GridPP53 collaboration meeting

Manchester, 09-11 April 2025

ATLAS Update

Sept 2024 to March 2025

Brij Kishor Jashal
Rutherford Appleton Laboratory
(On behalf of ATLAS UK Cloud team)

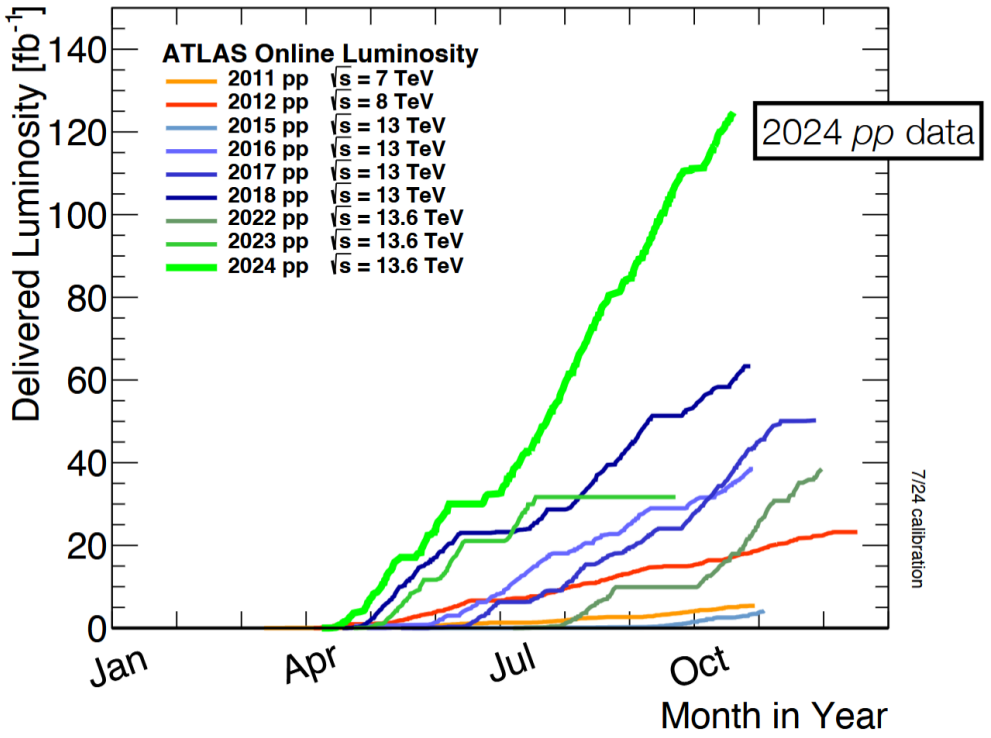
Overview:

- ATLAS UK Cloud performance since GridPP52 (01 Sep 2024 to 31 March 2025)
- Updates from ATLAS and ADC

ATLAS: Data Taking in 2024

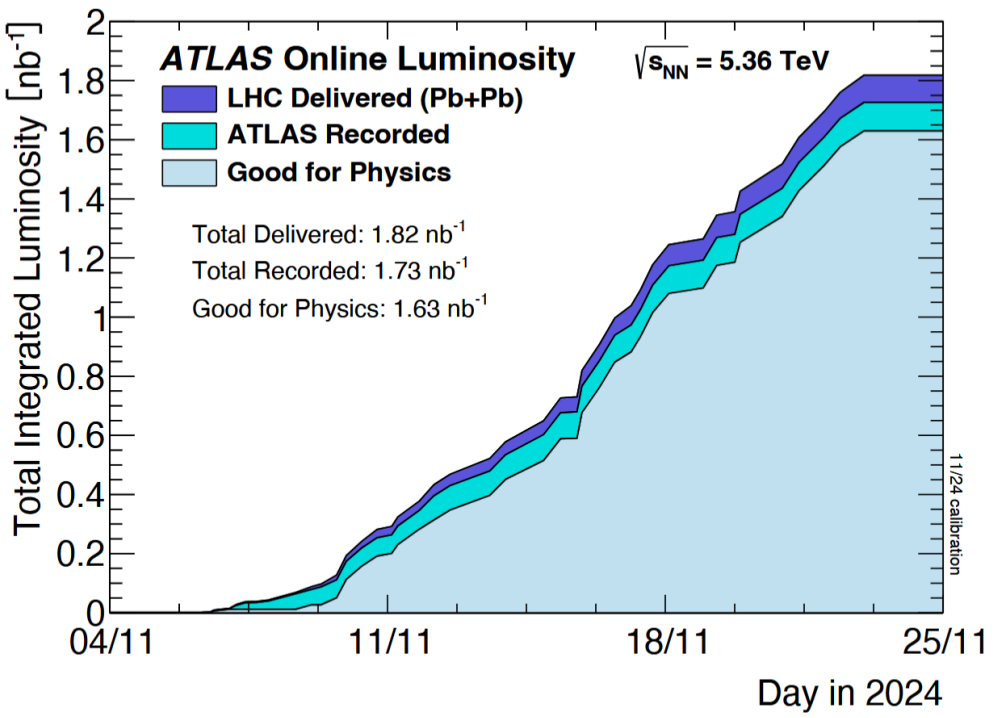
LHC and ATLAS luminosity above target — and smashed records in 2024

- Highest integrated luminosity achieved so far: 2x best previous year (2018)
- High recording efficiency (94%) despite of very high pileup conditions.
- Run 3 dataset, now at 195 (183) fb-1 delivered (recorded)



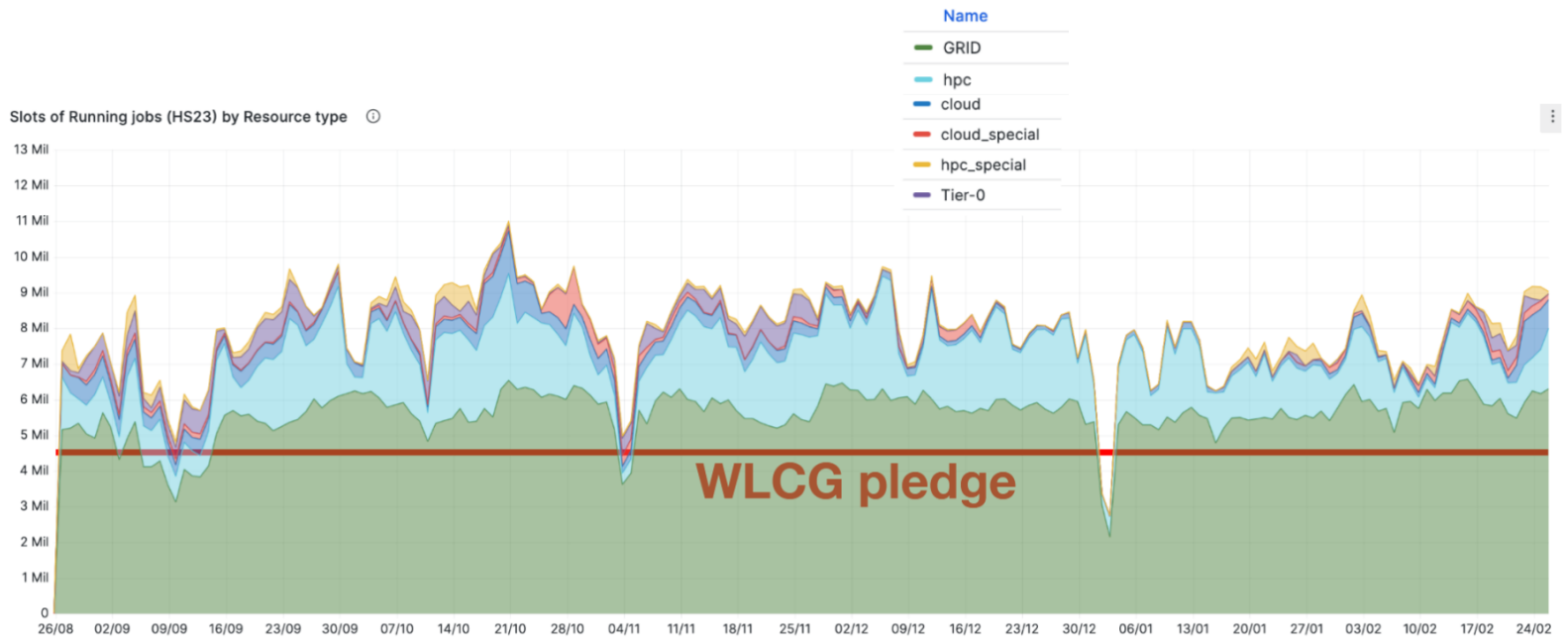
PbPb run:

