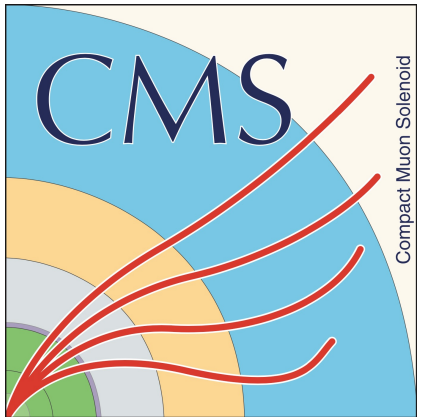
Science and Technology Facilities Council



CMS summary

Katy Ellis, CMS liaison at RAL Tier 1

23/03/22, GridPP47

Content

- General talk about CMS UK computing status, particularly Tier 1
  - Some emphasis on job efficiency
- Transfers to disk
- New tape system at RAL

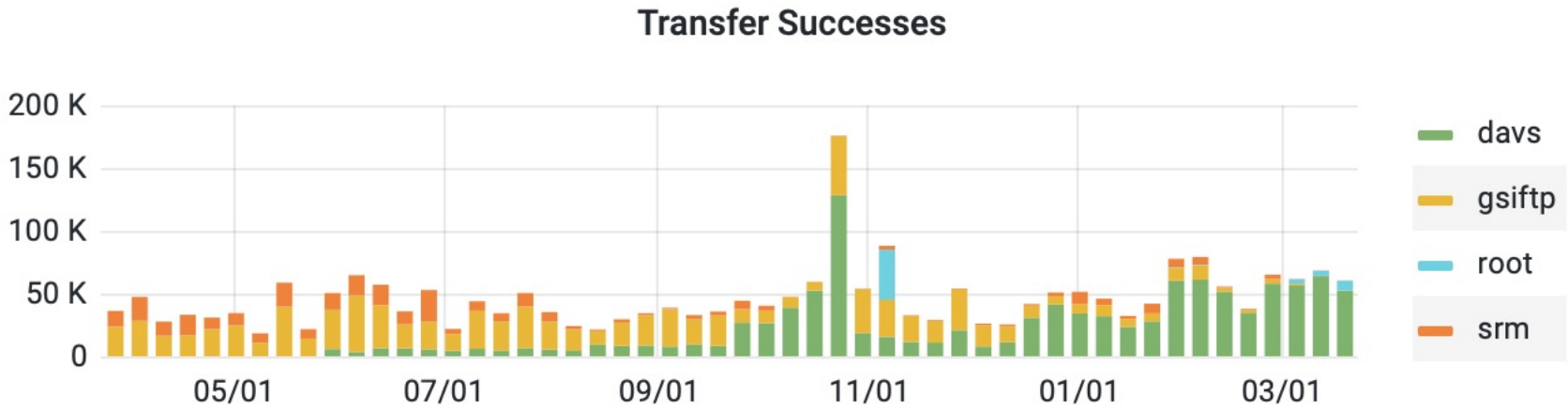
Antares tape migration

- A very significant part of work at T1 in recent months
- Preparation for tape metadata to be accessed from Antares (CERN Tape Archive – CTA) instead of the previous Castor tape system
- VOs tested with the pre-production instance
- Migration occurred in the week of 28th Feb 2022
- Dedicated talks to this coming up next…
- Not all plain-sailing in production with the CTA instance at CERN
  - Particularly with the Rucio ‘multihop’ functionality

WebDAV transfers

- CMS are replacing gsiftp transfers with the WebDAV protocol at all sites.
  - WebDAV based on XRootD with Third Party Copy capability
- Thanks to James Walder for getting this working with the Echo storage system at RAL.

Site	Webdav as primary
Tier 1	Y
IC	Y
RALPP	Y
Brunel	Y
Bristol	Y (last 2 weeks?)

