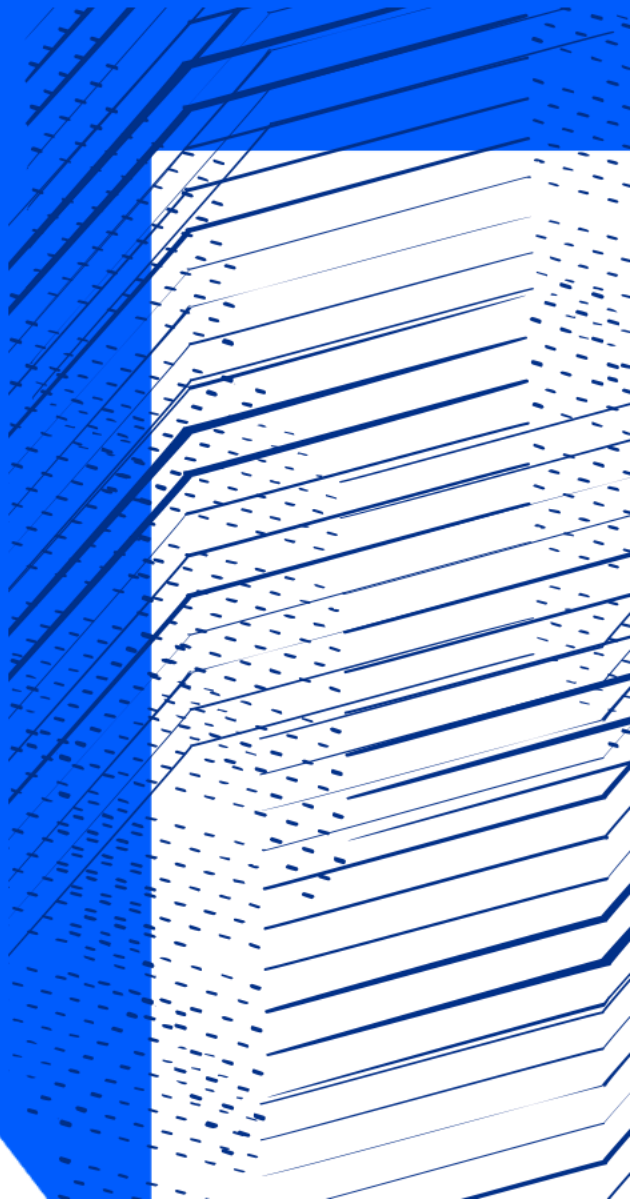
Science and Technology Facilities Council

