Science and Technology Facilities Council

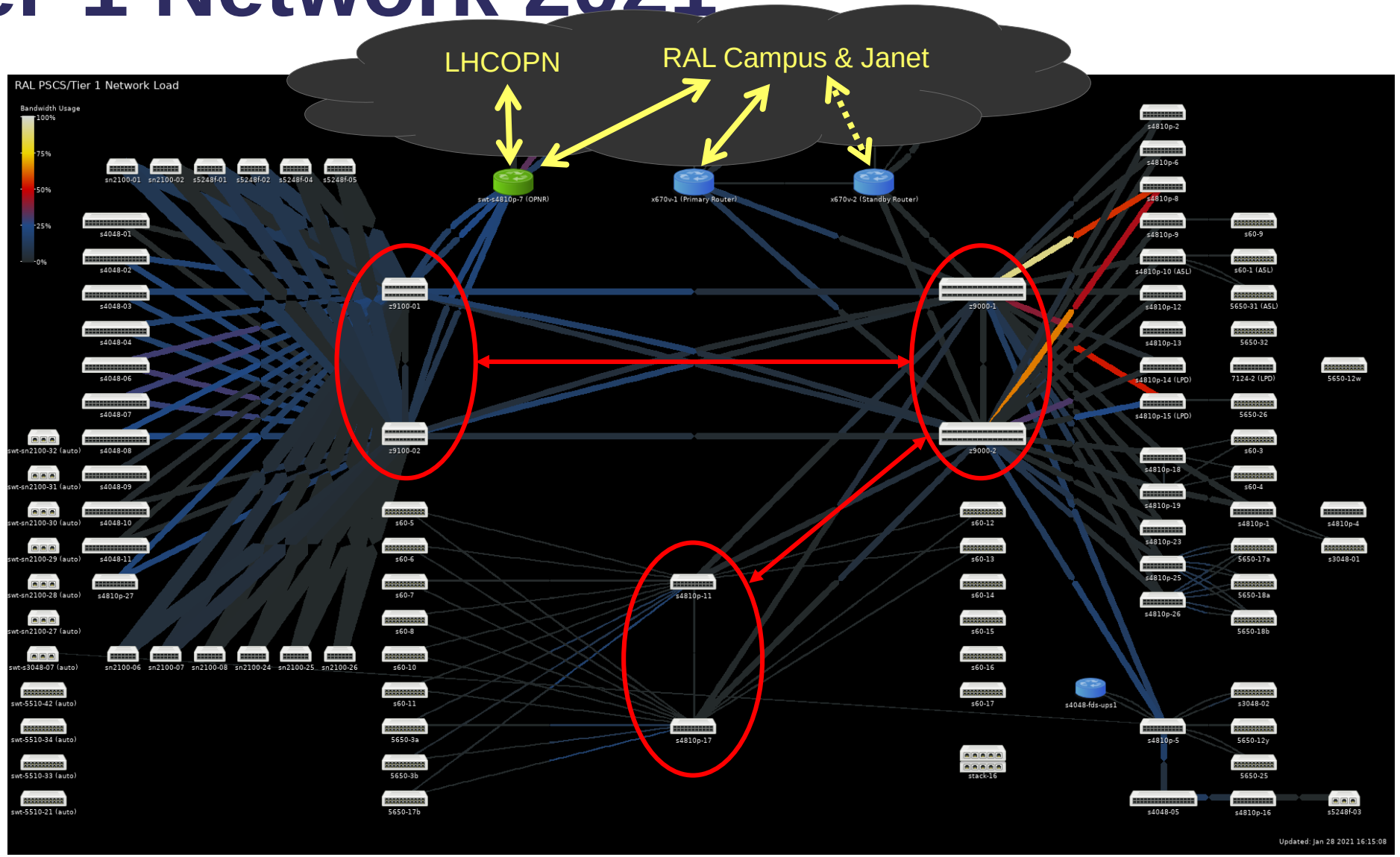
Tier-1 Network and general update

Alastair Dewhurst & James Adams on behalf of the Tier-1

