GridPP

UK Computing for Particle Physics

RAL Site Report

2023 UK HEP SYSMAN

University of Oxford, Department of Physics

22 June 2023

Darren Moore,

STFC UK Research and Innovation

CPU

- ~470 worker nodes in production.
- ~61K CPU cores
- 84% AMD (Rome/Milan)
- ~780K HEPSPEC06
- Highest spec nodes:
  - 2 x AMD EPYC 7763 64-Core Processor
  - 1TB RAM
  - 6TB NVMe

Batch Farm

- Tier1 testing token submission on ARC-CE's
- Condor 10 testing across Condor pool
- Worker nodes and Condor Central Managers migrated to Rocky 8
- Worker node environment upgraded to Docker 23

Echo

- Currently 265 hosts
- Providing 66.6 PB RAW,~48.4 PB useable
  - additional ~105 hosts to go in each providing ~400TB RAW each

CVMFS

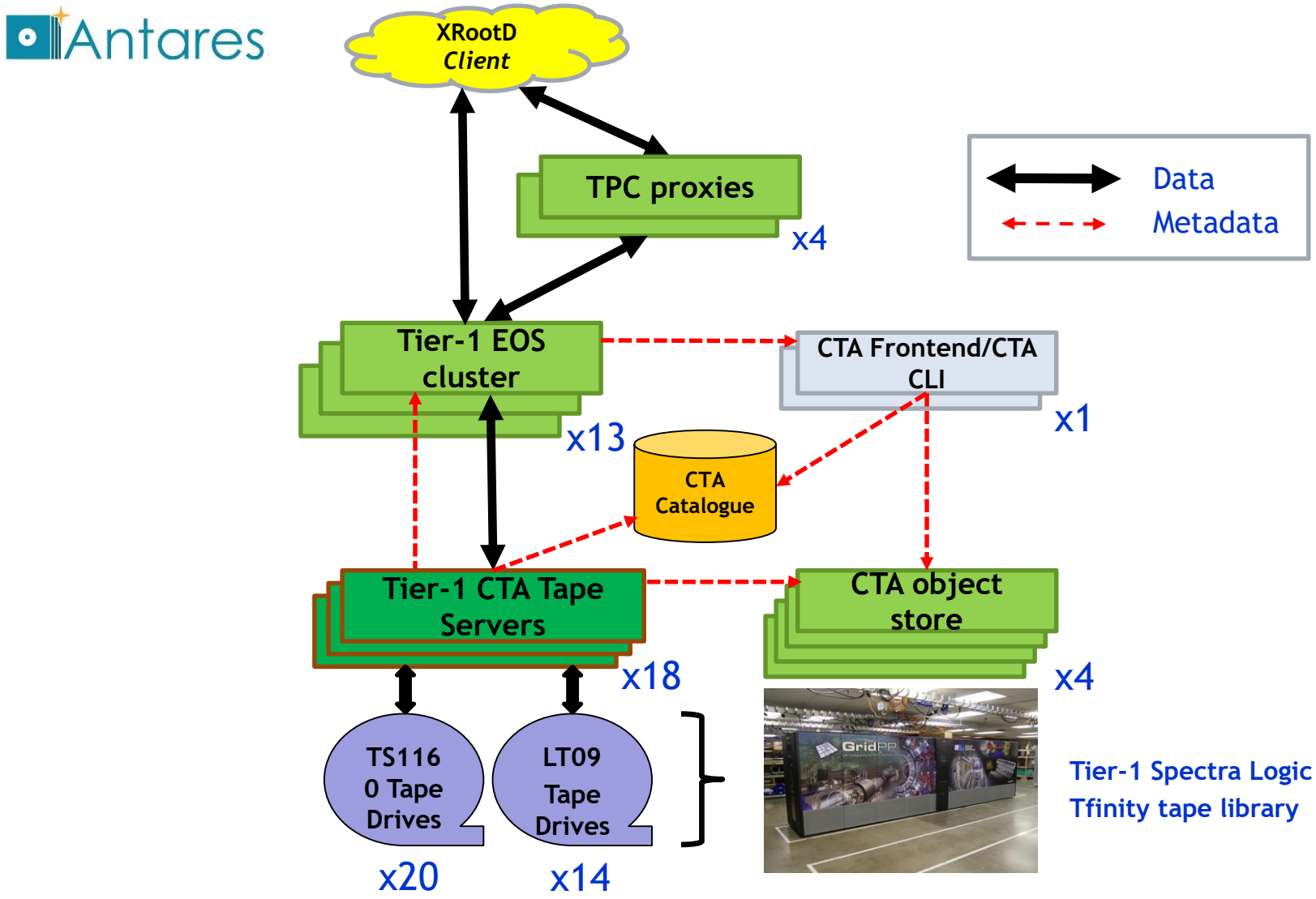
- Stratum-1:
  - 2 hosts, in a High Availability setup
  - each host has 55 TB of disk space, with (at this time) 94% of that used
  - moving to new architecture with all data stored on CephFS
- Stratum-0:
  - service composed by 2 hosts: Uploader and CVMFS server
  - total amount of user data being distributed (at this time): 3.6 TB