Tier-1 Job Efficiency

Alastair Dewhurst



Introduction

1 Accounting

2 Network Upgrade

3 Batch Farm Updates

4 Storage Improvements

Science and Technology Facilities Council

Science and Technology Facilities Council

Accounting

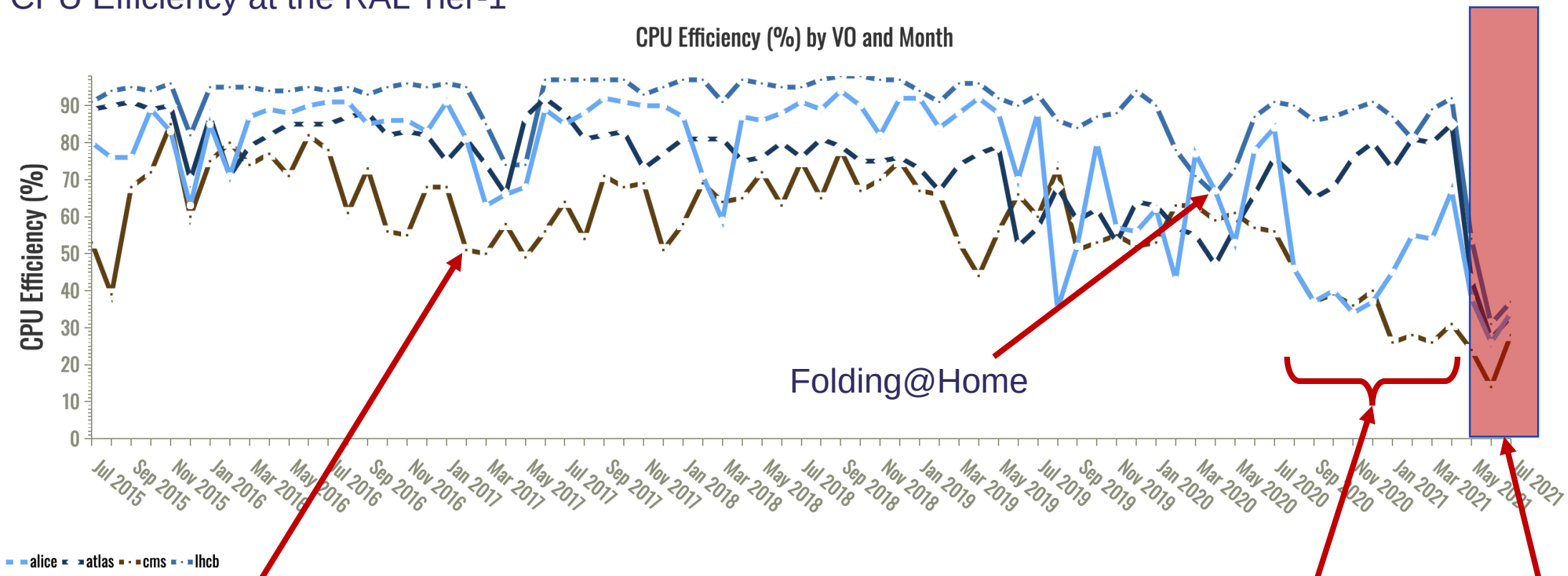
Measuring CPU Efficiency Accurately

Accounting

- The Tier-1 has struggled to provide accurate accounting information this year.
- This has been due to a variety of issues, which have often appeared after performing an upgrade.
  - Accounting problems often take time to be observable.
- Two outstanding issues:
  - Some job numbers are reported twice in May.
  - CPU Efficiency is wrong since April because many jobs are not reporting CPU usage (so it gets entered as 0).

Historical CPU Efficiency

CPU Efficiency at the RAL Tier-1



Historically CMS have had lower efficiency, although there is a clear drop since July 2020.

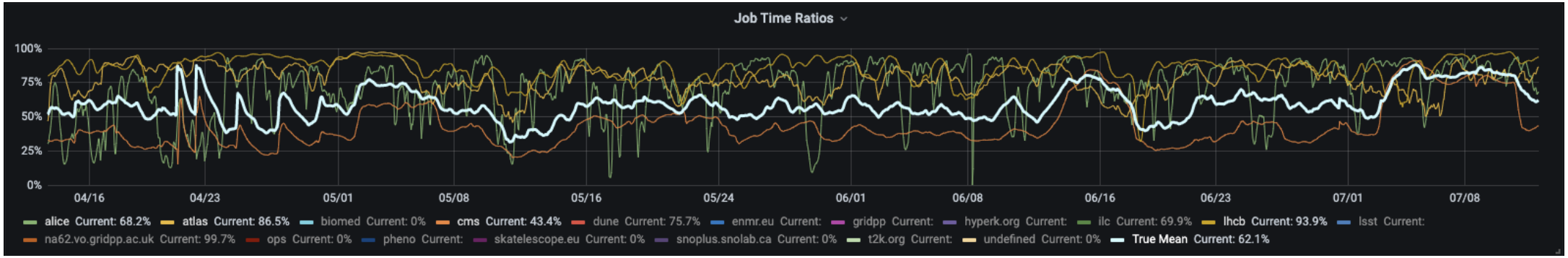
ALICE have more erratic efficiency in the last few years. They are running a wider variety of jobs which include more reads from both Echo and offsite.

Efficiency Calculation Incorrect

Accounting

- We have some manual calculations of efficiency in the last few months.
  - Most VOs in line with previous months.
  - Improvement in CMS (~50% efficiency)

Tier-1 Monitoring from April to July



EGI monitoring for last 6 months

Resource Centre RAL-LCG2 — CPU Efficiency (%) by VO and Month (LHC VOs)

VO	Jan 2021	Feb 2021	Mar 2021	Apr 2021	May 2021	Jun 2021
alice	45.1%	55.43%	54.73%	67.04%	38.33%	26.31%
atlas	73.53%	81.16%	80.36%	85.62%	43.58%	27.78%
cms	26.03%	28.55%	26.61%	31.42%	24.49%	14.48%
lhcb	87.89%	81.27%	89.05%	92.52%	54.47%	31.45%
Total	63.89%	64.81%	73.89%	75.43%	39.29%	24.45%

Improvements

- We are still trying to understand the cause of the failure to report CPU usage.
  - RAL does run jobs in docker containers which is less common among sites.
- We will review if we had significant numbers of 0 CPU usage jobs in the past.
  - Some months we were scratching our heads as to why efficiency was lower.
- Monthly Tier-1 meeting has been restarted to sanity check all accounting numbers.
- Test jobs to be run that:
  - Sleep (~0% CPU Efficiency).
  - Run a CPU intense process (~100% CPU Efficiency).
  - Request 2 cores and run a CPU intense process on one (~50% CPU Efficiency).

