NetWorker® and EDL® Capacity on Demand – Library Virtualization Challenges

EMC Proven® Professional Knowledge Sharing 2008

Erico Correia da Silva
Regional Technical Specialist
EMC Maintenance & Support Services – Brazil
silva_erico@emc.com

Author's Biography

Erico is a Regional Technical Specialist working in the EMC Maintenance & Support Services division since 2000, supporting the EMC BURA portfolio. He holds a master's degree in engineering, object oriented development and parallel processing. Erico has been working with IT support for 12 years and previously worked as a C/C++ developer for 6 years. He is a certified EMC Proven Professional with certifications in NetWorker and Network Management.

Table of Contents

Disclaimer: The views, processes or methodologies published in this compilation are those of the authors. They do not necessarily reflect EMC Corporation's views, processes, or methodologies.

Introduction

Virtualization is here to stay since it contributes to the effort to reduce TCOs (Total Cost of Ownership). Virtualization is even more relevant when talking about Backup, Recovery and Archive (BURA) solutions because of the advantages of backup to disk technologies combined with the ability to restore data after keeping tapes for several years. EMC Disk Library (EDL) is central in this scenario because of its ability to virtualize tape libraries commonly found in our data centers in a useful and user-friendly way. Similarly, NetWorker has an important role due to its flexibility and ability to be safely customized.

Capacity on Demand (COD) is one of EDL's most attractive features. It assures the best disk space usage by allocating disk space in chunks. However, you must customize the backup software to release tapes in a reasonable time to guarantee the best tape recycling, and therefore the best disk usage.

This article will introduce EMC Disk Library and discuss its integration with NetWorker. Neither the differences between the several EDL models nor their peculiarities are presented, since they are discussed in several documents available on Powerlink (http://powerlink.emc.com). This is not intended to be a best practice guide but, in conjunction with the references cited, you can acquire valuable information to implement, support, and administer BURA environments composed of both EDL and NetWorker. This allows you to take advantage of many features, especially Capacity on Demand.

NetWorker

EMC NetWorker is a central software solution for backup, restore, and archive. Companies worldwide use NetWorker as one of the primary tools to protect their data. IT professionals can easily configure and manage a backup environment so that the data can be remotely received then sent to central storage systems. The data can be sent to local storage devices (centralized control, distributed storage); and can be sent through the network or SAN. NetWorker supports a wide variety of operating systems, data bases and applications.

In a NetWorker environment, the administrator may centralize all catalogs and manage the backups from one site, reducing the need for specialized backup personnel in every site. With NetWorker, data is easily and quickly stored, and easily managed through the graphical interface and reports. In addition to the best speed to store data, NetWorker has also been elected as the fastest solution to retrieve data.

EMC NetWorker is an acknowledged industry leader in backup to disk technologies [1]. NetWorker is ahead of the competition for NAS NDMP backup to disk, data de-duplication (when integrated with EMC Avamar), Continuous Data Protection (when integrated with EMC RecoverPoint), making the most advanced backup technologies available to IT professionals without adding complexity to the environment.

Companies must retain backed-up data for years, meaning that new technologies frequently must coexist with old solutions. In this scenario, NetWorker is first with its easy integration with EMC Disk Libraries (EDL), taking advantage of all features and making management and support more comfortable.

EDL – EMC Disk Library

EMC Disk Library (EDL) is the industry leading virtual tape library in performance and scale [2]. Since 2004, EMC Disk Library has delivered the robustness of EMC storage systems to BURA environments. Based on the CLARiiON® CX3 Ultrascale series and Symmetrix® DMX series, integrating a powerful tape library emulation software, EDLs were widely accepted as the leader product in open systems tape virtualization [2]. IT professionals can deploy a new backup environment in hours, integrating the newest solutions with the older ones because of its simplicity and easy of use.

EDL Graphical User Interface (GUI) – Easy of Use

Users and administrators can use the disk library console to interact with the EDL, a graphical user interface (GUI) that can be easily installed in a regular desktop. Using this interface, the user can administer one or more EDLs, performing tasks such as adding and configuring clients, setting properties and managing the import/export of tapes [3].

An EDL delivers virtual tape libraries and/or drives in a Storage Area Network (SAN) environment. Some logical components are required:

- Virtual libraries
- Virtual tape drives
- Virtual volumes
- SAN clients
- Physical libraries (optional)

You can easily configure and/or manage components graphically by using the disk library console. The "Create Virtual Library" wizard illustrates the EDL GUI's ease of use. Start the wizard by right-clicking the "Virtual Tape Libraries" icon and selecting new (Figure 1).