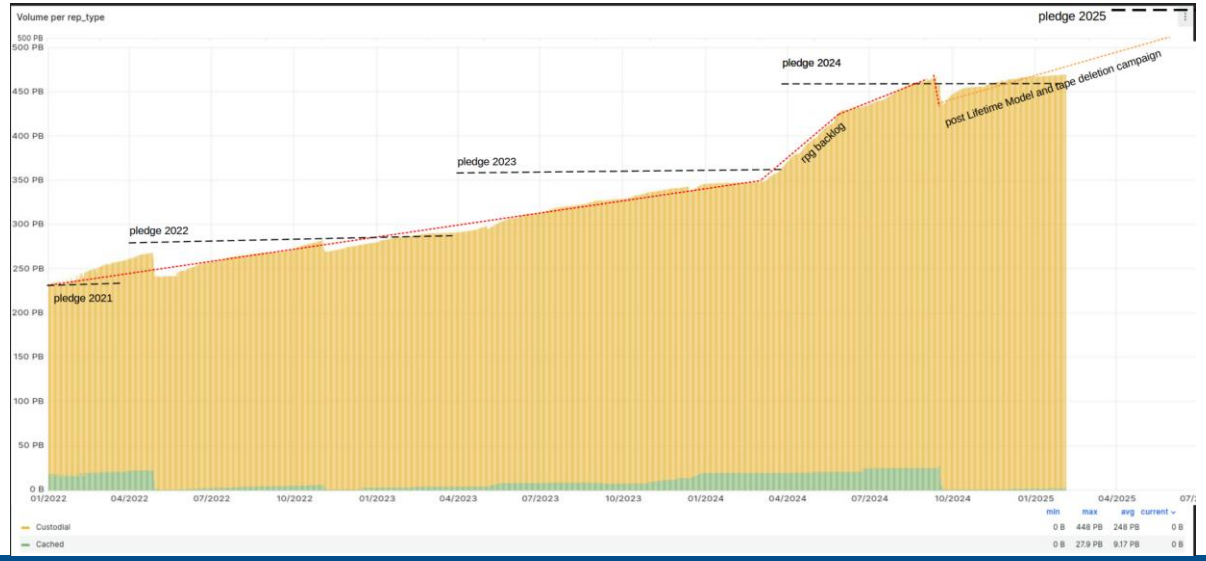
- Total delivered integrated luminosity of 1.82 nb-1, recorded 1.73 nb-1,
- Data quality efficiency for Pb-Pb data taking in 2024 was 97.7%

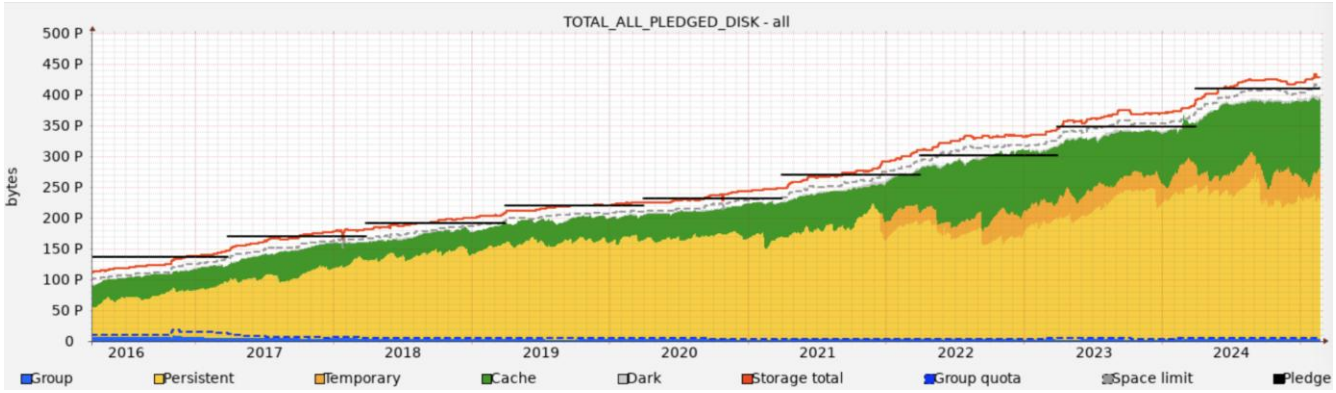


ATLAS ADC : 2025 readiness

Smooth computing operations
- WLCG sites continue to over deliver
- Regular deletions; largest persistent MC samples
- Ready for data taking this year with a RAW data size reduced from 1.8 to1.6 MB/event

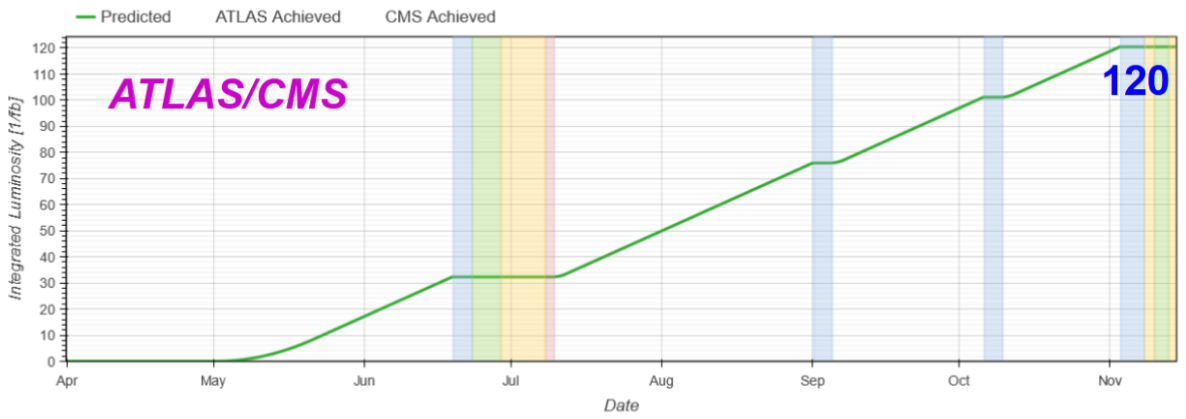


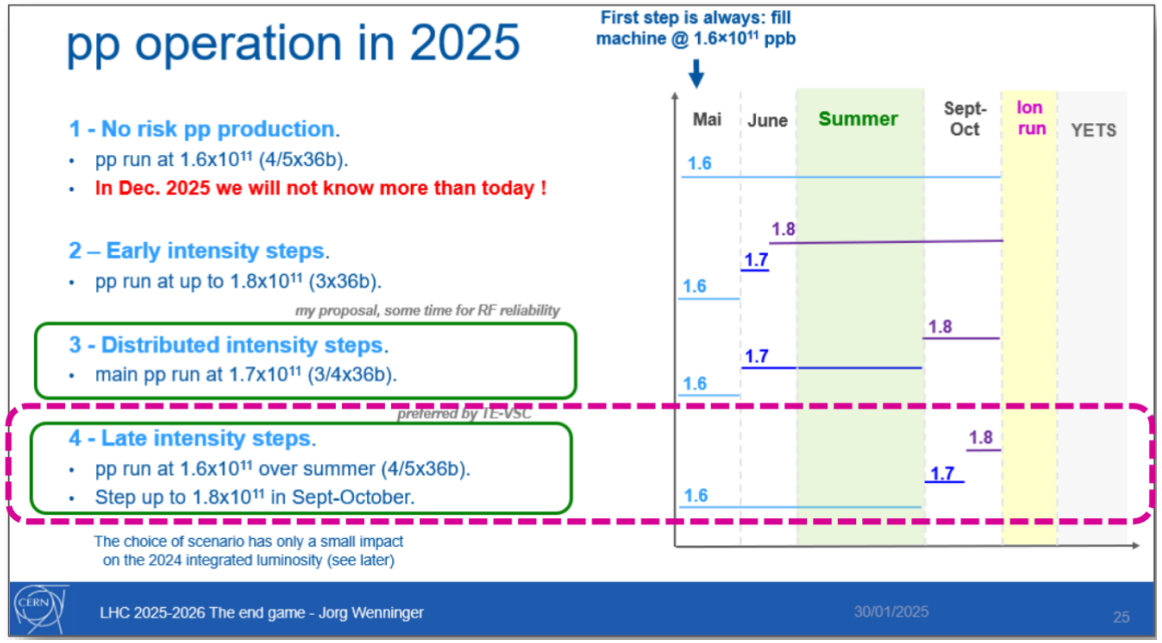




LHC Schedule 2025

- Experimental caverns closed on April 2nd
- First stable beams at 6.8 TeV in week of May 1st
- Intensity ramp up at 1.6×10^{11} p/b, full machine around mid-May.





