

# Solaris RAID failover problems

**Source:** <http://unix.derkeiler.com/Newsgroups/comp.unix.solaris/2004-06/1234.html>

---

**From:** Brian Mengel (*mengel\_at\_microserve.net*)

**Date:** 06/18/04

Date: Fri, 18 Jun 2004 14:26:11 -0400

Greetings,

I'm presently trying to configure a Sunfire V120 with 2 36GB drives. My intent is to mirror the drives completely, allowing for failover to the secondary drive in the event of a primary drive failure. The system has been patched with all of the relevant and current patches.

My only stumbling block at this point is having the server boot from the secondary disk after the primary has been removed. Below are the errors I receive on boot up, and copies of all information I can think to provide on my system, including my procedure for configuring the RAID.

Any help on this would be greatly appreciated as I have a partner waiting on this configuration!

TIA

-b

Credit to the following web sites for at least getting me this far:

[http://www.tech-recipes.com/modules.php?name=Recipes&rx\\_id=225](http://www.tech-recipes.com/modules.php?name=Recipes&rx_id=225)

[http://www.unix.ualberta.ca/SUN/SDS\\_cheatsheet](http://www.unix.ualberta.ca/SUN/SDS_cheatsheet)

LOM event: +23h2m3s host power on

€

Sun Fire V120 (UltraSPARC-IIe 648MHz), No Keyboard

OpenBoot 4.0, 1024 MB memory installed, Serial #

Ethernet address 0:3:ba:2b:61:5c, Host ID: 832b615c.

Drive not ready

Can't read disk label.

Can't open disk label package

Boot device: rootmirror File and args:

SunOS Release 5.9 Version Generic 64-bit

Copyright 1983-2002 Sun Microsystems, Inc. All rights reserved.

## comp.unix.solaris: Solaris RAID failover problems

Use is subject to license terms.

WARNING: md: d10: (Unavailable) needs maintenance

WARNING: forceload of misc/md\_trans failed

WARNING: forceload of misc/md\_raid failed

WARNING: forceload of misc/md\_hotspares failed

WARNING: forceload of misc/md\_sp failed

configuring IPv4 interfaces: eri0.

Hostname: cnr

metainit: cnr: stale databases

Insufficient metadevice database replicas located.

Use metadb to delete databases which are broken.

Ignore any "Read-only file system" error messages.

Reboot the system when finished to reload the metadevice database.

After reboot, repair any broken database replicas which were deleted.

At this point, I can remove the metadbs that were on the failed disk, but rebooting after this produces a panic:

WARNING: forceload of misc/md\_trans failed

WARNING: forceload of misc/md\_raid failed

WARNING: forceload of misc/md\_hotspares failed

WARNING: forceload of misc/md\_sp failed

configuring IPv4 interfaces: eri0.

Hostname: cnr

WARNING: md: d11: (Unavailable) needs maintenance

WARNING: md: d14: (Unavailable) needs maintenance

The system is coming up. Please wait.

WARNING: md: d15: (Unavailable) needs maintenance

WARNING: md: d13: (Unavailable) needs maintenance

checking ufs filesystems

/dev/md/rdisk/d3: is clean.

/dev/md/rdisk/d5: is clean.

starting rpc services: rpcbind done.

Setting netmask of eri0 to 255.255.255.128

Setting default IPv4 interface for multicast: add net 224.0/4: gateway

cnr

syslog service starting.

automount: /home: already mounted

volume management starting.

Jun 8 13:53:37 cnr metadevadm: Invalid device relocation information detected i

n Solaris Volume Manager

Jun 8 13:53:37 cnr metadevadm: Please check the status of the following disk(s)

:

Jun 8 13:53:37 cnr metadevadm: c1t0d0

panic[cpu0]/thread=2a1000d1d40: BAD TRAP: type=31 rp=2a1000d1500  
addr=3080006350

comp.unix.solaris: Solaris RAID failover problems

0 mmu\_fsr=0

sched: trap type = 0x31

addr=0x30800063500

pid=0, pc=0x108a4e0, sp=0x2a1000d0da1, tstate=0x880001602, context=0x0

g1-g7: 149a000, 7fff, 0, 1, 1, 10, 2a1000d1d40

000002a1000d1230 unix:die+a4 (31, 2a1000d1500, 30800063500, 0, 1485ff2, 0)