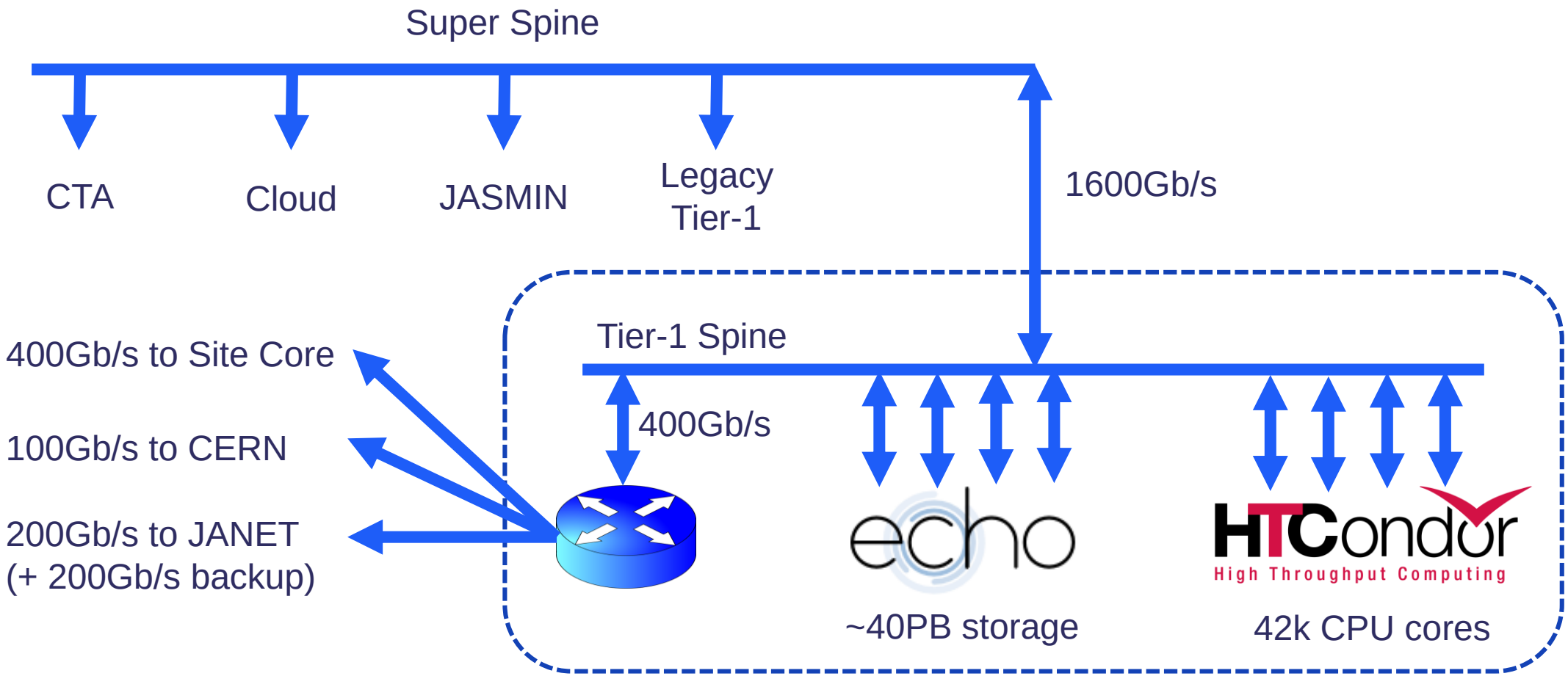
Introduction

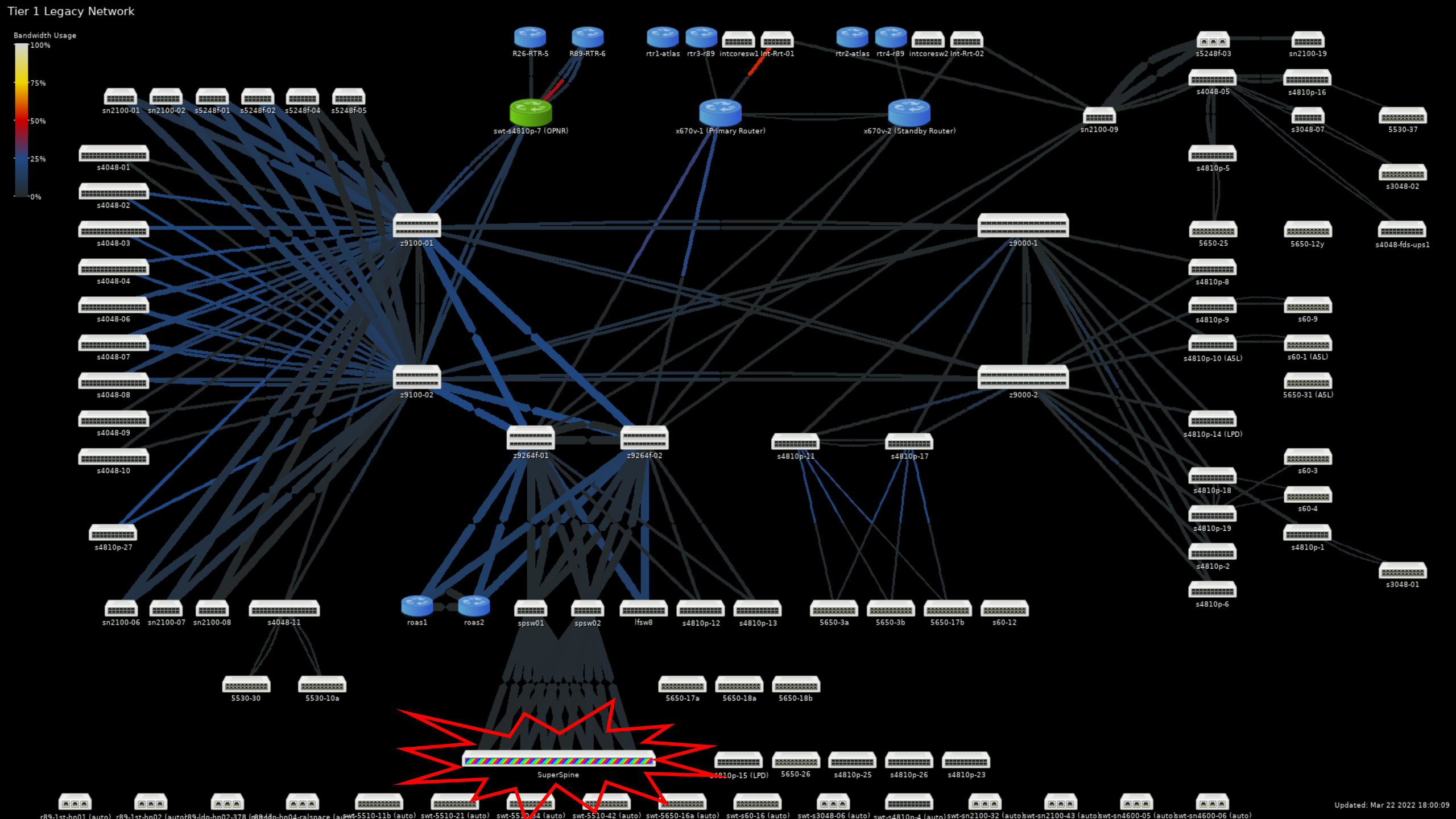
- Tier-1 Network
  - Major project to replace the entire Tier-1 network since September 2020.
- Antares
  - This will be covered in a talk by Tom Byrne on the setup and from Katy Ellis on how it has performed so far in the tape challenge.
- Procurement
  - CPU
  - Disk
- Future Plans

