

# ZFS User Conference 2017

Large Scale Homelab Backups

Mike Trogni

[miketrogni@gmail.com](mailto:miketrogni@gmail.com)

# Large Scale “Homelab” backups

- This is not meant to be a business/enterprise overview of backup, but rather a User-point of view presentation of options for backup for the homelab admin with a limited budget
- I will go over the various techniques available for backup that are available and affordable for a typical homelab datahoarder, and discuss pros and cons, leave a little time for Q&A

# Large Scale “Homelab” backups

- First, let’s define Large Scale.
  - My rule of thumb for Large scale backup requirements really depends on the state of the art of affordable SATA storage, if your desired amount of backup data exceeds the size of the largest affordable SATA drive (as of 3/2/17, Seagate Archive 8Tb is \$224), then you accumulated enough to make it large scale, congratulations!

# Large Scale “Homelab” backups

- Like many of you, my data storage requirements started off small, perhaps one conventional drive or workstation SSD, and then grew and grew.
- I was drawn to ZFS early on when it was ported to Mac OS X by Apple, then by Don Brady @ Zevo- the scalability and the bitrot protection, plus the scrubbing and raidz1/raidz2 support appealed.
- I quickly outgrew a ReadyNas appliance, and had a raidz1 failure by pulling out the wrong disk after a disk failed. Luckily, I had a backup for most of the data!

# Large Scale storage, peace of mind

- I am a hobbyist that includes having thousands of ripped CDs and nearly 700 Blu-Ray/DVDs, I felt naturally that I did not want to repeat the painful and lengthy process of re-ripping due to disk or volume failure.
- The Blu-Ray backup requirements imply large amounts of storage. My annual incremental storage consumption of ripped video ranges between 1Tb and 3 ½ Tb.

# 3-2-1 backup strategy

- 3 – Three copies of your data
- 2 - Two copies “local” to your primary location, 1 copy primary, 1 copy “cold storage”
- 1 – One Copy offsite. In case your house burns down or is robbed.
- See : [Backblaze 3-2-1](#)

# Large Scale “Homelab” backups

- I replaced ReadyNAS with Mac OS X ZFS, and eventually discovered and implemented FreeNas 9 at home. FreeBSD file server appliance that will support container virtualization with the next major release (FN 10).
- Backups saved my bacon several times over the course of my life. How to safely replicate my dataset ?(>22Tb currently).
- Here is the result of my multi-year backup investigation.

# Conventional Solutions

- USB3 external drives – throughput is very good, can be configured to be standalone striped vdevs, scrub-able by zfs to check for bitrot, but lack corruption repair, due to single stripe, excellent for cold storage or manual data replication.
- Disk + caddy – similar to external drive, striped vdev, get into a monthly habit of mounting the volume and backing up the data, preferably with a script (automate).
- Drawbacks : requires manual intervention to close the air gap and plug-in disks, scalable to size of affordable USB drive or caddy-mounted SATA drive (backups need to fit on single disk, partial backups can result).



# Rclone + Amazon Cloud, \$60/year

- RCLONE is available on multiple platforms (Freebsd, Linux, desktop os's), and support multiple back ends, in my case, Amazon Drive, \$60/year, unlimited-ish backups.
- I install rclone + my bash script to a FreeBSD jail running on my FN file server. Nightly, it incrementally uploads new and changed files into the Amazon cloud.
- Drawbacks – small files work fine, large files sometimes disappear and need to be re-uploaded. No checkpoint restart functionality currently with rclone. Also, no snapshot versioning as far as I can tell.

# LTO/LTFS tape backups

- A tried & true method for backup for decades
- LTO5/6 provides large & stable platform for cold storage. 2.5Tb native for LTO6, 1.5Tb for LTO5.
- LTFS is a rather new standard for writing to LTO5 /LTO6 tape. Basically allows treating the tape as if it were a disk (drivers required).
- Expense of hardware and tape itself main drawback but considered **safest option for backup/cold storage**.
- Most homelab-ers will not go to this expense. Small and large businesses hopefully do.

# The backup server +2X the disks

- [Why build one, when you can have two at twice the price?](#)
- I have an off-site backup server, filled with 8Tb Seagate Archives in a 8Tbx8 raidz2 configuration
- Updated nightly, pulls new/updated files via rsync, launched by cron
- Pros-Cons: easy to conceptualize, expensive to implement. Built in ZFS replication failed with my large mkv datasets (FN 9.3). . . I had a hard time getting “restart where left-off” with rsync for my large Blu-ray files.
- Never can get enough upload bandwidth from my Primary site (not on Google Fiber!)

# Large Scale “Homelab” backups

- **WD model**      **cpu**      **RAM**      **throughput**
  - MyCloud Live Duo      800 Mhz      256 Mb      200 Mbps/s
  - My Cloud EX2      1.2 Ghz      512 Mb      300 Mbps/s
  - My Cloud EX2 Ultra      1.3 Ghz DC      1 GB DDR3      600 Mbps/s
- 
- The current dual bay diskless EX2 Ultra lists for \$159.
  - I purchase 2 8Tb Seagate Archive SATA disks, and configure a spanned EXT4 ~15 ½ Tb, ssh, bash, rsync solution that can be plugged in and rsync'd on a weekly or monthly basis, or co-located at a friend's house and cron'd to pull down updates nightly.
  - Datasets in zfs correspond to shares in EXT4, 1 to 1 correspondence when pulling updates via rsync/cron.

# Large Scale “Homelab” backups

- TLDR :
  - tape (king of cold storage, requires drive & tapes)
  - Zfs stripes + Caddy (cold storage, 6Tb – 8Tb)
  - My Cloud EX2 Ultra with dual 8Tb SATA (16Tb ext4, ssh,rsync and cron support built-in)
  - USB3 external drives (cold storage, suitable for 6Tb- 8Tb, fast transfer)
  - Backup server with 8Tb Seagate Archives offsite (warm storage, twice the expense, bandwidth constraints)

# Expert tips from someone who has made the mistakes before

- Don't dump your data into 1 large pool; use datasets to segregate data (audio, video, photos in my case are all in different datasets). Keep datasets a reasonable size, perhaps create new datasets annually.
- Follow the 3-2-1 rule, backup data at least on a monthly (better: semi-weekly) basis to cold storage.
- Learn how snapshots work, they can save your bacon if you implement them **before** you screw up! Consider long expiry times for your snaps.
- Run SMART short daily checks and semi-weekly run long SMART tests. zfs Scrub semi-weekly or monthly.
- Replace drives that fail SMART tests quickly.
- Configure email so you get the nightly alerts!

# Share & Enjoy

- Enjoy the peace of mind of 3-2-1 backups!
- Customize my scripts for My Cloud EXT4 backups, cron rsync & striped vdev external backup.
- <http://trogninet.dyndns.org/scripts>
- Q & A