Getting the most performance from your HP StorageWorks Ultrium 960 tape drive white paper

Executive summary

The HP StorageWorks Ultrium 960[1] tape drive is arguably the world's fastest tape drive capable of speeds exceeding 160 MB/sec (576 GB/hr). To make use of this available performance requires a new way of thinking when designing backup solutions. This white paper explains how to get the best performance from the HP StorageWorks Ultrium 960 tape drive in a variety of technical configurations:

- Direct attach just a bunch of disks (JBOD)
- Entry-level storage area network (SAN)
- Enterprise-level SAN
- Over the network (OTN) backup
- Disk Staging Backup (D2D2T)

As a result of these tests, several recommendations and rules of thumb have emerged:

- Higher levels of concurrency (interleaving more data streams into a single tape stream) will be required to make best use of the available HP StorageWorks Ultrium 960 performance. Most backup software allows up to 32 streams to be interleaved to tape. However, interleaving more streams to tape can have a negative impact on restore performance.
- For entry-level JBOD and low-end disk array usage of HP StorageWorks Ultrium 960, it is important to understand what your disk subsystem is capable of delivering. This can be done using HP performance assessment tools downloadable from http://www.hp.com/support/pat or the performance tools are also embedded within the HP industry-leading Library and Tape Tools diagnostics.
- For servers in a SAN environment, the rule of thumb is 1.5 to 2.00 GHz of processing power and 1.00 GB of RAM per processor should be allocated for the backup process, where the backup process is reading from disk and writing to HP StorageWorks Ultrium 960 tape, using separate data channels (not sharing the same single fiber).
- OTN backup is not ideally suited to HP StorageWorks Ultrium 960 tape drives since Gigabit Ethernet (Gbe) itself has a 120-MB/sec limit, which is likely to become a bottleneck. In addition, high processing power is required to manage the TCP/IP stack. To improve performance consider:
  - Multiplexing more streams to the tape drive until the 120-MB/sec Gbe bottleneck is reached.
  - Using multiple network interface card (NIC) host bus adapters (HBAs) in the backup server and binding each of them to a processor.
  - Using TCP Offload Engines (TOES) to reduce processor loading.
  - Moving to a SAN-based backup architecture.
  - Moving to a Disk to Low cost Disk to Tape backup architecture (D2D2T) and performing the initial backup over the network to disk not tape, and then de-staging to tape.
- ISV parameter tuning can make a significant difference (up 60% improvement) in performance. See tuning recommendations for the major backup ISVs section in this white paper.
- D2D2T: Serial ATA (SATA) disks used as secondary disk storage arrays can have slower rotational speeds than their SCSI counterparts (for example, the HP StorageWorks Modular Smart Array [MSA] SATA drives are only 7.2K rpm) and although they are high capacity (250 GB), their performance reflects their pricing. Therefore, backing up to tape from a staged (secondary) disk array can be slower than backing up directly from the primary storage to tape in some circumstances. Ironically tape is now faster than disk! Disk staging is useful, however, for gathering several small files into a single object, or backing up slow networked hosts before the data is sent to tape. Both small files and slow hosts can cause very slow backups.

---

[1] The name is derived from adding together the capacity 800 GB (at 2:1 compressible data) and the speed 160 MB/sec (at 2:1 compressible data).

- For backups, use a block size/transfer size of 256 KB (most backup applications use 64 KB as default).
- Ensure the Ultra320 SCSI HBA for the tape drive is placed on a dedicated 133-MHz PCI-X bus and is not sharing the bus with other HBAs.

The HP StorageWorks Ultrium 960 tape drive offers a capacity of up to 800 GB per tape and also supports write-once read-many (WORM) media for use in compliance environments. It is available in internal, external, and rack mount configurations and will be deployed into HP automated tape libraries in early 2005. Data rate matching capabilities within the drive allow the drive to match its speed to incoming data from the host and remain streaming, ensuring optimal performance and reliability in all situations. (Data rate matching operates between host speeds of 27 to 80 MB/sec for uncompressible data.)

Figure 1. HP StorageWorks Ultrium 960 with Write Read media (yellow) and WORM media (yellow/grey)

The HP StorageWorks Ultrium 960 tape drive performance is best utilized for high-performance backup in a disk and tape SAN environment, backing up directly from a high-performance disk subsystem (from primary disk storage, split mirrors, snapshots, or clones). The HP StorageWorks Ultrium 960 is not ideal (performance wise) for OTN backup or backup from lower performance disk arrays. HP recommends matching appropriate LTO technology generations to your server capabilities. However, if you have capacity needs met by a higher generation product, you can utilize it in the knowledge that it will give you ample performance headroom in future.

Finally, tape continues to deliver the lowest $/GB of any storage technology and is no longer slow. Tape should always be the foundation of a data protection strategy. Rumors of tapes demise are unfounded.

Objectives and target audience

The main objective of this white paper is to educate and inform purchasers of the HP StorageWorks Ultrium 960 tape drive as to what levels of performance are achievable in different backup scenarios.

The emphasis is in showing what is typical and not what best-case scenarios are. The white paper highlights where the current bottlenecks to performance are and how these might be overcome both now and in the future.

The target audience for this white paper is system integrators and system architects and indeed anyone involved in getting the best backup performance out of their infrastructure investments. Throughout the white paper the HP StorageWorks Ultrium 960 may also be referred to as HP LTO3.

Introducing the HP StorageWorks Ultrium 960 tape drive

The HP StorageWorks Ultrium 960 tape drive is literally the world's fastest tape drive[2], with a native transfer rate of 80 MB/sec and a native capacity of 400 GB per tape. With 2:1 compressible data, this means the drive is capable of backing up data at 160 MB/sec (576 GB/hr) and storing up to 800 GB of data on a single piece of media.

At launch few systems will be able to take full advantage of this kind of performance. However, because the HP StorageWorks Ultrium 960 tape drive has a feature called data rate matching, the tape speed range can vary between 27 MB/sec and 80 MB/sec (1:1 data), so all users will benefit some degree due to the tape drive's ability to dynamically match its performance to the speed of incoming data from systems being backed up. In short, *the tape drive will no longer be the bottleneck to backup performance.*

The HP data rate matching means that as long as the data supply to the tape drive is a minimum of 27 MB/sec, the tape drive will continue streaming, hence ensuring best possible backup speed and improved media life through a reduced number of repositions (media passes).

Table 1.

Specification	HP StorageWorks Ultrium 460	HP StorageWorks Ultrium 960
Capacity (native)	200 GB	400 GB
Transfer rate (native)	Up to 30 MB/s	Up to 80 MB/s[1]
Data rate matching	10–30 MB/s	27–80 MB/s
Host interfaces	Ultra160, 2 Gb FC	Ultra320, 2/4 Gb FC
Head channels	8	16
Bit density	7.86 kbpi	10.2 kbpi
Number of tracks	512	704
Media length	609 m	680 m
WORM media support	NO	YES
Media compatibility		
Read	Ultrium Generation 1, 2	Ultrium Generation 1, 2, 3
Write	Ultrium Generation 1, 2	Ultrium Generation 2, 3

---

[2] HP classifies the performance of the HP StorageWorks Ultrium 960 as 80 X 1000 X 1000 bytes per second—80 MB/sec (that is, to base 10) in common with most other disk and tape drive vendors. Most applications, however, measure performance as 1024 X 1024 bytes per second, so for example in the HP StorageWorks Enterprise Virtual Array 5000 (EVA5000) performance test point, a transfer rate of 146.8 MB/sec as measured by HP OpenView Storage Data Protector was achieved. The equivalent transfer rate of the tape drive in the HP 1000 X 1000 formula is 153.9 MB/sec.

Table 1 shows the major changes from the previous generation HP StorageWorks Ultrium 460 tape drive. The increased performance comes from a combination of doubling the number of write/read heads and increasing the write/read channel performance. The capacity increase is achieved through longer tape and greater bit density on tape. The physical tape speed of HP StorageWorks Ultrium 960 is identical to that of HP StorageWorks Ultrium 460 in that it is 5.32 m/sec. The native Fibre Channel versions of the drive are only available in automated tape libraries and are not available as stand-alone tape drives.

Another useful innovation for HP StorageWorks Ultrium 960 is that it is the first Linear Tape Open (LTO) tape drive to support WORM media, hence making it an ideal, cost-effective solution for regulatory compliance.

Finally HP StorageWorks Ultrium 960 tape drives will write LTO3 format and LTO2 format (when the appropriate media is inserted) and read LTO3, LTO2, and LTO1 format media.

Basics of tape backup performance

Several factors critically affect the speed of backup from disk to tape:

- Data file size

  The larger the number of smaller files there are, the larger the overhead there is associated with backing them up. The worst-case scenario for backup is large numbers of small files due to system overhead of file accession.

- Data compressibility

  Incompressible data will back up slower than higher compressible data.

  Jpeg files, for example, are not very compressible, whereas database files can be highly compressible. The accepted standard for quoting tape backup specifications revolves around an arbitrary figure of 2:1 compressible data, which is what is used in the proof points detailed later in this white paper.

- Disk array performance

  It is often overlooked that you cannot put data on tape any faster than you can read it from disk. Backup is more sequential in nature than random (from a disk array access perspective).

  Disk array performance depends on the number of disks, RAID configuration, the number of Fibre Channel ports to access the array, queue depth available, and so on. HP has written several utilities to "read data" from disk arrays and deliver a performance value. This enables users to determine the throughput of their disk arrays when operating in backup mode and performing filesystem traversals typical of this activity.

  These performance assessment tools (pat) utilities can be downloaded from http://www.hp.com/support/pat or the performance tools are also embedded within the HP industry-leading Library and Tape Tools diagnostics.

- Backup software techniques

  Most enterprise backup solution software products allow multiple streams of data from different machines to be interleaved on the same physical piece of media. This technique can be called concurrency or multiplexing or interleaving depending on the ISV. This allows better bandwidth utilization of the tape drive during backup but can slow down restore performance because all the data is interleaved all the way down the tape. Therefore, the time spent to perform a single stream restore is higher due to other streams having to be read (and potentially ignored).

- Backup software tuning

  Most backup application software allows for the number of disk buffers or tape buffers to be changed to tune the backup software to particular circumstances. Remember, changing these values is a process of "trial and error" rather than a set of rules.

  In addition, most backup application software allows the block size on tape to be changed. For all HP StorageWorks Ultrium 960 backups, HP recommends the tape block size/transfer size be configured to 256 KB.

- Transfer size

  Transfer size is the overall total size of the SCSI transfer within a single SCSI command. In some operating systems, there is limit set on this. For example, in Microsoft® Windows® the default transfer size is 64 KB and to increase the overall transfer size above 64 KB in Windows, a registry entry called MaximumSGList (associated with the HBA) must be changed. Many modern HBAs already install their drivers with this registry value set appropriately. Check the registry entries.

- Fragmentation

  The more fragmented the files are on disk, the more random will be the disk access method, hence the backup will take longer. If your system does have a defragmentation utility, it is advisable to run it before full backups or on a regular scheduled basis to ensure files are largely contiguously arranged on disk.

- Backup application database

  It is sometimes overlooked that during backup every file backed up has an entry made in the backup database (sometimes called catalog). When this is located on a slow disk/array, it can slow down the backup process.

- SAN inter switch links (ISLs) ensure in your switched fabric that ISLs have sufficient bandwidth to support the backup traffic going through them. Trunk multiple ISLs together where possible.

- Physical interface limitations

  These are particularly important with HP StorageWorks Ultrium 960 since this pushes the limits even further. In general, you can expect the following bandwidth limitations:

  – **Gigabit Ethernet (Gbe)**—Expect no more than 120-MB/sec maximum. Enable Jumbo frame capability if your switch and NIC to reduce TCP/IP protocol overheads. Examine your routing tables and overall network topology—appropriate switching and routing can enable higher throughput and better use of network throughput capacity.
  – **2-Gb/sec Fibre Channel**—Expect no more than 200-MB/sec maximum from a single 2-Gb/sec Fibre Channel connection.
  – **SCSI burst rate**—Beware of disk drives quoted as Ultra320; Ultra 320 refers to the burst rate not the sustained rate. The typical sustained rate from a 15K rpm SCSI Ultra320 disk drive is around 80 MB/sec for raw sequential I/Os (that is, without file system read overhead).
  – **SATA disks**—These types of disks are lower cost, lower performance, and lower reliability than the SCSI disk previously listed but are useful for staging backups because they are high capacity. A typical 7200 rpm SATA disk drive has lower seek times than an the equivalent SCSI disk drive, a burst rate of around 150 MB/sec, and a sustained transfer rate of around 50 MB/sec.

Data protection architectures

This white paper considers tape backup under the following scenarios:

- **Direct Attached Storage (DAS)**—Disk storage is connected directly to the server by way of a SCSI interface and the tape drive is attached to the same server also by way of SCSI. The server itself runs the backup software.
- **Over The Network (OTN)**—In this scenario, multiple clients are backed up over the network to a centralized dedicated backup server. The data being backed up can reside on either the clients' "local" disks or the clients can themselves be sharing disk storage from a SAN-attached disk array. Disk agents are loaded on the remote clients and these "push" the data to the backup server. The backup server then writes the data to tape.
- **SAN-based backup**—In this scenario all the servers using shared storage in the SAN will have both the disk agent and the media agent from the backup application loaded. Hence the SAN-attached servers gather the data by the disk agent and then themselves push the data on a physical tape drive. In these types of scenarios, tape libraries are generally used and the tape drives shared amongst the servers wanting to back up. The bandwidth available to SAN-based backup is much higher because each server can have a guaranteed 2-Gb/sec link to the tape drives.

  The two examples used in this white paper are somewhat simplified and are aimed at showing the types of results that can be expected from the HP entry-level (HP StorageWorks Modular Smart Array 1000 [MSA1000]) and high end (EVA5000) disk arrays.

- **Backup Staging (D2D2T)**—With the advent of SATA disk drives promising lower cost/GB disk storage, a new technique in data protection has emerged. Here the primary disk data is backed up to the lower cost (secondary) disk array and then finally to tape. Note that tape is still the foundation of the data protection strategy offering removable, very low-cost storage in terms of $/GB. The advantage of the additional array is that single file restores are faster (as long as the data still resides on the secondary array). Full system streaming restores may not be any faster from disk—compare typical disk throughputs with streaming read performance from an HP StorageWorks Ultrium 960.

The HP "typical" file system and HPreaddata utility

In the following proof points, HP tried to create as typical a system as is possible so that the results shown are realistically achievable in similar situations. The proof points are all Windows based for simplification of equipment needs, but the lessons will still hold good for heterogeneous environments.

The datasets were developed using the HPcreatedata utility, which is downloadable from http://www.hp.com/support/pat, and the settings to build these datasets are shown in Figure 2.

Note the range of file sizes 64 K through to 64 MB and the compressibility of the data (2:1). The utility creates an equal distribution of files in each directory.