Antares

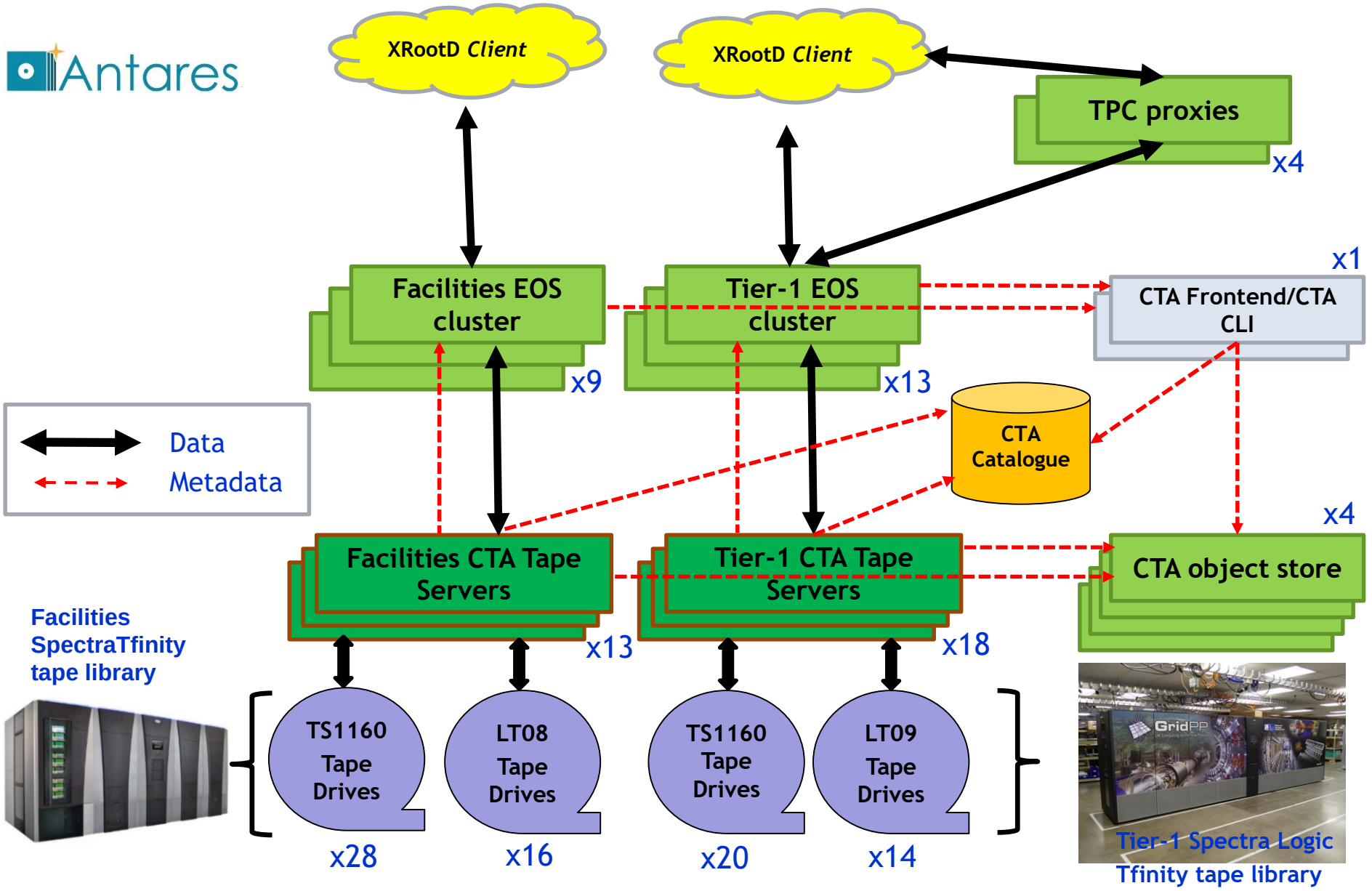
Antares

- CASTOR is dead!…there was much dancing, singing and celebration! No flowers by request.
- Migration of the last CASTOR instance (Facilities) completed last week.
- 2 EOS clusters, Tier-1 and Facilities, against a single CTA instance
