CVMFS Stratum 1 Performance

Dave Dykstra, dwd@fnal.gov

CernVM Workshop
17 September 2024

Fermilab

Stratum 1 hardware

- In the past 2 years, FNAL & OSG Stratum 1s have seen space used go up by about 50% to the 60T range
- Existing hardware aging and slowing down, time to replace
- FNAL chose to go with all-NVMe for replacement HA pair (and RAL has now too) but it may not be the only option
  - OSG funders in particular are looking for reduced cost
- Previous FNAL hardware was from the beginning much slower than OSG (at University of Nebraska-Lincoln)
  - Hardware RAID6 at FNAL vs ZFS RAID10 with NVMe metadata cache at OSG
  - However, lately OSG performance greatly slowing down, perhaps because of 85% space of 16 drives used, but recently 8 more drives added on both
    - Data not automatically redistributed, so in process of taking fresh snapshots

NVMe vs hard disk

- Initial benchmarks in April showed a ZFS RAID6 all-NVMe 4 to 12 times faster than one of the ZFS RAID10 systems
- Benchmarks repeated last week more rigorously, and after the additional drives added at UNL
  - Used 6 parallel untar of cernvm-prod.cern.ch (in April used 1 untar of sw-nightlies.hsf.org) to avoid cpu as bottleneck on all-NVMe
    - 63G each after untar, 1 million files each, average 63k bytes/file
  - Used fresh snapshot (so on low-populated drives) of same repo and made sure system was otherwise idle
  - Used slightly newer UNL system (5 years old instead of 6)
    - Current total usage is 69T, 404 million files, average 170k bytes/file, 121 repositories
  - Also did gc and check –ic
  - Intentionally not measured: download speed

NVMe vs hard disk

	6 untars	gc	check -ic
RAID6 NVMe	266s	103s	28s
RAID10 drives	552s (x2.1)	124s (x1.2)	603s (x21.5)

- When resilvering one of the older backup drives, untar test was dramatically slowed, more than triple the time at 1869s, and during a pool scrub it was 888s (1.6 times quiet system)

Network architectures

- Using networked filesystem is not a good choice because there is such a huge number of small files, resulting in extreme metadata usage
  - RAL tried CephFS, that was disastrously slow
  - BNL is using a NAS, but it is dedicated to only the Stratum 1
  - Ceph S3? Not measured, but also file-based so still lots of metadata
- Ceph RBD in principle could be OK because metadata still processed on the client, not file server
  - Advantage of delaying purchase of storage until needed, although that can also somewhat be done by leaving empty local slots

Local NVMe vs Ceph RBD NVMe

- Benchmarked 4 variations
  1. OSG Tiger Kubernetes cluster at U W Madison, production Ceph RBD 3x NVMe, ext4 filesystem, 10gbit client
  2. 2-node test Ceph cluster at Fermilab, also Ceph RBD 3x NVMe, xfs filesystem, 100gbit client (same machine used for local NVMe)
     - No time was spent on optimizing the server performance
  3. Same cluster and client at Fermilab, with 4+2 erasure coding, like RAID6 (half the raw storage as 3x, but twice as much write i/o)
  4. Same cluster and client, with 2x, like RAID10
     1. 2x is discouraged by Ceph documentation because of data loss risk

Local NVMe vs Ceph RBD NVMe

	6 untars	gc	check -ic
Local NVMe	266s	103s	28s
Tiger 3x RBD	273s-339s (x1.2)	191s (x1.9)	32s (x1.1)
FNAL 3x RBD	385s (x1.4)	92s (x0.9)	49s (x1.8)
FNAL 4+2 RBD	727s (x2.7)	94s (x0.9)	44s (x1.6)
FNAL 2x RBD	368s (x1.4)	94s (x0.9)	48s (x1.7)

- Repeated Ceph measurements varied quite widely on Tiger and somewhat on FNAL 4+2

How much speed needed?

- The worst performance problem most Stratum 1s experience is update delays caused by garbage collection
  - GC blocks updates, sometimes for many hours
  - It doesn't really matter if checks take a long time, even months
  - Here we have an opportunity with a software change to drastically reduce requirements on hardware speed: a GC that is either non-blocking or blocking for only short periods of time
  - Meanwhile there are probably repos that don't need GC that could be turned off

What's the best architecture choice?

- 3x Ceph RBD on production cluster is impressively close to local RAID6 NVMe, but storage cost is double so not justified
- 4+2 erasure coding (RAID6) Ceph RBD slower, double server i/o
- 2x Ceph RBD (RAID10) lower cost, faster, but not recommended
- Local RAID10 hard drives significantly cheaper than any NVMe and probably fast enough (especially after gc is changed to not block), and more consistent than Ceph RBD
  - Slightly higher risk of data loss than RAID6, but if there's a catastrophe data can be recovered from another machine
  - Avoid going over 70% capacity because of performance degradation
- For ultimate speed, reliability: local RAID6 NVMe although expensive