Experimental Computing Laboratory Meeting June 27, 2024, 12:30 PM

Advanced Computing Systems

Research Section

Jeffrey Vetter, Section Head

Steve Moulton, Systems Engineer

Aaron Young, Software Engineer

EPO Test on Monday 6/24 was successful

- Planned power outage in 5100, 116.
- Affected Systems were:
  - Oswalds
  - Pcie
  - Equinox
  - Milans

  And VMs:
  - Dragon
  - Intrepid
  - Spike
  - Aries
- The test was successful, and all systems are back online.

New ZFS Snapshot and Replication using zrepl

- See Backup & Storage | ExCL User Docs (ornl.gov).

A note about file system snapshot changes in ExCL.

The homegrown scripts to manage snapshots and replication have been replaced by zrepl (https://zrepl.github.io/). Zrepl handles both automated snapshot generation and file system replication. Snapshots are taken hourly, and ExCL file systems are replicated to the back up (old FS00) fileserver.

In the past, snapshots have been available as ~/.zfs/snapshot/(hourly,daily,weekly)*. These will continue to be available until they age out. The new snapshot format is

~/.zfs/snapshots/zrepl_yyyymmdd_hhmmss_000

where the hour is in UCT, not Eastern Daylight/Standard Time. This is a zrepl property to enable global replication consistency, and is not modifiable. If you deleted or made a destructive modification to, say, ~/.bashrc on June 11, 2024 at 3 PM, it should be available in ~/.zfs/snapshots/zrepl_20240611_185313_000/.bashrc, and in earlier snapshots. Note that ~/.zfs is not mounted (i.e., not visible) until you access it.

Snapshots do take space, so they are automatically deleted as they age, so that all hourlies are kepts for two days, one hourly from each day is kept for 30 days, and one hourly for each 30 day period is kept for 180 days. Snapshots are read only; you can copy files from them back into your home directory tree to restore them.

Given the current and historical instability of ORNL tape backup services, snapshots and a replicated file system server are the only forms of backup in ExCL. As stated elsewhere, you (as a user) are responsible for your own software and data storage management. While we make best effort to maintain file system integrity, We recommend that all software be maintained in git, and frequently pushed to gitlab (or your favorite git repository). Critical data should be replicated elsewhere.

For questions or general discussion, please contact excl-help@ornl.gov or the CCSD #excl slack.

ZFS/NFS Storage Quotas. See Backup & Storage.

- We have added quotas to the filesystems in ZFS to avoid runaway storage usage.
- Quotas in ZFS are easy to view and set at any level in the filesystem.
- The quota applies to all storage used by the filesystem, including snapshots of the data. Because of this, deleting files will not reduce disk usage until the snapshots age out.
- ZFS is smart and only stores changes in data. It also stores data in a compressed format, so your disk usage is less than your apparent amount.
- Home and project directories start with 512G per directory, and higher quotas can be requested via excl-help@ornl.gov. We can also help by giving a breakdown of file usage and helping clean up large usages.
- Use of /scratch/$USER for large build artifacts that are local to a node and do not require snapshotting can reduce ZFS storage usage.

ZFS Storage Deep Dive

```
[root@fs01 ~]# zfs list -r -t all -o space,refer,written pool/home/jum
NAME                                                  AVAIL   USED  USEDSNAP  USEDDS  USEDREFRESERV  USEDCHILD     REFER  WRITTEN
pool/home/jum                                         4.48T   527G      416G    111G             0B         0B      111G    5.45G
pool/home/jum@weekly-2024-05-05_00.00.02--8w               -  5.82G         -       -              -          -     65.3G    65.3G
pool/home/jum@weekly-2024-05-12_00.00.01--8w               -  1.63G         -       -              -          -     84.3G    24.8G
pool/home/jum@weekly-2024-05-19_00.00.02--8w               -  1.63G         -       -              -          -     84.3G    1.63G
pool/home/jum@weekly-2024-05-26_00.00.01--8w               -   434M         -       -              -          -     84.5G    9.17G
pool/home/jum@weekly-2024-06-02_00.00.02--8w               -   434M         -       -              -          -     84.5G     434M
pool/home/jum@weekly-2024-06-09_00.00.01--8w               -  1.22G         -       -              -          -      113G    30.5G
pool/home/jum@zrepl_20240611_035304_000                    -  1.25M         -       -              -          -      113G    1.47G
pool/home/jum@zrepl_20240612_035312_000                    -     0B         -       -              -          -      113G    1.26M
pool/home/jum@daily-2024-06-12_00.00.02--14d               -     0B         -       -              -          -      113G        0
pool/home/jum@zrepl_20240613_055311_000                    -  1.30M         -       -              -          -      113G    1.92M
pool/home/jum@zrepl_20240614_065310_000                    -  9.04G         -       -              -          -      114G    10.3G
pool/home/jum@zrepl_20240615_075311_000                    -  5.09G         -       -              -          -      129G    23.7G
pool/home/jum@zrepl_20240617_075309_000                    -  2.55M         -       -              -          -      132G    7.65G
pool/home/jum@zrepl_20240618_075311_000                    -  4.60M         -       -              -          -      132G    4.60M
pool/home/jum@zrepl_20240619_075312_000                    -  10.7M         -       -              -          -      132G    11.6G
pool/home/jum@zrepl_20240620_085311_000                    -  16.6M         -       -              -          -      132G    30.5M
pool/home/jum@zrepl_20240621_085318_000                    -  2.62G         -       -              -          -      332G     217G
pool/home/jum@zrepl_20240622_085852_000                    -  7.81G         -       -              -          -      341G    12.2G
pool/home/jum@zrepl_20240623_095307_000                    -     0B         -       -              -          -      341G    11.6G
pool/home/jum@zrepl_20240623_215317_000                    -     0B         -       -              -          -      341G        0
pool/home/jum@zrepl_20240623_235312_000                    -     0B         -       -              -          -      341G        0
```

```
[root@fs01 ~]# zfs get compressratio pool/home/jum
NAME           PROPERTY       VALUE  SOURCE
pool/home/jum  compressratio  2.12x  -
[root@fs01 ~]#
```

New Systems

- Zenith2, a second custom desktop similar to Zenith (aka Zenith1) in hardware.
  - Built and set up.
  - Uses a TinyPilot for remote management.
- Hudson, the H100 server has been racked and is ready for configuration.
- The MI300 system should ship June 26th.

Zenth2

Centos 7 Deprecation and Migration to Rocky 9

- We are continuing the migration of all infrastructure systems from Centos 7 to Rocky 9.
  - Support ends June 30, 2024.
- login-new.ornl.gov in the testing phase to replace login.ornl.gov after testing is completed.

Primary Usage Notes to be added to docs.excl.gov

- Working on creating a list of available nodes and how they are used and can be used.
- This would be a good landing page to see which nodes might want to run on.
- Good piece of missing onboarding documentation for new users to learn which worker nodes to leverage.
- If you are interested in contributing to this list, let me know.

Reminders and Notes

- Migrate data out of /noback directories.
- ExCL is not ready for Ubuntu 24.04 until the ansible VM is migrated off of centos7. The ansible version is too old to push to a Python3 system.

Questions/Discussion?