Tru64 UNIX Best Practice

Ensuring Redundancy of LSM Configuration Databases on a Fibre Channel

October 2001

Product Version:	**Tru64 UNIX Version 5.0 or higher or TruCluster Server Version 5.0 or higher, and HSG RAID Array Controller (ACS Version 8.5 or higher) or HSZ RAID Array Controller (VCS Version 1.01 or higher)**

This Best Practice describes how to manually distribute LSM configuration database copies on a Fibre Channel to ensure availability of the LSM configuration on systems running the Tru64 UNIX operating system.

Compaq Computer Corporation
Houston, Texas

Contents

Ensuring Redundancy of LSM Configuration Databases on a Fibre Channel

Ensuring Redundancy of LSM Configuration Databases on a Fibre Channel

This Best Practice describes how to manually distribute LSM configuration database copies on a Fibre Channel to ensure availability of the LSM configuration. This Best Practice applies to configurations with multiple redundant RAID arrays connected to the host or cluster over a Fibre Channel fabric.

In Tru64 UNIX Version 5.0 or later releases, LSM automatically manages the number and location of active copies of the configuration database. LSM maintains a minimum of four active copies per disk group, distributed over several adapters (if present).

However, in environments using Fibre Channel, LSM perceives the entire fabric as one adapter. This can result in situations where LSM places all the active copies of the configuration database on disks in the same RAID array. Even if LSM volume data is configured to be redundant across two RAID arrays, active copies of the configuration database are not handled the same as mirrored data and might not be duplicated across arrays.

There are two approaches to prevent this situation:

- If you have an existing LSM configuration, you can manually modify the LSM disks to force LSM to redistribute active copies of the configuration databases onto specific disks in different arrays to ensure redundancy and failover for LSM operation.
- If you have not yet created an LSM configuration, you can disable LSM's automatic configuration database management function and then individually configure specific disks to have copies as you initialize them for LSM use.

This Best Practice describes both approaches.

See the Tru64 UNIX Best Practices Web page for more information about Best Practices documentation:
http://www.tru64unix.compaq.com/docs/best_practices/index.html

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. To use this Best Practice, you must meet the requirements described in the following table:

Requirement	Description
Operating System	Tru64 UNIX Version 5.0 or higher, or TruCluster Server Version 5.0 or higher
System Configuration	Fibre Channel fabric and two or more HSG Array Controllers (with ACS Version 8.5 or higher) or HSV Controllers (with VCS Version 1.01 or higher). Configuration can include a mix of SCSI adapters and Fibre Channel. LSM software subsets are installed, but LSM might or might not be running yet. This Best Practice applies to both situations.
Impact on Availability	None. No shutdown required.

If you do not meet the previous requirements, see *Alternative Practices* for information.

Before You Begin

Before you apply this Best Practice, you should understand its implications on how you must manage the system in the future.

By default, the private region of an LSM simple or sliced disk is configured to be a candidate for storing an active copy of the configuration database for the disk group to which that disk belongs. LSM does not place an active copy on the disk unless it needs to ensure redundancy across buses. Because all simple and sliced disks are configured as candidates by default, LSM can enable an active copy on another simple or sliced disk if a disk with an active copy fails.

The steps in this Best Practice deconfigure all but a few disks within a disk group to limit the number of candidate disks. To maintain the minimum of four copies per disk group, LSM has to enable active copies on the remaining candidate disks, which are those you select based on their distribution among RAID arrays.

After implementing this Best Practice, LSM will no longer automatically maintain the number and location of active copies of the configuration database. You must keep track of changes to your configuration (specifically, disks or RAID arrays that are added, moved, or fail) and continue to manually configure disks as candidates for active copies of the configuration database as appropriate to maintain redundancy.

If you add or remove disks with active copies from the `rootdg` disk group, you must modify the `/etc/vol/volboot` file accordingly. This file is used when the system starts and tells LSM which disks contain copies of the configuration database for `rootdg`.

Also, to create new disk groups, you must manually configure specific disks to have active copies of the configuration database as you initialize them for LSM.

If the configuration includes SCSI adapters, this Best Practice affects the disks on each SCSI adapter as well as each RAID array.

