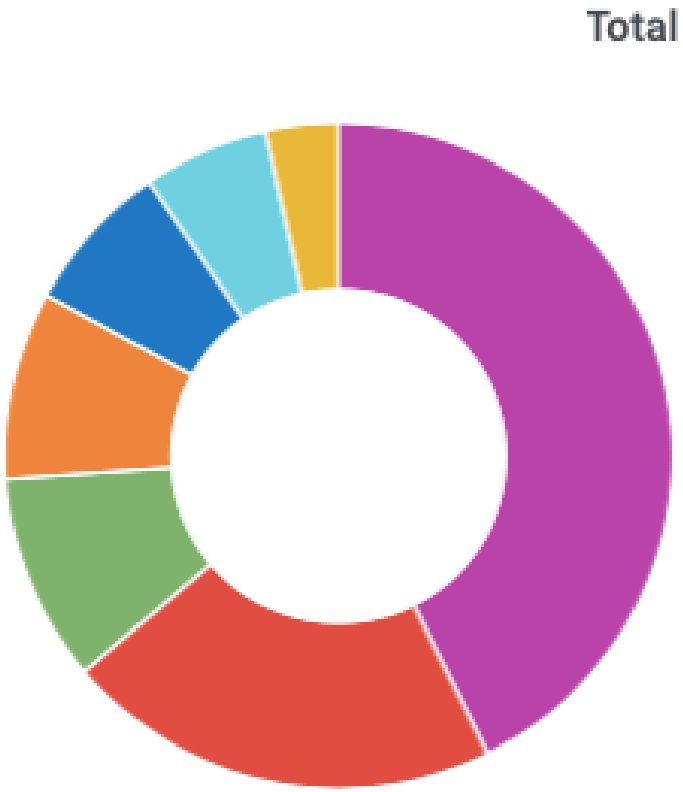
CMS Tier-1 Experiment sign off for Q1 2021

