Tru64 UNIX

Recovering Logical Storage Manager When the Cluster Root Domain Is Under LSM Control

October 2006

Product Version: Tru64 UNIX Version 5.1A, and TruCluster Server Version 5.1A or later

This Best Practice provides steps for recovering from a disaster that prevents the Cluster Logical Storage Manager (LSM) from starting.

1 Introduction

This Best Practice was created for the rare possibility that a disaster occurs that prevents the Cluster Logical Storage Manager (LSM) from starting. In the instance where the cluster wide root (/), the AdvFS domain cluster_root, is under LSM control and there is some sort of issue with the disk(s) underlying those LSM volumes or with the LSM metadata, this will prevent the system from (re)booting. This Best Practice provides you with two procedures for recovering if this issue should occur. The documented procedures in sections 3 and 4 are based on the state of the cluster when you realize that there is an unrecoverable situation in enabling LSM (vold).

1.1 Is This Best Practice Right for You?

Not all Best Practices apply to all configurations, so you must be sure that this Best Practice is appropriate for your system and circumstances. This Best Practice assumes that all normal steps have been taken to enable vold and that you have determined that the only step currently possible is a complete recovery of the LSM configuration. To use this Best Practice, you must meet the requirements described in the following table:

Requirement	Description
Operating System	Tru64 UNIX™ Version 5.1A, and TruCluster Server Version 5.1A or later
Logical Storage Manager	LSM must be installed, initialized, and you have incorporated the cluster_root as a LSM volume (cluster_rootvol)
Impact on Availability	The Best Practice's purpose is to provide a quick, efficient recovery so that your downtime is minimal
Additional Requirements	You must have superuser access and have some knowledge of LSM commands.

1.2 Requirements for Recovery Procedures

In order to use these procedures a certain minimum configuration is expected. Those are:

The procedure:	Minimum Requirements:
If the recovery is needed and the cluster is down and can not be booted	1) The standalone boot disk (used before forming the cluster) will need to be available and accessible as it will be necessary to boot standalone. 2) At least one disk on your cluster that contains a valid rootdg configuration database is needed.
If the recovery is needed and seen when the cluster is on-line	1) Free disks with enough available space in order to recover cluster root, usr, var, and all other rootdg volumes will be necessary while the cluster is still up and running 2) The ADVFS-UTILITIES license pak must be loaded and enabled

If you do not meet the previous requirements, see the LSM documentation for other recovery procedures to find one that meets your circumstance.

2 Error Messages

2.1 System Hanging on Boot at "Starting LSM"

Seeing this error does not necessarily mean that there is a problem in LSM. You need to check the other messages in the console preceding this to see if any errors are mentioned. This should pinpoint you into the correct direction needed in order to resolve the issue.

LSM will hang if it can not access the drives on boot up. Therefore the first thing to check with this problem is to make sure all hardware can be seen at the console and that it is seen as it was on prior boots in the same bus and target locations.

If you have covered all possibilities and there is still a hang on boot up, then the recovery procedure *Recovering When You Can Not Boot* should be considered.

2.2 Voldisk command will not complete on booted system

This issue can be due to the same issue seen previously. If you attempt to reboot one of the cluster members, you would see the system hang. Voldisk will also hang if it can not access all disks on the system. Therefore you should run the *hwmgr –view devices cluster* command to ensure that all disks are accessible. If that command hangs, check your hardware consoles to find the reason for not being able to access all of the hard drives.

2.3 ERROR: Cannot start cluster_rootvol volume, no valid plexes

If you see this message on booting, you will need to follow the recovery procedure listed for *Recovering When You Can Not Boot.* First you will want to fix any hardware issues associated with cluster_rootvol.

2.4 ERROR: IPC failure: Configuration daemon is not accessible

If you see this error when executing a *voldisk* or *volprint* command, you are getting this error because vold is not enabled. If LSM simply was not started, start vold using the *vold –k* command. If you are receiving this error and you suspect vold exited abnormally, attempt to restart vold: *vold -k*. If vold will not restart and cluster_root is under LSM control, do not reboot but rather execute the recovery procedure: *Recovery When the Cluster Is On-Line*

3 Recovering When You Can Not Boot

Following are the instructions for recovering your cluster when your cluster is down and you can not boot the cluster because of an issue with LSM. Before you follow these instructions, you will want to fix any hardware issues so that a recovery is possible.

1) Boot as a standalone using the original boot disk used before forming the cluster.

2) When the system is booted, login as root and execute the following command to set the hostname to the cluster's hostname. This is necessary so that the hostname will match the hostid on the LSM's disk's metadata. We need to initialize the /etc/vol/volboot file and need for it to contain the same hostid (cluster name).