Note

The examples in this Best Practice reflect the recommended practice of setting identifiers for each unit you create on the RAID array.

Further, for environments with multiple RAID arrays, we recommend using offsets for the identifiers in each array. For example, in the first array, set the identifiers to 01nn, and in the second array, set the identifiers to 02nn, and continue to increment the identifiers by 100 in each successive array. This simplifies the correlation of disks to their redundant controller pairs.

If you have not implemented this recommended practice, see *Alternative Practices* for another method of determining the disk-to-array correlation.

Applying the Best Practice

For an existing LSM configuration, determine which LSM disks are connected to the RAID arrays and to which disk groups they belong. Then determine which of these disks have active copies of the configuration

database, and either configure or deconfigure the disks to have copies appropriate for your availability requirements.

For a new LSM configuration, you must disable the automatic configuration database management function of LSM, which applies to the entire configuration (SCSI-attached disks and Fibre-attached disks). You must individually configure specific disks to have copies and ensure that each disk group has a minimum of four copies with at least one copy on each SCSI adapter or RAID array supporting the disks.

Modifying an Existing LSM Configuration

To modify an existing LSM configuration:

1. Determine the identity and connectivity of the accessible disks.
2. Identify the disk group membership for each LSM disk.
3. Configure and deconfigure specific disks to have copies.

You must be the root user to perform these steps.

Determine the Identity and Connectivity of the Accessible Disks

To determine the names of accessible disks and their associated RAID arrays:

1. Display a list of disks and their identifiers:

   ```
   # hwmgr -view devices | grep -i identifier
   ```

   Information similar to the following is displayed:

   ```
   338: /dev/disk/dsk275c    DEC      HSG80           IDENTIFIER=0101
   339: /dev/disk/dsk276c    DEC      HSG80           IDENTIFIER=0102
   340: /dev/disk/dsk277c    DEC      HSG80           IDENTIFIER=0103
   341: /dev/disk/dsk278c    DEC      HSG80           IDENTIFIER=0104
   345: /dev/disk/dsk282c    DEC      HSG80           IDENTIFIER=0109
   346: /dev/disk/dsk283c    DEC      HSG80           IDENTIFIER=0110
   347: /dev/disk/dsk284c    DEC      HSG80           IDENTIFIER=0111
   348: /dev/disk/dsk285c    DEC      HSG80           IDENTIFIER=0201
   349: /dev/disk/dsk286c    DEC      HSG80           IDENTIFIER=0202
   350: /dev/disk/dsk287c    DEC      HSG80           IDENTIFIER=0203
   351: /dev/disk/dsk288c    DEC      HSG80           IDENTIFIER=0204
   352: /dev/disk/dsk289c    DEC      HSG80           IDENTIFIER=0205
   353: /dev/disk/dsk290c    DEC      HSG80           IDENTIFIER=0206
   354: /dev/disk/dsk291c    DEC      HSG80           IDENTIFIER=0108
   355: /dev/disk/dsk292c    DEC      HSG80           IDENTIFIER=0105
   356: /dev/disk/dsk293c    DEC      HSG80           IDENTIFIER=0106
   357: /dev/disk/dsk294c    DEC      HSG80           IDENTIFIER=0107
   ```

   Note that the output might not list the disk identifiers in numerical order.

2. Note which disk is associated with which array.

   It might be helpful to make a table with headings similar to the following:

Identifiers in 0100s	Disk Group	Identifiers in 0200s	Disk Group
dsk275c (0101)		dsk285c (0201)	
dsk276c (0102)		dsk286c (0202)	
dsk277c (0103)		dsk287c (0203)	
dsk278c (0104)		dsk288c (0204)	
dsk292c (0105)		dsk289c (0205)	
dsk293c (0106)		dsk290c (0206)	
dsk294c (0107)			
dsk291c (0108)			
dsk282c (0109)			
dsk283c (0110)			
dsk284c (0111)			

Determine the Disk Group Membership for Each LSM Disk (Existing Configuration)

To display the names of LSM disks and their disk groups on an existing LSM configuration:

```
# voldisk list | grep online
```

