Introduction to Sun Interprise Volume Manager (SEVM) and Solstice DiskSuite (SDS) Differences

Joy Silva
SMCC Storage Products
March 1998

Software Overview of SEVM

Packages

- application SUNWasevm Sun Enterprise Volume Manager 2.5 AnswerBook
- system SUNWvmdev Sun Enterprise Volume Manager (header files)
- system SUNWvmman Sun Enterprise Volume Manager (manual pages)
- application SUNWvxva Sun Enterprise Volume Manager Visual Administrator
- system SUNWvxvm Sun Enterprise Volume Manager

Daemons

- vxconfigd - Responsible for maintaining Volume Manager (VM) disks and disk group configurations. The daemon communicates configuration changes to the kernel and modifies the configuration information stored on the disks. vxconfigd must be running before VM operations can be performed.

- vxiod - The volume extended I/O daemon controls and maintains VM I/O daemons. VM I/O daemons are not required for normal operation but increase performance.

Directories

- /usr/sbin - commands
- /etc/vx - control files and special utilities
- /opt/SUNWvxvm - man pages and libraries
- /opt/SUNWvxva - GUI files

Startup scripts

- /etc/init.d/
  vxvm-reconfig vxvm-recover vxvm-startup1 vxvm-startup2 vxvm-sysboot

Commands

- /usr/sbin/

vxassist	**vxdisk**	vxinfo	vxmake	**vxprint**	vxserial.old
vxconfigd	vxdiskadd	vxinstall	vxmend	vxrecover	vxstat
vxdctl	**vxdiskadm**	vxiod	vxnotify	vxsd	vxtrace
vxdg	vxedit	vxlicense	vxplex	vxserial	vxvol

Paths

PATH=/usr/sbin:/opt/SUNWvxva/bin
MANPATH=/usr/man:/opt/SUNWvxvm/man

Summary of Differences

- SEVM supports Dynamic Multi-Pathing (DMP) as of version 2.5. SDS has no plans to support DMP at this time, although the E10000 has alternate pathing (AP) software that SDS can use.
- SEVM is Veritas software, an open package which runs on multiple platforms. SDS is Sun Proprietary and future enhancements or software development may be limited.

Technical differences include:

1. Disk partitions
   - SDS - uses hard partitions
   - SEVM - uses soft partitions
2. Disk grouping
   - SDS - uses metasets in an HA environment
   - SEVM - uses disk groups

3 Disk configuration data
   - SDS - metadbs manually configured, must maintain quorum
   - VM - private areas on all disks, majority kept active for quorum

4. Metadevice/Volume creation and reconfiguration
   - SDS - metadevice creation is harder and requires more knowledge and planning
   - SEVM - volume creation with GUI is very easy requiring little knowledge and minimal planning
5. Disk replacement
   - SDS - several manual steps required
   - SEVM - easy menu driven program
6. Troubleshooting
   - SDS - simple
   - SEVM - complex
7. RAID capabilities
   - SDS - uses raid 1+0. Has ufs logging available. Can offline/online mirrors
   - SEVM - Currently only 0+1 is implemented although this feature will be available in VM version 3.x. Must use vxfs if filesystem logging is required. Snapshot utility is closest to offline/online but requires a complete mirror copy.

1. Disk Partitions

- SDS - Uses standard unix partitions as defined by the disk label.
  - Data is stored on set partitions so is easy to recover if the metadb configuration databases become corrupted.
  - Limit of 8 partitions per disk.
  - Partitions must be manually configured.

Example of metadevice info:
 # metastat -p d0
d0 -m d1 d2 1
d1 1 3 c1t0d0s2 c1t2d0s2 c1t4d0s2 -i 32b -h hsp001
d2 1 3 c1t1d0s2 c1t3d0s2 c1t5d0s2 -i 32b -h hsp001

- SEVM - Uses soft partitions call “subdisks” which are defined in the configuration databases.
  - subdisk information and data loss can occure if configuration information is corrupt.
  - Unlimited subdisks per disk.
  - Easy automatic creation and configuration.
  - The vxprint command is used to obtain subdisk information for a volume