Figure 2.
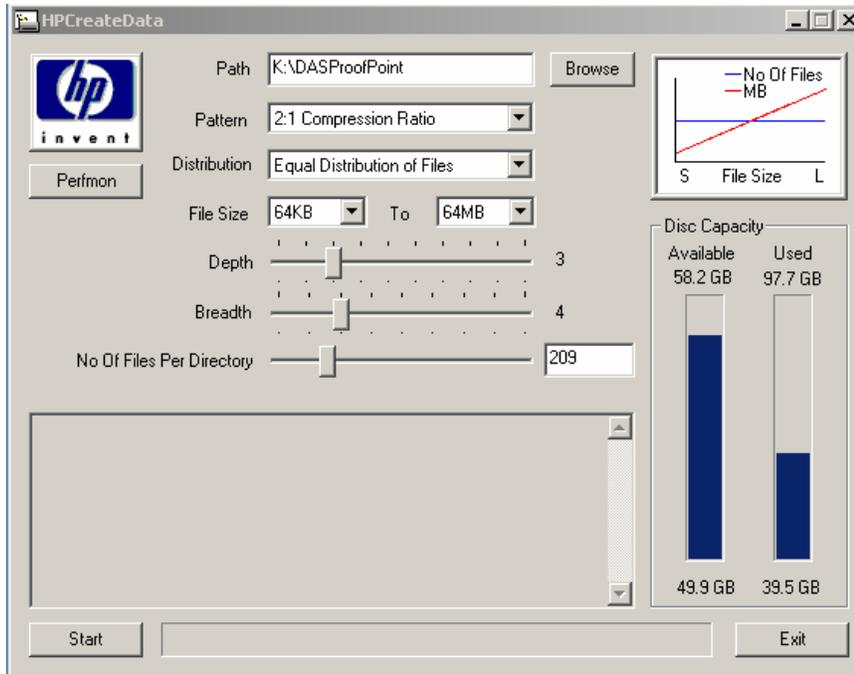
Figure 3.
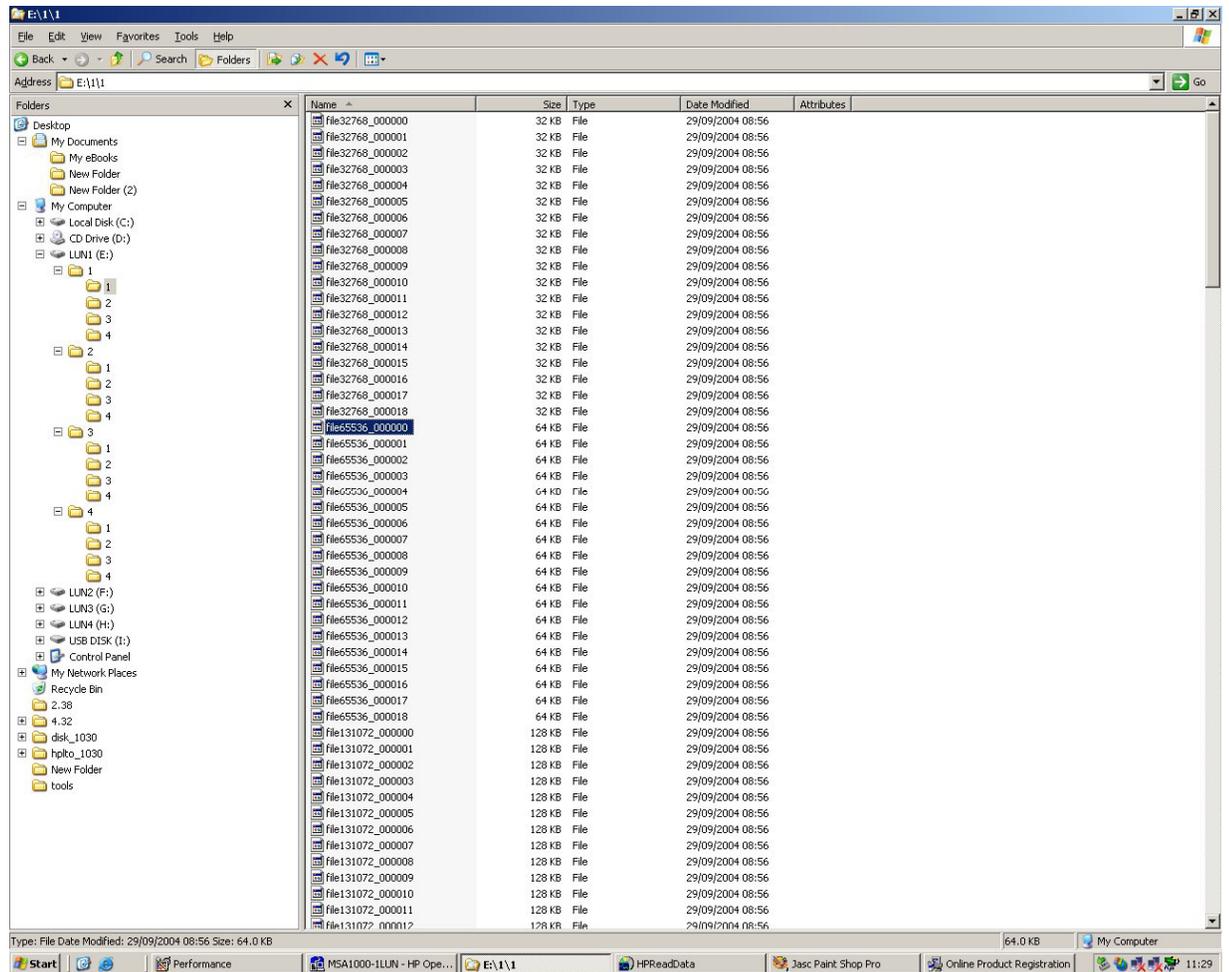
Using the HPcreatedata utility results in a directory and file structure similar to that shown in Figure 3.

For some of the restore tests used in this white paper, files were restored early in the directory tree and later in the directory tree to test the fast access capabilities. For example, file /1/1/file65536_000000 was chosen to be early in the directory tree (hence earlier in the backup and nearer the start of tape) and file 4/4/file65536_000000 was chosen because it was toward the end of the directory tree (hence later in the backup and further down the tape media).

The HPreaddata utility is useful in assessing the rate at which your disk subsystem can supply data, and this is ultimately what will limit the backup performance (providing the ISV backup application is efficient at buffering and writing the data to tape). A single instance of HPreaddata can read eight streams simultaneously from your array. To read more than eight streams, initiate multiple instances of HPreaddata. HPreaddata is available for Windows, HP-UX, Solaris, and Linux® and can be downloaded free from http://www.hp.com/support/pat or the performance tools are also embedded within the HP industry-leading Library and Tape Tools diagnostics.

A typical HPreaddata screenshot is shown in Figure 4.

Figure 4.
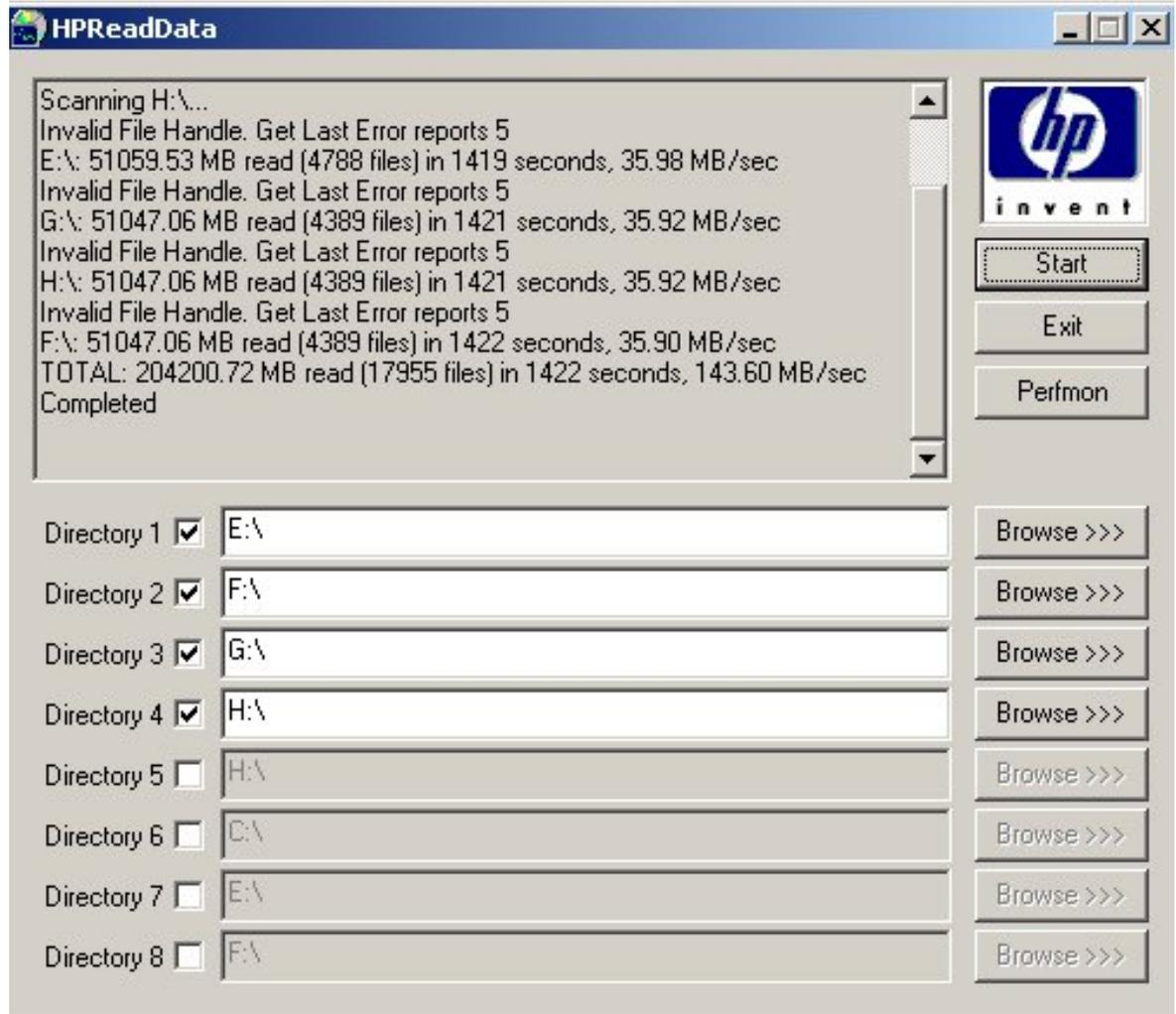
Figure 4 shows HPReaddata simultaneously reading 4 LUNs from the EVA5000 in a manner similar to the way a backup application will read files. The maximum read rate from this configuration is 143 MB/sec, so do not expect any higher backup transfer performance to tape than this figure.

The following series of proof points illustrates the level of backup performance you can expect from your HP StorageWorks Ultrium 960 tape drive in different backup configurations. Throughout the following proof points, the operating system used on all servers is Windows Server 2003 Standard Edition.

Throughout these tests a logical unit number (LUN) is a specifically addressable storage point presented to the operating system by the storage device. It can be presented as a drive letter or as a mount point.

The disk arrays configured in these examples are configured as RAID 0 to emulate the optimal performance the disk array can deliver for the given number of drives available.

Proof points

Direct Attach SCSI Disk (DAS)

Figure 5.

Direct Attach SCSI Disk (DAS)
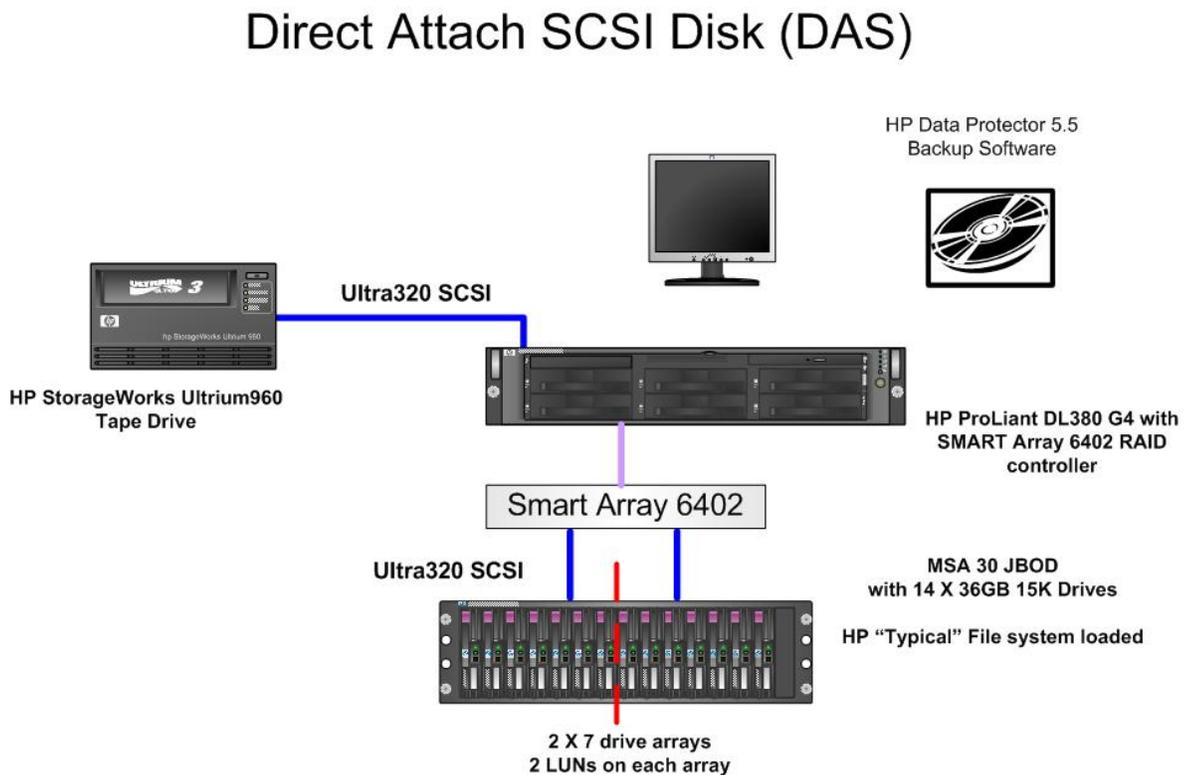
This scenario illustrates what a typical single server-based application such as CAD/CAM or File/Print server might look like, and what kind of backup and restore performance an HP StorageWorks Ultrium 960 can bring to this type of environment. The scenario looks at 2 X 7 drive disk arrays, but even the HP ProLiant DL380 G4 server used in the test has the capacity for a 6 disk drive disk array using an embedded SMART array controller so the scenario also represents what a minimal server internal disk array subsystem might look like.

Detailed configuration description

A disk array is composed of 14 X 36-GB 15k disks in an enclosure and driven by an HP SMART Array controller 6402 with two Ultra320 channels. The proof point shows the typical server with direct attached disks that could make use of the HP StorageWorks Ultrium 960 performance and capacity.

The disk enclosure is configured as two separate disk arrays, each of seven disks and each array configured as RAID 0. Two LUNs are created on each disk array. Each array has a separate Ultra320 SCSI channel.

HP ProLiant DL380 G4 server features 3.4 MHz with 1 GB of RAM, running Windows Server 2003.

Each LUN is loaded with HP "typical" file system as previously described. Each LUN is 50 GB.

HP OpenView Storage Data Protector 5.5 is configured with 256-KB block size, 32 disk buffers, and 20-GB segment size.

Results

Table 2.

Backup performance (MB/sec)			Processor Load (%)
1 LUN = 57.9MB/sec 2 LUNs (1 from each array) = 99.3 MB/sec 4 LUNs (2 from each array) = 101.1 MB/sec			The processor load during these backups varied from 34% to 55% for 2 and 4 LUNs and was between 18% to 40% for a single LUN.
Restore performance			**Processor load**
Backup type	**Full LUN restore speed**	**Single file access times***	Full LUN restore varied from 29% to 42% for 2 or 4 LUNs and was between 70% to 90% for a single LUN restore. Single file restores use minimal processing of 1% to 5%.
4 LUNs	39.7 MB/sec	$37^2/87^3$ sec	
2 LUNs	64.4 MB/sec	33/72 sec	
1 LUN	101.9 MB/sec[1]	20/56 sec	

When more than one LUN is backed up, the data is interleaved on a single tape.