Tier-1 Network 2021

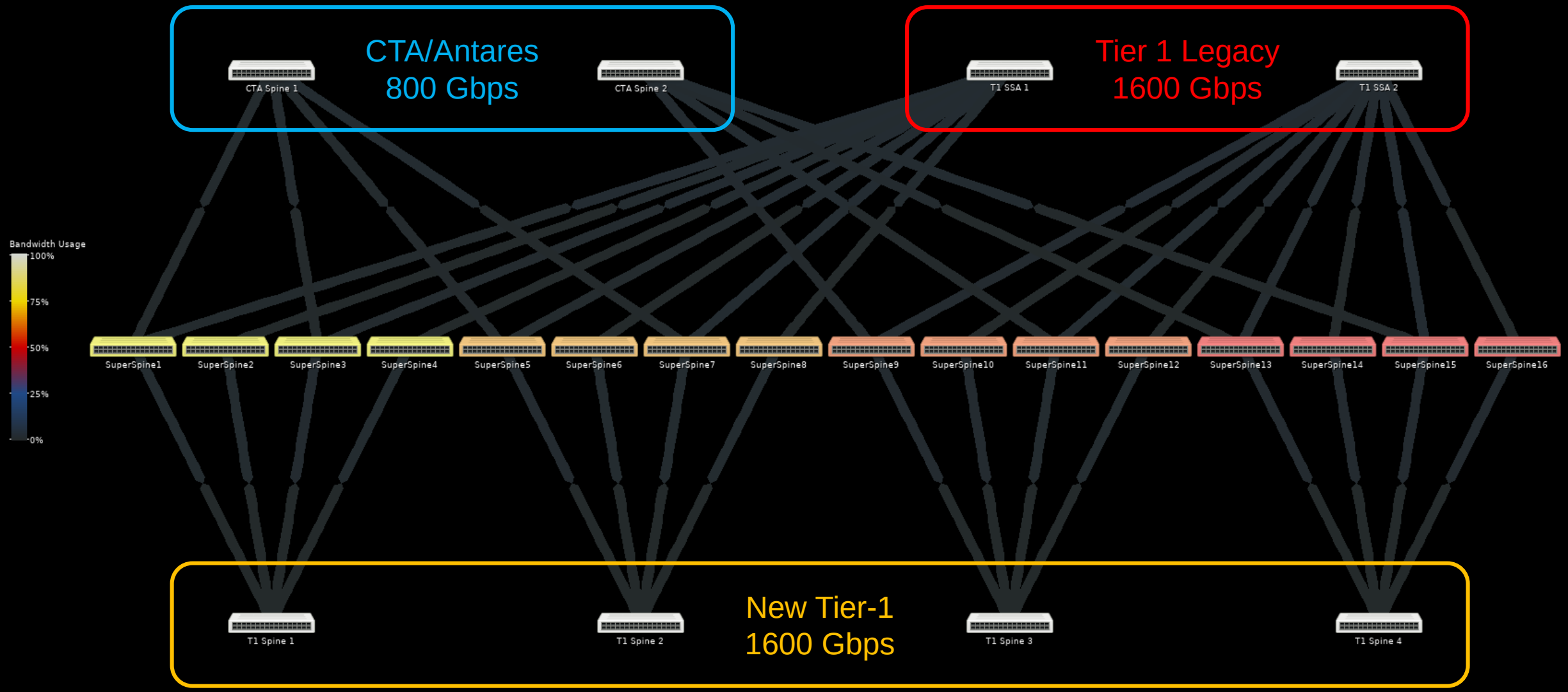


Core network plan

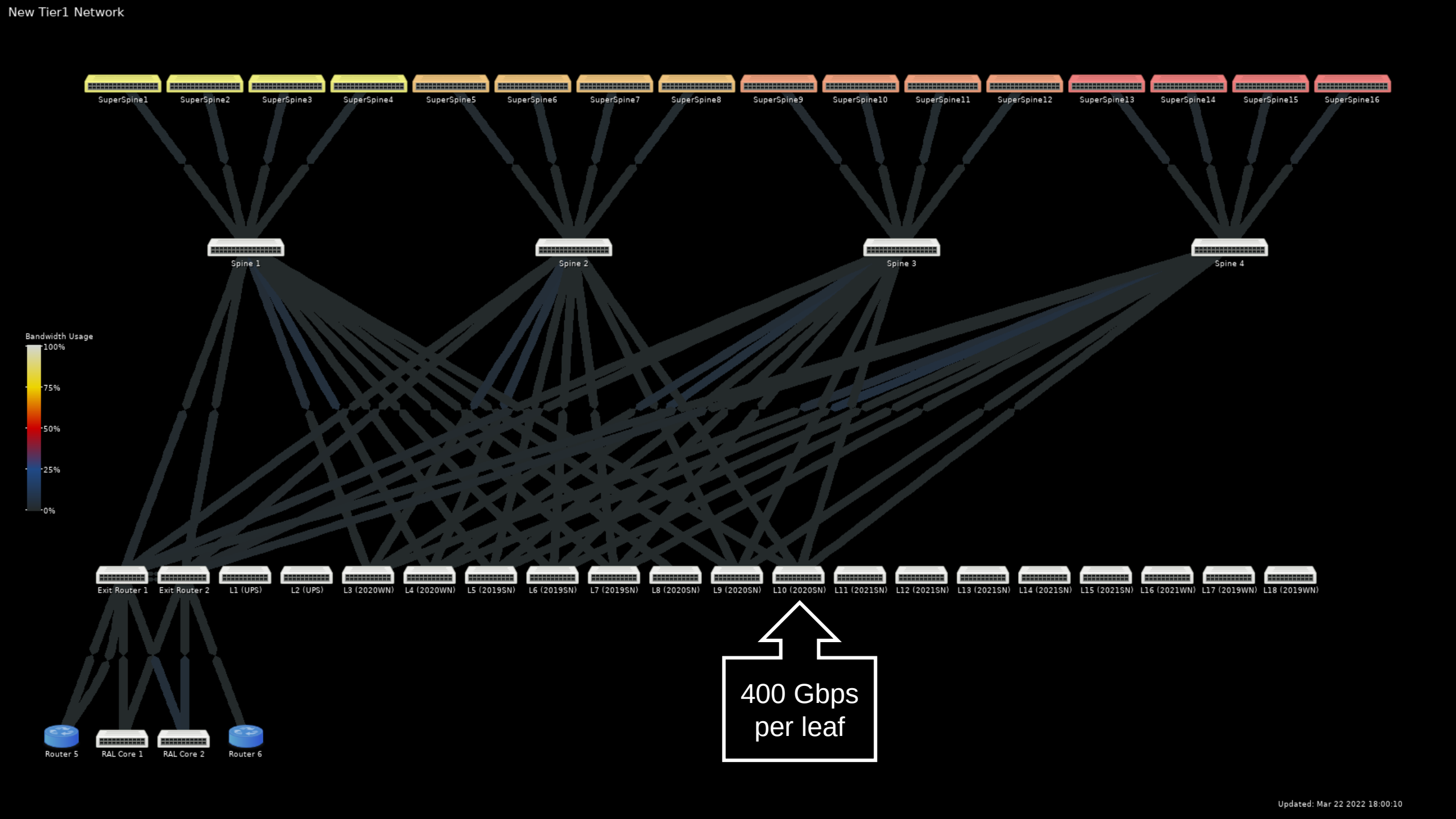



Tier 1 Legacy Network
Bandwidth Usage
100%
75%
50%
25%
0%
R26-RTR-5
R89-RTR-6
rtr1-atlas
rtr3-r89
intcoresw1
Int-Rrt-01
rtr2-atlas
rtr4-r89
intcoresw2
Int-Rrt-02
s5248f-03
sn2100-19
s4048-05
s4810p-16
s3048-07
5530-37
s3048-02
s4048-fds-ups1
sn2100-01
sn2100-02
s5248f-01
s5248f-02
s5248f-04
s5248f-05
swt-s4810p-7 (OPNR)
x670v-1 (Primary Router)
x670v-2 (Standby Router)
sn2100-09
s4810p-5
5650-25
5650-12y
s4048-01
s4048-02
s4048-03
s4048-04
s4048-06
s4048-07
s4048-08
s4048-09
s4048-10
z9100-01
z9100-02
z9000-1
z9000-2
s4810p-8
s4810p-9
s60-9
s4810p-10 (A5L)
s60-1 (A5L)
5650-31 (A5L)
s4810p-14 (LPD)
z9264f-01
z9264f-02
s4810p-11
s4810p-17
s60-3
s4810p-18
s60-4
s4810p-19
s4810p-1
s4810p-2
s4810p-6
s3048-01
s4810p-27
sn2100-06
sn2100-07
sn2100-08
s4048-11
roas1
roas2