%10-3: 0000000000000000 0000030000063508 000002a1000d1500  
000002a1000d13f8

%14-7: 0000000000000031 0000000000000001 00000000011557a8  
0000030000df1b30

000002a1000d1310 unix:trap+874 (2a1000d1500, 0, 10000, 10200, 308, 1)

%10-3: 0000000000000001 0000000000000000 0000000001436f38  
0000000000000031

%14-7: 0000000000000006 0000000000000001 0000000000000000  
0000000000000000

000002a1000d1450 unix:ktl0+48 (30000063508, 0, 20, 7fffffff8, 0, d0)

%10-3: 0000000000000003 0000000000001400 0000000880001602  
000000000102a0cc

%14-7: 000000000000003c 0000000001400090 0000000000000000  
000002a1000d1500

000002a1000d15a0 unix:kstat\_rele+20 (fffffffffffffff, 3, 4,  
30000393f28, 300003

66880, 2c0)  
%10-3: 0000000001428d08 0000030000df1b30 0000000000000000  
0000000001476000

%14-7: 000000000000003c 0000000001400090 000003000025d2e0  
00000300001c5a60

000002a1000d1650 md:md\_layered\_close+d4 (fffffffffffffff, 3, 4, 0, 0,  
10)

%10-3: ffffffffffffffff 0000000000000002 0000030000393f28  
00000000ffffffff

%14-7: 00000000ffffffff 0000002000000009 0000000000000000  
0000000000000018

000002a1000d1700 md\_stripe:stripe\_close\_all\_devs+dc (30000289d44,  
30000289d44, 0

, 20, 1485ff2, 0)  
%10-3: 0000000000000001 0000000000000001 0000000000000002  
0000000000000000

%14-7: 0000000000000002 0000030000289ce8 0000000001485a10  
0000030000289d58

000002a1000d17b0 md\_stripe:stripe\_close+88 (550000000b, 3, 4,  
30000393f28, 2, 0)

%10-3: 0000000000000000 0000000000000002 00000300004b928  
0000030000289ce8

%14-7: 00000300004b928 0000000000000015 0000000000000002  
0000000000000000

000002a1000d1860 md\_mirror:mirror\_probe\_close\_all\_devs+b8 (550000000b,  
300001faf

## comp.unix.solaris: Solaris RAID failover problems

```
b8, 1, 1, 300001fad88, 300001fade4)
%10-3: 00000000012c14e8 0000000000000001 0000000000000001
000000000000ffff
%14-7: 00000300001fad88 00000300001fade4 00000300001fafb8
00000300001fade4
000002a1000d1910 md_mirror:mirror_probe_dev+208 (108, 1, 1436f38,
1436f38, 1485f
f2, 0)
%10-3: 0000000000000004 0000000000000001 0000000000000001
000000000000ffff
%14-7: 000003000004b568 0000000000000001 00000300001fad88
00000300001fade4
000002a1000d19d0 md:md_probe_one+40 (3000135e580, 2a1000d1d40, 20,
1485fe8, 2a10
00d1d40, 0)
%10-3: 00000000012c83c4 000000000149a0a8 0000030000058d98
fffffffffffffff
%14-7: 0000000001400090 000000000142cf70 0000000001441800
00000300001eea48
000002a1000d1a80 md:md_daemon+220 (0, 149a078, 1436f38, 1436f38,
2a1000ddd40, 0)
%10-3: 00000000011be3fc 000003000135e580 0000000000000000
000002a1000d7d40
%14-7: 000000000149a0a8 000000000149a0a0 0000000000000000
0000030000174000
```

syncing file systems...

```
panic[cpu0]/thread=2a1000d1d40: md: writer lock is held
dumping to /dev/md/dsk/d1, offset 214827008, content: kernel
WARNING: /pci@1f,0/pci@1/scsi@8 (glm0):
    got SCSI bus reset
WARNING: md: d24: read error on /dev/dsk/c1t1d0s4
```

```
$ cat vfstab
```

```
#device device mount FS fsck mount
mount
#to mount to fsck point type pass at
boot options
#
fd - /dev/fd fd - no -
/proc - /proc proc - no -
/dev/md/dsk/d1 -- swap - no -
/dev/md/dsk/d0 /dev/md/rdisk/d0 / ufs 1 no -
/dev/md/dsk/d4 /dev/md/rdisk/d4 /var ufs 1 no -
/dev/md/dsk/d5 /dev/md/rdisk/d5 /home ufs 2 yes -
/dev/md/dsk/d3 /dev/md/rdisk/d3 /opt ufs 2 yes -
swap - /tmp tmpfs - yes -
```

