

# GridPP

UK Computing for Particle Physics

## RAL Site Report

HEPiX Spring 2022

25-28 March 2022

Martin Bly,

STFC UK Research and Innovation

- DC
- Networks
- Procurements
- Ceph
- Tape
- Cloud

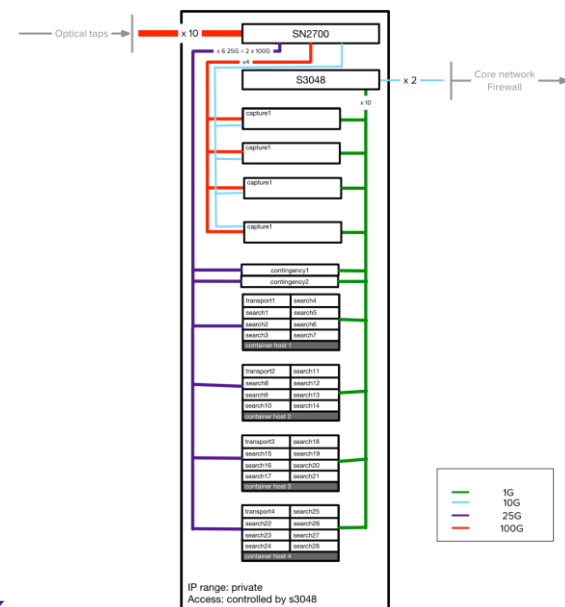
Thanks to colleagues for contributions

- DC operations back to pre-Covid level with small number of procedural changes track where work has been done
- UPS:
  - Additional UPS battery capacity installed to provide extra capacity
  - Work on rationalising the distribution of UPS sockets under racks to provide UPS where needed due to limited number of ways on distribution boards
    - Plan this year to install additional distribution board(s)
- New DC planning
  - The STFC Research Computing Centre (RCC) is currently going through RIBA stage 2 design and is subject to UKRI approval
  - The aim is to provide an energy efficient scalable design with a target PUE of 1.1
  - Phase I of the RCC is planned be operational mid 2025 with an IT Load capability of 3.75MW with Phase II following shortly after by 2028 at 6-11MW depending on requirements

- LHCOPN RAL to CERN on 100Gbps circuit, no ‘standby circuit’
  - Failover is via public Internet
  - Link to border doubled to 80Gbps pending roll out of 100Gbps routers
- Work to commission and join LHCONE progressing
  - Delayed by site network plans
- Tier1 Leaf/Spine network rollout in progress
  - New hardware generations joining L/S network
  - Some older hardware to be migrated to L/S
  - Oldest hardware to remain on legacy network to be decommissioned

- **Macro-Segmentation Project**
  - Point-to-point link routing
  - Initial gradual rollout late autumn 2021 was unsuccessful
    - Highlighted a number of hidden issues
  - More planning and testing required
  - Updating of router firmware
  - Scheduled big-bang rollout early summer (June '22)
- **Cyber Security is a major focus for UKRI and STFC**
  - Addressing security through auditing, with other projects in process
  - Priority is patching and patching policy
  - Emphasis on systems not covered by well established procedures and tools

- Monitor all STFC-RAL traffic and correlate with threat intelligence
  - Vital for STFC cybersecurity development
- Monitoring will include
  - 2x100Gb/s Janet
  - 1x100Gb/s LHCOPN links
- Hardware and networking in place
  - Optical taps available when needed
- See [Collaborative Incident Response](#) talk
  - <https://indico.cern.ch/event/1123214/contributions/4844820/>



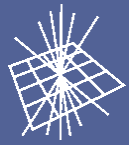
- Compute for FY21/22, joint procurement for Tier1, Cloud and JASMIN: 92 for JASMIN, 48 for Tier1, 30 for Cloud
  - 1U SuperMicro servers
    - 2 x AMD EPIC CPUs: Tier1 and cloud: 7763, JASMIN: 7643
  - 1TB RAM (Tier1), 2TB RAM (cloud), 512GB (JASMIN)
  - 7.68 TB Enterprise SSD (Tier1, Cloud), 480GB Enterprise SSD (JASMIN)
  - Mellanox ConnectX-5 10/25GB DP NIC
- GPUs (cloud)
  - 15 x 4U SM servers each with 4 x Nvidia Tesla 80GB A100 GPUs
  - 15 x 1U SM servers each with 4 x Nvidia Quadro RTX4000 GPUs
- Switches
  - Mellanox SN2410, SN2100
- Storage
  - Tier1 ECHO: 80 x Dell R740XD2 w/24 x 18TB SAS HDDs
  - Cloud: 15 x Dell R7525 w/24 x 3.84TB SATA RI SSDs
  - CTA EOS: 3 x Dell R7525 w/24 x 3.84TB SATA RI SSDs
- Significant delays to deliveries, installation and testing due to parts shortages, mainly high bandwidth NICs. Lead times now up in the 6-9 month range for some components
- Starting procurements for this FY now

- Echo is a large Ceph cluster, ~230 nodes, erasure coding
  - Currently deploying two generations of hardware on new leaf-spine network: adds 28PB raw, 20.3PB Useable at 8+3 EC
  - New capacity will be 79PB raw, 57.5PB useable.
  - WebDAV access now enabled on Echo, preparing for Run 3
- Backend Cloud services stable
  - New mountable CephFS storage - ‘Arided’ - to be deployed Q3 2022 to relieve capacity pressures on private cloud
- Investigating deployment with Cephadm
  - See talk by Kyle: “A new Ceph deployment using Cephadm at RAL”
    - <https://indico.cern.ch/event/1123214/contributions/4844818/>



- Castor service for Tier1 terminated
  - Continuing to run a small instance for local facilities users
- CTA service (Antares) now in service! (7th March)
  - All Tier1 data
  - c.f. Talk by George Patargias
    - <https://indico.cern.ch/event/1123214/contributions/4843405/>
- Libraries (SpectraLogic Tfinity)
  - Tier1 library expanded to 10 frames
  - Facilities library expanded to 11 frames
  - Space for additional expansion to 13 frames each
  - LTO8 and TS1160 used in production services
    - Tier1 currently using TS1160
  - LTO9 in commissioning for Tier1
    - LTO new to Tier1

- **Commissioned**
  - 1 x server with 4 x Xylinx U200 to allow use of FPGA tools in VMs
  - 2 x Mellanox SN4600C (64 x 100Gbps) spine switches
  - 8TB of RAM upgrades to selection of hypervisors to create large memory capability
  - Moved Rabbit service to new HW to handle additional load
- **Being commissioned**
  - 14 x servers with 4 x Nvidia Tesla V100 GPU, 384GB RAM, 240GB SSD
- **Future**
  - 30 x 1U compute nodes with 256 vCPU each
  - 15 x systems with 4 x a100 GPUs each, 100Gbps NIC and large local SSD
  - 15 x systems with 4 x rtx400 GPUs each, 25Gbps NIC, for visualisation
  - 'Arieded' Ceph cluster - 15 x R7525 with 2 x EPYC 7413, 256GiB RAM, 2 x SSD (system) and 24 x 3.84TB SSDS (data), 25Gbps NIC



**GridPP**

UK Computing for Particle Physics

Questions?