The administrator is then asked about key configuration settings such as virtual tape library type and name, virtual tape drive type, auto archive or remote copy, barcode patterns, Capacity On Demand (COD) settings, etc. The administrator can then create virtual tapes for the virtual library just created.
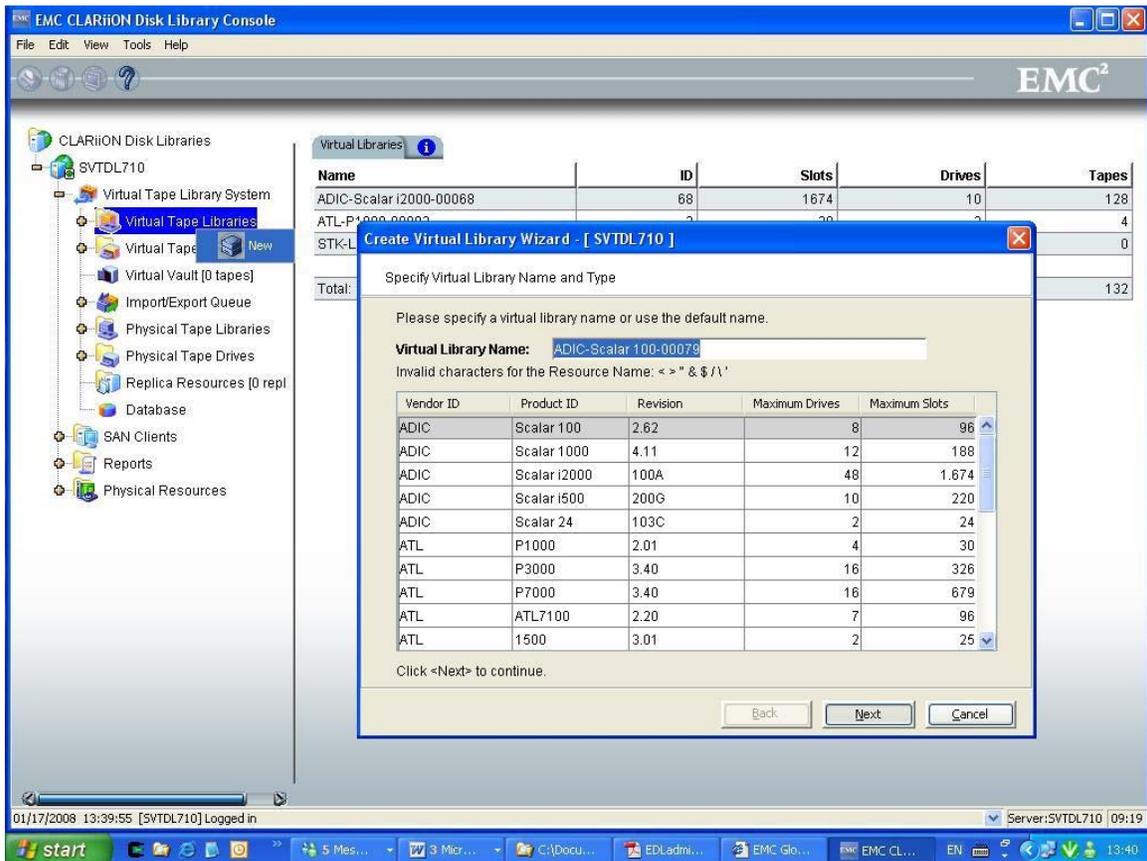

Figure 1 - Create Virtual Library wizard

Multi-System Management

The Disk Library console allows the administrator to group up to eight Disk Libraries in only one multinode group to facilitate the management of Virtual Libraries emulated by multiple EDLs. Using this capability the administrator can take advantage of the following features [4]:

- A group level Attention Required tab displaying a collective list of Attention Required events across all servers in the group
- A group level event log displaying a collective list of events across all servers in the group
- A group level X-ray that takes an X-ray of all servers in the group
- The ability to generate reports on all group members at once
- The ability to generate Group Reports to consolidate data collected from all the members of the group
- The ability to turn compression on/off on all servers in the group
- The ability to forward SNMP traps across all servers in the group
- Storage Monitoring across all servers in the group
- Auto-Save configurations across all members in the group
- Error forwarding across all members in the group

EDL Key Features

EDL has several features that improve the average value and usability of BURA environments, including and integrating with those built based on tape technologies. EDL introduced or enhanced some of the features in a BURA environment:

Consolidated media management

EDL has integration modules for both EMC NetWorker and Symantec Veritas NetBackup. Using these modules, the BURA administrator can use all functionalities of the backup solution with all benefits of the virtual library environment [2]. These modules are very important to implement tape cloning. Using the media management modules, the administrators can use the cloning capabilities of either NetWorker or NetBackup, being aware of the existence of a cloned tape and keeping track of its location [4].

Distance solutions

The mechanisms to move tapes offsite is one of the problems associated with regular physical tapes. Real tapes can easily be damaged or stolen during transportation, risking companies' data. Data can be sent from a source EDL to a remotely located target EDL with EMC Disk Library remote copy solutions. This can be done using regular TCP/IP connections, using bandwidth when the link between sites is relatively idle, and optimizing costs associated with data transfer.

Remote backup

Regardless of the two disk libraries, an EDL can be used as a remote target for backups, using standard connections such as LAN, SAN or WAN (Figure 2) [5].
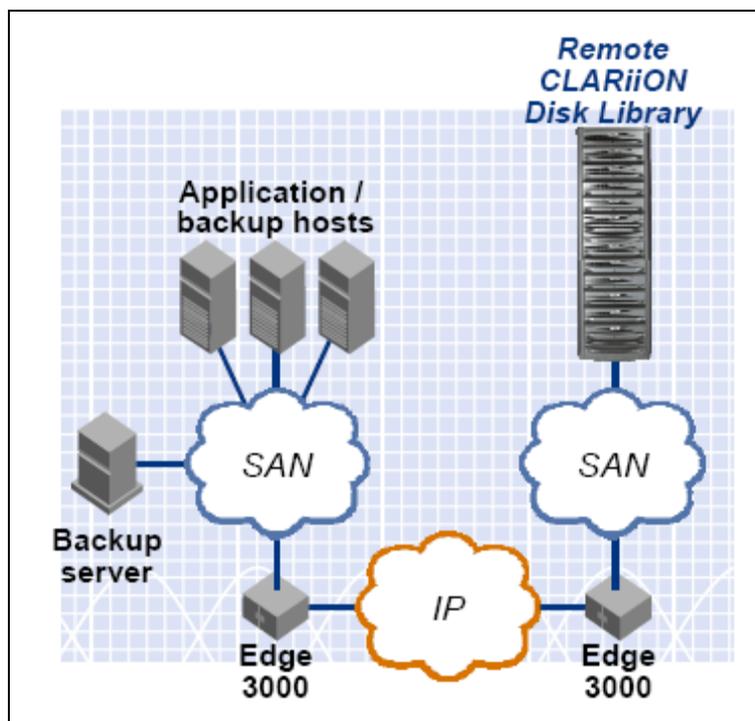
Figure 2 - EDL as a target for remote backups

Active Engine Failover

You can configure the Disk Library in an active-active cluster configuration. The active engine failover feature uses two servers to provide the ability to recover from a failure or a loss of connectivity to one of the servers. The two servers are peers and monitor each others' health by transferring control of the emulated resources to the remaining server.

