



Science and
Technology
Facilities Council

ATLAS: Ready for Run-3

J. Walder

GridPP 47

23/24 March 2022

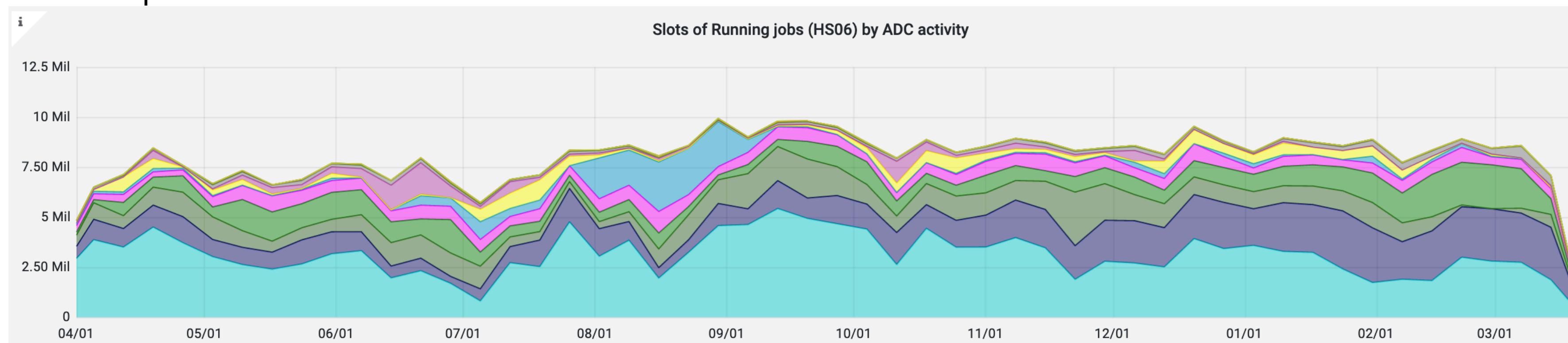
Overview

- ATLAS
 - Brief overview from last year
- ATLAS UK
 - Status in 21-22
- Updates at RAL
 - WebDav and (storage) Tokens
- Highlights and preparations for Run-3;
 - HS06 on the Grid
 - Lifecycle of a typical r22 MC reco job
 - Moving to Run-4
- Not mentioning many interesting things:
 - e.g. Activities in Google R&D
 - New storage opportunities using tape at NESE and projects like Seal

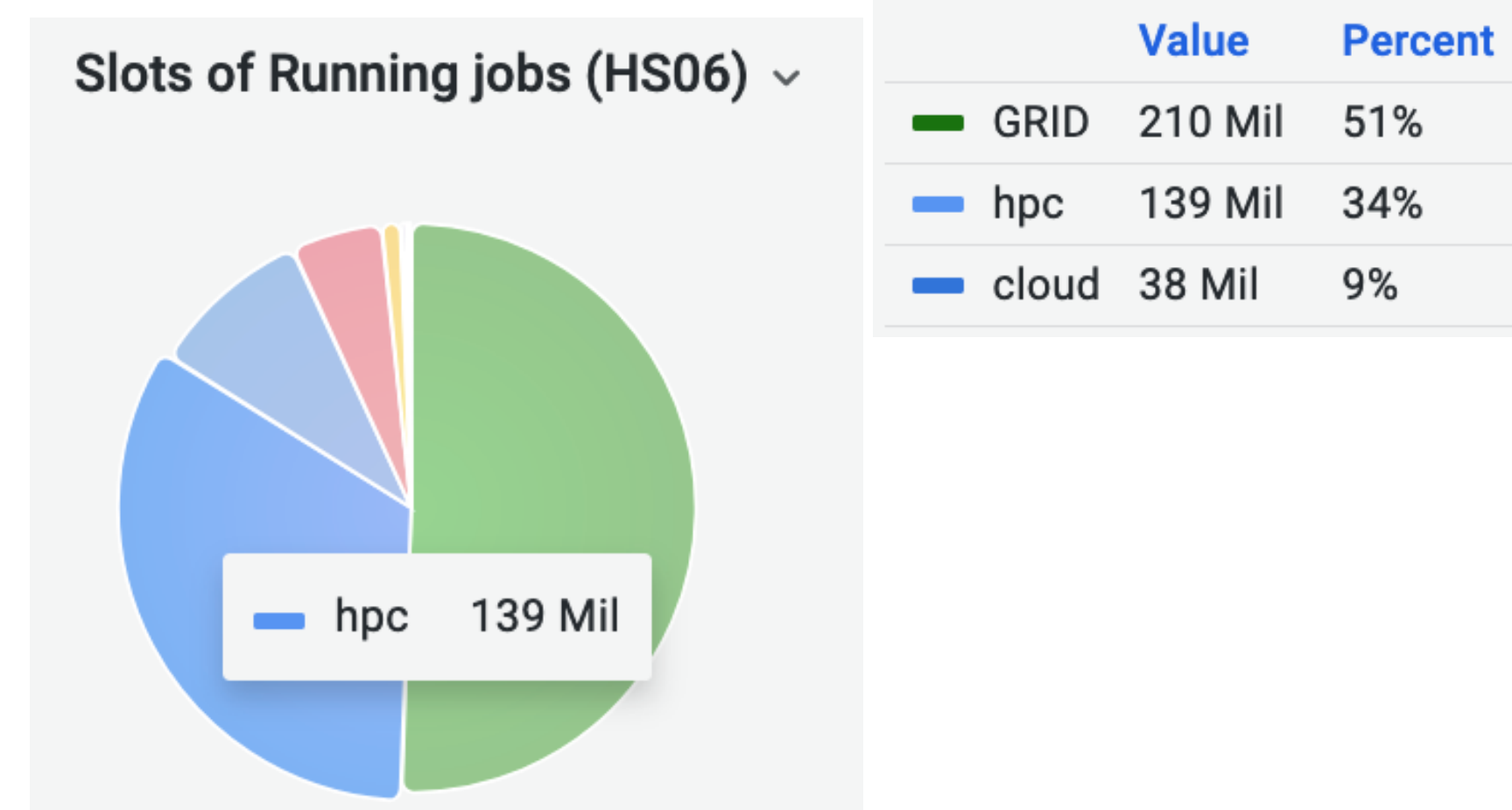
ATLAS Overview '21-22: Compute

- 8M HS06 average compute resources deployed to ATLAS (~660k slots)
 - Simulation (and its reconstruction) dominating most of the production
 - Run-2 full reprocessing performed (r22) and various Derivation production campaigns
 - New MC Resimulation task targeting improvements in events with decays of long-lived particles

	min	max	avg
MC Simulation Full	170 K	5.43 Mil	3.08 Mil
MC Reconstruction	248 K	2.91 Mil	1.46 Mil
MC Event Generation	1.58 K	2.67 Mil	946 K
Group Production	7.65 K	2.19 Mil	895 K
User Analysis	34.0 K	1.07 Mil	555 K
MC Simulation Fast	669	2.48 Mil	335 K
Data Processing	0	1000 K	267 K
MC Resimulation	0	1.57 Mil	200 K
Group Analysis	21.4 K	574 K	156 K
MC Merge	260	56.2 K	18.2 K
t0_processing	0	82.8 K	18.0 K
Testing	0	9.78 K	3.85 K
Event Index	30.3	6.82 K	1.27 K
t0_caf	3.35	4.68 K	749



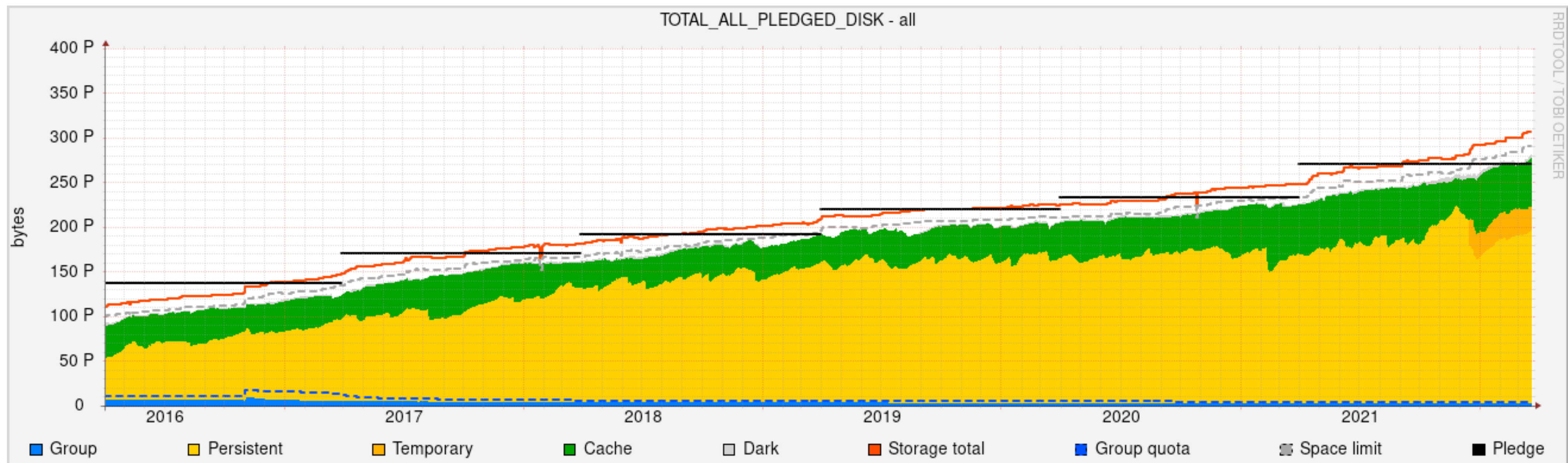
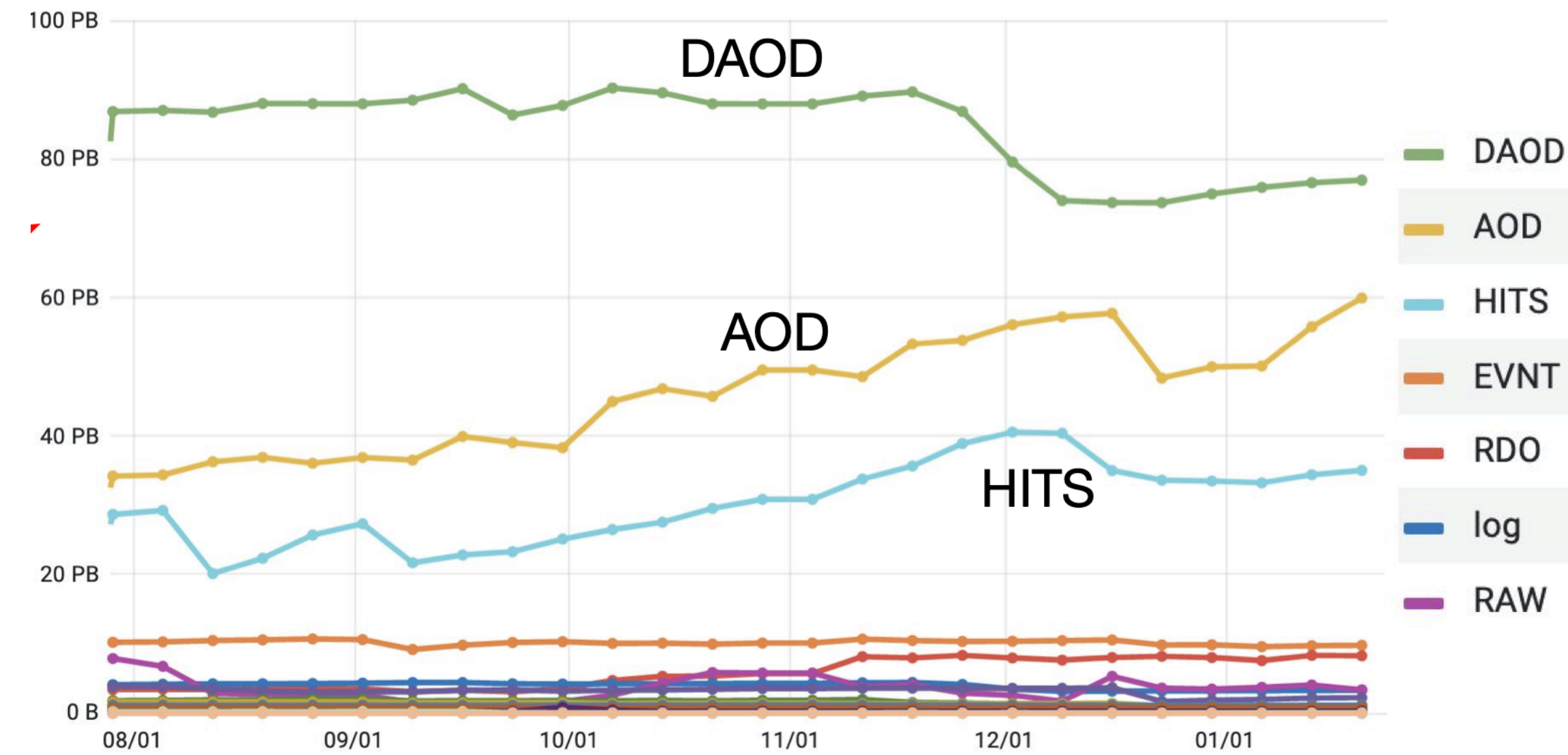
- ~ 55% of Grid resource delivered from T2 sites (30% from T1)
- HPC (opportunistic resources)
 - ~ 34% of total compute in last year
 - Although ~ 9% of submitted jobs
 - Mainly Simulation, but more recently demonstrating all production (and analysis) workflows
 - Anticipated that could reduce for '22



