

Release notes for Flash Recovery Tool Version 10.0(2)

Problem Description

After several months or years in continuous operation, underlying boot flash devices on NEXUS 7000 SUP2/2E supervisor boards may become unresponsive and go out of configuration. Even though the interruption is transient, it can eventually cause the boot flash device to get remounted as read-only by NXOS. Once system gets into this situation, further configuration copy procedures may fail. This issue is tracked using CSCus22805.

Background

Each NEXUS 7000 Supervisor board are equipped with 2 identical eUSB flash devices in a RAID1 mirror configuration. Called primary and mirror, they together provide for non-volatile repositories for storing boot images, startup configuration and other persistent application data.

Over years in service, one of these devices may get disconnected from the USB bus. This causes the RAID software to drop the affected device to be removed from its configuration. System still can function normally with the remaining working device.

However, if the second device also experiences similar issue and drops out of the RAID array, boot flash devices will be re-mounted as read-only preventing configuration copying. Even though there is no operational impact on systems running in this state, a reload of the affected supervisor is needed to recover from this situation. Moreover, the latest running configuration may be lost in the event of a power outage.

Problem Symptoms

To display the current status of the RAID devices, please run the following command. If the system has a standby supervisor, attach to it first and run the command as well.

```
switch# show system internal raid | grep -A 1 "Current RAID status info"
Current RAID status info:
RAID data from CMOS = 0xa5 0xe1
switch#
```

Last number in the RAID data indicate the number of disks failed.

```
0xf0 ==>> No failures reported
0xe1 ==>> Primary flash failed
0xd2 ==>> Mirror flash failed
0xc3 ==>> Both primary and mirror failed
```

Recovery Steps

While system can operate with only one flash device, it's highly recommended to recover and add the removed flash back into RAID configuration. Second flash device can also get into this condition over time triggering the read-only mode.

Single flash failure on active or standby

Systems running with single flash failure can be repaired while in service using the flash recovery tool.

Dual flash failure on the standby

This is a recoverable situation by reloading the standby and making sure that the flashes are healthy once it comes back online. You can use flash recovery tool to put both flashes back into service.

Dual flash failures on Active

This is not a recoverable situation unless there is at least one working flash on the standby. Otherwise, switch needs to be reloaded during next maintenance window. If the standby is healthy a switchover can be attempted and then recover the old active. Latest running configurations need to be saved external to the switch to be restored after reload.

Flash Recovery Tool

A flash recovery tool is available to be download from Cisco support site. This works as a custom plug-in that can be run using the "load" CLI.

- To run the tool, download and copy it to bootflash:/volatile:/slot0: and run the load command.
- Tool automatically fixes any single flash errors when present.
- If a standby available, it will copy itself to standby and run there.
- No side effects if there are no errors reported at the time.
- Tool will not attempt dual flash recovery either on active or standby.