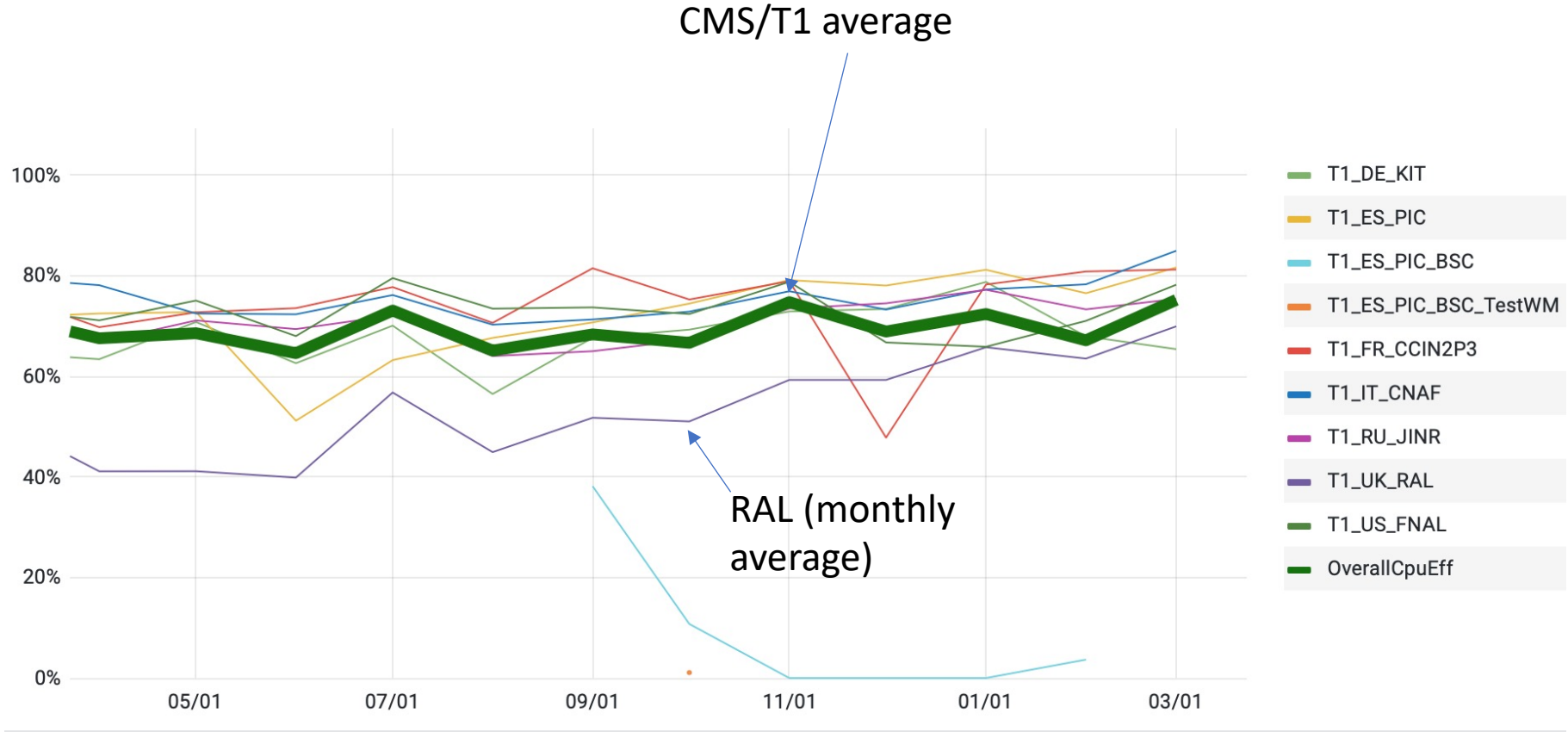


WebDAV issues at RAL T1

- A higher rate of failures for transfers using the WebDAV protocol than previously with gsiftp.
  - Particularly bad under high load
  - Tests fail when load is high, e.g last 30 days:

- This number of failures would be enough to keep T1 out of production work for CMS for the last few weeks
- Note that many transfers to tape will transit through Echo using WebDAV from now on.
- More hardware – “Echo gateways” have been added in the past week
  - But further analysis required on the WebDAV/Echo performance

Job efficiency - positive trend in last 1 year



Job efficiency depends on several factors, e.g.:

- Type of job
- Quality of code
- Amount of I/O
- Speed of access to inputs
- Ability of CPU

(CPU efficiency = CPU time / Core time)

Many changes in the past year to the network, and many more new CPUs with SSD instead of hard drives