Pledges for 2025

As compared to 2024 pledges,

- 16% increase in CPU
- 22% increase in Disk
- 32% increase in Tape (provisioned as needed)

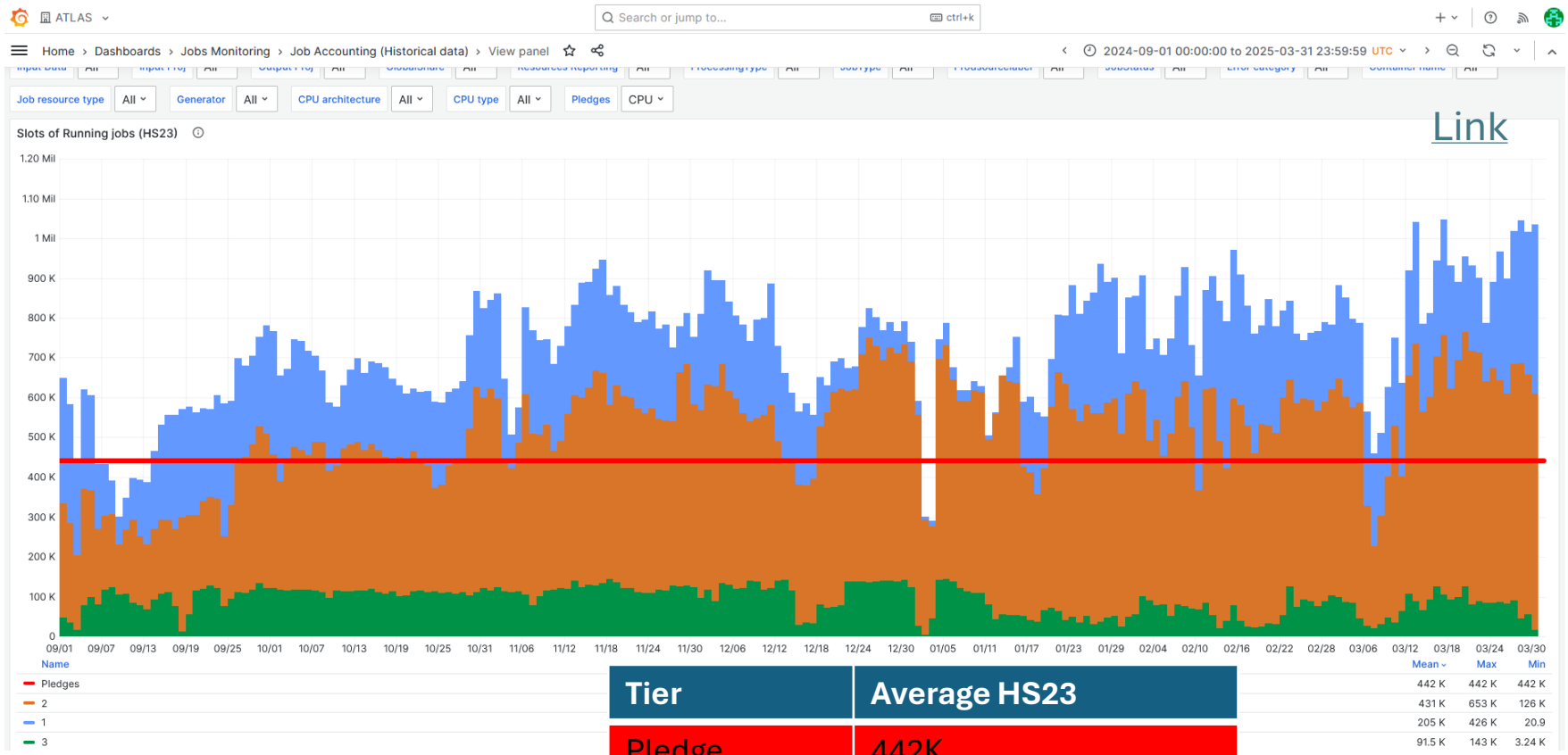
However, if we see the trend of previous years, we hope to continue to over deliver by 155% or more ☺

Edit	Federation	Tier	VO	Country	Year	Type	Pledge	Comment
	UK-T1-RAL	1	ATLAS	United Kingdom	2025	CPU	275186 HEPscore23	
	UK-T1-RAL	1	ATLAS	United Kingdom	2025	Disk	31306 TBytes	
	UK-T1-RAL	1	ATLAS	United Kingdom	2025	Tape	94422 TBytes	Tape provisioned as needed
	UK-London-Tier2	2	ATLAS	United Kingdom	2025	CPU	58934 HEPscore23	
	UK-London-Tier2	2	ATLAS	United Kingdom	2025	Disk	7442 TBytes	
	UK-NorthGrid	2	ATLAS	United Kingdom	2025	CPU	81967 HEPscore23	
	UK-NorthGrid	2	ATLAS	United Kingdom	2025	Disk	13118 TBytes	
	UK-ScotGrid	2	ATLAS	United Kingdom	2025	CPU	68234 HEPscore23	
	UK-ScotGrid	2	ATLAS	United Kingdom	2025	Disk	5534 TBytes	
	UK-SouthGrid	2	ATLAS	United Kingdom	2025	CPU	27807 HEPscore23	
	UK-SouthGrid	2	ATLAS	United Kingdom	2025	Disk	825 TBytes	

Compute: Occupancy vs pledge

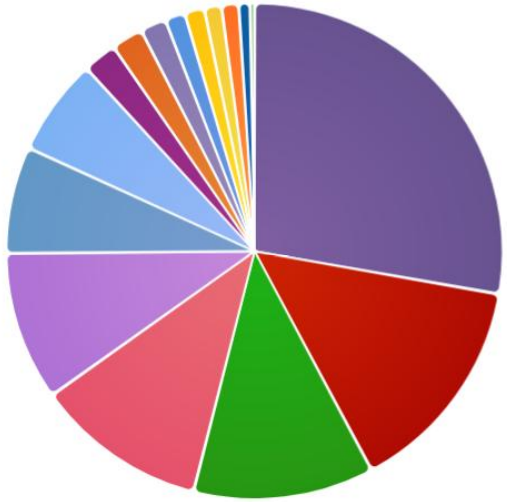
(01 Sep 2024 to 31 March 2025)

Delivered compute vs pledge at 164%



Tier	Average HS23
Pledge	442K
RAL Tier1	201K
Tier2s	431K
Tier3s	91.5K
Total delivered	723.5K

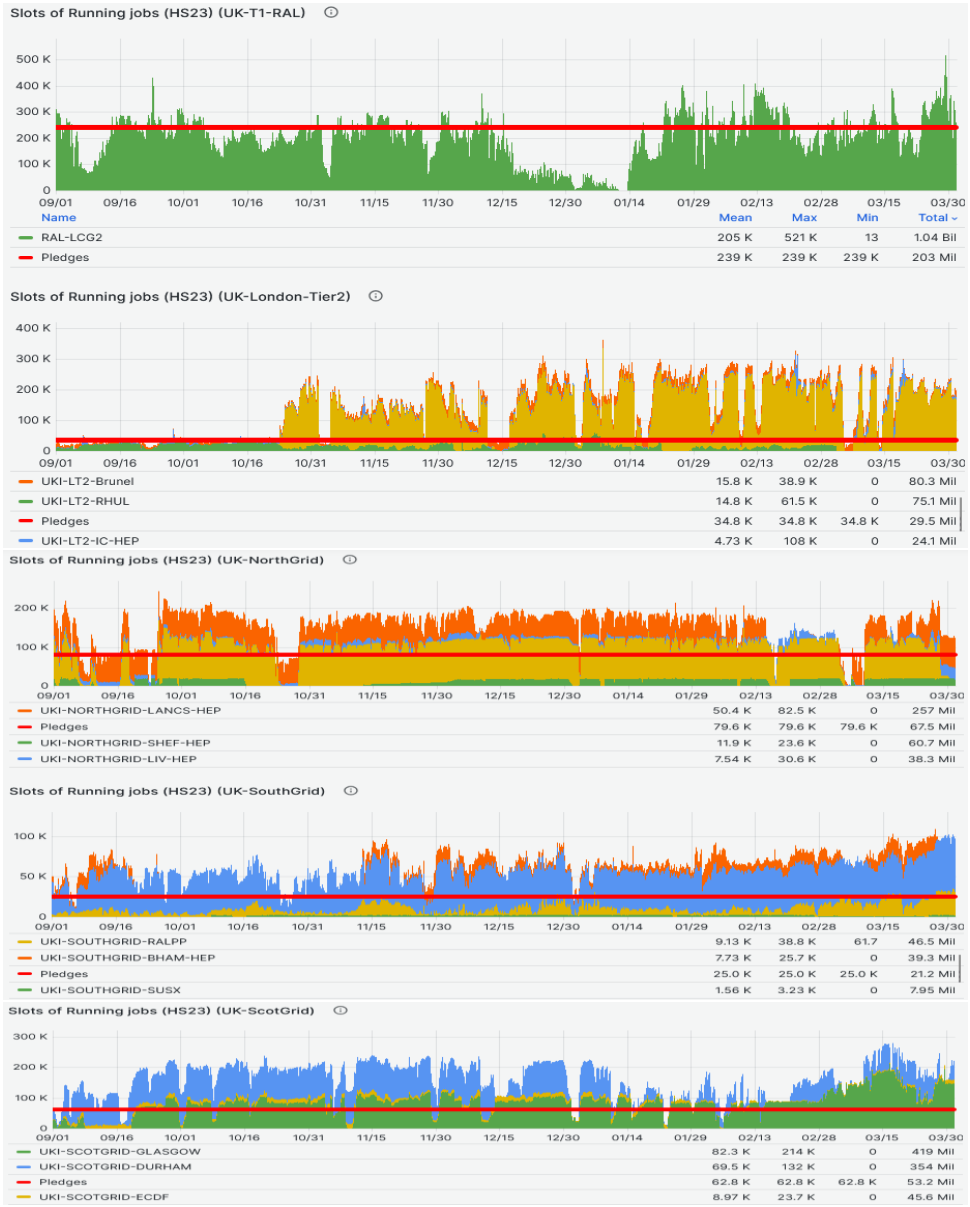
Wall clock time. All jobs (HS23 seconds)