Science and Technology Facilities Council

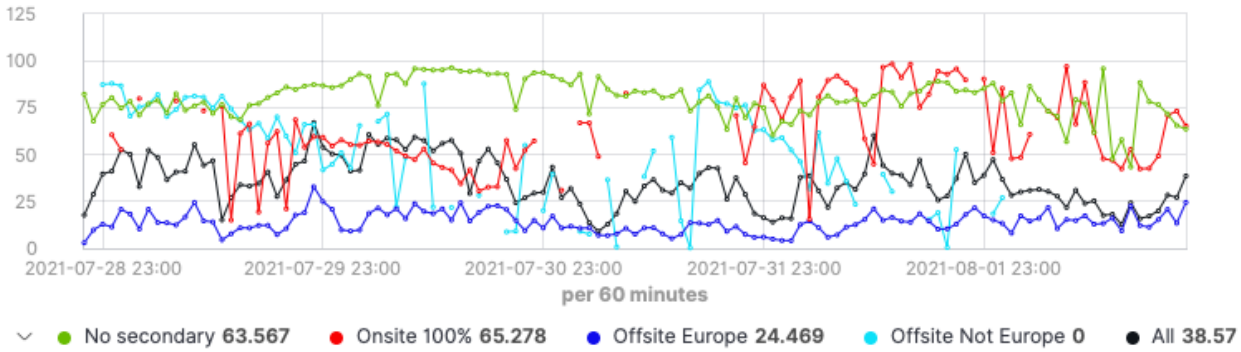
Network Upgrade

Spoilers Look away now if you don't want to know the conclusion to Katy's talk.

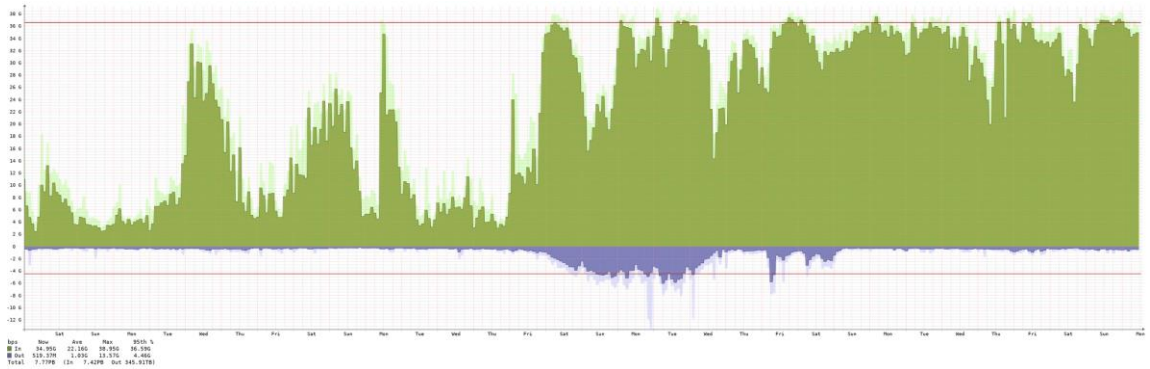
CMS Efficiency

- If CMS are pulling data from Offsite the job efficiency is lower.
- In the last few weeks, the 40Gb/s link that CMS-AAA traffic is routed via has been saturated.