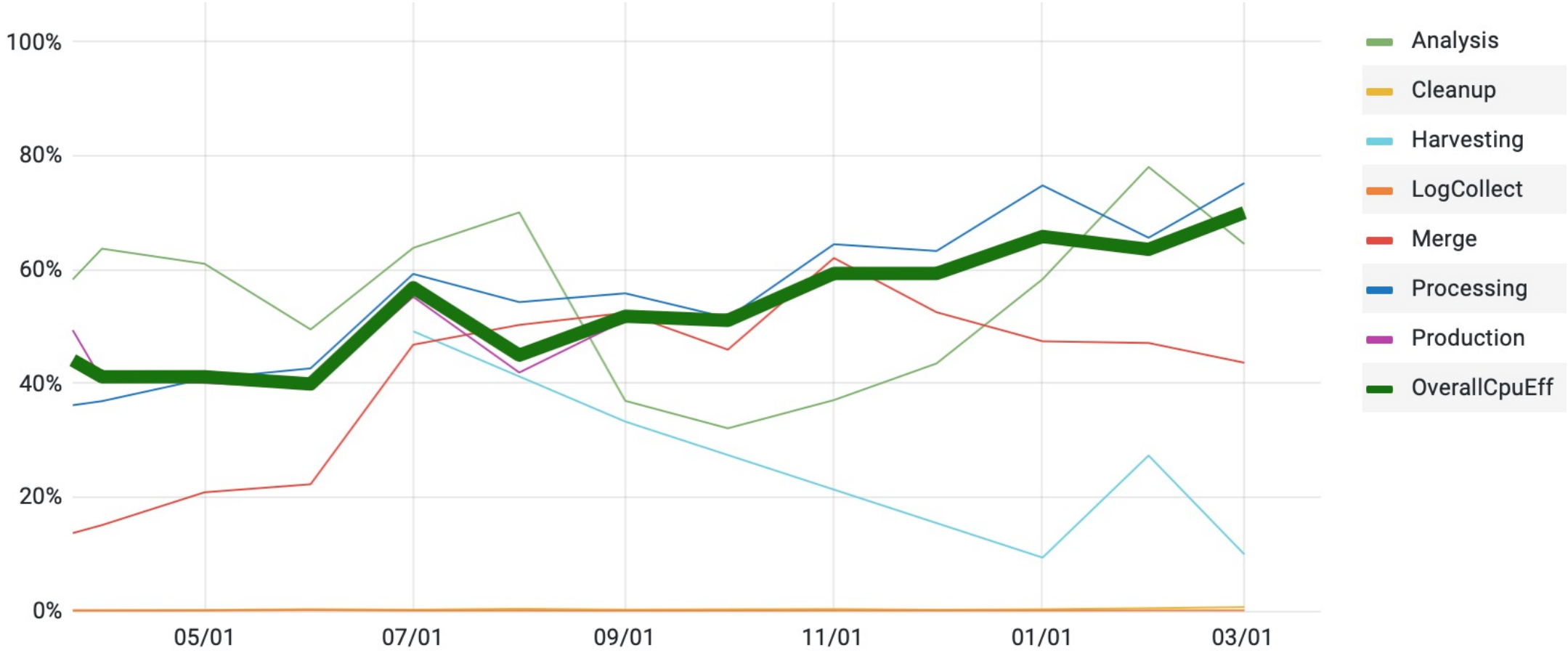
Some major changes in last 1-2 years

- New network design at Tier 1 – as described by Alastair
- Newest tranches of Worker Nodes have fast SSDs, larger memory and attached to the new network
- Newest disk storage nodes attached to the new network
- New firewall
- Etc.

- Coming shortly:
  - Connection to LHCONE, etc.

Please see the network/Tier 1 talk by Alastair Dewhurst: https://indico.cern.ch/event/1128343/contributions/4790689/