	Value	Percent
RAL-LCG2	3.71 Tri	28%
UKI-LT2-QMUL	1.92 Tri	15%
UKI-NORTHGRID-MAN-HEP	1.57 Tri	12%
UKI-SCOTGRID-GLASGOW	1.49 Tri	11%
UKI-SCOTGRID-DURHAM	1.29 Tri	10%
UKI-NORTHGRID-LANCS-HEP	925 Bil	7%
UKI-SOUTHGRID-OX-HEP	840 Bil	6%
UKI-LT2-Brunel	286 Bil	2%
UKI-LT2-RHUL	266 Bil	2%
UKI-NORTHGRID-SHEF-HEP	216 Bil	2%
UKI-SCOTGRID-ECDF	165 Bil	1%

Compute: Occupancy vs pledge

Site (Federation)	Average HS23	Federation pledge HS23
RAL (Tier1)	**205K**	**239K**
UKI-LT2-QMUL	**108K**	**34.8K**
UKI-LT2-RHUL	**15.8K**	
UKI-LT2-IC-HEP	**14.8K**	
UKI-LT2-Brunel	**4.73K**	
UKI-NORTHGRID-MAN-HEP	**85.3K**	**79.6K**
UKI-NORTHGRID-LANCS-HEP	**50.4K**	
UKI-NORTHGRID-SHEF-HEP	**11.9K**	
UKI-NORTHGRID-LIV-HEP	**7.54K**	
UKI-SOUTHGRID-OX-HEP	**46k**	**25k**
UKI-SOUTHGRID-BHAM-HEP	**9.13k**	
UKI-SOUTHGRID-RALPP	**7.73k**	
UKI-SOUTHGRID-SUSX	**1.56k**	
UKI-SCOTGRID-DURHAM	**69.5k**	**62.8k**
UKI-SCOTGRID-GLASGOW	**82.3.1k**	
UKI-SCOTGRID-ECDF	**8.97k**	

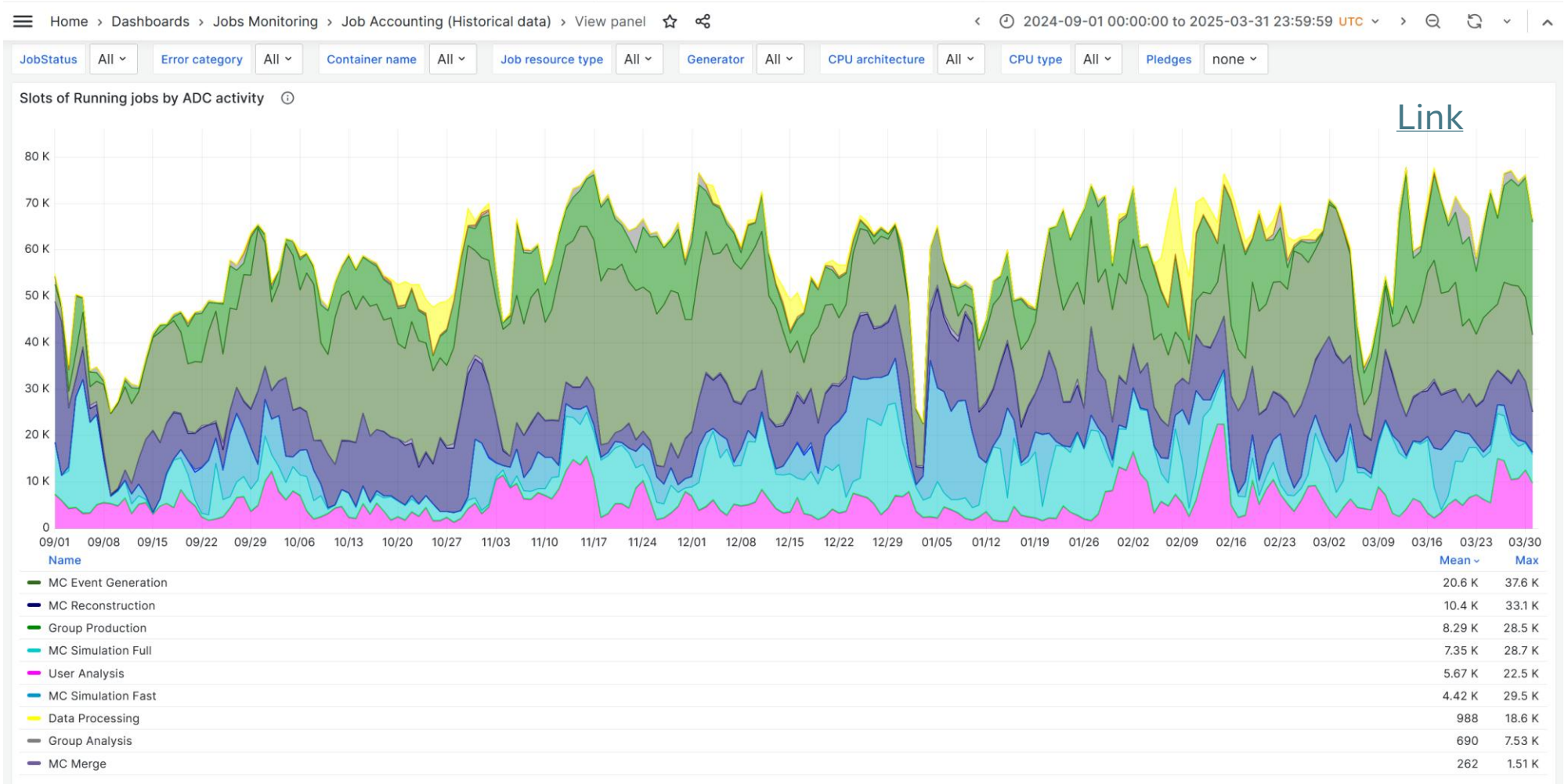


- CPU occupancy and job scheduling logic discussed in detail at, ATLAS S&C week talk Brij

- Actions taken thereafter have improved occupancy since Feb (107%).

Compute: workload type

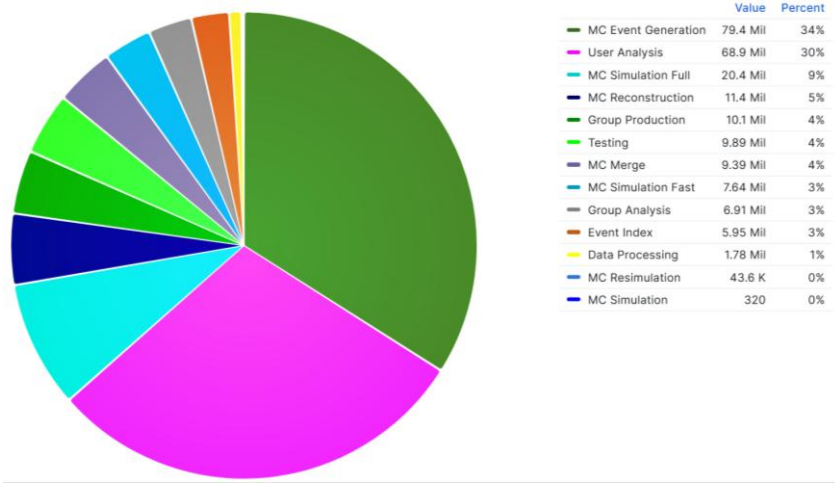
Slots of running jobs by ADC activity: Dominated by MC



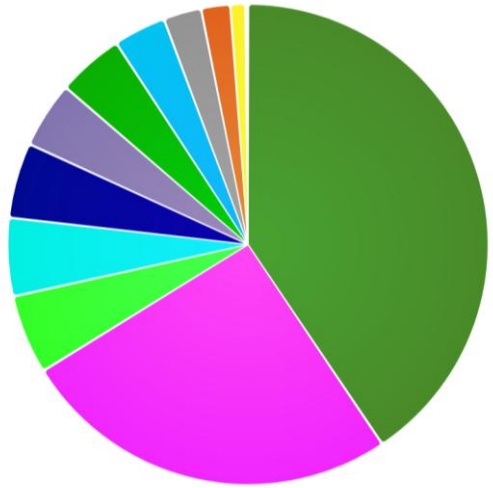
Compute: Efficiency global

Across ATLAS: Average CPU efficiency of good jobs for various job types at 92.3%

- As good as any other federation
- Similar mix of type of workflows in comparison to the global pool