Running the Flash Recovery Tool

Download the tool from Cisco support and copy it volatile: on the active supervisor.

```
switch# load volatile:n7000-s2-flash-recovery-tool.10.0.2.gbin
Loading plugin version 10.0(2)
#####
Warning: debug-plugin is for engineering internal use only!
For security reason, plugin image has been deleted.
#####
INFO: Running on active slot 6, checking if a ha-standby is available...
INFO: Standby present in slot 5. Checking drive status...
#####
Warning: debug-plugin is for engineering internal use only!
For security reason, plugin image has been deleted.
#####
INFO: Running on the standby in slot 5, Checking RAID status...
INFO: Primary=(sdd) Secondary=sdsc(sdc) Working=sdsc
WARNING: Attempting recovery of primary device sdd
INFO: Removing /dev/sdd from RAID configuration...
INFO: Resetting primary flash...
INFO: Found primary device sdd in 9 seconds.
INFO: Running health checks on the recovered device /dev/sdd...
```

```

INFO: Basic I/O tests passed. /dev/sdd looks healthy and responsive.
INFO: Verifying RAID configuration. Got primary=sdd Secondary=sdh
INFO: Adding sdd3 back into md3 RAID configuration...
INFO: sdc3 is already a part of md3.
INFO: Adding sdd4 back into md4 RAID configuration...
INFO: sdc4 is already a part of md4.
INFO: Adding sdd5 back into md5 RAID configuration...
INFO: sdc5 is already a part of md5.
INFO: Adding sdd6 back into md6 RAID configuration...
INFO: sdc6 is already a part of md6.
INFO: Resetting RAID status in CMOS...
WARNING: Flash recovery attempted on module 5.
INFO: A detailed copy of the this log was saved as volatile:flash_repair_log_mod5.tgz.
INFO: Recovery procedures complete on module 5.
INFO: Please check for any errors in previous messages.
INFO: Run 'show system internal file /proc/mdstat' and check 'up status' [UU] for all disks.
INFO: Run 'show diagnostic result module <slot#>' on all available supervisor slots.
INFO: And restart CompactFlash test (7) instances if not in running state.
Loading plugin version 10.0(2)
INFO: Now starting the flash recovery procedures on active.
INFO: Primary=sdc(sdc) Secondary=(sdb) Working=sdc
WARNING: Attempting recovery of secondary device sdb
INFO: Removing /dev/sdb from RAID configuration...
INFO: Resetting secondary flash...
INFO: Found secondary device sdb in 9 seconds.
INFO: Running health checks on the recovered device /dev/sdb...
INFO: Basic I/O tests passed. /dev/sdb looks healthy and responsive.
INFO: Verifying RAID configuration. Got primary=sdc Secondary=sdb
INFO: sdc3 is already a part of md3.
INFO: Adding sdb3 back into md3 RAID configuration...
INFO: sdc4 is already a part of md4.
INFO: Adding sdb4 back into md4 RAID configuration...
INFO: sdc5 is already a part of md5.
INFO: Adding sdb5 back into md5 RAID configuration...
INFO: sdc6 is already a part of md6.
INFO: Adding sdb6 back into md6 RAID configuration...
INFO: Resetting RAID status in CMOS...
WARNING: Flash recovery attempted on module 6.
INFO: A detailed copy of the this log was saved as volatile:flash_repair_log_mod6.tgz.
INFO: Recovery procedures complete on module 6.
INFO: Please check for any errors in previous messages.
INFO: Run 'show system internal file /proc/mdstat' and check 'up status' [UU] for all disks.
INFO: Run 'show diagnostic result module <slot#>' on all available supervisor slots.
INFO: And restart CompactFlash test (7) instances if not in running state.
switch#

```

Output messages are self-explanatory. Check for any errors reported in the output and report back to Cisco support. After running the tool any disks that are not participating in the RAID will be added back into it's configuration. Users will be able to check the latest status of the RAID arrays using the following command:

```

switch# show system internal file /proc/mdstat
Personalities : [raid1]
md6 : active raid1 sdd6[2] sdc6[0]
      77888 blocks [2/1] [U_]
      resync=DELAYED

md5 : active raid1 sdd5[2] sdc5[0]
      78400 blocks [2/1] [U_]
      resync=DELAYED

md4 : active raid1 sdd4[2] sdc4[0]
      39424 blocks [2/1] [U_]
      resync=DELAYED

md3 : active raid1 sdd3[2] sdc3[0]
      1802240 blocks [2/1] [U_]

```

```
[=>.....] recovery = 8.3% (151360/1802240) finish=2.1min s
peed=12613K/sec

unused devices: <none>
switch#
```

In the above output, the mirror flash was added back into the RAID configuration and the sync is in progress in the background to update the new disk with the up-to-date data from the working disk. The sync process may take few minutes to complete. All partitions should look like the following when it's complete:

```
switch# show system internal file /proc/mdstat
Personalities : [raid1]
md6 : active raid1 sdd6[1] sdc6[0]
      77888 blocks [2/2] [UU]

md5 : active raid1 sdd5[1] sdc5[0]
      78400 blocks [2/2] [UU]

md4 : active raid1 sdd4[1] sdc4[0]
      39424 blocks [2/2] [UU]

md3 : active raid1 sdd3[1] sdc3[0]
      1802240 blocks [2/2] [UU]

unused devices: <none>
switch#
```

If any of the disks are not showing the status as [2/2] [UU] (means 2 out of 2 disks are good and U stands for up and running), the recovery is incomplete.

Caveats

Dual flash failures

Tool will not attempt to recover from dual flash failures. Following error messages will be printed corresponding to the module when such a condition is encountered.

```
ERROR: Both disks are marked as failed. Cannot perform recovery.
ERROR: Please schedule downtime and reload supervisor card in slot <slot#> to recover.
ERROR: Please note that dual disk failure on active requires a switch reload to recover.
ERROR: Dual disk failure on standby can be recovered by reloading just the standby.
```

Recovery after Module Reloads

When an affected board with single or dual flash failure is rebooted, RAID recovery implemented in the startup scripts may not recover fully under certain conditions. If the recovery is incomplete, board may be running with only one flash in the RAID configuration with the second flash being unused. Even though “show system internal raid” CLI display the RAID status to be 0xf0, both disks may not be marked as up. Running the tool will recover from this condition.