[http://www.tech-recipes.com/modules.php?name=Recipes&rx\\_id=225](http://www.tech-recipes.com/modules.php?name=Recipes&rx_id=225)

[http://www.unix.ualberta.ca/SUN/SDS\\_cheatsheet](http://www.unix.ualberta.ca/SUN/SDS_cheatsheet)

## comp.unix.solaris: Solaris RAID failover problems

```
# metadb
  flags first blk block count
  a m p luo 16 8192
/dev/dsk/c1t1d0s6
  a p luo 8208 8192
/dev/dsk/c1t1d0s6
  a p luo 16 8192
/dev/dsk/c1t1d0s7
  a p luo 8208 8192
/dev/dsk/c1t1d0s7
  a u 16 8192
/dev/dsk/c1t0d0s6
  a u 8208 8192
/dev/dsk/c1t0d0s6
  a u 16 8192
/dev/dsk/c1t0d0s7
  a u 8208 8192
/dev/dsk/c1t0d0s7

# metastat
d4: Mirror
  Submirror 0: d14
    State: Resyncing
  Submirror 1: d24
    State: Okay
  Resync in progress: 0 % done
  Pass: 1
  Read option: roundrobin (default)
  Write option: parallel (default)
  Size: 20480121 blocks

d14: Submirror of d4
  State: Resyncing
  Size: 20480121 blocks
  Stripe 0:
    Device Start Block Dbase State Reloc Hot Spare
    c1t0d0s4 0 No Resyncing Yes

d24: Submirror of d4
  State: Okay
  Size: 20480121 blocks
  Stripe 0:
    Device Start Block Dbase State Reloc Hot Spare
    c1t1d0s4 0 No Okay Yes

d3: Mirror
  Submirror 0: d13
    State: Resyncing
  Submirror 1: d23
    State: Okay
  Resync in progress: 0 % done
```

comp.unix.solaris: Solaris RAID failover problems

Pass: 1  
Read option: roundrobin (default)  
Write option: parallel (default)  
Size: 11264211 blocks

d13: Submirror of d3  
State: Resyncing  
Size: 11264211 blocks  
Stripe 0:  
Device Start Block Dbase State Reloc Hot Spare  
c1t0d0s3 0 No Resyncing Yes

d23: Submirror of d3  
State: Okay  
Size: 11264211 blocks  
Stripe 0:  
Device Start Block Dbase State Reloc Hot Spare  
c1t1d0s3 0 No Okay Yes

d1: Mirror  
Submirror 0: d11  
State: Resyncing  
Submirror 1: d21  
State: Okay  
Resync in progress: 5 % done  
Pass: 1  
Read option: roundrobin (default)  
Write option: parallel (default)  
Size: 2097414 blocks

d11: Submirror of d1  
State: Resyncing  
Size: 2097414 blocks  
Stripe 0:  
Device Start Block Dbase State Reloc Hot Spare  
c1t0d0s1 0 No Resyncing Yes

d21: Submirror of d1  
State: Okay  
Size: 2097414 blocks  
Stripe 0:  
Device Start Block Dbase State Reloc Hot Spare  
c1t1d0s1 0 No Okay Yes

d0: Mirror  
Submirror 0: d10  
State: Resyncing  
Submirror 1: d20  
State: Okay  
Resync in progress: 0 % done  
Pass: 1

comp.unix.solaris: Solaris RAID failover problems

Read option: roundrobin (default)  
Write option: parallel (default)  
Size: 26622135 blocks

d10: Submirror of d0

State: Resyncing  
Size: 26622135 blocks  
Stripe 0:

