Lancaster Lookahead

GridPP46, Ambleside
3rd September 2021

Matt Doidge, Gerard Hand, Steven Simpson, Peter Love
Lancaster University

Overview of the only Gridsite in a Duchy (that I know of).

- Shared Cluster (SGE), root access, root responsibilities
  - Have the benefits of a veteran sysadmin running the cluster.
    - That’s not me, that’s Mike Pacey.
- Dedicated Machine Room with good room to expand.
  - Water cooled Rittal racks
- Access to an "Enterprise" VMWare instance
  - Well connected, backed up, performant VM provisioning
- ARC CE
  - Has been a pleasant experience.
- DPM SE
- New Faces: Gerard and Steven join the team

Niggles

- 10Gb link to JANET, upgrade is slow to happen
  - Move is gravitating to multiple 10Gb links, which has its own issues
- Shared cluster leads to a lack of agility, need for compromise.
  - Large scale changes (OS/Batch system) need to be carefully planned and coordinated with more then just grid user groups.
- DPM SE
  - For reasons explained later.
- Old monitoring, most bit-decayed beyond usefulness

(Grid)Engine Troubles

- S(on of)GE was showing its age a while ago
- scheduling on vmem gets us in trouble with atlas jobs
  - But that’s the only tool we have.
- No features like cgroups.
- Not likely to move batch systems for a while though (likely in the next OS shift)
  - most likely candidate to move to is SLURM
- creative ARC RTE usage may be needed in the short term to cope.

Restoring the Panopticon

- Thanks to the SL6 induced retirement of our ganglia and icinga box our monitoring had decayed down to a few monit instances, a statsd script and a central syslog server.
  - But sometimes it's better to start from scratch...?
- Steven working on a new system, built on Grafana, Prometheus and Loki (prometheus-node-exporter and promtail pushing out metrics).
- Aim to leverage the “native” CEPH metric exporters.

Any suggestions appreciated. Perhaps we should restart the conversation on this?

DPM Woes (1)

- Move to "dome" relatively painless at the time
- But the loss of pool node weightings made things even more rigid
- "Hotspotting" got even worse, and no good tools to deal with it.
  - DPM favours writing to the disk with the most available space, and new disk servers tend to be much larger.
- Most frequent job failures are due to overloaded disk not accepting output.

DPM Woes (2)

- We had an unreliable generation of 6TB HDD, lots of failures at the 4/5 year mark, very prone to throwing errors.
- Hit us very hard during the lockdowns, as we couldn't easily intervene.
  - lost 2 disk servers due to quadruple disk failures, had a large amount of zfs file corruption on another.
  - over a million files lost. we looked like berks.
- Over a PB of storage at Lancaster "unreliable"/on its last legs.

Just so it’s not all negative

DPM does a few things well - migrating this far ahead of the 2024 retirement wasn’t a total forgone conclusion.

- Over a decade of local experience
  - even if most of that was bit-decayed by the DOME migration.
- Still a lot of grid sites running it. Support in numbers.
- It does TPC really well.
  - With only a small amount of effort Lancaster’s DPM has been in the various TPC testing networks for ages.
- There are still some active users/supporters, and development isn’t as dead as previously reported.

CEPH to the rescue (eventually)