Information similar to the following is displayed:

```
dsk1         sliced    dsk1         rootdg       online
dsk275       sliced    dsk275       dg01         online
dsk276       sliced    dsk276       dg01         online
dsk277       sliced    dsk277       dg01         online
dsk278       sliced    dsk278       dg01         online
dsk282       sliced    dsk282       dg02         online
dsk283       sliced    dsk283       dg02         online
dsk284       sliced    dsk284       dg02         online
dsk285       sliced    dsk285       dg01         online
dsk286       sliced    dsk286       dg01         online
dsk287       sliced    dsk287       dg01         online
dsk288       sliced    dsk288       dg02         online
dsk289       sliced    dsk289       dg02         online
dsk290       sliced    dsk290       dg02         online
dsk291       sliced    dsk291       dg01         online
dsk292       sliced    dsk292       dg01         online
```

```
dsk293       sliced    dsk293       dg01         online
dsk294       sliced    dsk294       dg01         online
```

To display only the disks that belong to one disk group:

voldisk list | grep *diskgroup*

If you have several disk groups, add the disk group information to the connectivity table, as shown in the following example.

Identifiers in 0100s	Disk Group	Identifiers in 0200s	Disk Group
dsk275 (0101)	dg01	dsk285 (0201)	dg01
dsk276 (0102)	dg01	dsk286 (0202)	dg01
dsk277 (0103)	dg01	dsk287 (0203)	dg01
dsk278 (0104)	dg01	dsk288 (0204)	dg02
dsk292 (0105)	dg01	dsk289 (0205)	dg02
dsk293 (0106)	dg01	dsk290 (0206)	dg02
dsk294 (0107)	dg01		
dsk291 (0108)	dg01		
dsk282 (0109)	dg02		
dsk283 (0110)	dg02		
dsk284 (0111)	dg02		

Configure and Deconfigure Specific Disks

To configure and deconfigure specific disks, you must determine which disks in a disk group have active copies of the configuration database. If these disks are in the same RAID array, you must force LSM to place a copy on at least one disk in another array.

LSM attempts to maintain at least four active copies at all times. But you ordinarily cannot tell which specific disks LSM will place copies on, so you must limit the possibilities to only the disks you want; that is, at least one in each array, with a minimum of four.

1. Display which disks in a disk group have active copies of the configuration database and which disks are configured to potentially contain a copy:

   # **voldg list dg01 | grep "config disk"**

   Information similar to the following is displayed:

```
config disk dsk275 copy 1 len=2993 state=clean online
config disk dsk276 copy 1 len=2993 state=clean online
config disk dsk277 copy 1 len=2993 state=clean online
config disk dsk278 copy 1 len=2993 state=clean online
config disk dsk285 copy 1 len=2993 disabled
config disk dsk286 copy 1 len=2993 disabled
config disk dsk287 copy 1 len=2993 disabled
config disk dsk291 copy 1 len=2993 disabled
config disk dsk292 copy 1 len=2993 disabled
config disk dsk293 copy 1 len=2993 disabled
config disk dsk294 copy 1 len=2993 disabled
```

In the preceding output, the disks with an attribute of `state=clean online` are the disks with the active copies for this disk group. If dsk275, dsk276, dsk277 or dsk278 fails, LSM can place an active copy on any disk with an attribute of `disabled`.

To limit LSM to certain disks, you must deconfigure the private regions of all but the four disks you select based on their distribution among multiple RAID arrays.

If all disks in a disk group are connected to the same array, you can, if you have available storage, add one or more disks from a different array to that disk group.

Note

You cannot enable a nopriv disk to have a copy because nopriv disks do not have a private region to contain the configuration database. Only sliced or simple disks can contain copies of the configuration database for a disk group.

2. For each disk group:

   a. Choose four disks that are distributed among the arrays serving disks for that group.

      For example, in diskgroup dg01, dsk275, 276, 277 and 278 have active copies of the configuration database but all four are in the array with identifiers in the 0100 series. Choose two of those disks to deconfigure (for example, dsk277 and dsk278).

      Choose two disks in a different array to configure, if they are not already properly configured. For example, disks dsk285 and dsk286 are in another array and are currently configured to have a copy. If you deconfigure all but those two disks and the two you chose in the first array (dsk275 and dsk276), LSM will put active copies on dsk285 and dsk286.