Device Start Block Dbase State Reloc Hot Spare  
c1t0d0s0 0 No Resyncing Yes

d20: Submirror of d0

State: Okay  
Size: 26622135 blocks  
Stripe 0:

Device Start Block Dbase State Reloc Hot Spare  
c1t1d0s0 0 No Okay Yes

d5: Mirror

Submirror 0: d15  
State: Resyncing  
Submirror 1: d25  
State: Okay  
Resync in progress: 6 % done  
Pass: 1  
Read option: roundrobin (default)  
Write option: parallel (default)  
Size: 10238616 blocks

d15: Submirror of d5

State: Resyncing  
Size: 10238616 blocks  
Stripe 0:

Device Start Block Dbase State Reloc Hot Spare  
c1t0d0s5 0 No Resyncing Yes

d25: Submirror of d5

State: Okay  
Size: 10238616 blocks  
Stripe 0:

Device Start Block Dbase State Reloc Hot Spare  
c1t1d0s5 0 No Okay Yes

Device Relocation Information:

Device Reloc Device ID  
c1t1d0 Yes id1,sd@SSEAGATE\_ST336605LSUN36G\_3FP164ND00002231F68T  
c1t0d0 Yes id1,sd@SSEAGATE\_ST336607LSUN36G\_3JA0X6VG000073164H8P

# prtconf

System Configuration: Sun Microsystems sun4u  
Memory size: 1024 Megabytes

System Peripherals (Software Nodes):

SUNW,UltraAX-i2

- packages (driver not attached)
  - terminal-emulator (driver not attached)
  - deblocker (driver not attached)
  - obp-tftp (driver not attached)
  - disk-label (driver not attached)
  - SUNW,builtin-drivers (driver not attached)
  - dropins (driver not attached)
  - kbd-translator (driver not attached)
  - ufs-file-system (driver not attached)
- chosen (driver not attached)
- openprom (driver not attached)
  - client-services (driver not attached)
- options, instance #0
- aliases (driver not attached)
- memory (driver not attached)
- virtual-memory (driver not attached)
- pci, instance #0
  - pci, instance #0
    - ebus, instance #0
      - flashprom (driver not attached)
      - eeprom (driver not attached)
      - idprom (driver not attached)
      - SUNW,lomh (driver not attached)
    - pmu (driver not attached)
      - i2c (driver not attached)
        - temperature (driver not attached)
        - dimmm (driver not attached)
        - dimmm (driver not attached)
        - i2c-nvram (driver not attached)
        - idprom (driver not attached)
      - motherboard-fru (driver not attached)
    - fan-control (driver not attached)
  - lompp (driver not attached)
  - isa, instance #1
    - power, instance #0
    - serial, instance #0
    - serial, instance #1
  - network, instance #0
  - usb, instance #0
  - ide, instance #0
    - disk (driver not attached)
    - cdrom (driver not attached)
    - sd, instance #30
  - network, instance #1
  - usb, instance #1
  - pci, instance #1
    - scsi, instance #0
      - disk (driver not attached)

tape (driver not attached)  
sd, instance #0  
sd, instance #1  
sd, instance #2 (driver not attached)  
sd, instance #3 (driver not attached)  
sd, instance #4 (driver not attached)  
sd, instance #5 (driver not attached)  
sd, instance #6 (driver not attached)  
sd, instance #7 (driver not attached)  
sd, instance #8 (driver not attached)  
sd, instance #9 (driver not attached)  
sd, instance #10 (driver not attached)  
sd, instance #11 (driver not attached)  
sd, instance #12 (driver not attached)  
sd, instance #13 (driver not attached)  
sd, instance #14 (driver not attached)  
scsi, instance #1  
disk (driver not attached)  
tape (driver not attached)  
sd, instance #15 (driver not attached)  
sd, instance #16 (driver not attached)  
sd, instance #17 (driver not attached)  
sd, instance #18 (driver not attached)  
sd, instance #19 (driver not attached)  
sd, instance #20 (driver not attached)  
sd, instance #21 (driver not attached)  
sd, instance #22 (driver not attached)  
sd, instance #23 (driver not attached)  
sd, instance #24 (driver not attached)  
sd, instance #25 (driver not attached)  
sd, instance #26 (driver not attached)  
sd, instance #27 (driver not attached)  
sd, instance #28 (driver not attached)  
sd, instance #29 (driver not attached)  
SUNW,UltraSPARC-IIe (driver not attached)  
pseudo, instance #0