spsw01
spsw02
lfsw8
s4810p-12
s4810p-13
5650-3a
5650-3b
5650-17b
s60-12
5530-30
5530-10a
5650-17a
5650-18a
5650-18b
SuperSpine
s4810p-15 (LPD)
5650-26
s4810p-25
s4810p-26
s4810p-23
r89-1st-hp01 (auto)
r89-1st-hp02 (auto)
swt-5510-11b (auto)
swt-5510-21 (auto)
swt-5510-42 (auto)
swt-5650-16a (auto)
swt-s60-16 (auto)
swt-s3048-06 (auto)
swt-sn2100-32 (auto)
swt-sn2100-43 (auto)
swt-sn4600-05 (auto)
swt-sn4600-06 (auto)
Updated: Mar 22 2022 18:00:09

Superspine



Updated: Mar 22 2022 18:00:11


New Tier1 Network
SuperSpine1
SuperSpine2
SuperSpine3
SuperSpine4
SuperSpine5
SuperSpine6
SuperSpine7
SuperSpine8
SuperSpine9
SuperSpine10
SuperSpine11
SuperSpine12
SuperSpine13
SuperSpine14
SuperSpine15
SuperSpine16
Spine 1
Spine 2
Spine 3
Spine 4
Bandwidth Usage
100%
75%
50%
25%
0%
Exit Router 1
Exit Router 2
L1 (UPS)
L2 (UPS)
L3 (2020WN)
L4 (2020WN)
L5 (2019SN)
L6 (2019SN)
L7 (2019SN)
L8 (2020SN)
L9 (2020SN)
L10 (2020SN)
L11 (2021SN)
L12 (2021SN)
L13 (2021SN)
L14 (2021SN)
L15 (2021SN)
L16 (2021WN)
L17 (2019WN)
L18 (2019WN)
Router 5
RAL Core 1
RAL Core 2
Router 6
400 Gbps per leaf
Updated: Mar 22 2022 18:00:10

Complete

- Four spines and eight leaves installed and configured
  - eBGP routing throughout
- Two Exit Routers installed and configured
  - 400Gbps to RAL core, OSPF routing
  - 400Gbps for Firewall Bypass and LHCONE
- Two leaves of worker nodes running for four months
  - 112 nodes
  - 14,336 job slots
  - ~200K HS06

In Progress

- Getting storage leaves into production
  - More complex than compute
  - Both data and cluster networks must transit the super-spine
  - Gateways require LHCOPN routes and IPv6
- LHCONE
  - New Tier-1 Exit Routers will peer with LHCONE router
  - Cabling is in place and links are active
  - Expecting to start peering within the next few weeks*