When you are finished, only disks dsk275, dsk276, dsk285 and dsk286 will have active copies.

b. Deconfigure the configuration database copies on all other simple or sliced disks in the disk group:

```
# voldisk moddb dsknnn ... nconfig=0
```

For example:

```
# voldisk moddb dsk277 nconfig=0
# voldisk moddb dsk278 nconfig=0
# voldisk moddb dsk287 nconfig=0
# voldisk moddb dsk291 nconfig=0
# voldisk moddb dsk292 nconfig=0
# voldisk moddb dsk293 nconfig=0
# voldisk moddb dsk294 nconfig=0
```

This forces LSM to place active copies on only the remaining four disks, thereby distributing the copies among the arrays.

c. Confirm the changes by redisplaying the disk group information:

```
# voldg list dg01 | grep "config disk"
```

Only the disks you chose in step 2a should appear in the output:

```
config disk dsk275 copy 1 len=2993 state=clean online
config disk dsk276 copy 1 len=2993 state=clean online
config disk dsk285 copy 1 len=2993 state=clean online
config disk dsk286 copy 1 len=2993 state=clean online
```

d. Repeat the preceding steps for each LSM disk group.

If a disk with a copy fails, you must configure another disk in the same array to have a copy. This gives you control over which disks LSM can use to keep its active configuration database copies. It forces LSM to place an active copy on only that disk and maintain the distribution. LSM will no longer automatically distribute copies as disks fail.

To configure a disk to have a copy:

```
# voldisk moddb dsknnn nconfig=1
```

Initializing a New LSM Configuration

To prepare for a new LSM configuration:

1. Determine the identity and connectivity of the accessible disks.
2. Turn off the automatic configuration and load-balancing features of LSM.

3. Initialize disks specifically to have or not have a configuration database copy.

You must be the root user to perform these steps.

Disabling the Automatic Load-Balancing Feature of LSM (New Configuration)

To prevent LSM from automatically managing the configuration database copies:

1. Choose the disks you will use to initialize LSM, but do not initialize them.
2. Edit the `/sbin/lsm-startup` script.

   Change the `vold_opts=-k` line to read:

   ```
   vold_opts=-k -x noloadbalance -x noautoconfig
   ```
3. If you plan to use LSM for your root and swap devices on a standalone system, edit the `/sbin/lsmbstartup` script.

   Change the `vold -k -m boot` line to read:

   ```
   vold -k -m boot -x noloadbalance -x noautoconfig
   ```
4. Use the disks you identified in step 1 to initialize LSM, which creates the `rootdg` disk group:

   # **volsetup dsk*n* dsk*n* ...**
5. Add the disk names to the `/etc/vol/volboot` file:

   # **voldctl add disk dsk*n* ...**

Initializing New LSM Disks and Creating Disk Groups

To initialize LSM disks that will or will not have copies of the configuration database, do one of the following:

- To initialize a disk that can have a configuration copy (the default):

  # **voldisksetup -i dsk*n***
- To initialize a disk that will not have a configuration copy:

  # **voldisksetup -i dsk*n* ... nconfig=0**

To add disks to an existing disk group:

voldg -g *diskgroup* adddisk dsk*n* ...

To create a new disk group:

```
# voldg init diskgroup dskn ...
```

Alternatively, see *Creating a Disk Group with a Nonstandard Number of Configuration Database Copies*.

The following example shows the commands to initialize four disks with configuration copies and three disks with no copies, create a new disk group using the seven disks, and confirm that only the specified disks have active copies of the configuration database:

```
# voldisksetup -i dsk5 dsk6 dsk7 dsk9
# voldisksetup -i dsk10 dsk11 dsk12 noconfig
# voldg init kadg dsk5 dsk6 dsk7 dsk9 dsk10 dsk11 dsk12
# voldg list kadg
Group:     kadg
dgid:      998937447.1992.hostname.com
import-id: 0.1991
flags:
copies:    nconfig=default nlog=default
config:    seqno=0.1033 permlen=2993 free=2988 templen=3 loglen=453
config disk dsk5 copy 1 len=2993 state=clean online
config disk dsk6 copy 1 len=2993 state=clean online
config disk dsk7 copy 1 len=2993 state=clean online
config disk dsk9 copy 1 len=2993 state=clean online
log disk dsk5 copy 1 len=453
log disk dsk6 copy 1 len=453
log disk dsk7 copy 1 len=453
log disk dsk9 copy 1 len=453
```

If a disk with a copy is removed from the disk group, the number of configuration database copies is reduced by one, and no other disk receives a copy to replace it.

