GridPP

UK Computing for Particle Physics

RAL Site Report

HEPiX Fall 2019

NIKHEF – Amsterdam

14-18 October 2019

Martin Bly,

STFC UK Research and Innovation

Overview

- Scientific Computing Department
- Tier1 developments
- SCD Cloud
- Miscellany

Thanks to colleagues for contributions

Scientific Computing @ RAL

- 180 staff, 7500 users
  - Applications development and support
  - HPC & HTC Compute and data facilities
  - Systems administration, data services, numerical analysis, software engineering
- Tier1 for LHC
- JASMIN
  - Data Intensive Computing for climate and earth-system science
- DAFNI
  - Hub for national infrastructure analysis, modelling and simulation
- At Daresbury Lab:
  - Computational Science and Engineering
    - Biology and Life Sciences
    - Theoretical and Computational Physics
    - Computational Chemistry
    - Engineering and Environment

Tier1 Hardware

- CPU: ~316k HS06 (~31.5k cores)
- Castor disk-only: ~550TB (!) remaining
  - Retirement progressing, aim to switch off by end of 4Q19.
- Ceph: ~43PB raw / ~31PB usable
- VMWare:
  - 10 node cluster, two storage arrays, iSCSI
- Tape:
  - 10k slot SL8500 (one of two in system)
    - 80PB capacity (T10KD), ~52PB physics data, (~110PB total total)
    - ~70 T10K-B/C/D drives in system
    - Average ~110 tape mounts per hour
  - Spectra Logic Tfinity 7-frame library:
    - 1 master frame (includes drive bays) ; 2 end units; 1 drive frame; 3 media frames
      - 550 chambers, each holds 10 x LTO or 9 x TS1160 media, 60PB media installed
    - Drives:
      - 17 x LTO8 drives (one for test/verification/spare)
      - 16 x TS1160 drives
    - Just in production, ~300TB so far
- 2019/20 procurements for Storage and Compute in progress

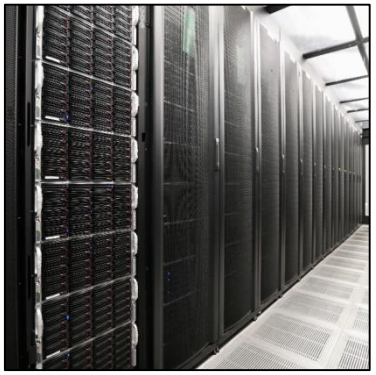
Castor

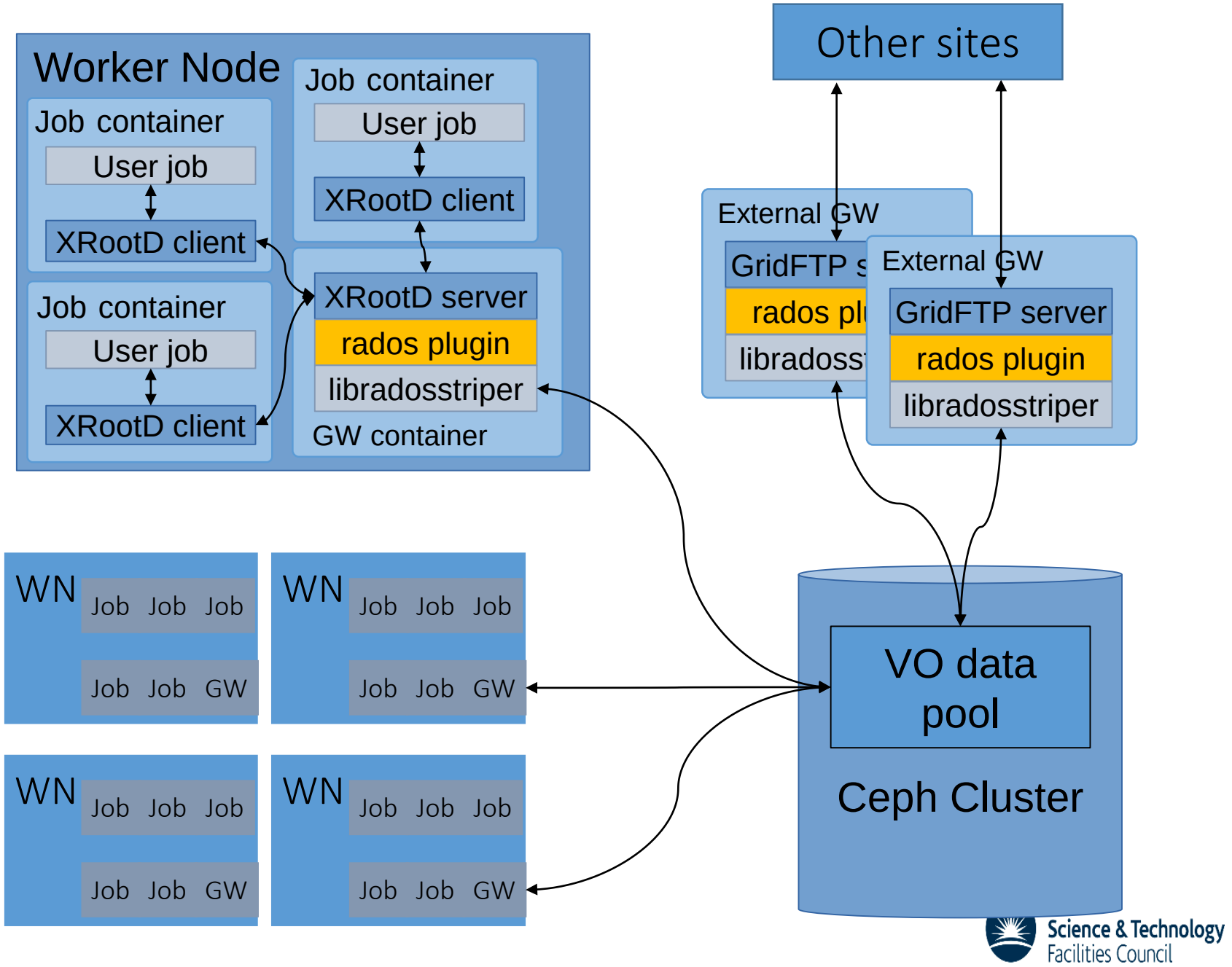
- Past 6 months:
  - Continued work to decommission ‘disk-only’ CASTOR pools in favour of Ceph-based Echo storage
    - ATLAS, CMS, Non-LHC users – previously completed
    - LHCb completed 7 May 2019
    - ~~ALICE 2nd Feb 2019~~ ‘by end 4Q19’
  - (New) Tape cache pool:
    - Full production service
    - All users migrated by early May 2019

