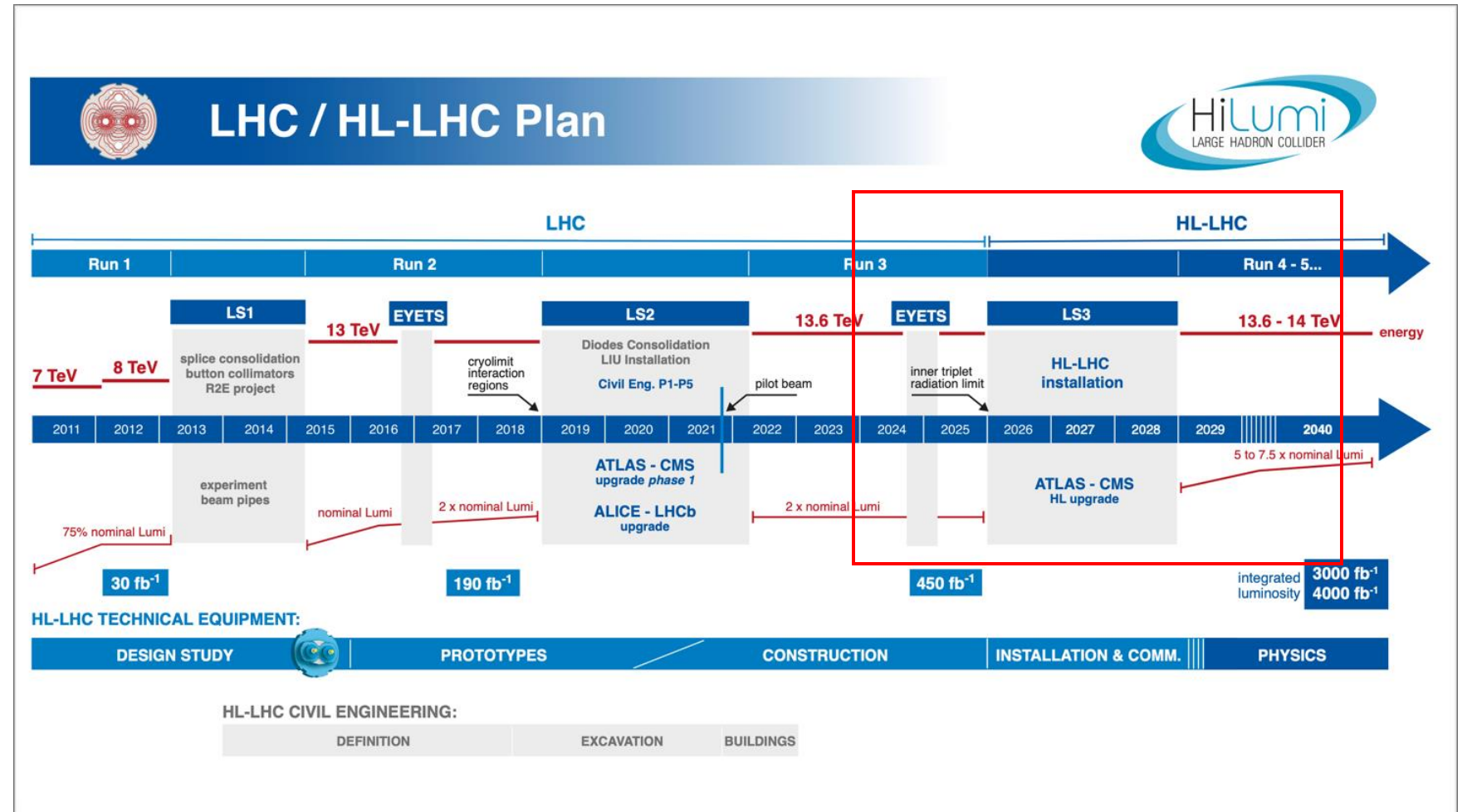


Network Status and Forward Look

Duncan Rand

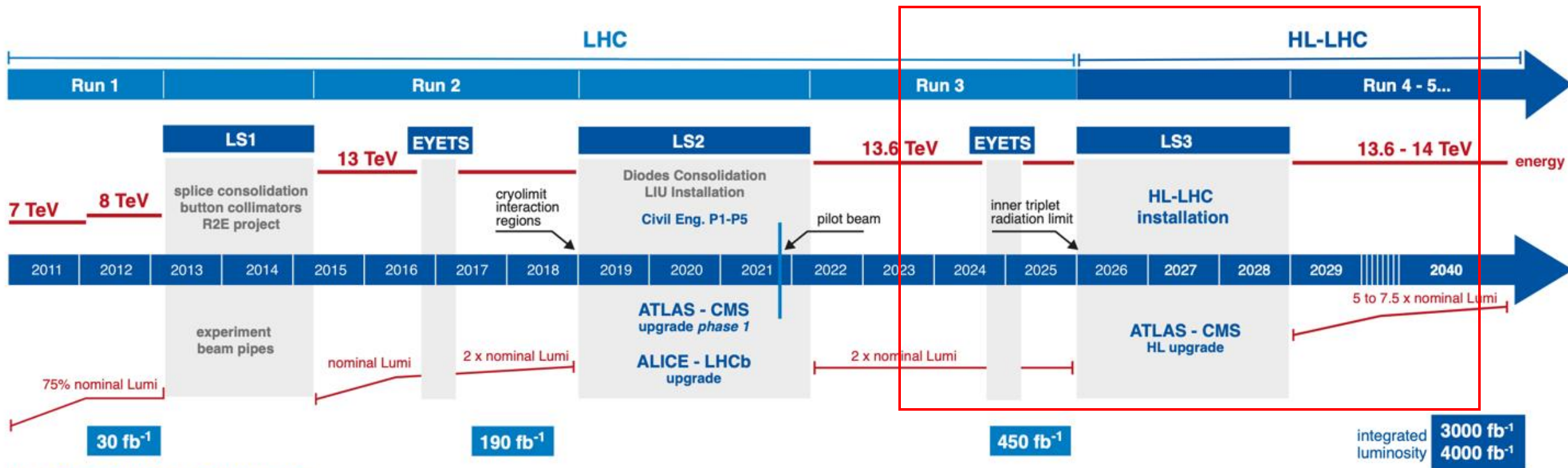
The LHC Schedule

- The LHC is currently in Run 3 of the LHC era which is scheduled to run through 2024 and 2025 and will finish at the end of 2025.
- The upgrade to HL-LHC starts with Long Shutdown 3 (LS3) at the beginning of 2026, scheduled to last for three years.
- The HL-LHC Run 4 is scheduled to commence at the start of 2029.





LHC / HL-LHC Plan



HL-LHC TECHNICAL EQUIPMENT:



HL-LHC CIVIL ENGINEERING:



ATLAS experiment requirements

- During the remainder of LHC Run-3 and beginning of LS3 ATLAS expects a growth of processing and analysis capacity of a factor of ~ 2 following the long-term trend of a 15-20% increase of resources per annum for both storage and CPU.
- The architecture of the distributed system is changing, with an increased concentration of storage at fewer sites and the emergence of diskless sites with larger sites needing to serve remote CPUs.
- Consequently, there will be an additional increase in the bandwidth requirements.
- The overall computing model in terms of formats, versions and selections was revised for Run-3 with the creation of a new type of “derived AOD” (DAOD) dataset type
- User analyses will access these DAODs directly over the network, which needs to be able to handle a fast analysis turnaround time.