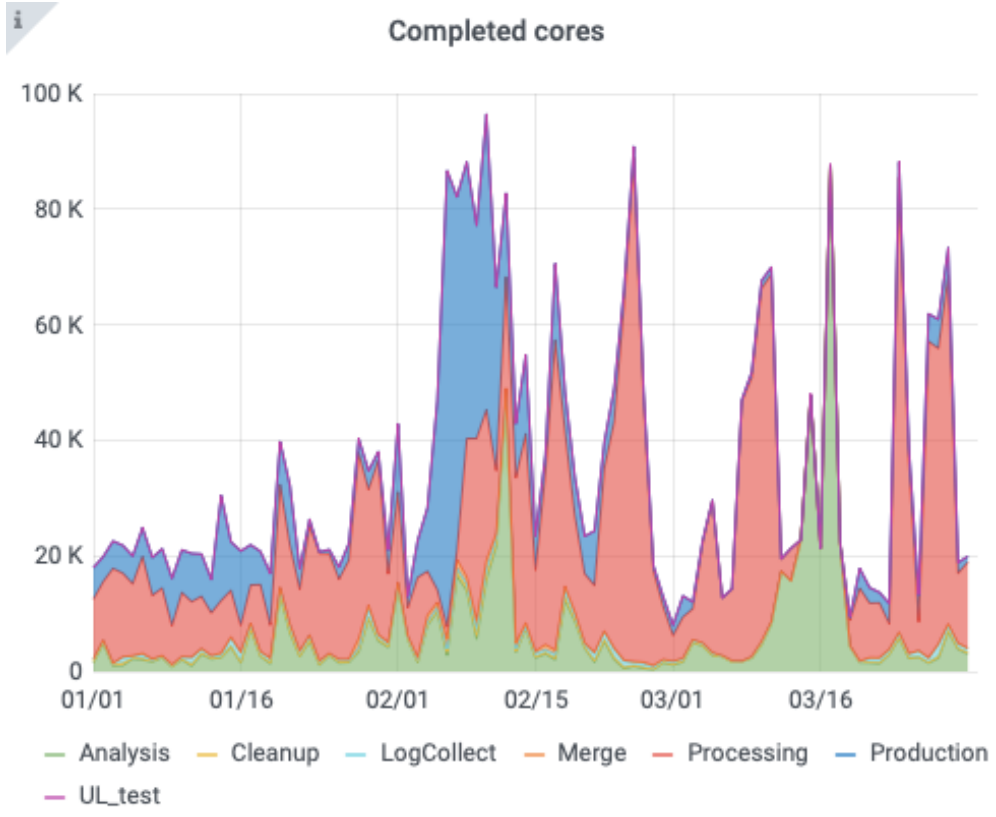
Katy Ellis, 28 Apr 2021

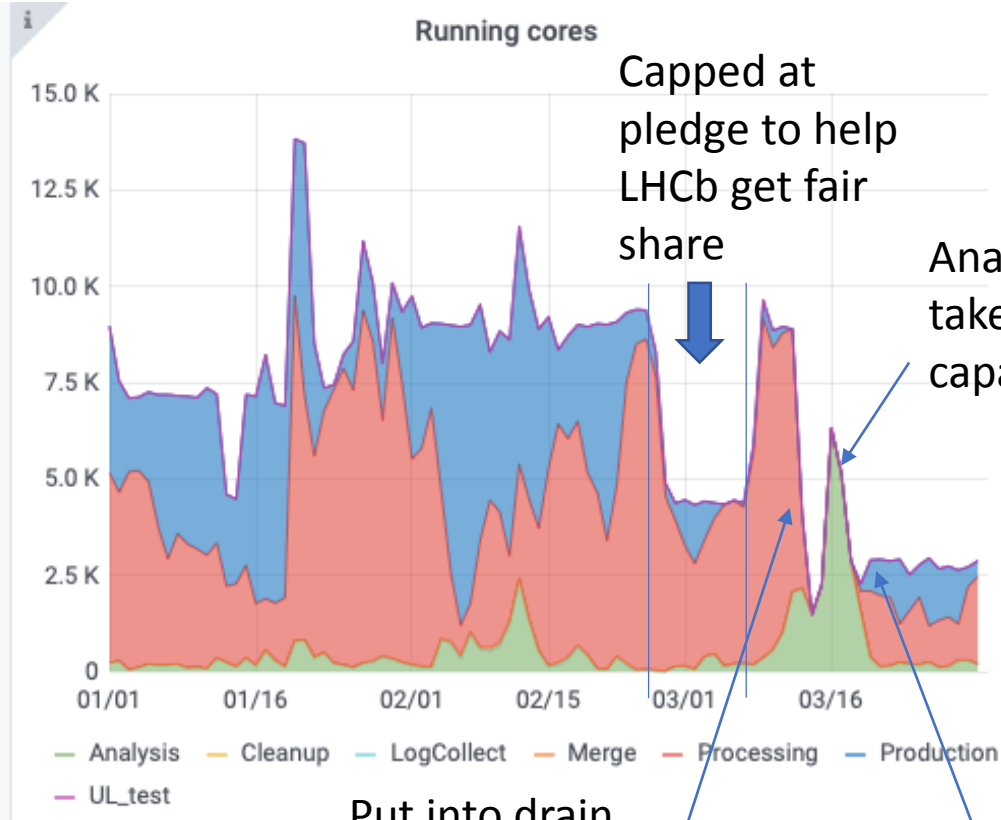
Completed jobs at T1s



	total	percentage
T1_US_FNAL	5921751	42.6%
T1_RU_JINR	2941326	21.2%
T1_DE_KIT	1402424	10.1%
T1_IT_CNAF	1269125	9.1%
T1_UK_RAL	1022571	7.4%
T1_FR_CCIN2P3	844563	6.1%
T1_ES_PIC	493103	3.5%

Completed and running cores at RAL





Capped at pledge to help LHCb get fair share

Analysis jobs take spare capacity

Put into drain due to extreme job inefficiency (disappearing WN xrootd containers)

Brought out of drain for prod with capped 3k cores

Running cores was easily above pledge on average over the quarter. Proportion of T1 completed cores is lower than proportion of T1 running cores due to lower efficiency of jobs at RAL. Completed cores includes failures.

Summary table of jobs, Q1

- https://monit-grafana.cern.ch/d/C8ewaCrWk/hs06-report?orgId=11&from=1609459200000&to=1617231599000

Site	Job Count	Failed jobs	CPU Eff	HS06CoreHr	CpuTimeHr	CoreHr
T1_US_FNAL	5811739	1553215	74.5%		31846325.35	42754077.22
T1_UK_RAL	1011627	276551	30.4%		3982392.07	13100684.88
T1_RU_JINR	2874015	402737	69.4%		17451127.46	25156886.78
T1_IT_CNAF	1239184	174307	76.9%		13329857.35	17344557.20
T1_FR_CCIN2P3	836165	83229	74.5%		7334848.60	9844656.55
T1_ES_PIC	498898	101227	71.0%		3611839.54	5087871.76
T1_DE_KIT	1368991	136049	70.3%		14400873.36	20485015.89

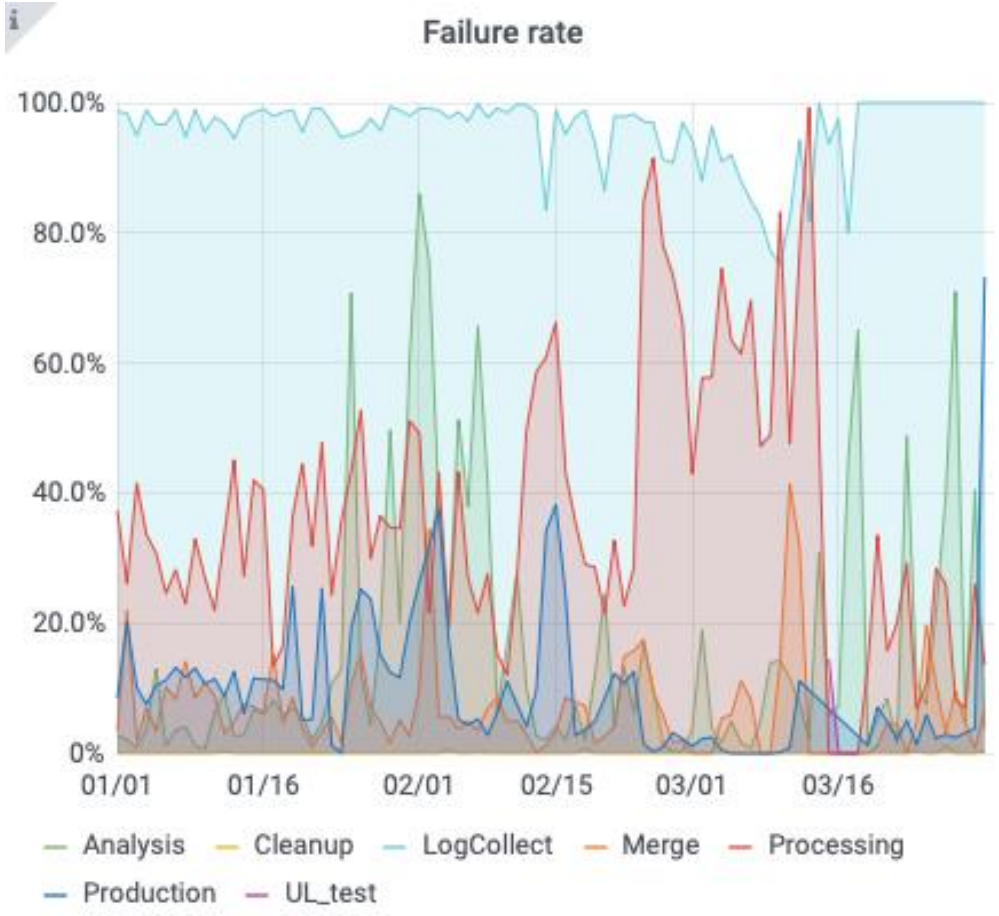
FNAL also has 27% failure rate, others (much) better

27% failure rate, (23/29% in Q3/4)

