NOVEMBER 26 2018

How to replace a failed disk in a ZFS mirror

ZFS, RAID, LINUX

I recently built a new file server for my media needs at home. Something I've been thinking about doing for *literally* years. I chose to go with ZFS as the storage technology after having used Linux software RAID for many years. I went with a mirrored setup for a lot of the reasons outlined in this article - performance, simplicity, and in particular, easy recovery from disk failures.

This is the setup I ended up with according to `zpool status`.

```
$ zpool status
  pool: storage
 state: ONLINE
  scan: none requested
config:

    NAME                                    STATE     READ WRITE CKSUM
    storage                                 ONLINE       0     0     0
      mirror-0                              ONLINE       0     0     0
        ata-WDC_WD80EFZX-68UW8N0_VJHDBDGX   ONLINE       0     0     0
        ata-WDC_WD80EFAX-68KNBN0_VAGASE7L   ONLINE       0     0     0
      mirror-1                              ONLINE       0     0     0
        ata-WDC_WD80EFZX-68UW8N0_VJHD6BAX   ONLINE       0     0     0
        ata-WDC_WD80EFAX-68KNBN0_VAGA5BPL   ONLINE       0     0     0
      mirror-2                              ONLINE       0     0     0
        ata-WDC_WD80EFZX-68UW8N0_VJHD982X   ONLINE       0     0     0
        ata-WDC_WD80EFAX-68KNBN0_VAG9X8YL   ONLINE       0     0     0

errors: No known data errors
```

Well, no sooner had I completed the ZFS setup (a very straightforward process) than one of my disks started reporting SMART errors. I don't think a disk that is weeks old

should do this, so I decided to start the RMA process.

And this is how I replaced the disk.

Replacing the disk

I started by physically removing the old disk, and replacing with a brand new one. I originally setup my pool using the disk id from `/dev/disk/by-id/`, so identifying the failed disk was very easy as the serial number is part of the device name

Once I started back up, I ran `zpool status` and saw this output.

```
$ zpools status
  pool: storage
 state: DEGRADED
status: One or more devices could not be used because the label is missing
    invalid.  Sufficient replicas exist for the pool to continue
    functioning in a degraded state.
action: Replace the device using 'zpool replace'.
   see: http://zfsonlinux.org/msg/ZFS-8000-4J
  scan: none requested
config:

    NAME                                    STATE     READ WRITE CKSUM
    storage                                 DEGRADED     0     0     0
      mirror-0                              ONLINE       0     0     0
        ata-WDC_WD80EFZX-68UW8N0_VJHDBDGX   ONLINE       0     0     0
        ata-WDC_WD80EFAX-68KNBN0_VAGASE7L   ONLINE       0     0     0
      mirror-1                              ONLINE       0     0     0
        ata-WDC_WD80EFZX-68UW8N0_VJHD6BAX   ONLINE       0     0     0
        ata-WDC_WD80EFAX-68KNBN0_VAGA5BPL   ONLINE       0     0     0
      mirror-2                              DEGRADED     0     0     0
        ata-WDC_WD80EFZX-68UW8N0_VJHD982X   ONLINE       0     0     0
        18311740819329882151                UNAVAIL      0     0     0  was

errors: No known data errors
```

ZFS noticed that it had a missing disk, and was now in a `DEGRADED` state, but crucially, everything was still working and available.

The next step was to find out what the *new* device is called. I did this by running `ls -1 /dev/disk/by-id/` and seeing which disk was new.

```
$ ls -1 /dev/disk/by-id/ | grep ata
ata-WDC_WD80EFAX-68KNBN0_VAGA5BPL
ata-WDC_WD80EFAX-68KNBN0_VAGA5BPL-part1
ata-WDC_WD80EFAX-68KNBN0_VAGA5BPL-part9
ata-WDC_WD80EFAX-68KNBN0_VAGASE7L
ata-WDC_WD80EFAX-68KNBN0_VAGASE7L-part1
ata-WDC_WD80EFAX-68KNBN0_VAGASE7L-part9
ata-WDC_WD80EFAX-68LHPN0_7HJSWL7F
ata-WDC_WD80EFZX-68UW8N0_VJHD6BAX
ata-WDC_WD80EFZX-68UW8N0_VJHD6BAX-part1
ata-WDC_WD80EFZX-68UW8N0_VJHD6BAX-part9
ata-WDC_WD80EFZX-68UW8N0_VJHD982X
ata-WDC_WD80EFZX-68UW8N0_VJHD982X-part1
ata-WDC_WD80EFZX-68UW8N0_VJHD982X-part9
ata-WDC_WD80EFZX-68UW8N0_VJHDBDGX
ata-WDC_WD80EFZX-68UW8N0_VJHDBDGX-part1
ata-WDC_WD80EFZX-68UW8N0_VJHDBDGX-part9
```

