

The background of the slide is a complex, abstract visualization. It features a dense network of thin, multi-colored lines (yellow, cyan, red, green) radiating from a central point, creating a starburst or network-like pattern. The lines are set against a dark, multi-colored background with shades of blue, purple, and red. The overall effect is that of a data visualization or a complex network graph.

# RAL PPD Site Report

HEPSYSMAN Summer 2019

24<sup>th</sup> May 2019

Chris Brew

Ian Loader



Science & Technology  
Facilities Council

# List of things to try

- CentOS7 + HTCondor + Docker + Singularity WNs
- FreeIPA
- Expand GPU test boxes into a production service
  - CentOS7 + Docker to allow for different users
- PerfSONAR Refresh – One box to rule them all
- Containerized reconfigurable cluster
- Puppet4
- NoMachine
- LHCONE



# Network/LHCOne

- RAL site connection now 2/100Gb/s, but next hop is still 2x40Gb/s
- Little change to Departmental physical network
  - Replaced Nortel/Avaya 55XX/56XX stacks for top of rack switches with Dell N1548 switches
- Deployed IPv6 on all grid storage, grid service and last batch of worker nodes
- Will join LHCOne shortly. Whole Grid subnet(IPv4/22, IPv6/64) will be in the DTZ.
- New IPv4/24 for stuff that cannot join LHCOne



# Accelerated Computing

- 4 GPU Nodes in batch cluster, no ssh access fo
- hepacc03
  - 4 x Tesla P100 connected by NVLink
- hepacc04-06
  - 2 x Titan RTX
- Interactive use by interactive jobs
  - `condor_submit -i \`
    - `-a 'request_gpus=[1..4]`
    - `-a 'request_cpus=[1..16]'`
    - `-a 'request_memory=[1..128000]'`
    - `-a 'requirements = (CUDADeviceName =?= "TITAN RTX")'`
- Interested in making grid available
  - Only P100
  - Via test HTCondorCE



# PerfSONAR Refresh

- 1 Box, 5 names, 5 Interfaces
  - Data Transfer Zone
    - hep-ps-bw-bp, 10Gb/s, IPv4+IPv6 Bandwidth
    - hep-ps-lt-bp, 1Gb/s, IPv4+IPv6 Bandwidth
  - Through Firewall
    - hep-ps-bw-fw, 10Gb/s, IPv4 Bandwidth
    - hep-ps-lt-fw, 1Gb/s, IPv4 Bandwidth





# The Department

- ~ 70 Staff
  - Joint Appointees, Visitors, Staff on LTA at CERN, UK Liaison Office at CERN
  - 3 Main LHC Experiments – Atlas, CMS, LHCb
  - Numerous “Small” Experiments e.g. Dark Matter, DUNE, T2K, Hyper-K, EDM
  - Hosting Department for MICE
  - Remit to provide support for UK Particle Physics
  - GridPP Tier 2 site



# The Team

- Team of 5:
  - Chris Brew (Management, Linux)
  - Ian Loader (Linux)
  - Kevin Dunford (Windows)
  - Will Pilcher (Windows + Linux)
  - Obinna Utobo (YinI, Windows)
  - Apprentice Next Year



# Core infrastructure - Rooms

2 Machine Rooms (22 Racks)



Remote – 16 Racks in the Atlas Centre

- Core Network, Disk, Workers, Xrootd, Reverse Proxies, PerfSONAR, HyperVisors etc.

Local – 6 Racks in the R1 basement

- Everything else, including Windows servers, Interactive Linux, NFS, Mercury, bits & bobs





# Core Infrastructure - HyperVisors

- Windows 2012r2 Failover Cluster
  - 8 Hyper-V Servers (6 Production, 2 Test)
  - 6TB Network Attached Storage
  - Now Hosting 60 VMs
  - Deploying Windows 10 by PXE and 3<sup>rd</sup> party S/W via SCCM



# What we run

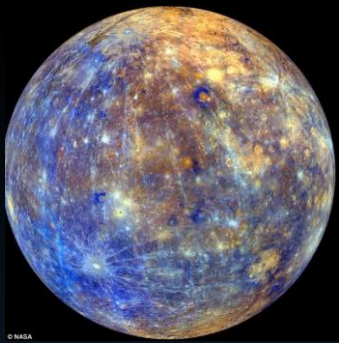
- Condor Batch System
  - Used by both Grid and Local Users
  - 172 (+8 +4) Nodes, 5060 (+32 +32) Cores, 10GPUs 54,000 HS06
  - Multicore Enabled, used by CMS and Atlas
  - 12 new nodes (+768 cores, +8,600HS06)
  - Almost all C7 (11 nodes, 240 cores SL6)
- dCache Storage
  - dCache 3.2
  - 59 Storage Servers, 4.2PB includes 20 old servers reprovisioned as 2 copy resilient pools for local storage.
  - 5 out, 5 in 20x12TB disks as single raidz2 ZFS dCache pool



# End User Devices

- Laptop + Docking Station rather than Desktop + Laptop
- Windows 10 + SCCM
- Mac, will be joining another Depts. JAMF Service
- Linux SL6, CentOS7 or Ubuntu in our LDAP/NFS Domain
- A handful of self “managed” other flavours of Linux





# Mercury

## Fast Analysis Cluster

- 8 node BeeGFS Clusters with 45TB of shared storage (20TB rust + 25TB SSD) cross mounted
  - 5 x 2 Intel E5-2430 v2 (6+6@2.5GHz) and 64GB RAM
  - 3 x 2 Intel E5-2620 v4 (8+8@2.1GHz) and 132GB RAM
- Used for data intensive work
- 4 batch slots on each, only for local jobs



# Storage

- Home Area – VM with iSCSI NAS
- Software Area – NFS mount of ZFS volumes rsync'd to second server (want to look at DRBD)
- Data Area – 400TB of old dCache space, still in dCache, NFS4 mounted, 2 copies of files
- Two very old (2006) ex-dCache NFS Disk servers for scratch





# Monitoring

- Ganglia
- Nagios
- LibreNMS – interest in Central Service
- Pakiti – Currently v2, interest in central v3
- ELK
  - dCache Billing via FileBeat
  - Metric Beat?
  - Condor/ArcCE (please)
- PuppetBoard





Science & Technology Facilities Council  
Particle Physics Department