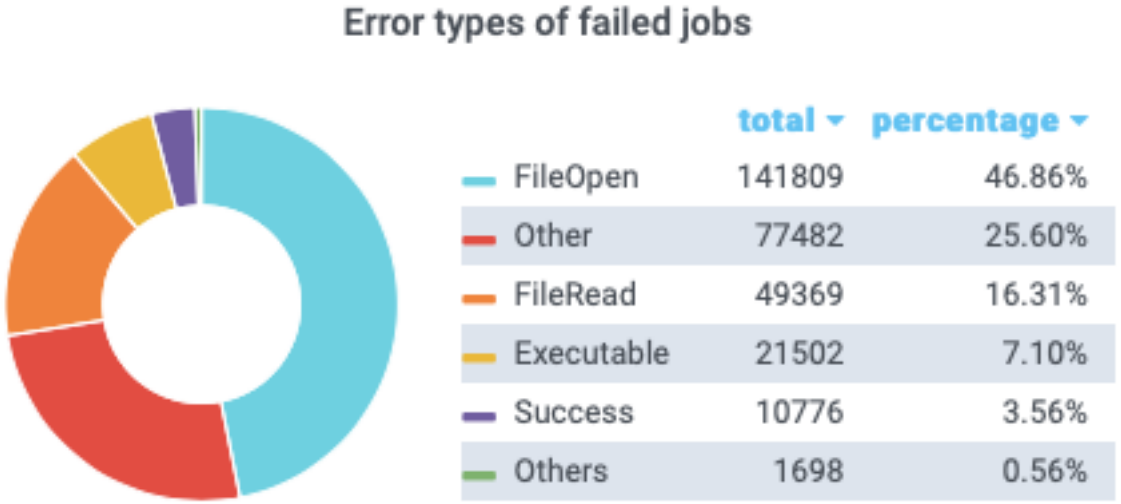
Continuing to fall (was 48/46% in Q3/4)

I don't trust this number

Failed jobs

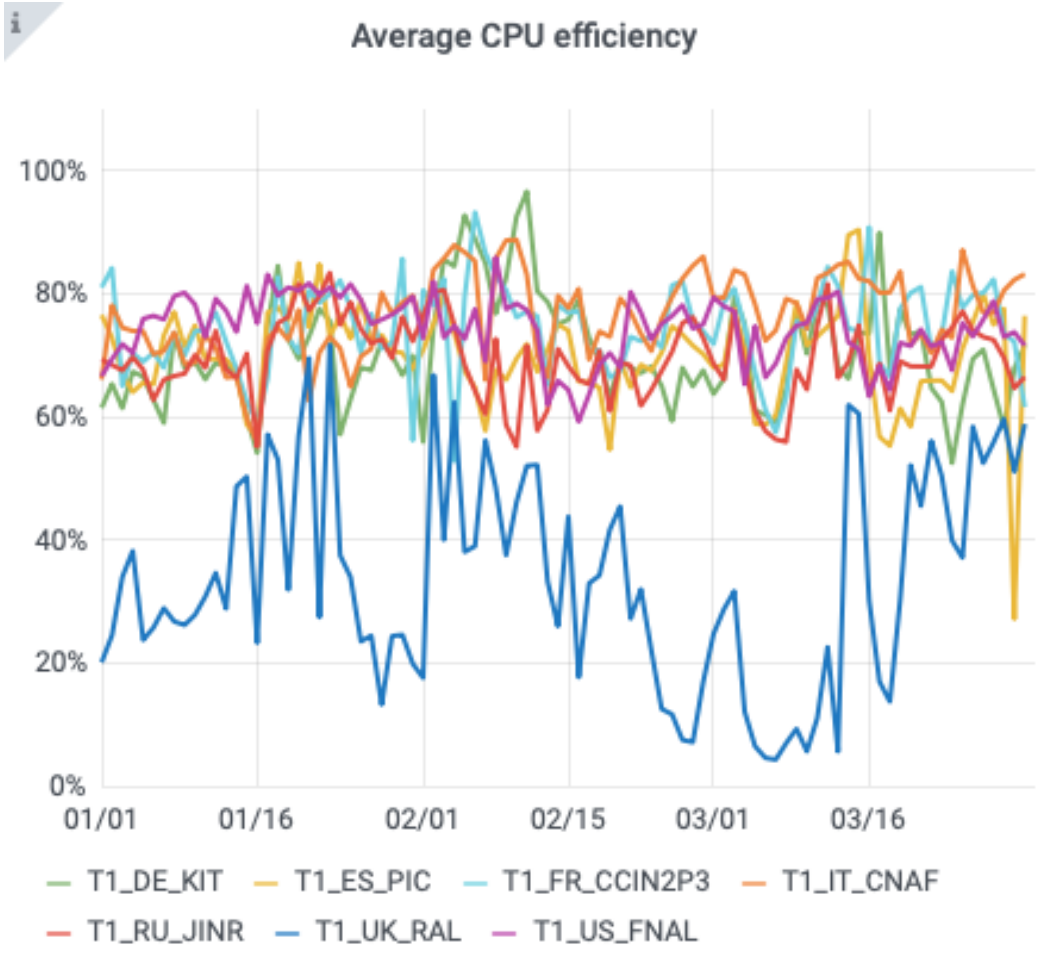


N.B. Virtually all LogCollect jobs fail due to a known problem (16% of all failures this quarter).

