bd NOG
Bangladesh Network Operators Group

OpenZFS

ZFS file system

Insert the title of your subtitle Here

Rezaul Karim
Redsky Solutions Pvt. Ltd.

Disclaimer:

The scope of this topic here is not to discuss about the architecture of the ZFS rather Features, Use Cases and Operational Method.

What is a File System?

- File systems are an integral part of any operating systems with the capacity for long term storage.
  - present logical (abstract) view of files and directories.
  - facilitate efficient use of storage devices.
  - Example-NTFS,XFS,ext2/3…etc

What is a File System?

ZFS is a combined file system and logical volume manager designed by Sun Microsystems and now owned by Oracle Corporation. It was designed & implemented by a team at Sun Microsystems led by Jeff Bonwick and Matthew Ahrens.

Why ZFS?

ZFS is scalable, and includes extensive protection against data corruption, support for high storage capacities, efficient data compression.

Available OS Platform for ZFS

bd NOG
Bangladesh Network Operators Group

ZFS is available for Solaris (and its variants), BSD (and its variants) & Linux (Canonical’s Ubuntu Integrated as native kernel module from version 16.04x)

Features:

- It’s a 128-bit file system.
- Auto Healing.
- Dynamically expandable.
- East to extend capacity.
- different RAID levels supported.
- It supports Multi Size Disks.
- It is Copy On Write (COW) File System and support very less costly snapshots & clones.

Features:

- File System Replication, Export & Import.
- System can be easily re-import in a new system/data volume.
- Supports Quota and it is modifiable on the fly.
- It supports file system level compression.
- No FSCK is required to check disk error.
- Much more less time need to rebuild (re-silver) failed disks.
- Support read*/write* cache.
- Alternative to and much more better than LVM on Linux.

Limitation/Controversy

- ZFS is limited to running on a single server in contrast to distributed or parallel file systems.
- ZFS need more RAM and may need more CPU while using de-duplication feature.
- In the Linux community, there are various opinions on licensing with respect to the redistribution of the ZFS code and binary kernel modules under a general public license (GPL).

Commercial Product Using ZFS

bd NOG
Bangladesh Network Operators Group

- DELPHIX
- NEXENTA
- IXSYSTEMS TRUENAS/FREENAS
- OSNEXUS
- SYNETO
- JOYENT MANTRA

Some Examples Commands

Installation:

- apt install zfsutils-linux

Pool Creation:

- zpool create -f vol1 /dev/sda3 ; creating stripe/raid0 pool-volume
- zpool create -O ashift=12 -f vol1 mirror /dev/sdc /dev/sdd ; raid0
- zpool create -O ashift=12 -f vol1 raidz /dev/sdc /dev/sdd /dev/sde ; raid5
- zpool create -O ashift=12 -f vol1 raidz2 /dev/sdc /dev/sdd /dev/sde /dev/sdf; raid6
- zpool create -O ashift=12 -f vol1 raidz2 /dev/sdc /dev/sdd /dev/sde /dev/sdf /dev/sdg
- ; triple parity raid can survie 3-disk failure at a time

Pool Manipulation

- zpool status ; see pool status
- zpool list ; see pool list, multiple pool can be created in a single box

- zpool add vol1 raidz /dev/sdf /dev/sdg /dev/sdh ; expand a raidz/raid5 volume

- zpool add -f vol1 cache /dev/sdj ; add SSD read cache L2ARC
- zpool add -f vol1 cache mirror /dev/sdj /dev/sdk
-
- zpool add -f vol1 log /dev/sdl ; add SSD write cache ZIL
- zpool add -f vol1 log mirror /dev/sdl /dev/sdm

File System/Dataset Manipulation

- zfs list
- zfs create vol1/iso
- zfs create vol1/kvm
- zfs careate vol1/lxc
-
- zfs set compression=on vol1
- zfs set atime=off vol1
- zfs set autoexpand=on vol1
-
- zfs set quota=10G vol1/lxc
- zfs set reservation=10G vol1/lxc
- zfs set mountpoint=/mnt/data vol1/iso

Snapshot & Snapshot manipulation

- zfs create vol1/lxc/container1
- zfs snapshot vol1/lxc/container1@version1
- zfs snapshot vol1/lxc/container1@`date +%Y%m%d-%H%M%S`
-
- zfs list -t snapshot |grep container1
-
- date_time=`date +%Y%m%d-%H%M%S`
- zfs snapshot -r vol1/lxc/container1@$date_time
-
- zfs rollback -r vol1/lxc/container1@version1
- zfs destroy -r vol1/lxc/container1@version1

Handling Disk Failure

- zpool replace [poolname] [old drive] [new drive] ; replace old/damaged disks
- zpool replace vol1 /dev/sdh /dev/sdx

- watch zpool status
- zpool clear vol1
- zpool status

Any Questions ?

Thank you

bdNOG
Bangladesh Network Operators Group