Sun Enterprise™ 10000 Server Floating Tape Library Solution

By Enrique Vargas - Enterprise Engineering

Sun BluePrints™ OnLine - January 2000

http://www.sun.com/blueprints

Sun Microsystems, Inc.
901 San Antonio Road
Palo Alto, CA 94303 USA
650 960-1300 fax 650 969-9131
Part No.: 806-4394-10
Revision 01, January 2000
Sun Enterprise™ 10000 Server Floating Tape Library Solution

Introduction

This is the fourth article in the Dynamic Reconfiguration (DR) series. It compiles the results of a consolidated effort between Sun Microsystems, Inc. and VERITAS Software Inc. to determine the precise requirements and appropriate methodology to provide two separate Sun Enterprise™ 10000 server (Starfire™) Dynamic System Domains (DSDs), each configured as a Sun StorEdge Enterprise NetBackUp™ master server, with shared access to a Sun StorEdge™ L3500 tape library without any disturbance to their operating system instances.

There are an increasing number of Starfire server customers running mission critical datacenters which deal with data warehousing legacy/enterprise data and database/file/application servers that require backups and restores to move large amounts of data to and from tape media in a short period of time.

The Floating Tape Library solution provides mission critical customers with an alternative method for reducing their backup/restore time windows. It also helps them to obtain the best return on investment from the purchase of expensive tape libraries. This solution currently involves manual interaction. However, this could be automated in the future when the appropriate scripting interface becomes available for DR. Until then, we recommend that this methodology be used when backup/restore time windows are critical and a network client/server backup/restore setup is not sufficient.

The methodology introduced in this article was developed from a lab exercise. It addresses the appropriate DR requirements as well as the sequential steps that are needed to detach a tape library from a source DSD and attach it to a destination DSD. To avoid excessive content, this article does not include setup details for the
different software modules involved, and it may not provide sufficient information to implement a full backup/restore solution in your datacenter. Contact Enrique.Vargas@Sun.COM to obtain additional details.

If you have already implemented DR in your datacenter and are interested in sharing your experiences, or if you would like to see us address a topic that will empower you to take better advantage of DR, please give us your feedback through the "Tell Us" button at the end of this article on http://www.sun.com/blueprints.

Strategy

The strategy adopted for the Floating Tape Library exercise involved the implementation of two single-system board DSDs named flinstones and marmaduke within a single Starfire server (see FIGURE 1). The System Service Processor (SSP) managing the Starfire platform which sustained these two domains was running the SSP software version 3.1.

DSDs are logical, hardware-isolated server entities that can add (attach) or remove (detach) hardware resources through the use of Dynamic Reconfiguration (DR) software and without incurring any down time on their operating system instances. DSD's electrical isolation provides a fault-fencing mechanism which allows applications to run independently of each other within the same Starfire platform.

Each DSD in this lab exercise had access to its own mirrored Unipack boot disk running the Solaris™ 2.6 Operating Environment. Each of the system boards was populated with:

- four 250Mhz UltraSParc-II CPUs (each CPU hosting one Megabyte of external cache)
- one Gigabyte of memory

Both DSDs in this lab exercise had shared access to a mirrored Sun StorEdge™ A5000 Array which sustained seven separate filesystems. Each filesystem mapped to an individual tape drive on the tape library to perform efficient parallel work. Mirroring and disk-group shared access were implemented using the VERITAS VxVM version 2.5 volume manager software. Mirroring was used to protect the filesystem data from hardware outages. Shared access was used to instantly migrate the filesystem data between hosts to perform backup and restores on an alternate host.

**Note** – Disk groups are migrated between hosts using VxVM’s `vxdg` utility using the import/export option.
A single StorEdge L3500 tape library, populated with fourteen DLT tape cartridges, was attached to a system board. The tape library’s seven DLT tape drives and a single robotic interface mapped perfectly to the four on-board SBus DWIS interfaces which ended up sustaining two SCSI devices each.

The Sun StorEdge™ L3500 tape library represents tape-drive technology committed to the high-end spectrum of the Solaris Operating Environment. The Sun StorEdge L3500 tape library can support a maximum of seven DLT 7000 tape drives and it provides random access to a maximum of 100 DLT cartridges, each having an uncompressed capacity of 35 Gigabytes for a total native capacity of 3.5 TeraBytes. This access is provided through a robotic interface implementing the SCSI-2 medium changer command set.