ATLAS: Storage

- Storage, as always, remains critical
 - Opportunistic (HPC) compute resources don't provide storage.
 - ATLAS requested if 2022 pledged TAPE resources could be brought forward
- TAPE storage (data carousel) continues as integral part of data operation models
 - Some workflows already treating (intermediate) outputs as transient (AF3).
- Regular lifetime and obsolescence campaigns run to recover space.

(Primary) Data on Disk



ATLAS UK: Comparison

- UK provided 15% of ATLAS GRID resources since April 21 (According to Monit)

- Reduces to 9% if all HPC and Cloud resources also included

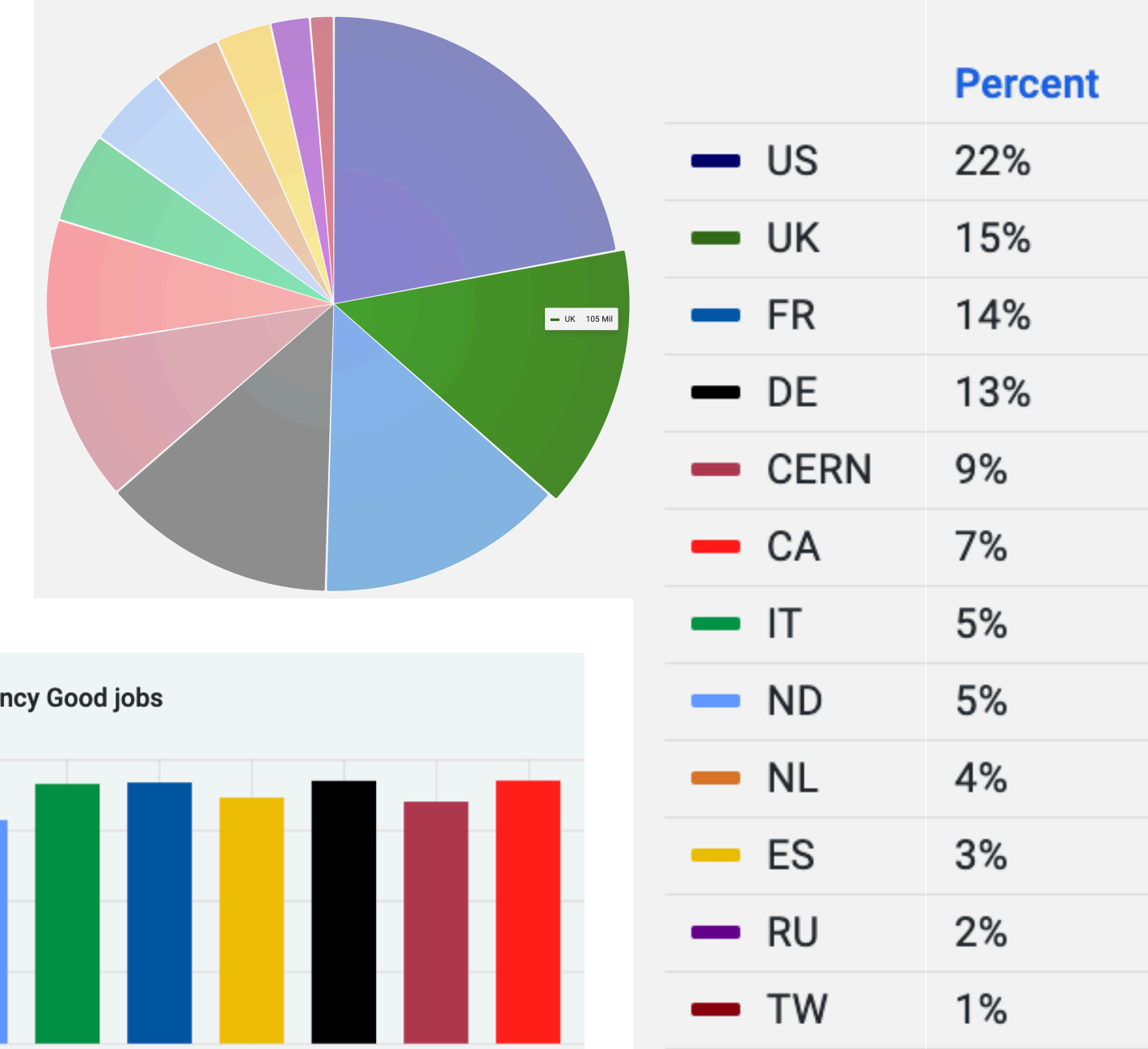
- General performance metrics comparable to global averages:

- Similar CPU efficiencies

- Slight differences in failure rates and failure types

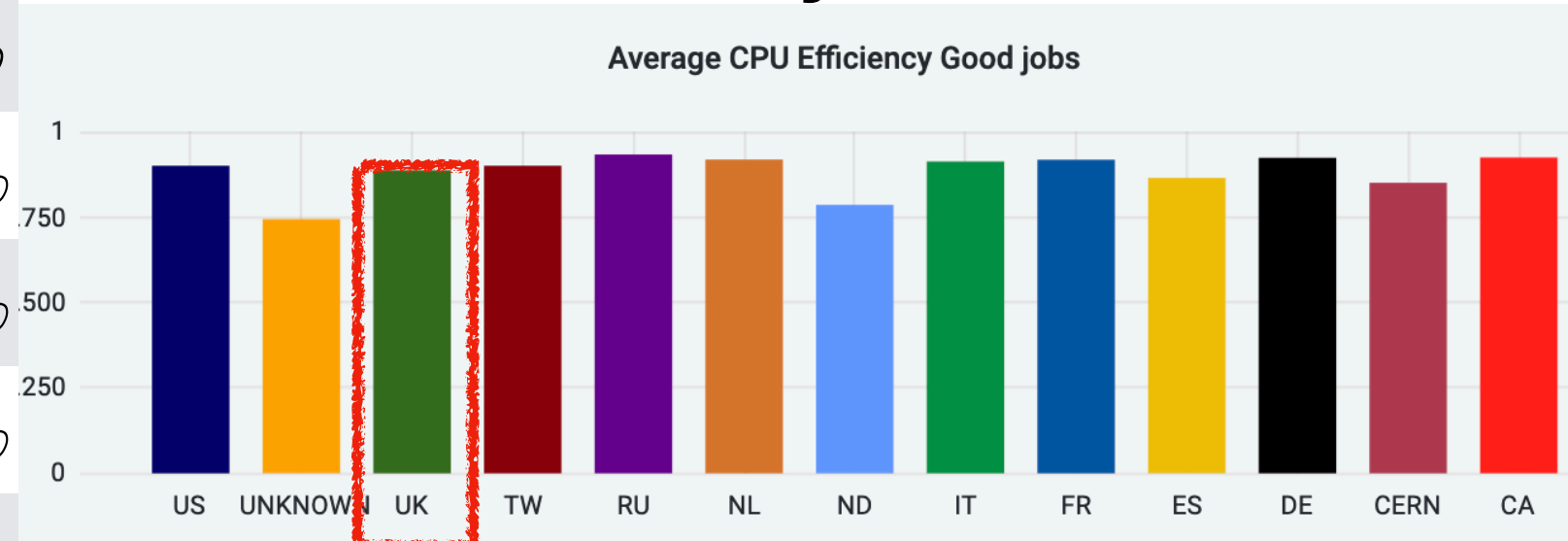
- Can be sensitive to the job-mix in the averaged values

GRID resources by cloud



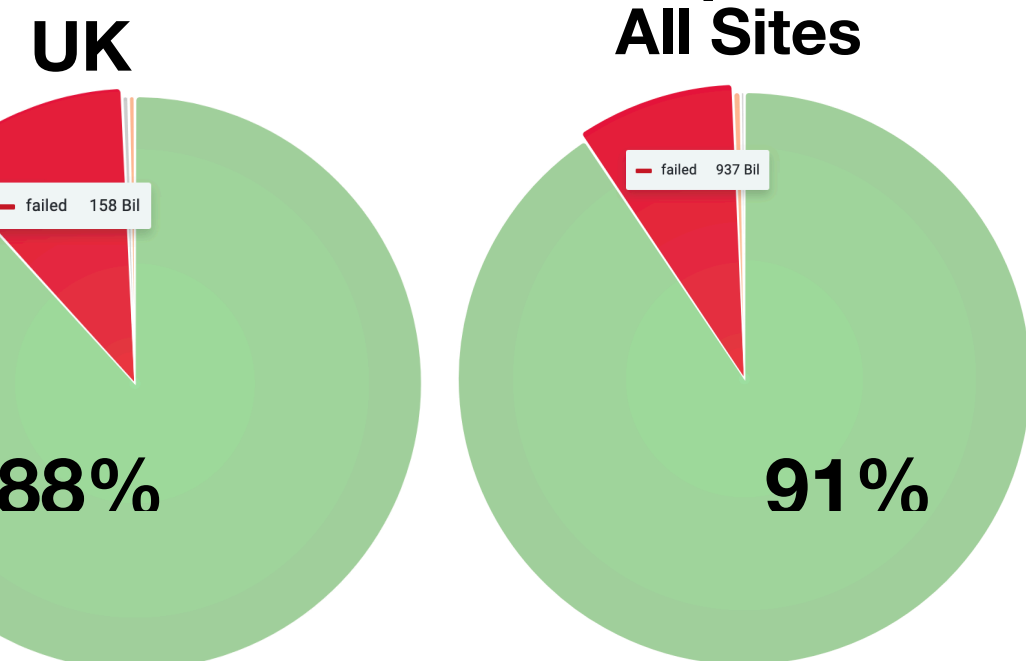
Job Mix (Leading)	UK [%]	ALL [%]
MC 16 evgen	26%	19%
MC 16	18%	17%
MC 16 simul	15%	15%
User Analysis	10%	13%
MC Derivations	8%	9%
MC Other	6%	7%
Data Derivations	5%	5%
Reprocessing default	4%	6%

