Tier 2 Operations during Lockdown.

Matt Doidge, for **GridPP45** - 20 October 2020
with help from Daniela and Simon.

Working from home is nothing new for us.

- Between kids, pets, waiting for home deliveries and good old fashioned things breaking out of hours we've all done more than our fair share of working from home.
- For me it's meant a shorter commute, a different office mate but the same lunchtime companions.

Commute not pictured as no-one wants to see that.

Not without some hitches.

- My biggest gripe by far is working with the remote console.
  - Moving to wired (over power) networking at home helped ease this a bit.
  - But still almost unusably laggy a lot of the time.
  - Console access is X-forwarding a browser from a cluster machine hooked into our separate IPMI network, so maybe this is the best we can expect?
- Loss of “casual communication”.
  - The bandwidth from bumping into people around the department was surprisingly high.
  - Alleviated with the adoption or uptick in the use of (many) chat channels.
    - Mattermost, keybase, Teams, just plain hanging out at the end of virtual meetings.
- The commute may be short, but an emergency response wouldn’t be.
  - At the height of lockdown buses were hourly - ETA to the site could be up to 90 minutes.
  - Even with buses running normally now I would want 45 minutes for a rush to campus.
    - Luckily this hasn’t been tested yet.

The Big Problem.

Was loss of access to our machine rooms and delivery spaces.

- Spare Parts, Replacements, New Kit - how to get them?
  - Luckily the lockdown hit for us just after we got our 2020 kit delivery.
    - But we still needed to install it ourselves.
  - We've not been able to send back replaced bits either.
    - Vendors have been understanding, but may start running out of patience.
- For a while machine room access was for emergencies only whilst policy was implemented.
  - Then only a select few got their "papers" to allow access to the University.
    - A side pain for Lancaster was this period coincided with a huge reduction in public transport - any trip onto campus required a lot of planning.

In essence for the first month of lockdown we didn't really try to go onto site.

The Other Problem - Dropped Projects

The sudden shift in conditions and restrictions everyone had to work under caused some projects (especially if they were at the “conversation” stage) to be stalled.

- An early example for us we lost alignment between the move to a new CE and the move of our cluster’s shared filesystem - a move made necessary by OS incompatibilities.
  - There was a few weeks with some pretty dodgy “mount then export as nfs” shims in place.
- We’re embarrassingly only just getting our new Perfsonar Box online.
- Conversations about upgrading the site network link just fell away, and still need to be restarted.

Access Protocols (1)

- Machine room access at Lancaster requires several (reasonable) hoops to be jumped through:
  1. That you have a buddy on site.
  2. That security is notified of the access need (they also need to unlock the doors).
  3. Social Distancing measures are followed.
  4. That the room is left “fallow” if possible between visits (ideally at least 72 hours).

Access Protocols (2)

This leads to a few working practices we've adopted:

- Work is "batched" - we wait until a critical mass of odd jobs accrue.
- Visits arranged to coincide with someone from the IT Services' Technical Infrastructure Group having a similar set of tasks elsewhere in the building (the buddy).
  - Prevents unnecessary travel.
- As much as possible is done in each visit.
  - A (justified) fear of wasted trips.

It's important to note that this procedure fundamentally hasn't changed for us since leaving lockdown - there's just more buses and it's slightly easier to arrange a buddy.

Deploy the Daniela - I.C.’s lockdown.

