



Whitepaper:

Comparing Host-Based D2D to Virtual Tape Libraries for Backup and Restore

Executive Summary

Tape technologies have been the dominant force in backup and recovery since almost the beginning of modern computing, however, the landscape is finally changing. With the arrival of next-generation ATA disk technology, newer, low-cost backup and recovery methods are being offered by vendors promising to solve many of the problems that IT administrators have faced with tape backups and restores. The market has produced a proliferation of disk-based offerings ranging from simply including disk arrays as a “middleman” in the backup path to complicated imaging and snapshot systems that store their content on disk. With such a wide field of offerings, it is useful to focus in on only a couple at a time and see how they compare. In this discussion on we will take a closer look at two of the most promising of these new backup technologies, namely host based disk-to-disk (D2D) and virtual tape libraries (VTL).

At the outset, new host-based D2D backup strategies using ATA disk arrays seemed to offer the most promise, using low-cost disks and providing increased data transfer speeds for backup and recovery. As many IT administrators rushed to implement D2D, they discovered that there are a number of challenges with D2D that may not have been readily apparent, including integration issues, storage formats, file system size and performance problems, fragmentation concerns and others. IT administrators should be aware of these issues as they strive to reap the benefits of D2D in their backup and restore environments.

Another technology that is capturing the attention of IT administrators is that of virtual tape libraries (VTL). VTL takes elements from both traditional tape backup and the newer ATA disk-based technology and combines them to provide a solution optimized for existing backup environments. A virtual tape library is essentially a disk-based library that emulates standard tape library and tape formats. Acting like a tape library in the environment, with the performance of modern ATA disk, this new VTL is truly the “best of both worlds”.

In the following discussion we will first take a look at problem of traditional backup and restore from tape, another quick look at the advent of ATA technology, then delve deeper into these two new backup and restore technologies. We will investigate the benefits and challenges of each technology and present information that will help executives and IT decision makers determine which is the better technology to improve their backup environment.

The Nightmare of Traditional Tape Backup and Restore

Backup and restore from tape has been a source of consternation for IT administrators for years. One has only to make a quick search of the Internet for the topic “tape backup problems” to see the proliferation of articles and discussions about the challenges of traditional tape backups. After looking at some of the top entries, some of the problems described include:

- Tape failures during backup
- Failed notifications of incomplete backups
- Inability to locate an appropriate and current backup tape
- Cost of tape media
- Tape backup overruns into production time
- Inadequate restore speeds from tape

In spite of all of these difficulties, organizations have been left with few alternatives to traditional backup technologies until recently. To add to the list of common problems with tape backup and restore, IT administrators are now faced with a growing mountain of data that must be protected.

Improvements in computing hardware like processor and memory speeds, lower cost hardware and software, and application development advances that improve computing capabilities have all led to the creation of more (and larger) data. Historically, data consisted primarily of text, word processor documents, spreadsheets and the like. The total amounts of data created in an organization were relatively small in size and were not very difficult to backup and restore using tape media. Today organizations use more complex applications to create images, databases, 3D geophysical maps, movies and other sizeable data. The total quantity of this data within organizations has become rather daunting for IT administrators, and is increasingly difficult to manage in terms of tape backup and recovery following data corruption or system failures.

Over the past few years, IT administrators have begun to look to disk-based solutions to solve the nightmares of tape backup and recovery. The most prevalent problem with implementing disk-based systems has been the cost of hard drives in comparison to tape media. In order to utilize a disk-based solution, IT administrators were often faced with doubling their total disk capacity to create an online backup, yet still having to further backup to tape for a complete data protection scenario. This approach would solve the problem of backup overruns and tape restore speeds, but still carried a cost that was typically insurmountable. These nightmares of traditional backup soon set the stage for new technology that could offer the data protection of tape, the performance of disk, at a cost that was still in that same ballpark as the tape technology.