ATLAS experiment requirements

- This means that on the 4 year timescale, that the large (“nucleus”) ATLAS sites (QMUL, Manchester, Lancaster and Glasgow) will need at least ~80 Gbit/s for the Tier-2 traffic alone (and hence a greater capacity at the campus site overall).
- It is possible that some of these sites may require 100 Gbit/s on this timescale.
- The UK needs to monitor the rate of growth of the network bandwidth at the larger ATLAS sites.
- For the smaller lower storage (“satellite”) sites, operational considerations require the site to have a demonstrated connection to a nucleus site of greater than 1.6 Gbit/s.
- In the UK, the satellite sites are actually relatively large and should be able to provide connections to ~4 UK nucleus sites
- Therefore, a bandwidth of 4-8 Gbit/s at the smaller sites should be sufficient with headroom for other traffic

CMS experiment requirements

- The current network provision was sufficient for CMS at its Tier 1 and Tier 2 sites in 2023, and requirements are not expected to change significantly during the remainder of Run 3.
- However, we should note that in both 2022 and 2023 the amount of experimental data produced was below expectations and we anticipate significantly more in 2024.
- CMS transfer throughput to the UK in the last year was dominated by jobs reading data directly from CERN, mostly to RAL Tier 1 and using the AAA system.
- CMS stream large volumes of ‘premix’ libraries from CERN to sites running production jobs
- Improved XRootD monitoring should soon hopefully give us a much more accurate idea of how much data is being streamed via AAA, worldwide.
- Although people often like to say that CMS is using more network bandwidth for this method, the data has to get to the site running the jobs somehow, whether streamed via AAA or in a TPC transfer via FTS.
- Jobs using AAA only read the parts of the file needed.

Scale of HL-LHC requirements and WLCG data challenges

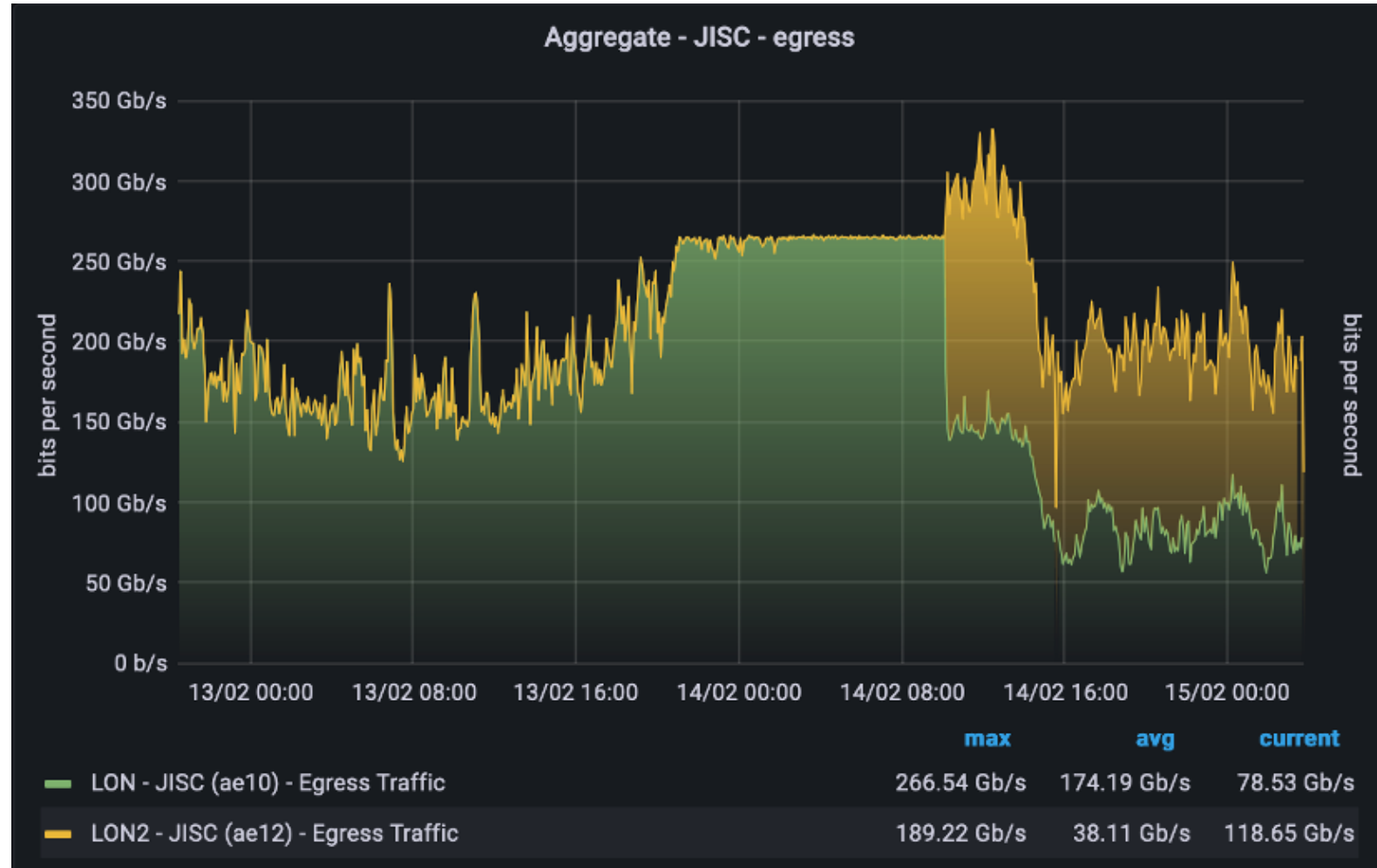
- The original plan was for the 100% rate during the last DC to be half the minimal scenario envisaged for HL-LHC startup (~2029)
- But minimal scenario now abandoned
- The first DC ran in 2021 with a target of 10% of the DC28 rate and in early 2024 a second challenge was run at 25% of the DC28 rate
- LHC experiments calculated the ingress and egress rates for Tier-1 and Tier-2 sites - shown together with extrapolations for the upcoming DC in 2026 and 2028.
- The rate displayed for the Tier-1 are the threshold goals set for DC24, i.e. 85 Gbit/s and 170 Gbit/s
- Can we naively use these to anticipate DC requirements until 2028?
- Core Tier-2s okay with 100G up until ~2026 but might have to move to 200G after?