- Tier-1 EOS cluster: 13 x 1.5TB SSD nodes
- Facilities EOS cluster: 2 x 3.5TB + 7 x 1.5TB SSD nodes
- EOS: 4.8.88-1,CTA: 4.7.14-1

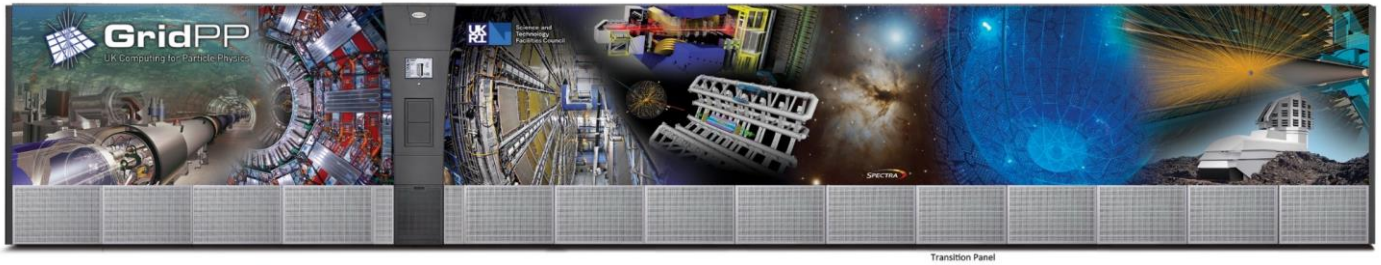
Antares - Current Setup



Antares - Future setup



Tape library (Asterix and Obelix)



