

GridPP

UK Computing for Particle Physics

RAL Site Report

HEPiX Fall 2016 - LBNL Berkeley

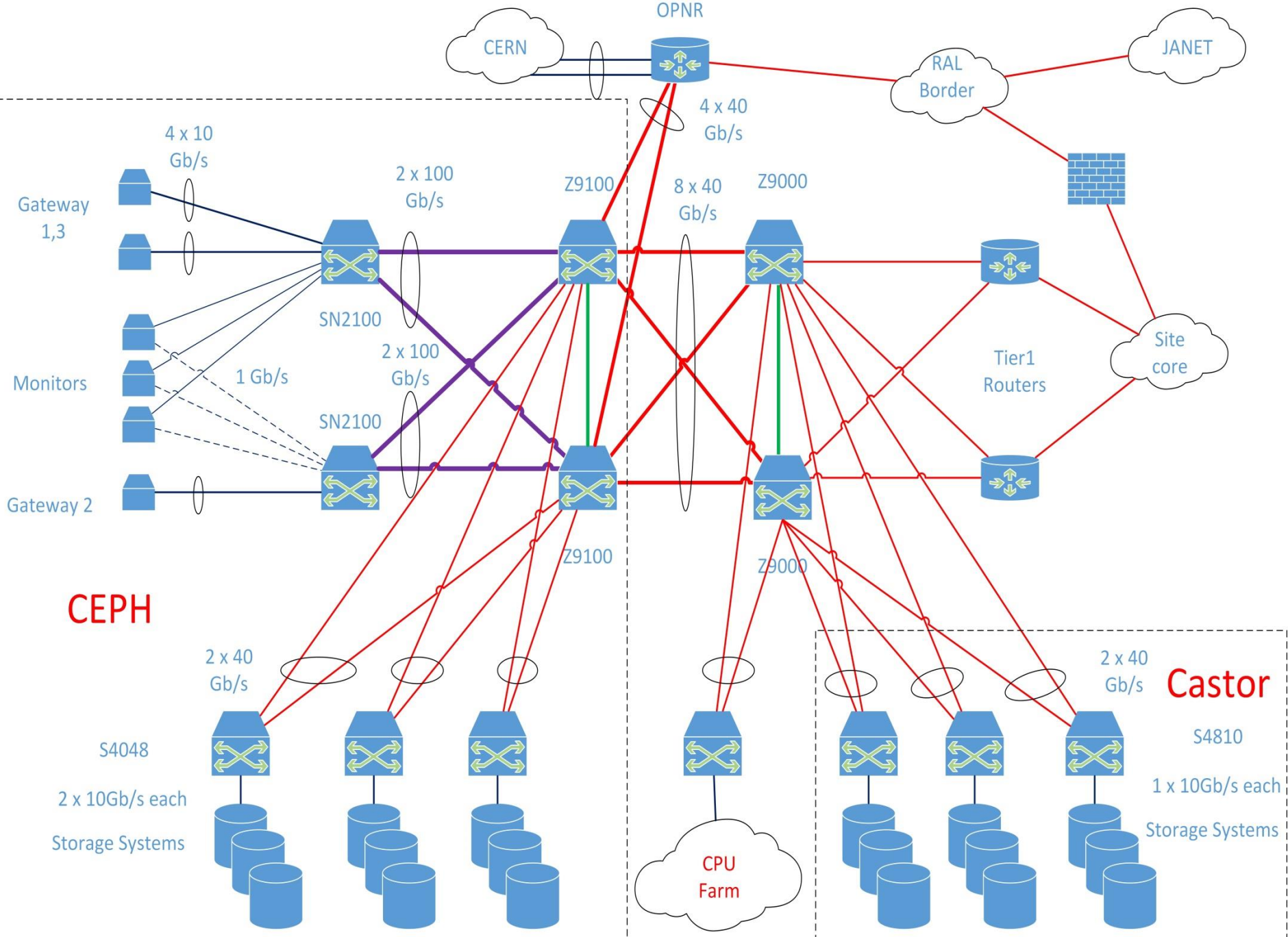
17-21 October 2016

Martin Bly, STFC-RAL

- Hardware
- Networking
- Services
- Storage
- CVMFS

- CPU: ~240k HS06 (~24k cores)
- Storage:
 - ~16.5 PB useable in Castor
 - ~13.3PB raw for Ceph
- Tape: 10k slot SL8500 (one of two in system)
 - 44PB (T10KC/D)
 - Migrations to D-only:
 - Atlas: 1364 tapes, ~6.5PB - completed, ~4 months
 - LHCb: 600 tapes, ~3PB - waiting installation of new D drives
- Procurements
 - To come...
 - At least 5,900 HS06 and ~2PB