Both servers can be active, each one controlling part of the virtual libraries. In case of a failure, one of them can control all virtual libraries, taking ownership of the unavailable virtual devices and restarting the emulation services. External actions may be required and will vary from one backup software program to another. In a failover event, backups will fail or suspend and a failure notification will be delivered to the backup operator.

Tape Shredding

Using the US Department of Defense (DoD) 5220.22-M algorithm, the Tape Shredding feature assures data erasure by overwriting disk data with multiple patterns of bits, making it difficult or even impossible to recover data from a virtual tape. This feature is very attractive because of the many regulatory security constraints.

Tape Stacking

This feature allows up to six virtual tapes to be moved to one 'real' tape. The relation between virtual and real tapes was one-to-one in the past. Exporting (moving from virtual to real) would be troublesome if the virtual library had several small tapes. This would cause a high real tape usage and consecutive loss of money. Resolve this issue by using consolidated media management software.

Optimized Disk Usage

EDL backend storage space can be either a CLARiiON or a DMX, bringing the reliability of EMC storage systems to the BURA environment. The data is RAID-protected and features like Capacity On Demand (COD) assure improved disk space utilization.

Backup software can use EDL's device side compression feature. This acts as the hardware compression of the real tape devices, minimizing the disk space required to store the data. The administrator has the option to enable client-side compression in several backup software programs. This will use backup clients' resources to compress the data before sending it to virtual tapes. EMC does not recommend using both client-side and device-side compression because it will not increase storage capacity but will decrease EDL performance [4].

You can easily manage disk usage through the GUI by using the disk library console or the command line interface.

COD – Capacity On Demand

More and more people are talking about "thin provisioning" in the storage backstage. For storage arrays, thin provisioning means creating logical disks, but only allocating physical disks when the host attached to the logical disk actually tries to write data to disk [6]. For an EDL, similar functionality is done in a per virtual tape basis, this feature is called "Capacity on Demand", or simply COD.

EDL Capacity on Demand is the feature where a virtual tape is delivered to the backup software as if it has its final total capacity. In reality, disk allocation is made in chunks, in an allocate-upon-write manner. The process follows [7] (Figure 3):

1. Tape is created initially allocating one disk chunk (5 GB by default), called extent
2. When the first extent is filled with backup data, EDL automatically and transparently allocates another extent
3. Data will continue to be written and more extents will be allocated as needed, until writes are complete or the final virtual tape capacity is reached
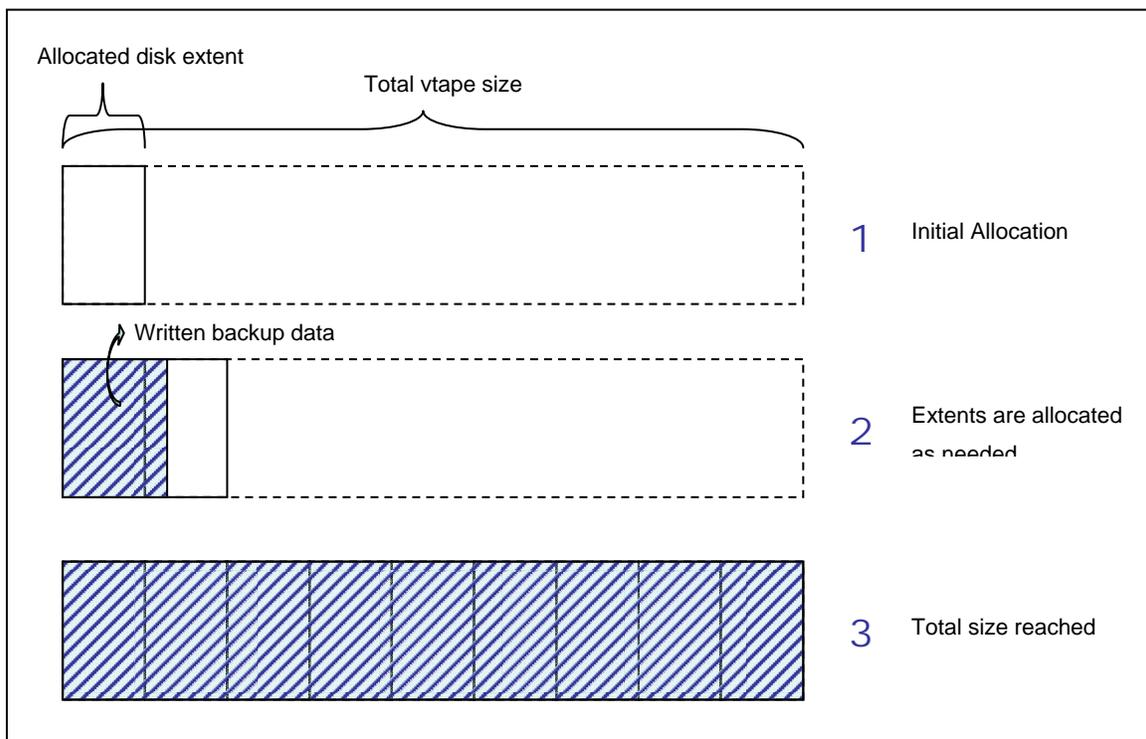

Figure 3 - Capacity On Demand disk allocation

Virtual tapes emulating large capacity tapes have extents larger than the default 5 GB to achieve maximum use of its total size. COD can be disabled so that tapes will have their total capacity allocated in disk space since their creation.