- 15 frames - longest in Europe/UK!
- 20 x TS1160, 16 x LTO9
- Capacity: 250PB

- 13 frames
- 28 x TS1160, 17 x LTO8, 6 x LTO9
- Capacity: 212PB

Network

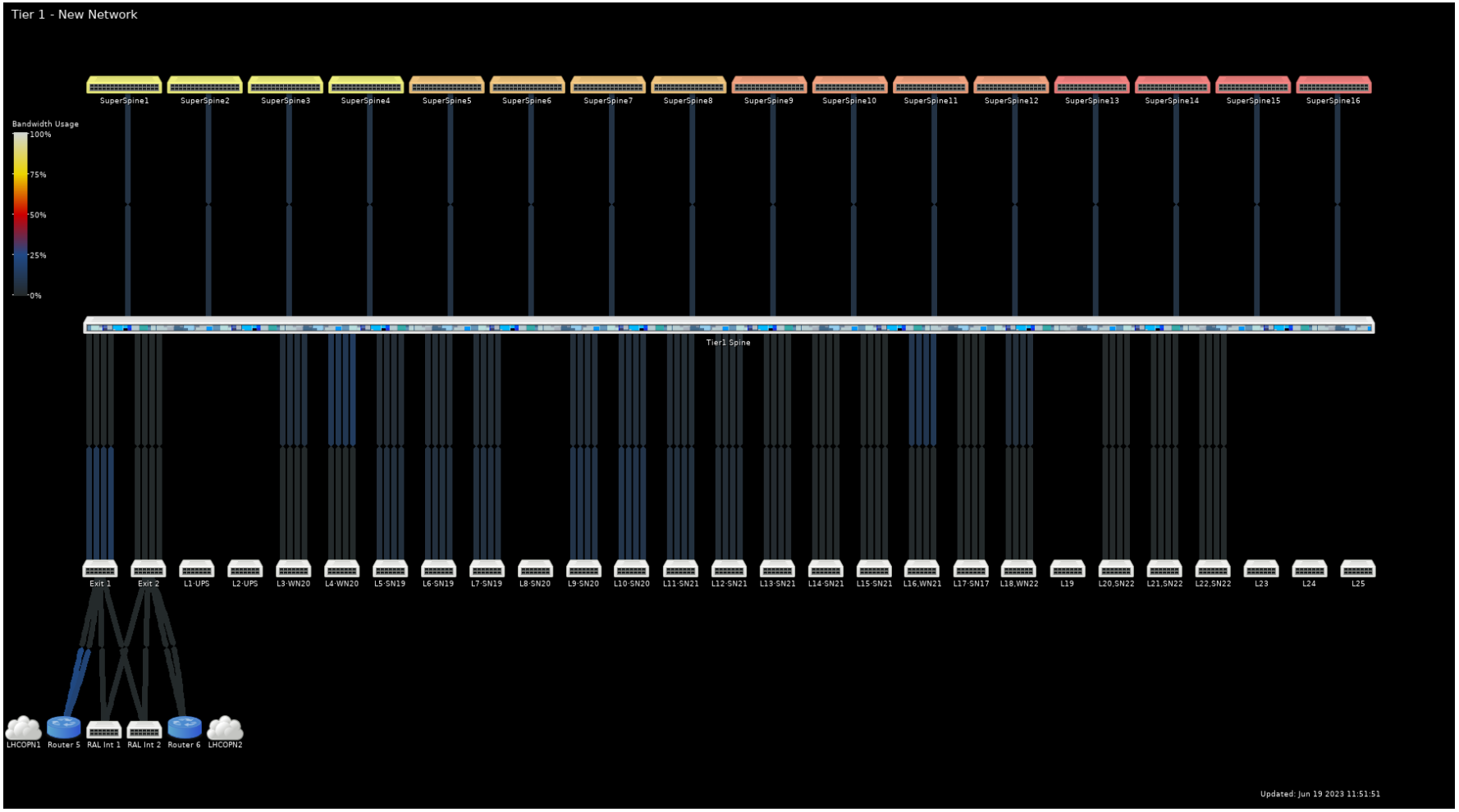
- New network for the Tier-1 alongside the “legacy” network
  - Fully-routed eBGP ECMP architecture
  - Mellanox switches running Cumulus Linux
  - Joined to legacy network by SCD SuperSpine
  - 4×100Gbps to each leaf, 16 ×100Gbps to SuperSpine
  - IPv6 by default
- Started July 2021
- Connected to SCD SuperSpine October 2021
- Connected to RAL site November 2021
- First worker nodes live December 2021

- Legacy network meltdown in October.