Note – Sun Microsystems, Inc. currently offers the Sun StorEdge™ L11000 tape library which supports a maximum of 16 DLT drives and a total of 326 cartridges in the same library to achieve a total native capacity of 11.4 TeraBytes.

Shared access to the tape library was achieved through use of the DR software. FIGURE 1 shows the marmaduke DSD having current ownership of the tape library, which means that it has ownership of the following resources:

- eight UltraSPARc-II CPUs (distributed between system boards 6 and 7)
- two Gigabytes of memory (distributed between system boards 6 and 7)
- seven DLT tape drives (associated with system board 7)
- one tape library robotics interface (associated with system board 7)

Both DSDs were loaded with the Sun StorEdge Enterprise NetBackUp™ version 3.1.1 software. The Sun Storedge Enterprise NetBackUp software represents Sun Microsystems’ high-performance, industrial-strength backup and restore solution which extends to UNIX and PC network clients using TCP/IP. The StorEdge Enterprise NetBackUp software master/slave architecture enables centralized administration and provides scalability to service multiple clients and multiple servers.

Both DSDs for this lab exercise were configured to be Sun StorEdge Enterprise NetBackUp master servers, which means that each DSD has its own private version of the Sun StorEdge Enterprise NetBackUp database and could provide backup/restore services to network clients and slave servers.
All tape device links in each domain were created (executing a `boot -r` or simply by executing the `drvconfig;links; tapes` sequence) while each domain had ownership of the tape library through DR. The 14 available tape cartridges inside the tape library were configured and partitioned into two separate pools of seven tapes each using the Sun StorEdge Enterprise NetBackUp software to provide secured access to each DSD. Use of the DR technology enabled the physical attachment of tape devices to the master server host, thus eliminating the need for:

- additional network controllers and their respective I/O slots
- increased latency involved with disk-network-tape data transfers
- TCP/IP processing overhead
FIGURE 1  Floating Tape Library Connectivity Diagram
The Solaris™ Operating Environment Pre-requisites

There are a number of Solaris Operating Environment pre-requisites which have to be fulfilled to help facilitate the success of DR operations. Such pre-requisites change with different versions of the operating system and depend on the devices present on the system board to be attached/detached. The pre-requisites addressed by this article reflect our particular domain configuration. You should review the “Starfire DR-Detach and DR-Attach Requirements” BluePrints Online article published in 8/99 to see the complete list of pre-requisites for any configuration.

Operating System Patches

A full list of recommended patches for the Solaris 2.6 Operating Environment and current revisions can be obtained from your local service provider. The following list of patches applied to both DSDs during this lab exercise reflects basic requirements:

- 105181: Kernel update patch with several DR fixes included
- 106381: Network console (netcon) fixes
- 106048: dr and hswp driver fixes
- 106284: dr_daemon patch fixes

Memory Interleaving

DR requires that memory interleaving be disabled between system boards. Memory interleaving helps improve memory subsystem performance by reducing the probability of hot spots or contention by a single memory bank through having sequential access spread out to multiple memory banks. Memory interleaving is disabled by default, but the .postrc file on the SSP can enable interleaving through the mem_interleave_ok directive (postrc(1M)).
dr-max-mem OBP Variable

In the Solaris 2.6 Operating Environment the \texttt{dr-max-mem} OBP environment variable is used as a vehicle to enable DR functionality. The default value for the \texttt{dr-max-mem} OBP environment variable is 0. It must be set to 1 to enable DR. The following line was executed on both the \textit{flinstones} and \textit{marmaduke} domains from the OBP prompt:

\texttt{<#XX> ok setenv-dr dr-max-mem 1}

\begin{verbatim}
dr-max-mem = 1
\end{verbatim}

\textbf{Note} – The operating system needs to be rebooted once the \texttt{dr-max-mem} OBP environment variable is set in order to be acknowledged.

Methodology for Detaching the Tape Library

This section describes the steps that must be followed to detach system board 7, which houses the Tape Library, from the \textit{marmaduke} domain.