	DC24 (25%)		DC26 (50%)		DC28 (100%)	
Site	Ingress	Egress	Ingress	Egress	Ingress	Egress
RAL-LCG2 (min)	85	85	170	170	340	340
RAL-LCG2 (flex)	170	170	340	340	680	680
Imperial	24	7	49	14	97	29
QMUL	36	23	72	47	144	93
Glasgow	14	12	28	23	56	46
Lancaster	25	29	49	57	98	115
Manchester	36	23	72	47	144	93
RAL-PPD	10	6	20	12	40	24
Brunel	4	1	8	1	15	3
Bristol	3	1	7	1	13	2

Site ingress and egress rates for data challenges (Gbit/s). Note, QMUL did not take part in DC24, but is a similar size to Manchester and thus the figures for Manchester have been used.

The GÉANT-Janet link during Data Challenge 24

- An issue of concern for Jisc prior to DC24 was the possibility of filling the 300G link between the GÉANT and Janet networks
- This consists of two links - a primary connection *Lon* and a secondary reserve connection *Lon2* - carry general IP and LHCONE traffic, but not CERN –RAL T1
- However, during the first part of the DC24 the OPN was inoperable due to a damaged cable under the North Sea with the consequence that traffic which usually flows over the OPN was rerouted onto the LHCONE and hence also used the GÉANT-Janet links.
- GÉANT to Jisc peering hit its capacity of ~265 Gbit/s and remained fully loaded from 9pm on 13th Feb until 10am on 14th Feb when the general IP traffic was moved away to the GÉANT-Janet backup link, leaving the main link for just LHCONE
- The cap at 265 Gbit/s was a GÉANT-side configuration issue
- While one of the three 100G links comprising the aggregate did hit 100G, the other two between them could not exceed a cap of 165G