When the EDL has no disk space available (100% of disk usage), new extents will not be created and an end-of-tape is returned to the backup software generating tapes smaller than the initial total tape capacity presented to the software. This may be critical if virtual tapes are migrated to real tapes in a one-to-one relationship.

When to use COD

The EDL COD feature may not be necessary in NetWorker environments where tape pools' organization drives to optimal and fast tape utilization. In other words, COD may be unnecessary if the virtual tape is filled up quickly. The BURA administrator must keep in mind that COD is ideal for leaving disk space for other tapes that would be allocated in advance for a newly created virtual tape. COD is recommended if the backup organization generates virtual tapes that will not fill up in a backup cycle, for example for incremental backups of small files with less frequent modification.

BURA configuration and organization is complex. Optimizing backup settings to fill virtual tapes quickly is not trivial for enterprise environments. In the field, companies with business constraints that drive to BURA configurations with several pools, schedules, client groups, etc., make the COD feature very attractive to BURA administrators and architects.

COD and disk usage administration

From another perspective, COD demands more careful disk space administration. The EDL administrator must assure that disk usage will not grow faster than originally sized for the EDL, especially because of the need to have disk space for extending virtual tapes.

Disk space and virtual tape creation must evaluate and respect the three following points:

- **Number of available virtual tapes:** As in any BURA environment without tape virtualization, the administrator must analyze and control the tape consumption. In an EDL environment, tapes will not be purchased but created. Disk space must be reserved to create new virtual tapes when needed.
- **Disk space for COD virtual tapes extension:** In an EDL environment with COD enabled, part of the disk space must be kept available for possible virtual tapes extensions.
- **Tape recycling:** Tapes are commonly used, released and reused in every BURA environment. The recycling of tapes is not less important in an EDL environment because expanding the disk space requires much more time and work than simply buying and adding more tape cartridges.

COD releases the disk space whenever a tape is relabeled. Most of the BURA administrators would think that disk space is released whenever all backups in a virtual tape have expired. This is conceptually acceptable, but practically unfeasible due to the need for a very deep integration with the backup software. Such an integration would limit the number of backup software programs capable of taking advantage of the COD feature.

In summary, if disk usage gets to a critical point, the EDL administrator in a COD enabled environment should re-label the recyclable tapes. Of course, this task can be automated through a simple script for most backup systems.

Migrating data to/from real tapes

When talking about disk libraries, BURA administrators have to be prepared to send backup to other media in order to keep the data for longer periods of time. Should the data be kept only for historical purposes due to regulatory constraints, the administrator may prefer to send the data to an alternative media such as real tape, and then send it offsite. The disk space can then be used for other backups that may need to be restored faster.

If the data is already backed up to a virtual tape and must then be sent over to real tape, the EDL may be attached to a real tape library. You can then 'export' tapes to real tapes.

Where to connect the real library

In an EDL environment, the real tape library is accessible throughout several connection types. Each type of connection has its advantages and disadvantages. We will discuss some of them.

Disk library directly connected to the EDL

In this configuration, the SAN infrastructure is not used, saving SAN resources such as connection ports, bandwidth, etc. This also avoids using NetWorker server CPU cycles, giving control of the migration tasks to the EDL. The limited available bandwidth is a disadvantage because the consolidated media management (the integrated storage node) is not in place.

The EDL emulation software does copies with EDL server hardware. The number of fibre ports or SCSI ports available for the real library is limited, fixing the maximum rate for exporting/importing tapes [8][9].
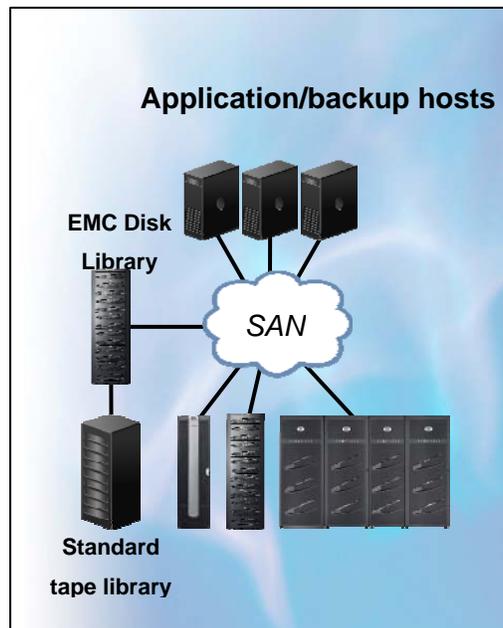
Figure 4 - Disk library directly connected to the EDL

Here are the advantages of EDL consolidated media management:

- Clones can be done under NetWorker control
- Possible clones at the save set level
- Real tape type must not match the virtual tape type
- Up to 6 tapes can be exported to the same real tape (tape stacking)

Disk library available over the SAN

The best approach is to attach the real library to a Storage Node using the SAN if you have a large amount of data being sent to real tapes on a daily basis. Backups are done to virtual tapes and NetWorker is used for cloning the data from virtual tapes to real tapes, out of the backup window [8].
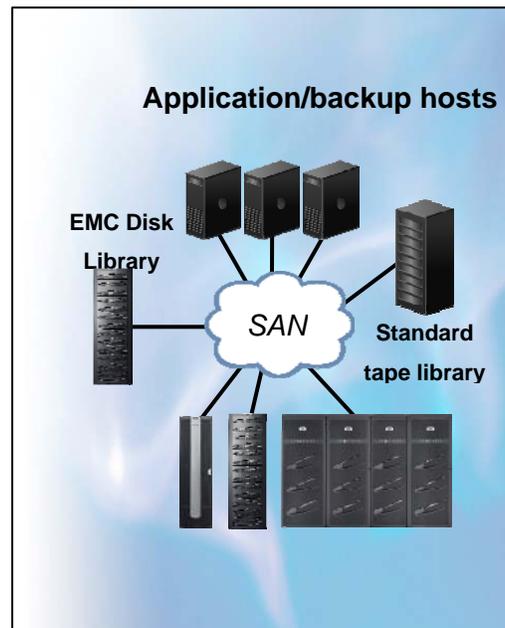
Figure 5 - Disk library available through the SAN