Different job types at RAL T1

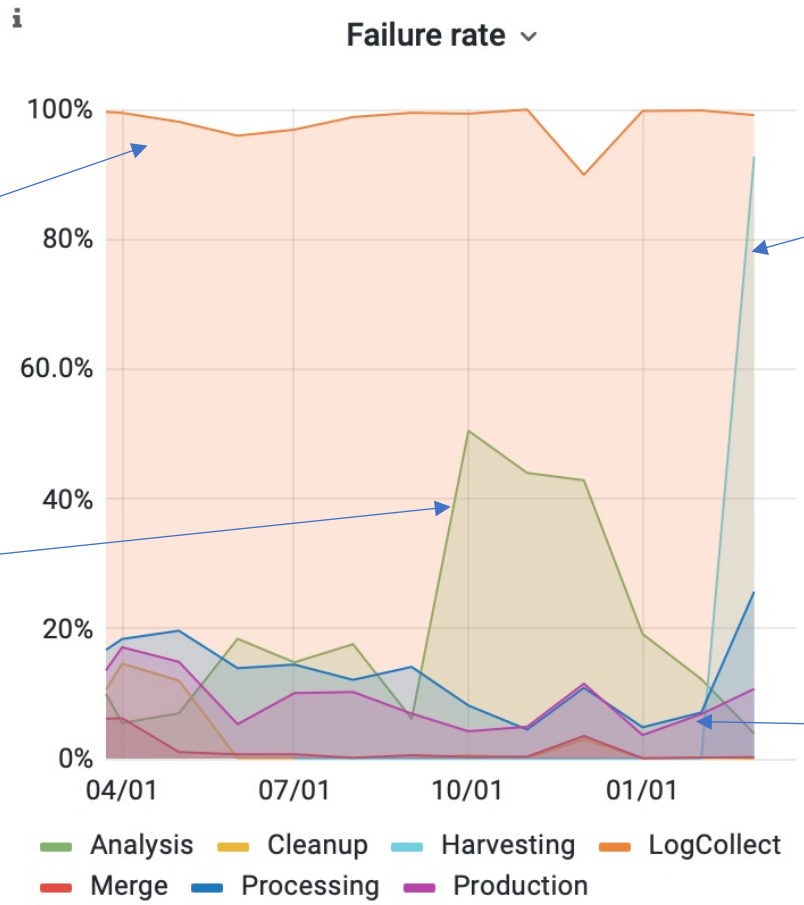


Job failures in last 1 year

LogCollect jobs:
Almost all fail for known reason – CMS to fix

User Analysis jobs:
Jobs unable to access local storage

Identified this via HammerCloud jobs, which use the same infrastructure as user analysis jobs – fixed by a change in that code



Harvesting jobs:
To be investigated ASAP!

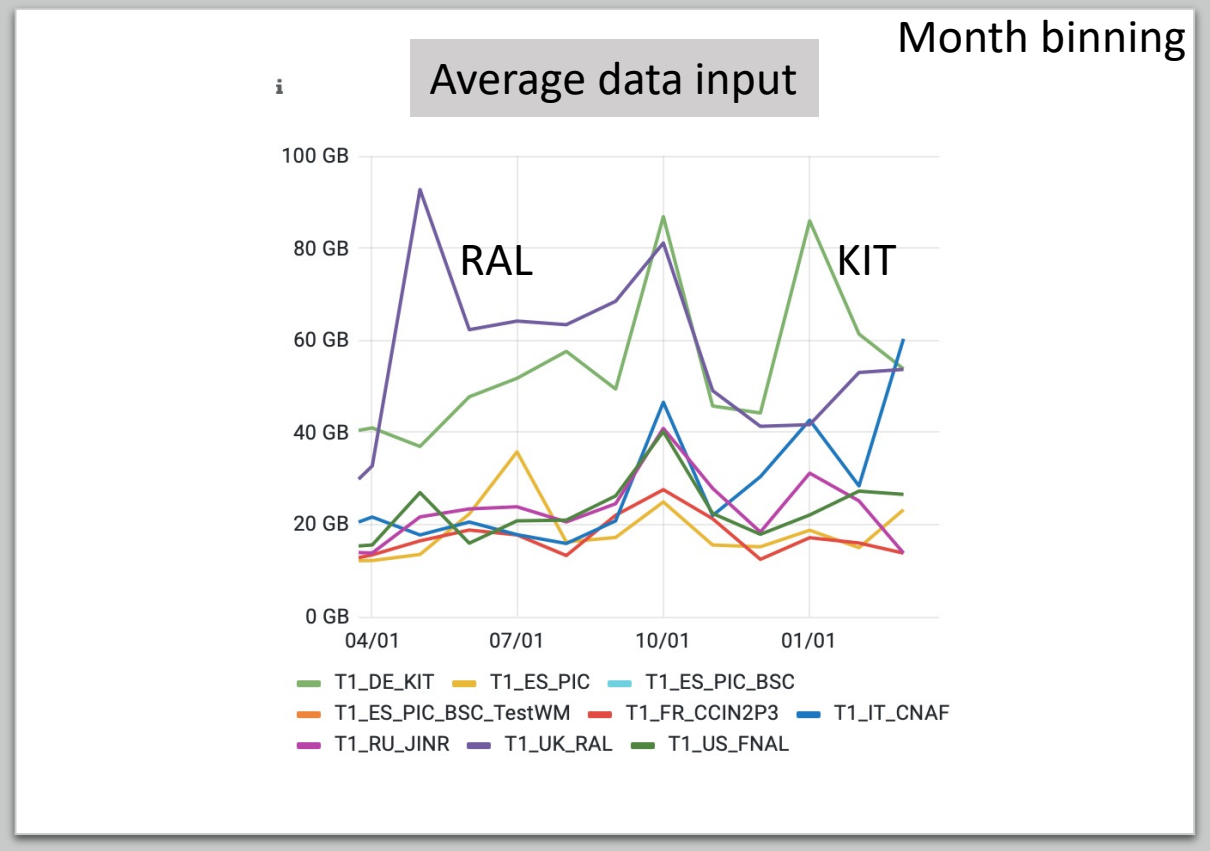
Production and Processing (vast majority of jobs):
General improvement?

