

# Keys to Successfully Architecting your DSI9000 Virtual Tape Library

---

By Chris Johnson  
Dynamic Solutions International



**dsi** dynamic  
solutions  
international

July 2009

# Section 1

---

## Executive Summary

Over the last twenty years the problem of data archiving has grown exponentially and has been compounded as internal and external policies add complexity. IT departments are faced with the problem of how to manage and backup all of the data in a shrinking window. To help combat the challenges presented by the changing backup requirement, Dynamic Solutions International (DSI) recently announced the DSI9682 Standard edition and DSI9882 Enterprise edition Virtual Tape Library (VTL) appliances to compliment an expansive line up of VTL systems. The new systems are part of the 4th generation VTL systems released by DSI. The new appliances add significant breadth to the DSI VTL product portfolio. With so many VTL choices from DSI, which unit is the right choice for your environment?

DSI provides two architecture lines for its DSI9000 VTL systems, each complementing both Storage Area Networks (SAN) and Network Attached Storage (NAS) implementations.

The first DSI9000 VTL architecture is based on a straightforward appliance system. The appliance-based VTL products are arranged in three families; Value-line, Standard edition and Enterprise edition systems. Each appliance is a self contained system consisting of the VTL application and storage. The appliance systems start at 3TB and, with expansion, the largest system supports 144TB of native storage capacity.

The second DSI9000 VTL architecture is based on a head unit or cluster system. The head unit-based VTL products are arranged in two families; Standard edition and Enterprise edition systems. Each VTL head unit is designed for optimum expansion and flexibility. The head units provide for the VTL application only and do not contain any storage beyond that which is required to house the application. The head unit systems will seamlessly support hundreds of terabytes of external storage.

In order to make the right choice for your environment you need to analyze both your current and future requirements. You must apply your knowledge regarding a physical tape environment while keeping in mind that a VTL system is not quite tape or disk. This white paper will discuss key considerations in moving to virtual tape backup and will propose the key questions to consider when evaluating the different DSI VTL systems.

## Key Questions

1. How many physical tapes are being created today and what is the expected growth?
2. How much data is represented by the current backup and what is the expected growth?
3. Will the virtual tape system be required to keep your backup history on disk for a day, a week, a month, a year or longer?
4. What is the current window for backup and is the requirement being met by physical tape drives?
5. Is data replication a requirement?
6. Is high availability a requirement?

### How many physical tapes are being created today?

The first question looks easy, but as you apply it to a new technology, the answer will change the size of the VTL and the dynamics of other features provided by a virtual tape system such as data replication and off-site valuating. When switching to a DSI9000, many tape emulation options are available. By applying some detective work to the current physical backup, an emulation should be selected that will maximize VTL storage and simplify replication and archive.

The DSI9000 VTL systems all employ capacity on demand technology. Virtual tape volumes are created in pools and provided with attributes that set the initial starting size and allowable expansion; these attributes do not have to match what is available in the physical world. You could have an LTO4 tape that is limited to 200GB, while a physical LTO4 tape will hold up to 1.6TB of compressed data.

The default initial starting size of a virtual volume is 1GB. The largest DSI9000 systems can support up to 64,000 virtual volumes in a single unit. Initializing 64,000 cartridges at the base size would require 64TB of storage before you ever write your first volume.

If you are currently using an older tape technology such 36-track, DLT, or 9840 you may be creating many hundreds of tapes. Some of the tapes might be created on an interval basis such that the physical media may be only partially filled, while other tapes may be filled and rolling to continuation volumes. In most situations, simply applying a one-for-one rule will use unnecessary storage and add complexity to the backup and archive process. Setting capacity on demand features to match your environment will allow you to maximize your storage utilization.

## How much data is represented by the current backup?

People automatically look at the number of tapes being created and multiply that by the maximum capacity of the tape. If you are creating one hundred DLT8000 cartridges today with a capacity of 80GB per cartridge, you must be backing up 8TB of data, right? Unfortunately, that is not the case. Getting back to the discovery process for the previous question, many of the volumes may only be partially full. It is important to know the size of the data sets that are creating the backup. It is also vital to understand how each of these data sets creates backup media. Are the tapes incremental or full and how often are they created?

Outgrowing your storage is an ongoing, persistent, nagging issue. First, there is normal data growth that most companies experience, and then there is the unexpected growth presented by a new line of business or service, or perhaps an acquisition. This makes buying the correct amount of storage an educated guessing game. Fortunately, the DSI9000 systems scale well. However, if you chose the wrong starting point you may find that you cannot avoid a fork lift upgrade. If you implement a second VTL, you may find that you have to start subdividing and reconfiguring your backup jobs for two independent systems.

A second virtual tape system may not handle your data growth well, and leads you to a system that is in many ways similar to having multiple physical tape libraries. And sneaker net is difficult with a virtual tape cartridge. To avoid these problems, the DSI9000 virtual tape library systems will dynamically expand the RAID. The DSI9000 within each product family will auto-discover and virtualize any newly added storage into a single pool. For the VTL appliances, a storage increase involves bringing the unit down and adding drives. For the head units, storage can be added on the fly without disruption to your environment.

## Will the virtual tape system be required to keep your backup history on disk for a day, a week, a month a year or longer?