* Single file access times were performed by taking a file midway in the directory tree and toward the end of the directory tree. The speed of these single file restores is dependent on the efficiency of the backup software and the fast seek time algorithm of the tape drive. The full LUN restore performance includes overwriting the disk subsystem with the files from tape.

[1] An unexpected result. The RAID 0 write performance on a single LUN appears to be much faster than the read performance on a single LUN. This is a caching feature of the HP SMART Array 6402 Controller.

[2] Restore time of a single 64-KB file early in the directory tree.

[3] Restore time of a single 64-KB file late in the directory tree.

Observations

Even a relatively modest array of seven drives can use the backup performance of the HP StorageWorks Ultrium 960 tape drive. If the partitions are on different disk arrays, then these can be read simultaneously and HP OpenView Storage Data Protector can interleave the sources onto tape for even better performance, utilizing the tape drive performance more effectively.

Full LUN restore performance is better when there are fewer LUNs interleaved on the tape. The results vary from 40 MB/sec for four LUNs to more than 100 MB/sec for a single LUN tape. Despite the data being interleaved, the HP StorageWorks Ultrium 960 tape drive fast search algorithm can still deliver a single file restore within an average of 72 seconds after the tape media is loaded into the tape drive.

With seven (2 drive) LUNs configured and staggered across the two Ultra320 channels, the HPreaddata utility reading the seven LUNs simultaneously showed the array could supply data through the file system at 190 MB/sec. The backup of these seven LUNs ran at 378 GB/hr (105 MB/sec). The difference between HPreaddata and the actual backup performance is the overhead in file buffering in memory and writing into tape format by the ISV application.

So in this scenario the bottleneck to backup performance is the number of LUNs that can practically be configured, but then the number of LUNs is dependent on the server function/configuration and the efficiency of the RAID controller.

Figure 6.

SAN Backup with MSA Disk array

Configured as 4 x 7 drive disk arrays. 1 LUN on each array.

Dual U320 Link to MSA30 expansion box. Single 2Gb Fibre connection to the Array

MSA1000 Disk Array with 28 36GB/15K Disk Drives,

HP typical filesystem installed

Ultra320 SCSI

2 Gigabit Fibre channel

HP ProLiant DL380

HP Data Protector 5.5 Backup Software

Ultra320 SCSI

HP StorageWorks Ultrium960 Tape Drive
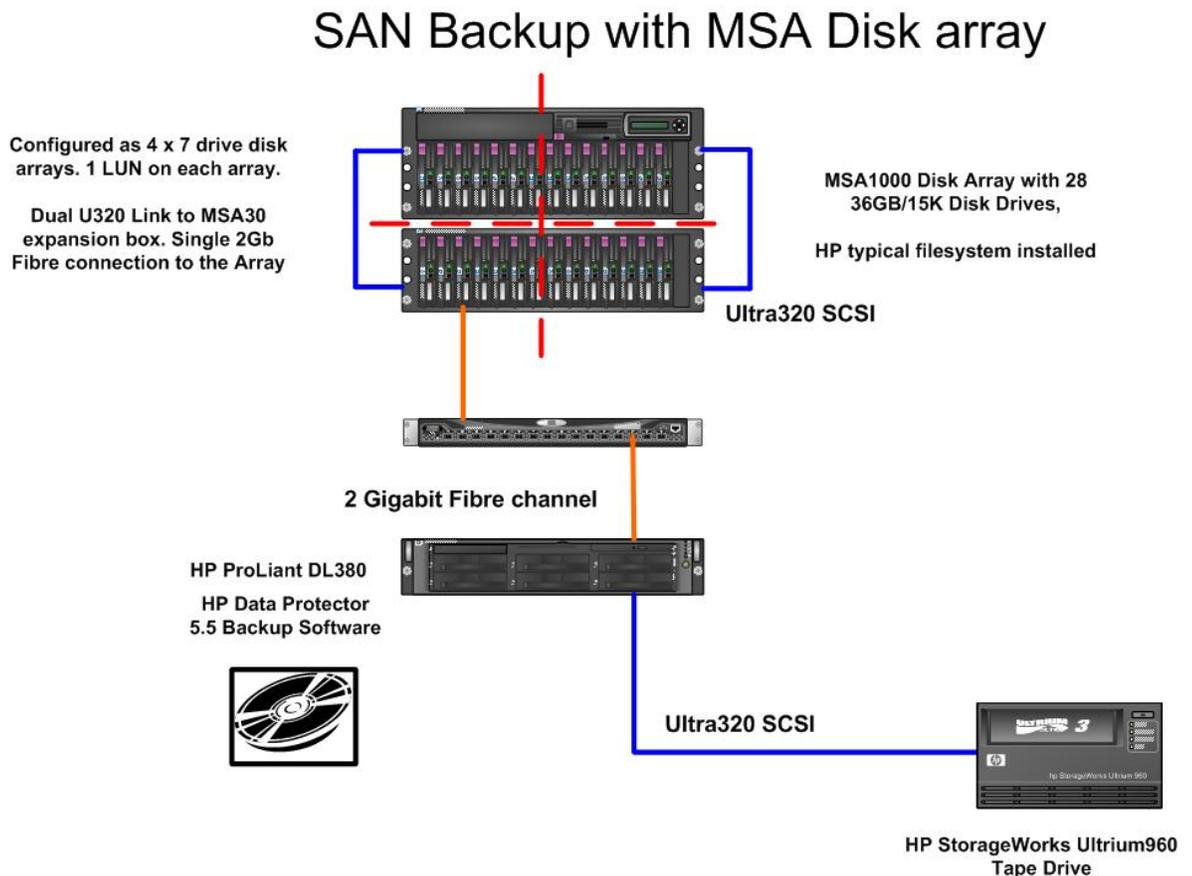
This scenario represents an entry-level SAN deployment. The MSA1000 SAN-based disk array can expand up to 42 disk drives, and utilizes an embedded Fibre Channel switch. It has a single 2 Gb/sec Fibre connection. A 28-drive model was selected to try to be "typical." The SAN-based array is split into 4 x 7 drive disk arrays by using the HP array configuration utility, and then the arrays are presented to the host. Each array has a single LUN on it, which in turn is loaded with the HP "typical" file system dataset. This scenario could emulate a small business SAN with different departments being allocated for one of the LUNs, for example, R&D, marketing, finance, and human resources.

Detailed configuration description

A disk array is composed of 28 X 36-GB 15k Ultra320 disks in an enclosure and driven by an MSA1000 controller. The MSA1000 controller is Ultra3 (Ultra160) based. The expansion boxes (HP StorageWorks Modular Smart Array 30 [MSA30] enclosures) have dual Ultra320 channels. The proof point shows the typical entry-level SAN deployment that could make use of the HP StorageWorks Ultrium 960 performance and capacity.

The disk enclosures are configured as four separate disk arrays, each of seven disks and each array configured as RAID 0. Four LUNs are created, one on each disk array.

HP ProLiant DL380 G4 server features 3.4 MHz with 1 GB of RAM, running Windows Server 2003.

Each LUN is loaded with HP "typical" file system as previously described. Each LUN is 50 GB.

HP OpenView Storage Data Protector 5.5 is configured with 256-KB block size, 32 disk buffers, and 20-GB segment size. When more than one LUN is backed up, the data is interleaved on a single tape.

Results

Table 3.

Backup performance (GB/hr)			Processor load (%)
1 LUN = 48.8MB/sec) 2 LUNs = 88.9 MB/sec 4 LUNs (1 from each array) = 104.0 MB/sec			The processor load during these backups varied from 47% to 55% for 2 and 4 LUNs and was between 15% to 25% for a single LUN.
Restore performance			**Processor load**
Backup type	**Full LUN restore speed**	**Single file access times***	Full LUN restore varied from 29% to 42% for 2 or 4 LUNs and was between 54% to 85% for a single LUN restore.
4 LUNs	33.3 MB/sec	31/54 sec	Single file restores use minimal processing of 1% to 5%.
2 LUNs	60.8 MB/sec	22/120[2] sec	
1 LUN	97.2 MB/sec[1]	20/53 sec	

* Single file access times were performed by taking a file midway in the directory tree and toward the end of the directory tree. The full LUN restore performance includes overwriting the disk subsystem with the files from tape.

[1] An unexpected result. The RAID 0 write performance on a single LUN appears to be much faster than the read performance on a single LUN. This is a caching feature of the HP SMART Array 5i Controller contained within the MSA1000 disk subsystem.

[2] Result indicates the single file recovery was located toward the end of the media and hence took longer to search to. The maximum locate time is about 140 seconds.

Observations

The secret to utilizing the performance of an HP StorageWorks Ultrium 960 is to source fast data streams and then interleave them on the tape. Fast data streams generally mean reading from separate LUNs, and if the LUNS are on separate disk arrays, even better. In this example are four LUNs on four separate arrays giving an overall throughput of more than 374 GB/hr (104 MB/sec).

However, you will notice that the overall throughput for four LUNs on four separate disk arrays is not that much faster than four LUNs on the two disk arrays shown in the Direct attach SCSI Disk scenario, which only utilizes 14 disks. Why? The answer is that the bottleneck in performance in this scenario resides in the MSA1000 controller, which is Ultra160 based, hence even though there are more disks and more arrays, the bandwidth of the subsystem is limited to 160 MB/sec. Even though the disk drives are Ultra320 capable, the subsystem is forced to Ultra160 speeds. The MSA1000 controller will be upgraded to be Ultra320 capable during 2005.

With 14 (2 drive) LUNs configured, the HPreaddata utility showed the array could supply data through the file system at 159 MB/sec when reading 14 streams. When the number of streams was reduced to eight, the read throughput was 154 MB/sec, indicating a throughput bottleneck very close to 160 MB/sec, the MSA1000 controller limit.

SAN using EVA

- **Disk group:** The number of disks that act as an allocation pool to provide various RAID level virtual disks (LUNS).
- **Virtual disk:** A distinct logical disk that can be presented to a host.
- **RAID type:** The types of virtual disks created can be RAID 0 or RAID 1 or RAID 5. RAID 0 was used in these tests.
- **Disk leveling:** Performance and capacity can be grown "online" by the addition of extra disks and virtual disk leveling as shown in Figure 8.

Figure 7.

Virtualization on the EVA

Host

Lun

SAN

Physical disk are pooled in Disk Groups

Virtual Disks are carved out of a Disk Group

A Virtual Disk is presented to a host as a Lun

Enterprise Virtual Array

Virtual Disk

Disk Group

11/9/2004

page 13

Figure 8.

Virtual Disk Leveling

Symptoms

- Disk spindles becoming a throughput bottleneck
- No additional storage space available

…Add more disks

VRAID 0
VRAID 5
VRAID 1*

+

Logical

Low Redundancy Virtual Disk
Med Redundancy Virtual Disk
High Redundancy Virtual Disk

- Disks Running at Optimum Throughput (dynamic load balancing)
- Additional Storage Space Available

Free space (virtual space = 3 Disk)
VRAID 0
VRAID 5
VRAID 1*

*RAID V1 uses even numbers of disks

11/9/2004

page 16

Figure 9.

SAN BACKUP with EVA Disk Array

EVA 5000 with 48 Disks, Single Disk Group. VRaid0.
Clients 1,2,3 have 4 LUNs each allocated to them.
HP "typical" file system loaded

HP Data Protector
5.5 Backup
Software
Disk and Media
Agents operating
on Clients 1,2 & 3.

HP Secure Path - balances the load
across all 4 available fibre channel
connections

2 Gigabit Fibre channel

DL380G4 Client 1
3.4GHz, 1GB RAM

DL380G3 Client 2
2.8GHz, 512MB RAM

ML370G4 Client 3
2 way 3.4GHz, 1GB RAM
+ Data Protector Cell manager

Ultra320 SCSI

HP StorageWorks Ultrium960
Tape Drives

This scenario shows what kind of backup performance can be expected in a true enterprise environment. The scenario simulates the type of performance that will be achievable with HP StorageWorks Ultrium960 tape libraries early in 2005.

Detailed configuration description

The EVA5000 is one of the HP enterprise class disk arrays. For these tests, a 2 controller 6 shelf (2C6D) model with 48 X 72 GB 2-Gb/sec Fibre Channel disk drives installed was used. The EVA5000 performs best in terms of IOPs when the intensity level (number of requests for I/O) is high, hence this test is run with multiple servers accessing the EVA5000 simultaneously. The EVA also performs best when all the disks are in a single disk group, so in this example the disk group consists of 48 disks. Twelve LUNs (Vdisks) are created on the array and presented to the three client servers (four to each server). The RAID level applied to these LUNs is VRaid 0 to get the best performance out of the available drives. RAID 5 would be used in practice but would require more drives to achieve the same performance levels because of the parity overhead.

Figure 10.
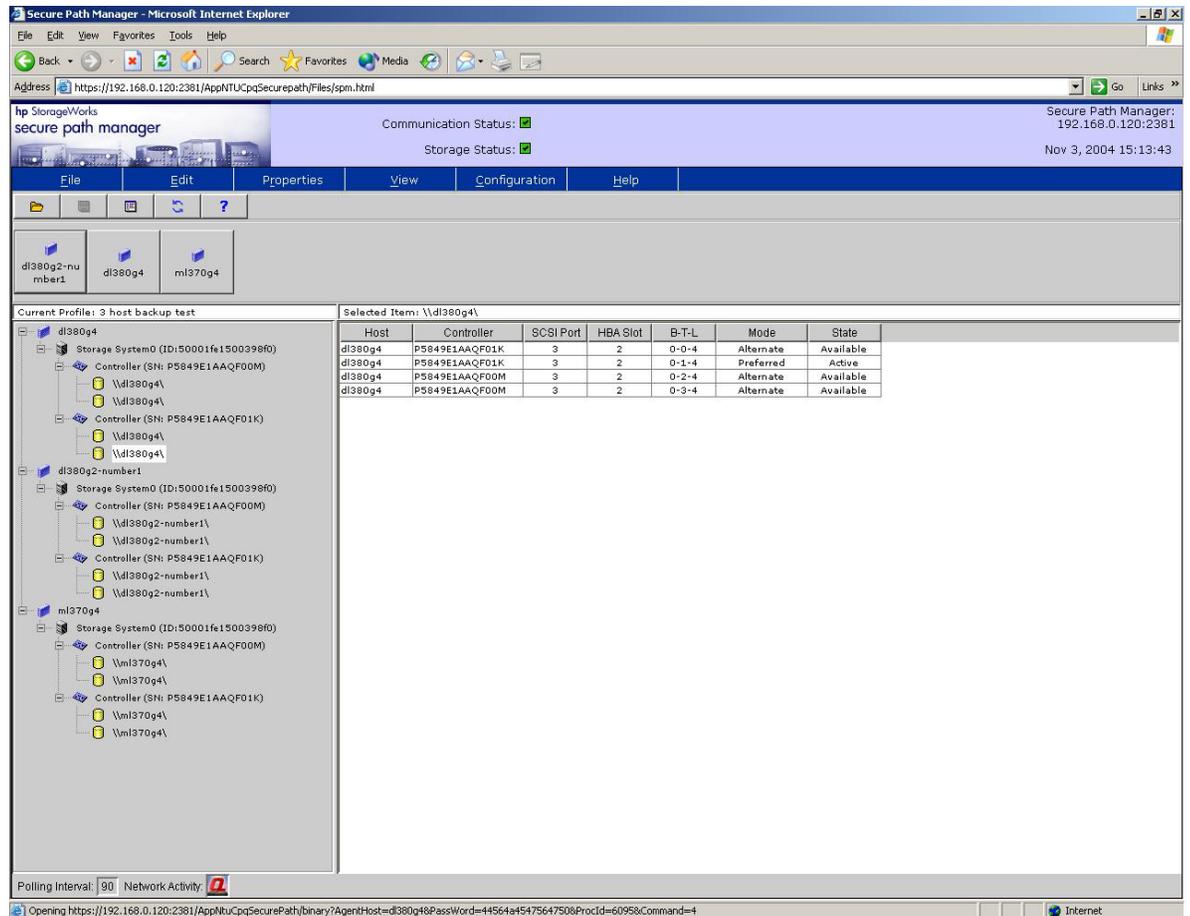
For best performance from the EVA5000 it is also required to dynamically load balance the available fibre outputs across the available servers using HP StorageWorks Secure Path Manager. This enables consistency (binding) each time a server is re-started because it binds a particular LUN on the disk array to a particular server using a secure (preferred) path. The other paths are still available in case of a particular path failing. Figure 10 shows how LUN4 on the DL380 G4 server is bound to Bus/Target/LUN (B/T/L) 0-1-4; the other three LUNs are bound to the other three paths. The HP typical file system is installed on each of the 12 LUNs.

