Using Sun StorEdge™ A3x00/A1000 With Sun Enterprise Volume Manager™

Please
Recycle

Contents

Using Sun StorEdge™ A3x00/A1000 With Sun Enterprise Volume Manager™

The Sun StorEdge™ A3x00/A1000 subsystem in conjunction with RAID Manager 6 (RM6) is compatible with Sun Enterprise Volume Manager™ 2.4, 2.5, 2.6, and 3.0.1. To ensure compatibility between Sun StorEdge A3x00/A1000 and Sun Enterprise Volume Manager, you must do the following.

- The StorEdge A3x00/A1000 and Sun Enterprise Volume Manager installation sequence must be followed exactly as documented in the section "Installation of the Sun Enterprise Volume Manager and Sun StorEdge A3x00/A1000.
- Sun Enterprise Volume Manager volumes configured using devices from the HW RAID subsystems cannot be part of the root disk group, rootdg. Configured HW RAID subsystems LUNs can belong only to non-rootdg disk groups.
- When Sun Enterprise Volume Manager encapsulates a device that has a mounted file system and entry in the /etc/vfstab file, Sun Enterprise Volume Manager updates the /etc/vfstab entry for that device to contain the Sun Enterprise Volume Manager device node name. Sun Enterprise Volume Manager is not compatible with the old Sun StorEdge A3x000/A1000 device node naming convention. As an example, Sun Enterprise Volume Manager is not compatible with:

  `dev/RAID_module01[0-7] and /dev/RAID_Module01/0s[0-7]`

  But it is fully compatible with Solaris™ Operating Environment device node naming convention:

  `/dev/rdsk/c3t4d0s0 and /dev/dsk/c3t4d0s0`

- Follow the Sun StorEdge A3x00 controller error recovery procedure as documented by the Sun StorEdge A3x00/A1000 subsystem documentation.

Caution – The error recovery procedures must be followed according to the Sun StorEdge subsystem documentation. Improper error recovery procedures can cause incompatibility with Sun Enterprise Volume Manager.

Installation of the Sun Enterprise Volume Manager With the Sun StorEdge HW RAID Subsystem

---

Caution – Any deviation from these steps may cause your HW RAID subsystem to be incompatible with Sun Enterprise Volume Manager. The Sun Enterprise Volume Manager should only be installed after the following steps have been completed and validated.

---

1. **The Sun StorEdge HW RAID subsystem is properly attached to the host computer.**
2. **The RAID Manager 6 software is properly installed including any patches as documented in the RM6 Installation Guide.**
3. **The Sun StorEdge A3x00 Logical Unit Numbers (LUNs) are properly configured using the RM6 software.**

Note – If you are using RM6 6.1, 6.1.1, 6.1.1 Update 1, 6.1.1 Update 2, or 6.22, you only need to perform steps 1-3, then proceed to Step 7.

---

4. **The host system is rebooted using the -r flag to rescan for attached devices upon reboot.**
5. **Upon reboot, the RM6 software has recognized the HW RAID subsystem LUNs and created the appropriate Solaris Operating Environment device nodes.**
6. **After your host comes back, bring up RM6 and verify that all the LUNs you created are there.**

7. **Install the Sun Enterprise Volume Manager software packages and any of its mandatory patches as documented in the Sun Enterprise Volume Manager Installation Guide.**

Using Sun Enterprise Volume Manager With Sun StorEdge A3x00/A1000

This section describes some benefits in using the host-based Sun Enterprise Volume Manager software and controller-based HW RAID. When the Sun Enterprise Volume Manager is used in conjunction with the Sun StorEdge HW RAID subsystems, synergy can be attained. An increase in the availability, performance, and manageability of the combined configuration can be realized.

Performance Analysis and On-line Load Balancing Between LUNs

Using the Sun Enterprise Volume Manager statistics-gathering capability allows administrators to analyze the I/O loads and responsiveness of volumes in the configuration and to move storage between disks. Since each LUN with the Sun StorEdge HW RAID subsystem looks like a single disk to the Sun Enterprise Volume Manager and host, it is not possible to identify the load on each spindle. But understanding your application and what you are trying to do helps a great deal in configuration of the HW RAID Subsystem before you install Sun Enterprise Volume Manager on top of the Sun StorEdge LUNs. To get maximized top performance and/or availability out of your HW RAID subsystem, use Sun Enterprise Volume Manager's performance analysis and on-line reconfiguration capabilities.

Before using Sun Enterprise Volume Manager's on-line reconfiguration utility, you might want to explore reconfiguration within the HW RAID subsystem on the Sun StorEdge A3x00. If one of your HW RAID controllers is taking a pounding I/O-wise and the other controller is idle, you might want to reconfigure at the HW RAID level (for example, moving the affected LUN over to the other controller).