Example of volume info:
 # vxprint -ht vol01
Disk group: test

```
V  NAME        USETYPE     KSTATE   STATE    LENGTH   READPOL   PREFPLEX
PL NAME        VOLUME      KSTATE   STATE    LENGTH   LAYOUT    NCOL/WID
MODE
SD NAME        PLEX        DISK     DISKOFFS LENGTH   [COL/]OFF DEVICE   MODE

v   vol01        fsgen       ENABLED   ACTIVE   25135104 SELECT    -
pl  vol01-01     vol01       ENABLED   ACTIVE   25136144  STRIPE    3/128   RW
sd test02-01   vol01-01    test02     0         8378640    0/0       c1t0d3   ENA
sd test03-01   vol01-01    test03     0         8378640    1/0       c1t1d3   ENA
sd test04-01   vol01-01    test04     0         8378640    2/0       c1t2d3   ENA
```

Example of mounted volumes/metadevices:
 # df -k

```
depth:/home/joys/photon% df -k
Filesystem              kbytes      used       avail      capacity   Mounted on
/dev/dsk/c0t0d0s0       1704901     1443156    91255      95%        /
/dev/md/dsk/d0          6211010     9          558990     1%         /d0
/dev/vx/dsk/test/vol01  11817908    569736     4938750    54%        /vol01
```

2. Disk Grouping

- SDS - SDS in a High Availability (HA) environment allows the use of "metasets", these are groups of disks which can can be be moved between hosts for failover purposes.

- SEVM - All disks are created in "disk groups"
  - Each host must have a default "rootdg" disk group which cannot be shared.
  - Data disk groups (non-rootdg) can be easily deported and imported between hosts.
  - Its best to keep rootdg for boot disk mirroring only, or use a local non-array disk. This configuration allows easier maintenance and upgrade procedures.
  - Disk group information can be obtained with the "vxdg" and "vxdisk" command.

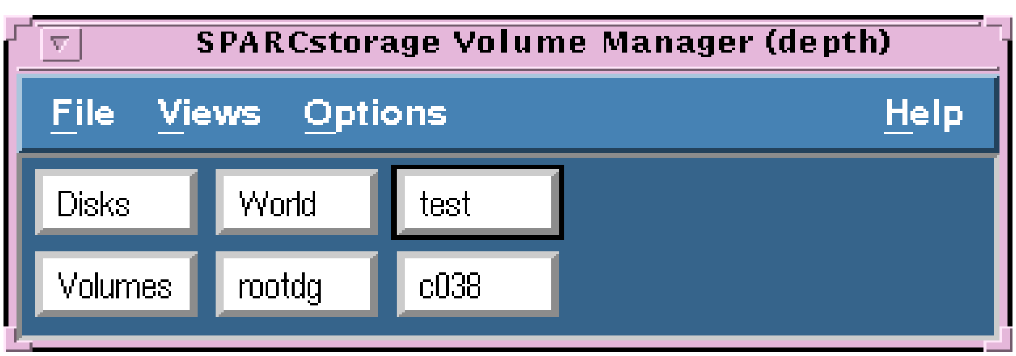
Example:
 # vxdg list

NAME	STATE	ID
rootdg	enabled	865465081.1025.rootdg
test	enabled	864406424.1126.depth

\# vxdisk -g test list

DEVICE	TYPE	DISK	GROUP	STATUS
c1t0d3s2	sliced	test02	test	online
c1t0d4s2	sliced	test08	test	online
c1t1d3s2	sliced	test03	test	online
c1t1d4s2	sliced	test09	test	online
c1t2d3s2	sliced	test04	test	online
c1t2d4s2	sliced	test10	test	online
c1t3d3s2	sliced	test05	test	online
c1t3d4s2	sliced	test11	test	online
c1t4d3s2	sliced	test06	test	online
c1t4d4s2	sliced	test12	test	online
c1t5d3s2	sliced	test07	test	online
c1t5d4s2	sliced	test13	test	online

Example of main GUI:



3. Disk Configuration Data

- SDS - Uses separate manually created partitions to contain the “metadb” or metadevice configuration information.
  - metadbs must maintain a quorum of copies, minimum of 3

Example:
 # metadb

flags	first blk	block count	
a m pc luo	16	1034	/dev/dsk/c0t0d0s4
a pc luo	1050	1034	/dev/dsk/c0t0d0s4
a pc luo	2084	1034	/dev/dsk/c0t0d0s4

- SEVM - All disk have a “private region” for configuration information
  - When disks are “initialized” for use by SEVM the private (slice 3) and public (slice 4) areas are created. Private for config info, public for subdisks and data.
  - A majority of private regions are kept enabled for quorum.