ATA Takes the Stage to Increase Speed While Lowering Cost

Prior to about 2001, PC (and PC Server) basically two types of hard drives were available, Small Computer System Interconnect (SCSI) and Integrated Drive Electronics (IDE). SCSI disks were typically used in server systems or those systems requiring higher transfer rates. Early SCSI interfaces were capable of only around 5MB/sec transfer, while later versions of SCSI supported 80MB/sec or more.¹ Concurrent development has also been taking place over the years in the Fibre Channel arena, with Fibre Channel drives now capable of 200MB/sec transfer speeds. These drives are now used in many high-end storage arrays and to provide optimal data transfer performance when combined with 2Gbps Fibre Channel network components. Recently, with the latest advances in Serial SCSI, this interface is capable of up to 300MB/sec data transfer.² While the performance of these SCSI and Fibre Channel drives is attractive, their cost has typically made them feasible for only primary PC, server and more expensive storage systems.

The IDE interface was later renamed Advanced Technology Attachment. Maxtor held patents on Parallel ATA technology, and has drove its development to the point that it was capable of 100MB data transfer speeds by the year 2003. According to Maxtor, parallel ATA was reaching its limits at that speed.³ To get beyond 100MB/sec, a cooperative industry group including Maxtor, Intel, Seagate, APT, Dell and IBM began developing the Serial ATA specification. Serial ATA is currently capable of 150MB/sec, and has resulted in the production of new Serial ATA drives and interfaces. In the near future, according to Seagate, Serial ATA will be capable of up to 600MB/sec data transfers, making it a very attractive, low cost, hard drive technology.⁴

¹ The SCSI Trade Association, "SCSI Laying the Ground Work for Performance", Dec. 1997

² The SCSI Trade Association, "Serial Attached SCSI and Serial ATA Compatibility", 2002

³ Maxtor, "Serial ATA Whitepaper" 2002

⁴ Seagate, "SEAGATE SAYS RELEASE OF BREAKTHROUGH SERIAL ATA INTERFACE WILL ENABLE 600-MB-PER-SECOND DATA TRANSFERS AT TRADITIONAL ATA COST", Aug. 2001

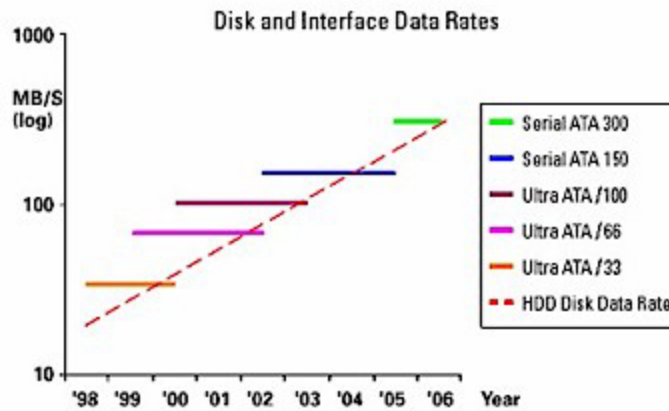


Figure 1: ATA Disk and Interface Data Rates 1998-2006 ⁽³⁾

ATA technology today offers the highest performance-to-cost ratio of any other disk alternative available on the market today. This fact has now spurred the development of a number of ATA disk-based products to assist organizations in overcoming their tape backup and recovery problems.

Could Disk-to-Disk (D2D) End Tape Backups?

Redundant Arrays of Independent Disks (RAID) aggregates the performance of multiple disks together. It is not only the speed of an individual disk that counts, but rather the speed of a disk group (i.e. the media transfer rate of an ATA 7200RPM hard disk is around 35MB/sec. With 10 disks in a group, you'd get ~350MB/sec). With the advances in data transfer rates being applied to lower cost hard drives, a number of vendors quickly began to devise products and methods for backup based on RAID arrays using ATA technology. The combined performance of disks in RAID arrays well outpaces that of writing to traditional tape. D2D became a new buzzword in the market, with promises of a new panacea for the data and backup headaches of IT administrators everywhere. Higher backup speeds, online backups and restores and additional data protection without the difficulties of tape looked promising.