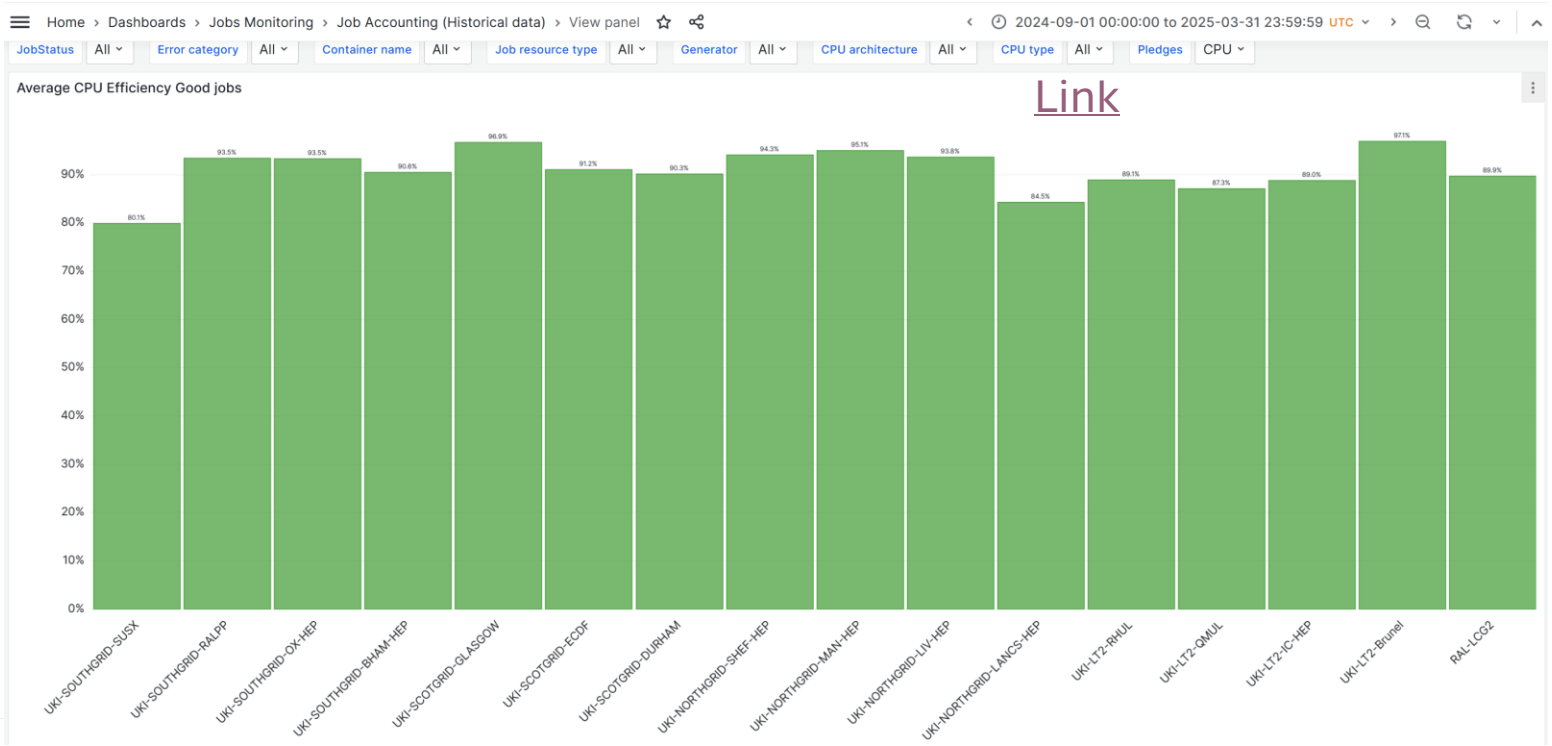
	Value	Percent
MC Event Generation	79.4 Mil	34%
User Analysis	68.9 Mil	30%
MC Simulation Full	20.4 Mil	9%
MC Reconstruction	11.4 Mil	5%
Group Production	10.1 Mil	4%
Testing	9.89 Mil	4%
MC Merge	9.39 Mil	4%
MC Simulation Fast	7.64 Mil	3%
Group Analysis	6.91 Mil	3%
Event Index	5.95 Mil	3%
Data Processing	1.78 Mil	1%
MC Resimulation	43.6 K	0%
MC Simulation	320	0%

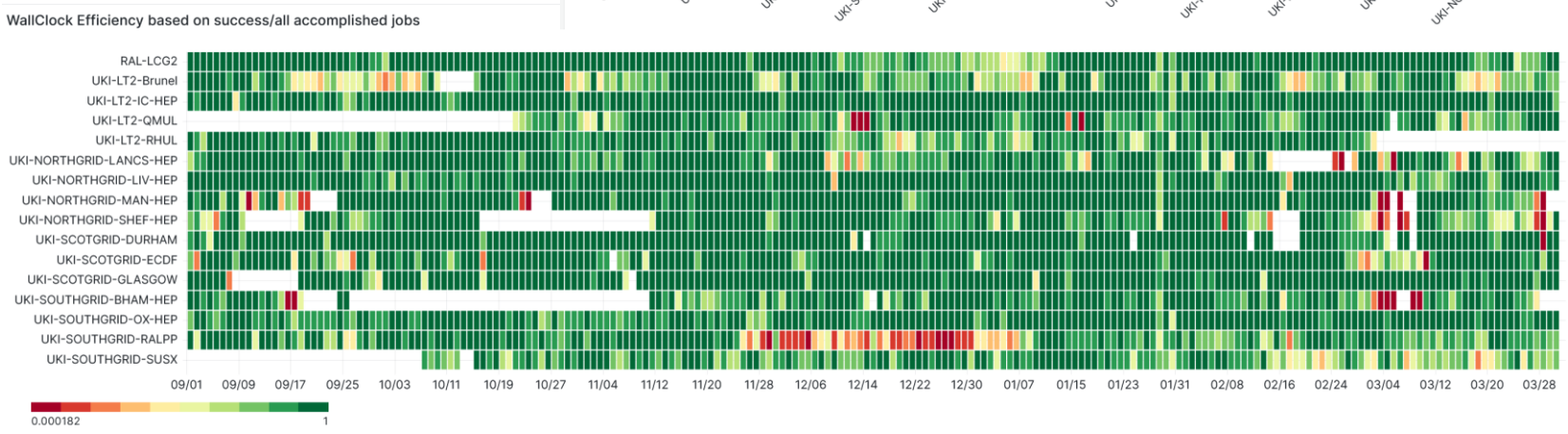


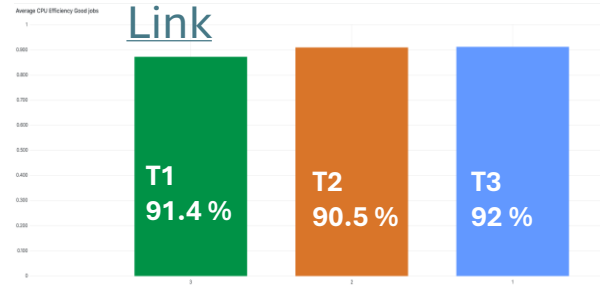
	Value	Percent
MC Event Generation	12.0 Mil	41%
User Analysis	7.55 Mil	26%
Testing	1.55 Mil	5%
MC Simulation Full	1.53 Mil	5%
MC Reconstruction	1.47 Mil	5%
MC Merge	1.30 Mil	4%
Group Production	1.29 Mil	4%
MC Simulation Fast	1.02 Mil	3%
Group Analysis	731 K	3%
Event Index	554 K	2%
Data Processing	273 K	1%
MC Resimulation	5.64 K	0%
MC Simulation	222	0%

Compute: Efficiency: UK sites

- Fairly stable operations and efficiencies across all UK federations and sites.