Increasing Capacity On-line by LUN Concatenation

The size of a single file system or database tablespace is limited by the maximum size of a single LUN in a controller-based HW RAID subsystem. To create very large file systems or tablespaces, administrators can use Sun Enterprise Volume Manager to combine LUNs from multiple RAID controllers into one large “super-LUN” volume.

On-line Data Movement Between LUNs

The backup / reconfigure / reload cycle required to change the layout of HW RAID LUNs causes data access interruption and possible data loss. To reduce the likelihood of data loss, administrators can choose to construct the destination LUN according to the desired parameters (if extra disks are available), then copy data directly from the old location to the new one. However, even this requires the interruption of data access, since it is necessary to prevent changes to the old copy of the data after it has been copied to the new location.

If the data copying is performed by adding the new Sun StorEdge HW RAID LUN as a Sun Enterprise Volume Manager mirror of the data, however, all writes are delivered to all mirrors, keeping all copies up to date at all times. Once the mirror synchronization is complete, the mirror set can be separated, and the original LUN can be removed or used for other data. The data will have been kept online without the interrupting access throughout the operation.

Stable Backup

The method described above for data movement can also be used to provide consistent stable backup without interrupting user access. An additional mirror of data can be created, or an existing mirror can be detached from the updating applications, and the data instance can be backed up to tape or other offline storage pool. This guarantees a backup that is internally consistent at a single point in time. Upon completion of backups, the additional mirror space can be freed, or the mirror reattached to the live data volume and resynchronized.

Dynamic Partitioning

The dynamic partitioning capability of the Sun Enterprise Volume Manager is especially useful with large disks presented by the system for each LUN. The Solaris Operating Environment has a hard limit on the number of slices into which a disk can be partitioned. For the HW RAID Modules prior to 2.6 5/98, the maximum numbers of LUNs within a RAID Module was 8. With the release of 2.6 5/98 the maximum number of LUNs supported within a HW RAID Module unit is 16, provided certain steps are followed.

Sun Enterprise Volume Manager and Sun StorEdge HW RAID Supported Configurations

VxVM RAID 1 (Mirroring) and Sun StorEdge HW RAID Modules Configurations

- Sun Enterprise Volume Manager used to mirror between Sun StorEdge HW RAID Modules Non-Mirrored LUNs
- Sun Enterprise Volume Manager mirroring between multiple Sun StorEdge HW RAID Modules subsystems
- Sun Enterprise Volume Manager 3-way mirroring (triple mirroring)
- Sun Enterprise Volume Manager mirroring across Sun StorEdge HW RAID LUNs configured as RAID 0 (striping)
- Sun Enterprise Volume Manager mirroring across Sun StorEdge HW RAID LUNs configured as RAID 5

Sun Enterprise Volume Manager RAID 0 (Striping) and Sun StorEdge HW RAID Modules

- Sun Enterprise Volume Manager striping across Sun StorEdge HW RAID LUN(s) configured as RAID 5
- Sun Enterprise Volume Manager striping across multiple Sun StorEdge HW RAID Modules
- Sun Enterprise Volume Manager striping across Sun StorEdge HW RAID LUN(s) configured as RAID 1 (Mirroring)

Sun Enterprise Volume Manager Hot Relocation/Hot Sparing and Sun StorEdge HW RAID Hot Sparing

Sun Enterprise Volume Manager hot relocation and hot sparing allows the host system to automatically react to I/O failures on redundant (mirrored or RAID 5) VxVM objects and restore redundancy and access to these objects. Sun StorEdge HW RAID hot sparing allows the RAID Module to automatically react to I/O failures for RAID 1, 3, and 5. This is done within the HW RAID controller(s) and hidden from the host as well as Sun Enterprise Volume Manager.

If a disk failure occurs within the HW RAID Module, and the hot sparing option is configured, the HW RAID Module provides the disk failure redundancy. If data redundancy is provided by VxVM RAID 5 or mirrored configuration, then Sun Enterprise Volume Manager Hot Sparing or Hot Relocation can also provide disk or partial disk failure redundancy protection.

The most complete level of disk redundancy is achieved with both Sun Enterprise Volume Manager Hot Relocation/Hot Sparing and Sun StorEdge HW RAID hot sparing enabled. Issues that can arise out of this type of configuration are discussed in further detail later in this document.