Tier-1 Evolution

- Tier-1 migrating from legacy network to new network
- First half of GridPP7 (2024 – 2025):
 - The RAL site link to Janet will be upgraded to 4 x 100 Gbit/s (resilient)
 - Tier-1 batch farm will be made dual stack
 - EOS frontend to Antares will be made dual stack and put on both the LHCOPN and LHCONE
- Second half of GridPP7 (2026 – 2027):
 - The current Tier-1 spine will be replaced after 5 - 6 years in production with 400 Gbit/s capable switches.
 - We will upgrade the LHCOPN link to 400 Gbit/s once this falls to the same level as the 2 x 100 Gbit/s links.
- RAL site link will be monitored and upgraded when necessary (up to 800 Gbit/s), expected to be in 2027 - 2028

Expected requirements

- We expect that the CERN Tier-0 to RAL Tier-1 network requirement will rise to approximately 400 Gbit/s by 2026
- The Tier-1 connection to Janet is in the process of being upgraded to a 400 Gbit/s resilient link. This should be adequate for the remainder of Run-3
- In the next 2 years, the “large” Tier-2 sites (Glasgow, Imperial, Lancaster, Manchester, QMUL, RAL) are likely to need up to 100 Gbit/s for all LHC traffic and the relevant connections to Janet are already or being upgraded.
- We expect the “other” Tier-2 sites to require 20 Gbit/s for LHC traffic, though there are some “medium” sites that currently have a high network I/O rate and a connection bandwidth similar to the larger sites.

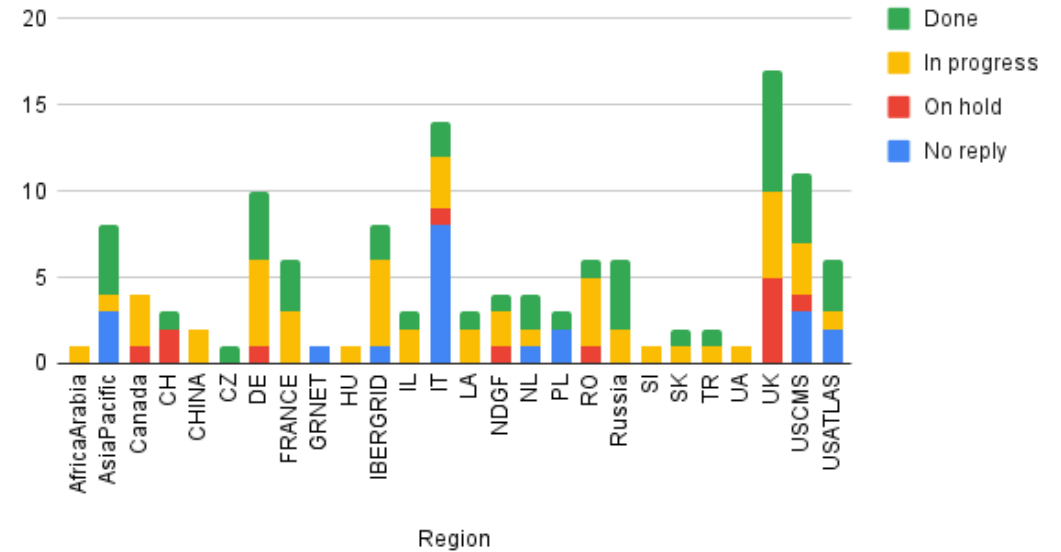
GridPP site WAN connectivity

- GridPP sites are connected to Janet network at a variety of different capacities and architectures
- 400G site
 - RAL-LCG2
- 100G sites
 - Imperial College, QMUL, RALPPD
- 40G sites
 - Brunel, Lancaster, Manchester, Durham
- 20G sites
 - Liverpool, Glasgow, Edinburgh, Oxford, Bristol
- 10G sites
 - RHUL, Sheffield, Birmingham, Sussex
- Message: site network teams need to be ready to upgrade the site WAN connection possibly multiple times if requested
- The monthly Jisc [Research Network Engineering Community](#) calls were set up to discuss such matters