Example of SEVM managed disk label:
 # prtvtoc /dev/rdsk/c1t0d0s2

```
* /dev/rdsk/c1t0d0s2 partition map
*
*                          First      Sector     Last
* Partition Tag  Flags     Sector     Count      Sector     Mount Directory
2            5   01           0       4154160    4154159
3           15   01           0          1520       1519
4           14   01        1520       4152640    4154159
```

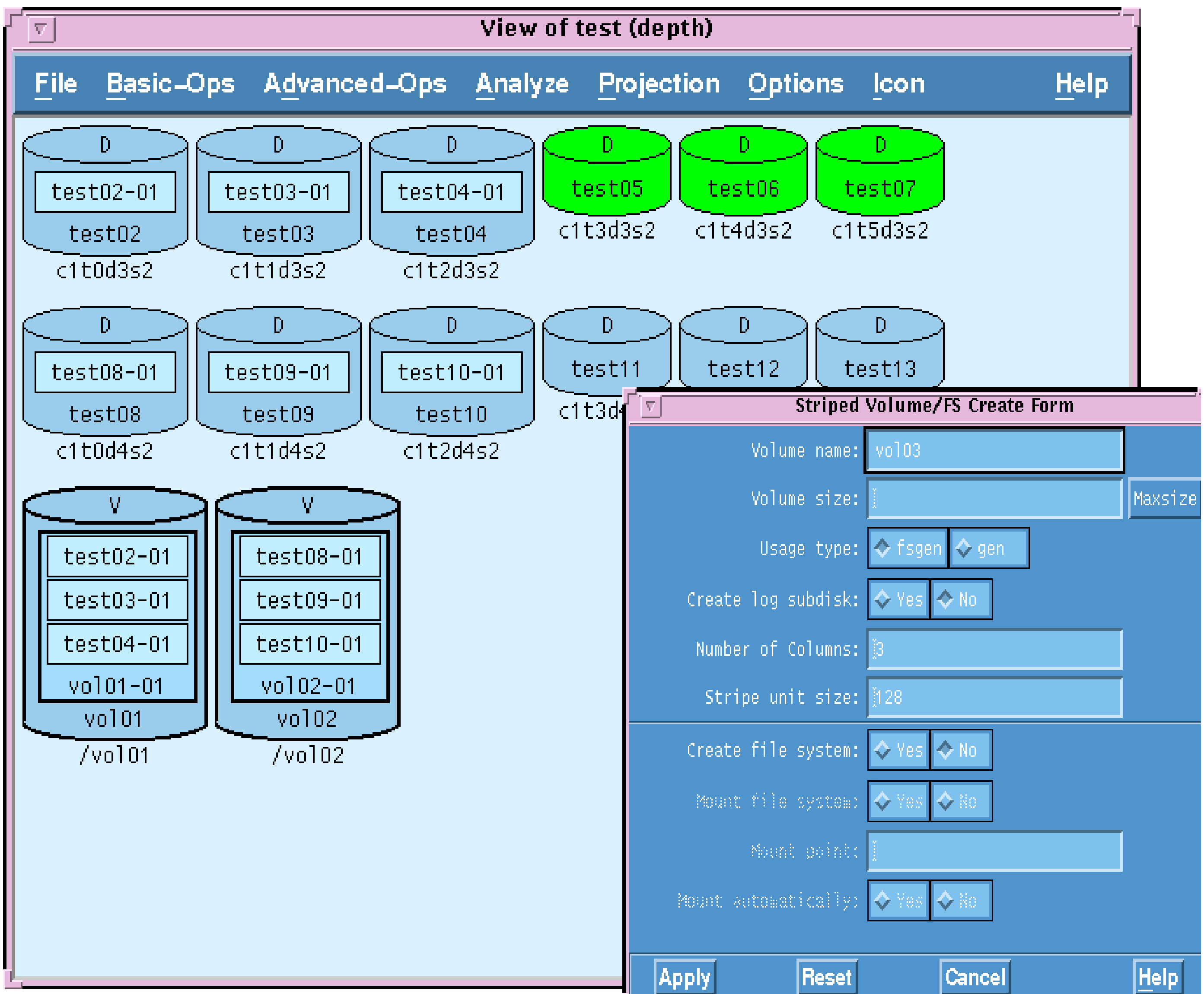
Example of disk group list:
 # vxdg list test

```
Group:    test
dgid:     864406424.1126.depth
import-id: 0.1474
flags:
copies:   nconfig=default nlog=default
config:   seqno=0.2402 permlen=1563 free=1550 templen=9 loglen=236
config disk c1t0d3s2 copy 1 len=1563 state=clean online
config disk c1t0d4s2 copy 1 len=1563 disabled
config disk c1t1d3s2 copy 1 len=1563 state=clean online
config disk c1t1d4s2 copy 1 len=1563 disabled
config disk c1t2d3s2 copy 1 len=1563 state=clean online
config disk c1t2d4s2 copy 1 len=1563 disabled
config disk c1t3d3s2 copy 1 len=1563 state=clean online
```