Sun Enterprise Volume Manager and Sun StorEdge HW RAID Module Unsupported Configuration

- Sun Enterprise Volume Manager RAID 5 and Sun StorEdge HW RAID LUN(s) configured as RAID 5

  Using RAID 5 at both levels in the I/O subsystem can result in poor performance for no significant gain in reliability or availability. Use of this configuration is not supported.

Known Problems, Issues, and Answers To Questions

- When trying to initialize a HW RAID LUN with Sun Enterprise Volume Manager 2.4 using the vxdiskadm script, at the point where this LUN is being added to the diskgroup, you get the following message:

  ```
  NOTICE:vxvm:vxio: Disk cXtYdZs2: Unexpected status on close 19
  ```

After the HW RAID LUN is brought under Sun Enterprise Volume Manager control, this LUN can be used as normal. The problem appears to be the NOTICE message. This is fixed in Sun Enterprise Volume Manager 2.5.

- Q: What is the meaning of this message below?

```
WARNING: It is possible that VxVM is not bound to the RDAC
device nodes.You need to reboot or reset the configuration
daemon (vxconfigd -k -r reset) to ensure proper binding.
```

A: In a nutshell, it is telling you that VxVm 2.5 has detected some new HW RAID LUN(s) and if you want them brought under VxVm control, then you need to run vxconfigd -k -r reset in order to bring the Sonoma disk(s) under Sun Enterprise Volume Manager control. Or you can use vxdiskadm to bring the HW RAID LUN(s) under Sun Enterprise Volume Manager control. If you don't want the disk under VxVM control, don't worry about the message.

- Q: How do I remove logical units from Veritas control?

A: The deletion of LUNs under the configuration application requires obtaining exclusive access to the LUN(s) being deleted. If any program has any slice of a LUN open, exclusive access cannot be acquired. Mounted devices must be unmounted. LUNs configured under Veritas must be removed from Veritas control before they can be deleted.

This is also true of other RAID Manager 6 operations that require exclusive access. These include Reset Configuration, fixing multiple drive failures with the Recovery Guru, formatting a LUN with Options->Manual Recovery->Logical Units, and Firmware Upgrade->Offline in Maintenance/Tuning.

Removing a LUN from under Veritas control involves the following steps:

1. **Remove a LUN from a disk group**:

```
        vxdg [-g groupname] rmdisk <diskname>
example:  vxdg -g blinky -k rmdisk disk01
```

2. **Remove the LUN from Sun Enterprise Volume Manager control**

```
        vxdisk rm <devname>
example:  vxdisk rm c1t0d0s2
```

3. **Take the physical LUN offline:**

```
vxdisk offline <devname>
```

example: `vxdisk offline c1t0d0s2`

Q: Can I stripe across both controllers on a Sun StorEdge A3x00?

A: You can, but you will totally hose up controller failover. What you can do is create two 8+1 RAID 5, assign one to each controller, and then use VxVM to stripe them in the host. This delivers about 70 mb/sec single stream (a single LUN can deliver at most ~35 mb/sec in a single stream, using an 8+1 RAID 5).

- Q: I'm confused — what is the relationship of Sun Enterprise Volume Manager to controller failover and how does Sun Enterprise Volume Manager react to a failover?

  A: The basic issue underlying this question is a misunderstanding of how the various layers of software and firmware work together. If one understands these layers, it is a lot easier on the people you support with the StorEdge HW RAID Modules.

  Let's look at this in a model similar to the OSI Networking model. There are a whole bunch of layers, starting with the application at the top. Let's say the application is SAP, although it could be anything. SAP is built on top of a DBMS, such as Oracle or Informix. SAP doesn't know what's living underneath the DBMS, just that *something* is under there.

  The DBMS puts all of its stuff into a bunch of tablespaces, which are either files residing in a file system, or raw devices. A file system actually resides on a raw device, so the DBMS may be bypassing a layer here. Examples of raw devices are /dev/rdsk/c0t0d0s0, /dev/vx/rdsk/vol05, /dev/md/rdsk/md23, and /dev/rdsk/c13t3d0s2. They have slightly different names, but they are, from either Oracle or UFS's point of view, the same thing: a UNIX® raw device.

  For example, you can newfs any of these devices, then mount them as a file system. Or you can disk init them (that's a Sybase term) and plop a tablespace onto the device. I used UFS as the file system here, but it could be anything, such as VxFS or anything else.

Note that one of those devices was /dev/vx/rdsk/vol05, which happens to be a VxVM volume. It can be any type of volume: it could be a RAID 5 <choke>, a mirrored plex, a simple volume, etc. The point here is that it's just a raw device from the perspective of the next layer up.