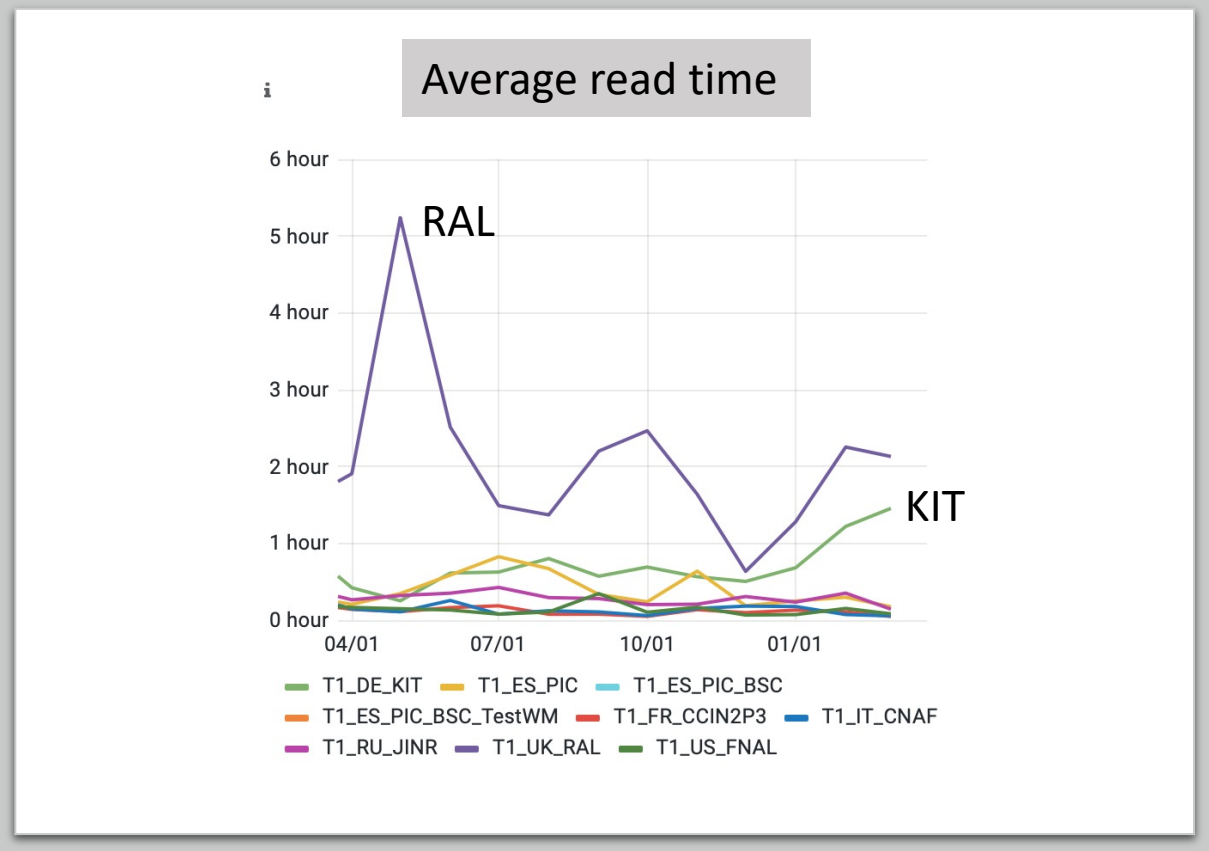
“Lazy-Download” and read times

- Tier 1 jobs still using the Lazy-Download feature which streams larger chunks of data.
  - Applies to data streamed from both onsite and offsite
  - Interested in testing alternatives