The BURA architect has to consider that the same real library must not be shared between the EDL emulation software and the NetWorker Storage Node (no matter if it is the consolidated media management). In [9] the authors emphasize this with the following example:

"If two physical tape libraries are directly connected to the EDL (each on a separate FC port), the following configurations are possible:

- Both libraries are dedicated to the EDL software and no library is dedicated to the storage node software
- Both libraries are dedicated to the storage node software and none are dedicated to the EDL software
- One library is dedicated to the EDL software and one library is dedicated to the storage node software."

EDL export/import tape x NetWorker clone

Both NetWorker and EDL can migrate data from virtual tapes to real tapes. Each method has its advantages and disadvantages. For example, for a large amount of clones (more than 2 to 3 TB per night [9]), the NetWorker clone process is recommended because of configuration bottlenecks where exports are done purely by the EDL. The administrator must use caution to avoid resource contentions when the NetWorker clone process is used. We will briefly describe the EDL and NetWorker processes.

EDL export/import

Data may be migrated to/from real tapes using only the EDL software when the real library is attached to the EDL. This is useful when the BURA architect wants to avoid using another NetWorker Storage Node license to create real tapes, especially when the amount of data is not large. This can be done automatically or manually. Either way, the real and the virtual tape must have the same barcode and be the same type.

EDL moves the virtual tape from the virtual library to the virtual vault and then copies the data to the real tape. The administrator can configure a grace period during which to keep the emulated tape. There are two export options: copy or move. If the data is supposed to be moved, the virtual tape in the virtual vault will be removed at the end of the process. Since the tape must first be sent to the virtual vault, backup software cannot use the tape until it is injected again in a library (the NetWorker real tape can also be injected in any NetWorker controlled real library).

To read the real tape, choose one of the following methods:

- **Copy import:** data is completely copied from the real tape to a virtual tape and then it can be injected to a virtual library for read purposes. Depending on the drive technology, this may require a long time before being able to read the data;
- **Direct-access import:** the EDL creates a direct path between the real tape drive and the backup software, avoiding the need to create a virtual tape first. This is the best option for a single and quick restore.

"**Recycle import**" is a third import method to reuse the tape. It reads only a small amount of data needed for the backup software to determine whether the tape can be recycled or not.

NetWorker clone

The NetWorker's software clone feature can be used either through the EDL integrated storage node or through a regular production storage node. The real library may be connected over the SAN. NetWorker clones can be done automatically or manually. Two storage devices must be available for NetWorker clone operations. A virtual and a real tape device will be used for the clone operation to/from real tapes. The BURA architect must take this into account when defining the number of virtual tape drives he will create. Also, the destination volume must belong to a clone pool [10].

The BURA administrator may decide to clone during the regular backup window, but caution must be taken to avoid nsrmmd resource contentions [10]. For example, if a clone process is started for a saveset in a tape that is being used for backup purposes, the clone operation will stop, waiting for this tape to be released. Contentions will also affect backup operations if the tape is being used for clone purposes while the backup is started. The BURA architect must avoid such situations when creating the backup configuration and the clone script. Automatic clone operations are even more sensitive to nsrmmd resource contentions.

There is an alternative to avoid nsrmmd resource contentions, and therefore make the usage of the automatic clone feature and make clones during the backup window possible. This alternative consists of respecting a one-to-one relation when associating NetWorker client groups (nsr_group) to a media pool (nsr_pool). That is, any NSR pool resource must be associated to one and only one NSR group. In the field, this was shown to be a very effective solution but it dramatically increased the complexity involved when creating the backup configuration, especially for environments with several different backup schedules. Such one-to-one relationships drive to a configuration with a numerous media pool list. BURA architects know that the number of media pools must to be minimal to achieve optimal tape usage. In an EDL environment, if tapes do not get full, disk space is being wasted. It is recommended that the administrator enable COD.

EDL as a tiered storage

The BURA architect may think of an EMC Disk Library as a first tier from where data can be restored quickly. Following this idea, backups would be cloned to real tapes as a second tier. With this approach, we can list the following items for observation:

- **EDL x NetWorker staging feature:** From a NetWorker perspective, the Disk Library is like a regular real tape library. In this sense, the NetWorker stage feature cannot remove the save sets right after they have been moved to the second tier (real tape). For an EDL, disk space is released when all save sets of a virtual tape expire and it is overwritten. Therefore, this feature is not recommended [9].
- **Disk space management:** If the EDL is a first tier for backups, the administrator must keep in mind that disk usage varies according to the way virtual tapes are recycled. If one small save set has not expired yet, the disk space of the virtual tape will not be released even if all other save sets have been expired.
- **NetWorker clone retention:** The Administrator may configure different retention policies for the multiple clones of a save set, making the save sets on real tapes available for a longer period. However, if the original is the first tier (the EDL), as soon as it expires, the browse portion of the save set will also expire. It will not be possible for users to browse the cloned save set for restore.

NetWorker clone and browse retention

When data is cloned from virtual to real tapes, the cloned save sets can have an expiration date for the original and another for the clones. This is useful for keeping data for days or weeks on the EDL, and keeping data on real tapes for months and years. The restore procedure with this configuration is not the same if the original save set expires.

NetWorker allows different retention periods for originals and clones, but both shares the same browse policy. Browse capabilities cannot be kept longer than the original save set. After the expiration of the browse information, save sets have to be scanned or browse indexes must be restored.