CPU Efficiency

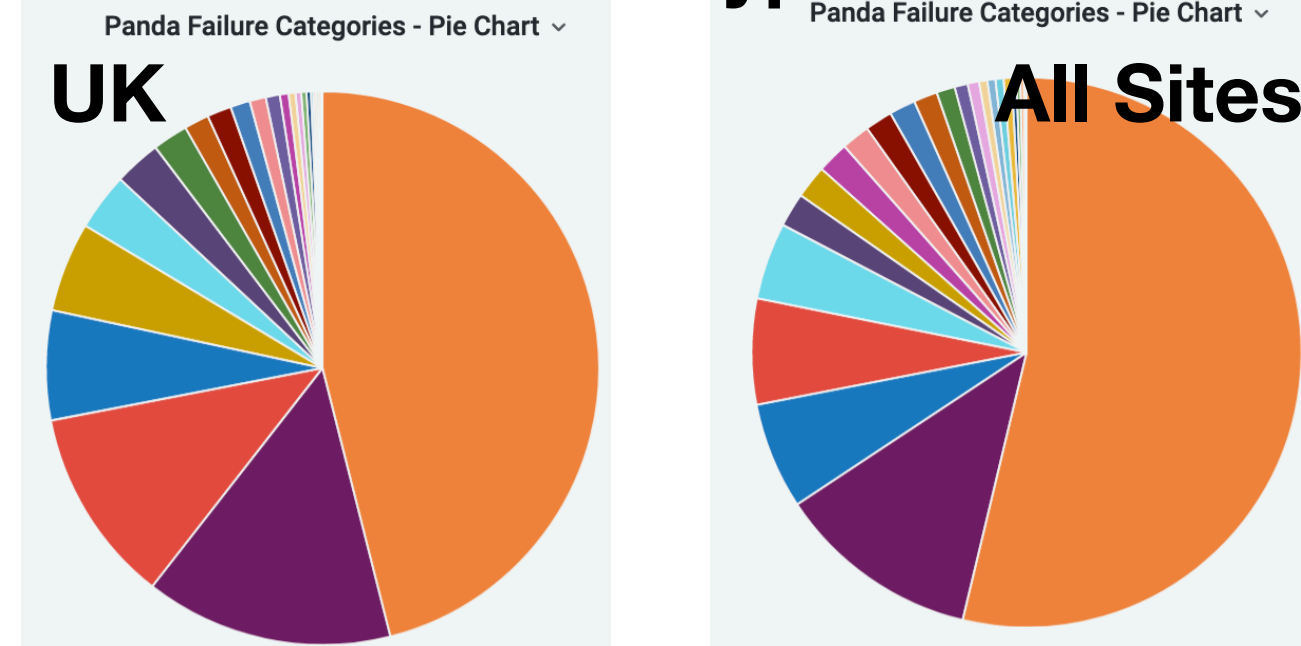


(Non grid excluded)

Wallclock consumption



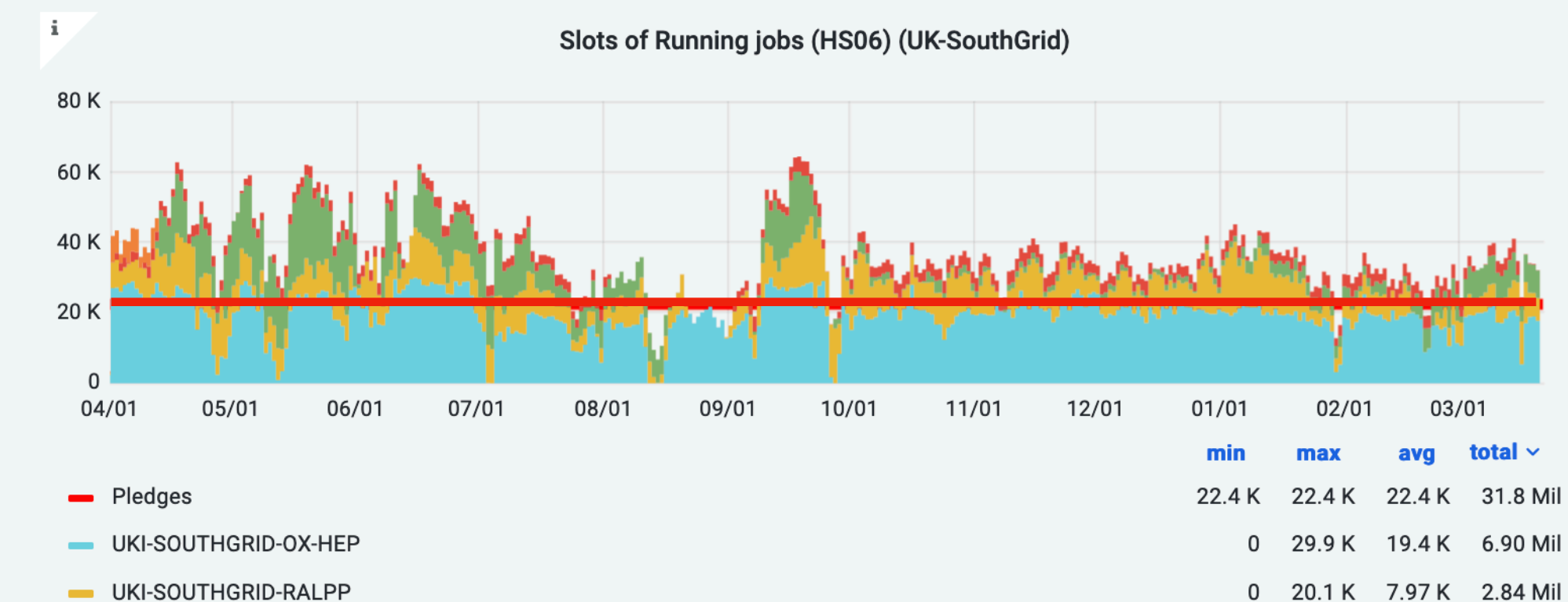
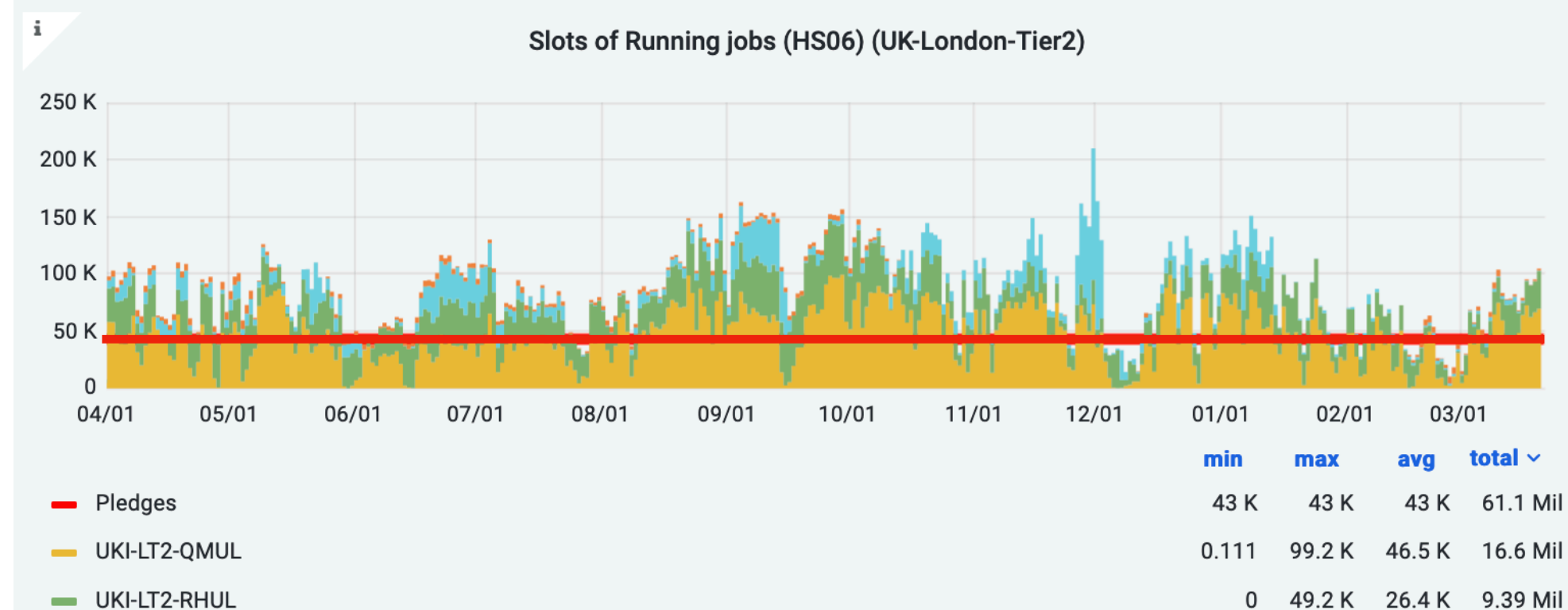
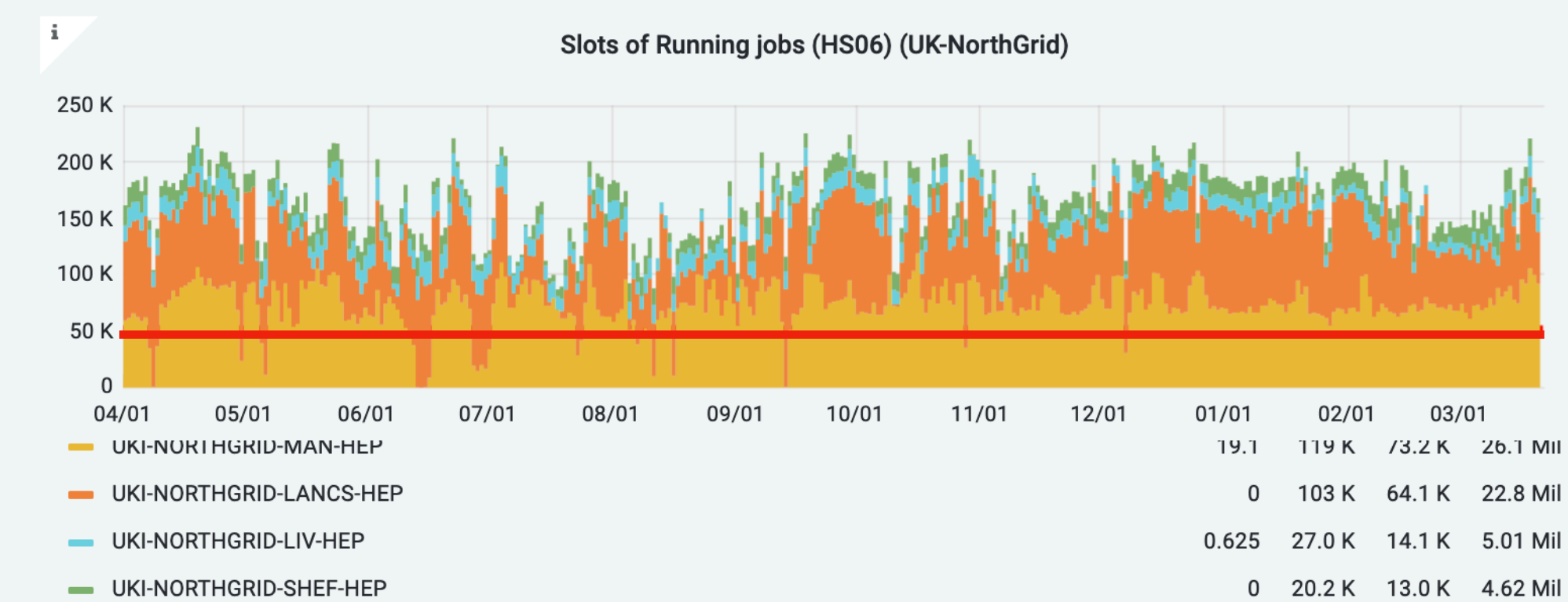
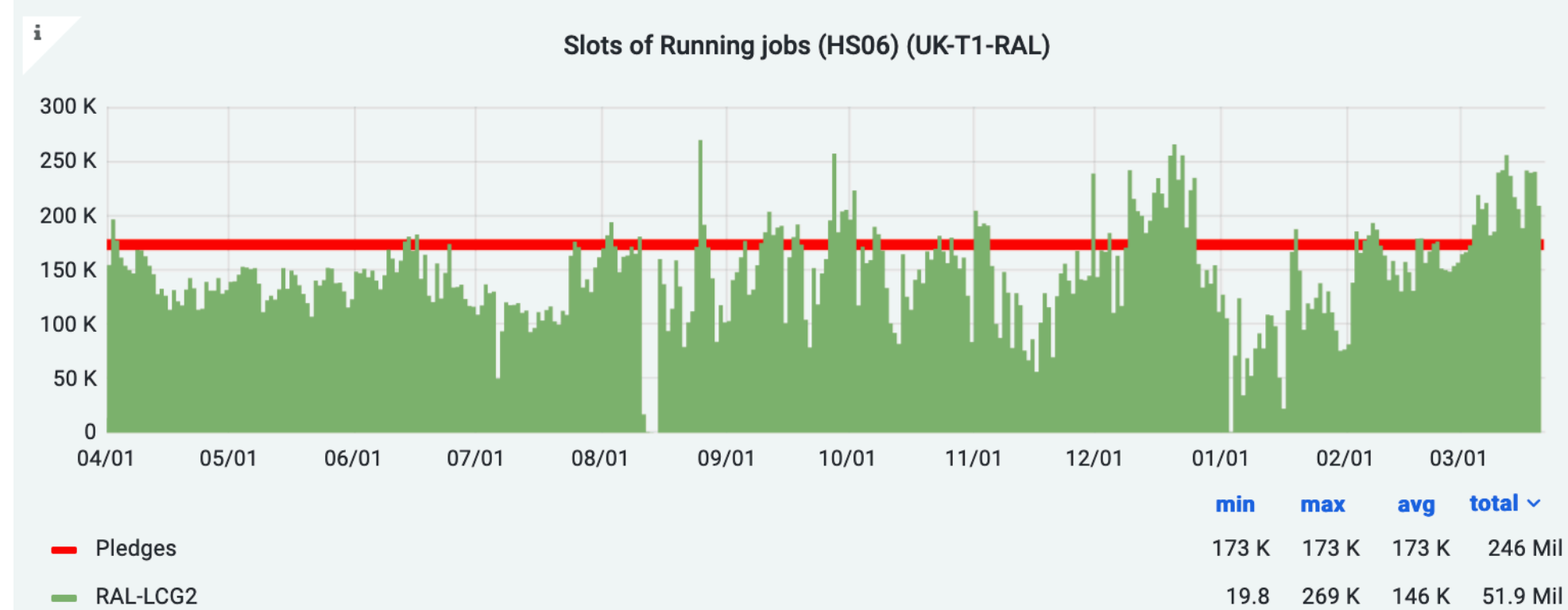
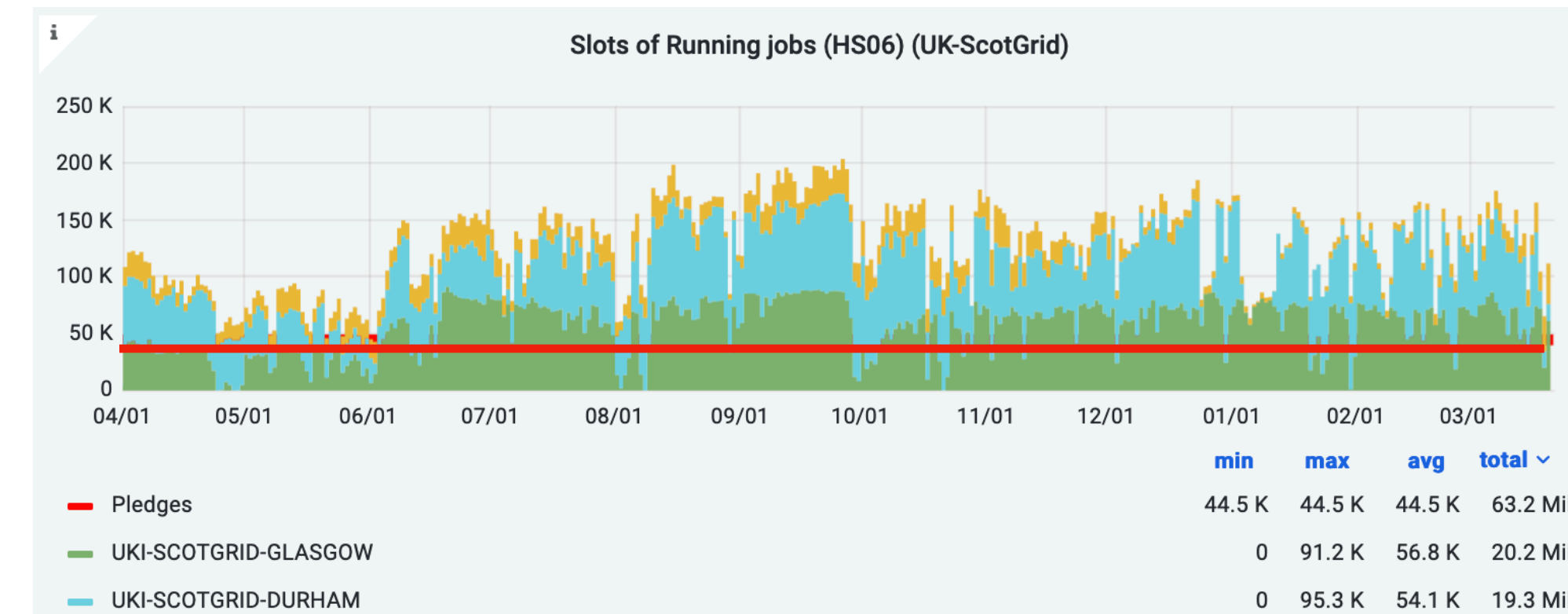
Failure types