CMS job efficiency as a function of location

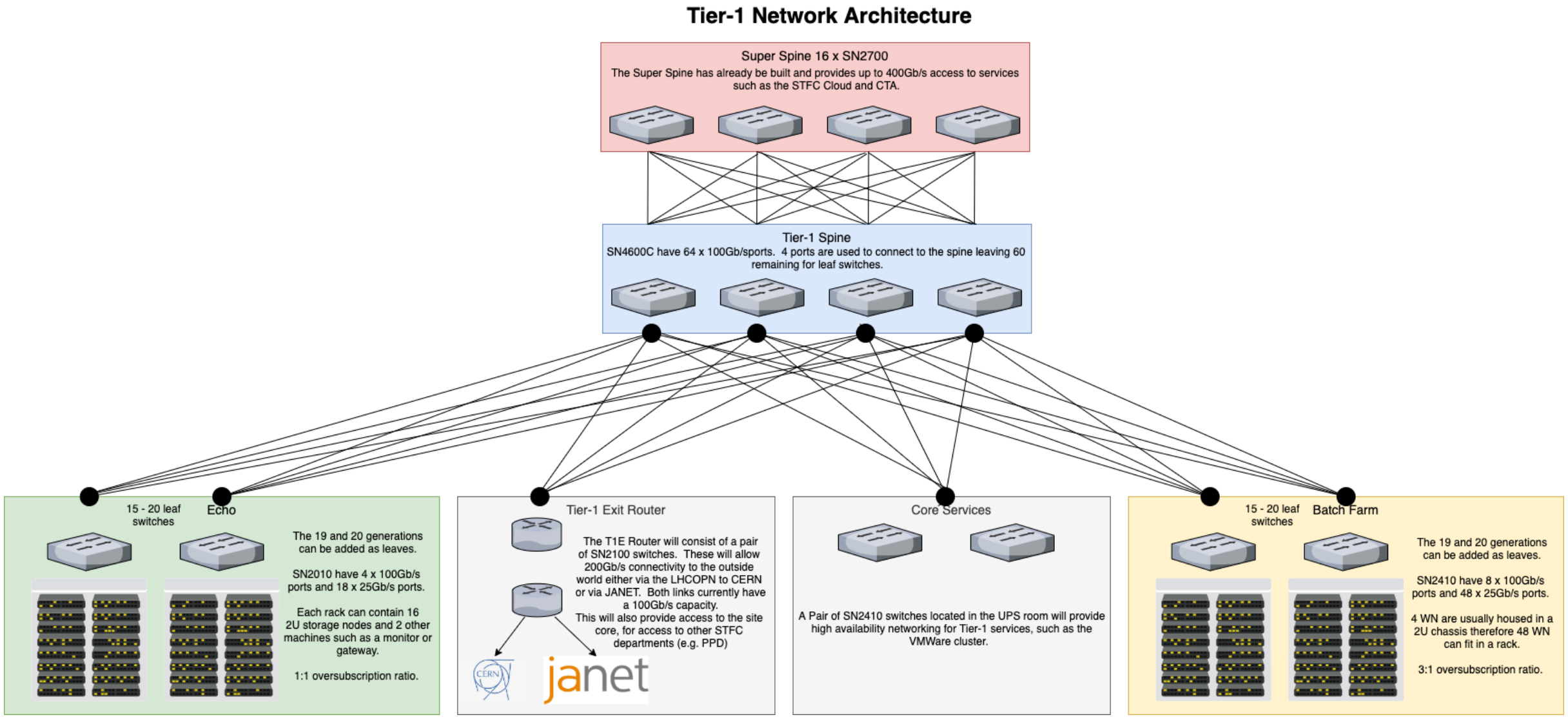


CMS AAA traffic at RAL

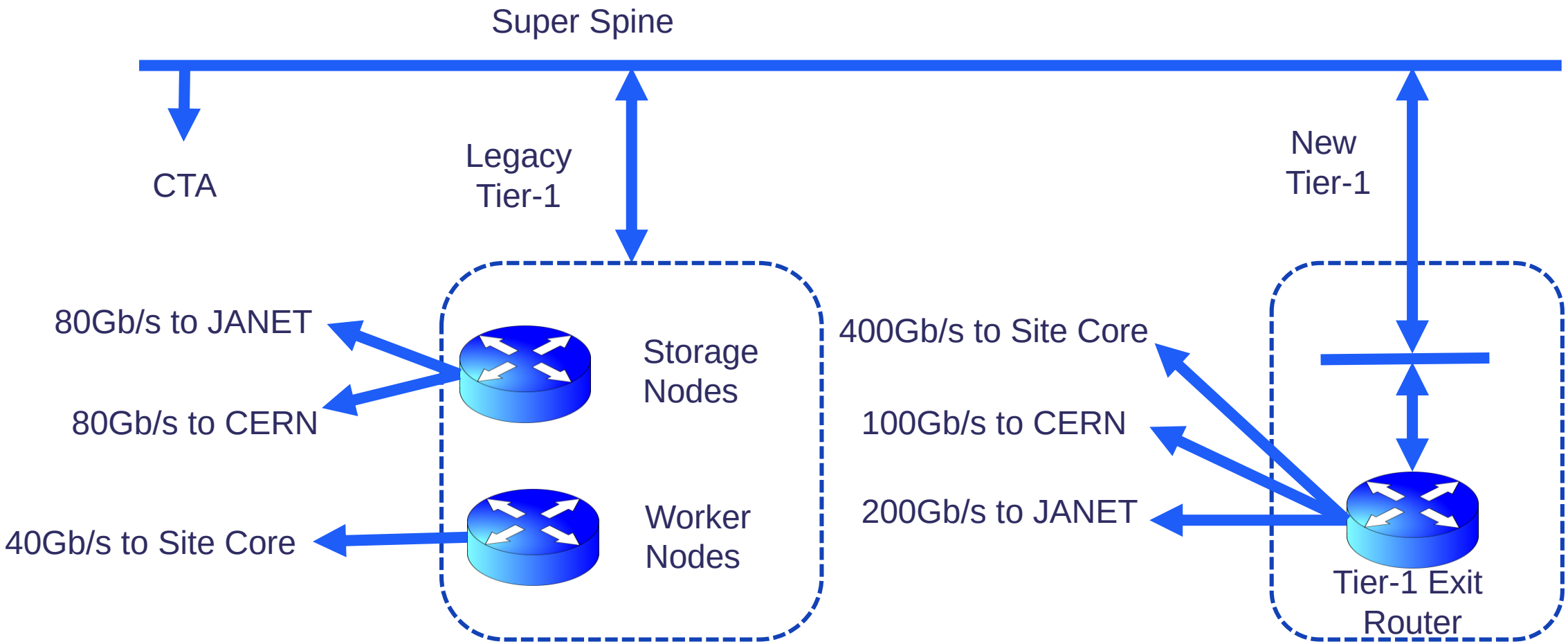


Designing a new Tier-1 Network

- We decided to build a new network for the RAL Tier-1 with a Spine / Leaf topology.
  - New CTA Tape system is on a separate network pod.
- Wanted to provide a uniform experience for end users:
  - Dual Stack everywhere.
  - All machines accessible to the outside world will be on the LHCOPN and LHCONE.
- We choose a single vendor (Mellanox) with the Cumulus OS for all systems.
  - Also required storage and CPU nodes to come with Mellanox NICs.
- Hardware purchased in the last two years can be easily added to new network:
  - Review other hardware on case by case basis.