The BURA architect can use the strategy outlined below to avoid scanning the tapes:

1. **Configure backups to have the same retention for original, clone and browse info**: Both original and clones will have the longest retention. The browse policy will have same retention as the clone; as long as there is a restorable save set, its restore can be done with browse capabilities because the browse information (CFI) will not be discarded first;

2. **Maintain information on how long to keep the original save set**: This can be kept in the NSR pool name or in the comment field of the NSR pool resource. For example if the original have to be kept for a week and the clone for a year, all policies will be set for a year of retention but the pool name will have a suffix such as "Keep 7 days".

3. **Maintain a script to manually recycle virtual tapes with original save sets**: This script would use NetWorker report capabilities (mminfo) to query the catalog and search for tapes that have been kept in the EDL for the amount of time specified in their pool name, containing only successfully cloned save sets, then would mark such tapes as "eligible for recycling". Table 1 outlines the algorithm of the script.

Table 1 - Possible logical steps for a script for manually recycling tapes

Script logical step	Main commands

1. List all pool names searching for the Keep keyword.	`nsradmin -i - > /tmp/pools.txt << EOF`
	`show name`
	`print type: NSR pool`
	`quit`
	`EOF`
	`grep "Keep" /tmp/pools.txt`

2. Loop for all pools found	`Example: pool="Historical Keep 7 days"`
2.a) extract the desired retention for originals from the pool name	`Retention for originals = 7 days`

2.b) Query for tapes for this pool that have save sets from before the amount of time specified in the pool name (vtl name = "EDL")	`mminfo -a -r volume -q`
	`"location=EDL,sscreate<7 days ago,`
	`pool=Historical Keep 7 days" > tapes2b`

2.c) Query for tapes for this pool that have save sets generated after the amount of time specified in the pool name	`mminfo -a -r volume -q`
	`"location=EDL,sscreate>7 days ago,`
	`pool=Historical Keep 7 days" > tapes2c`

2.d) Generate another list with tapes listed in step 2.b that were not listed in step 2.c	`fgrep -v -f tapes2c tapes2b > tapes2d`
For each tape listed in step 2.d	`For example EDL000`

Query for save sets that were not cloned successfully	`mminfo -a -q 'volume=EDL000,`
	`ssrecycle=false,copies<2,`
	`incomplete=false' -r`
	`'volume,ssid,client' > savesetstoclone`

If there is no save set that was not successfully cloned, mark tape as eligible for recycling	`if [! -s savesetstoclone]`
	`then`
	`    nsrmm -y -o notreadonly EDL000`
	`    nsrmm -y -o recyclable EDL000`
	`fi`

Next tape listed in step 2.d	
Next pool with the keyword	

Table 1 includes an algorithm to identify and mark tapes for recycling. The main purpose of this table is to illustrate how the mminfo command can be used to take advantage of NetWorker intelligence to identify tapes and save sets that have been kept in the EDL for the desired amount of time. For example, mminfo query settings such as "copies<2" and "incomplete=false" are very important to safely identify save set without a clone ("copies<2") and/or with a successful status ("incomplete=false").

Using the strategy above, the EDL disk space used for virtual tapes with original save sets would be released without destroying the browse information for the save set. This would make it possible to do browseable restores for NetWorker save sets present only in real tapes.

Long term and short term, where to store the data

EDL is ideal for environments where backups are kept for shorter periods, where virtual tapes may be recycled on a regular basis, freeing up disk space for new backups. If part of the backups must be kept long term, administrators and architects must deal with the complexity involved in sending data to real tapes. BURA professionals have to consider the following trade-offs when deciding where to store the data:

- Virtual tapes (disk space) should be used for long term backups only if the restore frequency justifies the data being close at hand;
- Even short term backups may be sent straight to real tapes if the EDL has limited capacity and the likelihood of needing a restore is very low. In this case, the real tapes' consumption and their cost must be considered;
- For long term backups, if it is likely that a restore may occur, it is a good idea to send the data to the EDL first then clone it to real tape. If save set restores (no browse) of such backups is impracticable, the BURA architect must decide whether to store it first in a real tape (slower backup) and then clone it to the EDL, or deal with the complexity involved in keeping the browse capabilities when releasing the virtual tape with the original save set. Depending on the save set characteristics, re-creating the entries in the client file index (thus recovering the browse capabilities) may be a quick step before a restore [10];
- The BURA administrator must identify and separate backups needed for disaster recovery purposes from those needed to keep historical data. Backups with historical data generally have a very low likelihood of being recovered. These are good candidates for sending straight to real tapes. EDL is still a good first place to store this data if a faster backup is required, though.

In fact, there is no rule for deciding where to store the data. BURA architects have to make the decisions based on the speed needed for backup and restores, the real tape consumption, the amount of disk space available in the EDL, and the need for browseable restores.

EDL disk space optimization and monitoring

The EDL disk organization is a compound of RAID 6 groups (4+2 or 8+2 combinations, with disks reserved as hot spares) [11], see Figure 6. RAID 6 implements two independent sets of parity information for each parcel of data. It was introduced in the last EDL models to improve data availability. With RAID 6, media errors are tolerated even while rebuilding after a single drive failure.