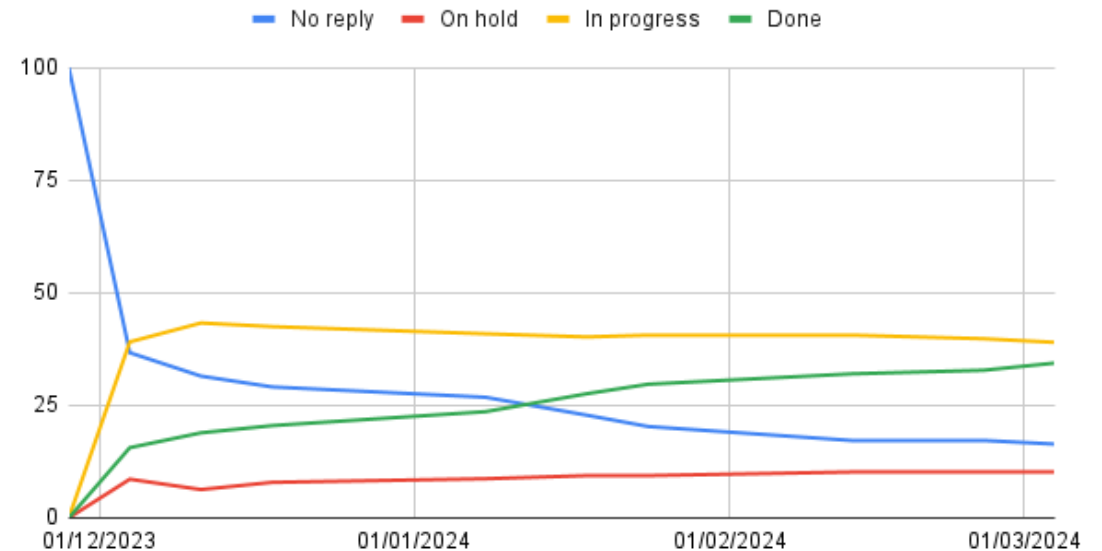
IPv6 readiness

- Sites with storage have made it dual stack (Antares EOS frontend being made dual-stack)
- WLCG Management Board have now requested that WLCG sites make CE's and worker nodes dual-stack by June 2024
- Being tracked <https://twiki.cern.ch/twiki/bin/view/LCG/Wlclgv6>
- In the UK: Brunel, Imperial, QMUL, Manchester, Durham, Oxford, RALPPD are 'done'
- RAL Tier-1 will have dual-stack WN by the summer
- WLCG summer workshop in DESY will have a session on this
- Also, a IPv6 R&D project tracking legacy v4 usage ran in DC24, looked at OPN links, especially CERN – KIT
- Network monitoring should measure both v6 and v4

Tier-1/2 IPv6 CE/WN deployment status [05-03-2024]

