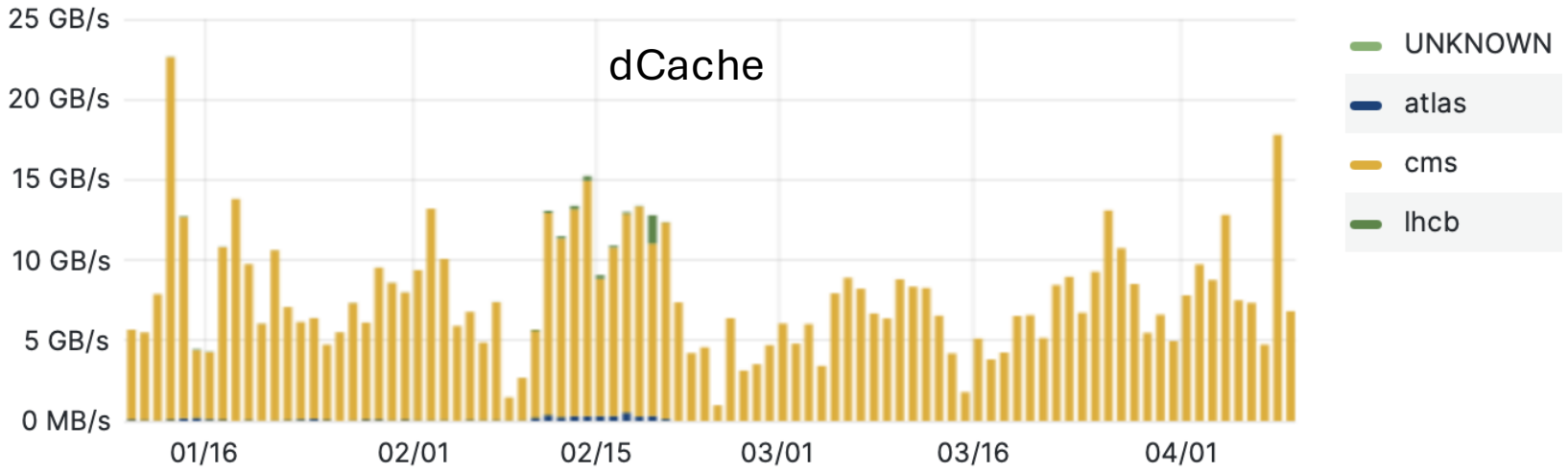
Monitoring update (I)

- Shoveler XRootD monitoring
  - Little progress; not rolled out at more sites yet
  - The CERN-based devops has recently been granted much more time to work on this
  - CMS is also putting in more effort
  - To do examples:
    - Failure monitoring
    - VO labelling for multi-VO servers (could come with tokens?)
    - Improve collector performance
    - Combine with the dCache monitoring (also needs verifying)
  - Good news: Shoveler verified vs ALICE MonALISA monitoring
  - Katy made use of Shoveler monitoring during mini-DC

Shoveler vs dCache (last 90 days)



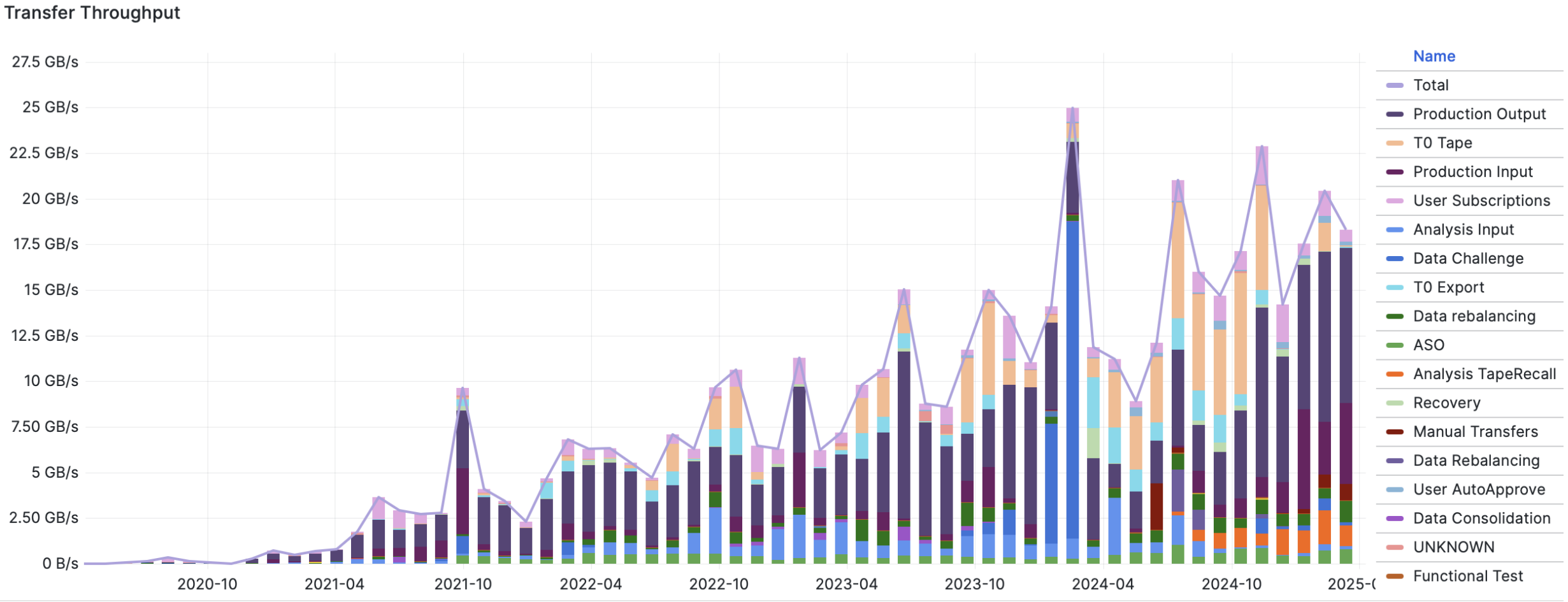
Remote reads only
Many sites not yet monitored



