

OU XROOTD SITE REPORT

HORST SEVERINI
XROOTD WORKSHOP 2023
MARCH 2023

Outline

- Introduction
- Computing and Storage Hardware
- Network
- XRootD Configuration
- CephFS Testing

Introduction

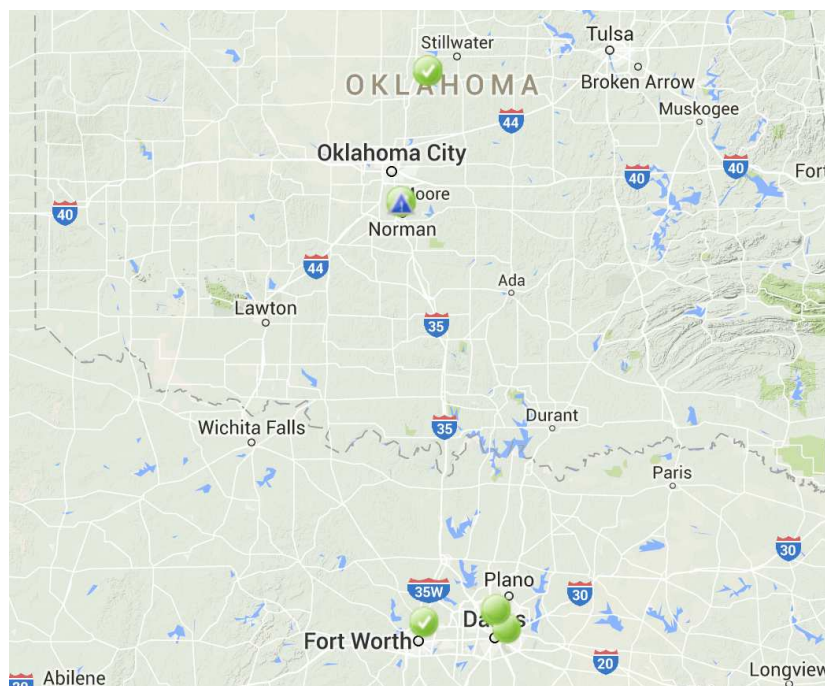
- Planning to migrate from xrootd to ceph storage in the next year
- Unfortunately not as much to present on these tests as intended
- Had to spend much time on upgrading HTCondor-CE GK
- EL9 osg-36 brand new, installed from osg-upcoming-testing
- Took much longer to iron out teething bugs
- Most likely first ATLAS EL9 GK world wide
- Therefore this talk covers only first initial tests, done yesterday!



US ATLAS SWT2 Center

- University of Oklahoma
 - Oklahoma Center for High Energy Physics (OCHEP)
 - OU Supercomputing Center for Education and Research (OSCER)

- University of Texas Arlington
 - Chemistry and Physics Building (CPB)
 - Arlington Regional Data Center (ARDC) in Fort Worth



OU_OSCER_ATLAS Tier 2 Hardware

- 83 Nodes (5600 Slots) – 2 GB RAM per Slot
- 10 Support Nodes (1 GK, 1 DTN/Proxy, 1 redirector, 7 storage)
- 700 TB of usable xrootd storage (7 T630s with 16 8 TB drives, RAID6, xfs)
- SALT (CentOS 7.8), SLURM 20.11, OSG 3.6, XRootD 5.4.2
 - Upgrading to EL9 and SLURM 22 this summer
- Part of generic OSCER HPC cluster
- Rest of OSCER Schooner Hardware
 - 850 Nodes (about 25k Slots) – 2-4 GB RAM per Slot
 - Opportunistically available for ATLAS production



Network

- OU connected at 100 Gbps to I2 and ESnet via OneNet
- OSCER connected at 100 Gbps to OneNet
- Everything ipv6 ready except for HTCondor-CE GK
- SE on 50 Gbps DMZ – OFFN (Oklahoma Friction Free Network)
 - Dual 25-gig Bonded
 - iperf3 test over 40 Gbps
 - WAN xrootd transfers to storage nodes over 25 Gbps

XRootD Configuration

- Currently, se1.oscer.ou.edu proxy server for 700 TB xrootd cluster
 - Pretty stable and performant
 - Occasionally, one or two of the 7 servers gets overloaded with open connections, causing transfer timeouts
 - Restart of xrootd (not cmsd) on these nodes eventually fixes this
 - Not fully understood
- Plan to migrate to cephfs file system after warranty expires
 - Part of 9.5 PB (and growing) OSCER ceph file system
 - 26 R740xd2 OSD nodes, 8+3 erasure coding
 - 14 GB/s total throughput
 - Very performant, reasonably priced
 - \$90 per usable TB, good for 7 years



Current XRootD Ceph Testing Status

- Created 10 TB cephfs partition, /xrd_test
- Mounted on grid0 (retired old EL7 GK)
- xrootd 5.5.2 installed on grid0
- Unfortunately, grid0 went down during maintenance, hasn't come back up yet
- Plan B: mount /xrd_test on se1 (current SE proxy)
- Then, attempt to either incorporate this cephfs partition into current xrootd proxy server, or stand up separate test xrootd server
- At first, permissions problems mounting on se1, but now solved!
- Brought up separate xrootd server on port 64000

XRootD Ceph Testing

- Alternate XRootD server config very similar to proxy config:

```
[hs@ouhep1 se]$ diff xrootd-cephfs.cfg xrootd-se.cfg
2d1
< xrd.port 64000
7,8c6,7
< all.export /xrd_test
< #pss.origin dms.oscer.ou.edu:1094
---
> all.export /xrd
> pss.origin dms.oscer.ou.edu:1094
11c10
< #ofs.osslib libXrdPss.so
---
> ofs.osslib libXrdPss.so
```


XRootD Ceph Testing continued

- Simple gfal testing:

```
gfal-copy --copy-mode pull https://se1.oscer.ou.edu:1094/xrd/srm-test/  
twentyfivegiga https://se1.oscer.ou.edu:64000/xrd_test/atlasdatadisk/  
twentyfivegiga1
```

- Can get up to 400 MB/s transfer speeds

- Hiro's FTS transfer tests:

```
Parallel transfer tests against  
https://se1.oscer.ou.edu:64000/xrd_test/atlasdatadisk/
```

- Get up to 3.5 GB/s transfer speeds, but average closer to 1.5 GB/s

- Done without any tuning so far

- Also, ATLAS Rucio production transfers going on at the same time

Summary and Conclusions

- OU XRootD storage quite stable and performant
- Able to transfer 3+ GB/s, which is probably close to current available hardware/network limit
- Initial xrootd-ceph setup successful
- Looking forward to performance improvement with this new setup
- Hopefully get closer to 50 Gbps current network limit
- Possibly also test s3-connector, although current OSCER ceph configuration doesn't support any block operations
- Also still to do: Token Auth for XRootD (Working fine on new EL9 UT Condor-CE)