   # /sbin/hostname <cluster name>

   Note

   If you do not know the exact cluster name (e.g. foo.hp.com or foo), you can get the hostname from the LSM's disk's metadata by issuing the following command:
   # /usr/lib/lsm/bin/volprivutil list <raw partition containing LSMpriv>

3) Now start LSM on the booted standalone system by doing the following:

   # volinstall /* To create the special device files needed for LSM */
   # voliod set 2 /* This starts 2 voliods */
   # vold –k –m disable /* Start vold in disabled mode as no volboot file exists */
   # voldctl init /* Create the /etc/vol/volboot file setting the hostid as the hostname */
   # voldctl enable /* The command will scan all disks and enable vold */

4) Now that LSM is up and running, execute:

   # voldisk list /* To ensure all disks are enabled. If needed, execute the appropriate LSM commands to recover the drives. */
   # volprint /* To ensure that all volumes are enabled/active. You will most likely need to start the volumes. To do this, execute the following:
   # volume -g rootdg startall */

5) Also if not all disks in a non-rootdg disk group are available, you will need to manually import the disk group using the –f (force) switch. LSM is designed to only import a disk group if all disks are accessible. This is to prevent data integrity issues. However if you know that a disk is unavailable you can manually import the disk group using:

   # *voldg –f import <disk group name>*
   Example:
   # *voldg –f import datadg*

Note

If this still fails because there are no valid configuration copies for the disk group, then *volrestore* can be used to recover. This assumes *volsave* was executed at one time containing this disk group's latest configuration. If this is not the case, a manual reconfiguration of the disk group will need to be done.

6) Once the volumes have been synched and everything is enabled/active, you can then reboot the cluster.

7) Once the cluster is rebooted, run *voldisk list* and *volprint* commands on each cluster member to confirm the configuration is in sync and is as expected.

4 Recovery When the Cluster Is On-Line

Following are the instructions for recovering your cluster when at least one of your cluster members is running and you can still access your LSM volumes successfully but can not get LSM (vold) to start. Before you follow these instructions, you will want to fix any hardware issues so that a recovery is possible and all ensure that all other avenues have been investigated to be sure that a recovery is necessary.

1) From one cluster member, do an *addvol* to cluster_root, cluster_usr and cluster_var, adding new unused non-LSM storage to those domains. Example:

   # addvol /dev/disk/dsk20a cluster_root

2) After adding new non-LSM storage to cluster_root, cluster_usr and cluster_var, from one cluster member, do a *rmvol* of the LSM volumes. Example:

   # rmvol /dev/vol/rootdg/cluster_rootvol cluster_root

   Note

   If you have any other domains in rootdg, you will also want to do an *addvol* and a *rmvol* on those as well. If a volume in rootdg is used as a raw volume for a database or other purpose, you will want to ensure you have a good backup of the data. If any volumes are created exactly as they were previously, no data is lost. However all precautionary measures should be taken.

If you DO NOT have primary swap configured under LSM, go to Step 5.

3) Now you need to remove primary swap from LSM control. On each cluster member, edit the /etc/sysconfigtab file and update the "swapdevice" entry with the full path to the swap partitions.

   Note

   The partitions that you will use need to have the fstype as swap in the disklabel. One will be needed for each cluster member. An example of the /etc/sysconfigtab entry:

   vm:

   swapdevice: =/dev/disk/dsk20b

4) Because you are changing primary swap storage, you will now need to reboot your cluster. If there are any entries pointed directly to an LSM volume in the /etc/fstab file, you will need to remove these lines so that LSM does not attempt to use them on the reboot. Also before rebooting, ensure that you have recent backups of all data in case irreconcilable problems occur.

5) To ensure a clean start of LSM in the next step we want to reset the kernel configuration databases in the kernel. To do this the following needs to be done:

    a. Unmount all non-system file systems that use LSM volumes.

    b. Execute: # *vold –k –r reset –d* /* From one cluster member only */

    c. Execute on all possible cluster members: # *voldctl stop*

6) Now using a free partition of the free disk(s) used for the recovery, do a *volsetup* command using the following syntax: #*volsetup -o force <disk partition>* Do this command on only one cluster member to initialize LSM, creating a new rootdg disk group.

7) Run the *volsetup –s* command on the remaining cluster members. This will synchronize the remaining members with the new LSM rootdg.

8) If there were volume(s) in the rootdg disk group which were not part of the system volumes, you can recover them by issuing a *volrestore* command. The syntax for that command is as follows:

    # *volrestore -g rootdg -v <vol_name>*

    Note

    This assumes *volsave* was executed at one time containing the volume(s). If this is not the case, a manual reconfiguration of the volume will need to be done.

9) The non-rootdg disk groups should import automatically during volsetup. If all disks in a non-rootdg disk group are not available, you will need to manually import the disk group using the –f (force) switch. LSM is designed to only import a disk group if all disks are accessible. This is to prevent data integrity issues. However if you know that a disk is unavailable you can manually import the disk group using: *# voldg –f import <disk group name>* Example:

   # voldg –f import datadg

   Note

   If this still fails because there are no valid configuration copies for the disk group, then *volrestore* can be used to recover. This assumes *volsave* was executed at one time containing this disk group's latest configuration. If this is not the case, a manual reconfiguration of the disk group will need to be done.

10) Execute a *voldisk list* and *volprint* command on each cluster member to confirm the configuration is in sync and is as expected. You may need to do the following: *# volume -g <disk group> startall* command for each non rootdg disk group to start the volumes and get them to an enabled/active state.

11) At this point, you can mount all of your file systems that contain LSM volumes as all volumes should be in an enabled/active state.

12) From one cluster member, you can now do a *volmigrate* of cluster root, usr and var and mirror the resulting volumes.

13) Now encapsulate and mirror swap. Do this one at a time from each cluster member. This will allow the cluster to stay on-line and will also allow you to fix any problem that may arise. We have also seen issues that have arisen in attempting this task on more than one member at a time while testing.

14) Run the *voldisk list* and *volprint* commands again on each cluster member to confirm the configuration is in sync and is as expected.

Legal Notice