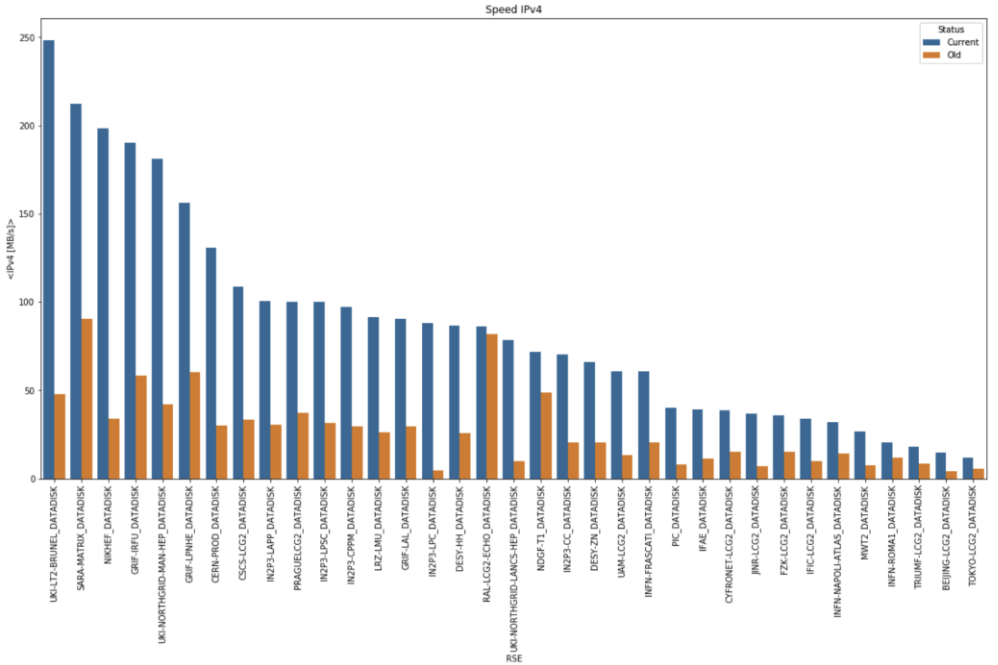

Tier-1 Network Architecture
Super Spine 16 x SN2700
The Super Spine has already be built and provides up to 400Gb/s access to services such as the STFC Cloud and CTA.
Tier-1 Spine
SN4600C have 64 x 100Gb/sports. 4 ports are used to connect to the spine leaving 60 remaining for leaf switches.
15 - 20 leaf switches
Echo
The 19 and 20 generations can be added as leaves.
SN2010 have 4 x 100Gb/s ports and 18 x 25Gb/s ports.
Each rack can contain 16 2U storage nodes and 2 other machines such as a monitor or gateway.
1:1 oversubscription ratio.
Tier-1 Exit Router
The T1E Router will consist of a pair of SN2100 switches. These will allow 200Gb/s connectivity to the outside world either via the LHCOPN to CERN or via JANET. Both links currently have a 100Gb/s capacity.
This will also provide access to the site core, for access to other STFC departments (e.g. PPD)
CERN
janet
Core Services
A Pair of SN2410 switches located in the UPS room will provide high availability networking for Tier-1 services, such as the VMWare cluster.
15 - 20 leaf switches
Batch Farm
The 19 and 20 generations can be added as leaves.
SN2410 have 8 x 100Gb/s ports and 48 x 25Gb/s ports.
4 WN are usually housed in a 2U chassis therefore 48 WN can fit in a rack.
3:1 oversubscription ratio.

Tier-1 Network Progress



Firewall upgrade

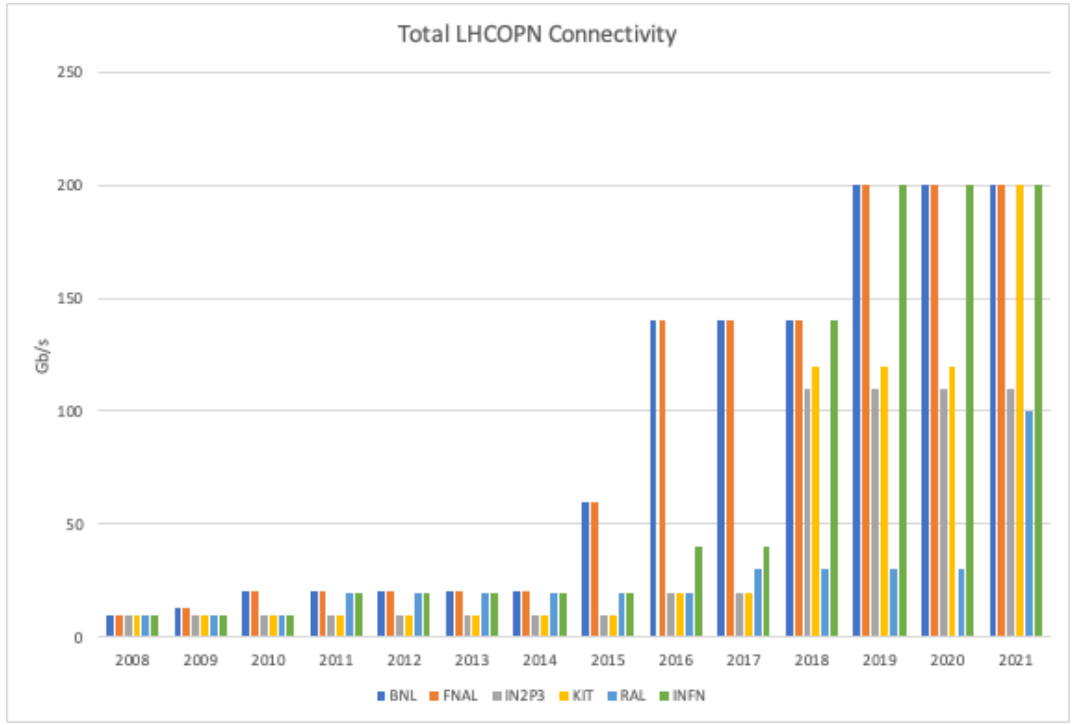
- From July 2020 - April 2021, due to everyone working from home there was significant load on the firewall affecting download speeds.
- Firewall upgrade completed in April 2021.
  - No obvious difference observed between IPv4 and IPv6