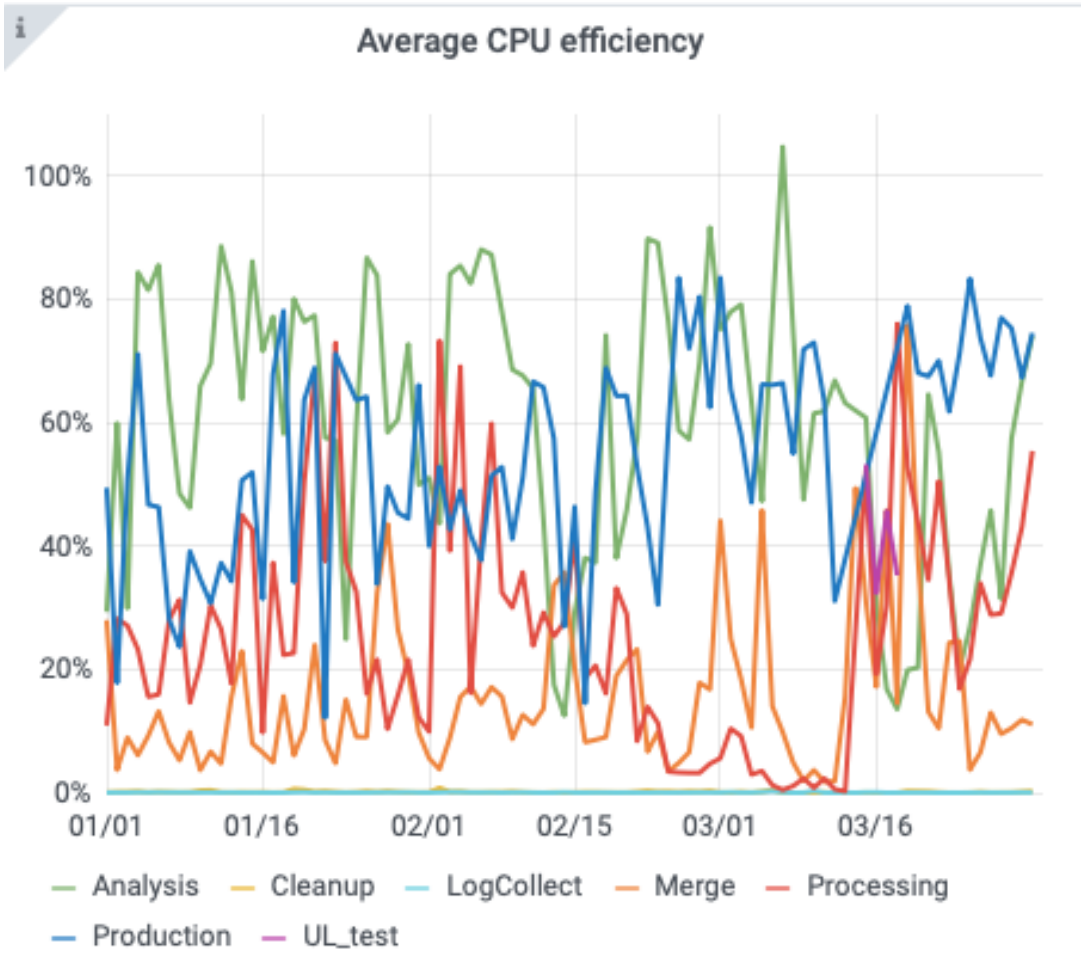


- Failures are dominated by file access issues (FileOpen/FileRead)
- This could be onsite or offsite reads...but…
- Jobs of type Processing caused a lot of failures – these are typically high-I/O, reading many files and a lot of this comes from offsite, streaming over AAA.