Many vendors soon began to offer drive arrays, NAS devices, SAN products, and other configurations built around these new, low-cost ATA drives. Storage was now available for around a penny a MB, where it had previously been as high as ten cents or more! IT administrators rushed to get their hands on this new technology. After all, D2D offered a number of benefits over traditional tape backup, including:

- Faster writes of backup data to disk
- Faster restores from disk
- Familiar PC Server hardware
- Familiar RAID storage technology

While cheaper RAID arrays were certainly attractive, IT administrators did not readily abandon their tape backup strategies. For existing IT environments of any size, moving from a tape backup infrastructure to a disk-based strategy was more difficult than it looked. Most enterprise organizations have a sizeable investment in their existing backup infrastructure, and have likely spent considerable effort to get their backups functioning as well as possible using existing tape technology.

Business leaders and IT administrators are hesitant to throw away such an investment, and are loathe to upset their existing backup processes just for the promise of increased performance. In order for D2D to be successful in replacing tape backups, it needs to be implemented in such a fashion so as not to interrupt the existing environment, and to somehow leverage the sizeable investment organizations have made in their backup infrastructure.

Host-Based D2D Offers Some Promise

Vendors of traditional host-based backup software and hardware were quick to recognize the potential of host-based D2D and moved to ensure that they did not lose their customers to new players in this arena. Leading backup application vendors like Veritas and Legato, for example, began to offer the ability to point backups to RAID arrays instead of to tape libraries or tape drives. They also began to create vaulting applications that would allow the RAID arrays to be further backed up to tape in the traditional fashion, foreseeing that organizations might not easily forego their dependence on tapes.

These new offerings from the traditional backup application vendors seem at first glance to be exactly what IT administrators need, namely a method that will allow them to continue to use their existing backup software and still gain the advantages of the lower cost ATA RAID arrays. Host-based D2D adds benefits above and beyond generic D2D to include:

- Integration with existing backup and restore applications
- Ability to execute multiple, simultaneous backup and/or restore operations using disk-based systems
- Ability to combine disk snapshots with backup and recovery operations
- Flexibility in combining disk-based staging with tape backups
- Ability to consolidate storage and backup data pools

The Challenges of Host-Based D2D

While the benefits of a host-based D2D solution are significant, upon closer inspection however, host-based D2D solutions still have some inherent problems. Many challenges involved in host-based D2D involve the integration of the various pieces necessary for the solution. Most of the application vendors do not sell RAID arrays, servers or the other necessary network hardware (and possibly software) to ensure compatibility among the various pieces.

When we look at the installation requirements for both host-based D2D and VTL, it becomes apparent that much more configuration is required of the D2D solution. The following table illustrates the steps required for each:

Installation Steps Required for Each Solution	
Virtual Tape Library	Host-Based D2D
Assumption: Customer has existing enterprise backup application (i.e. Veritas or Legato) and was using a traditional tape library or tape drives for backups prior to the purchase of new system.	
1. VTL Arrives –VTL arrives fully configured to the customer’s specifications, including RAID arrays, hardware and included software.	1. RAID Array system arrives.
2. System is installed and plugged into existing Fibre Channel or Ethernet network	2. IT administrator or vendor will have to install hardware, hard drives, cabling and other

infrastructure.

3. Backup is executed using enterprise backup application. Since VTL emulates previously used tape library or backup, the backup application sees VTL as existing library. Changes to policies rarely required.
4. Backup job is verified. Installation and test complete.

physical hardware necessary for RAID system.

3. D2D system is directly connected (or via switch) to Host backup application server that will utilize its storage volumes.
4. IT administrator decides on RAID types, configures RAID sets
5. IT administrator installs file system required by host backup application (i.e. NTFS for Windows, NFS for Solaris) and partitions the volumes created in the RAID sets.
6. IT administrator configures security access, zones, etc.
7. IT administrator configures host backup application to utilize RAID system as target for backup operations.
8. New backup job is created and run.
9. Backup job is verified. Installation and test are complete.