However, you can initialize a disk to have up to two copies of the configuration database:

```
# voldisksetup -i dsk5 nconfig=2
```

When a disk with a copy fails, LSM can put a replacement copy on any disk that has been configured to contain two copies.

Creating a Disk Group with a Nonstandard Number of Configuration Database Copies

As mentioned previously, LSM attempts to maintain a default number of configuration database copies per disk group, with a minimum of four. In a Fibre Channel environment, LSM maintains only the minimum number because it sees the whole fabric as one adapter.

However, you can specify a different number of copies when you create a disk group, but only then. Once the disk group exists, you cannot change the number of copies it will maintain.

For example, if you plan to create a disk group distributed over five RAID arrays and want to ensure LSM maintains two copies of the configuration database in each array, configure two disks in each array to have a copy, then create the disk group with those disks and set the number of copies for the disk group to 10. (In the following example, disks dsk45 and dsk46 are in one array, dsk55 and dsk56 are in another, and so on:)

```
# voldisksetup -i dsk45 dsk46 dsk55 dsk56 dsk65 dsk66 \
dsk75 dsk76 dsk85 dsk86 nconfig=1
# voldg init diskgroup dsk45 dsk46 dsk55 dsk56 \
dsk65 dsk66 dsk75 dsk76 dsk85 dsk86 nconfig=10
```

The second command line forces LSM to put an active copy on each of those 10 disks. You can then add more disks to the disk group. If some of those additional disks are configured to potentially have a copy and one of the original disks with a copy fails, LSM will place a copy on one of those disks to maintain 10 copies for the group.

LSM might not place a copy on another disk in the same array as the disk that failed. To ensure the correct distribution of active copies of the configuration database, specify `nconfig=0` when you initialize more disks to add to that disk group. Then manually reconfigure one disk at a time as a replacement candidate, in the same array, when a disk with a copy fails.

Verifying Success

After you apply this Best Practice, you can verify whether it was successful in any of the following ways:

- Display the properties of an LSM disk to see whether or not it has an active copy of the configuration database or the potential to have one:

  # **voldisk list dsk*n* | grep configs:**

  Information similar to the following is displayed:

  ```
  configs:   count=1 len=2993
  ```

  Disks with an active copy or the potential to have one show an attribute of `count=1`. Disks without the potential to have a copy show `count=0`.

- Display the properties of a disk group:

  # **voldg list *diskgroup***

Information similar to the following is displayed:

```
Group:     kadg
dgid:      998340823.1935.hostname.com
import-id: 0.1934
flags:
copies:    nconfig=default nlog=default
config:    seqno=0.1030 permlen=2993 free=2989 templen=2 loglen=453
config disk dsk11 copy 1 len=2993 state=clean online
config disk dsk12 copy 1 len=2993 state=clean online
log disk dsk11 copy 1 len=453
log disk dsk12 copy 1 len=453
```

Only disks that contain or have the potential to contain a configuration database copy appear in the output of the `voldg list` command:

– Disks with an active copy show an attribute of `state=clean online`.

– Disks with the potential to have an active copy show an attribute of `disabled`.

Confirm that the output lists only the disks that you configured to have active copies or the potential to have one.

- Turn off one of a pair of redundant RAID arrays and confirm that your LSM configuration is still available.

Note

This could cause volumes to undergo resynchronization when you turn on the RAID array. Do this at your discretion.

If the Best Practice was not successful, see *Troubleshooting* for information about identifying and solving problems.

Troubleshooting

If you determine that the Best Practice was not successful, as described in *Verifying Success*, use the following table to identify and solve problems:

Problem	Possible Solutions

You receive the following error when trying to modify the number of configuration copies on a disk:
`lsm:voldisk: ERROR: voldisk: Device dsk`*nn*`: \`
`Modifying database failed:  \`
`Bad record name`
	The `voldisk moddb` command requires the disk access name, assigned by the operating system — not the disk media name.
	Check the disk access name with the `voldisk list` command. Disk access names appear in the `DEVICE` column in the output.