CPU efficiency – including failed jobs

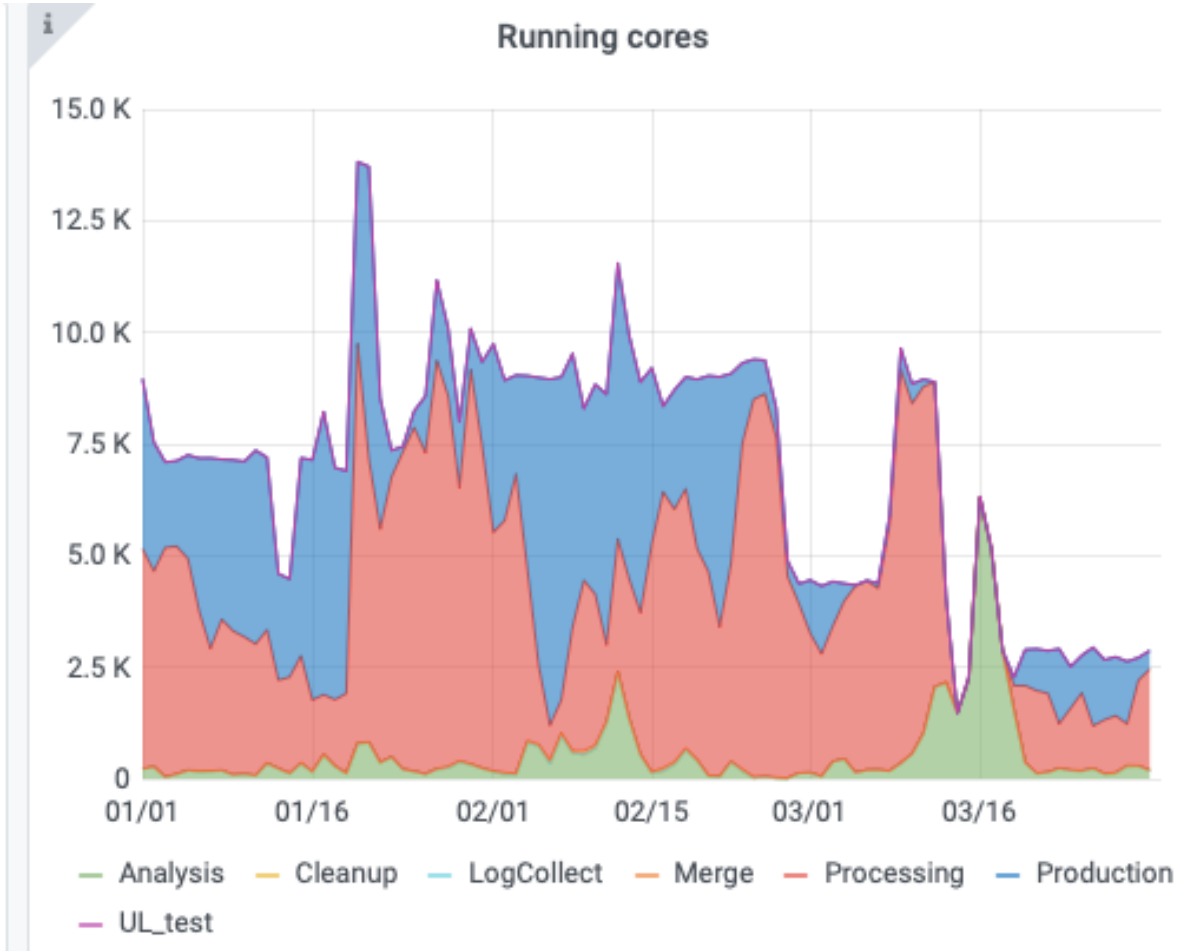


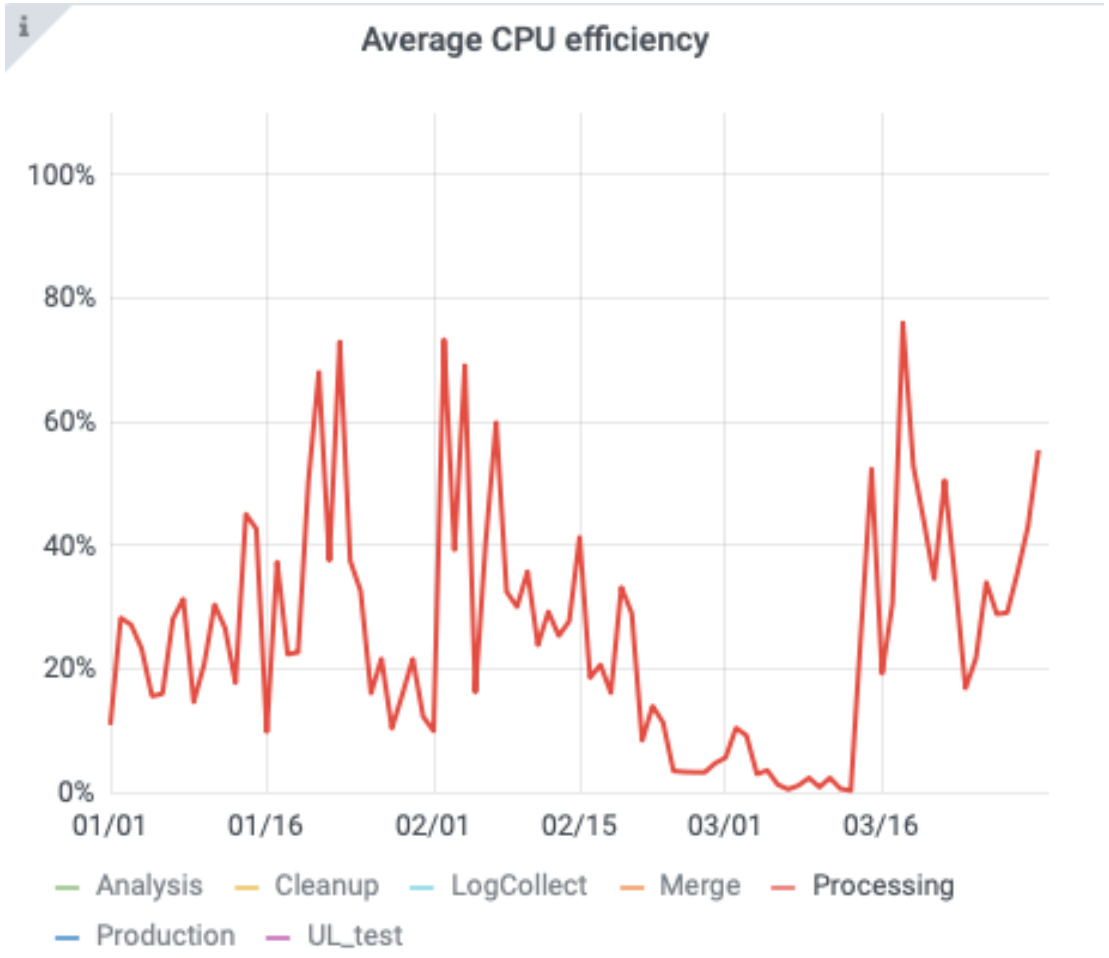
All Tier 1s



At RAL, split by job type

CPU efficiency – including failed jobs



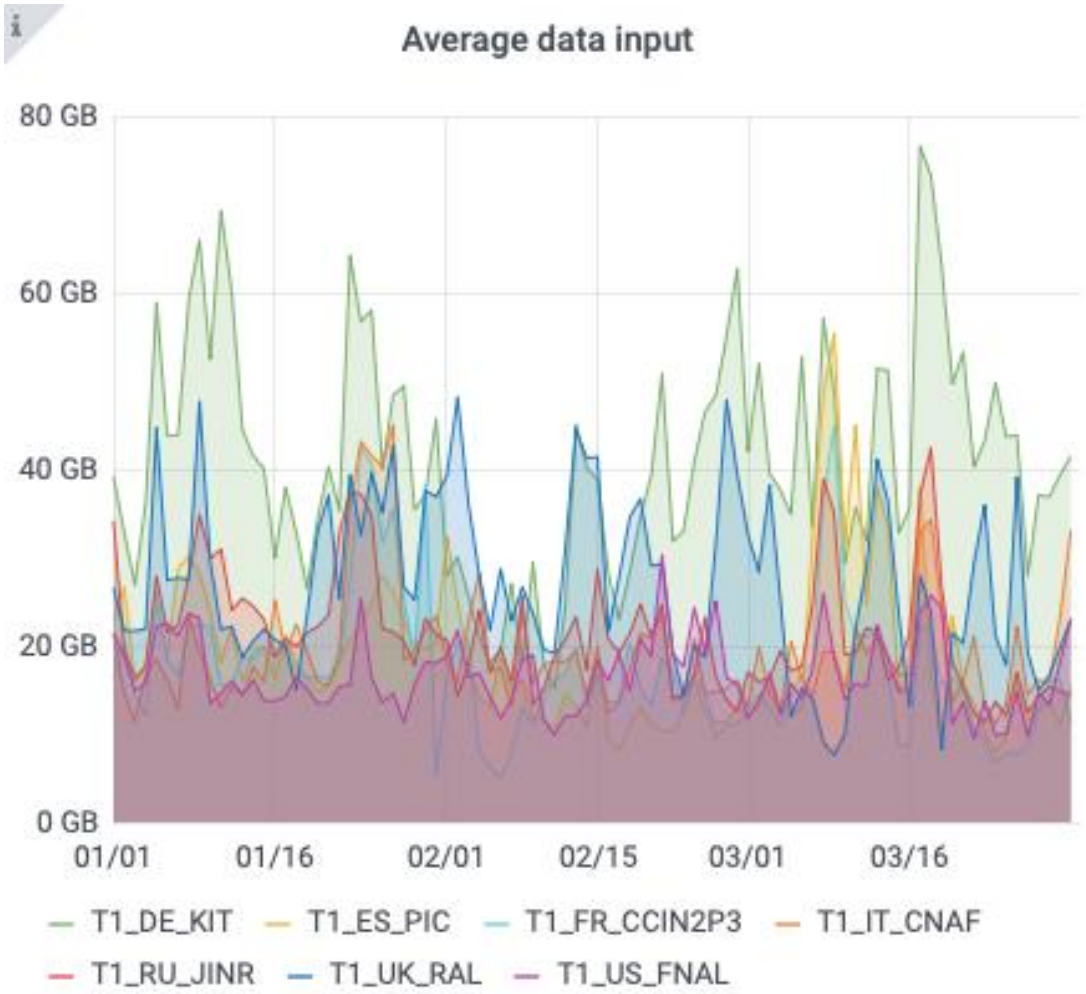


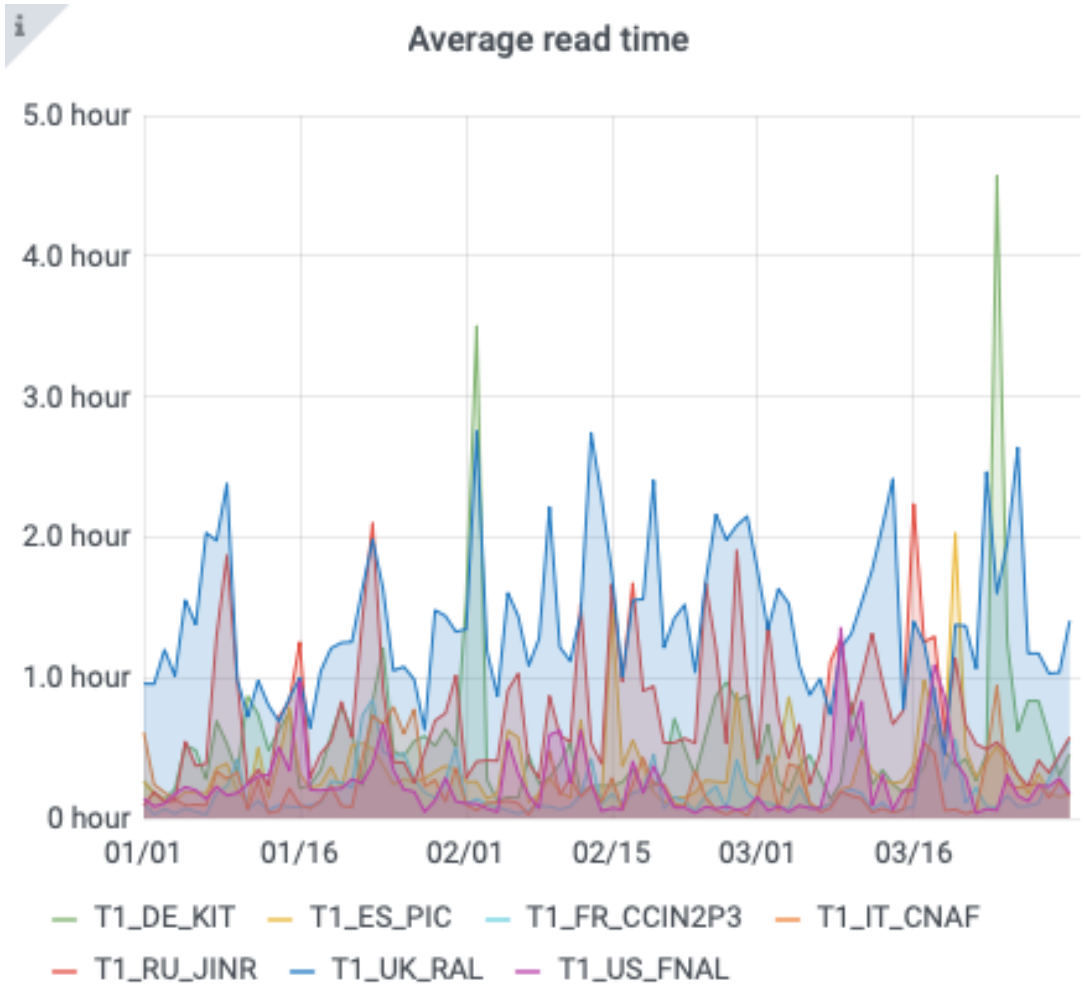
At RAL, split by job type