Failure Type	total	percentage
TaskBuffer Error: Timeout	5 Mil	46%
Pilot Error	2 Mil	14%
TaskBuffer Error 300	1 Mil	11%
TaskBuffer Error	681 K	6%
Pilot/PanDA Error: Get error	559 K	5%
Execution Error 65	356 K	3%
Job Dispatcher Error: lost heartbeat	291 K	3%
Pilot/PanDA Error: Put error	220 K	2%
Pilot/DDM Error: put error	155 K	1%
Pilot/DDM Error: get error	150 K	1%
Pilot/PanDA Error	122 K	1%
DDM Error	101 K	1%
Pilot/PanDA Error: killed by panda server	87 K	1%

ATLAS UK: Compute

- UK T2 federations continue to deliver above pledge
- RAL finishing year well but struggled ATLAS pledge:
 - single-/multi-core partitioning in a multi-VO environment complex
 - maintaining high job-throughput with necessary high-throughput data transfers (to get output data from Echo to final destinations) (issues can quickly lead to large FTS Backlogs)
 - Use of multi-job pilots tested and may be improving the situation

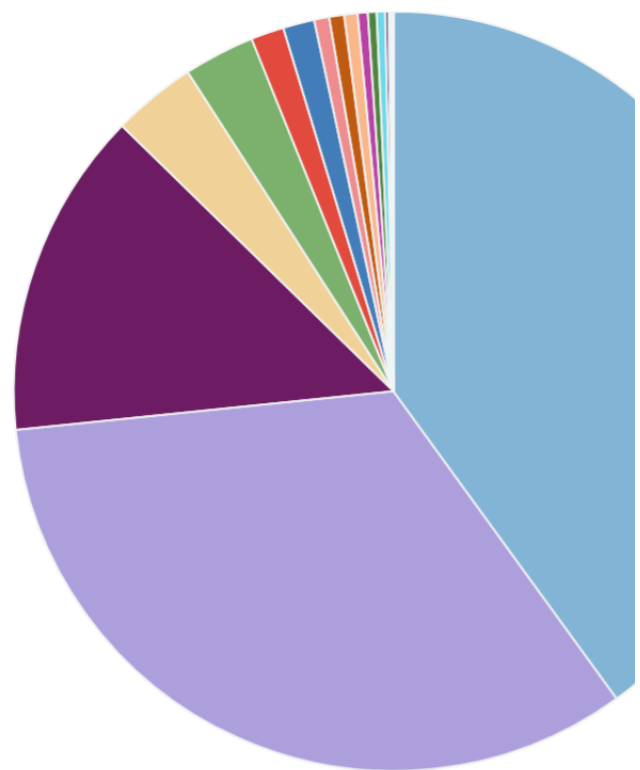


ATLAS UK: Storage

- Over 35PB now stored on TAPE at RAL (Antares)
- 30PB of used disk in UK (excluding Localgroupdisk).
- Recent increase of QMUL from ~ 3.5 – 7.0 PB starting to be used
 - *Impressed ADC coordinators*
- Derivation formats (then AOD) dominate used disk by volume
 - EVNT and log dominate by count