Next

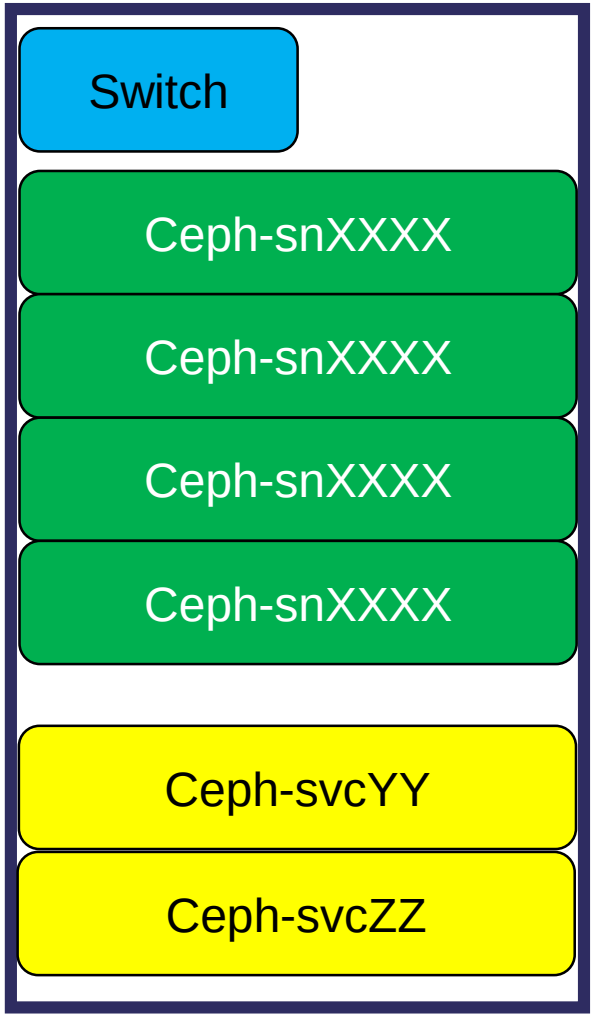
- IPv6
  - Cannot be achieved within STFC addressing plan
  - SCD has requested a new /48 from JISC
- Move of LHCOPN to new network
  - Move cable directly to a Tier-1 Exit Router
  - Requires Firewall bypass to be moved at the same time
  - Will reverse routes for OPN traffic over super-spine

Procurement

- Procurement has been extremely challenging this year.
  - It doesn't look like it is getting better any time soon.
- The Tier-1 placed orders in July and August 2021.
  - Some components from these orders are scheduled for delivery in August 2022!!
  - By working with vendors closely and being flexible in the components we allow, we also placed (smaller) order in December 2021 which were fulfilled by January 2022.
- I have run joint SCD procurements this year total £4.6 million. For the Tier-1:
  - £630k on CPU (£4.37 / HS06 which is a 24% increase in cost)
  - £490k on Disk (£29.72 / raw TB which is 15% decrease in cost)
  - £550k on Tape
  - £150k on Network

Disk

- For Echo for GridPP + IRIS + Others we purchased 5 racks of storage (~34PB).
  - Installed on 28th February 2022, network cabling to start this week.
- Each Rack:
  - 1 Mellanox SN2010 ToR switch
  - 16 x Storage Servers
  - 2 x Gateways / Monitors
- Each Storage server:
  - 24 x 20TB HDD, 1 x 4TB SSD, 1 x 480GB SSD (for OS)
  - 2 x Intel Xeon Silver 4214R 2.4G, 12C/24T
  - 256GB memory
  - 25Gb/s NIC