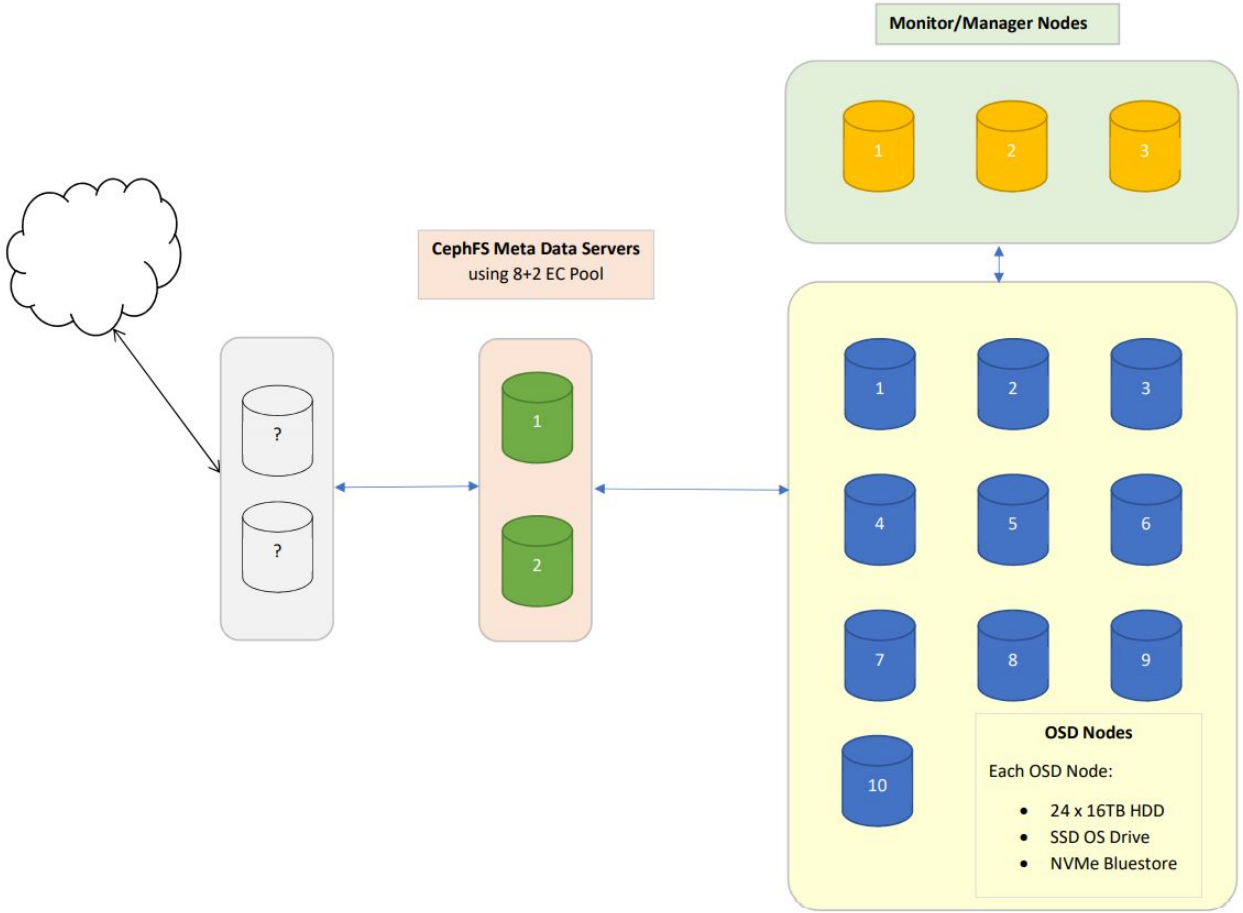
- A surprise bit of EoY funding allowed us to purchase the core of what could be a new SE.
- Drew up some with the help of Alastair.
- Gerard got the "pleasant" job of becoming our CEPH expert.
  - Although the benefit of CEPH being “non-grid” is we could just unleash Gerard at it, without first initiating him in the secret workings of dark and unholy grid technologies.
- Sadly lots of delays in getting hardware online - some covid related, some due to a problem with the NICs in the new storage when we tried to bios boot them.
  - If we had been on time this would have been a technical CEPH talk

Our CEPH Specs

(Trying to not be vendor specific)

- Typical multiple HDD bay commodity storage node (we have 24 x 16TB)
- A reasonable CPU (we have mid-range Xeon Silvers)
- A lot of RAM (256GB)
- A pair of (software mirrored) 1.6 GB NVMe drives to hold the journal.
  - Along with the large amount of RAM this makes it a lot less “commodity”.
  - The “budget” version would swap out 2 of the 16TB drives for SSDs.
  - This is as small an NVMe as we could get away with.
- Pair of 25Gb NICs.
  - Unlike our DPM storage there’s no need for the data nodes to talk to the outside world, so there’s scope to do something cunning with the spare NIC.

A cunning CEPH plan



Gerard’s CEPH Thoughts

- Ceph is a resilient, scalable storage solution running on commodity hardware.
- Open source so no licencing to worry about.
- Allows Object, Block and File Storage.
- It is designed to have no single point of failure.
- Device failures can be easily managed without downtime.
- Automated systems manage the cluster to ensure reliability of data.
- Large number of configuration options make life a bit more complicated.

“It's easy enough to get setup but I suspect it's going to be a lot of tinkering to get it tuned to how we like it.”

What goes in the front?

- A big hole in our plan is what to put in the front of all this storage?
- First thought is some simple xroot server mounting the cephfs volume?
- Load concerns, so perhaps some load balancing between a few servers? Round robin DNS? HAPROXY? Something a bit weirder?
- EOS as a frontend has been mentioned in some talks, but does that way madness lie?
- Locally we're probably just going to mount cephfs on the worker nodes and encouraging access via file://, which should do a lot to reduce load.

Why CEPH(FS)?

- "Single server level redundancy" is no longer a viable option (IMO)
- CEPH or HDFS were the options - CEPH won
  - HDFS appears to be losing support
  - CEPH has a lot of support
  - And we have a lot of UK-grown experience
- CEPHFS over a raw object store seemed a logical choice (especially as I don't "get" object stores).
- A griddy shim over a something that looks, acts and smells like a filesystem feels like the way forward for grid storage in general.
    - The “Neo-Classic” SE

The Run 3 Elephant in the Room

- If all goes swimmingly then we will have a production ready, stress tested CEPH backed SE ready at about the same time Run 3 starts
  - And of course that's if it all goes swimmingly.
- Reluctance to migrate data from atlas? Hopefully will be straight forward due to the "boring" setup.
- We're going to have to run DPM in parallel for a while (the "Glasgow model"), for other VOs. Not too worried, as DPM works well when it's not stressed.

Questions or Comments?

Hopefully this will kickstart some conversations, feel free to prod Gerard, Steven, Peter or myself about anything chatted about here.