## Description

Add raid to your Solaris system

## Directions

Once you're done that, you'll need to determine how you want to lay out your disks. The following assumes that:

- 1 – You have 2 disks – c1t0d0 (disk0) and c1t1d0 (disk1).
- 2 – The system installed only on disk0, and disk1 is unused.
- 3 – Each disk has the following slices: (for a 36G sample drive)

0 – / 12GB

1 – swap 2GB

- 2 – whole-disk
- 3 – /opt 5.3 GB
- 4 – /var 9.6GB
- 5 – /home 4.8GB
- 6 – unassigned 64-MB
- 7 – unassigned 64-MB

Adjust the above to match whatever your preferred layout is. This is only for a simple example. Slices 6 and 7 are for Meta-Database logging. If you don't have 128MB of free space to spare, then try and make some space (ie., sacrifice some swap if you have to).

You need to duplicate your layout from disk0 to disk1. It's fairly important that the disk geometry matches. Metadevices work at the block-level of the disk, and if one disk has fewer blocks than the other you'll wind up making a mess. Once you're sure you're ready to proceed, dump the layout from disk0 to disk1 thusly:

```
prtvtoc /dev/rdisk/c0t0d0s2 | fmthard -s - /dev/rdisk/c0t1d0s2
```

Second, you need to create your meta-databases. This is for logging, and all but eliminates the need for fsck to run after a dirty shutdown. Do the following:

```
metadb -af -c 2 /dev/dsk/c1t0d0s6 /dev/dsk/c1t0d0s7  
metadb -af -c 2 /dev/dsk/c1t1d0s6 /dev/dsk/c1t1d0s7
```

This adds (-a) 2 (-c for count) meta-databases in each of the slices. If you have more disks, you can span the databases across multiple disks for better performance and fault-tolerance.

The next step is to create your raid-devices. In a two-disk system, you're stuck with Raid0 and Raid1. Since Raid0 is almost pointless (you're doing this for redundancy, remember?!), we'll go with Raid1 – mirrored disks.

We'll deal with the following raid devices and members:

```
d0 – / mirror  
d10 – /dev/dsk/c1t0d0s0  
d20 – /dev/dsk/c1t1d0s0  
  
d1 – swap  
d11 – /dev/dsk/c1t0d0s1  
d21 – /dev/dsk/c1t1d0s1  
  
d3 – opt  
d13 – /dev/dsk/c1t0d0s3  
d23 – /dev/dsk/c1t1d0s3
```

```
d4 – var  
d14 – /dev/dsk/c1t0d0s4  
d24 – /dev/dsk/c1t1d0s4
```

```
d3 – home  
d13 – /dev/dsk/c1t0d0s5  
d23 – /dev/dsk/c1t1d0s5
```

The device names are somewhat arbitrary. In a simple setup like this, I use d0 to match up with a mirrored slice0, and d10 to indicate member 1 of d0 (member 1 d0 = d10, member 2 d0 = d20).

So create the raid devices and members:

```
metainit –f d10 1 1 /dev/dsk/c1t0d0s0  
metainit –f d20 1 1 /dev/dsk/c1t1d0s0  
metainit –f d0 –m d10
```

```
metainit –f d11 1 1 /dev/dsk/c1t0d0s1  
metainit –f d21 1 1 /dev/dsk/c1t1d0s1  
metainit –f d1 –m d11
```

```
metainit –f d13 1 1 /dev/dsk/c1t0d0s3  
metainit –f d23 1 1 /dev/dsk/c1t1d0s3  
metainit –f d3 –m d13
```

```
metainit –f d14 1 1 /dev/dsk/c1t0d0s4  
metainit –f d24 1 1 /dev/dsk/c1t1d0s4  
metainit –f d4 –m d14
```

```
metainit –f d15 1 1 /dev/dsk/c1t0d0s5  
metainit –f d25 1 1 /dev/dsk/c1t1d0s5  
metainit –f d5 –m d15
```

This initializes the devices. The command "metastat" will show you that the devices exist, but the mirror-halves aren't attached. So let's attach them:

```
metattach d0 d10  
metattach d1 d11  
metattach d3 d13  
metattach d4 d14  
metattach d5 d15
```

You've just attached the first half of the mirror. Yes, this is the disk that you're currently running on. Your data is still there.