On the backup side, HP OpenView Storage Data Protector 5.5 is installed and the ML370 G4 server is used as the Cell Manager (database repository). HP OpenView Storage Data Protector disk and media agents are installed on Clients 1 and 2 and also operate on Client 3 so that these clients can read from disk and write to tape autonomously, but under the monitoring control of the Cell Manager (ML370G4).

During the backup, the four LUNs for each client are interleaved on the same physical media (concurrency level = 4).

HP OpenView Storage Data Protector offers another important feature—automatic load balancing. This feature ensures that the backup job is evenly distributed across available tape drives, ensuring maximum performance and the lowest possible backup window.

The three clients are:

- ProLiant DL380 G4 3.4 GHz, 1.0-GB RAM
- ProLiant DL380 G3 2.78 GHz 512-MB RAM
- ProLiant ML370 G4 2x 3.4 GHz 1.0-GB RAM

The throughput was also measured on the fabric switch.

Ports 1–4 were the EVA ports. Ports 10, 11, and 12 were the Client ports.

Results

Table 4.

Test	Configuration	Client 1 (DL380 G4) MB/sec	Client 2 (DL380 G3) MB/sec	Client 3 (2-way ML370 G4) MB/sec	Total throughput MB/sec	Disk FC ports (4) MB/sec	Tape FC ports (3) MB/sec
Throughput (MB/sec)	Back up all 12 LUNs simultaneously to 3 tape drives. (4 LUNs per drive)	146.8	118.48	138.57	403.85 (1.453 TB/hr)	Average 110 per port	Average 145 per port
Processor loading %		29 to 47	57 to 76	20 to 30			
	Restore time in min/sec	**Throughput**					
Restore LUN 1 for Client 1 with full overwrite. (4 LUNs interleaved on a tape)	9 min 11 sec	40 MB/sec					
Restore single 64-K file for Client 1	95 sec	64 KB/sec					

Observations

The results were very impressive—a cumulative backup throughput of 403.85 MB/sec (1.45 TB/hr) from just three tape drives.

EVA5000 performance in terms of MB/sec peaks at about 525 MB/sec with a fully loaded EVA, so it is unlikely that more than four LTO3 drives would be required in terms of optimal backup performance. However, more than four LTO3 drives may be required to back up the higher volumes of data simultaneously.

If one analyzes the amount of client processing power that is used during this backup, it roughly equates to 1.5 GHz of processing power required to read four LUNs from the disk array and write them to tape. A 3.4-GHz machine would only be capable of sustaining two HP StorageWorks Ultrium 960 tape drives at best performance, and this would consume most of the server bandwidth. A single LTO3 tape drive allocated to a 3.4-GHz server would leave 50% of the processing power still available for functional use during backup. There would seem to be a case for allocating no more than one LTO3 tape drive to a server, unless that server is a multi-processor. Utilizing the server for online activities while performing a backup may also reduce potential backup throughput if the online activity is counter to typical backup I/Os occurring.

Clients 1 and 3 have reasonable memory provision, but Client 2 has only 512 MB. The reduced performance on Client 2 is due to insufficient memory. A further set of tests will prove this conclusively.

Restore performance is also good: a single LUN being restored from a concurrency tape of value 4 at a very acceptable 40 MB/sec. The restore speed could be improved by backing up only one LUN per tape, but then backup performance suffers.

Single file restore for a file early in the directory tree is good at 95 seconds.

Sizing servers for use in SAN-based backup

One of the key questions nearline solution architects always ask is how much processing power and RAM do servers used in a SAN-based backup environment need to read from a disk array and write to tape, while still having processing power to spare to run its prime function (database, file print, web server, and so on). Some additional tests were run with one of the previously mentioned servers to try to get a "rule of thumb" on this server sizing issue. The results shown in Table 5 are for various processor and memory configurations of the 2.8-GHz DL380 G3 server.

Results

Table 5.

Total throughput for 4 LUN backup using DL380 G3 MB/sec	Processor loading range %	Estimated average processor load %	Number of processors	Memory installed (MB)
112.2	43–62	52	1	512
136.71	65–79	72	1	1024
2 tape drive backup 2 LUNS to each tape drive				
Drive 0 = 83.33	86–95	90	1	1024
Drive 1 = 86.98				
Total = 170.31				
4 LUNs				
94.45*	34–45	40	2	512
Re-run 92.74*	34–45	40	2	512
4 LUNS				
140.08	38–48	45	2	1024
2 tape drive backup 2 LUNS to each tape drive				
Drive 0 = 72.63	48–59	54	2	1024
Drive 1 = 75.17	48–59	54	2	1024
Total = 147.8				

* This result. is explained by the two processors being starved of memory.

Observations

Working from the fact that a 2.8-GHz process with 1.00-GB RAM uses 72% of its processing power to read from disk and write to tape equates to 2.8 x .72 = 2.00 GHz being required for this operation.

To drive an LTO3 in a SAN-based backup environment requires 2-GHz processing power and 1.00-GB RAM and any processing power required for other operations should be built in on top of this requirement.

Two processors and 1024 MB of memory cannot run the drives as hard as one processor and 1024 MB of memory. As the number of processors increases, the amount of memory required should also be increased appropriately.

Test thesis with another test point

Two drives take 54% processor load from 5.6-GHz processing power and 1.00-GB RAM.

$$5.6 \times .54 = 3.02$$

The GHz processing power required for two drives is approximately 1.5 GHz per drive.

Summary

In a SAN environment at least 1.5 to 2.00 GHz of processing power and 1.00-GB RAM per processor is required to drive an LTO3 tape drive over separate disk array and tape drive connections.

OTN backup

Gigabit Ethernet (Gbe) has a theoretical maximum transfer rate of around 120 MB/sec, so with a single HP LTO3 tape drive being capable of up to 160 MB/sec, there is a strong possibility that a single Gbe connection can become the bottleneck to backup performance. The tests were designed to use a high-performance disk array to tax the network infrastructure. The EVA5000 in this configuration was capable of supplying data at more than 200 MB/sec when reading four LUNs.

Some customers readily consolidate disk storage but still use OTN backup because they will not deploy a tape SAN. These tests will show the limitations of this approach.

Where the clients being backed up over a network have much lower performance disk subsystems then large-scale interleaving would need to be deployed to get the best performance from the HP StorageWorks Ultrium 960 tape drive.

The real issue is that network backup no longer scales in performance. It can seem to offer a cost-effective backup solution because it uses existing network infrastructure and most servers come with one or two Gbe ports today, but you must add in the cost of the network backup server and the processing power it requires to handle multiple TCP/IP stacks. The result is slow backups in most cases. Migrating to a SAN-based backup scenario gives far better scalability for the cost of a Fibre Channel HBA and a switch port associated with each SAN client.

Several tests were run, including:

- 12 LUNs to a single HP LTO3 tape drive
- 12 LUNs to 3 x HP LTO3 tape drives (four LUNs to each drive)
- 3 LUNs (one from each client) to a single HP LTO3 tape drive
- 3 LUNs (one from each client) to a 3 x HP LTO 2 tape drives (HP StorageWorks Ultrium 460)

For each test the processor loading of the clients, the processor loadings of the backup server, and the network load balancing were observed. In some tests, two of the backup server NICs were active, and in others all three were active, which is a result of the automated network load balancing.

The overall backup throughput was measured as reported by HP OpenView Storage Data Protector.

In addition to the physical test setup, Windows Network Load Balancing was enabled across the three Gbe ports on the DL580 server. Windows Interrupt Affinity utility was used to bind each of the three NIC cards in the DL580 server to a given processor in the DL580 backup server, so that the processing power was as balanced as possible.

Finally HP OpenView Storage Data Protector load balancing was deployed in the case where the three tape drives were run.

Detailed configuration description

A 4-way DL580 G2 server with 4 x 2.8-GHz processors is used as the backup server. Installed in the DL580 G2 server are 2.00 GB of RAM. The DL580 G2 server has three Gbe connections into the Gbe switch.

NIC7170 Gbe cards or their embedded equivalent are used throughout. The DL580 server has three separate PCI buses and a total of six PCI-X 100-MHz x 64-bit slots. The Dual NIC cards are on separate PCI busses and the three tape drives are driven from two Dual Ultra320 controllers on the third PCI bus. Redundant/adaptive load balancing across the three NICs is also deployed.

The data source is the same LUNs used in the EVA5000 tests and contains HP typical file system data at 2:1 compression. These groups of four LUNs (per client) can be read at over 200 MB/sec from the EVA5000 with HPreaddata. The processor and network loadings on the DL580 server were plotted using Windows Performance monitor.

Figure 11.

Over the Network (OTN)

EVA 5000 with 48 Disks, Single Disk Group, VRaid0 LUNs. Clients 1,2,3 have 4 LUNs each allocated to them. HP "typical" file system loaded

HP Data Protector 5.5 Backup Software

HP data Protector Disk Agents only installed on Clients 1.2,3 to ensure Backup takes place over the Network.

LUNS1-4

LUNS5-8

2 Gigabit Fibre channel

LUNS9-12

DL380G4 Client 1
Single Processor
3.4GHz, 1.00GB RAM

DL380G3 Client 2
Single Processor
2.8GHz, 512MB RAM

Gigabit Ethernet

ML370G4 Client 3
Dual Processor 3.4GHz,
1.00GB RAM

1000TX

Ultra320 SCSI

HP StorageWorks Ultrium960 Tape Drives

DL580
4 way 2.8 GHz,
2.00GB RAM
Backup Server

Results

Table 6.

Test type	Overall total throughput MB/sec	Average processor loading% DL380 G4 server	Average processor loading% DL380 G3 server	Average processor loading% ML370 G4 server	4-way processor loadings % DL580 server
Back up 12 LUNs to 1 LTO3 tape drive	122 (conc =12)	100	100	92 (2 procs)	1 processor at 90% servicing NIC associated with ML370 server 2nd Processor at 50% load handling clients 1 & 2 2 NICs in DL580 server active
Back up 12 LUNs to 3 LTO3 tape drives 4 LUNs per tape	145 (conc = 4)	100	100	98 (2 procs)	1 proc at 90% servicing NIC associated with ML370 server 1 processor at 64% loading 1 processor at 56% loading 3 NICs in DL580 server active 1 at 80% loading 2 at 35% loading
Restore type	**Time**	**Throughput**			
Restore single LUN (DL380 G4 LUN1) from 12 LUN tape	25 min 43 sec	14.2 MB/sec			
Restore single 64-K file to DL380 G4 LUN1 from 12 LUN tape	238 sec	64 KB/sec			
Restore single LUN (DL380 G4 LUN1) from 4 LUN tape	9 min 9 sec	40.08 MB/sec			
Restore single 64-K file from 4 LUN tape	70 sec	64 KB/sec			

Observations

The task of transferring data over Gbe is a very processor-intensive operation. Even the 2-way 3.4-GHz ML370 G4 server has reached high processing load during these tests. The reason is that each of the clients is running four Storage Data Protector disk agents (vdba)—one for each LUN being accessed from the EVA5000 and backed up. The disk agent is also responsible for pushing that data over the Gbe connection. When observing the vdba processor loading in the clients, each disk agent took about 23% processor loading on average, hence the saturation effect.

To get the most accurate picture of network loading it is best to measure the individual processors loading through performance monitor as shown in Figure 12.

From these results it seems there is little overall benefit in adding additional tape drives because the throughput is not increasing significantly, being limited by the client's processing power in accessing four high-performance LUNs each.

It is unlikely that customers would invest in additional processing power in production servers solely for the purpose of network backup.

Figure 12 shows the network and processor loading for the 12 LUN backup to a single tape drive. The solid lines represent the network throughput on each of the three network channels; the dotted lines indicate the individual processor loadings for the four processors in the DL580 server.

Processor 0 (purple) has little load and is running the base HP OpenView Storage Data Protector functions. Processor 1 has no load because it is assigned to a NIC card that is not deployed at present because of the automatic load balancing that is in operation.