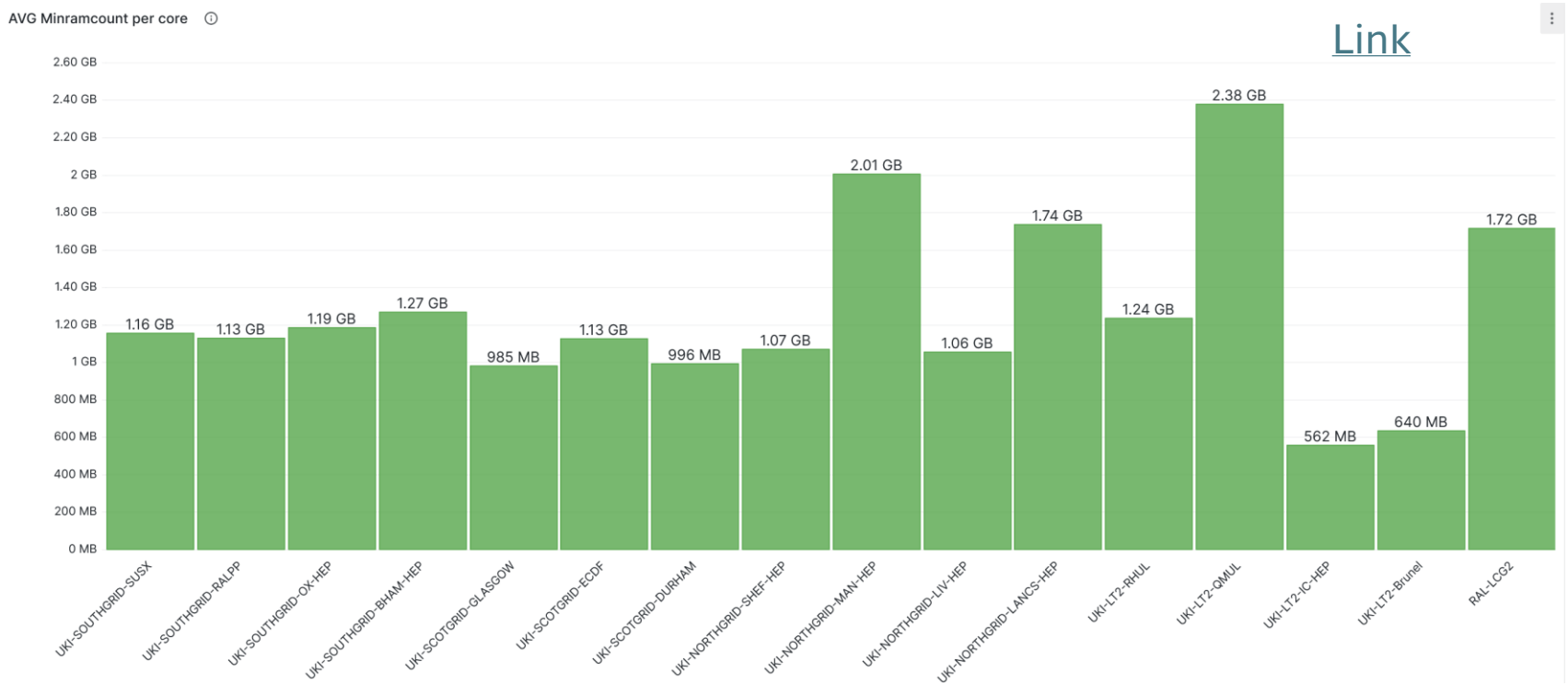




Improving memory utilization

Balance Low, High, VHIMEM to improve memory utilization

- MeanRSS is the mean amount of real memory available to job slot at a site: typically, 2-3 GB but varies dramatically between sites.
- MaxRSS is the maximum of mount of real memory available to a job slot at a site: typically, 6 to 8 GB
- Pilots and HTCondor are both using cgroups to kill jobs using more memory than they request

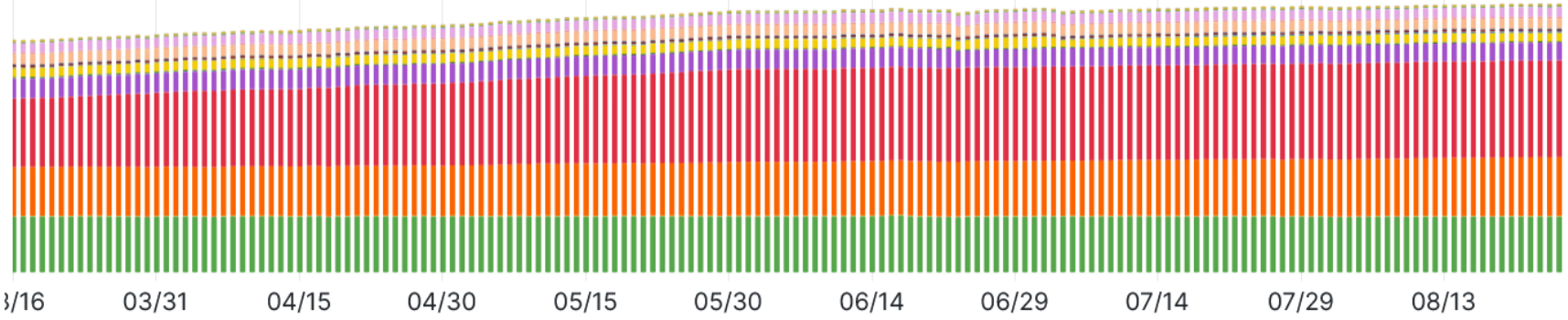


- Updates applied at CRIC for MeanRSS and MaxRSS PQ settings.
- If these numbers don't reflect your memory per core configuration, then please let me know.

Storage

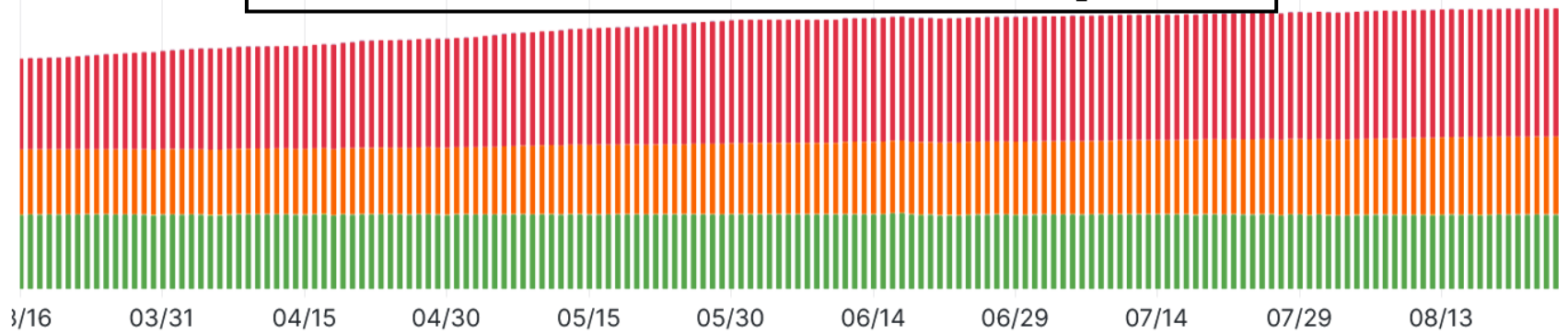
Volume per rse

ATLAS UK storage
~ 120.6 PB total (51.2 PB DISK + 69.4 PB Tape)



Volume per rse

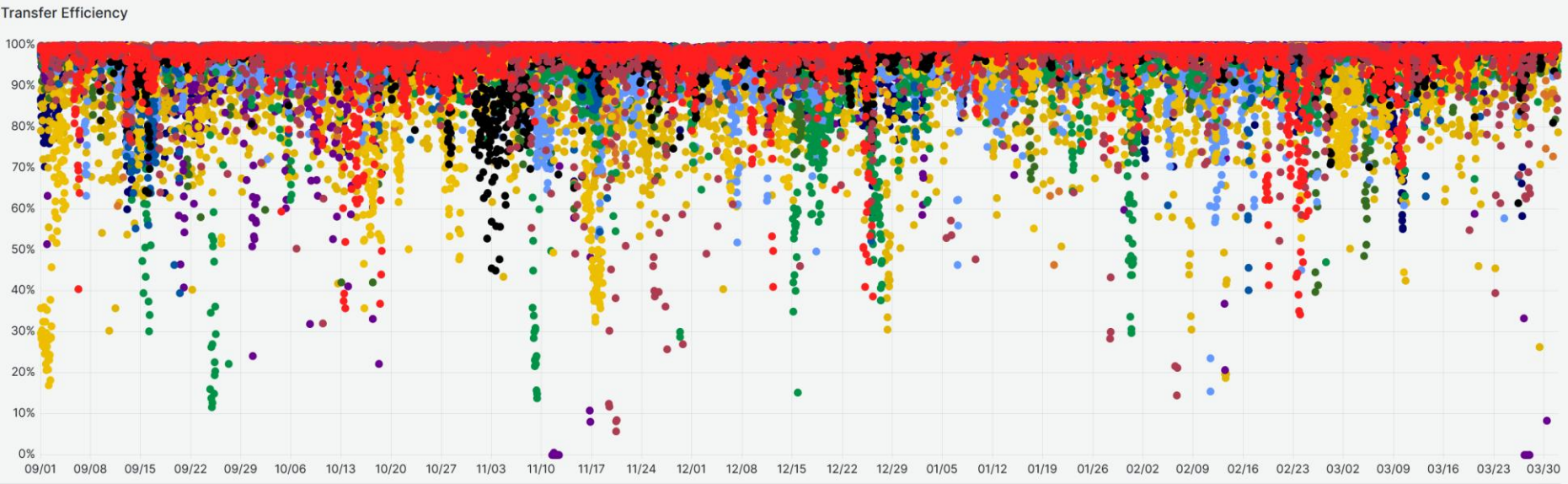
Tier-1 RAL-LCG2
~94.7 PB total (25.3 PB DISK + 69.4 PB Tape)