Next, you need to ensure the system will use the metadevices. The root-filesystem is easy: (this command automatically edits the vfstab to make the root device d0)

metaroot d0

Next, you need to edit /etc/vfstab to change the swap device to use /dev/md/dsk/d1 as swap, along with the other partitions. While you're in there, turn on logging under the mount options for the root filesystem (d0). Double-check that you haven't screwed up. Save and exit if it all looks good. Below is a sample

```
cat /etc/vfstab
```

```
#device device mount FS fsck mount
mount
#to mount to fsck point type pass at
boot options
#
fd - /dev/fd fd - no -
/proc - /proc proc - no -
/dev/md/dsk/d1 -- swap - no -
/dev/md/dsk/d0 /dev/md/rdisk/d0 / ufs 1 no -
/dev/md/dsk/d4 /dev/md/rdisk/d4 /var ufs 1 no -
/dev/md/dsk/d5 /dev/md/rdisk/d5 /home ufs 2 yes -
/dev/md/dsk/d3 /dev/md/rdisk/d3 /opt ufs 2 yes -
swap - /tmp tmpfs - yes -
```

Once you're done, issue the following:

```
lockfs -fa
init 6
```

Watch your system come up. There will be some new messages, most notably the kernel complaining about not being able to forceload three raid modules:

```
forceload of misc/md_trans failed
forceload of misc/md_raid failed
forceload of misc/md_hotspares failed
```

You can ignore these messages. They're harmless. Basically, you haven't created any raid-devices that require those modules so they're refusing to load.

Now that your system is up (You didn't mess up vfstab, did you?!), you need to finish off the process. Log in and do this:

```
metattach d0 d20
metattach d1 d21
metattach d3 d23
metattach d4 d24
metattach d5 d25
```

You'll notice that your system is now a little slower, both commands

## comp.unix.solaris: Solaris RAID failover problems

took a moment to return, and your disks are going nuts. Look at the output of "metastat" and you'll see why – your disks are syncing.

You'll need to install the bootsector to your second disk so that you can boot from it. This is fairly easy to do:

```
installboot /usr/platform/^uname -i`/lib/fs/ufs/bootblk  
/dev/rdisk/c1t1d0s0
```

Now you'll want to store some information about the disks so that we can boot from it remotely in the event of a primary disk failure.

```
ls -l /dev/dsk/c1t0d0s2  
lrwxrwxrwx 1 root root 44 Jun 2 11:57  
/dev/rdisk/c1t0d0s2 -> /dev/rdisk/c1t0d0s2 ->  
../../devices/pci@1f,0/pci@1/scsi@8/sd@0,0:c,raw  
ls -l /dev/dsk/c1t1d0s2  
lrwxrwxrwx 1 root root 44 Jun 2 11:57  
/dev/rdisk/c1t1d0s2 -> /dev/rdisk/c1t1d0s2 ->  
../../devices/pci@1f,0/pci@1/scsi@8/sd@1,0:c,raw
```

These two pointers to the hard disks will be used to set up the system to boot to the mirror disk in the event that the primary fails. The system will need to be rebooted if the primary disk fails in operation, but it should fall to the secondary disk automatically on boot if everything is set up correctly. Take the links above and edit them to meet the format below, these two lines need to be placed in a file called myaliases in the root directory:

```
cat ./myaliases
```

```
devalias rootdisk /pci@1f,0/pci@1/scsi@8/disk@0,0  
devalias rootmirror /pci@1f,0/pci@1/scsi@8/disk@1,0
```

Next, execute the following commands to set up the nvram on the box to store these aliases for use at boot time:

```
eeeprom nvramrc=""`cat ./myaliases`"  
eeeprom boot-device='rootdisk rootmirror'  
eeeprom diag-device='rootdisk rootmirror'  
eeeprom "use-nvramrc?=true"
```