4. Metadevice/Volume Creation

- SDS - Metadevices require advance configuration planning and manual setup
  - metadb quantity, partition selection and creation
  - manual formating of disk labels and md.tab setup
- SEVM - Automatic creation of rootdg and private slices
  - vxinstall script used for initial installation and creation of rootdg
  - additional disk group and volumes easily created with user friendly GUI
  - Difficult command line setup and reconfiguration

Example of GUI:



5. Disk Replacement

- SDS - Several manual commands required
  - repartition disk
  - restore any metadbs
  - metareplace command

- SEVM - Menu driven program with "vxdiskadm" script, GUI or commands

Example:
 # vxdiskadm
Volume Manager Support Operations
Menu: VolumeManager/Disk

1	Add or initialize one or more disks
2	Encapsulate one or more disks
3	Remove a disk
4	Remove a disk for replacement
5	Replace a failed or removed disk
6	Mirror volumes on a disk
7	Move volumes from a disk
8	Enable access to (import) a disk group
9	Remove access to (deport) a disk group
10	Enable (online) a disk device
11	Disable (offline) a disk device
12	Mark a disk as a spare for a disk group
13	Turn off the spare flag on a disk
list	List disk information
?	Display help about menu
??	Display help about the menuing system
q	Exit from menus

Select an operation to perform:

6. Troubleshooting

- SDS - Simple with few commands and options
  - metastat often gives recovery command

Example of device in error:

```
# metastat d5
d5: RAID
    State: Resyncing
    Resync in progress: 21% done
    Hot spare pool: hsp001
    Interlace: 32 blocks
    Size: 12456400 blocks
Original device:
    Size: 12456864 blocks
Device          Start Block  Dbase State        Hot Spare
c1t0d1s2            1850     No    Okay
c1t2d1s2            1850     No    Okay
c1t4d1s2            1850     No    Okay
c1t3d1s2            1850     No    Resyncing    c1t5d1s2
```

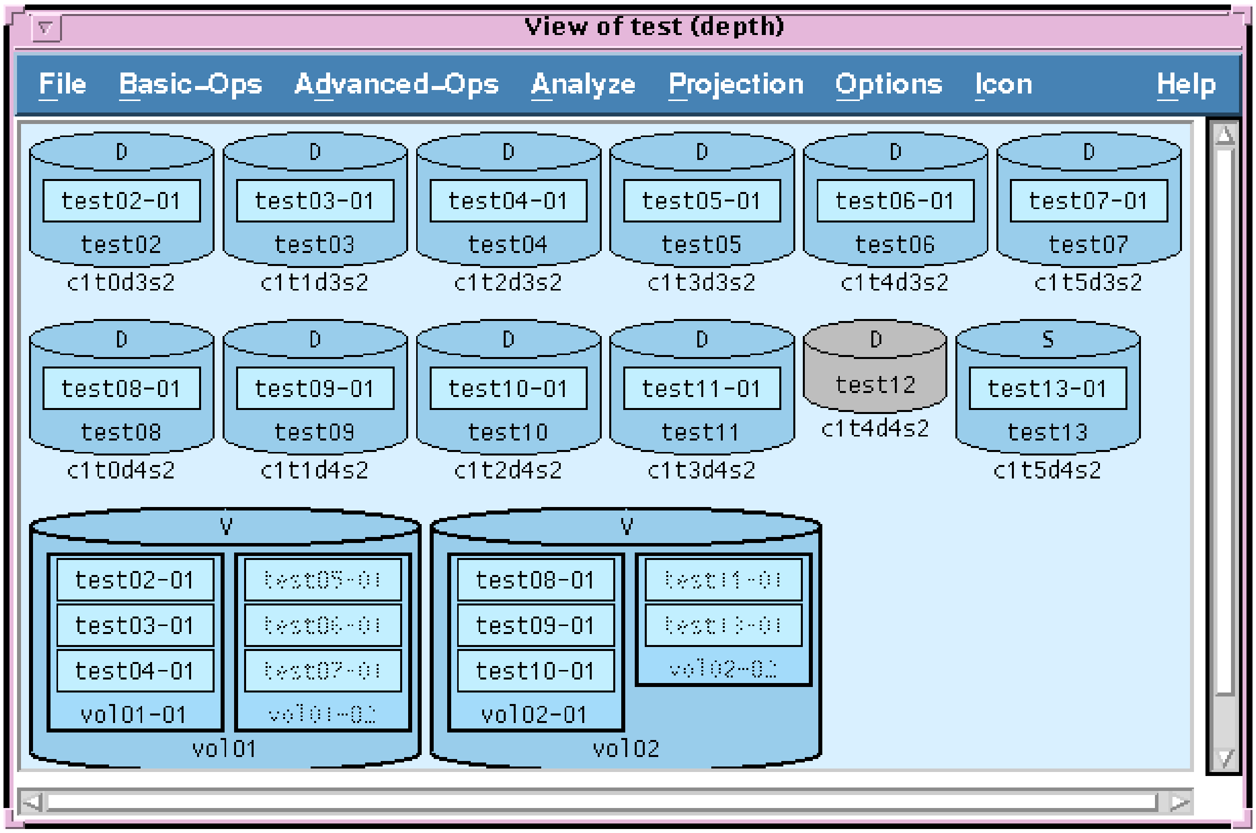
- SEVM - Can be complex with dozens of commands, options and possible scenarios
  - Multiple possible states of subdisks, plexes and volumes require training and experience to understand, recognise and troubleshoot
  - Loss of configuration info can be impossible to recover from (keep vxprint -ht output)