Problem	Possible Solutions
Disk does not have an active copy or the potential but should	For all of these problems, enter the following command line for each incorrect disk:
Disk has an active copy or the potential but should not	`# voldisk moddb nconfig=`*n* `dskn`
Not enough disks in a disk group have active copies or potential	
Too many disks in a disk group have active copies or potential	
Disks with copies or potential are not properly distributed among RAID arrays or SCSI buses	

Alternative Practices

Although this Best Practice is the recommended method for ensuring redundancy of LSM configuration databases in a Fibre Channel environment, you can use an alternative method if your system does not meet the requirements described in *Is This Best Practice Right for You?*.

- If you are running a version of the operating system earlier than Version 5.0, LSM does not automatically manage the configuration database copies. If you are considering upgrading to Version 5.0 or higher and your environment will not contain Fibre Channel adapters, LSM will manage the databases correctly.
- If you are running Tru64 UNIX Version 5.0 or higher (or planning to upgrade) and your environment contains a mix of SCSI buses and a Fibre Channel, you could use only those disks connected to the SCSI buses for LSM use. LSM correctly sees each SCSI bus as a separate adapter and manages the configuration database copies appropriately. (However, with multipath SCSI, LSM can run into the same problem as with Fibre Channel.)

If you did not set identifiers for the units on your RAID arrays, you can either assign identifiers and then reboot the system or perform the following steps to determine the connectivity of disks to arrays.

Note

Once the disk's worldwide ID (WWID) is constructed, it does not change even if you move the physical disk from one array to

another. Therefore, use the following procedure only if you have not moved disks or replaced HSG or HSV controllers.

To determine the connectivity of devices without an identifier:

1. To find the worldwide ID for the controllers or controller pairs on the system:

   # **hwmgr -show scsi -full | grep WWID**

   The WWID of each disk connected to those controllers is constructed from the WWID of the controller.

   In this example, the WWIDs for the HSG controller pair are:

   WWID:02000008:5000-1fe1-0005-a720
   WWID:02000008:5000-1fe1-0005-a730

2. To find all the disks whose WWID contains the same hex digit groups as the controller or controller pair:

   # **hwmgr -show scsi -full**

   For example, the world-wide ID for dsk277 is:

   WWID:01000010:**6000-1fe1-0005-a720**-0009-9500-5463-01a0

   The world-wide ID for dsk182 is:

   WWID:01000010:**6000-1fe1-0005-a730**-0009-9500-5451-0178

   It might be helpful to make a table with the controller world-wide ID numbers as column headings, and list each disk connected to each node name in the table rows.

   For example, your table could look like the following:

6000–1FE1–0005–A720	Disk Group	6000–1FE1–0005–A730	Disk Group
dsk7		dsk134	
dsk8		dsk135	
dsk9		⋮	
⋮		dsk150	
dsk49		dsk151	
dsk50		⋮	
⋮		dsk199	

6000–1FE1–0005–A720	Disk Group	6000–1FE1–0005–A730	Disk Group
dsk100		dsk201	
dsk101		⋮	
⋮		dsk259	
dsk133		dsk260	

3. Proceed with the steps in *Determine the Disk Group Membership for Each LSM Disk (Existing Configuration)* and *Configure and Deconfigure Specific Disks*.

Note

The table in *Determine the Disk Group Membership for Each LSM Disk (Existing Configuration)* lists output where an identifier was set for each unit in the RAID array. When you follow the steps in that section, use the table you began here to note the disk groups of each disk.

Comments and Questions

We value your comments and questions on the information in this document. Please mail your comments to us at this address:

best_practices@zk3.dec.com

Legal Notice

Compaq shall not be liable for technical or editorial errors or omissions contained herein. The information in this document is provided “as is” without warranty of any kind and is subject to change without notice. The warranties for Compaq products are set forth in the express limited warranty statements accompanying such products. Nothing herein should be construed as constituting an additional warranty.