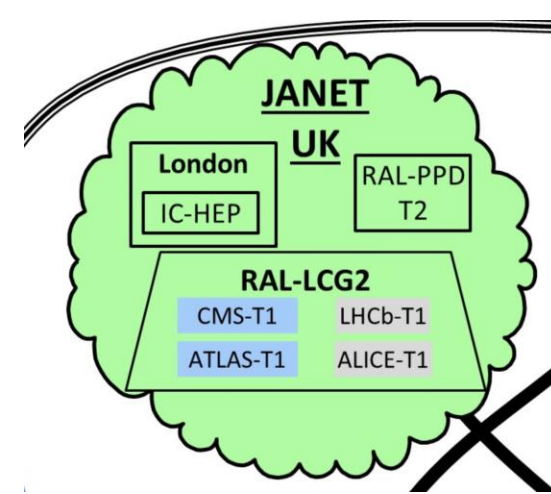


Tier-1/2 CE/WN IPv6 deployment status vs. time



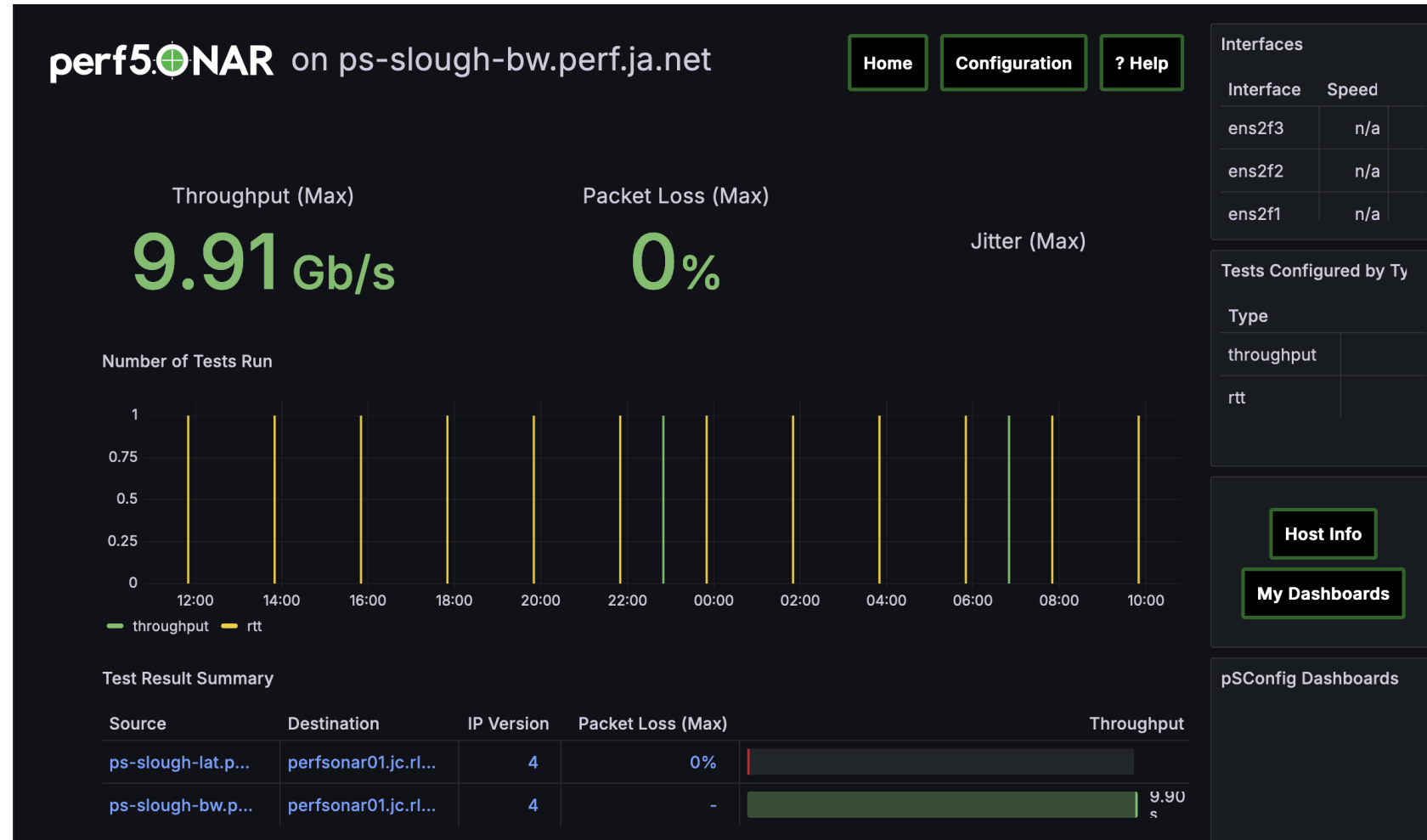
LHCONE statement

- GridPP has historically had no bandwidth limitation issues that would, in themselves, motivate the use of LHCONE and this remains the case today
- GridPP has hitherto not specifically called for its Tier-2 sites to join LHCONE
- We note that Jisc have always been ready and willing to configure LHCONE for GridPP if we had requested it.
- Joining the LHCONE means that you can give more trusted access to a significant fraction of WLCG sites
- Principally for reasons of uniformity with the rest of WLCG, since August 2023 GridPP has modified its policy for core Tier-2 sites.
- GridPP now requests core sites to consider joining LHCONE if they can easily do so.
- GridPP does not specifically request its non-core sites to join LHCONE
- As in the past, any non-core site wishing to join LHCONE is endorsed to do so.



perfSONAR

- The WLCG is probably the largest user community using perfSONAR
- GridPP has been running perfSONAR for several years and continues to maintain and upgrade its perfSONAR infrastructure.
- Upgrading to the latest version of the pS software and a recent OS is strongly recommended
- Many problems with see with perfSONAR appear to be related to the implementation of IPv6 at the site
- The Jisc perfSONAR node in Slough is now running a beta version of 5.1 with new Grafana web page
- Looking to build a new Grafana dashboard to replicate the venerable UK MaDDash



<https://ps-slough-lat.perf.ja.net/>

Jisc

- Jisc is in close communication with GridPP about its anticipated networking requirements
- Jisc is aware of the significant data rates expected with the advent of the High Luminosity upgrade to the LHC.
- It is also aware of the data challenges being planned by the WLCG leading up to the start of the HL-LHC in 2029
- There is an ongoing programme of upgrading the connections of individual GridPP sites to the Janet network as and when requirements indicate it is appropriate.
- Also, following from this, upgrades to the Janet backbone, and the GEANT peering
- Jisc also assists with the CERN-RAL OPN and facilitates LHCONE implementation for two GridPP sites