Host-based D2D has considerably more steps involved in the installation than a VTL system. As we can see above, the VTL system is designed to be plug and play. The D2D system being a collection of pieces from various vendors essentially brings all of the challenge of primary storage configuration into the backup arena. IT administrators should be aware of the potential installation requirements, know the limitations of the hardware, filesystems, RAID types etc. before deciding to implement this kind of solution.

Compatibility presents some challenge for IT administrators wishing to implement host-based D2D. The IT administrator must typically assemble a collection of devices from different vendors, configure each of these to operate as needed, then connect the hardware and applications so that they will interoperate properly. Incompatibility could potentially be a serious problem if the wrong devices or applications are coupled. Further exacerbating the problem is the issue of technical support, should any one of the pieces fail. This scenario could easily result in finger pointing among vendors, leaving the IT administrator to act as referee, while his data protection is in jeopardy.

Even though compatibility and integration provide the most significant problems with a host-based D2D solution, there are other problems inherent in the solution as well. The D2D capabilities of the backup/restore software in conjunction with off-the-shelf RAID arrays rely on file systems like NTFS and NFS that were created for use in PC server systems. These file systems were designed to manage many small files for many users. The problem when these file systems are used for tape-style backups is that they take a performance hit when performing the larger write operations typically executed by tape backup applications. Additionally, when combined with tape backup applications, the storage system's processor takes an additional, unnecessary performance hit, as the application must interact with the file system's file allocation tables to determine a particular file's location.

As mentioned, general-purpose file systems used with most D2D RAID arrays are not optimized for tape backups. Three other problems exist with these general-purpose file systems relative to backup systems.

1. Most general-purpose file systems are subject to significant file system fragmentation over time just as typical servers are. Fragmentation of the file system increases over time and degrades the performance of the reads and writes on the system.
2. General-purpose file systems like FAT, NFS, and NTFS are subject to file system, partition, and individual file size limitations. There is an additional limitation as to the total number of files allowed on a file system. These limitations can impose restrictions on the scalability of backup volumes used by host-based D2D applications. When the limitation is reached, organizations will incur additional expense and difficulty by having to move data, create new volumes, and reconfigure their backup applications.
3. The other major problem with many D2D file systems is security of the data that resides on them. Hackers with malicious intent and viruses can compromise general-purpose file system security. Any D2D file system could be subject to these same security concerns, which could lead to a failure in the data protection scenario. As we will see in the following section, these problems can be avoided when using a more advanced disk-based backup system called a Virtual Tape Library (VTL).

Virtual Tape Libraries Offer the “Best of Both Worlds”

IT administrators have enough problems to face in their existing environments without creating new ones by integrating multiple vendor products in an attempt to improve their backup system. Another, more advanced ATA disk-based backup solution has appeared on the market, one that provides all of the benefits of a traditional backup system with the benefits of a host-based D2D system, while removing the problems provided by each. The VTL offers IT administrators the best of tape technologies while delivering performance, reliability and scalability of disk. It represents a solution that is far superior to traditional tape-based data protection systems.

A VTL is a disk-based backup system that emulates one or more tape libraries and/or tape formats. The VTL is a completely integrated solution, which includes the tape emulation server, the ATA RAID arrays, and a file system optimized for tape backups. In many cases additional proprietary hardware and software is included by the vendor that is designed to provide additional backup and restore functionality or enhanced performance. The VTL can also provide the same media management (tape cataloging, media expiration etc.) capabilities that the administrator is used to, while most host-based D2D systems fail in this regard.

Because the VTL is a completely integrated solution, the issues mentioned in the previous sections about compatibility, integration and installation are no longer a concern for the IT administrator. As such, the VTL should arrive from the vendor configured and ready to install into the existing network environment. In most cases, the regular backups should be possible in a very short time (even within minutes), with little or no changes to the current backup policies set in the existing backup application. When the backup job runs, it will deposit the data onto the VTL just as it would onto a tape library, only many times faster. Restores will also be many times faster.