At the same time, VxVM is working *only* with raw disk devices (for example, /dev/rdsk/c13t3d0s2 and /dev/rdsk/c0t0d0s0). The nature of the devices isn't really known to VxVM, other than that they have a specific size and name. In particular, one of our devices, /dev/rdsk/c13t3d0s2, could turn out to be a Sonoma LUN, which might be 36GB in size. The controller (meaning c13) may be a virtual controller, in that it may be a virtual composition of c7 and c8, as aggregated by RDAC. VxVM does not know the nature of these devices, as they simply present the appearance of a raw device. You could, for example, create a VxVM plex mirroring /dev/rdsk/c13t3d0s2 and /dev/rdsk/c15t0d0s2, as long as they are the same size. How would these be managed? Via RM6, which commands the controller(s) to construct LUNs.

In this case the LUNs I've picked are 4+1 LUNs consisting of 9GB disks, which is why they're 36GB in size. Since VxVM is using a virtual controller, either c7 or c8 could fail, and VxVM will never know, because that is hidden behind the virtual controller interface. A member disk in c13t3d0s2 might fail, and again VxVM would never know, since the controllers will at least run the RAID 5 in degraded mode and might even spare the failing drive. The only way the mirrored plex would detach (in this case) is if any of the following series of events occurred:

- *Two* (or more) members of c13t3d0s2 failed. Thus the RAID 5 LUN is semantically no longer viable.
- Both c7 and c8 fail, thus making the entire LUN unavailable to VxVM, and the entire LUN just disappears.
- The administrator explicitly detaches a plex.

- Q: During the LCS program for Sun™ Cluster 2.1 and Sun StorEdge A3x00, it was discovered in HA configurations using Sun Enterprise Volume Manager that DMP is *not* supported under SC2.1. Is that because of DMP's own dual-path algorithms, coupled with RDAC's own dual-pathing?

  A: If there was a issue with the Sun StorEdge A3x00 on a host with Sun Enterprise Volume Manager 2.5 and even if there were *not* any Sun StorEdge A5000s in the configuration, DMP would still be active and can cause timing issues with Sun StorEdge A3x00s. Therefore, in

configuration with Sun StorEdge A3x00 and Sun Enterprise Volume Manager 2.5 and no Sun StorEdge A5000(s), it is recommended that you turn *off* DMP. The procedure is documented in the Sun Enterprise Volume Manager documentation. This issue is *not* specific to Sun Cluster 2.1.

- If you are mirroring between Sun StorEdge A3X00s (double or triple mirroring), you should throttle down the rdac retry count. This enables rdac to retry on its alternate path depending on the number of I/Os queued up and pass on to sd that it has failed and therefore is able to detach the plex and continue I/O over on the other Sun StorEdge A3x00. There is a small script (five lines) to set this variable and this variable will be included in the next spin of RM6 and be a variable in the rmparams file. This will be fixed in patch 106707-01. You can get this script and its README off the SunSolve℠ web page.

- By default Sun Enterprise Volume Manager does Hot Relocation on the free space within your Sun StorEdge A3x00/A1000/A5000 or any other disk subsystem under controller by Sun Enterprise Volume Manager. As an example of an extreme case, if you were triple mirroring Sun StorEdge A3x00 and you *lost* a Sun StorEdge A3x00, it would Hot Relocate the *entire* Sun StorEdge A3x00.The only way to recover from this operation would be to restore *or* go to the URL http://spider.aus/utils/utils-vxvm.html, where there are two utilities that you can use to recover from a Hot Relocation operation. The name of the two utilities are: vxreconstruct and vxunrelocate. Or you can disable Hot Relocation by using the following command:

```
vxedit set reserve=on Volume_Manager_Disk
```

Note – Hot Relocation works on failed subdisk.

- The only time Sun Enterprise Volume Manager's Hot Sparing would kick in is when you lost the whole Sun Enterprise Volume Manager disk. There is a tremendous amount of I/O activity during a Hot Spare operation. If you need Hot Spare Disk(s), it is recommended that you do this within the HW RAID Controllers. One reason is to cut down on the overhead between the host and the RAID Subsystem. It is strongly recommended that you make your Hot Spare Disk(s) the biggest size as the largest disk you have in your RAID Subsystem. As an example, if you had a 4+1 RAID 5 and you were using 9-GB drives, your Hot Spare Disk should be at least 9 GB in size. If you use a smaller disk, there will be no Hot Sparing for that LUN and you'll be running in degraded mode.

Note – The last two items are not specific to Sun Cluster 2.1.

- Q: I was running a configuration with a Sun StorEdge A3500 with 16 LUN support turned on and using VM to mirror my LUNs. I had reinstalled my RM6 software and when my system rebooted, I couldn't see a bunch of my LUNs and VM was giving “alt used” messages all over the place.