- The Jisc Network Performance Team was set up to help Janet-connected sites make best use of the Janet Network, especially when transferring the large volumes of data often found in research computing such as in the WLCG
- The Janet Netperf Team works closely with GridPP to the mutual benefit of both parties

Global Science Network Forum

- Several other large data volume experiments anticipated to start up between now and 2030
 - SKA (to be discussed at LHCONE meeting in Catania)
 - VRO (Vera Rubin Observatory - telescope in Chile using IRIS resources)
 - ITER?/STEP (fusion)
- Also other sites and communities such as
 - EBI (bioinformatics),
 - Diamond Light Source (synchrotron science, cryo-EM)
 - CEDA (climate)
 - Isambard-AI, Darwin (Artificial Intelligence)
 - Other HPC, e.g. Exascale
- Jisc is interested in trying to predict anticipated usage of Janet
- Tracking them, for example via RNE Community calls
- Will they want their own overlay networks like LHCONE e.g. an SKAONE?
- Side meeting to discuss this at the TNC summer networking meeting in Rennes

Summary

- Run3 of the LHC era will continue through 2024 and 2025, then there will be a three year shutdown and HL-LHC starts in 2029
- ATLAS have refactored their architecture resulting in five UK nuclei sites which will require significant network connectivity (~100 Gbit/s) and the remainder less (20 Gbit/s)
- CMS are happy with current connectivity, expect significantly more data in 2024, and continue to use the demanding AAA paradigm
- The Tier-1 connection to Janet is in the process of being upgraded to a 400 Gbit/s resilient link, which should be adequate for the remainder of Run-3
- The summary of the scale of connection required is:
 - **Tier 1: 400 Gbit/s**
 - **Tier-2 (large): 100 Gbit/s**
 - **Tier-2 (other): 20 Gbit/s (and in some cases possibly 40 Gbit/s)**
- GridPP now requests the remaining core sites to consider joining LHCONE if they can easily do so. Any non-core site wishing to join LHCONE is endorsed to do so.
- GridPP thanks Jisc for its continued pro-active engagement with the LHC programme
- Message: site network teams need to be ready to upgrade the site WAN connection if requested