Monitoring update (II)

- Fireflies / packet marking
  - Give information about transfers for future network monitoring
    - VO
    - Activity
    - Source/dest
  - CMS ticketed sites to request roll-out of fireflies
    - At T1 Jyothish worked really hard going back and forth with developers
    - Discovered a bug and new configuration for CMS
    - Applies also to other VOs

Transfer watch (FTS, CMS, last 5 years)



Binning is 1 month

UK CMS

UK T2 site status - Brunel

- Breaking news!
- Data centre had an issue with the air con which caused a fire in one of the cabs.
- Due to this it was recommended by the fire service that we turn off all the devices in the data centre in order to bring down the temperature.
- Downtime declared for the Brunel site.

UK T2 site status - Bristol

- Person power increased from 0.2 to ~0.5 FTE - still familiarising new hire with local infrastructure
- New storage system in place (CephFS), working well except for meltdown end of March/beginning of April 2025
  - Many thanks for the GridPP support to resolve!
- Updated batch system, added new monitoring setup (ganglia -> prometheus + thanos + grafana)
- Remaining CentOS7 nodes migrated to RHEL 7 for now -> migration to Rocky 9 follows
- Site back online with two missing bits:
  - Accounting to be updated (storage and compute)
  - HTCondor job routes to be updated: fair share, additional settings (e.g. for CMS containers to work with --contain)
- Reminder: The Bristol cluster is shared ~30% GridPP and 70% from local funds - eliminating storage is a no-go. Using XRootD+CephFS is simple enough *when it works*.

UK site performance

	CPU eff	Failures
T1_US_FNAL	70.8%	18.27%
T1_UK_RAL	73.3%	9.97%
T1_RU_JINR	73.4%	12.16%
T1_IT_CNAF	72.3%	13.32%
T1_FR_CCIN2P3	75.0%	14.13%
T1_ES_PIC	70.8%	32.20%
T1_DE_KIT	73.4%	12.74%

