

Test report for zpool and spares on OpenIndiana 148

```
# zpool create -f mimatedata \  
  raidz2 c4t1d0 c4t2d0 c4t3d0 c4t4d0 c4t5d0 \  
  spare c4t6d0 c4t7d0 \  
  log mirror c6t0d0s0 c6t1d0s0 \  
  cache c6t0d0s1 c6t1d0s1 \  
# zfs set compression=on mimatedata \  
# zfs create -o compression=off mimatedata/tmp
```

Then fill up the disk a bit

```
# mkfile 1024g /mimatedata/tmp/file1
```

While this is running, unplugged a c4t5d0, after some time resilver starts

```
root@mime-oi:/mimatedata/tmp# zpool status mimatedata  
pool: mimatedata  
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 Fri Mar  4 12:05:17 2011  
  15.0G scanned out of 64.0G at 223M/s, 0h3m to go  
  2.99G resilvered, 23.46% done  
config:  
  
   NAME                STATE          READ  WRITE CKSUM  
mimatedata             DEGRADED       0     0     0  
  
   raidz2-0            DEGRADED       0     0     0  
     c4t1d0             ONLINE         0     0     0  
     c4t2d0             ONLINE         0     0     0  
     c4t3d0             ONLINE         0     0     0  
     c4t4d0             ONLINE         0     0     0  
     spare-4           REMOVED        0     0     0  
     c4t5d0             REMOVED        0     0     0  
     c4t6d0             ONLINE         0     0     0 (resilvering)  
logs  
  mirror-1            ONLINE         0     0     0  
    c6t0d0s0           ONLINE         0     0     0  
    c6t1d0s0           ONLINE         0     0     0  
cache  
  c6t0d0s1            ONLINE         0     0     0  
  c6t1d0s1            ONLINE         0     0     0  
spares  
  c4t6d0              INUSE          currently in use  
  c4t7d0              AVAIL  
  
errors: No known data errors
```

Then, after resilver, I unplugged c4t4d0, waited for resilver, and then unplugged c4t2d0. Theoretically, the pool should work well after these changes, but in practice, it becomes unreadable, just hanging:

```

root@mime-oi:/mimedata/tmp# zpool status -v mimedata
  pool: mimedata
  state: DEGRADED
status: One or more devices are faulted in response to IO failures.
action: Make sure the affected devices are connected, then run 'zpool clear'.
   see: http://www.sun.com/msg/ZFS-8000-RC
  scan: resilvered 24.9G in 0h9m with 0 errors on Fri Mar  4 12:24:14 2011
config:

```

| NAME | STATE | READ | WRITE | CKSUM |
|----------|----------|------------------|-------|-------|
| mimedata | DEGRADED | 1 | 0 | 0 |
| raidz2-0 | DEGRADED | 6 | 0 | 0 |
| c4t1d0 | ONLINE | 0 | 0 | 0 |
| c4t2d0 | ONLINE | 12 | 572 | 0 |
| c4t3d0 | ONLINE | 0 | 0 | 0 |
| spare-3 | REMOVED | 0 | 0 | 0 |
| c4t4d0 | REMOVED | 0 | 0 | 0 |
| c4t7d0 | ONLINE | 0 | 0 | 0 |
| spare-4 | REMOVED | 0 | 0 | 0 |
| c4t5d0 | REMOVED | 0 | 0 | 0 |
| c4t6d0 | ONLINE | 0 | 0 | 0 |
| logs | | | | |
| mirror-1 | ONLINE | 0 | 0 | 0 |
| c6t0d0s0 | ONLINE | 0 | 0 | 0 |
| c6t1d0s0 | ONLINE | 0 | 0 | 0 |
| cache | | | | |
| c6t0d0s1 | ONLINE | 0 | 0 | 0 |
| c6t1d0s1 | ONLINE | 0 | 0 | 0 |
| spares | | | | |
| c4t6d0 | INUSE | currently in use | | |
| c4t7d0 | INUSE | currently in use | | |

```

errors: Permanent errors have been detected in the following files:

```

```

<metadata>:<0x23>

```

```

root@mime-oi:/mimedata/tmp# cfdiskadm | grep ^sata
sata0/0::dsk/c6t0d0      disk      connected  configured  ok
sata0/1::dsk/c6t1d0      disk      connected  configured  ok
sata0/2                  sata-port empty      unconfigured ok
sata0/3                  sata-port empty      unconfigured ok
sata0/4                  sata-port empty      unconfigured ok
sata0/5                  sata-port empty      unconfigured ok
sata1/0::dsk/c4t0d0      disk      connected  configured  ok
sata1/1::dsk/c4t1d0      disk      connected  configured  ok
sata1/2                  sata-port empty      unconfigured ok
sata1/3::dsk/c4t3d0      disk      connected  configured  ok
sata1/4                  sata-port empty      unconfigured ok
sata1/5                  sata-port empty      unconfigured ok
sata1/6::dsk/c4t6d0      disk      connected  configured  ok
sata1/7::dsk/c4t7d0      disk      connected  configured  ok
root@mime-oi:/mimedata/tmp#

```

Update 2010-03-04 14:15 CET

I just tested on another system. This one, not in production yet, has a mirrored **rpool** and a 14-drive RAID10 pool named **tos-data**. I started a copy from a Windows machine into this CIFS share just to generate some traffic. Then I did a `zfs detach` of one side of each of the mirrors for **tos-data** and created a new 5-drive `raidz2` pool name **jalla** with two dedicated spares. I started a `dd` to fill it up and plugged one drive, waited for it to resilver and plugged another, again waited for the resilver to finish and plugged the third. The server now hangs on *all* pools. I've also tested removing drives from mirrors and waiting for them to resilver to spares. This seems to work as expected, although I doubt booting from one will work without `grub` being installed.

Conclusion (updated)

It seems ZFS doesn't treat spares as good replicas for the pool. If two drives dies or are removed and then are replaced by spares in RAIDz2, resilver finishes, removing the third drive will make *all the pools* on the server unavailable. I guess this is not as intended, as it effectively reduces the value of spares to a mere minimum.

Roy Sigurd Karlsbakk <roy@karlsbakk.net>