- Tier1 WAN
 - UKLR - R.I.P. 😊
 - Replaced with S4810P stack, aka the OPNR
 - 40Gb Active/Passive firewall bypass to JANET
 - New record for off-site (all RAL) traffic @ 40Gbs
 - 2 x 10Gb Active/Active to CERN (OPN)
 - relieving saturation, but still occasional congestion
 - 4 x 40Gb to Tier1 core
 - PerfSONAR packet loss still with us
- LAN
 - Core expanded: 2 x Forec10 z9100, 32 x 40/100Gbs
 - Gateway facility for Ceph - 2 x Mellanox SN2100, 16 x 40/100Gbs
 - Redundant 100Gbs links to core
 - Ceph gateways to be 4 x 10Gbs
- IPv6
 - RAL network team now engaged to provide infrastructure for full site IPv6 capability (DNS, FW etc).
 - OPNR provides more options for Tier1 IPv6 work to meet the wLCG targets



- **Load balancers**

- Since April have been using load balancers based on Keepalived & HAProxy in front of FTS3
- Also used for OpenStack cloud (under development) and to be used for some services associated with Ceph (e.g. Dynafed)
- Talk on Thursday
<https://indico.cern.ch/event/531810/contributions/2314174/>

- **Monitoring**

- Using Telegraf, InfluxDB & Grafana to monitor over 900 hosts (worker nodes, cloud, HyperV hypervisors, some grid services)

- **Mesos**

- Investigating running services & compute resources (e.g. worker nodes) on Mesos
- Talk on Friday
<https://indico.cern.ch/event/531810/contributions/2328936/>

- v2.1.4
 - Continues in stable operation
- v2.1.15 in test
 - Testing nearly completed
- Starting integration work with the new Echo (Ceph) service

- New data cluster with the aim to replace disk-only Castor and other 'disk' services
- HW:
 - 63 data nodes, 36 x 6.0TB ea, 13.6PB raw, 9.9PB useable, 8+3 erasure coding,
 - 3 monitors and 3 gateways for S3/Swift, XrootD and GridFTP
- SW:
 - SL7, ceph "jewel" release
 - latest release, fixes erasure coding bugs present in "hammer"
 - XrootD 4.4.
- Network:
 - Public and Private data network for each storage node
- See talk for details
 - <https://indico.cern.ch/event/531810/contributions/2298934/>

- **Stratum-0 (part of the EGI Infrastructure)**
 - new server ready to replace existing server (VM)
 - 32GB RAM, 12TB disk, 2x E5-2407 @ 2.20GHz
 - cvmfs v2.3.2-1
 - - 28 repositories
 - will soon look into setting up config-egi.egi.eu repository
 - new 'africa-arabia.org' CVMFS domain to be supported via this mechanism
- **Stratum-1**
 - replicates many cern.ch repositories, also egi.eu and nikhef.nl domains)
 - 2-node HA cluster currently cvmfs-server v2.2.3-1
 - Recent upgrade to HW
 - 64 GB RAM, 55TB storage, 2x E5-2620 v3 @ 2.40GHz

- CVMFS Uploader
 - provides upload area for egi.eu repositories
 - GSI interface (gssh, gscp)
 - rsync mechanism between Stratum-0 and Uploader
 - soon to be replaced by the cvmfs-rsync tool
- CVMFS clients
 - upgraded batch workers in Summer to v2.2.3-1
 - Reverted back to v2.1.20 in September because of series of issues at CVMFS clients and at squids level
 - plan to upgrade to v2.3.2-1 by end of year
- CernVM Users Workshop (6-8 June) at RAL well attended
 - <https://indico.cern.ch/event/469775/>
- Plan to look into 'protected CVMFS repositories', some use cases at STFC/EGI - will probably contact OSG colleagues