UK site performance

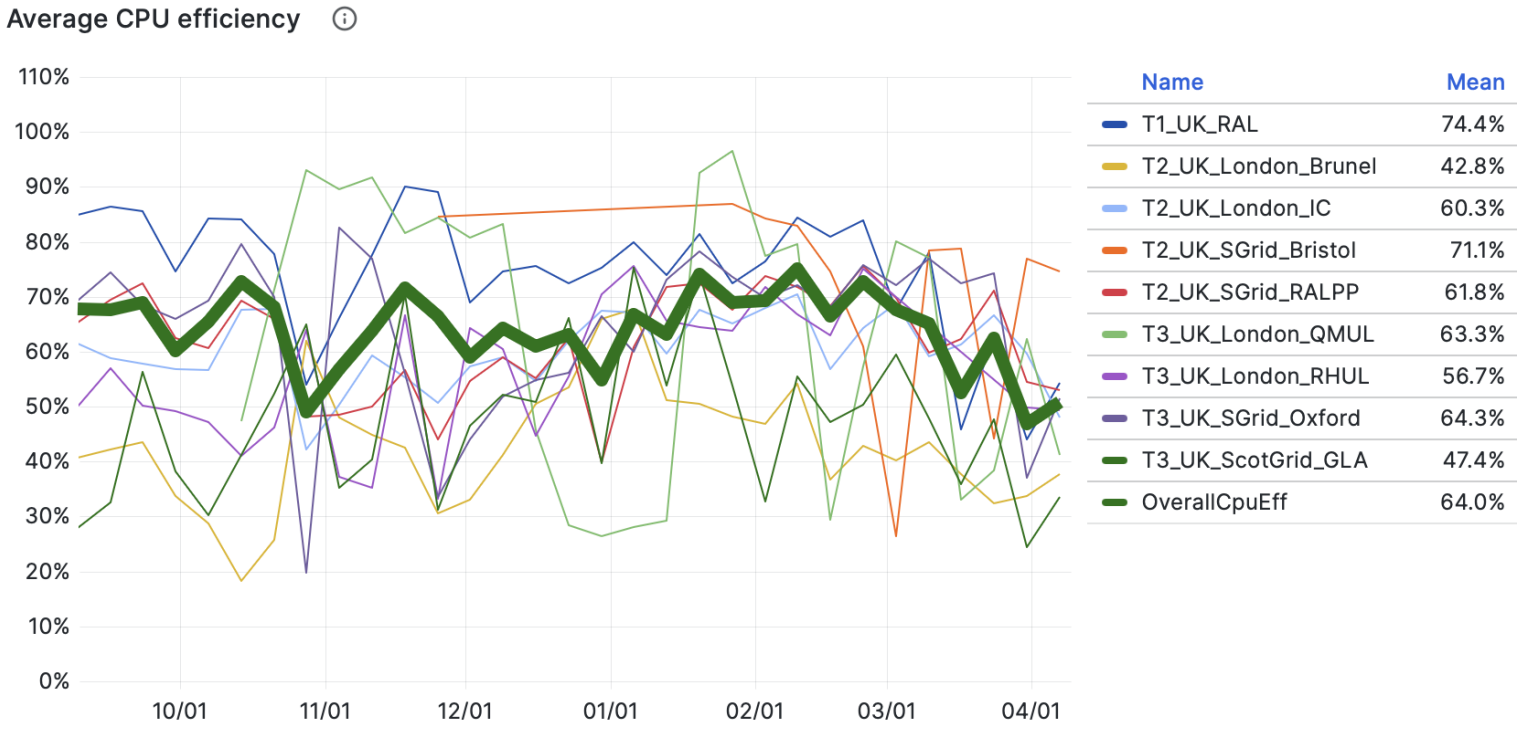
	CPU eff	Failures
T1_US_FNAL	70.8%	18.27%
T1_UK_RAL	73.3%	9.97%
T1_RU_JINR	73.4%	12.16%
T1_IT_CNAF	72.3%	13.32%
T1_FR_CCIN2P3	75.0%	14.13%
T1_ES_PIC	70.8%	32.20%
T1_DE_KIT	73.4%	12.74%

	CPU eff	Failures
T2_UK_SGrid_RALPP	61.7%	37.26%
T2_UK_SGrid_Bristol	40.5%	17.98%
T2_UK_London_IC	60.6%	28.07%
T2_UK_London_Brunel	43.4%	32.81%

Reminder: Tier 2 sites typically run different job mix to Tier 1s – more User Analysis

UK site performance

	CPU eff	Failures
T1_US_FNAL	70.8%	18.27%
T1_UK_RAL	73.3%	9.97%
T1_RU_JINR	73.4%	12.16%
T1_IT_CNAF	72.3%	13.32%
T1_FR_CCIN2P3	75.0%	14.13%
T1_ES_PIC	70.8%	32.20%
T1_DE_KIT	73.4%	12.74%



CMS O&C updates

Review of Workload Management (incl. CRAB)

- Last formal review took place in 2018-2019
- A new review is in progress
  - What is needed for CMS in Run-4?
  - Existing CMS software? New software? Other VO software?
- Will make a recommendation to CMS
- Planning to report during spring
- CMS O&C will then make a decision
- Workload Management development team also have some well-developed opinions

Other areas of study

- Can we reduce e.g. the amount of AODSIM we keep?
- Can we share build system with other VOs?
- How to improve MinBias/pile-up mixing? ML solution?
- How to improve application memory usage (CMSSW)?
- 16-core pilots on the Grid?

- CDR (Conceptual Design Report) is with the special readers

DOMA status

Data Organisation, Management and Access

DOMA mini-challenges for data movement

- “Capacity” challenges
  - E.g. can your site/country hit the DC24 target rates?
  - Can it saturate the network bandwidth?
  - What are the bottlenecks, when does it break?
- “Capability” challenges
  - E.g. can we employ new technology to improve file transfer?
    - Is it good enough to roll out at many sites?
    - Examples: Jumbo frames, SDNs
  - E.g. what happens when we try to change what we’re doing?
    - Can the existing system cope?
    - Examples: Large file transfers, token auth

Summary

- A busy period for data-taking
- CMS continue to study scalability for the future
- But not to forget operational issues of the present…
- Tricky times for some UK sites
- Mini-challenges for data transfers are go!