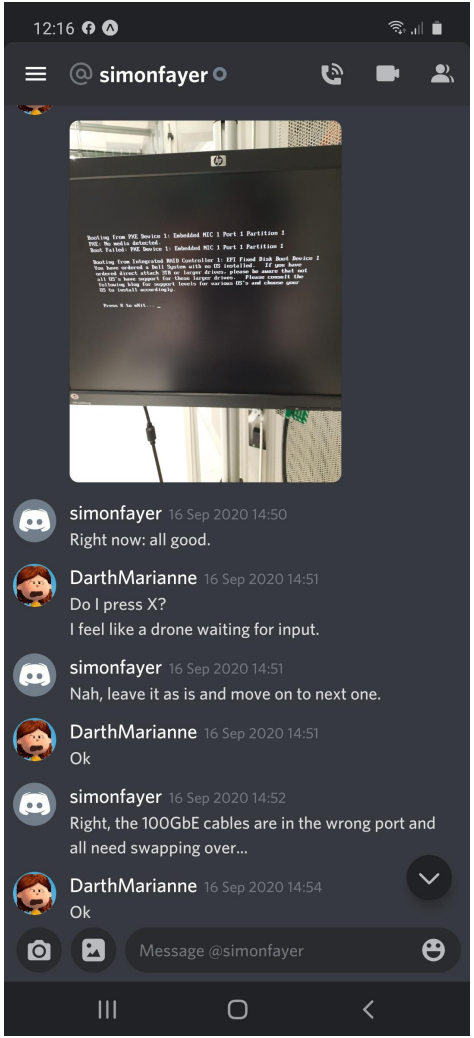
Daniela and Simon had the foresight to dash down to Slough before the lockdown hit, taking the pressure off.

Had some good communication to make sure that they made the most out of their few visits.

It being a bit awkward to get to their machine room is nothing new to them!

To paraphrase “most of our kit can trundle on for months without any maintenance”, their lockdown unexciting and devoid of good warstories.

Will need a multi-person visit though to rack kit up.



How Lockdown helped kill a disk server (1).

The problem with the lockdown environment is that it compounds delays - which for us led to a series of events leading to the loss of 120TB of data. The timeline of this disaster went something like:

- The upgrade to CentOS7 meant we dropped the zfs health monitoring for a generation of disk servers.
- This wasn’t caught until we had racked up a good number of faulty disks across all 8 servers (which was a weirdly high number).
- Working from home meant the documents containing the machine’s serial numbers and our spare part supply were not to hand.
- In the time it took to arrange a machine room visit more disks died. Hot spares were deployed.
- Many disk servers were set to RDONLY in DPM, to protect them and mitigate any problems - but this lead to a drop in Capacity and general storage service quality (the amount of juggling and hassle this caused was immense).

How the Lockdown helped kill a disk server (2)

- By the time a machine room visit was arranged we had 14 dead 6TB disks in one generation, and 2 on site spares. These are deployed to the most at risk (IMO). No opportunities for cannibalising extra spares.
- The vendor was slightly slow to respond (due to their own issues), and did not have anywhere near enough disks to cover us. Put in an order for replacements from their suppliers.
  - I foolishly did not insist on them sending what they have.
- Replacements took a month to arrive - then another week to arrange a machine room visit.
  - Another three disks fail in this time - two in the same server…
- Replacement disks are deployed - but then a fourth disk fails in the stricken server due to the stress of resilvering.

Lessons learnt.

Need more onsite spares. We had two for a generation of 8x36-bay servers. I'd suggest we want at least double. Oddly our out-of-warranty kit did better as I had a decommissioned server to raid for spare disks.

Keep better (and accessible records) - you shouldn't be looking up serial numbers from the back of the actual machines in 2020.

Monitor the health of the monitoring. I needed to update my "morning coffee routine" when I spotted a hole in our visibility, but no replacement for alerts.

This kind of issue is not unique to us - Gareth and Sam report similar fears and trials at Glasgow with their DPM - they note that CEPH is much easier to manage/recover from disk losses..

Wrap up.

Compared to many us Tier 2 admins could do our jobs pretty well during the lockdown (and what came after).

- But then I suppose we're technically working remotely even when in the office.

It's not been without its challenges though, mainly revolving around access to our machines and facilities to receive deliveries.

- And for many these pressures haven't actually eased much recently - receiving and deploying new kit will still represent a challenge for at least a few more months to come.

Any Questions/War Stories of your own?