Run Frequency

This tool can be scheduled to run at regular intervals to recover from flash errors as soon as they occur. However, minimum recommended frequency is once a week.

Simultaneous Executions

To protect the flash integrity, tool prevents itself from running while invoked simultaneously from multiple login sessions. It prints the following error messages for those conditions:

```
ERROR: Another instance of flash recovery tool may be already running on module <slot #>.
ERROR: If this is an error, please delete volatile:/FLASH_RECOVERY_IN_PROGRESS on module
<slot#> and retry.
```

Execution while disk Sync in Progress

When a recovery is attempted, tool automatically adds the recovered disk back into RAID configuration. But, it may take few minutes before the data on the working master disk to sync completely to the newly configured disk. “show system internal file /proc/mdstat” will display the latest sync status for all the recovered disk partitions. Execution of repair tool will print the following error messages in this case:

```
ERROR: Disc sync from a previous run is still underway on module <slot#>.
ERROR: Please wait for all the discs to complete the sync.
ERROR: For current sync status try: 'show system internal file /proc/mdstat'
```

GOLD Failure Notifications

On all 6.1(x) releases, diagnostic software running on NXOS will flag eUSB related errors under its CompactFlash test umbrella. Syslog notifications are sent when the CompactFlash test detects flash errors. GOLD’s diagnostic CLIs also can be used to confirm current test status:

```
2014 Nov 14 02:25:00 N7K %DEVICE_TEST-2-COMPACT_FLASH_FAIL: Module 5 has failed test
CompactFlash 20 times on device Compact Flash due to error The compact flash power test
failed
2014 Nov 14 02:25:00 N7K %MODULE-4-MOD_WARNING: Module 5 (Serial number: JAF1645ANLQ)
reported warning due to The compact flash power test failed in device DEV_UNDEF (device
error 0x0)
```

Ideally, one could wait for these syslog messages and run the recovery tool in response to recover from it. If the tool is unable to recover the flash, please contact Cisco Support for additional help.

```
show diagnostic result module 5 detail
```

```
Module 5: Supervisor module-2 (Standby)
```

```
7) CompactFlash E
```

```
Error code -----> DIAG TEST ERR DISABLE
Total run count -----> 27510
Last test execution time ----> Thu Jul 24 01:44:53 2014
First test failure time -----> Wed Jul 23 16:15:04 2014
Last test failure time -----> Thu Jul 24 01:44:58 2014
Last test pass time -----> Wed Jul 23 15:45:04 2014
Total failure count -----> 20
Consecutive failure count ----> 20
Last failure reason -----> The compact flash power test
                             failed
Next Execution time -----> Thu Jul 24 02:14:53 2014
```

Please make sure to restart the CompactFlash test once the recovery tool completes successfully. To restart an error disabled test, disable monitoring, clear result and enable monitoring with the following CLI.

```
no diag monitor module <sup slot#> test 7
diag clear result module <sup slot#> test 7
diag monitor module <sup slot#> test 7
```

However, starting 6.2(1) release, GOLD will not detect single disk failures. Please use “show system internal raid” command for periodically monitoring the health of these flashes.

RAID Status Check after a Recovery

Check the latest status of the RAID array the following command:

```
switch# show system internal raid | grep -A 1 "Current RAID status info"
Current RAID status info:
RAID data from CMOS = 0xa5 0xf0
switch#
...
```

Last byte of the CMOS data printed as 0xf0 indicate that the repairs were successfully completed, RAID array reconfigured and disk sync initiated. However, run “show system internal file /proc/mdstat” to make sure it displays the [2/2] [UU] status for all disk partitions.

```
switch# show system internal file /proc/mdstat
Personalities : [raid1]
md6 : active raid1 sdd6[1] sdc6[0]
      77888 blocks [2/2] [UU]

md5 : active raid1 sdd5[1] sdc5[0]
      78400 blocks [2/2] [UU]

md4 : active raid1 sdd4[1] sdc4[0]
      39424 blocks [2/2] [UU]

md3 : active raid1 sdd3[1] sdc3[0]
      1802240 blocks [2/2] [UU]

unused devices: <none>
switch#
```

Revision History

| Version | Date | Comments |
|---------|----------------|---|
| 10.0(1) | April 04, 2015 | Initial Release |
| 10.0(2) | April 24, 2015 | Several enhancements from first release. 1) Dual flash recovery issue after the reload on NXOS Release 6.1(1) is fixed. 2) Make sure to always reset the flash part and check its health before configuring it back into the RAID array. 3) Collect run logs & extra debug data from the last run and leave a copy under volatile: for further analysis. |

