Site Report HPC2N

NDGF All Hands Kastrup

2025-11-04

New HPC2N LHC OPN switch(router)

- Purchased together with new/replacement dCache disk pools
- Dell S5448F-ON 48x100GbE, 8x400G QSFP56-DD
- Will be used as a 25G/100G switch to begin with
  - Aiming for 400G in a pilot project later on
- Quirks so far
  - PSU has C15 (high-temperature) connectors, shipped with C15-Shuko cables...
  - Shipped with SONIC, we purchased it with OS10
  - Squealing fans, don't know yet if it's broken speed control or...
  - Needs front-to-back airflow to handle many 400G modules
- SUNET still wants us to do BGP, will migrate old stuff first

Tape/backup

- Same as last time
- IBM TS4500 library (2550 slot capacity)
- 6x TS1155 tape drives (JD tapes, 15T, 360 MB/s)
- 6x TS1170 tape drives (JF tapes, 50T, 400 MB/s)
- Dell R750
  - 2x100G Ethernet
  - 4x32G FC
  - Approx 30T NVMe for DB and incoming stgpool
  - A few TB of SAS SSD for log mirrors etc
  - Approx 250T spinning disk for on-disk backup storage
  - 256 G RAM, 2xIntel Gold 6334 (total 16 cores @ 3.6 GHz)

New dCache disk pools

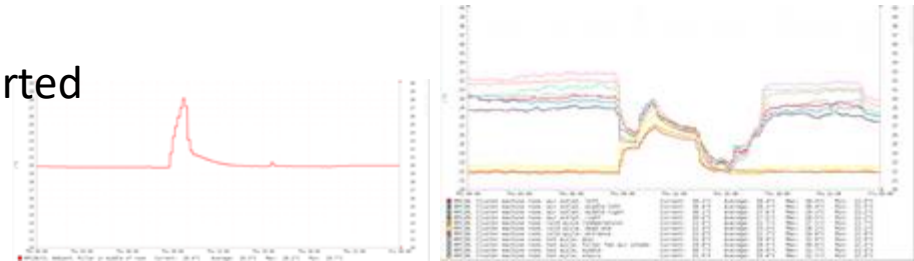
- 7x Dell R760xd2
  - 28x24T HDD, PERC H965i, 128 G RAM, 2x 12core Xeon Silver 4510
- Slow and confused delivery, but finally got all Dell stuff
- But that doesn't matter much since APC/Schneider racks are two weeks delayed.
- Amusing myself with preparing as much as possible for quick deployment when we get the racks/PDUs...

WLCG compute – Aigert (aka g-ce01)

- Same as last time
  - 31 PowerEdge R6625, 256 cores/node, 3GB/core, 3.2TB/node
  - OS – Ubuntu 22.04 Jammy
    - Only one minor service host (grid-home & user mapping) still focal
  - ARC 7.1
- Datastaging still stalling with ARC 7.1
  - Nordugrid bug 4191
  - Reduced core count since load seems to exacerbate the problem
  - Maybe the ARC disk cache is too small?

This should not affect us … ☺

- Change of electric substation for our building was planned for 2025-07-03
- Since our cluster machine room (C-room) has its own substation and the equipment in the infra-structure machine room (I-room) would go on diesel power, we should not be affected
- It did not go as planned
  - In the morning, we noticed that HPC2N was not accessible. Later onsite the two not on vacation discovered C-room had no electricity., The I-room did have UPS electricity but still the (main) network equipment had no power. And heat was slowly increasing
  - Start to scramble to contact people on vacation. Figured out the power cables to the switches were plugged in to the wrong power outlets. Got power to switches. But still no network.
  - Just after 11 has the battery to the card reader to the I-room ran out. Had to use emergency handle to get out which left to machine room unlocked. Got campus security to lock it.
  - Close to lunch, managed to contact central IT (in person). After some inspection they find problems in switch configuration, and they got the network up. Started updating our web of the incident and the progress
  - In the afternoon redundant power comes back. Lights work in C-room but still no power to the machines. Electrician comes by and help turn on the power. With power back, people (many on vacation) starts working on getting servers/services up
  - 17:00 Most of our services up and running again
- So, in the end, all our services was affected and all running jobs were aborted



It did affect us …

- What did happen
  - UMU switches in I-room was not connected UPS/diesel power due to an earlier temporary rewiring to have power at an earlier stage . Likely a miss in communication between nikke and UMU central IT people managing those switches.
  - Unclear why C-room lost power. Most likely a combination of the fact that we also lost cooling tripping the hard cutoff and later Vattenfall deciding that it was a good time to do revision of the upstream station.
- Unable to get information to our user
  ⇨ Figure out a way to inform our users without depending on our own infrastructure 🫣
- Unclear what (network) equipment was ours and what was central ITs
  ⇨ Put stickers on their network equipment ✓
- Took too long to contact central IT
  ⇨ Document contact points to central IT network people ✓
- When not full staff, low knowledge of power, cooling, network etc not present
  ⇨ Quarterly walkabouts in the data centers ✓
  ⇨ Improve documentation of big picture infrastructure and dependencies 👷
  ⇨ Automatic push of mkdocs-wiki repository from git server to appropriate servers 👷
- The battery for card readers had way too short run time
  ⇨ Issue reported and being corrected, turned out to be a problem in the entire building (should last 24h at least) ✓