Why efficiency was so incredibly low in March

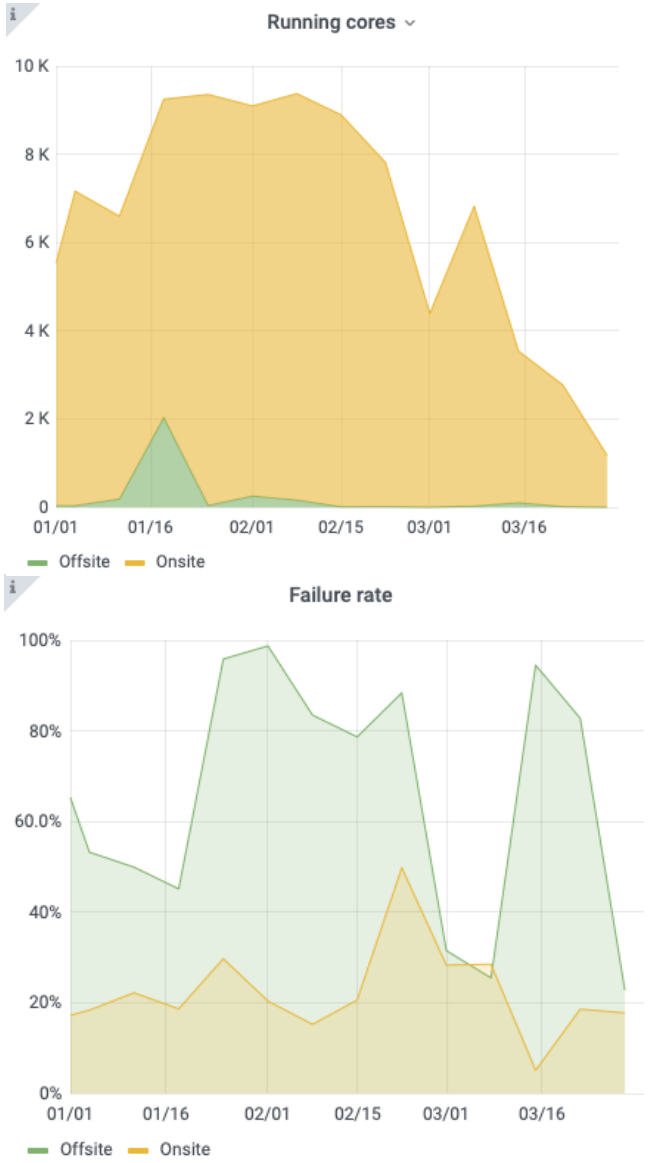
- Starting before 24th Feb :–
- As a result of a bug in the kernel that had been updated on some WNs
- WNs would sometimes reboot and return without XRootD gateways
- Jobs start with XRootD gateway containers missing
  - Unknown why jobs still scheduled – supposed to look for ‘healthy’ gateway
- CMS jobs started but when file access was required went into idle state
- Even if gateway was present, the job remained in this state until the pilot expired
- Efficiency = CPUhours / Corehours
- Core hours were O(100s) for many jobs; efficiency was <0.1%

Input data and read time (all jobs)