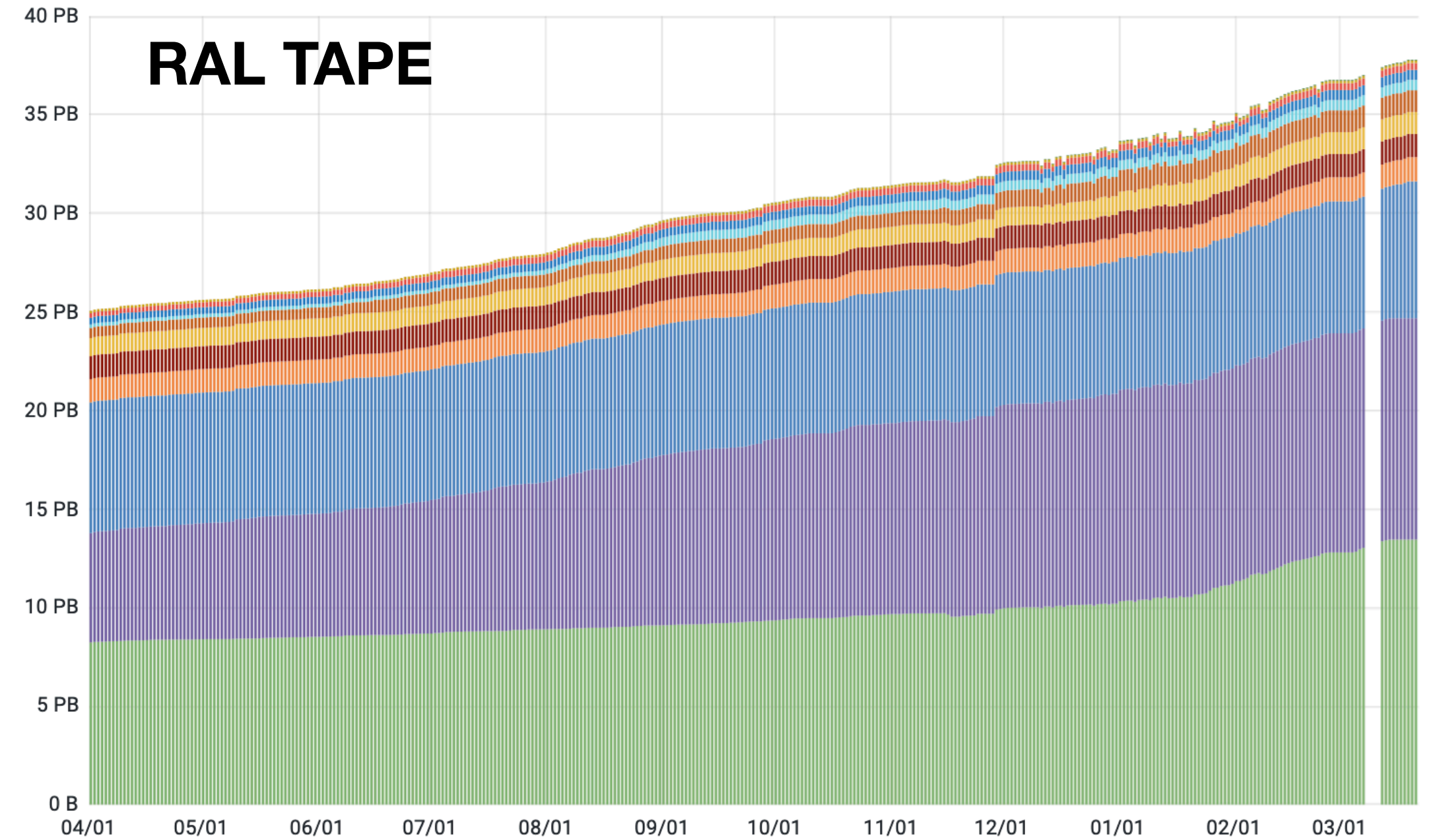
Disk: Type by Volume

	current	percentage
DAOD	12 PB	40%
AOD	10 PB	33%
HITS	4 PB	14%
EVNT	1 PB	4%
user	896 TB	3%
log	417 TB	1%
RDO	395 TB	1%



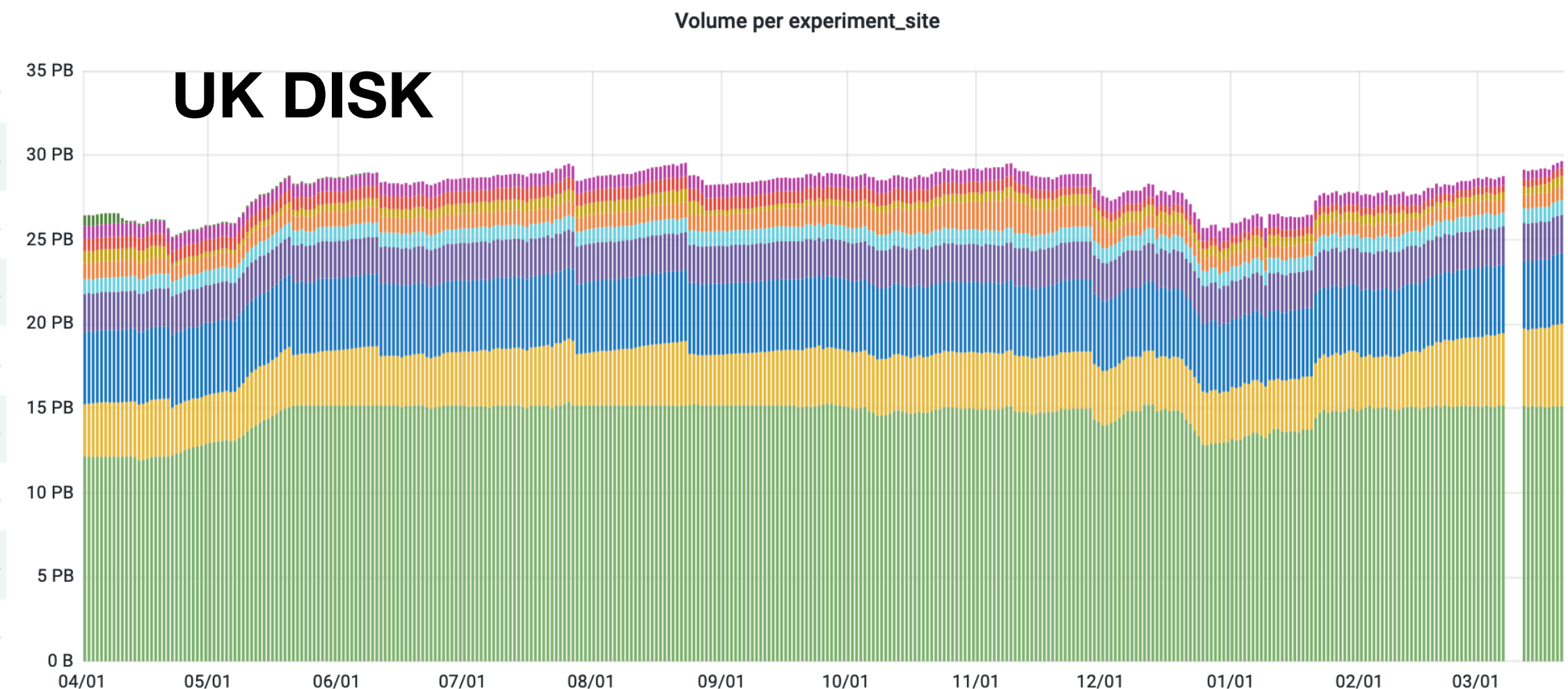
current

AOD	13.4 PB
HITS	11.2 PB
RAW	6.94 PB
DRAW	1.23 PB
other	1.18 PB
DAOD	1.11 PB
RDO	1.10 PB
DESD	526 TB



current

RAL-LCG2	15.1 PB
UKI-LT2-QMUL	4.90 PB
UKI-NORTHGRID-MAN-HEF	4.21 PB
UKI-SCOTGRID-GLASGOW	2.24 PB
UKI-LT2-RHUL	891 TB
UKI-NORTHGRID-LANCS-H	829 TB
UKI-SOUTHGRID-RALPP	607 TB
UKI-NORTHGRID-LIV-HEP	462 TB
UKI-SCOTGRID-ECDF	396 TB



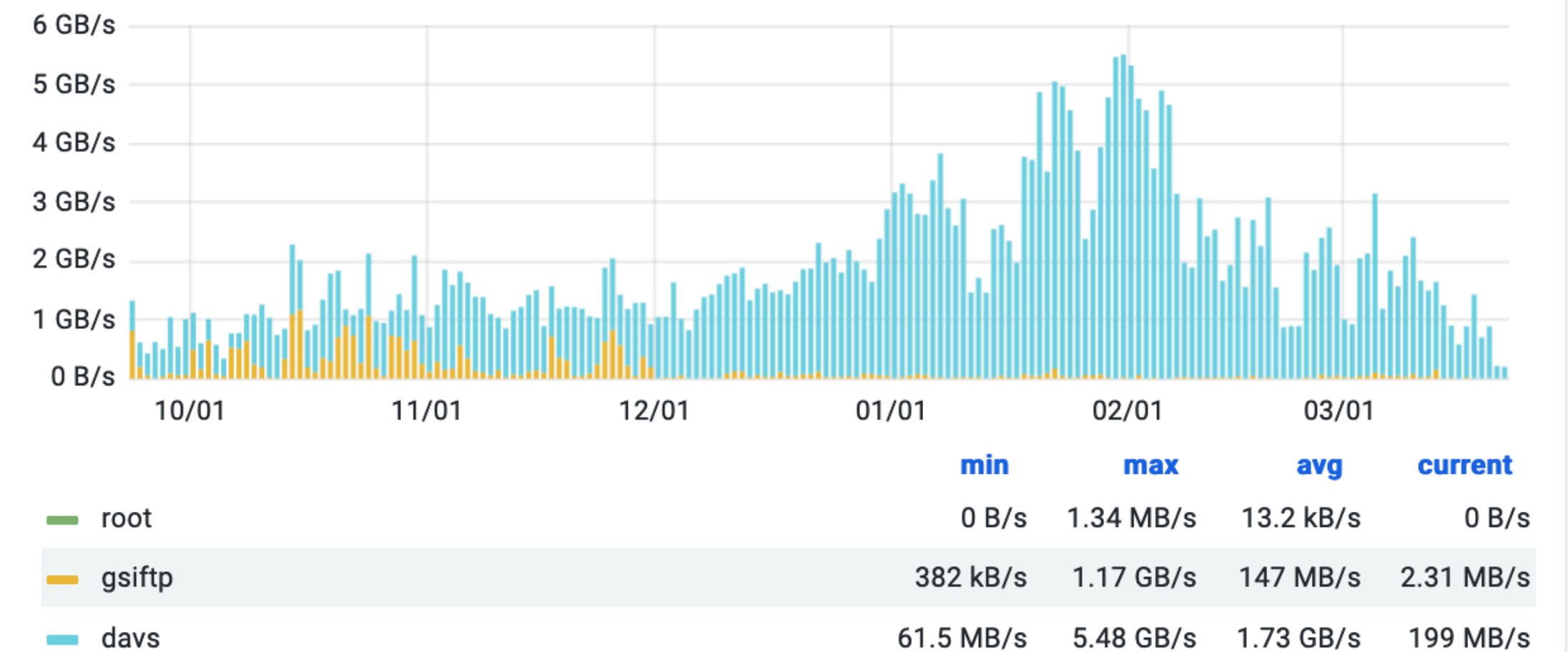
Operational Challenges/Achievements

- Most observed operational problems related to Data and IO: (e.g. Transfer and deletions errors)
 - Does impact compute (i.e. running jobs, staging, etc.)
- Migration to WebDav across almost all sites
 - Exposes new issues / sets of problems for sites to learn:
 - e.g. Lancaster OOM errors in httpd services
 - QMUL: tunings numbers of allowed connections
- Storage:
 - Started campaign* of DPM migration:
 - Glasgow done (For ATLAS)
 - Oxford storage decommissioning
 - Lancaster in progress
 - Brunel (next)
 - Who's next ... ?
 - Data loss events; tending to be correlated near times of DPM commissioning (e.g older hardware ?)
 - So far mainly non-unique or log files are unrecoverable
 - Xcache:
 - *The cause of, and solution to, all your storage problems (?)*
 - External and 'transparent' internal caches;
 - Tendency to fall over when no-one is looking ...
 - Improved monitoring / fault detection?
 - Auto-restarting scripts usual last-resort method
 - Total of 131 GGUS tickets opened under "ATLAS UK";
 - A few long-standing tickets (necessary developments or hard to pin down issues).
 - Otherwise, closed in timely manner.

