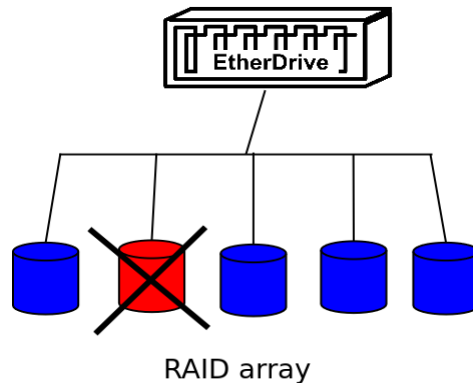


# Dealing with disk failures in a Coraid SR

## Objective

In this document we will explore some situations where a disk failure occurs within a Coraid SR and how to remedy the problem. We will also talk about some of the logging events that the SR will communicate through syslog.



## Problems and their solutions

For these configurations, we will assume that you are accessing the Coraid SR via cec, KVM, or a serial console. If you need help with configuring cec, please visit our [cec documentation](#). Later on, we are analyzing syslogs sent from a Coraid SR to a syslog server. Please follow our [syslog documentation](#) to enable syslog on your server.

## Single-disk failure within a RAID 5 LUN

In this example we have a three disk RAID 5 with a hot spare. We want to see what would happen if we “failed” a disk, so we will use the *fail* command to emulate a disk failure.

```
LD shelf 2> list -l
7 600.138GB online
7.0 600.138GB raid5 normal
7.0.0 normal 300.069GB 2.21
7.0.1 normal 300.069GB 2.22
7.0.2 normal 300.069GB 2.23
LD shelf 2> spare
2.20 300.069GB
LD shelf 2> fail 7.0.0
raid device 7.0.0 has failed
spare 2.20 allocated for 7.0.0
beginning recovery of disk 7.0.0
LD shelf 2> list -l
7 600.138GB online
7.0 600.138GB raid5 recovering,degraded 2.12%
7.0.0 replaced 300.069GB 2.20
7.0.1 normal 300.069GB 2.22
7.0.2 normal 300.069GB 2.23
LD shelf 2> spare
```

This shows how the RAID 5 array would act if one disk would fail. If another disk would fail in this situation before the recovery of disk 7.0.0, it would cause the LUN itself to fail. As this disk failure is only emulated, you could add the failed disk (now 2.21) as a spare and then wait for the LUN to recover from a degraded (but functional) state. Note that performance will be slower until the parity has been fully written to the new disk.

If a single disk were to fail when there is no spare allocated on the shelf, there are a couple options:

- a) Add a spare disk to the spare pool with *spare*. The LUN will then begin a rebuild using the new spare as a replacement for the failed disk as in the example above.
- b) Physically replace the failed disk, then replace it with itself using the *replace* command which will start a LUN rebuild.

### Allocating a new spare

With a RAID 1, 5 or 10, it is recommended to allocate at least one spare. If possible we recommend that you allocate multiple spares for added redundancy. In this example, as a continuation the previous section, we are allocating spares when there are none.

```
LD shelf 2> list -l
7 600.138GB online
7.0 600.138GB raid5 normal
 7.0.0 normal 300.069GB 2.20
 7.0.1 normal 300.069GB 2.22
 7.0.2 normal 300.069GB 2.23
LD shelf 2> spare
LD shelf 2> spare 2.19 2.21
LD shelf 2> spare
2.19 300.069GB
2.21 300.069GB
```

Here we allocate two global spares for allocation in any LUN that needs a replacement. You can allocate as many or as little as you prefer, depending on your needed redundancy and disk configuration. Notice that the LUN member disks are not linearly aligned; this is not a problem at all and we recommend that you do not worry about changing it. If you would still like to reorder the disks of the LUN please follow the steps outlined below.

### Reordering the member disks of a LUN

Please note that this is **NOT** a recommended solution as you are putting additional stress onto the disks by spinning them up and down. Additional stress means putting your data at risk.