Figure 12.
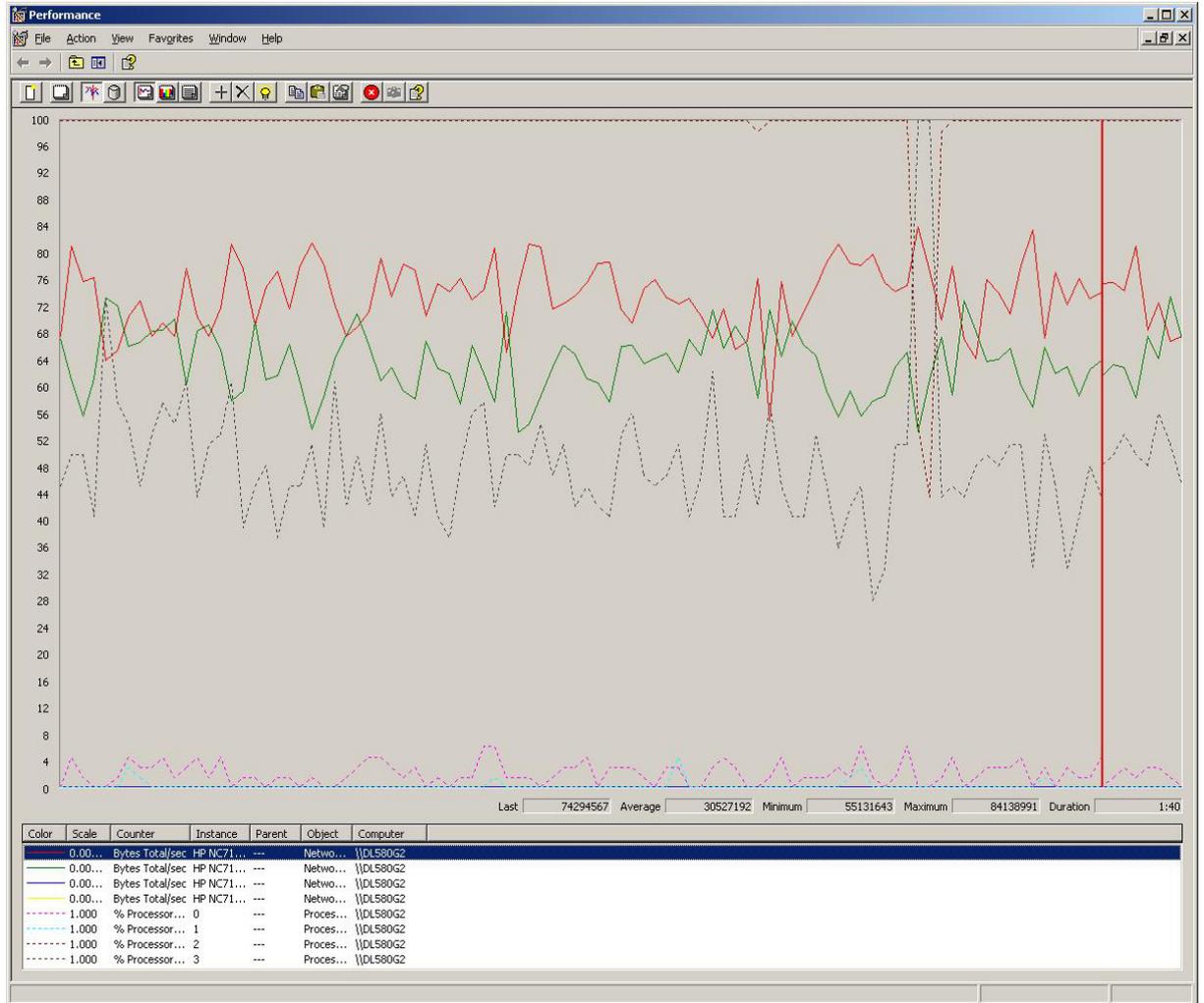
Two solid lines (red and green) indicate that the network load balancing has used two NICs running at around about 65% and 75% loading each. Processor 2 in the DL580 server is running at 90–100% since this is receiving data from the 2-way ML370 G4 server.

In Figure 13, looking at the DL580 server loading during backup to three LTO3 tape drives, you can see three solid lines (red, blue, green) indicating the three NIC cards have been deployed. Blue is running at 85% load (associated with ML370 G4 server) and the red and green networks are running at about 35% load. The associated processing load matches the network loads, with one DL580 processor (associated with ML370 G4 NIC) at 90–100%, and the other two running about 60% load. For all this extra activity, there is only an 18% increase in throughput because it is the client's read I/O to the LUNs on the EVA5000 that cannot take place any faster and is being limited by processing power.

Figure 13.
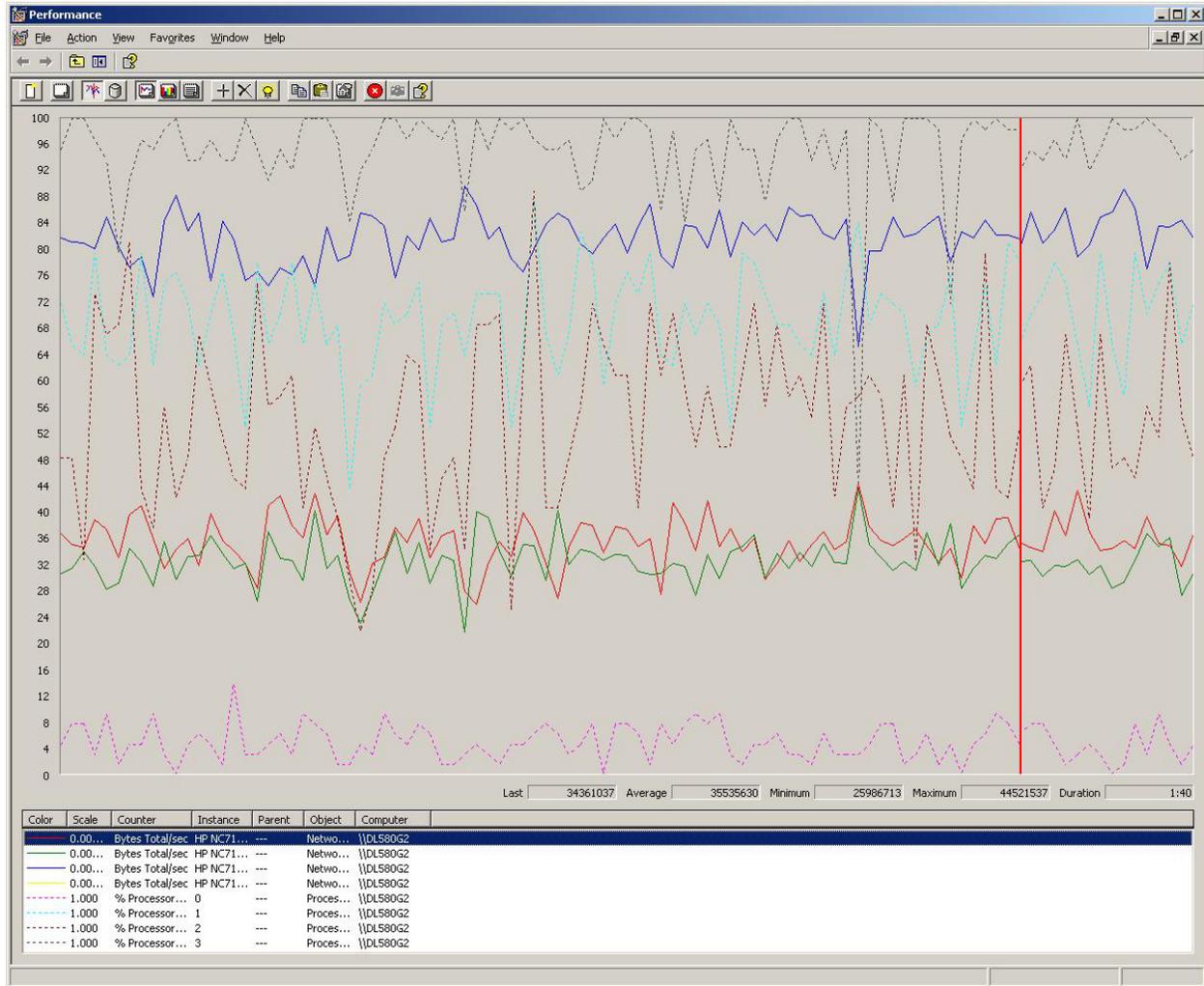
Some additional tests were run to see how much processing load could be reduced by if only one LUN (one vdba process) was run on the clients (see Table 7). Finally the tests were re-run using HP LTO2 tape drives.

HP LTO2 tape drives have a native transfer rate of 30 MB/sec and a native capacity of 200 GB.

As predicted, the client processor loading became much more reasonable when only a single LUN from the EVA is backed up at once (instead of four), and was reduced even further when 3 x HP LTO2 drives were used. Using three LTO2 drives also did not invoke the third NIC card in the load balancing.

Table 7.

Test type	Overall throughput MB/sec	Average processor loading% DL380 G4 server	Average processor loading% DL380 G3 server	Average processor loading% ML370 G4 server	4-way processor loading% DL580 server
Back up 3 single LUNs to 1 LT03 tape drive	120 (conc =3)	64	75	35	1 proc at 90% servicing NIC associated with ML370 server. 2nd Processor at 60% loading 2 NICs in DL580 server active
Back up 3 single LUNs to 3 x LTO2 tape drives 1 LUN per tape	127 (conc =1)	55	70	35	1 at 52% 1 at 48% 1 at 68% 2 NICs in DL580 server active

In terms of processor loading in the clients, and overall network backup throughput performance, 3 x HP LTO2 drives offer a better solution than 1 x HP LTO3 drive and can provide 50% more aggregate capacity, but at a higher overall cost.

Summary

1. Using a Gbe infrastructure to back up high-performance disk subsystems with LTO3 is not recommended. Not only does it put a very high processing load on the client servers but it does not make best use of LTO3 performance. A SAN-based tape system delivers better scalable performance, providing the client machines allocate 1.5–2.00 GHz and 1.00-GB RAM per processor for the backup process.
2. Gbe infrastructure is best suited for backing up systems with slower disk performance and then interleaving them on tape to make best use of the available tape performance. Earlier generations of HP LTO technology are ideally suited to this type of environment.
3. If your disk subsystem is high performance do not try to back it up over the network, you will not get the best backup performance that is possible.

Improving OTN backup performance

1. Reduce TCP/IP Receive processor loading by using TCP/IP Offload Engines (TOES). See Appendix B for some testing HP has performed in this area.
2. Back up slow hosts over the network to a lower cost secondary disk array. This process may take some time but when complete, restores can be made directly from the secondary disk (if data still resides there) and a full backup to tape can take place at faster rates from the consolidated image on secondary disk storage than it can by going direct to tape over the network. The impact on the client's processing power is also significantly reduced as the backup to lower cost secondary disk storage takes place over the network. Many ISVs are now recommending backup to secondary disk and then to tape as a means of staging backups on the most appropriate medium to optimize infrastructure costs and performance.

Disk to Disk to Tape (D2D2T)

With the high capacity and lower cost offered by SATA and Fibre Channel ATA (FATA) disk technologies, many customers are now considering implementing backup to low cost disk arrays before backup to tape.

The use of secondary disk arrays for backup is best suited to environments where:

- The business dictates rapid single file restore capabilities (seconds to minutes). It is generally quicker to restore from a secondary disk subsystem than it is from tape (minutes to hours). However, the secondary storage array is not infinite in capacity and only the most recent backups may still reside on the disk array. In addition, disk to disk backup is no substitute for off-siting media.
- The hosts can only supply data at a relatively modest rate (10–20 Mb/sec). The backup image can be gradually produced on the secondary array without having to interleave multiple streams to tape and without the need for a large powerful dedicated backup server.
- Small file backup to tape has always been a performance limiter. With backup to disk, a complete backup image of small files can be constructed and then passed to tape at much higher transfer rates than if the small files were transferred directly to tape.

Table. 8.

Feature	SATA disks	SCSI disks
Mean time to failure (MTTF)	500,000 hrs @ 20% duty cycle	1,200,000 hrs @ 100% duty cycle
Burst transfer rate	150 MB/sec	320 MB/sec
RPM	7.2K	10 K or 15 K
Queuing	Non-Tagged—serial execution	Tagged—optimized seeks–better performance
Capacity	250 GB	146 GB
Warranty	1 year	3 years

While the SCSI disk MTTF looks high, this figure decreases when there are many disk drives bound together in an array. By comparison have an MTTF of 250,000 at 100% duty cycle but are used in smaller volumes within an automated tape library. Therefore, do not assume that disk is automatically more reliable than tape.

The key point is D2D2T has its place and with proper administration can improve the data protection process but it is not a replacement for tape. Tape is still the foundation of a robust data protection strategy.

HP OpenView Storage Data Protector 5.5 provides a comprehensive implementation of D2D2T called "advanced backup to disk" (Figure 14).

Figure 14.

Advanced Backup to Disk
Disk Staging Example

Disk Agent

Media Agent

Disk Agent

Stream to tape

Backup

Disk Agent

Tape

NEW:
direct restore from tape

Restore:
Fast restore from disk if
data still available there

Backup 1: Store on disk
and copy to tape on
scheduled basis; expire
backup on disk as disk fills
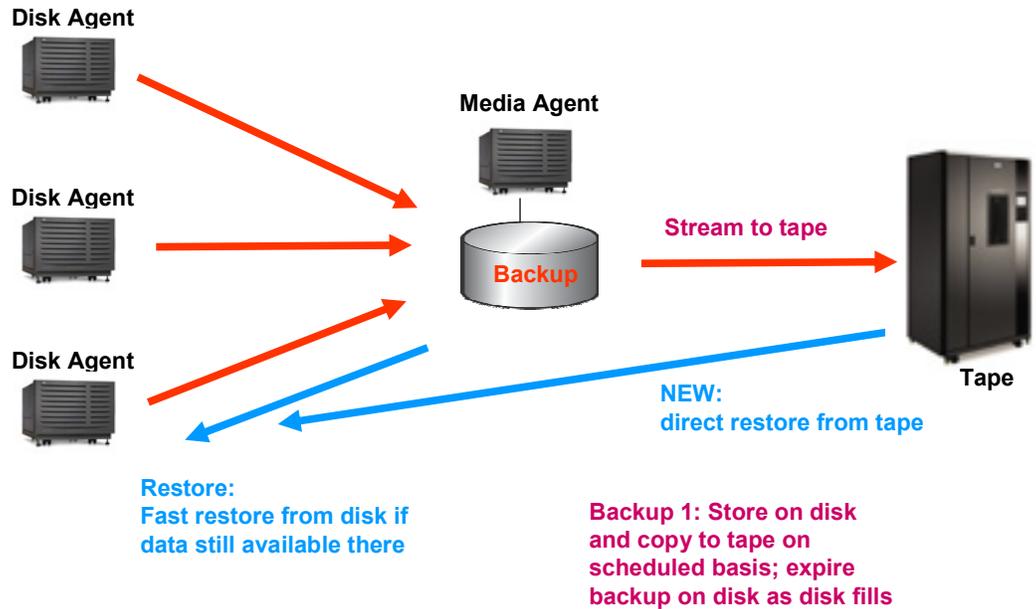
Features

- Automatically selects file from disk stage for restore, if available
- Single file restore directly from tape, if not on disk
- Automatic space management
  - Space creation
  - Retention management
    ->automatic space re-use
- Restores and backups in parallel
- GUI Wizard for creation
- Disk array independent

Use cases

- Continuous backup of transaction log files (tape drives not in start/stop mode)
- Backup of slow clients without multiplexing
- Resource sharing
- Virtual tape library
- Tapeless branch office backup
- Improving small file backup performance.

HP OpenView Storage Data Protector provides a variety of disk-based data protection alternatives (Figure 15).

Figure 15. ZDB = Zero Downtime Backup IR = Instant Recovery

Positioning backup to disk technologies

	ZDB / IR Replication	Advanced Backup to Disk Virtual Tape, Disk Stage D2D2T	File Device Virtual Tape
	t0 t1 t2		
How?	mirror, snapshots etc	backup to disk	backup to disk
Instant Recovery	yes	no	no
On-line Application b.	yes	yes	yes
Easy single file restore	no	yes	no
Easy setup	yes	yes	no
Incremental backup	no	yes	yes
Cheap disk	no	yes	yes
Operating systems	HP-UX, Win, Linux, Solaris	all	all
Licensing	per TB	per TB	Per file device "drive"
Use case	7* 24 DBs, appl	many	limited
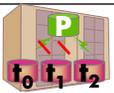
November 16, 2004   HP Confidential   11

Figure 16.

Staged Backup using HP MSA1500 SATA Disk Array

MSA1000 Primary Disk Array
with 14 x 72GB/15K Disk
Drives, Configured as RAID0
4 LUNS configured
Single Fibre connection to the
Array

HP typical filesystem installed
HP small file test data installed

BACKUP TO A
SECONDARY
ARRAY

Low Cost SATA based disk array
MSA1500 with 12 x 250GB/7.2K rpm
SATA Disk Drives.
Configured as RAID0
4 LUNS configured

2 Gigabit Fibre channel

HP ProLiant DL380 G2
1.26GHz, 2.00GB RAM

BACKUP TO
TAPE FROM
SECONDARY
ARRAY

HP Data Protector
5.5 Backup Software
with advanced
backup to disk
licence

Ultra320 SCSI

HP StorageWorks Ultrium960
Tape Drive

Detailed configuration description

Both the MSA1000 and HP StorageWorks Modular Smart Array 1500 (MSA1500) were configured as RAID 0 and had four LUNs.