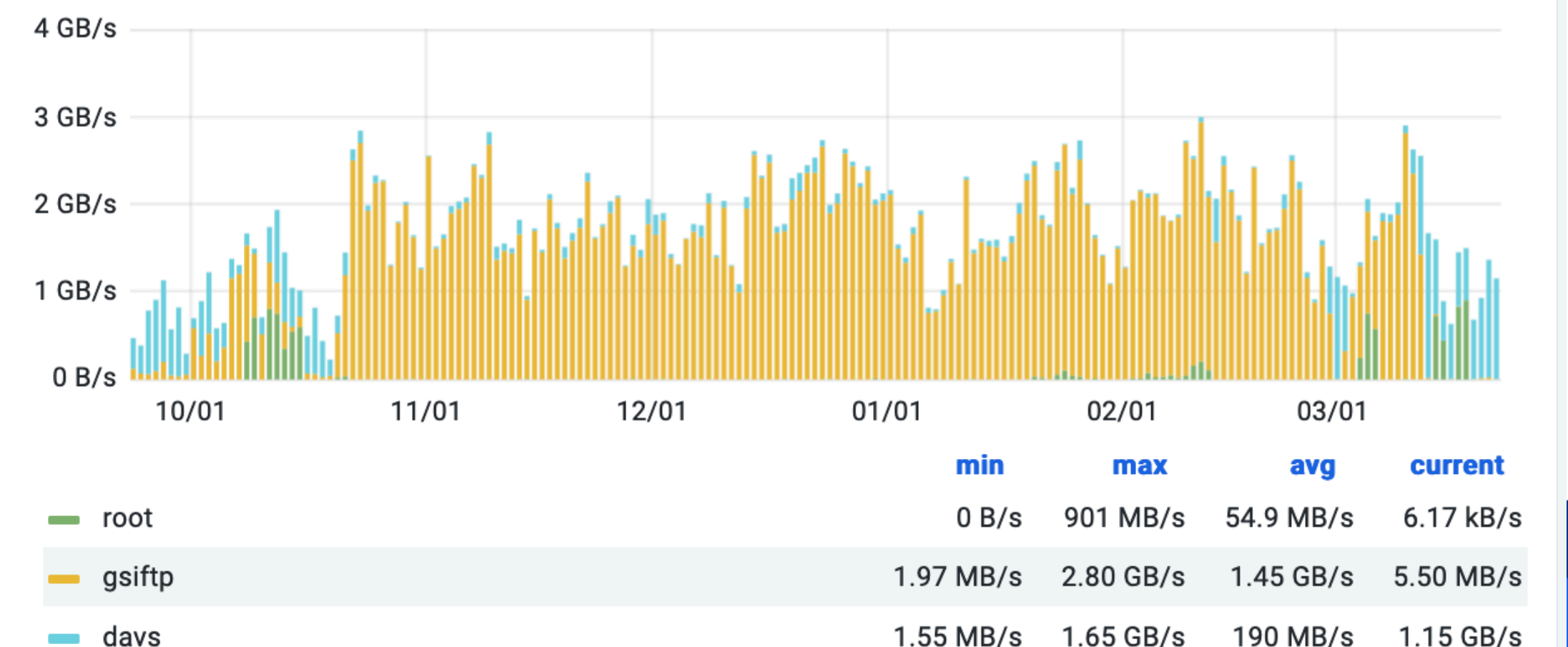
RAL: operating with Davs (1)

- ATLAS will be removing gridFTP as a required protocol shortly.
- US most likely first region to actively remove it from their sites
- RAL; moved to WebDav for reads for some time;
 - Has shown good throughput at points
 - (e.g. periods of large recalls from Tape using Multihop)
- Enabling davs as primary preferred protocol (i.e. prioritising writes via davs) only became feasible after migration of Castor to Antares (related to SRM+gridFTP vs SRM+https)
- Writes to RAL via davs enabled recently (at nominal priority)
- Failure rate is currently higher than with gridFTP, related to some xrootd service instabilities that are under investigation
- Other VO's also using davs (also for the Tape challenge)

Reads from RAL Transfer Throughput



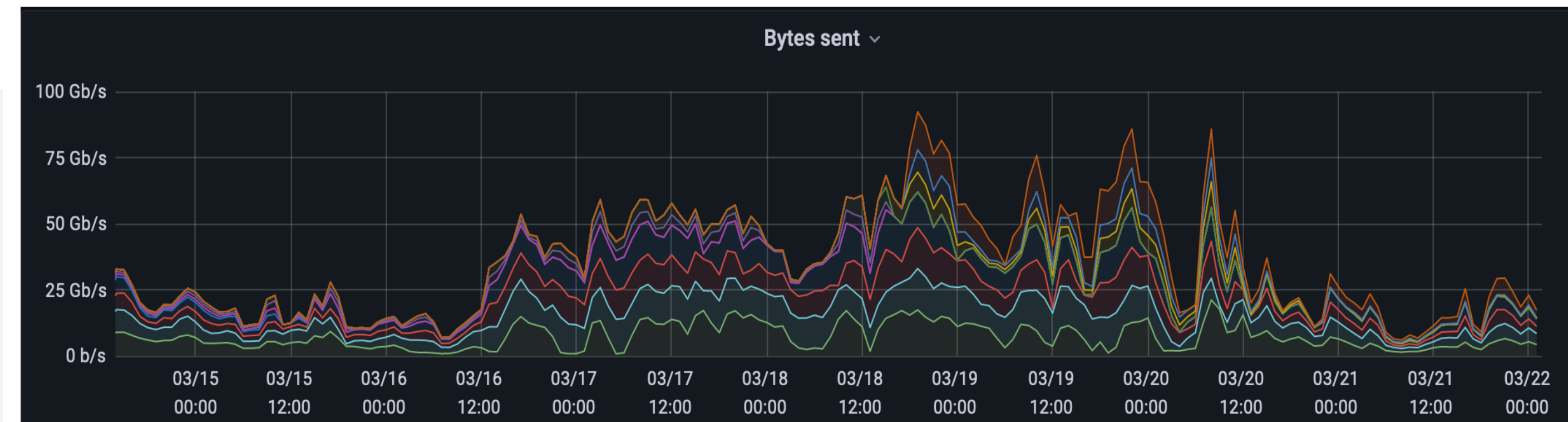
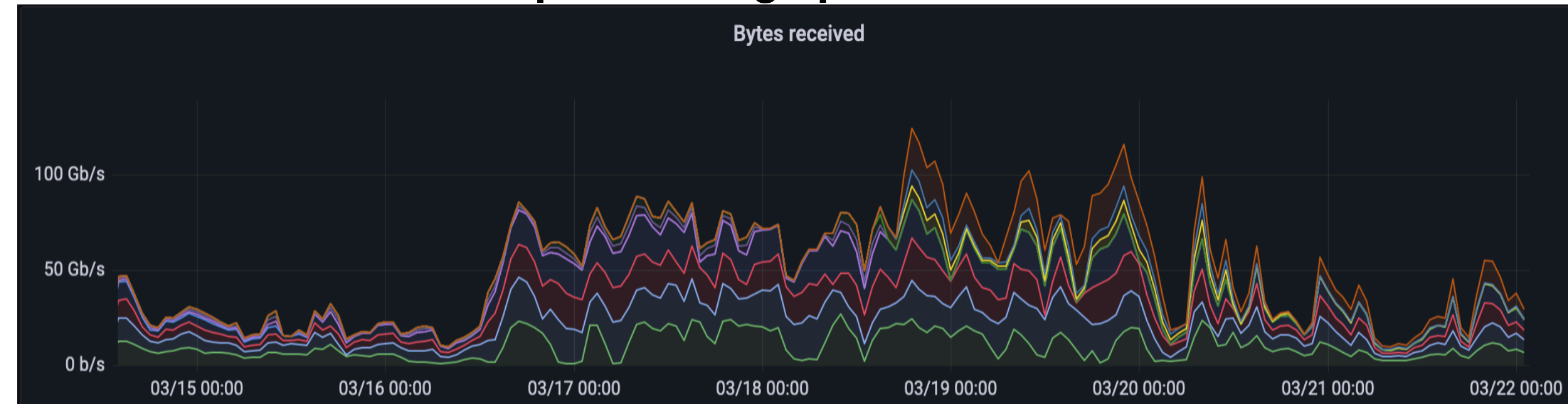
Writes to RAL Transfer Throughput



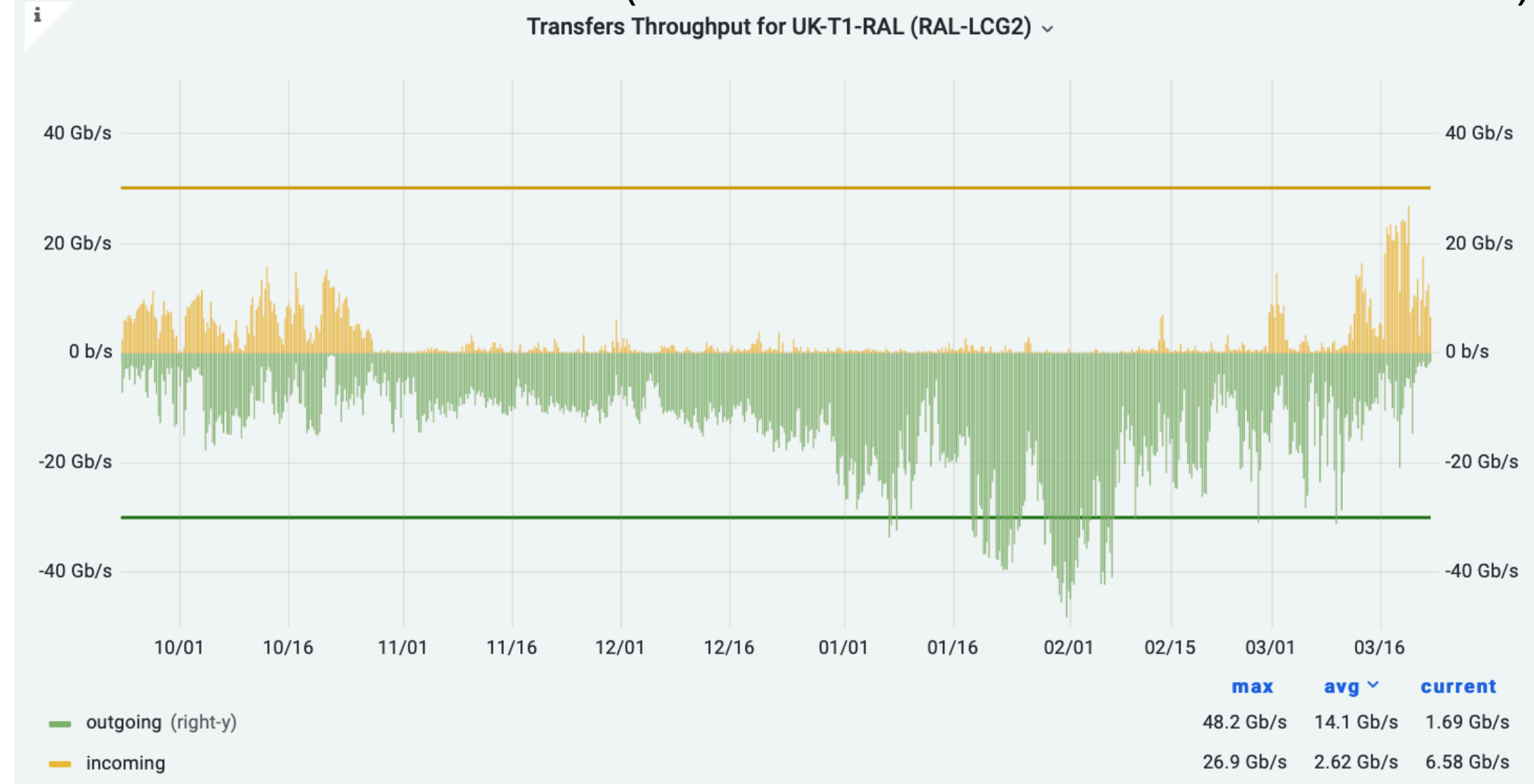
RAL: operating with Davs (2)

- Peak rates of ~ 120Gb/s observed passing through the Echo Gateways during the TAPE challenge (largely from LHCb Tape challenge workflow)
- Additional hardware now added to help the load
- Further developments / improvements are ongoing / planned

Tape challenge period



Davs over last 6 months (FTS transfers for ATLAS/CMS/LHCb)



WLCG Tokens compliance test-suite

- RAL participating in JWT compliance tests for WLCG token support for storage endpoints
 - Daily suite of tests run (results posted to DOMA BDT mailing list)
- WLCG JWT compliance tests using Robot Framework for its test automation
- Bonn making good progress with their XrootD config
 - RAL almost there ... (and some work on the xrootd devs side)
- Overall, strong progress recently in improving overall compliance pass scores.