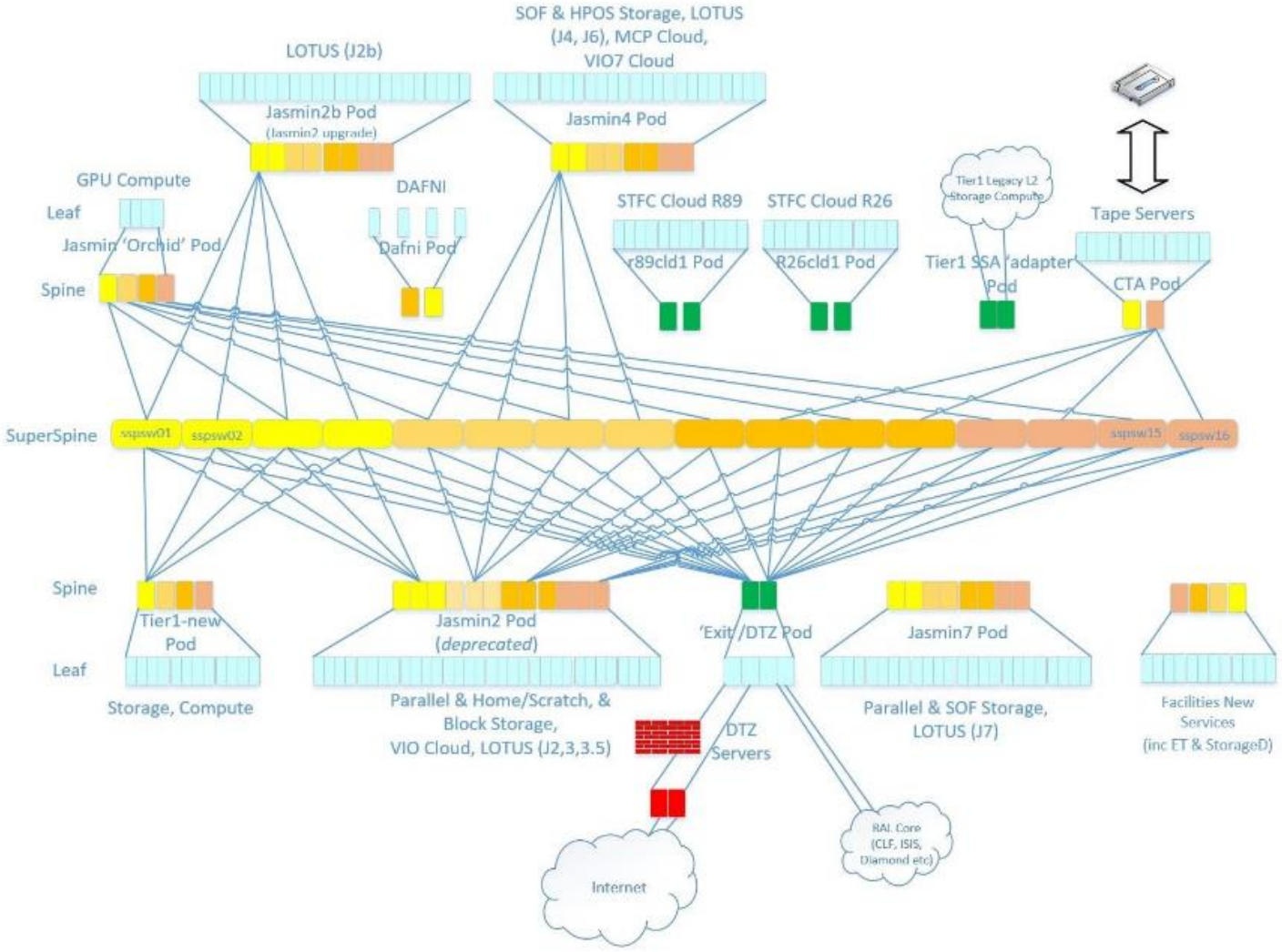
- Second 100Gbps OPN link now active, but not handling traffic yet.

GridPP
UK Computing for Particle Physics

Tier-1 New



Superspine


SOF & HPOS Storage, LOTUS (J4, J6), MCP Cloud, VIO7 Cloud
LOTUS (J2b)
Jasmin2b Pod
(Jasmin2 upgrade)
Jasmin4 Pod
GPU Compute
Leaf
Jasmin 'Orchid' Pod
Spine
DAFNI
Dafni Pod
STFC Cloud R89
r89cld1 Pod
STFC Cloud R26
R26cld1 Pod
Tier1 Legacy L2 Storage Compute
Tier1 SSA 'adapter' Pod
Tape Servers
CTA Pod
SuperSpine
sspsw01
sspsw02
sspsw15
sspsw16
Spine
Tier1-new Pod
Leaf
Storage, Compute
Jasmin2 Pod (deprecated)
Parallel & Home/Scratch, & Block Storage, VIO Cloud, LOTUS (J2,3,3.5)
'Exit' /DTZ Pod
DTZ Servers
Jasmin7 Pod
Parallel & SOF Storage, LOTUS (J7)
Facilities New Services (inc ET & StorageD)
Internet
RAL Core (CLF, ISIS, Diamond etc)