CPU

TIER 1					
1024US-TRT Supermicro CSE-819UTS-R1K02P-A Chassis Supermicro H12DSU-iN Motherboard	**Supermicro 1024US-TRT 1U Node, 2x7763, 1TB RAM, 1x7.68TB SSD, DP 25GbE** 1U Server with 4 x 3.5" Drive Bays and Redundant Power 2 x AMD Epyc 7763, 64 Core, 2.45GHz Processors 1TB DDR 4 3200MHz ECC Registered Memory (16 x 64GB) 1 x 7.68TB Enterprise Class SSD Drive Mellanox MCX512A-ACAT Connect-X5 EN Dual Port 10/25GbE SFP28 Connection IPMI 2.0 BMC with RJ45 Connection 2 x 10Gb Ethernet Built-In (RJ45 Connection) Supermicro Out of Band (OOB) Manager License 3 Year On-Site Warranty Next Business Day		£ 12,570.34	48	£ 603,376.32
	Installation *Basic Rack installation and cabling* *Includes ancilliary items - CAT 6 Cables (Management), Power Cables, Blanking Panels, Cable Labels and ties, Delivery to Site, Removal of packaging and Project Management*		£ 13,115.43	1	£ 13,115.43
Total for Tier 1					£ 616,491.75

- CPU were installed in December 2021!
  - Didn't include the Mellanox 25Gb/s NICs (which are now due in August 2022).
  - We have ordered Intel 10Gb/s NICs which will be installed in the interim.
- For our 18 generation of WN we have ordered 2TB 1DWPD SSD to replace the HDD.
- **We are currently planning to run Rocky Linux on these machines.**

Disk Cache

- I also purchased 3 x SSD Storage nodes:
  - 24 x 4TB SSD
  - 2 x AMD 7413 2.65GHz, 24C/48T
  - 256GB memory
  - Mellanox 25Gb/s NIC
- These are to be used for data intensive applications
  - 15 of these nodes were purchased for an IRIS funded CephFS cluster.
  - EOS nodes (for Antares)
  - XCaches
- Will loan one server to Oxford, for their Xcache testing.
- Will run a local Xcache.

Future Plans - Atlassian

- I have been leading a project to migrate the whole of Systems Division to using Atlassian
  - We already use OpsGenie for our oncall

Product	Licence Type	Support	User Band	Max no. of Users	List Price (£)	Discount (%)	Discount (£)	Price (£)
Jira Software Standard Academic	Cloud	12 Months	New Licence	100	2,921.76	10.00%	292.18	2,629.58
Jira Service Management Premium Academic	Cloud	12 Months	New Licence	100	15,582.70	10.00%	1,558.27	14,024.43
Confluence Standard Academic	Cloud	12 Months	New Licence	100	2,142.62	10.00%	214.26	1,928.36
Email This Issue for Jira	Cloud	12 Months	New Licence	100	1,558.27	10.00%	155.83	1,402.44
Atlassian Access Academic	Cloud	12 Months	New Licence	100	0.00	0.00%	0.00	0.00
					22,205.35		**2,220.54**	**19,984.81**

- We are paying consultants to develop an integration between JIRA and GGUS (“Email This Issue for Jira”).
- Some groups will be migrated in April, with Tier-1 migration starting in May (and probably taking most of the year).

Future Plans - Container orchestration

- In the GridPP6 bid, in the future development section, I wrote that we needed effort to develop ways to automate the way we manage the growing number of servers in Echo.
  - We didn't get funded in GridPP6.
  - We did get a little bit of funding in ExCALIBUR 1a, but we didn't get funding for 1b.
- We need to find a way of doing this:
  - We run ~20 Echo gateway supporting:
    - TPC: GridFTP, XRootD, WebDav
    - Data Access: XrootD locally, and remotely via XCaches, AAA etc
- The recommended way to run Rucio is via Kubernetes.
  - CERN run other Grid Services (e.g. FTS) in Kubernetes.

Science and
Technology
Facilities Council

Questions?