Marking removed tape drives unavailable within Sun StorEdge Enterprise NetBackUp software

Marking tape drives unavailable within the Sun StorEdge Enterprise NetBackUp application is recommended to make sure missing resources are accounted for when the application is restarted. For our lab exercise there is no sense in bringing the Sun StorEdge Enterprise NetBackUp application back on line since we are indeed removing all available resources with a DR-Detach operation. The following command line is executed on the \textit{marmaduke} domain to remove all detached tape drives from Sun StorEdge Enterprise NetBackUp control:

\texttt{/usr/openv/volmgr/bin/vmoprcmd -down X<CR>}, where \texttt{X=0-6}
Bringing down the Sun StorEdge Enterprise NetBackUp application to relinquish I/O resources

The DR-detach operation requires that no open device instances be associated with the system board to be removed. By bringing down the Sun StorEdge Enterprise NetBackUp application, the robotics interface and all seven drives associated with the Tape Library are closed device instances. The following command lines are execution choices to be executed on the marmaduke domain in order to bring down the Sun StorEdge Enterprise NetBackUp application:

```
/etc/rc0.d/K77netbackup<CR>, or
ps -ef|grep openv |awk '{print $2}'|xargs kill<CR>
```

Ensuring that all tape drives to be removed are sitting at BOT

The st (SCSI tape) driver will verify that all tape drives associated with the system board to be detached are either empty or populated with a tape cartridge sitting at BOT (beginning of tape) before it services a DDI_DETACH call generated by a DR-Detach operation. The following lines are executed on the marmaduke domain to make sure all tape devices associated with the system board to be detached are rewound to the BOT point and their status is later verified:

```
mt -f /dev/rmt/X rewind<CR>, where X=0-6
mt -f /dev/rmt/X status<CR>, where X=0-6
```

Unloading the Sun StorEdge Enterprise NetBackUp sg driver

The Sun StorEdge Enterprise NetBackUp application is bundled with it’s own sg (SCSI generic) driver to control the tape library robotic interface. The sg driver is a non-dr-compliant driver (it does not support the DDI_DETACH, DDI_ATTACH, DDI_SUSPEND, DDI_RESUME functions) and it must be manually unloaded before starting a DR-Detach operation. The following command line is executed on the marmaduke domain to unload the sg driver:

```
modinfo|grep "sg "|awk '{print $1}'|xargs /usr/sbin/modunload -i<CR>
```
Initiating DR-Detach on the SSP through Hostview

The DR-Detach operation is initiated by invoking the Hostview GUI from the SSP through the following command line:

```
/opt/SUNWssp/bin/hostview&<CR>
```

Once the Hostview GUI is displayed, the *marmaduke* domain is selected from the View menu (see FIGURE 2).

**Note** – If you are logging in remotely into the SSP, make sure the `DISPLAY` shell-environment variable is appropriately set and exported to display Hostview’s GUI locally. DR operations can be also be achieved through the `dr(1M)` command line interface monitor.
FIGURE 2  Hostview Domain Selection
Once the *marmaduke* domain is displayed through Hostview, system board 7 is selected by clicking on it and the Board->Detach function is invoked from the Configuration menu (see FIGURE 3).

**FIGURE 3**  Hostview DR-Detach
Once the DR-Detach function is invoked on marmaduke’s system board 7, the “Detach Board and Domain Selection” window is displayed to verify board and source-domain entries. Once the “execute” button is pressed on the “Detach Board and Domain Selection” window, the drview “Dynamic Reconfiguration” window is displayed (see FIGURE 4). The all button is then pressed to extract system status.

FIGURE 4  Hostview’s drview detach menu display
Once the all button is pressed from the drview window (FIGURE 4), the following windows are displayed:

- **DR OBP Configuration** (FIGURE 5): It provides OBP’s onboard device tree information as well as general system information. The “Memory Detach” entry must be “enabled” in order for the DR-Detach operation to succeed.

- **DR CPU Configuration** (FIGURE 5): It provides information about onboard CPU participation in processor sets (psrset(1M)) and bound threads (pbind(1M)). Processor sets and bound threads associated with onboard CPUs must be torn down before the DR-Detach operation.