- For more detailed look at the evolution of the tape archival service, see George’s talk on Thursday:
  - https://indico.cern.ch/event/810635/contributions/3593326/

Echo Storage

- Big Ceph cluster for Tier1 object storage (and other users)
  - Density and throughput over latency
    - EC 8+3
    - 64MB rados objects
- 21PB for Tier1, capacity 31PB useable
- Stable running, no major incidents
- Version: Luminous
  - Upgrade to Mimic is pending a required bug fix
  - Previous issues with OSD data balancing full resolved by upmap balancer




Worker Node
Job container
User job
XRootD client
Job container
User job
XRootD client
Job container
User job
XRootD client
XRootD server
rados plugin
libradosstriper
GW container
Other sites
External GW
GridFTP s
rados pl
libradoss
External GW
GridFTP server
rados plugin
libradosstriper
WN
Job Job Job
Job Job GW
WN
Job Job Job
Job Job GW
WN
Job Job Job
Job Job GW
WN
Job Job Job
Job Job GW
VO data pool
Ceph Cluster
Science & Technology Facilities Council

Batch Systems

- ~31,500 slots over ~800+ nodes providing around 316,000 HS06.
- Running HTCondor
  - Starts jobs inside Docker container images that we maintain
  - decouples the execution environment from the base OS
  - Very close to making CentOS 8 images available for jobs to run in
- Backfilling to SCD cloud now automatic and working well, generally at least 600 cores, often 1200-1800 cores
  - Working on exposing GPUs for use
- Central suspension via Argus now applied to all services

SCD Cloud

- Cloud activities now within a new group within Scientific Computing Department focussed delivering dynamic infrastructure and cloud services to users
- IaaS platform built on OpenStack
  - ~9k vCPU cores
  - Users 160+ projects
  - Several flavours of GPUs for Virtualisation or Compute
- Delivers dynamic compute resources to scientists across STFC and externally
  - LSST, EUCLID, AENEAS, UKAEA
  - ISIS (neutron source), DLS (Diamond Light Source), CLF (Lasers), ASTEC, UKATC, RALSpace, Hartree
- Recruiting for a systems administrator, advert up soon, contact alexander(dot)dibbo(at)stfc(dot)ac(dot)uk

Networking

- IPv6 issues:
  - Tier1 firewall bypass now directly connected to 100Gb/s site border routers, but only at 40Gb/s
    - Didn't solve IPv6 packet loss issues ☹
  - JANET (September) found a bad router card which was replaced
    - This fixed the IPv6 problems from the Tier2s ☺
  - After a problem with the Tier1 site-internal IPv6 traffic, one of the site core switches was rebooted.
    - This fixed the remaining problem for T1 to PPD-T1 traffic ☺
- GridPP is developing a new standard hardware config to refresh the UK's PerfSonar infrastructure and would like input from the community
  - Particularly if anyone has evidence that specific CPU architectures make a significant difference.
  - https://indico.cern.ch/event/849677/contributions/3570742/attachments/1911907/3159545/GridPP-perfSONAR-refresh-01.pdf
  - See me…

Grid Services

- Short staffed
  - Catalin moved on to EGI, other staff changing roles within SCD
- Most grid services are being supported, but not much development going on
- Vacancies for two Grid Service SysAdmin posts which close the Monday after HEPiX (21st October):
  - https://www.topcareer.jobs/Vacancy/irc250600_9896.aspx
- (See James Adams)

Other activities

- Participating in WLCG Security Operations Centre WG
  - Lead by David Crooks at RAL
  - C.f. David's presentation on Wednesday:
    - https://indico.cern.ch/event/810635/contributions/3593242/

- Enterprise Virtualisation
  - Migration off Hyper-V cluster to VMware cluster completed
  - Migration of standalone Hyper-V systems to be completed during November

Questions?