RAID 6 must to be deployed if you plan to take advantage of the new 1 TB drives. RAID five is still an option for customers where data availability is not the priority and 1 TB drives will not be used.
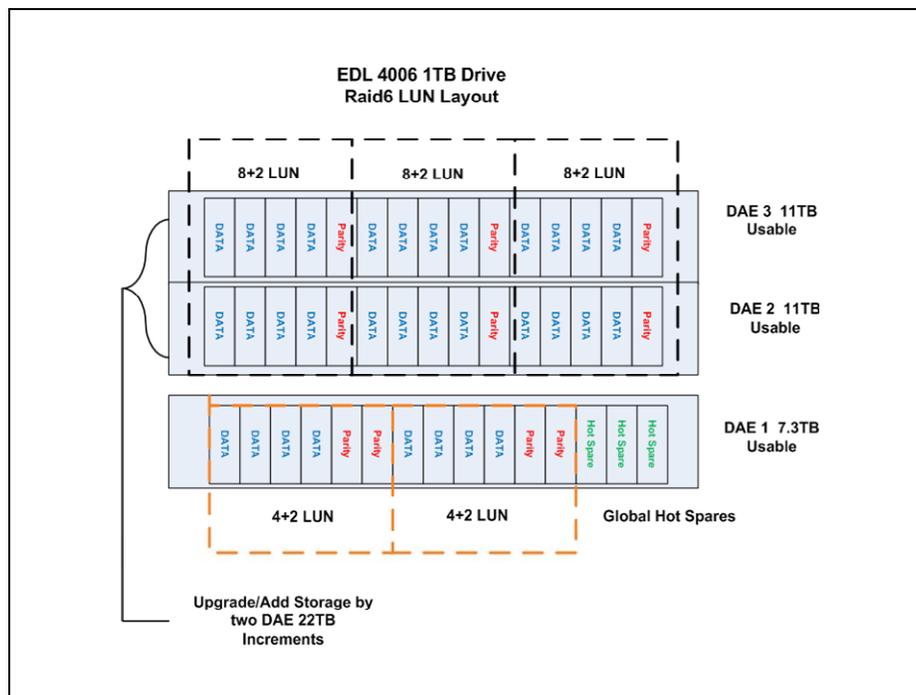
Figure 6 - EDL 4006 1TB drive RAID 6 LUN layout

Disk usage must be monitored closely in an EDL environment. The administrator can use the following tools [12]:

- Check storage disk leds for non green leds;
- Use the ConnectEMC error forwarding service, enabling EMC to automatically identify potential and existing problems online;
- For EDL based on EMC CLARiiON systems, use Navisphere® manager to monitor the storage system;
- For EDL based on Symmetrix, health check is done automatically and any possible maintenance is made by an EMC Maintenance and Support Services representative;
- Monitor the Disk Library console event log for storage specific error/warning messages;
- For EDL based on EMC CLARiiON systems, use Navisphere manager to enable automatic alerting; for EDL based on Symmetrix, use the Solutions Enabler appropriate tool. SNMP traps (that can be sent straight to the EDL engine), email notification and paging are the supported alerting methods;
- Configure the email alerts feature to send errors to your chosen mail boxes.

In conjunction with the COD functionality, EDL's newest version has a feature called "Proactive Service". When the LUN is full, COD will pick up disk space for an extent from the next available LUN. The "Proactive Service" will then pretend the LUN is full whenever a rebuild or equalize is in place. This will make the virtual tape to go to another LUN. This feature avoids writing backup data to a potentially bad LUN.

NetWorker, EDL and large media types

The administrator must ensure that sure tapes will be reused in an optimal fashion in EDL environments, especially where COD is in place. The BURA administrator has to monitor and take some action to force another piece of media to be used when big media types such as LTO3 are being used, if NetWorker is the backup software.

If small save sets with short term retention periods are mostly to be used, NetWorker will probably take some time to recycle a media, keeping disk space allocated longer than really needed. For example, let's assume a daily full backup with 10GB and expiration of a week. By the end of the 8th day, the first backup will be expired but the backup of the 8th day will still use the same piece of virtual tape as used in the first day. Depending on the backup configuration, this will continue to happen for all subsequent backups while the media is inside the virtual library. This disk space (that has been kept allocated even after backups' expiration) can be very valuable in a heavily used EDL, especially when COD is enabled because the disk space is not completely allocated in advance.

It is best to avoid using the same media for a long period of time even in real tape environments, although for real libraries the ejection and injection of tapes is very frequent and this problem is not usually noticed.

One option is to use NetWorker to mark tapes as read only on a regular basis. For our previously given example, if the tape was marked as read only after the 7th backup, the 8th backup would have gone to a new piece of virtual tape, and the disk space allocated for these first seven backups would be released as soon as the 7th backup expired. Figure 7 shows a very small script that can be used for marking NetWorker tapes as read only.

```
##    Find all tapes in library EDL, from pool "Historical Keep 7 days"
## with save sets created before 7 days ago and mark them as read only.

for a in `mminfo -a -r volume -q "location=EDL,sscreate<7 days ago,
pool=Historical Keep 7 days"`
do
      nsrmm -y -o readonly $a
done
```

Figure 7 - Script for marking tapes as write protected

Performance

Performance is more and more critical due to the minimized backup windows and aggressive Recovery Point Objectives (RPOs) and Recovery Time Objectives (RTOs) in enterprise BURA environments. EDL associated with NetWorker can deliver sensible improvement in backup and restore performance (see Figure 8) [13]. Performance is no longer limited due to sequential access tapes and robotics technologies using disk libraries.
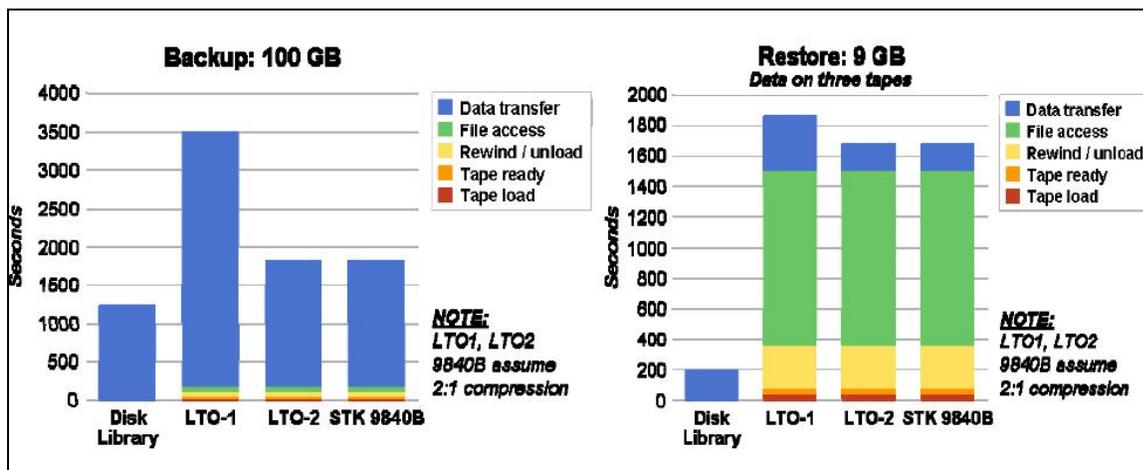