Network upgrade

- How did our network get into this state?
  - We are about 3 years behind other Tier-1s.
- How do we ensure this doesn't happen again?
  - ~£900k investment in network over last 2 years.
  - CHEP paper with design out to 2031.
  - Jonathan Churchill is SCD network architect

Network Backward Look



No Tier-1 Manager

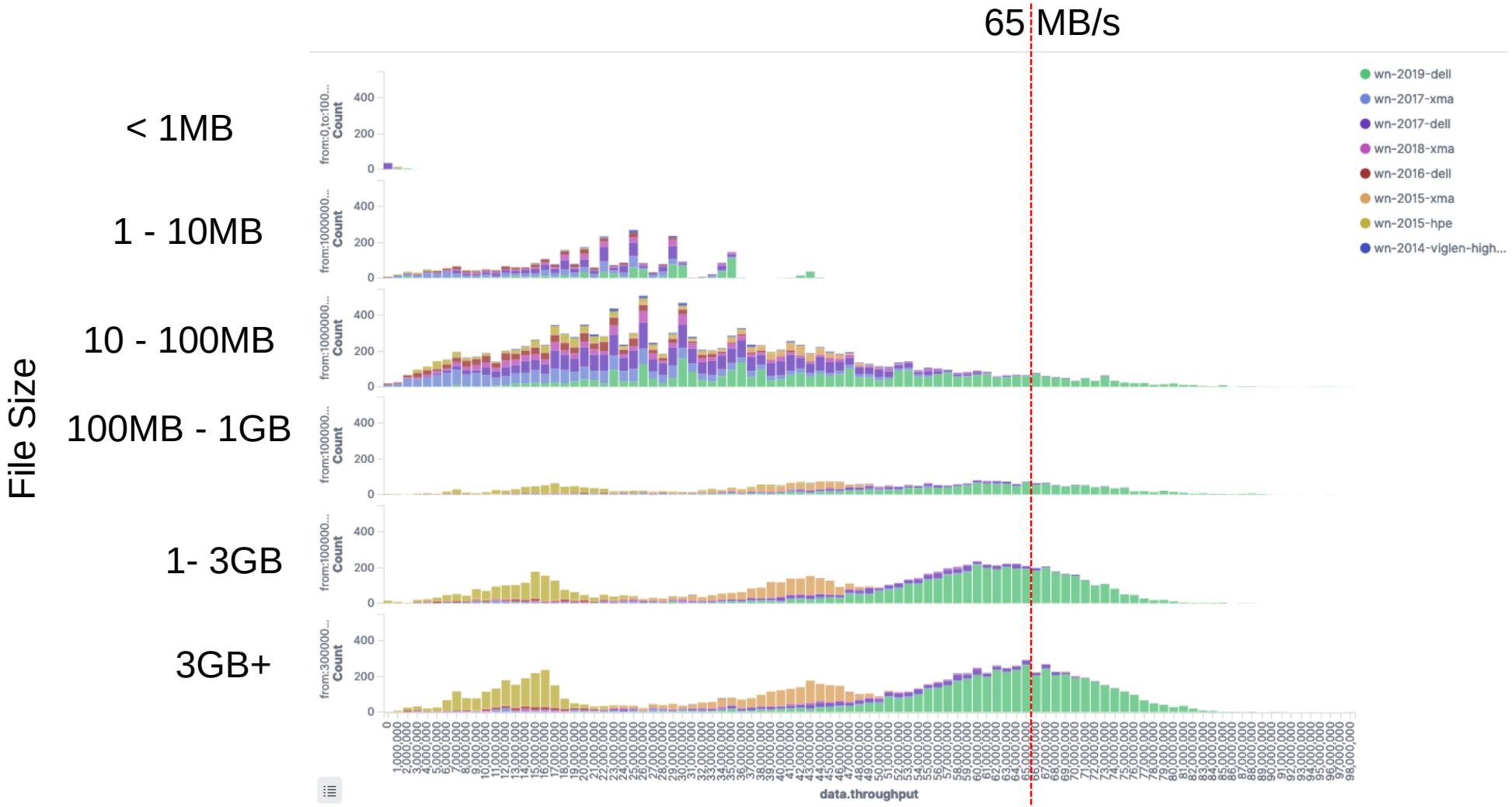
Science and Technology Facilities Council