The simplest of backup schemes; grandfather, father, son can store three complete versions of the same data. If you plan to keep two weeks or a months worth of full backups on virtual tape, then your storage requirements may blow through your wallet in a hurry. Each time you store a full copy of your backup on secondary storage, it consumes an equivalent amount of space to your primary storage. Consider the following scenario:

- You have 1 TB of primary data on which you do full backups daily
- You plan to keep these full backups of this data on disk for two weeks

If you take the 1 TB of primary data times the 14 days of full backups you will be storing approximately 14 TB (raw without compression) of data. When you consider using either RAID 5 or RAID 6 architecture that also adds storage overhead, you may spend more money on your secondary arrays than you did on your primary storage.

A VTL system that incorporates data de-duplication technology will only store the block (or byte) level changed data. So now if you have three days where the data did not change at all, only one copy would be stored. And, if you have 14 days where the data only changes marginally, only the changed data will be stored. This can significantly reduce your storage requirements. A 10:1 de-duplication ratio may drop the 14 TB full two week storage requirement to a manageable 1.4 TB, dramatically reducing the amount of disk space required, and the overall cost of the solution.

## **What is the current window for backup and is the requirement being met by physical tape drives?**

In the physical world, tape drives are expensive to buy and maintain, but it is necessary to have enough units to complete backups within the appropriate window. If the window shrinks or a new application server is added and the data expands, more drives are added to complete the task – too few drives and you cannot meet the window, and too many drives is a significant cost issue. With a physical tape library this is a call to the vendor. With a VTL, you simply add another virtual drive or two to the configuration.

Current shipping physical tape drives like LTO4 are fast, but come at a price. Not just for the drive, but also the infrastructure required to support it. The new drives need to be fed data in order to maintain a sustained throughput. Any starting and stopping of the data stream will significantly degrade the performance and the life of the physical drive. HP maintains that a single LTO4 drive per bus rule must be adhered to. This is an expensive proposition as HBA's and fibre channel switch ports are not inexpensive. In a virtual tape library, the same rules do not apply to an emulated LTO4 drive as they do to their physical counterpart, and virtual tape libraries provide for hundreds of drives to be defined. More virtual drives can be added per port as the data throughput does not have the impact of sustained performance.

In an environment with ten DLT8000 drives, the system has a combined throughput of 120MB/sec compressed. The smallest DSI9000 appliance has a sustained performance rating of 400MB/sec. A single VTL can replace the performance of many physical drives and have a significant positive impact on backup window requirements.

## **Is data replication a requirement?**

Today, many industries have legislated requirements for securing a copy of data away from the primary location, and companies can draw significant penalties if they do not do so. Transmission of data will incur costs for the needed bandwidth, so again some consideration should be given to how much, or what classification of data is to be replicated. With a replicating virtual tape system, you can also backup primary data located at your second site and have it sent back to the first site as a secondary copy. This results in protection for your critical business data at both sites.

A VTL system that is using data de-duplication for storage efficiency can also use that technique for replication. Now, the system is only required to send the change data across the WAN. The DR site can reassemble the pieces and re-construct the most recent full in the event of a request for restore. This technique allows you to keep all of your retention on your disk-based backup system. Doing this reduces or eliminates the issues associated with tape and provides for much faster communication and better utilization of resources.

With a replicating virtual tape system, you can also backup primary data located at your second site and have it sent back to the first site as a secondary copy. This results in protection for your critical business data at both sites.

Replication may not accomplish all of your data movement and retention requirements. Your best bet for long term archive is still tape. If you need to create an archive volume that must be kept for a year or more, you don't want to have to bring this data back through the host to create a physical volume. The VTL must provide for the ability to make a copy of the backup data to a physical tape. This is referred to as disk-to-disk-to-tape (D2D2T). Because the physical volume is needed for long term archive and will most likely be stored offsite, the virtual tape system needs to be able to encrypt the data as well.

## **Is high availability a requirement?**

A tape library system that is being used for active data operations to support your business may have a negative impact if downtime occurs. A VTL system that is configured with head units can have some redundancy built in, with automated failover in the event that a unit is unavailable. A pair of head units can be interconnected, such that one unit takes over for the other, and provides access to all resources (virtual tape volumes and drives).

With all of the advanced methods to protect data with a DSI9000 virtual tape library, high availability is another option for sites that cannot lose the secondary storage for even five minutes. If your application requires you to have constant access to the to the DSI9000 virtual tape system, a second head can be configured in a clustered environment as a passive standby node. The clustered systems have a shared heart beat, in the event of a system failure with the primary unit, the passive standby system will initiate a takeover of all the VTL functions. The failover environment will be available within three minutes providing data access to host systems.

Failover is only available in a clustered head unit DSI9000 virtual tape library configuration. Appliances do not provide for this feature.

## Summary

There are a number of virtual tape library choices available from DSI from small 3TB systems to large head units in a clustered environment that will support petabytes of storage. It is important to understand the questions as they pertain to your environment and how each approach will benefit your goals while keeping in mind the costs associated with implementation and ongoing management. Archiving to disk is a real solution that should be investigated and implemented without the fear that you are headed down the wrong path. It is vital to the project to work with a vendor that understands your environment and has the global knowledge necessary to fill in the gaps.

When done right, a virtual tape library system will easily integrate into your environment and silently protect all of your backup data with easy ongoing management.

For more in-depth information, visit [www.dynamicsolutions.com](http://www.dynamicsolutions.com) or contact us at one of the following:

- [info@dynamicsolutions.com](mailto:info@dynamicsolutions.com)
- [sales@dynamicsolutions.com](mailto:sales@dynamicsolutions.com)