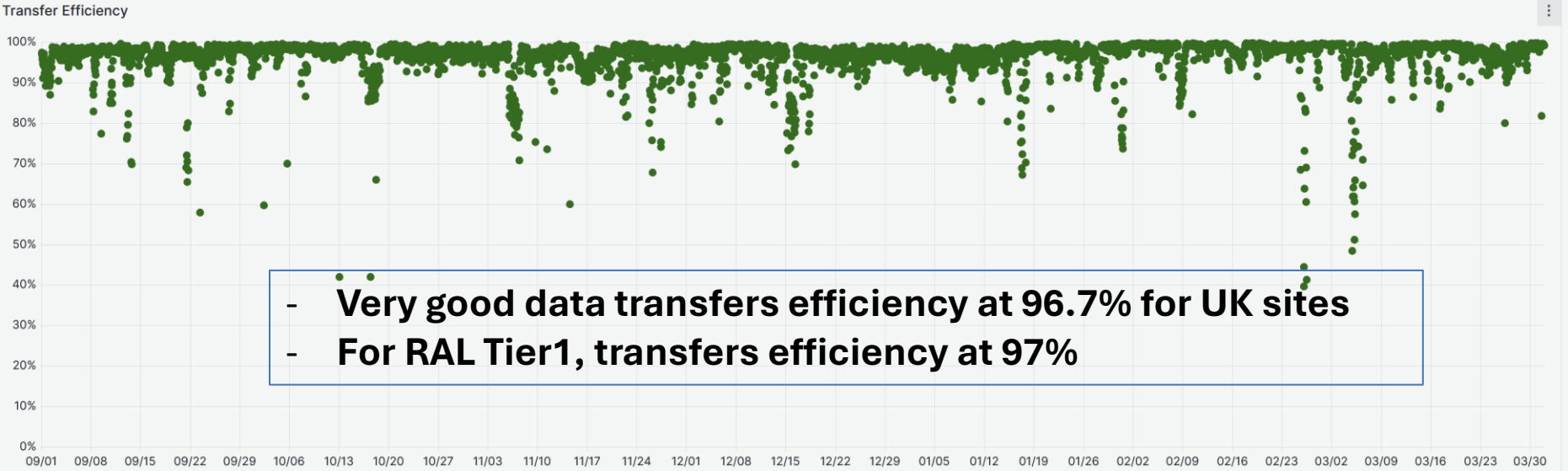


2024-08-25 00:00:00

RAL-LCG2_MCTAPE	43.1 PB
RAL-LCG2_DATATAPE	26.3 PB
RAL-LCG2-ECHO_DATADISK	25.1 PB
UKI-LT2-QMUL_DATADISK	8.05 PB
UKI-NORTHGRID-MAN-HEP_DATADISK	4.34 PB
UKI-SCOTGRID-GLASGOW-CEPH_DATADISK	4.20 PB
UKI-NORTHGRID-LANCS-HEP-CEPH_DATADISK	3.95 PB
UKI-NORTHGRID-MAN-HEP_LOCALGROUPDISK	730 TB
UKI-NORTHGRID-LANCS-HEP-CEPH_LOCALGROUPDISK	637 TB
UKI-NORTHGRID-LANCS-HEP-CEPH_SCRATCHDISK	636 TB
UKI-SOUTHGRID-RALPP_DATADISK	585 TB
UKI-SCOTGRID-GLASGOW-CEPH_SCRATCHDISK	546 TB
UKI-LT2-QMUL_SCRATCHDISK	447 TB
UKI-NORTHGRID-LIV-HEP_DATADISK	401 TB
RAL-LCG2-ECHO_SCRATCHDISK	199 TB
UKI-NORTHGRID-LANCS-HEP-CEPH_PHYS-BEAUTY	179 TB
UKI-SCOTGRID-GLASGOW-CEPH_LOCALGROUPDISK	139 TB
UKI-SOUTHGRID-RALPP_LOCALGROUPDISK	**130 TB**
UKI-NORTHGRID-MAN-HEP-CEPH_DATADISK	97.4 TB
UKI-NORTHGRID-LIV-HEP_SCRATCHDISK	24.9 TB
UKI-NORTHGRID-MAN-HEP_SCRATCHDISK	14.9 TB
UKI-LT2-QMUL_LOCALGROUPDISK	13.3 TB
UKI-SOUTHGRID-RALPP_SCRATCHDISK	12.7 TB
RAL-LCG2-ECHO_TEST	3.50 TB
UKI-NORTHGRID-MAN-HEP-CEPH_SCRATCHDISK	1.19 TB
UKI-NORTHGRID-LIV-HEP_LOCALGROUPDISK	120 GB
UKI-NORTHGRID-MAN-HEP-CEPH_LOCALGROUPDISK	94.8 MB
RAL-LCG2_TAPE-TEST	0 B

Transfers performance





- **Very good data transfers efficiency at 96.7% for UK sites**
- **For RAL Tier1, transfers efficiency at 97%**

Name	Mean
NL	98.0%
FR	97.7%
RU	97.4%
CERN	97.4%
CA	96.9%
UK	96.7%
DE	96.7%
US	96.6%
IT	94.8%

Networking outlook

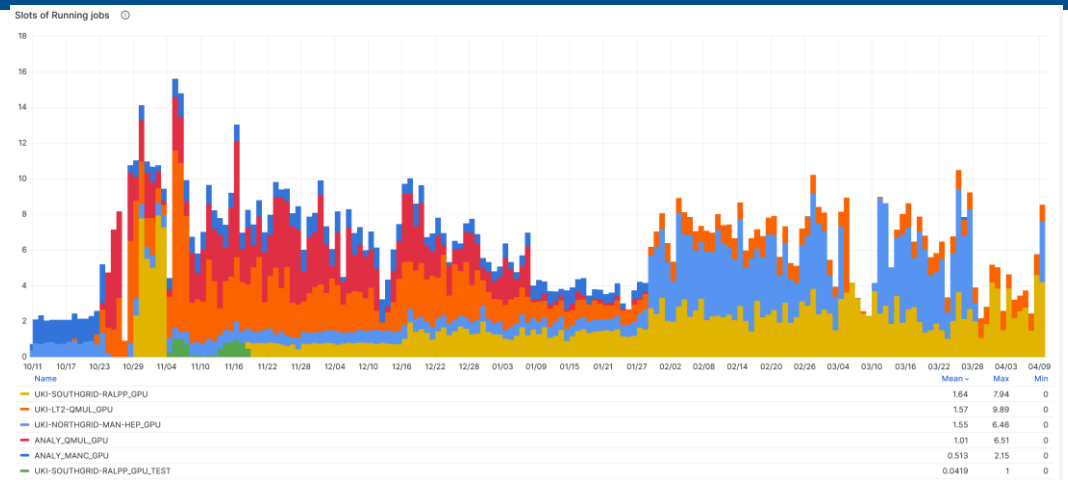
- Is this the correct picture as of today ?
- Depending on the need, preparing to scale up bandwidth by 2027 to:
  - Tier-1: 400 Gbit/s (dedicated)
  - Tier-2 (large): 100 Gbit/s
  - Tier-2 (other): 20 Gbit/s (and in some cases 40 Gbit/s)
- Scientific Network Tags (scitags) is an initiative promoting identification of the science domains and their high-level activities at the network level.
- ATLAS
  - Enabled at CERN EOS ATLAS production since January and RAL xrootd gws since last week.
  - An ATLAS campaign is pending support in dCache