The other file system problems discussed previously regarding performance and security are also no longer a concern. This is due to the fact that the file system for a VTL is typically proprietary and designed to act similar to a tape system with performance of ATA disk. ATA disks perform well when executing streaming reads or writes, but their throughput drops drastically once random seeks occur. For example, SEPATON's Disk Dynamic File System allows large I/O streams to execute efficiently and has the built-in infrastructure to dynamically balance performance across all available disks in their VTL appliance. The Disk Dynamic File System has the important side effects of not only sustaining maximum throughputs, but also dynamically load balancing I/O streams without any requirement for performance “tuning”. The following diagram represents the comparison between the backup of 20GB of data using a host-based D2D system and a VTL (SEPATON's S2100). Note the resulting performance increases possible with a VTL and its optimized file system.

Looking at the two graphs below, it is notable that while the VTL (S2100) was able to complete the backup of 20GB of data over twice as fast as an off the shelf host-based D2D configuration, the same VTL (S2100) was able to backup 80GB of data over 2 ½ times faster than the D2D system. This clearly shows that the file system performance advantage actually increases as the quantity of data being backed up increases!

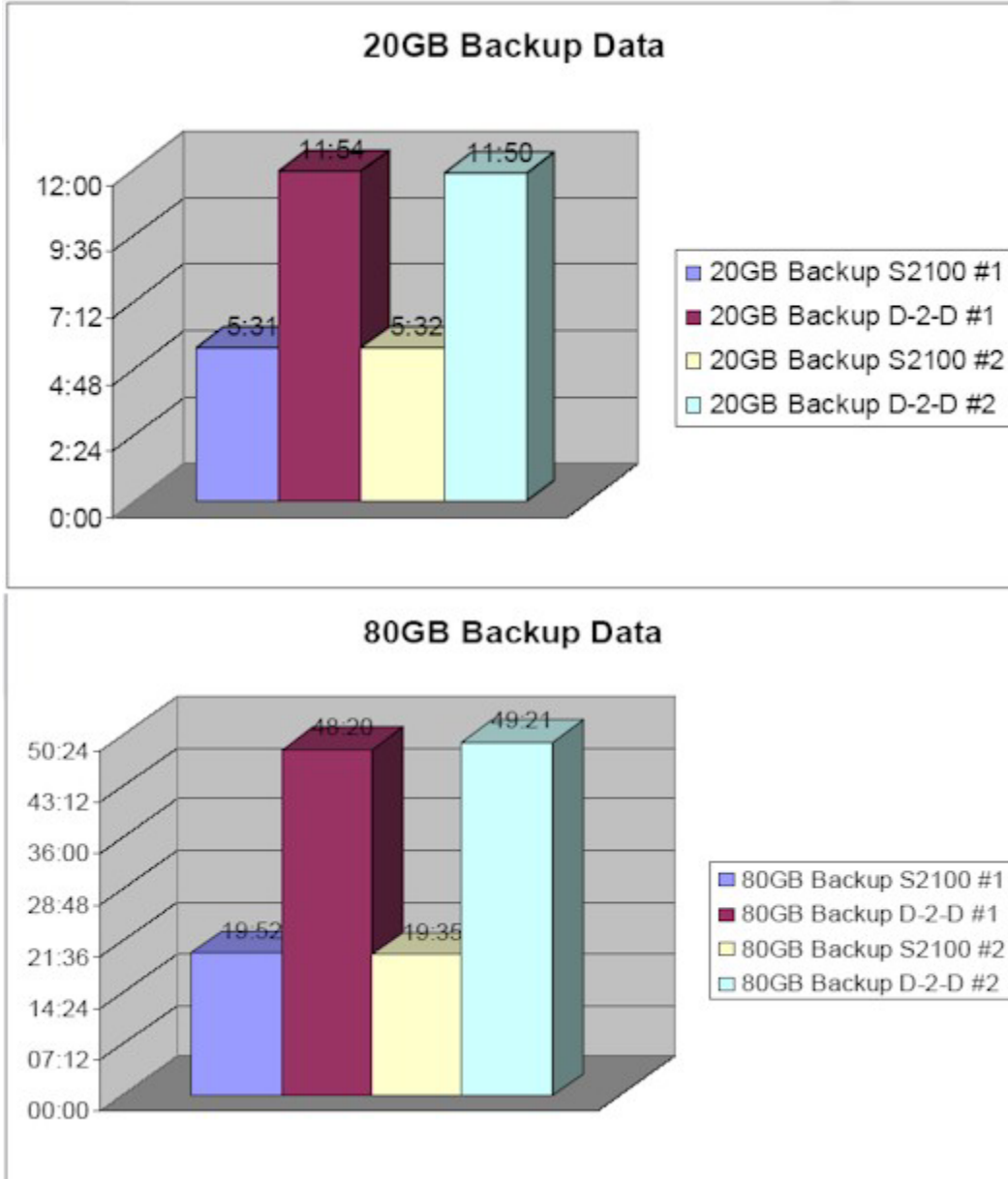


Figure 2: Time in Minutes to Backup 20GB and 80GB of Data

(For more information regarding the comparison tests that resulted in the above graphs, please visit <http://www.sepaton.com> and download the whitepaper titled "S2100-ES vs. Traditional D2D Backup")

One of the reasons that the VTL outperforms the D2D system during backup is that the backup application writes a backup to the VTL disk in much the same manner as it would to tape...in long continuous streams. The D2D on the other hand writes individual files, and must record each file location and meta data to the file system before moving on to the next. This same method the VTL uses for writing streams of data instead of individual files prevents the VTL from encountering the limitation on the number of files its file system can support. The proprietary nature of the VTL file system overcomes the other file system limitations that the D2D experiences with its general-purpose file system.

With regards to security, accessing the file system of the VTL is like accessing that of tape. A user cannot gain access to a tape directly, nor can a virus-infected file infect a tape, as it cannot access the files as it would on a typical file system. The file system in most VTL appliances prevents viruses or other security threats from wreaking havoc on the system. A typical host-based D2D system and its general-purpose file system will not have the ability to provide this security, resulting in potential data loss for the organization -- the very problem it is supposed to avoid.