The new disk is the one on line 8 - `ata-WDC_WD80EFAX-68LHPN0_7HJSWL7F`. It stands out in this example as all the other disk serial numbers start with "V".

I now needed to tell ZFS to replace the missing disk with this one.

```
f storage 18311740819329882151 /dev/disk/by-id/ata-WDC_WD80EFAX-68LHPN0_7HJSWL7I
```

ZFS automatically started the resilvering process (copying data to the new disk). I wasn't sure how long that would take...

```
$ zpool status
  pool: storage
 state: DEGRADED
status: One or more devices is currently being resilvered.  The pool will
    continue to function, possibly in a degraded state.
action: Wait for the resilver to complete.
  scan: resilver in progress since Thu Nov 15 17:01:06 2018
    7.97G scanned out of 7.51T at 233M/s, 9h22m to go
    2.56G resilvered, 0.10% done
config:
```

```
11
12      NAME                                      STATE     READ WRITE CKSUM
13      storage                                   DEGRADED     0     0     0
14        mirror-0                                ONLINE       0     0     0
15          ata-WDC_WD80EFZX-68UW8N0_VJHDBDGX     ONLINE       0     0     0
16          ata-WDC_WD80EFAX-68KNBN0_VAGASE7L     ONLINE       0     0     0
17        mirror-1                                ONLINE       0     0     0
18          ata-WDC_WD80EFZX-68UW8N0_VJHD6BAX     ONLINE       0     0     0
19          ata-WDC_WD80EFAX-68KNBN0_VAGA5BPL     ONLINE       0     0     0
20        mirror-2                                DEGRADED     0     0     0
21          ata-WDC_WD80EFZX-68UW8N0_VJHD982X     ONLINE       0     0     0
22          replacing-1                           DEGRADED     0     0     0
23            18311740819329882151                UNAVAIL      0     0     0  w
24            ata-WDC_WD80EFAX-68LHPN0_7HJSWL7F   ONLINE       0     0     0  (
25
26  errors: No known data errors
```

The resilvering completed in 5 hours and 53 minutes. A figure I'm very satisfied with. In this mirrored setup the data is at risk whilst resilvering completes, so the quicker, the better.

```
1   $ zpool status
2     pool: storage
3    state: ONLINE
4     scan: resilvered 2.50T in 5h53m with 0 errors on Thu Nov 15 22:54:41 201
5   config:
6
7       NAME                                    STATE     READ WRITE CKSUM
8       storage                                 ONLINE       0     0     0
9         mirror-0                              ONLINE       0     0     0
10          ata-WDC_WD80EFZX-68UW8N0_VJHDBDGX   ONLINE       0     0     0
11          ata-WDC_WD80EFAX-68KNBN0_VAGASE7L   ONLINE       0     0     0
12        mirror-1                              ONLINE       0     0     0
13          ata-WDC_WD80EFZX-68UW8N0_VJHD6BAX   ONLINE       0     0     0
14          ata-WDC_WD80EFAX-68KNBN0_VAGA5BPL   ONLINE       0     0     0
15        mirror-2                              ONLINE       0     0     0
16          ata-WDC_WD80EFZX-68UW8N0_VJHD982X   ONLINE       0     0     0
17          ata-WDC_WD80EFAX-68LHPN0_7HJSWL7F   ONLINE       0     0     0
18
19  errors: No known data errors
```

ZFS is easy to setup and use for the most part. It *feels* solid. Stable. If all disk replacements are this easy I will be very happy.

You can email **hello@jordanelver.co.uk** or tweet me at **@jordelver**

Find me on GitHub and Stack Overflow