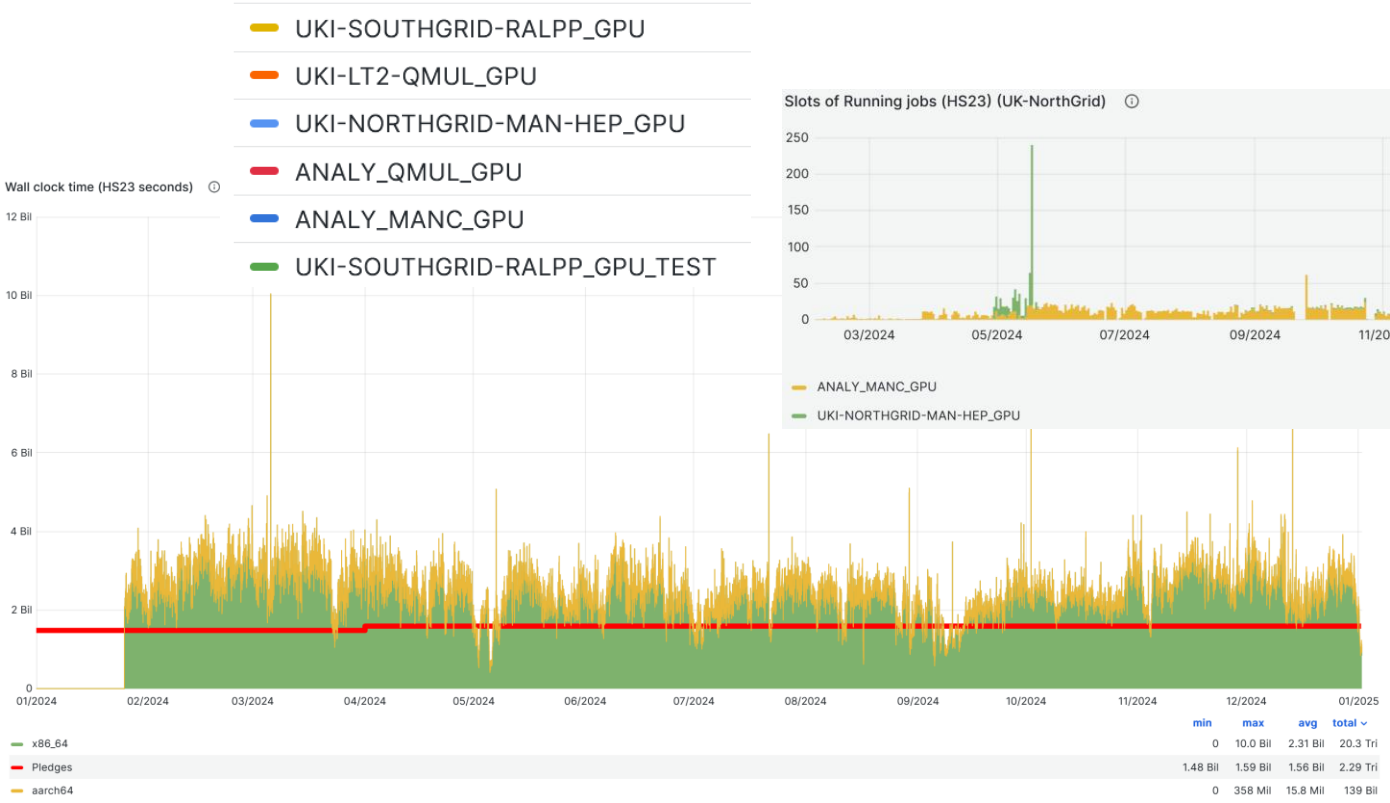
Site	WAN link
UKI-RAL-LCG2	2 x 100 Gbit/s LHCOPN links, 4 x 100 Gbit/s WAN + LHCONE (Shared)
UKI-LT2-IC-HEP	100 Gbit/s shared
UKI-LT2-QMUL	100 Gbit/s dedicated
UKI-NORTHGRID-LANCS-HEP	100 Gbit/s dedicated
UKI-NORTHGRID-MAN-HEP	40 Gbit/s dedicated
UKI-SCOTGRID-GLASGOW	2 × 20 Gbit/s
UKI-LT2-Brunel	4x10 Gbps
UKI-LT2-RHUL	10 Gbit/s dedicated
UKI-NORTHGRID-LIV-HEP	2 x 10Gbps
UKI-NORTHGRID-SHEF-HEP	10 Gbit/s
UKI-SCOTGRID-DURHAM	4 x 10 Gbit/s
UKI-SCOTGRID-ECDF	2 x 25 Gbit/s shared
UKI-SOUTHGRID-BHAM	2 x 10 Gbit/s
UKI-SOUTHGRID-BRISTOL	3 x 10 Gbit/s
UKI-SOUTHGRID-OX-HEP	2 x 10 Gbit/s
UKI-SOUTHGRID-RALPPD	2x100 Gbit/s (shared with RAL-LCG2)

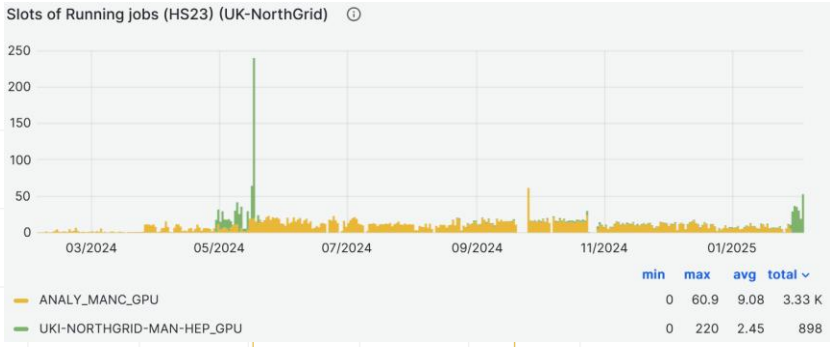
Non x86 computing resource.

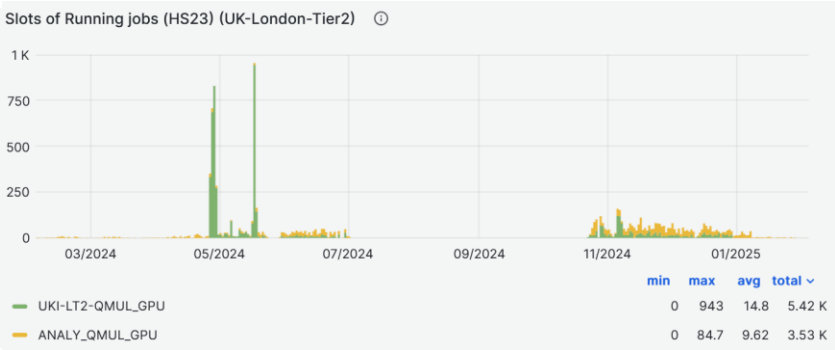
Significant number of non-x86 resources:

- Production ARM queues at Glasgow (~ 4K cores)
- New test ARM queue at Tier-1, ATLAS tests smooth.
- 5 GPU queues at
- Non Grid GPUs for local users at RALPP, Manchester, Glasgow, QMUL [list] (more in Jyoti's talk)









Updates from ATLAS

- **ATLAS Documentation Weeks** (started in 2024) with some very fruitful results.
  - Documentation report about documentation report
  - PanDA and Rucio documentation
  - Documentation service now hosted by CERN IT based on mkdocs-material (replacement of Twiki)
    - https://atlas-computing.docs.cern.ch/
    - https://gitlab.cern.ch/atlas/computing-docs/atlas-computing-docs

- **Security incidents** affecting ATLAS MUST be sent to atlas-csirt@cern.ch ONLY
  - “Cyber-Security Incident Response Team”
  - Experiment security liaison will follow-up on a need-to-know basis
  - Do not send to public mailing lists
  - Mention to CSIRT any threads/follow-ups with your internal cyber security programmes

- **PanDA Analysis Data Carousel**
  - Purpose: Optimize data staging for analysis tasks using DC machinery.
  - New components: DATA_CAROUSEL_REQUESTS and RELATIONS tables.
  - Integrated DC in PanDA for prioritized tape data staging.
  - Challenge is expanding the Data Carousel for production use and efficient tape management.

Updates from ATLAS: ATLAS Analysis Infrastructure

Discussion on Analysis facilities later.

- LHCC initiated a review on analysis facilities; ATLAS and CMS collaborated on responses.
- Focus on scaling infrastructure for HL-LHC needs while maintaining operational efficiency.

Current Run 3 Analysis Model:

Data Formats: AOD, DAOD_PHYS, DAOD_PHYSLITE, nTuples.

- Expected storage: ~120 PB in Run 3.
- Adoption: 80% of analyses rely on DAOD_PHYSLITE.

Workflows:

- Include reconstruction, calibration, event filtering and ML workflows

Resource Usage:

- ~10% of CPU cycles used for analysis workflows in ATLAS
- Majority of usage is via SSH and batch but Interactive usage accounts for ~10% of compute resources. (and growing)

Updates from ATLAS: ATLAS Analysis Infrastructure

Challenges and Innovations for Run 4 and Beyond

Challenges:

- Data increase O(10x) data increase will strain current patterns.
- Increase in MC data to achieve physics goals.
- Anticipated volume: 7-10 EB by Run 5.

Evolving Infrastructure and Analysis Facility R&D

- Adoption of efficient data formats (e.g., RNTuple).
- Enhance caching infrastructure for better interactive analysis.
- Automated tools for seamless batch and interactive processing.
- Transition to Kubernetes and cloud-native platforms for flexible resource orchestration.
- Expand Kubernetes-based AFs to include Jupyter and interactive analysis platforms.
- On-demand provisioning capabilities accelerators (GPUs, FPGAs) in interactive environment.
- Burst workloads from AFs to Tier facilities resources as needed.

- Compute resource delivery reached 164%, supported by stable storage and network operations, with the ATLAS UK Cloud's performance significantly exceeding pledged targets.
- Challenges faced and major upgrades carried out by ATLAS cloud sites not covered in this talk but continuous enhancements in infrastructure, software, and services across multiple sites are ensuring smooth and reliable resource delivery to WLCG and ATLAS.

Thank you