More Miscellaneous

- Migration of Nagios to Icinga
  - Work in progress
- On-going development of InfluxDB
  - Ongoing roll-out as the default time series monitoring.
  - Ganglia still scheduled for decommissioning Dec 2019.
- Security Challenge started 2019-03-15
  - Yes, a Friday afternoon

Icinga

"Migration of Nagios to Icinga
- *Work in progress"*

- Icinga
  - Nagios has "*ceased to be! 'E's expired and gone to meet 'is maker! 'E's a stiff! Bereft of life, 'e rests in peace!"*
  - Icinga now in production for all SCD alerting requirements.
  - Running in a HA configuration
  - The RIG and Tier-1 instances of Icinga are to be merged (Q4 2023)

InfluxDB

"On-going development of InfluxDB

Ongoing roll-out as the default time series monitoring.

Ganglia still scheduled for decommissioning Dec 2019."

- Entered SCD production early 2020 for SCD Time Series DataBase (TSDB) requirements.
- V1.8.x became EOL May 2021.
  - Still being used with various risk mitigations being in place.
- Consultation started end of 2022 on replacement/upgrade
- Replacement to be chosen before Q4 2023.
- Psst! it might be Victoria metrics
- Ganglia is no longer a production service, it has been suggested there are one or two instances left in the Badlands of SCD.

Security

"Security Challenge started 2019-03-15

- *Yes, a Friday afternoon"*

- *Security Challenge started 2023-03-23*
  - *This time it was a Thursday*
  - *Intrusion was first detected 2nd day of GridPP47.*
  - *Generally considered to be a successful challenge.*

Cloud

- The STFC Cloud is an IaaS platform based on OpenStack (running on Rocky 8).
- Open to users across STFC facilities, IRIS and other partner organisations it has seen enormous growth over recent years with over 2000 users across over 500 active projects.

Compute: CPU

- ~64k logical cores
- ~300 TB memory
- ~680 HVs of various types, from 28 cores with 90GB of RAM to 128 cores with 2TB of RAM

Compute: GPU

- 55 x rtx4000 GPU servers (4 x rtx4000 NVidia cards in each) – ideal for visualisation
- 32 x v100 GPU compute servers – (4 x v100 NVidia cards in each) – for compute
- 28 x a100 GPU compute servers – (4 x a100 NVidia cards in each) – for compute
- 17 x a4000 GPU compute servers

Cloud

Storage:

- For volumes – 181 of 231 TB used (185TB is the 80% threshold) - Sirius
- 1304TB for “local disk” VMs – just on compute
- Manilla service (Arided) – 400 TB available

Network:

- Hypervisors and storage connected at 10 or 25 gb depending on age
- 4x40gb uplink to RAL site core.
- Virtual network provided in two ways:
  - Internal on the STFC network for internal users. - Better performance
  - Private Project networks (for external users or access outside of STFC)

The STFC Cloud also offers:

- Cluster API support for Kubernetes deployment
- Jupyter training deployments for users, currently providing approximately 1000 training days per year.

Cloud

What's coming online soon:

- 9 x supermicro a4000 hosts– 8 X NVIDIA a4000 cards in each GPU node for remote visualisation (about to be delivered)
- 34 x Supermicro 2021 CPU compute – 2 x AMD epyc 7763 (256 logical CPU), 2TB RAM, 1 x 7.68 TB local SSD, 25Gb/s NIC
- 104 x Lenovo 2022 CPU compute - 2 x AMD epyc 7763 (256 logical CPU), 1TB RAM, 100Gb/s NIC
- More Deneb and Arided (Manila Cloud service) and Sirius storage.
- Migration to kola-ansible and Scientific OpenStack
- Deployment of Ironic OpenStack component to offer baremetal support

GridPP
UK Computing for Particle Physics

Any Questions?