The source data resides on the MSA1000 and consists of two sets of data.

- HP typical file system consisting of 64-KB to 64-MB file sizes—around 20 GB on each of the four LUNs.
- Small file test data file sizes 4 KB to 64 KB—around 20 GB (859,320 files in 3,905 folders)—on a single LUN only. This data is for an additional test to demonstrate a particular usage model of the secondary storage array.

Both arrays are connected to the server by a single Fibre Channel link.

It is important to set expectations about the SATA disk array performance. It is not designed for primary online storage because of the drive reliability and performance specifications.

The HP OpenView Storage Data Protector Advanced Backup to Disk license enables the MSA1500 to be configured as a "file library," so the initial backup from MSA1000 to MSA1500 takes place as a normal backup job with the file library being specified as the destination. The file library can be configured in much the same way as a tape device, with block size, segment count, disk buffers, and so on. It is important to realize that the backup file created on the MSA1500 is to all intent and purpose a tape image file. The benefit of this will be seen later when examining small file backup performance.

Using HP OpenView Storage Data Protector Advanced Backup to Disk, the migration of the backup on the MSA1500 to tape is actioned by means of a "copy" style function. It can be interactive (manual) or scheduled to occur at a certain time or even directly after the initial backup to the MSA1500 has completed.

Results

Table 9.

Data set	Backup transfer type	Transfer rate MB/sec	Processor loading average %
HP typical file system 64 K–64 MB	Single LUN on MSA1000 to HP LTO3 directly	47.95	34
HP typical file system 64 K–64 MB	Single LUN on MSA1000 to MSA1500	38.68	64
HP typical file system 64 K–64 MB	Single LUN on MSA1500 to HP LTO3	34.25	49
Restore tests			
HP typical file system 64 K–64 MB	Single LUN MSA1500 to MSA1000	33.1	53
HP typical file system 64 K–64 MB	Single File 64-K file from MSA1500 to MSA1000	8 seconds	N/A
HP typical file system 64 K–64 MB	Single LUN from HP LTO3 to MSA1000	87.3	97
HP typical file system 64 K–64 MB	Single 64-K file from LTO3 to MSA1000	45 seconds after tape loaded	N/A
Small file tests			
HP small file test data 4 K–64 K	MSA1000 to HP LTO3 directly	12.6	39
HP small file test data 4 K–64 K	MSA1000 to MSA1500	11.0	48
HP small file test data 4 K–64 K	MSA1500 to HP LTO3	33.1	38
Multi-LUN Backup tests			
HP typical file system 64 K–64 MB on all 4 LUNS	4 LUNS from MSA1000 to MSA1500	53.44	95
HP typical file system 64 K–64 MB on all 4 LUNS	4 LUNS from MSA1500 to interleaved onto HP LTO3	44.77	47

Observations

For the HP typical file system backup to the tape drive directly is faster than transfer to secondary disk and then to tape.

Restoring a total LUN from HP StorageWorks Ultrium 960 tape is faster than restoring it from secondary disk array (87 MB/sec versus 33 MB/sec). For single file restore, a secondary disk storage mechanism is ideal. However, for full system/LUN streaming restores, the raw capability of the Ultrium 960 will work in favor of faster restores. It is important to design your use of disk technology for data protection based on your protection and restoration needs.

Single file restore from the MSA1500 is faster than from HP StorageWorks Ultrium 960 tape (8 sec versus tape load plus 45 sec).

The transfer rate from the MSA1500 is above the low end of the HP LTO3 data rate matching range, so the tape drive will continue streaming and provide optimal backup performance and capacity to match the secondary disk array performance.

The processor used in this test is only 1.26 GHz (DL380 G2 server), so in more modern machines the processing loads will be much less than those recorded here.

There is almost a 3x performance improvement when backing up small files from the MSA1500 as opposed to backing them up directly from source because when backing up from source, each file in the tree structure must be opened and closed. This presents a huge overhead, especially if the file is only 4 KB in size. The image on the MSA1500 is a single large tape image and when transferred to tape needs only one open and close operation; the rest of the time is spent transferring data.

There does not seem to be an increase in backup performance to tape if interleaving multiple sources from the MSA1500 on tape, which is thought to be due to read performance limitations within the MSA1500.

Table 10 shows a summary of when to use secondary disk storage for staging backups to tape.

Table 10.

Best use of secondary disk	Best use of HP StorageWorks Ultrium 960 tape
Random single file restores (if data still on secondary disk)	Full LUN restores
Concatenating OTN backups from slow hosts into a tape image on disk	High-capacity backups direct from high-performance disk arrays
Concatenating small files on disk before backup to tape.	Full system recovery.
Doing backup and restore in parallel	

ISV backup application tuning

Terminology

- **Block size**—The physical size of a block of data on tape. Every block can be a fixed size, for example, 64 KB (fixed mode), or the block size can vary for every transfer (variable mode).
- **Transfer size**–The data transferred in a single SCSI write or read command. The larger the transfer size, the more efficient the data is transferred. For example, a transfer size of 640 K is better than 10 transfers of 64 K because each of the smaller transfers has a command and other SCSI phase overheads associated with it.

In Windows, a transfer size limit of 64 KB exists unless the MaximumSGList parameter for the HBA is changed in the registry. Similar restrictions apply in other operating systems.

What makes life difficult is that some ISVs refer to block size when they really mean transfer size, so you must be careful. If the transfer size is large enough (128 KB minimum), then the tape block size is not an impediment to performance. The HP Ultrium drives have no dependency on tape block size limiting performance, so as long the transfer size is appropriately large, the drive will deliver good performance.

HP strongly recommends that for HP LTO3, the tape block size be set to a minimum of 64 KB and the transfer size be set to a minimum of 128 KB. With some ISV packages the only way to change the transfer size is to change the block size, in which case HP recommends 128 K or 256 K for best performance.

Tuning recommendations for the major backup ISV vendors

HP OpenView Storage Data Protector

A free download of the single server version of HP OpenView Storage Data Protector 5.5 ships with every HP StorageWorks Ultrium 960 tape drive. Several backup software suppliers require you to purchase (at extra cost) additional modules (typically called "Shared/SAN Storage" or similar) to use tape drives in a SAN-based configuration as previously shown. Storage Data Protector does not require additional modules to be purchased and represents exceptional value for the money.

The main performance tuning parameters are under the Devices tab in the dropdown list of functions.

1. Under Devices, right-click the tape device you want to tune.
2. Click **Properties.**
3. Select the **Settings** tab.
4. Click **Advanced.**

Under the Settings tab you can set the concurrency value of the tape drive, that is, how many streams it is allowed to interleave on a single tape. This is used to improve backup performance when there are several slow hosts supplying data.

The concurrency level should be set to maximize the performance of your tape drive and reduce the backup window. However, for very high levels of concurrency the restore performance can suffer severely. A concurrency value of about 12 is a good optimal starting point for HP StorageWorks Ultrium 960.

Under the Sizes tab you can set:

- **Block size** for the format on tape (recommend 256 KB for HP LTO3)—Sets transfer size.
- **Segment size**—Is the space between filemarks. Writing filemarks causes the data buffers to flush to tape and can impede performance if written too often, but filemarks are used to segment data on tape to improve fast search restore performance. A recommended typical setting is 4000 MB or 8000 MB. This means a filemark is written every 4 GB or 8 GB of data written to tape.
- **Disk agent buffers**—Is the number of buffers used on remote clients when data is read from the disk subsystems, generally used for streaming slower tape drives. Use default setting for HP LTO3.

Each time a backup is run under Storage Data Protector, the "network load" prompt is displayed.

For SAN-based backup where data passes directly from the client to the tape device, the network load is not important (since only communication data goes by way of the LAN). However, when performing OTN backup, the load prompt of selecting high/medium/low network load determines how much extra load the backup will place on the LAN. For faster backups select the high load option.

- **Match tape drive block size to file library block size**—When using a secondary subsystem as a file device (file library) for backups, it is important to give it the same block size as the tape device that will be used in the final part of the data protection process. So in the case of HP StorageWorks Ultrium 960 tape drive, this would be 256 KB. Failure to match the file library block size to the tape block size results in additional conversion overheads that will reduce backup performance to tape.

VERITAS Backup Exec

- Primarily used on Windows platforms
- Under Devices, right-click for the drive properties, then select the **Configuration** tab.
- Parameters are available for tuning include:
  - **Block size**—Recommended setting is 256 KB for HP LTO3.
  - **Buffer Size**—The default is the preferred amount of data sent to the drive on each read or write command (also known as the transfer size). The buffer size must be an even multiple of the block size. Depending on the amount of memory in your system, increasing this value might improve drive performance. The parameter change to buffer size greater than 64 KB on Windows can only take effect if there is also a registry change to a parameter called "MaximumSGList," or Maximum scatter/gather List. This must be made for the HBA, which is connected to the HP Ultrium tape drive.

    Use regedit from the run command line to start the registry editor.

    Navigate the directory tree to

    \\HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\adpu160m\Parameters\ DeviceX\MaximumSGList

adpu160m is an example of an Adaptec 29160 HBA registry entry configured into the system.

The REG_DWORD MaximumSGList may not exist for a given HBA because the default is being used, so it may be necessary to create a NEW key from regedit.

The formula for calculating the value of MaximumSGList is as follows:

MaximumSGList = (Tape Blocksize (bytes)/4096) + 1

Or, set MaximumSGList to a value of ff (hex) if in doubt.

- **Buffer Count**—The default is the preferred number of buffers allocated for this device. Depending on the amount of memory in your system, increasing this value may improve drive performance. Each type of drive requires a different number of buffers to achieve maximum throughput. If you change the buffer count, you might need to adjust the high water count accordingly.
- **High Water Count**—The default is the preferred number of buffers to be filled before data is first sent to the drive and any time after that if the drive underruns. The high water count cannot exceed the buffer count. A value of 0 disables the use of high water logic, that is, each buffer is sent to the drive as it is filled. The default setting provides satisfactory performance in most instances. In some configurations, throughput performance may be increased when other values are specified in this field. If you increase or decrease the buffer count, the high water count should be adjusted accordingly. If a drive has a high water count default of 0, it should be left at 0.
- **Interleaving (Multiplexing/concurrency)**—Unfortunately VERITAS Backup Exec does not support interleaving, which could limit the available performance when using the HP StorageWorks Ultrium 960 tape drive.

VERITAS NetBackup

Multiplexing

Multiplexing must be set in two places in NetBackup:

- Storage unit—Maximum multiplexing per drive (1–32).
- Schedule—Specifies the maximum number of backups from the schedule that you can multiplex on a single drive in the configuration.

The effect of multiplexing can be tuned/monitored in the bptm logs. If these values are changed, the changes do not take effect until the next time a schedule runs.

Fragmentation (segment size)

Fragmentation can be configured on the storage units and can be 0, or any integer between 50 and 2000 MB.

Server performance

VERITAS NetBackup has a number of tunable parameters that allow users to optimize the speed of their backup.

- Number and size of shared data buffers. NetBackup uses 8 shared data buffers for multiplexed backup, 16 shared data buffers for non-multiplexed backup, 12 shared data buffers for multiplexed restore, and 16 shared data buffers for a non-multiplexed restore.

  To change these, create the following files:
  – <install path>\NetBackup\db\config\NUMBER_DATA_BUFFERS—used for multiplexed and non-multiplexed backups
  – <install path>\NetBackup\db\config\NUMBER_DATA_BUFFERS_RESTORE—used for multiplexed and restores
  The default size of these buffers is 64 KB.

  To change the size of the buffers create the following file:
  – <install path>\ NetBackup\db\config\SIZE_DATA_BUFFERS
  The size is in bytes, for example, 64 KB = 65536.

In effect this is the tape block size and transfer size. HP recommends this value be set to 256 KB for best performance with HP LTO3.

The two most important parameters with regard to tape transfer rates are:

- The number of tape transfer buffers (default 16)
- The size of the tape transfer buffers (default 64 KB)

To find the most appropriate values for the number and size of buffers, the NetBackup log files must be analyzed, in particular the BPTM logs. To enable BPTM logging, the properties of the media server (*the machine that the tape drive is directly attached to*) must be edited.

1. Expand the Host Properties within the NetBackup Administration Console on the NetBackup management server (if different form the media server).
2. Find the appropriate media server and select the properties dialogue by right-clicking the selected media server.
3. When the properties of the media server are being displayed, select the **Logging** tab.
4. Disable all logging except BPTM by setting the dropdown menus to "0 (minimum)." Logging of all services and activities can be detrimental to system performance.
5. Enable full BPTM logging by selecting **5 (maximum)** in the BPTM dropdown box.
6. Create the specific directories for the log files to be saved in by running the "mklogdir" batch file:

   C:\Program Files\VERITAS\NetBackup\Logs\mklogdir.bat

   Directories for all the different log files will be created, in this instance BPTM is the only relevant directory. This file must be run on the media server to which the tape drive is connected.

Analyzing logs

With BPTM logging enabled, a backup job must be run to create a log file.

In the following example, a single large file was backed up off a fast Fibre Channel disk array. When the job completes, the following directory must be opened on the media server:

C:\Program Files\VERITAS\NetBackup\Logs\bptm

The logs are in plain text format with no file extension, with the name format of:

DDMMYY, for example, 281102 (28 November 2002)

To view, open in Notepad or any similar text viewer.

The following examples are from the BPTM log:

10:46:08.222 [3200.3164] <2> io_init: using 65536 data buffer size

10:46:08.222 [3200.3164] <2> io_init: CINDEX 0, sched KB for monitoring = 30000

10:46:08.222 [3200.3164] <2> io_set_recvbuf: setting receive network buffer to 132096 bytes

10:46:08.222 [3200.3164] <2> io_set_recvbuf: receive network buffer is 132096 bytes

10:46:08.222 [3200.3164] <2> io_init: using 16 data buffers

The first example shows the buffer size at 65536 (64 KB) and number of buffers 16.

10:55:38.585 [2560.3160] <2> write_data: received first buffer (65536 bytes), begin writing data

10:56:44.445 [2560.3160] <2> write_data: writing block shorter than BUFF_SIZE, 13312 bytes

10:56:44.460 [2560.3160] <2> write_data: writing short block, 13312 bytes, remainder 0

10:56:44.460 [2560.3160] <2> write_data: waited for full buffer 11,000 times, delayed 11,000 times

The second example shows the BPTM service had to wait 11,000 times for a full buffer to write to tape. These delays dramatically slow data transfers to tape and also indicate that more buffers are required.

Results of HP testing with HP LTO2 tape drives

Table 11.

Number of tape transfer buffers	Number of waits	Transfer rate to tape (MB/sec)
16	11,000	15
24	3,000	16–17
32	0	30–32
48	0	30–32

These results indicate that having 32 buffers has the effect of canceling any waits, meaning there are always full buffers available to be written to tape. In this example, the net effect is an increase of 100% in the data transfer rate.

The optimal setting should always be used, as any more buffers than is necessary would just waste memory space, for example, 48 buffers at 65536 bytes each equals 3145728 bytes, or 3 MB.