“Lazy-Download” and read times

- Tier 1 jobs still using the Lazy-Download feature which streams larger chunks of data.
  - Applies to data streamed from both onsite and offsite
  - Interested in testing alternatives



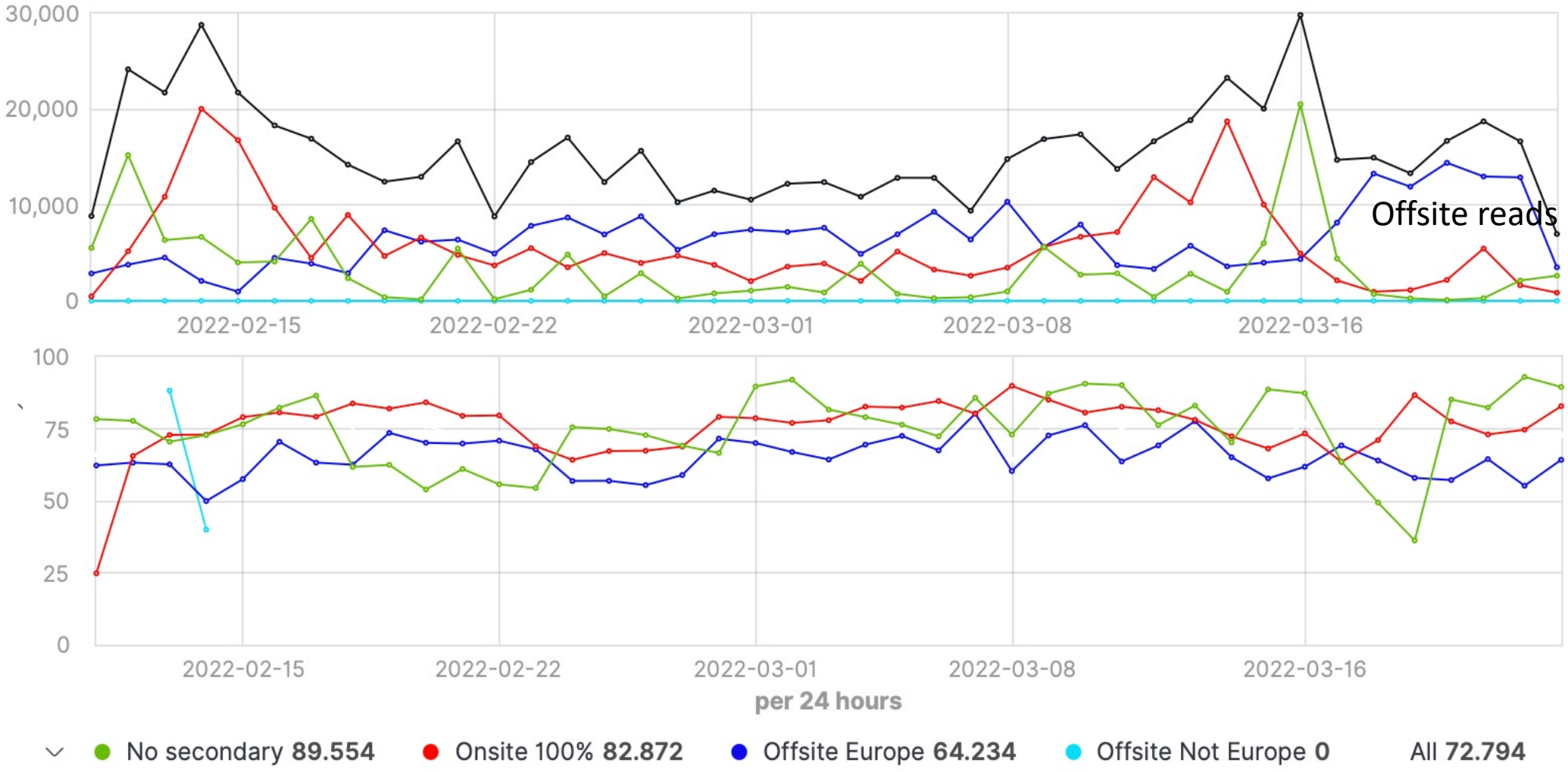


