Table of Contents

Determining if a Repository is old world or New World.

A rough tally is maintained on progress on NetappToCephDiskServer but definitively. Check where e.g cvmfs-atlas resolves to. If lxcvmfsNN < 50 it is old world. If new world then it is > 50.

Old World Stratum 0s.

- Contained in the hostgroup **cvmfs/lxcvmfs**.
- One node per respoitory e.g cvmfs-atlas.cern.ch -> lxcvmfs37.cern.ch.
- In puppet cvmfs is configured by the now deprecated class cvmfs::server which is included from the hostgroup.
- For each repo there is netapp volume e.g CVMFS-nfs01.cern.ch:/vol/CVMFS2/atlas . This is mounted
    - rw on the stratum 0 node. (lxcvmfs37)
    - ro on the stratum 0 webserver node. i.e **cvmfs/zero** hostgroup. It is from here the content is served under cvmfs-stratum-zero.cern.ch.

New world Stratum 0s.

- Contained in the hostgroup **cvmfs/lx**
- Can be many cvmfs repos per node, e.g cvmfs-test and cvmfs-opal.cern.ch both resolve to lxcvmfs53.cern.ch
- In puppet there are two defined types per stratum 0. * cvmfs::zero - this contains generic stratum 0 configuration that could be used on all sites everywhere. * hg_cvmfs::private::localzero - contains CERNisms e.g mounting up the file systems and such like.
- Each lxcvmfs node contains a configuration in e.g **fqdn/lxcvmfs53.cern.ch.yaml** a create_resources call is made to load of a list of **cvmfs::zero** and **hg_cvmfs::private::localzero** instances. One per repo on each node.
- Each new world stratum 0 node runs apache as well. The hosts zero05 and zero06 behind cvmfs-stratum-zero.cern.ch both reverse proxy requests back to the lxcvmfs node.

New world Stratum 0s Filesystem

- Each repository trypically has two CEPH filesystems named
    - 20150702-stratum0-ams - This is the live stratum 0 ceph file system mounted on the cvmfs-ams node.
    - 20150702-backup-stratum0-ams - THis the backup ceph volume of the first.
- Both ceph volumes are a zpool each called **ams.cern.ch** containing one filesystem **ams.cern.ch/data** which is the actual data.
- Snapshots:
    - The zfs filesystem cvmfs.cern.ch/ams on cvmfs-ams.cern.ch is snapshoted via **/etc/cron.*/zfs-snap-shot** to produce hourly, daily, weekly snapshots. These are purged by the same scripts.
    - The zfs filesystem cvmfs.cern.ch/ams on cvmfs-ams.cern.ch is pushed incementally using zrep to one of the backup machines. e.g backup-cvmfs01.cern.ch in the hostgroup **cvmfs/backup**. The push is done via the script **/etc/cron.hourly/zrep_cron.sh**. WHen the push is done it is incremental. It also pushes all the snapshots that have been made via the above scripts.
    - On the backup server the snapshots from the zfs-snap-shot script are not purged so another cronjob **/usr/local/bin/purge-snapshots.sh** runs on the backup server to purge these.
- More detail on stratum 0s setup this way can be learnt from looking at the setup in NewRepo.

Stratum 0 file system upsizing.

- I tried detaching a ceph volume and resizing it. ZFS was afterwards not happy so until tested don't do this.
- Instead read the man pages and google, very standard zfs operation.
    - ♦ Create a new bigger ceph volume and attach it to node.
    - ♦ Attach as mirror to existing ceph volume **zpool attach ams.cern.ch virtio-SMALL virtio-BIG**
    - ♦ Remember to set the myid property on the volume as always.
    - ♦ Monitor **zpool status ams.cern.ch** to check how mirroring is doing.
    - ♦ Once mirror complete drop the old ceph volume. **zpool detach ams.cern.ch virtio-BIG**
    - ♦ Expand the volume. See **zpool status** to understand and the **zpool online -e ams.cern.ch virtio-BIG**
    - ♦ Increase the quota on the zfs filesystem **ams.cern.ch/data**

Stratum 0 Failure Senarios

Someone just remove the contents of Stratum 0

Writers of the stratum 0 have permissions to e.g *rm -rf /var/spool/cvmfs/ams.cern.ch/data. In this case zfs snapshots are your friend basically.

- Find a suitable snapshot and rewind to it. See google.
- To get backups to work again you probably need to delete snapshots from the destination also.

Loosing a Stratum 0 Node.

This assumes you still have a CEPH volume with all the data.

- Install a new node via NewRepo.
- Attach the existing ceph volume.
- Run zfs import , it will complain about a bad shut down and then you force it. I have never tried it.

zrep replication fails.

Particularly after a crash of backup server or sending node zrep may stop doing replications and report

```
sending ams.cern.ch/data@zrep_000135 to backup-cvmfs01.cern.ch:ams.cern.ch/data
cannot receive incremental stream: destination ams.cern.ch/data has been modified
since most recent snapshot
```

The fix here is to roll the destination system back to the last good snapshot that was made so ignoring the partial changes.

```
# zfs list  -t snap | grep ams
ams.cern.ch/data@zrep_000097                                          136K      -  2.24T  -
ams.cern.ch/data@zrep_000098                                          136K      -  2.24T  -
ams.cern.ch/data@zrep_000099                                          136K      -  2.24T  -
ams.cern.ch/data@zrep_00009a                                          136K      -  2.24T  -
ams.cern.ch/data@zrep_00009b                                          136K      -  2.24T  -

# zfs rollback  ams.cern.ch/data@zrep_00009b
```

Hopefully after that a normal zrep should work again.

Stratum One

The stratum one is stand alone diskserver at the end of **cvmfs-backend.cern.ch** and currently 3 reverse squid proxies behind **cvmfs-stratum-one.cern.ch**. They are deployed in the hostgroups.

- cvmfs/one/backend
- cvmfs/one/frontend/live

A script **it-cvmfs/stratum1/create-frontend-machine.sh** exists for deploying frontend squid servers into **cvmfs/one/frontend/spare**.

Once deployed there are a couple of manual steps which I must puppetize one day.

- Stop squid **systemctl stop squid.service**
- Create cache directories **squid -z**
- Reboot and run puppet. If good then stick the node in **cvmfs/one/frontend/live** to have to join the alias cvmfs-stratum-one.cern.ch.

If you are mega unlucky and the cvmfs-backend fails then panic quite frankly. CvmFS should work without it since the other stratum ones should cope with the load. Find a big disk server and install into **cvmfs/backend** hostgroup. It will take days to be ready via cron jobs even once the host is installed. It will only be noticed by CvmFS clients once the cron jobs have finished.

This topic: CvmFS > MachineResources
Topic revision: r11 - 2015-07-27 - unknown