Batch farm updates

Batch farm

- Since ~April 2021 Tom Birkett has been the batch farm manager.
- Several upgrades have been performed, including:
  - ARC-CEs
  - HTCondor
  - Docker
- Several performance tweaks as well, to balance the load with the capability of the nodes.

WN Download speeds from Echo

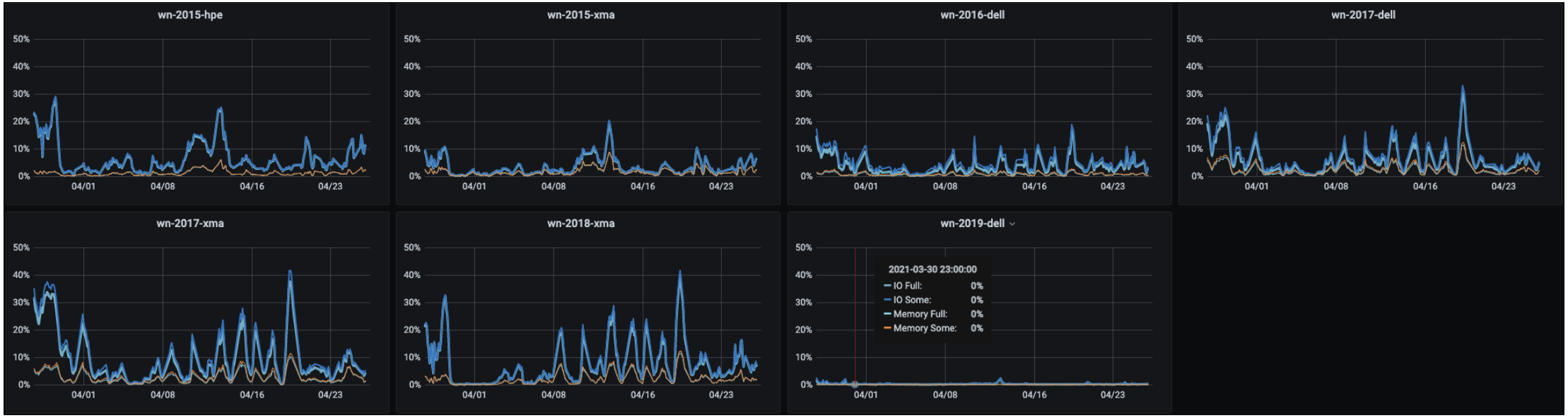


Transfer Speed in April 2021

Alastair Dewhurst, 2nd September 2021

WN Load

- I/O contention on WN is one reason why transfers can be slow.
  - All non-SSD generations show some contention.
- Reducing number of jobs run on older generations to reduce I/O contention.



Scheduling Issues

- Chris Brew created a test job to test CPU efficiency.
  - He ran a test on an empty 2017 WN (64 job slots total)

Processes	Efficiency
8	95.6
16	90.8
32	59.5
64	26.4

- Our batch farm is configured to reserve some nodes for multi-core only jobs.
  - This concentrates CMS (high I/O jobs) on to certain nodes.
- Even though SSD nodes currently appear to remove any I/O contention, the more we can spread load the better.