- **DR Device Configuration** (FIGURE 5): It provides all onboard device instances and device drivers. If *Alternate Pathing* (AP) is being used it will display the alternate device instances the I/O stream will automatically switch to after the DR-Detach operation.

---

**FIGURE 5** drview OBP, CPU and Devices configuration
AP provides an additional software mechanism in support of DR to provide redundant disk and network controllers (and their respective physical links) to sustain continuous network and disk I/O when resources are removed (detached) from a DSD. While AP is supported on the Starfire server, it is not a prerequisite for DR.

Once the all button is pressed from the drview window (FIGURE 4), the following windows are displayed:

- **DR Memory Configuration (FIGURE 6):** Displays memory configuration and verifies that there is no memory interleaving between boards in the system (DR-Detach will fail with memory interleaving enabled). The "Memory Detach" entry must be "enabled" in order for the DR-Detach operation to succeed.

- **DR Unsafe Devices (FIGURE 6):** Displays devices sustained by non-dr-compliant drivers. The Sun StorEdge Enterprise NetBackUp sg devices would show up in this window if the sg driver had not been previously unloaded.

FIGURE 6    drview Memory, DR Unsafe Devices configuration
Once the \textit{drain} button is pressed from the \textit{druview} window (FIGURE 4), it will provide memory drain progress status through the DR Memory Configuration window. The memory drain is a process by which all dirty memory pages resident on the board to be detached get immediately flushed to the swap area (either memory or disk), while all non-pageable memory gets copied to the remaining system memory.

\textbf{Note} – All free pages associated with the system board to be detached are locked to prevent further use by applications.
Once the complete button is pressed from the `drview` window (FIGURE 7), it starts displaying progress until the board detach is successfully completed (FIGURE 8).

The `reconfig` button is used to invoke the “`drvconfig;devlinks;disks;ports;tapes`” command sequence to clean up devices that are permanently removed. The `reconfig` button should not be pressed since we are reattaching system board 7 at a later time and reusing all of its associated devices. Once the board detach is complete, the Hostview GUI displays the `marmaduke` DSD’s association with a single system board (FIGURE 9).

![Dynamic Reconfiguration](image)

**FIGURE 8** drview detach completion
FIGURE 9 Hostview’s display of the marmaduke DSD after the DR-detach operation
Bringing the Sun StorEdge Enterprise NetBackUp application back online:

For our lab exercise there is no sense in bringing the Sun StorEdge Enterprise NetBackUp application back on line since we have indeed removed all available resources with the DR-Detach operation. In the event that the Sun StorEdge Enterprise NetBackUp application needs to be restarted with reduced tape resources, the sg driver must be manually loaded and the Sun StorEdge Enterprise NetBackUp application restarted using the following command lines:

```
/usr/sbin/modload /kernel/drv/sg<CR>
/etc/rc0.d/S77netbackup<CR>
```

Methodology for Attaching the Tape Library

This section introduces the sequential actions involved in the attachment of system board 7, which houses the Tape Library, to the *flinstones* domain.

Bringing down the Sun StorEdge Enterprise NetBackUp application and unloading sg to acknowledge new tape resources:

If the Sun StorEdge Enterprise NetBackUp application is already running, it must be brought down since its bundled sg (scsi generic) driver, which controls the robotics interface, is non-dr-compliant. Non-dr-compliant drivers do not support the DDI_DETACH, DDI_ATTACH, DDI_SUSPEND, DDI_RESUME functions and can not acknowledge new device instances brought in through the DR-Attach operation. The following command lines are executed to bring down the Sun StorEdge Enterprise NetBackUp application and to unload the sg driver:

```
/etc/rc0.d/K77netbackup<CR>, and
modinfo|grep "sg " | awk '{print $1}' |xargs /usr/sbin/modunload -i<CR>
```
Initiating DR-Attach on the SSP through Hostview

The DR-Attach operation is initiated by invoking the Hostview GUI from the SSP through the following command line:

```
/opt/SUNWssp/bin/hostview&<CR>
```

Once the Hostview GUI is displayed, the flinstones domain is selected from the View menu (see FIGURE 2).