Figure 8 - EDL x other technologies - Performance comparisons

EMC Disk Libraries are, in most cases, far faster than real tapes. Disk Libraries still have the advantage of not having to wait robotic arms movement even for the fastest tape drive. The advantages are especially noticeable for restores, where files are spread on several different tapes, causing mechanical seek delays [13].

For single stream writing, EDL 4x06 can reach a 150MB/s throughput. Such performance will be reached if the BURA infrastructure (Host, Network, SAN) can leverage this speed without creating different bottlenecks [13]. Using a multiple stream, multiple virtual drives configuration, EDL 4x06 I/O speed can reach up to 2400 MB/s (EDL 4406 can go over 2400 MB/s). All EDL front-end target ports should be used and properly zoned to optimize the I/O distribution over the virtual devices.

Virtual Tape Creation

In order to optimize performance, when virtual tapes are created, they are distributed over the available LUNs on a round-robin basis [9][13]. This guarantees predictable performance of all the EDL hardware, because I/O is distributed as equally as possible over the back-end ports and other resources. Performance predictability is achievable for environments without COD enabled, and where virtual tapes are completely allocated at the time of virtual library creation.

If the environment respects these two constraints, all tapes' disk space mapping is made following the EDL tape creation algorithm. Preventing administrators from creating virtual tapes will cost a heavy price; preventing use of the COD feature may be a more difficult decision. Field experience shows that if most of the virtual tapes are created at the very beginning, the performance impacts for tapes created afterwards are not reasonable, even when COD is on.

Designing NetWorker for Performance

NetWorker flexibility makes it possible to design an environment that completely fulfils business requirements. During the design phase, the BURA architect has to follow these steps [14]:

- **Define technical and business questions**: understand business requirements, define RPO and RTO, and identify available infrastructure and constraints.
- **Identify backup system bottlenecks**: identify any possible front-end bottleneck (Host CPU, file size, application nature); any central bottleneck (NetWorker server and/or Storage Node resources) and any back-end bottleneck (Storage Device in use).
- **Eliminate backup system bottlenecks**: Once bottlenecks have been identified, they must be addressed. In regular backup systems (using real tape devices), the BURA architect designs the environment to make the storage device limit the only bottleneck and making the tape device perform the best. For an EDL environment, this approach would drive the architect to configure only a few virtual tape devices, based on the EDL front-end performance limits. Due to the virtual nature of an EDL, a better approach would be to design the environment based on the EDL front-end limits, no matter the number of virtual tape devices that can be used.

  You can refer to the article "NetWorker Best Practices – Designing for Performance", where Steinberg discuss the above items in depth [14].

Conclusion

New technologies are developed to attend to business requirements as the backup window shrinks. However, BURA is the IT area where moving to new technology is not easy. Old technologies must be kept while data must be available for restores. In this scenario, tape library virtualization fits very well by bringing the reliability and robustness of advanced disk based techniques now available. With EDL systems, the virtual tape usage is carefully dimensioned and easily monitored, lowering the overall TCO of the BURA environment. Many of the ideas mentioned resulted from intensive implementation efforts made by EMC Technology Solutions Group in conjunction with the EMC Maintenance and Support Services team and the EMC NetWorker Critical Response engineers. Therefore, we highly recommend that BURA professionals fully understand the characteristics outlined here, and apply what is indicated by the best practices guides cited as references.

References

[1] **Why EMC NetWorker**, EMC Corporation, October 2007

[2] **Why EMC Disk Library**, EMC Corporation, October 2007

[3] **EMC CLARiiON Disk Library Administrator's Guide v3.0**, EMC Corporation, November 2006

[4] **EMC CLARiiON Disk Library Online Help v3.1**, EMC Corporation, January 2008

[5] **CLARiiON Disk Library Implementation and Management training – student resource guide**, EMC Corporation, July 2006

[6] **What is thin provisioning? Which applications benefit most?**, Howard Marks, in http://searchstoragechannel.techtarget.com/expert/KnowledgebaseAnswer/0,289625,sid98_gci1282191_tax304818,00.html?bucket=ETA&topic=304818, January 2008

[7] **Knowledge Base Article emc134272**, EMC Corporation, in powerlink.emc.com, May 2006

[8] **EMC Disk Library Best Practices with EMC NetWorker**, EMC Corporation, December 2006

[9] **EMC Disk Library with NetWorker, Best Practices Planning**, EMC Corporation, January 2008

[10] **EMC® NetWorker Release 7.4 Service Pack 1 Multiplatform Version Administration Guide**, EMC Corporation, December 2007

[11] **EMC Disk Library (EDL) 3.1 Technical PreSales**, EMC Corporation, October 2007

[12] **EMC Disk Library Version 3.1 Release Notes**, EMC Corporation, November 2007

[13] **EMC Disk Library DL4106, DL4206 and DL4406 Version 3.1 – Best Practices and Planning**, EMC Corporation, January 2008

[14] **NetWorker® Best Practices – Designing for Performance**, Matt Steinberg, August 2007