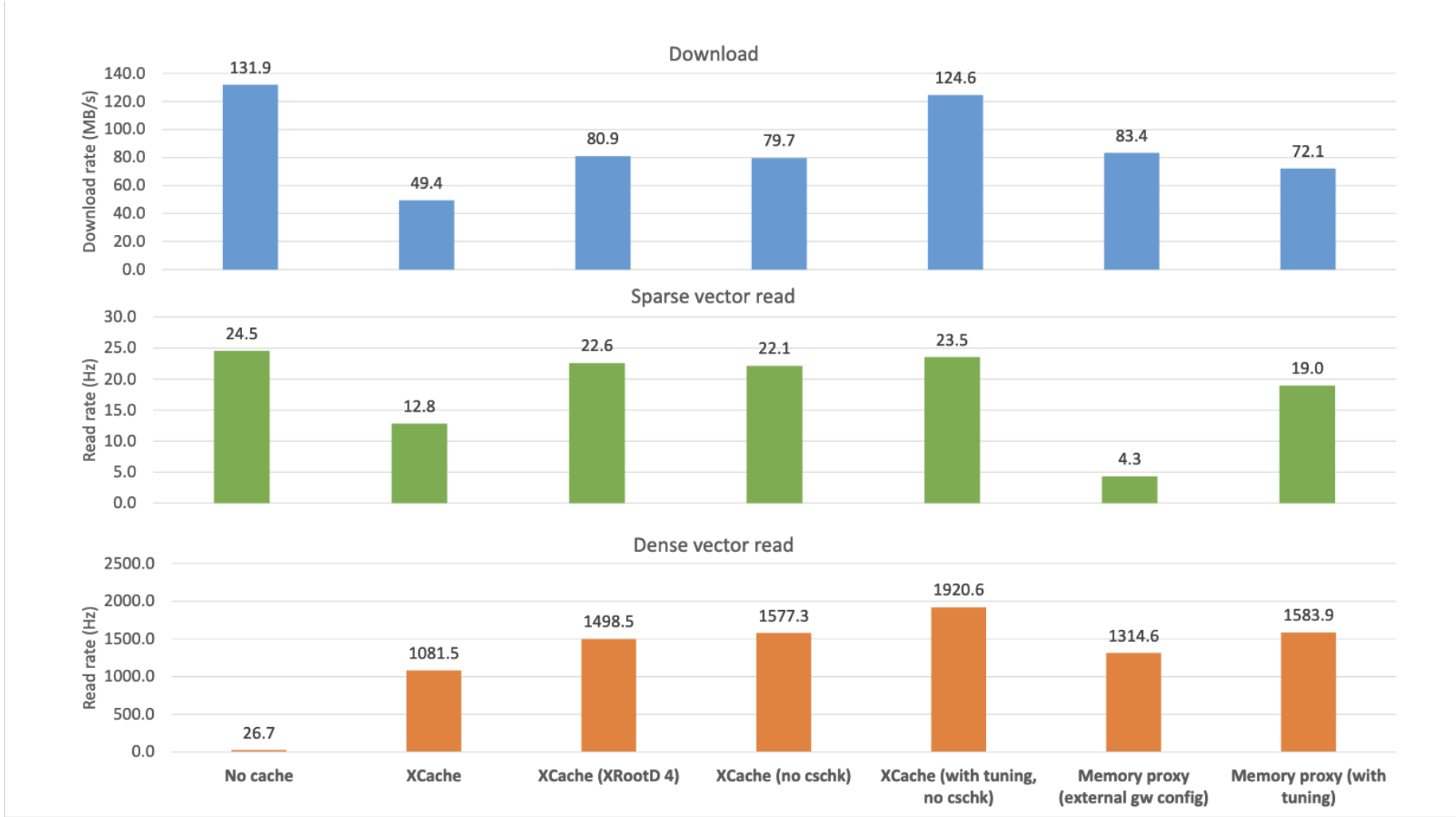
Science and Technology Facilities Council

Storage improvements

XRootD 5.X

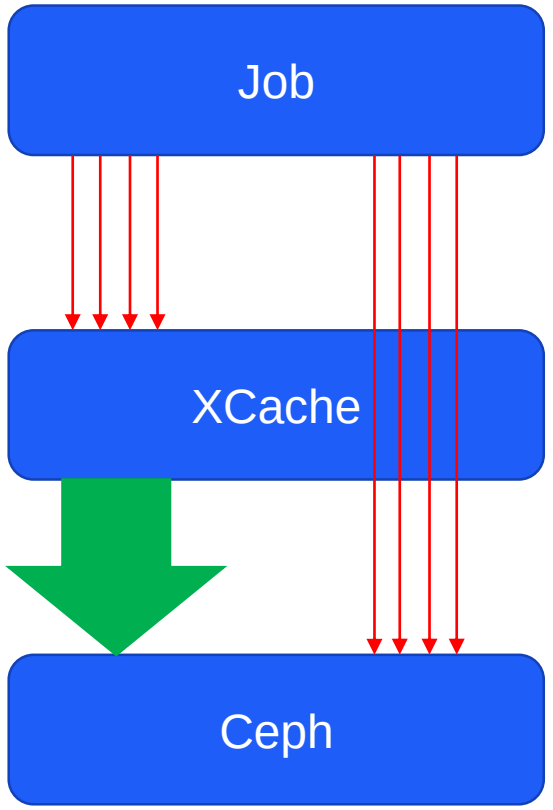
- In April 2021 a stable version of XrootD 5 was released.
  - We had tested and rolled it out within 6 weeks.
- XrootD 5 contained many improvements for Echo.
  - XRootD 5.2.0 introduced strict checksumming by default which reduced performance. This has been disabled.
  - Cache settings have been re-tuned.
  - TPC now works and we are waiting scale tests from ATLAS

Performance tuning



Vector Reads

- When Echo was first deployed it was assumed that it would be faster at transferring entire objects.
- XCaches were deployed on WN to fetch blocks of data.
  - This also protects against pathological jobs.
- Analysis found that XCaches worked well but if it is busy it can “pass through” the request.
- Vector reads appear to trigger job failures.



Future work

- Sarah Byrne (no relation) is starting in September to work with Ian Johnson on improving code base.
- Will be working on consolidating code.
- Will remove strict locking of files for reads.

XCache

XrdCeph

LibradosStriper

Science and
Technology
Facilities Council

Questions?