Example of volume with error:

```
# vxprint vol02
Disk group: test

TY NAME       ASSOC     KSTATE    LENGTH   PLOFFS  STATE    TUTIL0  PUTIL0
v  vol02      fsgen     ENABLED   2097152  -       ACTIVE   ATT1    -
pl vol02-01   vol02     ENABLED   2099648  -       ACTIVE   -       -
sd test08-01  vol02-01  ENABLED   699840   0       -        -       -
sd test09-01  vol02-01  ENABLED   699840   0       -        -       -
sd test10-01  vol02-01  ENABLED   699840   0       -        -       -
pl vol02-02   vol02     ENABLED   2099616  -       STALE    ATT     -
sd test11-01  vol02-02  ENABLED   1049760  0       -        -       -
sd test13-01  vol02-02  ENABLED   1049760  0       -        -       -
```

Example of same error from GUI view:



7. RAID Features

- UFS Logging - Used for fast fscks after system crash (seconds instead of hours)
  - SDS - Incuded
  - SEVM - Not available unless the vxfs filesystem is purchased and used.

- RAID5 Logging - Protects data and parity from possible corruption after a system crash.
  - SDS - Default on all RAID5 devices. Cannot disable.
  - SEVM - Optional (recommended).

- Hot Spare Disks - Will automatically replace a failing data disk.
  - SDS - Can be assign to individual disks and metadevices. Will replace entire disk even for a partial disk failure.
  - SEVM - Disks are assigned to entire disk group. Choice of 2 modes:
    - Hot sparing - Entire is disk is replaced but on solid disk failure only (retest in slice 3 fails). Partial disk failures and not hot spared.
    - Hot relocation - Protects against partial failure. Will use any unused disk space to relocate a failing subdisk. Can scatter subdisks with a multiple disk or array failure.

- Mirror Offline/Online - Ability to temporarily remove a mirror then reattach without doing a full resync. Useful for backups and sharing of data.
  - SDS - Yes
  - SEVM - No. The closet task is a “snapshot” which creates a tempory additional mirror, a full sync is required.

- RAID 1+0 - Disks are mirrored before being stripped. This protects from multiple disk failures and increase recovery time on resyncs.
  - SDS - Standard
  - SEVM - Not available yet. Planned for future release (Nov 98?).

Resources

Sun Educational Services Training - http://suned.sun.com/suned/

- SA-345 Volume Manager with SPARCstorage Array - 2 days

Veritas Training - http://www.veritas.com/

- Volume Manager - 3 days
- Performance Tuning and Recovery for Volume Manager - 2 days
- File System - 1 day

Documentation

- Sun Enterprise Volume Manager Users Guide - 805-1603
- Sun Enterprise Volume Manager Administrator's Guide - 805-1607
- Sun Enterprise Volume Manager Software Installation Guide - 805-1604
- Sun Enterprise Volume Manager Software Release Notes - 805-1605