All of the preceding examples and settings are only relevant to standard file-by-file backups. It is recommended that if application backups are being performed such as Microsoft Exchange server or SQL Server, a full analysis of the BPTM log is carried out to ascertain the optimum buffer settings. For example, tests have shown a minimum of 64 data buffers are needed with an Exchange online backup as opposed to 32 in file-by-file.

Note

After any changes are made to the configuration files, it is recommended that the current log is renamed to a backup log. If NetBackup detects an existing log, it will append to the end of the file, which can be confusing. If NetBackup does not detect an existing log, it will create a new one the next time a backup is performed. The effects of any changes made can then be viewed in the separate log and compared if necessary to the backup log.

NetBackup Network Performance

If after tuning the buffer size and the number of buffers the network backup performance is not as expected consider changing the following parameters.

NetBackup also has a tunable parameter that can be used to adjust the size of the network communications buffer, when data is transferred over the network to the media server.

The size of these parameters is calculated as follows:

- Backup (data buffer size * 4) + 1024
- Restore (data buffer size * 2) + 1024

Assuming the default buffer size in NetBackup is 65536, this sets the backup network buffer to 263168 bytes for backup and 132096 bytes for restores. Set as follows:

- \NetBackup\NET_BUFFER_SZ for backups
- \NetBackup\NET_BUFFER_SZ_REST for restores

If you change the data buffer size, verify that the network communication buffer size has been changed accordingly in line with the preceding formula.

Network Client Performance Tuning

This parameter is changed on the client. This tunable parameter is the counterpart to the network buffer previously described and is set on the client. The parameter is referred to as the SO_SNDBUF. The default value is 32 (but an extra kilobyte is added internally to the value specified).

From Host Properties in the NetBackup administration console in 4.5 or from the configure NetBackup GUI in earlier releases, open the Client Properties dialog for the client on which the parameter is to be changed.

From the Windows client tab, set the Communication buffer size field to a default of 32. Values as high as 128 have been shown to increase performance significantly.

Another way to set this value is to do a registry change:

Software\VERITAS\NetBackup\CurrentVersion\Config

Add or modify the BufferSize value, which is a REG_DWORD value. Set it to 0x20.

VERITAS NetBackup also offers the ability (like HPreaddata) to assess theoretical client disk access performance using the bpbkar32 utility.

bpbkar32 –nocont *<filepath>* > NUL 2 > NUL

VERITAS NetBackup also recommends checking the client machines involved in network backup for:

- Disk fragmentation
- Virus scanning—Impacts potential client disk I/O achievable
- NetBackup notify scripts (bpstart_notify.bat and bpend_notify.bat)
- NetBackup software location—Try not to back up the data on the same physical disk drive as the NetBackup installation as performance will be affected, especially if the logging feature is enabled
- Open transaction manager—When enabled, causes a delay while information is gathered about the volumes being snapped and for the drives to quiesce
- Job Tracker—Causes a delay while estimate of data to be backed up is performed

Common restore performance issues

- Improper multiplex settings—Only use the value required to stream the drive.
- NetBackup catalog performance—For faster restores, this catalog must reside on a fast disk subsystem tuned for fast read performance.
- Fragment size—Affects where tape markers are placed and how many are placed. Too few tape markers will slow single file restore performance. A typical setting is 2048 MB (2 GB between file marks).
- MPX_RESTORE_DELAY setting—NetBackup can perform multiple restores simultaneously from the same multiplexed tape. The default is 30 seconds. If large files are multiplexed on the tape, consider increasing this value.

Legato Networker

Legato Networker prefers to utilize variable mode devices—do not use fixed mode.

Although the terminology is "block size," what Networker actually writes is (1 x block size) KB per write. That is, to influence transfer size, set the block size appropriately.

- The blocksize (translates into transfer size) is set in the device resource as of Legato Networker 7.0 (in earlier versions they are set as environment variables).

Look for "Device block size" under the Configuration tab on Windows or in the "Details" view on UNIX.

- In Legato Networker, the concurrency (number of streams a tape drive can interleave) is referred to as target sessions and can be set to a maximum of 512. From the Media menu, select devices, select target sessions, and change as appropriate.
- Legato Networker can support up to 512 backup jobs simultaneously, which is referred to as Server Parallelism.

Legato strongly recommends using a block size of 192 KB or 256 KB for LTO3 and this can be implemented as follows by clicking **details** (UNIX GUI) or **Options/Customize/Display Hidden Attributes** (Windows GUI). You can easily change the device block size attribute. Note that the actual block size used is set when the tape is labeled, so if you change the block size you must re-label the tape.

Computer Associates BrightStore Arcserve Backup (BAB)

Note:
BAB performance tuning parameters described in this document are for advanced users only. It should not be necessary for most users to change these parameters and it is also not recommended. Inappropriate use of these parameters can potentially degrade performance and also destabilize the system.

The following answers are with respect to BrightStor ARCserve Backup r11.1. Specific configuration parameters and options may not be applicable for other releases.

The default setting for block size in BAB is 64 KB. This cannot be changed.

In Windows, BAB will use a transfer size to 512 KB for LTO, if this supported by the SCSI/Fibre controller driver. In UNIX/Linux, 64-KB transfer size is used by default.

BAB will uses 64-K transfer size for most tape drives. This can be increased to the maximum transfer size supported by the SCSI adapter up to a maximum of 512 KB by setting the following registry key to 1. Note this is not needed for Ultrium tape drives as BAB automatically sets the transfer size to 512 KB, even without the presence of this key.

> HKLM\SOFTWARE\ComputerAssociates\BrightStor ARCserve Backup\Base\ TapeEngine\ Config\MaximizeRWShots= 1(DWORD)

BAB writes filemarks only when it appends to a session, except for the session that spans tapes, so there is no ability to change the fragmentation size.

There are no tape or disk buffers to change in BAB. BAB automatically and dynamically detects the physical memory available in a system and optimizes the cache sizes accordingly. It is at least 2 MB per backup/restore job.

BAB supports many advanced features for enhancing performance such as multiplexing, multi-stripped backups (TAPE RAID), backup-to-disk, image backup (small file backup performance enhancement), and others.

Users may enable multiplexing through BAB backup manager when a backup job is being created. The number of concurrent streams per job is four by default, and can be configured to be up to 32.

For OTN backup, the process priority of the backup agent and its memory usage can be manually adjusted to reduce resources used by the agent.

For OTN backup, BAB agents automatically tune all performance parameters except network load balancing. BAB recommends using operating system settings for network load balancing. The TCP/IP buffers are internally tuned to maximize performance. User can disable buffer optimization made by the BAB agent by setting the following registry key and in this case TCP/IP default buffer size is used.

> HKEY_LOCAL_MACHINE\SOFTWARE\ComputerAssociates\BrightStor ARCserve Backup\ClientAgent\Parameters\SetWSSendBuffSize= 0(DWORD)

If users are sharing tape libraries with multiple BAB servers, the purchase of BAB SAN option is required. If users are using the SAN library as direct attached without sharing it with other BAB servers, SAN option is not required.

For Disk-to-Disk backup with BAB, multiple streams (multiplexing) to secondary disk storage is supported. Multiple streams are not supported from the secondary disk storage to a tape device. The second option is usually not needed and also not recommended in many cases as simultaneous reads from the same disks can actually reduce performance due to disk thrashing. Disk staging is recommended under the following circumstances.

If the user wants fast backups without the overhead of multiplexing, disk staging is recommended. The assumption is that the disk is fast enough to meet the backup window, thus the user shrinks the backup window without penalizing restore performance.

Most of the restores are file/directory-level restores. These kind of restores should be very fast from a disk as compared to tape because they avoid the unload/load tape time, seek time, mounting spanned tapes, and so on.

Restores, Merge, Scan can be done even while backups are being done to the disk. This enables a user to not wait for other operations to be finished.

BAB offers the following tuning parameters for Disk-to-Disk backup:

- Users can override the default amount of data written in one write operation (chunk size) to disk. The greater the value, the lesser would be disk fragmentation when multiple streams are writing to disk.
- Chunk size can also increase backup throughput depending on disk array's capabilities.
- Advanced algorithms are used to automatically optimize on various other I/O parameters to increase the throughput.

Tivoli Storage Manager

Tivoli Storage Manager (TSM) differs from most tape backup applications in that its default behavior is to back up to disk then migrate this data to tape. This has many implications for backup performance and tuning options.

Communications

The TSM client can transfer data to the server in two ways: shared memory and TCP/IP. These can have an effect on performance.

If the client is on the same machine as the server and attached tape drives, then the best method of communication is by direct "Shared Memory," which means that data is not passed over a network link. Another option is TCP/IP, which is required for network attached clients, but may also be configured for local clients (using loop back), which may adversely affect performance.

The COMMmethod option is configured in the dsmserv.opt file on the server and in the dsm.sys file on the client.

If TCP/IP is used, several server and client configuration options can improve performance. These options are set in the dsmserv.opt file on the server and dsm.sys or dsm.opt files on the client.

- TCPWindowsize

  This specifies the amount of data to send or receive before acknowledgment must be obtained from the receiver. The default is 16 but can be increased, dependant on operating system. It is recommended that on UNIX systems, this is set to 64, while Windows systems should limit this to 63.

- TCPnodelay

  Setting this option disables buffering of small TCP packets, so send them immediately. This may increase network usage but will improve performance.

- TCPbuffsize

  This specifies the size of the TCP communications buffer when reading from disk and putting data into the TCP buffer. A larger buffer size can increase performance but requires more memory. It is recommended that this be set to 32.

Client configuration options

These options are configured in the client's dsm.opt or dsm.sys files:

- Compression

  TSM can compress files. If clients are sending data over the network to the server, then it is advisable to use compression to reduce the amount of data sent over the network. If, however, the client and server are on the same machine, it is better to disable compression and rely on the tape drive compression, which will not impact performance.

- Largecommbuffers

  This specifies the size of the communications buffers on the client. Turning this option on results in 256-KB blocks being used (32 KB if disabled). This will improve backup performance to tape.

Other performance considerations are:

- Backup to disk storage pool

  Is the default behavior for TSM. Data is backed up to disk before being migrated to tape when high water thresholds are met.

  Setting the capacity of the disk storage pool correctly results in a useful buffer to keep the tape drive stream writing. Experimenting with this value is essential as the size of the disk storage pool that provides best performance will depend on the speed and number of clients being backed up, the performance of the server, and the network performance. In some cases it is sensible to back up directly to tape, especially if the client is on the same machine as the tape drive.

- Bufpoolsize

  Specifies the size of the buffer used to store the application database. Though not directly related to backup performance, an amount of server memory is allocated to cache the database so overall server performance will be affected. Setting this to a large value (limit is 2 GB) is recommended.

- Movebatchsize

  Specifies the number of files in each server transaction, for example, a migration to tape. Set this as large as possible (= 1000).

- Movesizethresh

  Specifies the amount of data in each server transaction. Set this as large as possible (= 500).

- Maxsessions

  Specifies the number of clients that can simultaneously connect to the server. The default is 25, the minimum is 2. The effect of this will depend on the performance of the clients and the server. A large number of clients directly backing up to tape will ensure high tape drive utilization, but may be limited by the available memory or processor bandwidth in the server.

- Txngroupmax

  Specifies the number of files that a client can send to the server in one attempt. Set this to the maximum of 256.

- Uselargebuffers

  Increases the size of the client/server communication buffers. Setting this increases the buffer size to 256 KB and will improve performance.

LAN-free data movement

If the data source and tape library are attached to a Fibre Channel SAN, it is important to correctly configure the data path. If no special configuration is used, all data from a client will still be sent by the Ethernet network by way of TCP/IP rather that over the SAN.

To configure data to be transferred over the SAN, a storage agent must be installed on the client machine. The server still maintains the database, and meta-data is still sent over Ethernet but the storage agent on the client handles the transfer of the actual data over the SAN.

Configuration of a LAN-free backup is complicated, so reference to the TSM manual is recommended.

Yosemite TapeWare

This software ships in the box with every HP StorageWorks Ultrium 960 stand-alone tape drive. The configurable parameters for TapeWare are contained in the file twtapdev.ini:

- shortErase=Y
- mediaClass=9
- developmentLevel=1
- deviceBlockSize=65536
- senseDataCleanFlag=Y
- forcePageMode=Y
- OBDRshift=Y

TapeWare can stream data from up to eight sources simultaneously. It automatically assigns a new stream to each drive letter and also automatically uses existing streams for directories. This may be incorrect if the drive letter is on a partition rather than a physical disk, or if physical disks are mapped to directories rather than drive letters.

This can be modified when choosing the backup source. Right-click the drive letter or directory and select **Properties.** From the Storage tab, select either **Create New Stream** or **Use Existing Stream.** To fully occupy the bandwidth available with the HP StorageWorks Ultrium 960, multiple drives/directories must be backed up simultaneously. Hence, the data is interleaved on the media. This is excellent for backup performance but on restore there will be a slowing down as the restore searches for file components down the tape. It is worth considering your usage of restores as to whether this is an issue for your environment.

Summary and conclusions

HP StorageWorks Ultrium 960 tape drive is arguably the world's fastest tape drive capable of speeds exceeding 160 MB/sec (576 GB/hr). To make use of this available performance requires a new way of thinking in designing backup solutions.

As a result of the scenarios tested in this white paper, several recommendations and rules of thumb have emerged:

- Higher levels of concurrency (interleaving more data streams into a single tape stream) will be required to make best use of the available HP StorageWorks Ultrium 960 performance. Most backup software allows up to 32 streams to be interleaved to tape. Interleaving more streams to tape can have a negative impact on restore performance.
- For entry-level JBOD and disk array usage of HP StorageWorks Ultrium 960, it is important to understand what your disk subsystem is capable of delivering. This can be done using HP performance assessment tools downloadable from http://www.hp.com/support/pat or the performance tools are also embedded within the HP industry-leading Library and Tape Tools diagnostics.

- For servers in a SAN environment, the rule of thumb is 1.5 to 2.00 GHz of processing power and 1.00 GB of RAM per processor should be allocated for the backup process, where the backup process is reading from disk and writing to tape on separate channels.
- OTN backup is not ideally suited to HP StorageWorks Ultrium 960 tape drives since Gbe itself at 100 MB/sec becomes a bottleneck. In addition, high processing power is required to manage the TCP/IP stack. To improve performance consider:
  - Multiplexing more streams to the tape drive until the 100-MB/sec Gbe bottleneck is reached.
  - Carefully examining your network topology and routing tables. There may be easy performance gains to be had in balancing network fiber usage.
  - Using multiple NIC HBAs in the backup server and binding each of them to a processor.
  - Using TOES to reduce processor loading.
  - Moving to a SAN-based backup architecture.
  - Moving to a Disk to Low cost Disk to Tape backup architecture (D2D2T) and performing the initial backup over the network to disk not tape.
  - Aggregating multiple networked clients. This remains a good strategy for delivering on the performance capability of earlier generation Ultrium products.
- ISV parameter tuning can make a significant difference (up 60% improvement) to performance. Details of how to performance tune the most popular ISV backup software are contained in this white paper.
- D2D2T: SATA disks used as secondary disk storage arrays are only 7.2K rpm and although they are high capacity (250 GB), their performance reflects their pricing. So backing up to tape from a staged (secondary) disk array can be slower than backing up directly from the primary storage to tape. Ironically tape is now faster than disk. Disk staging is useful, however, for gathering lots of small files into a single object, or backing up slow hosts before the data is sent to tape. Both small files and slow hosts can cause very slow backups.
- For backups, use a block size/transfer size of 256 KB (most backup applications use 64 KB as default).
- Ensure the SCSI Ultra320 HBA for the tape drive is placed on a dedicated 133-MHz PCI-X bus and is not sharing the bus with other HBAs.