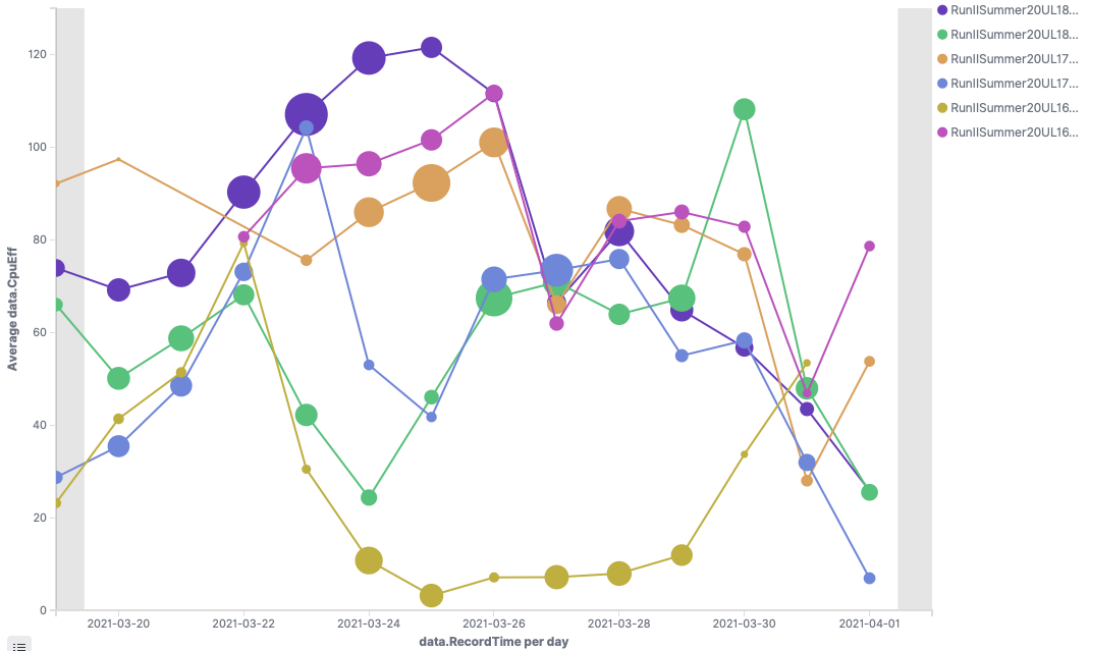
Offsite reads

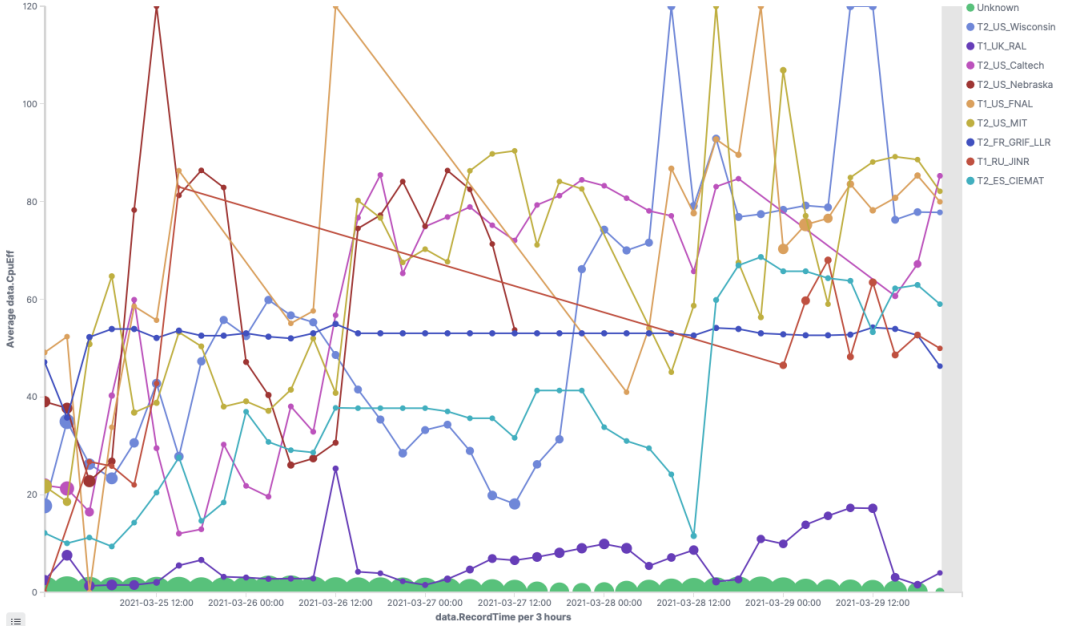


- Not a lot of jobs flagged ‘Offsite’ this Q but many of those that are, failed
- The problem with the Onsite/Offsite flag here is that it only applies to the Primary dataset
- Many jobs use significant secondaries
  - Most of which are offsite
- However, we can still see here a much worse failure rate than when primary dataset onsite
- As reported last quarter, the measured data rate from offsite reads through the firewall to the batch farm is much lower than expected for many sites
  - E.g. 1MB/s or even less

Onsite reads

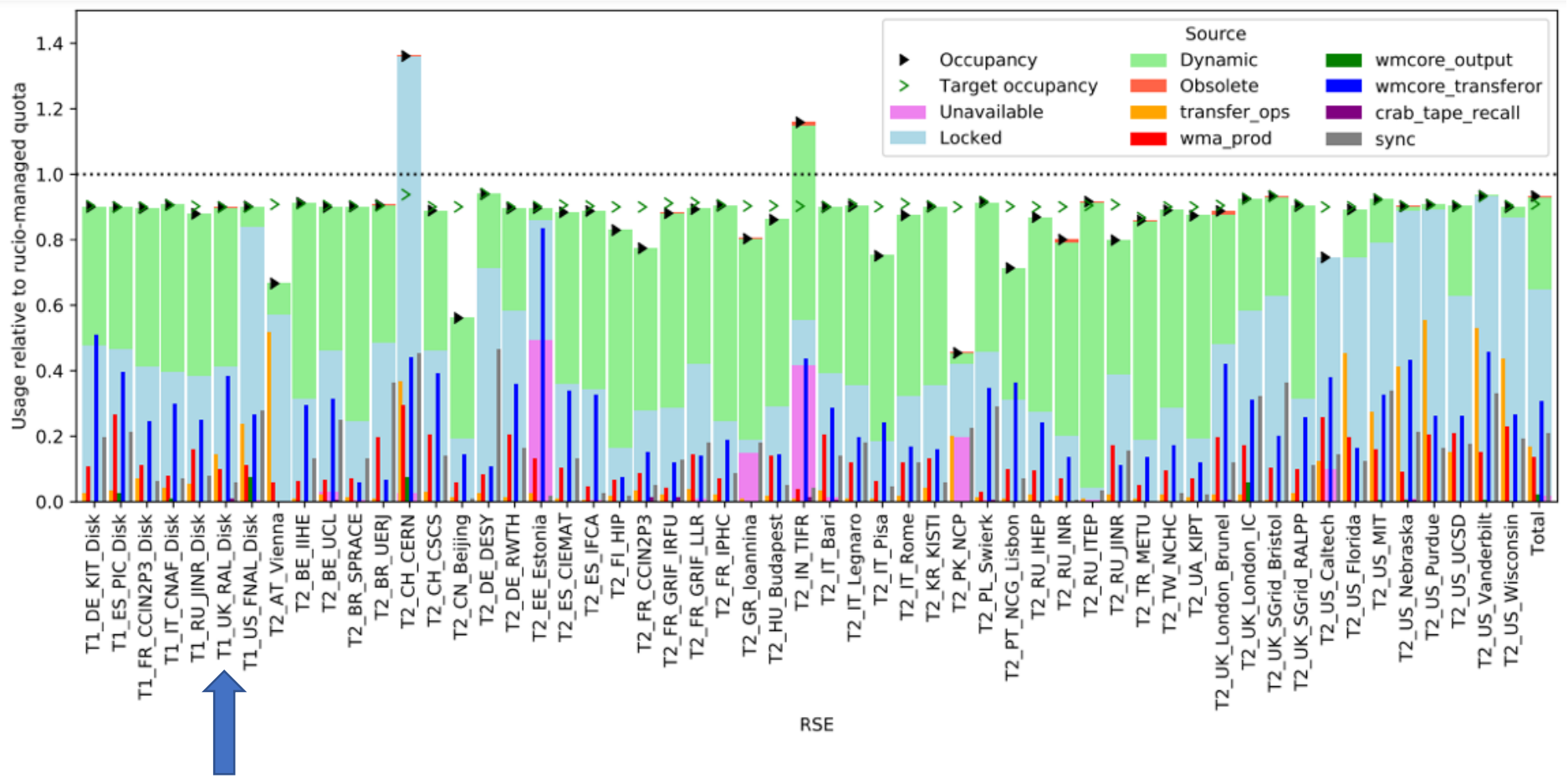
- The ‘vector read’ problem is likely an issue for CMS, as for LHCb.
- A test of onsite reads was made in March
  - Moved one of the 500TB ‘premix’ libraries to RAL which should enforce only onsite reads for the ‘20UL16’ campaigns.





Disk usage

Previous monitoring has disappeared (due to retirement of PhEDEx in Q4 2020).



Tape usage

- Previous monitoring has disappeared (due to retirement of PhEDEx in Q4 2020).

- RAL tape is still full and therefore not taking new transfers
- There have been some issues with the ‘loadtests’
- CMS data was being migrated to the new ‘Spectra’ tape system in Q1 – no problems to report!
  - I did a clean up of dark data before this started.
- Did not take part in the tape challenge, but will do in the next round.

Summary

- CPU usage:
  - Number of cores in use is over pledge on average.
  - Failure rate is still higher than many other T1s – we can do better.
  - Efficiency remains low and believed to be at least partially related to on and off site reads. New firewall from 21st April hoped to improve offsite read rates.
- Disk usage is high and being managed.
- Tape is still full at RAL and we have not accepted new writes since Dec.