  A: When you reinstalled your RM6 software, a copy of the rmparams file was installed specifying by default that 8 LUNs are supported. You need to change the variable System_MaxLunsPerController=8 to specify System_MaxLunsPerController=16, and then do a reconfig reboot. The “alt used” is a VM message telling you that it is using the alternate label. If you have done nothing else to your configuration, everything will return to normal after the reconfig reboot. In the worst case, you would have to rebuild those VM disk by hand using vxprint -ht as a guide.

- Q: I have a Sun StorEdge A3500 with VM, and as a test, I switched the cables around on the back and on reboot I got the following:

```
VxVM general startup...
vxvm:vxconfigd: ERROR: enable failed: Error in disk group
configuration copies
        No valid disk found containing disk group; transactions are
disabled.
vxvm: Vold is not enabled for transactions
        No volumes started
vxvm:vxconfigd: ERROR: enable failed: Error in disk group
configuration copies
        No valid disk found containing disk group; transactions are
disabled.
The system is coming up.  Please wait.

Array Monitor initiated
RDAC daemons initiated
volume management starting.
RDAC Resolution Daemon locked in memory
vxvm:vxrecover: ERROR: IPC failure: Configuration daemon is not
accessible
```

Then I lost all his data.

After moving the cables back to the original slots, I got the following:

```
WARNING: /sbus@3,0/QLGC,isp@0,10000/sd@5,0 (sd35):
        i/o to invalid geometry
WARNING: /sbus@3,0/QLGC,isp@0,10000/sd@5,1 (sd365):
        i/o to invalid geometry
```

A: This is bugid 4180291, which was first discovered in VM 2.6 and has been verified as fixed with VM 3.0.1. They are trying to back port the fix to VM 2.6. Because of this, any data that is involved will be lost.

- Q: I have a Sun StorEdge A3500 installed on my host and I went to install VM 2.6 and during the package installation of SUNWvxvm, my system panicked.

  A: This is bugid 4216620. There is a workaround but it goes against what we have been preaching to the field about installation order. Before you begin to install VM 2.6, do a touch of /kernel/drv/ap and when you install VM 2.6, it will see “ap” and not install DMP and continue with the installation without problems. With RM6 6.22 as part of the installation, it will do the touch /kernel/drv/ap if ap is not present.

Sun Microsystems Computer Company
A Sun Microsystems, Inc. Business
901 San Antonio Road
Palo Alto, CA 94303 USA
650 960-1300
FAX 650 969-9131
http://www.sun.com

Sales Offices

Argentina: +54-1-317-5600
Australia: +61-2-9844-5000
Austria: +43-1-60563-0
Belgium: +32-2-716-7911
Brazil: +55-11-5181-8988
Canada: +905-477-6745
Chile: +56-2-638-6364
Colombia: +571-622-1717
Commonwealth of Independent States:
+7-502-935-8411
Czech/Slovak Republics:
+42-2-205-102-33
Denmark: +45-44-89-49-89
Estonia: +372-6-308-900
Finland: +358-9-525-561
France: +33-01-30-67-50-00
Germany: +49-89-46008-0
Greece: +30-1-680-6676
Hong Kong: +852-2802-4188
Hungary: +36-1-202-4415
Iceland: +354-563-3010
India: +91-80-559-9595
Ireland: +353-1-8055-666
Israel: +972-9-956-9250
Italy: +39-39-60551
Japan: +81-3-5717-5000
Korea: +822-3469-0114
Latin America/Caribbean:
+1-650-688-9464
Latvia: +371-755-11-33
Lithuania: +370-729-8468
Luxembourg: +352-491-1331
Malaysia: +603-264-9988
Mexico: +52-5-258-6100
Netherlands: +31-33-450-1234
New Zealand: +64-4-499-2344
Norway: +47-2218-5800
People's Republic of China:
Beijing: +86-10-6849-2828
Chengdu: +86-28-678-0121
Guangzhou: +86-20-8777-9913
Shanghai: +86-21-6247-4068
Poland: +48-22-658-4535
Portugal: +351-1-412-7710
Russia: +7-502-935-8411
Singapore: +65-438-1888
South Africa: +2711-805-4305
Spain: +34-1-596-9900
Sweden: +46-8-623-90-00
Switzerland: +41-1-825-7111
Taiwan: +886-2-514-0567
Thailand: +662-636-1555
Turkey: +90-212-236 3300
United Arab Emirates:
+971-4-366-333
United Kingdom: +44-1-276-20444
United States: +1-800-821-4643
Venezuela: +58-2-286-1044
Worldwide Headquarters:
+1-650-960-1300