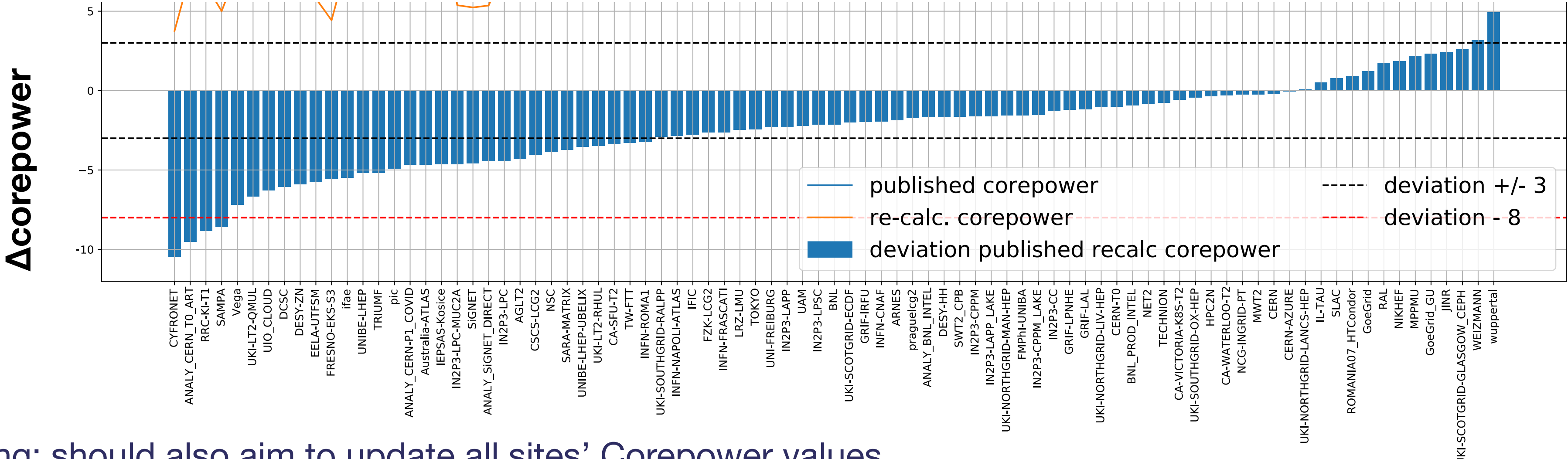
Total Statistics	Total	Pass	Fail	Skip	Elapsed	Pass / Fail / Skip
All Tests	168	125	43	0	00:03:48	
Statistics by Tag	Total	Pass	Fail	Skip	Elapsed	Pass / Fail / Skip
se-cern-eos	21	8	13	0	00:00:40	
se-cnaf-amnesiac-storm	21	21	0	0	00:00:22	
se-fnal-dcache	21	19	2	0	00:00:46	
se-infn-t1-xfer-storm	21	21	0	0	00:00:19	
se-nebraska-xrootd	21	0	21	0	00:00:00	
se-prometheus-dcache	21	19	2	0	00:00:34	
se-ral-test-xrootd	21	17	4	0	00:00:31	
se-ubonn-xrootd	21	20	1	0	00:00:35	
Statistics by Suite	Total	Pass	Fail	Skip	Elapsed	Pass / Fail / Skip
JWT compliance tests	168	125	43	0	00:04:08	
JWT compliance tests . se-cnaf-amnesiac-storm	21	21	0	0	00:00:24	
JWT compliance tests . se-cnaf-amnesiac-storm . Audience	4	4	0	0	00:00:03	
JWT compliance tests . se-cnaf-amnesiac-storm . Basic Authz	17	17	0	0	00:00:19	
JWT compliance tests . se-infn-t1-xfer-storm	21	21	0	0	00:00:21	
JWT compliance tests . se-infn-t1-xfer-storm . Audience	4	4	0	0	00:00:03	
JWT compliance tests . se-infn-t1-xfer-storm . Basic Authz	17	17	0	0	00:00:16	
JWT compliance tests . se-prometheus-dcache	21	19	2	0	00:00:36	
JWT compliance tests . se-prometheus-dcache . Audience	4	4	0	0	00:00:07	
JWT compliance tests . se-prometheus-dcache . Basic Authz	17	15	2	0	00:00:27	
JWT compliance tests . se-fnal-dcache	21	19	2	0	00:00:50	
JWT compliance tests . se-fnal-dcache . Audience	4	4	0	0	00:00:05	
JWT compliance tests . se-fnal-dcache . Basic Authz	17	15	2	0	00:00:41	
JWT compliance tests . se-cern-eos	21	8	13	0	00:00:42	
JWT compliance tests . se-cern-eos . Audience	4	4	0	0	00:00:07	
JWT compliance tests . se-cern-eos . Basic Authz	17	4	13	0	00:00:33	
JWT compliance tests . se-nebraska-xrootd	21	0	21	0	00:00:04	
JWT compliance tests . se-nebraska-xrootd . Audience	4	0	4	0	00:00:00	
JWT compliance tests . se-nebraska-xrootd . Basic Authz	17	0	17	0	00:00:00	
JWT compliance tests . se-ral-test-xrootd	21	17	4	0	00:00:34	
JWT compliance tests . se-ral-test-xrootd . Audience	4	4	0	0	00:00:04	
JWT compliance tests . se-ral-test-xrootd . Basic Authz	17	13	4	0	00:00:27	
JWT compliance tests . se-ubonn-xrootd	21	20	1	0	00:00:37	
JWT compliance tests . se-ubonn-xrootd . Audience	4	4	0	0	00:00:05	
JWT compliance tests . se-ubonn-xrootd . Basic Authz	17	16	1	0	00:00:31	

HS06 on the Grid

- Dedicated local order used for various benchmarking and performance studies
 - SPOT: 3 kHS06 sec/ event simulation of ttbar events on dedicated server
- Grid: Where most jobs will end up running
 - GRID average: 4.7 kHS06 sec / event - average mc16 13TeV simulated event in 2020 on the GRID
- Impacts on Resource estimation requirements
- Use Hammercloud infrastructure to run test jobs against ATLAS queues

- ▶ pos. deviation = perf underestimated
- ▶ neg. deviation = perf. overestimated

- Recalculated corepower per site based on normalising to 3kHS06s/evt
- Also noted however that scaling between benchmarking and test cases with hardware generations don't scale linearly.