Drawing Some Conclusions About VTL

The information provided regarding the difficulties of tape backup was not likely news to anyone, in particular to IT administrators who have been living with the problem for years. It is important to understand the specifics of the tape challenge, however, in order to seek out the best possible solution or technology. ATA disk is just such a technology, and it was shown that the best possible read and write performance comes at the best possible price with this technology. In other words, the performance to cost ratio is higher with ATA than any other disk technology that might be used in place of tape. The cost has finally dipped low enough to spur the development of disk-based solutions to the tape backup nightmare.

We saw that host-based D2D provides significant performance benefits, and many of the benefits that are being sought by IT administrators looking to implement disk-based backups. Host-based D2D integrates with existing backup applications, and provides performance enhancements, but as we discussed, the integration, compatibility and support concerns could outweigh (and out-cost) the benefits of the solution. Moreover, the ongoing "care and feeding" of a D2D solution will quickly erode any staffing cost benefits that were originally realized.

Virtual tape library technology is the most promising technology available that will allow IT administrators to leverage the existing investment in their backup infrastructure while moving forward to gain the advantage of ATA disk enhancements.

Virtual tape library solutions are:

- Easy to implement, often installed and backing up using existing applications in minutes
- Easy to integrate, as they look like a standard tape library to your applications
- Flexible, as there are a number of scenarios with which they will work
- Scalable, like disk arrays, in that they can be expanded to back up hundreds of terabytes
- HIGH PERFORMANCE! Can be 10 times faster than traditional tape libraries.

VTL technology captures all of the benefits discussed for host-based D2D, then adds even more. The choice for IT decision makers should be clear. Virtual Tape Library appliances are the best choice for disk-based data protection.

Contact SEPATON at 508.490.7900 or via email at sales@sepaton.com today for more information!

About SEPATON

Based in Southborough, MA, SEPATON, Inc. develops and markets a family of data protection solutions to medium and large data centers that must manage multiple terabytes of data daily. SEPATON's data protection solutions integrate seamlessly into existing customer environments.