**Note** – If you are logging in remotely into the SSP, make sure the DISPLAY shell-environment variable is appropriately set and exported to display Hostview’s GUI locally. DR operations can be also be achieved through the `dr(1M)` command line interface monitor.
Once all the domains are displayed through Hostview, system board 7 is selected by clicking on it and the Board->Attach function is invoked from the Configuration menu (see FIGURE 10). Notice that system board 7 is not displaying a color border which means that it is not currently participating in any domain.

**FIGURE 10** Hostview DR-Attach
Once the DR-Attach function is invoked on system board 7, the “Attach Board and Domain Selection” window is displayed to verify board and source-domain entries. The \texttt{flinstones} domain name is entered in the “Target domain” window and the “execute” button is pressed to display drview’s “Dynamic Reconfiguration” window (see FIGURE 11).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{drviewAttachMenu.png}
\caption{Hostview’s drview attach menu display}
\end{figure}
Once the "init attach" button is pressed from the dview window (FIGURE 11), it will provide POST (Power On Self Tests) progress status through dview’s Dynamic Reconfiguration window (FIGURE 12).

Note – The Starfire server DR allows an increased test coverage on all system board components by temporarily increasing the POST level through modification of the SSP’s postrc(4) file.)
Once the complete button is pressed from the drview window (FIGURE 12), it starts displaying progress until the board detach is successfully completed (FIGURE 8).

The reconfig button is used to invoke the "drvconfig;devlinks;disks;ports;tapes" command sequence to clean up devices that are permanently removed. The reconfig button should not be pressed since we are reattaching system board 7 at a later time and reusing all of its associated devices. Once the board detach is complete, the Hostview GUI displays the marmaduke DSD’s association with a single system board (FIGURE 9).
Once the complete button is pressed from the *drview* window (FIGURE 12), it starts displaying progress until the board detach is successfully completed (FIGURE 13).

The *reconfig* button is used to invoke the "*drvconfig;devlinks;disks;ports;tapes*" command sequence to create appropriate device links for newly attached devices. The *reconfig* button should not be pressed since the device files associated with newly attached devices should already be in place. Once the board detach is complete, the Hostview GUI displays the *flinstones* DSD's association with a single system board (FIGURE 14).

![FIGURE 13: dview attachment completion](image-url)
FIGURE 14  Hostview’s display of the *flinstones* DSD after the DR-Attach operation

Restarting Sun StorEdge Enterprise NetBackUp and loading the *sg* driver:

Once the DR-Attach operation is complete, the *sg* driver needs to be reloaded and the Sun StorEdge Enterprise NetBackUp application restarted to acknowledge newly attached tape resources using the following command lines:

```
/usr/sbin/modload /kernel/drv/sg
/etc/rc0.d/S77netbackup
```
Marking attached tape drives available within Sun StorEdge Enterprise NetBackUp:

If the newly attached tape drives had been marked unavailable within the Sun StorEdge Enterprise NetBackUp application during the DR-Detach process, the following command line should be executed on all newly attached tape drives from the *flinstones* domain:

```
/usr/openv/volmgr/bin/vmoprcmd -up X<CR>
```

where X=0-6

---

**Conclusions**

This Sun BluePrint OnLine article has provided you with an application example of managing resources within the same Starfire platform. (For more information on resource management, please refer to the “Resource Management” BluePrint, Prentice Hall, ISBN #0-13-025855-5. This book is available through http://www.sun.com/books, amazon.com, fatbrain.com and Barnes & Noble bookstores.)

This Floating Tape Library exercise has helped you understand DR prerequisites as well as the steps required for executing DR operations in your datacenter. It should be emphasized that this solution should only be implemented if the latency introduced by TCP/IP during network backups/restores does not meet the restrictions established by existing backup/restore time windows.

Contact Enrique.Vargas@Sun.Com for additional details.

---

**Author’s Bio**

Enrique Vargas brings a wealth of large systems experience to Sun where he specializes in high-end UNIX offerings including the Sun Enterprise 10000 server and Sun Cluster™ Technologies. Enrique came to Sun from Amdahl Corporation where he focused on UTS® (mainframe UNIX) and high-end Solaris Systems.