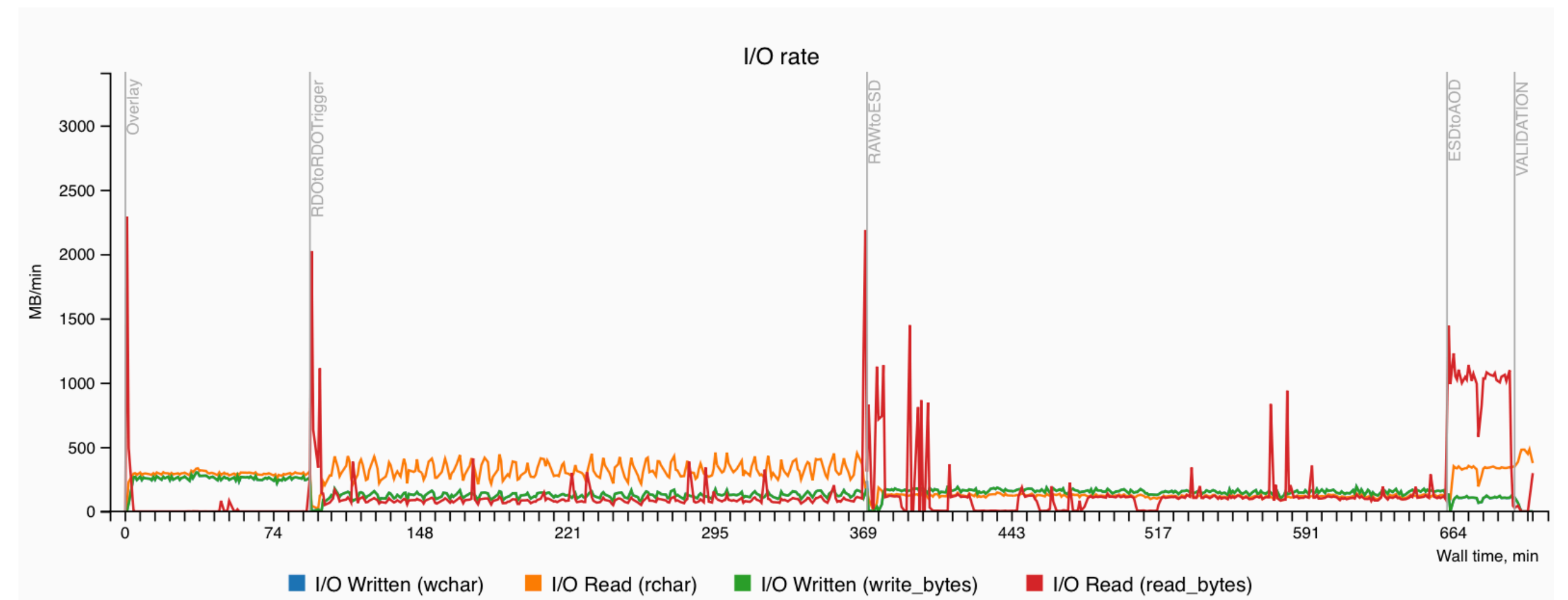
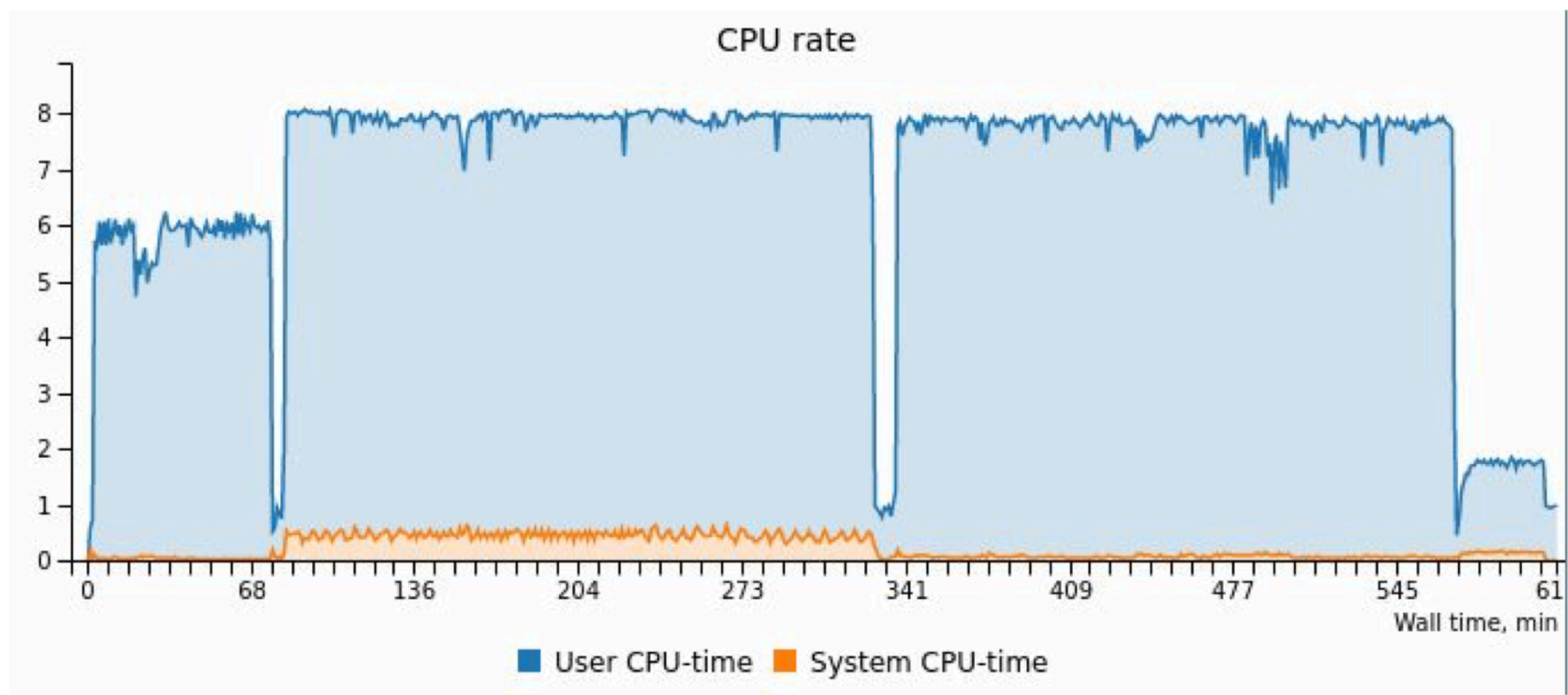
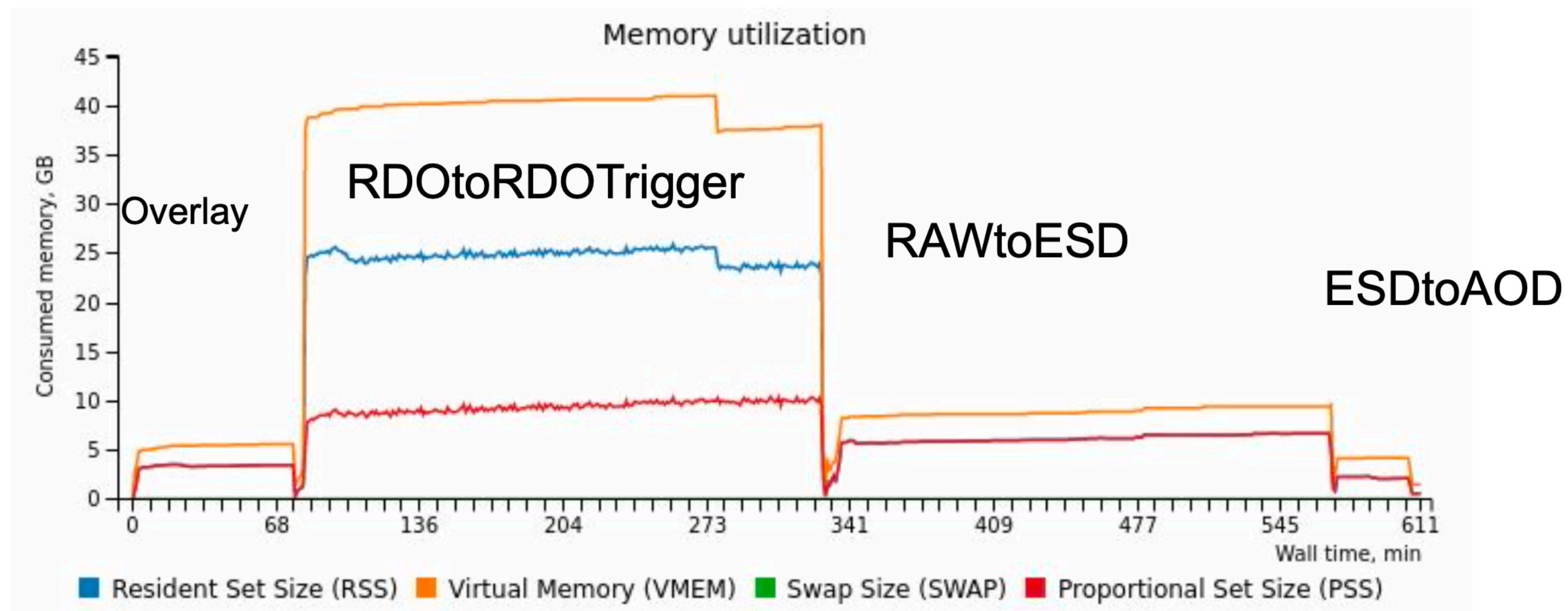


- Ongoing effort on new benchmarking; should also aim to update all sites' Corepower values.

Michael Boehler

MC reco job lifecycle

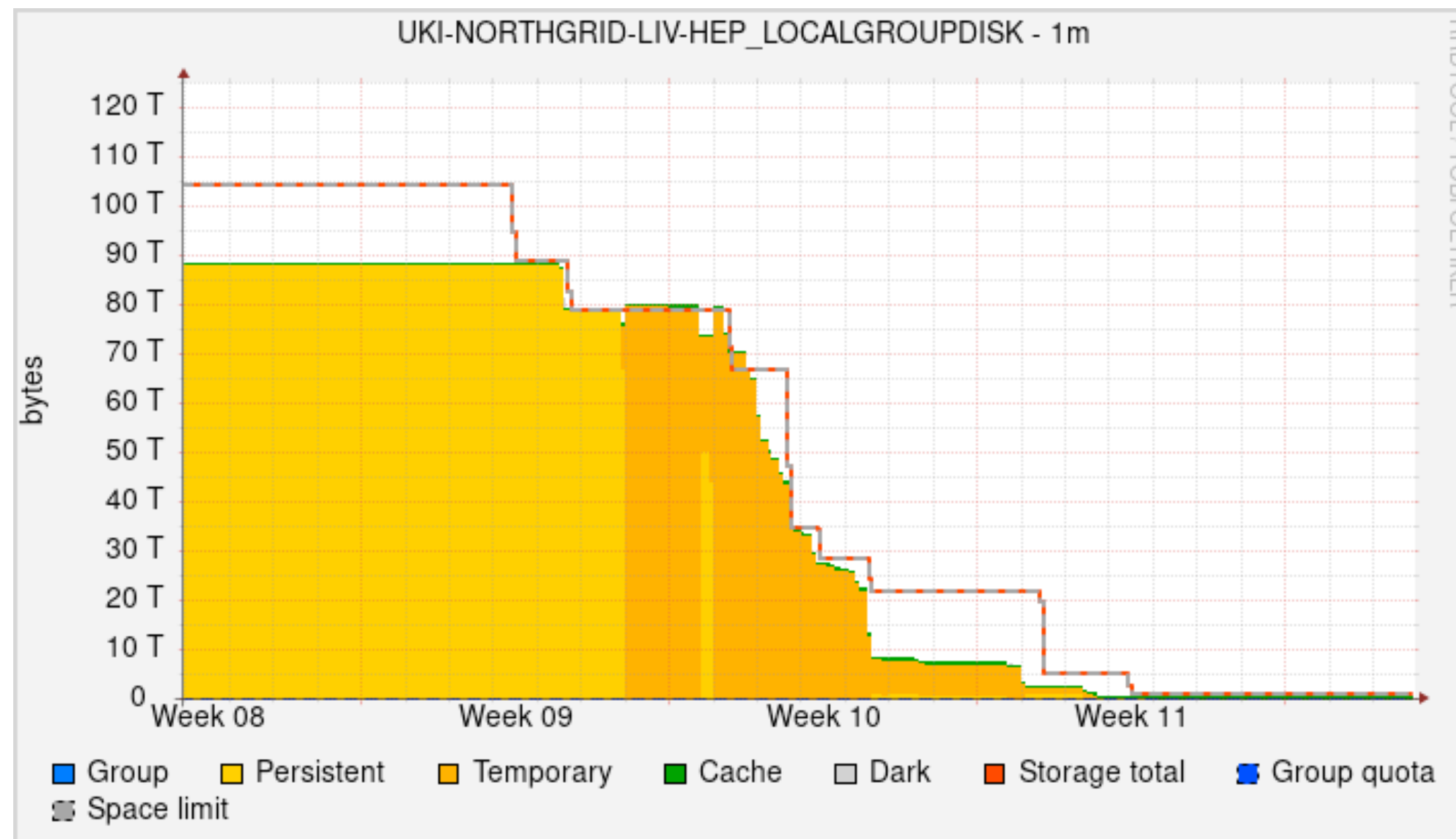
- Reconstruction jobs manage a number of different steps within the job.
- Plots of the anatomy of CPU, memory and IO for rel 22 MT Athena MC reco task.
- CPU generally well utilised over the lifetime of job



I/O plot from similar job

Storage migrations / decommissioning

- Number of UK sites have / are / will be migrating to new storage solutions
- Managing the data movement can take time and needs coordination with experts
 - Let Cloud team know as early as possible to help mitigate any issues
- Similarly, extended downtimes;
 - please also communicate with advanced notice where possible.



Timeline to Run-4

Distributed Computing				
MID	DID	Description	Due	
DC-1		Transition to tokens	Q4 2025	
	1.1	Submission from Harvester to all HTCondor CEs with tokens	Q1 2022	
	1.2	All users move from VOMS to IAM for X509	Q4 2022	
	1.3	All job submission and data transfers use tokens	Q4 2025	
DC-2		Storage evolution	Q4 2025	
	2.1	No GridFTP transfers at any site	Q1 2022	
	2.2	SRM-less access to tape	Q4 2025	
	2.3	Recommended transition plan from DPM completed	Q4 2021	
	2.4	Transition plan from all DPM sites	Q4 2022	
	2.5	All sites moved away from DPM	Q2 2024	
DC-3		Next operating system version	Q2 2024	
	3.1	Ability to run on "future OS" on grid sites	Q4 2022	
	3.2	Central services moved to "future OS"	Q4 2023	
	3.3	(CentOS 7/8 EOL)	Q2 2024	
DC-4		Network infrastructure ready for Run 4	Q4 2027	
	4.1	Network challenge at 10% expected rate	Q4 2021	
	4.2	Network challenge at 30% expected rate	Q4 2023	
	4.3	Network challenge at 60% expected rate	Q4 2025	
	4.4	Network challenge at 100% expected rate	Q4 2027	

DC-5		Integrating next generation of HPCs	Q2 2023
	5.1	Integration of at least 2 EuroHPC sites	Q4 2022
	5.2	Integration of next generation US HPCs for production	Q2 2023
DC-6		Exploratory R&D on GPU-based workflows for next generation HPC	Q4 2023
DC-7		HL-LHC datasets replicas and versions management	Q2 2024
	7.1	Replicas and versions detailed accounting	Q4 2022
	7.2	DAOD replicas reduction	Q4 2023
	7.3	DAOD versions reduction	Q2 2024
DC-8		Data Carousel for storage optimization	Q4 2023
	8.1	Investigate with sites the cost of Tape infrastructure and the estimated cost in case of sensible increase of read/write throughput	Q4 2022
	8.2	Reduce the AOD on disk to 50% of the total AOD volume, using Data Carousel to orchestrate the stage from tape for DAOD production.	Q4 2023
DC-9		Disk management: secondary(cached) dataset	Q2 2023
	9.1	Evaluate the impact on job brokering and task duration if disk space for secondary data is reduced	Q2 2023