Onsite/offsite reads

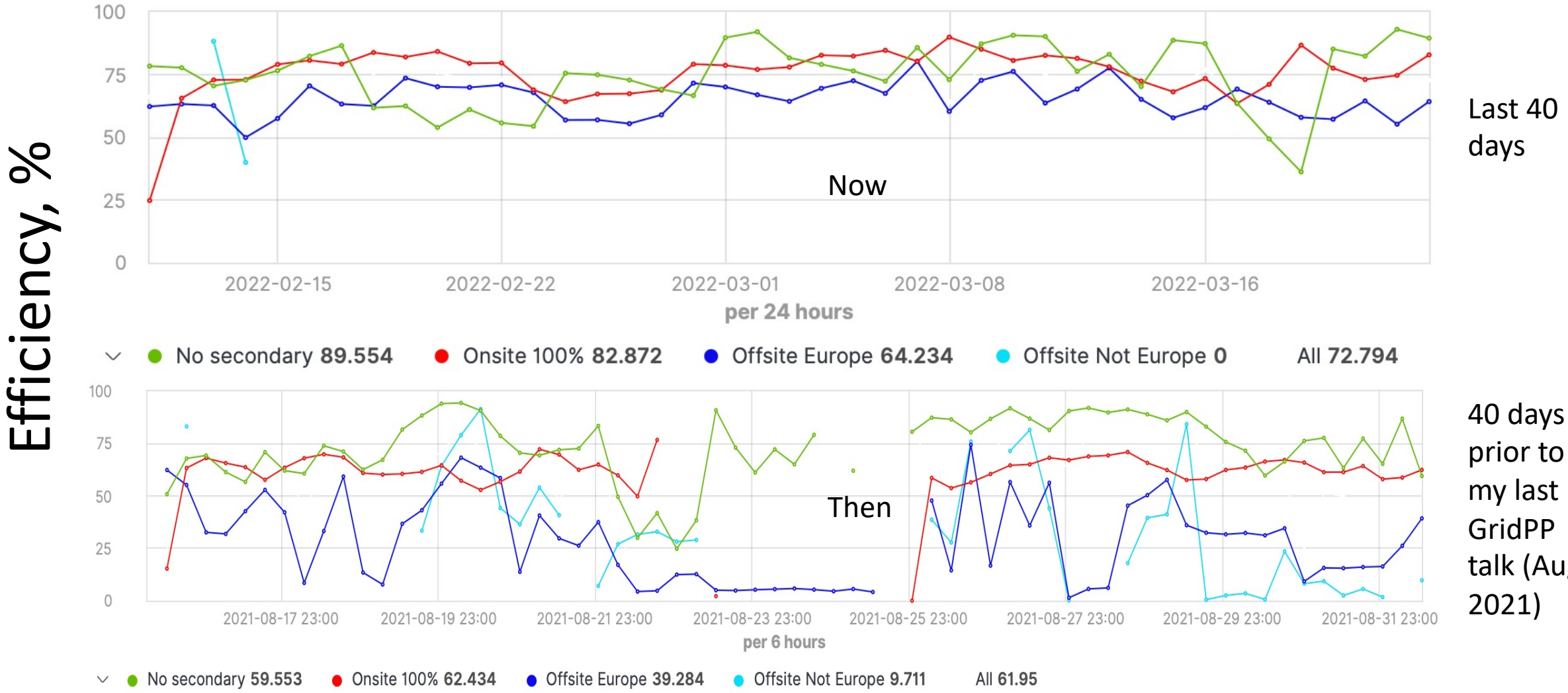
Secondary input data can be very large. These plots compare efficiency of jobs that run without secondary input files, with secondary input files onsite, and with secondary input files offsite.