If the order of the disks is important to you, what follows is a method to reallocate the disks in a linear order. Ensure that any LUN rebuilds finish before reordering the respective elements. Before halting the system, make sure that first the LUN has been unmounted from the host server.

```
LD shelf 2> list -l
7 600.138GB online
7.0 600.138GB raid5 normal
7.0.0 normal 300.069GB 2.20
7.0.1 normal 300.069GB 2.22
7.0.2 normal 300.069GB 2.23
LD shelf 2> spare
2.19 300.069GB
2.21 300.069GB
LD shelf 2> halt

System halted. Hit enter to reboot:
```

To rearrange the drives, you now have some options:

- a) Swap the drives with the disk carriers unchanged. This works with unnumbered disk carriers.
- b) Swap the drives with the disk carriers swapped. This works with numbered disk carriers. This is so the physical, external numbering is the same as the internal software numbering.

After swapping the drives 'list -l' output should be as it was previous to the failure & rebuild.

```
LD shelf 2> list -l
7 600.138GB online
7.0 600.138GB raid5 normal
7.0.0 normal 300.069GB 2.21
7.0.1 normal 300.069GB 2.22
7.0.2 normal 300.069GB 2.23
```

## Interpreting log files

We recommend that you setup remote logging to a syslog host. For help on how to do that, please reference our [syslog documentation](#). Now we will show some of the useful logs that will help you administer and understand the Coraid SR. Here are some examples of logs that are sent from the Coraid SR.

- **Testing logging**

The following example is produced using the *syslog* command on the SR. This is how you can send a test message from the SR to ensure that the syslog function is operating normally.

```
On the SR:  
LD shelf 2> syslog Hello World!  
Hello World!
```

```
On the syslog host:  
Sep 11 12:46:54 192.168.0.15 shelf_2 Hello World!
```

- **Failed disk**

Coraid's SR uses RAIDShield™ to detect soft failures in disks that are part of a LUN. Here we see the detection of a soft failure and then a hard failure of the device. This would demand a replacement or additional checking of the disk and the onlining of the LUN after the replacement.

```
On the syslog host:  
Sep 11 13:22:09 192.168.0.15 shelf_2 raidshield: unrecoverable error on disk 7.0.1 at block 1547078  
Sep 11 13:35:04 192.168.0.15 shelf_2 raid device 7.0.1 has failed  
Sep 11 13:35:04 192.168.0.15 shelf_2 unrecoverable failure on raid 7.0  
Sep 11 13:35:04 192.168.0.15 shelf_2 warning: config save error for 7.0.1
```

- **Failed spare**

RAIDShield™ also probes the spare or spares on the SR. This functionality has been added to later versions of the Coraid firmware. Here we see what was logged when we physically removed the spare, then later placed the drive back into the SR. If the spare fails to respond over a long period of time, you may have a failed disk. In this case you can simply use *rm spare* to remove the disk from the spare pool, physically replace it, then add the new disk to the spare pool using *spare*.

```
On the syslog host:  
Sep 23 11:24:26 192.168.0.15 shelf_2 Warning: spare disk 2.18 is not responding to periodic check.  
  
Sep 23 11:25:26 192.168.0.15 shelf_2 Spare disk 2.18 is now responding.
```

- **Spare allocation**

In this example we see that a disk in a redundant LUN has failed. As we have a spare allocated, it begins the recovery. If the syslog server receives this log message it is necessary to replace the disk that has failed and to allocate a replacement spare.

```
On the syslog host:  
Sep 11 13:12:58 192.168.0.15 shelf_2 raid device 7.0.0 has failed  
Sep 11 13:12:58 192.168.0.15 shelf_2 spare 2.20 allocated for 7.0.0  
Sep 11 13:12:58 192.168.0.15 shelf_2 beginning recovery of disk 7.0.0
```

- **Failure of a RAID**

Below we see a LUN failing because there is no spare (or the spare is not big enough) to replace the failed disk. It will need a spare allocated to rebuild the array.

*On the syslog host:*

```
Sep 11 13:18:40 192.168.0.15 shelf_2 raid device 7.0.0 has failed  
Sep 11 13:18:40 192.168.0.15 shelf_2 aborted recovery of disk 7.0.0  
Sep 11 13:18:40 192.168.0.15 shelf_2 no spare large enough for 7.0.0
```

*Additional Information*

If you are having difficulty setting up anything contained within this document or have questions/comments about the document itself, please contact us at [support@coraid.com](mailto:support@coraid.com). For more information, please refer to the [SR software user manual](#).

©2008 [Coraid, Inc.](#)