Core count

Efficiency, %



Compare (offsite reading) job efficiency

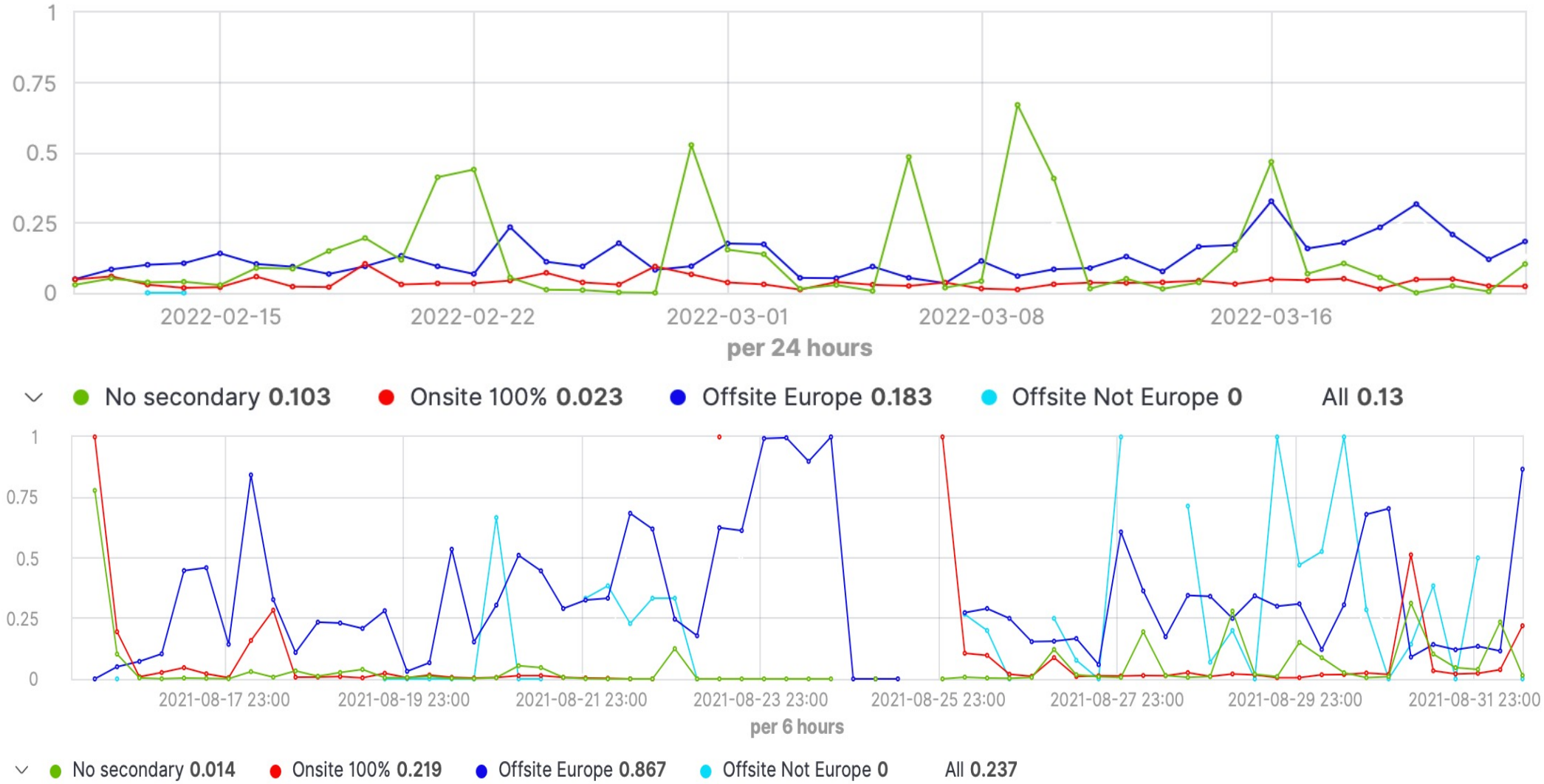


Last 40 days

40 days prior to my last GridPP talk (Aug 2021)

Job failures – now and then

Proportional failures



Last 40 days

40 days prior to my last GridPP talk (Aug 2021)

Conclusions

- Vast improvement in efficiency and failure rate of jobs pulling in data from offsite
  - Recent months put RAL much closer to the CMS T1 average efficiency
  - Further improvements anticipated
- Migration to Antares tape appears to have gone well
  - Testing continues
- Issues with performance of WebDAV transfers causing problems for site-to-site transfers with Echo disk
  - Potential to impact Antares too