The HP StorageWorks Ultrium 960 tape drive offers a capacity of up to 800 GB per tape and also supports WORM media for use in compliance environments. It is available in internal, external, and rack mount configurations and will be deployed into HP automated tape libraries early in 2005. HP data rate matching capabilities within the drive allow the drive to match its speed to incoming data from the host and remain streaming, ensuring optimal performance and reliability in all situations. (Data rate matching operates between host speeds of 27 to 80 MB/sec.)

The HP StorageWorks Ultrium 960 tape drive performance is best utilized for high-performance backup in a disk and tape SAN environment, backing up directly from a high-performance disk subsystem (from primary disk storage, split mirrors, snapshots, or clones).

The HP StorageWorks Ultrium 960 is not ideal (performance wise) for OTN backup or backup from lower performance disk arrays. However, some customers may want to use HP StorageWorks Ultrium 960 tape drives in these environments purely for the capacity that it offers. Providing a minimum data rate of 27 MB/sec can be supplied to the HP StorageWorks Ultrium960 in these environments. The tape drive will still stream and optimize the performance accordingly. Finally, tape continues to deliver the lowest $/GB of any storage technology and is no longer slow. Tape should always be the foundation of a data protection strategy. Rumors of tapes demise are unfounded.

Future developments in tape backup performance

Tape technology itself is no longer the bottleneck to tape backup performance but rather it is disk performance, other system connections, and environmental factors that are likely to be the cause of poor performance. Following is a list of future enhancements that will enable more systems to take advantage of the HP StorageWorks Ultrium 960 performance:

- Disk rotational speed increases to 22K rpm (2006)—More throughput per disk spindle.
- 10 Gbe—Likely to be trunking links and infrastructure only initially. This will take some time to permeate into the mainstream networking arena (expect initial introduction in 2005). Across the network will become viable as a good strategy for utilizing HP StorageWorks Ultrium 960 performance with 10 Gbe.
- TOES and Remote Direct Memory Access (RDMA) capability within NICs—These enhancements allow more of the TCP/IP stack processing to be done in hardware and applications to move data directly between memory spaces on NIC cards. These help to drastically reduce the processor load involved in network protocols and, therefore, allow more of the processing power to be made available to the backup application, hopefully resulting in increased throughput. See Appendix B.
- 4-Gb Fibre Channel—During 2005 HP will launch a 4-Gb Fibre Channel version of the HP StorageWorks Ultrium 960, which together with the HP StorageWorks Extended Tape Library Architecture (ETLA), will allow optimum load balancing use of HP StorageWorks Ultrium 960 tape drives within HP tape libraries. Over time, disk arrays will also migrate to a 4-Gb interface, allowing greater bandwidth from the disk arrays.
- PCI-X 2.0 and PCI-Express—Most server backplanes today have PCI-X interconnect. At the highest level of operation a 64-bit, 66- MHz PCI-X slot is capable of a shared 533-MB/sec burst rate around 490 MB/sec sustained. However, with future developments of 4-Gb Fibre Channel and dual 4-Gb fibre cards, there is a distinct possibility that PCI-X bus itself could become a bottleneck. PCI-X 2.0 has taken the clock frequency up to 133 MHz, enabling a backplane burst transfer rate of 1066 MB/sec. Meanwhile, PCII-Express—due for release in early 2005–will offer a high-speed serial backplane link instead of the existing parallel interfaces used in PCI-X. PCI-Express is specified to offer backplane speeds of between 500 MB/sec (x 1 implementation) and 8 GB/sec (x 16 implementation) over the next few years.

Figure 17.

LTO Technology Roadmap

Gen 1	Gen 2	Gen 3	Gen 4	Gen 5	Gen 6
100GB	200GB	400GB	800GB	1600 GB	3200 GB
Up to 20MB/s	Up to 40MB/s	Up to 80MB/s	Up to 160MB/s	Up to 200 MB/s	Up to 300 MB/s
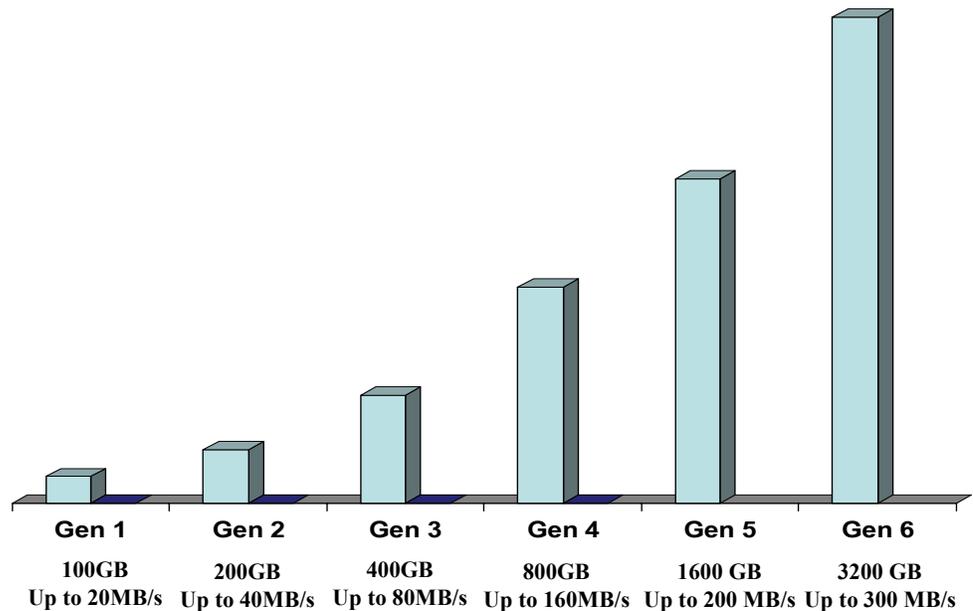
Figures are uncompressed. Roadmap is an estimate of the LTO Program's current intent and are subject to change without notice.

The technology provider companies (TPC) who develop LTO technology have recently approved Generation 5 and Generation 6 technologies. Customers can purchase an HP StorageWorks Ultrium 960 tape drive solution today and have the confidence that they are buying into a technology with a long roadmap into the future.

Appendix B: TOES

One way of reducing the processor overhead in OTN backup is to use TCP Offload Engines (TOES), which perform some of the receive processing in hardware. HP conducted some tests with Alacritech Gbe TOES using HP LTO2 tape drives in a backup server environment.

Figure 18.

Throughput Verification

Used HP's TapeTest utility to test Disk File Read Rate.
Local File System: ~55MB/s
Network File System: ~30MB/s

TapeTest simply reads data from logical drive. It does not pre-fetch, compress or multi-thread disk reads.

~30MB/s

~55MB/s

Media
Backup
Server

FC Sw   2Gb   Client   1GE   GE Sw   1GE   NIC

DL380G2

ProLiant DL580G2
4P, 1.4GHz Xeons

EVA
>80 Spindles

3
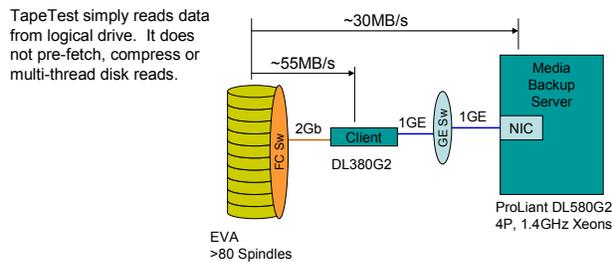
Figure 19.

Test Parameters

Constants:
• Operating System   =   Windows Server 2003
• Backup Software   =   Veritas Backup Exec 9.1
• Tape Drives per Client   =   1 to 1 Mapping
• Tape Drives   =   LTO2, 60MB/s @ 2:1
• Server Under Test   =   4P 1.4GHz, DL580G2

(FC HBA and GE NICs on Separate 64b/100MHz PCI-X Bus Segments)

Variables:
• File Set   =   "Large" & "Small"
• File Compressibility   =   1:1 & 2:1
• Number of Clients   =   1, 2, 4, 6
• Clients to NICs   =   6 to 1 NIC, and 3 to 2 NICs (x2)
• NIC Types   =   GE (NC7771) & Alacritec (TOE)

4

Figure 20.

Single NIC Network Config
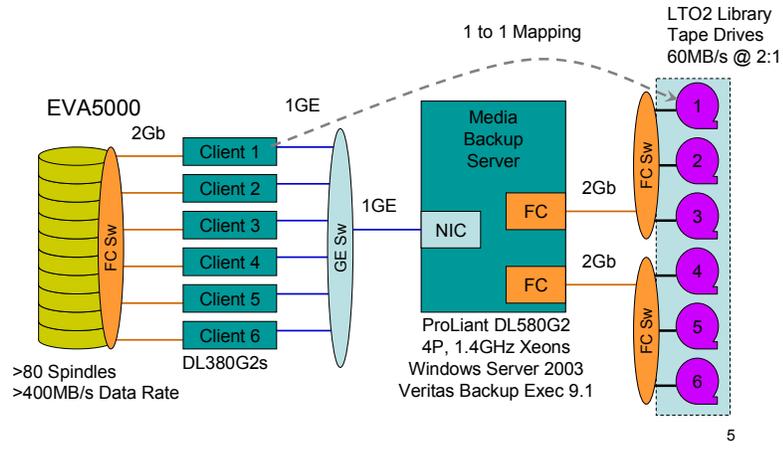
Figure 21.

Dual NIC Network Config

Figure 22.

Network Tape Backup, DL580G2 4P 1.4GHz Xeon, no HT, 6 LTO Drives
WinSrv03, Veritas Backup Exec 9.1, Large Compressible File Set

8

Figure 23.

Network Tape Backup, DL580G2 4P 1.4GHz Xeon, no HT, 6 LTO Drives
WinSrv03, Veritas Backup Exec 9.1, Small Compressible File Set
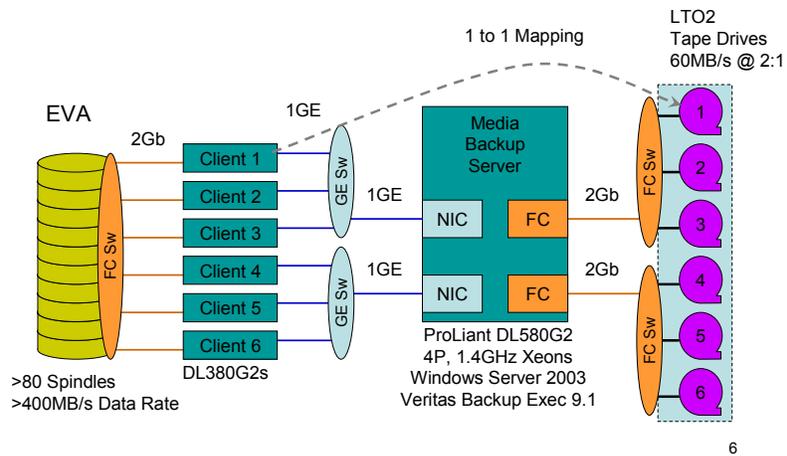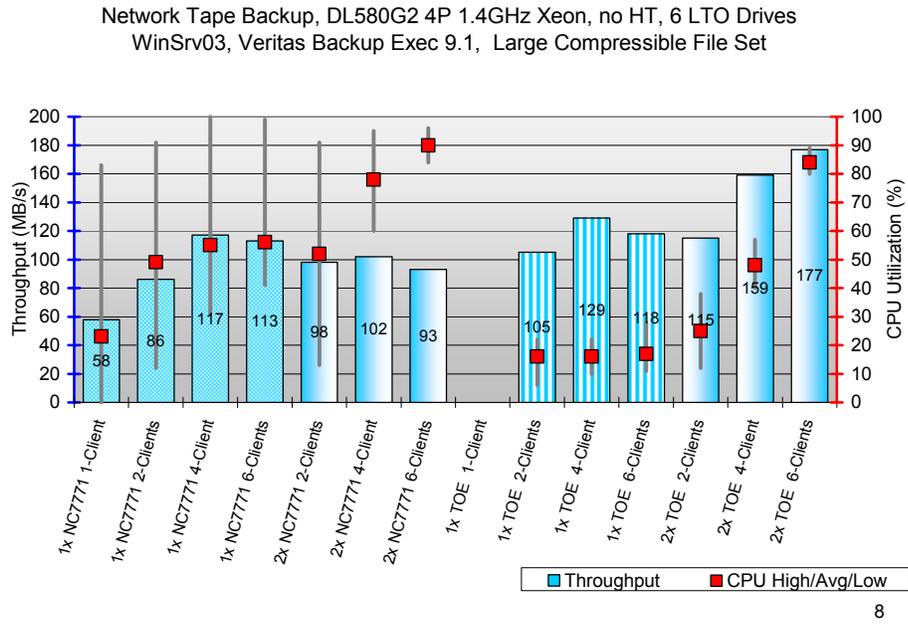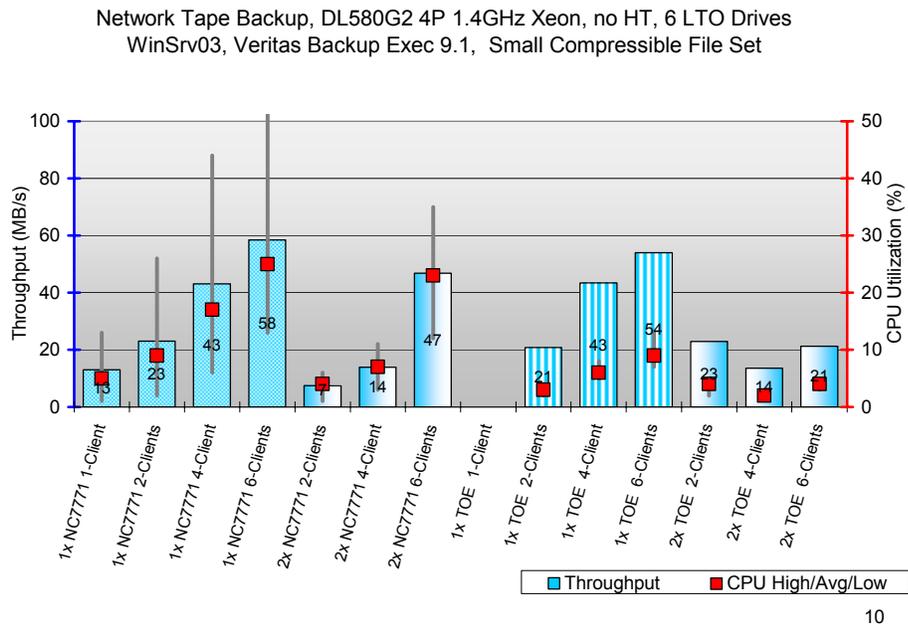

10

Conclusions

- Two TOE NICs performance 50% greater than best of 1GE NIC, 2GE NICs, or 1GE TOE.
- Backup Exec appears to have some overhead managing two NICs (GE and TOE).
- For Small File Set, Alacritech NICs did not help performance. CPU was not bottlenecked on media server. The bottleneck is on client file system.
- The best throughput for a single NIC was 117 MB/sec at 58% DL580 4P loading and 177 MB/sec for 2 x TOES at 85% DL580 4P loading.
- This type of performance (using two TOES) is a much better match for HP LTO3 performance capabilities.
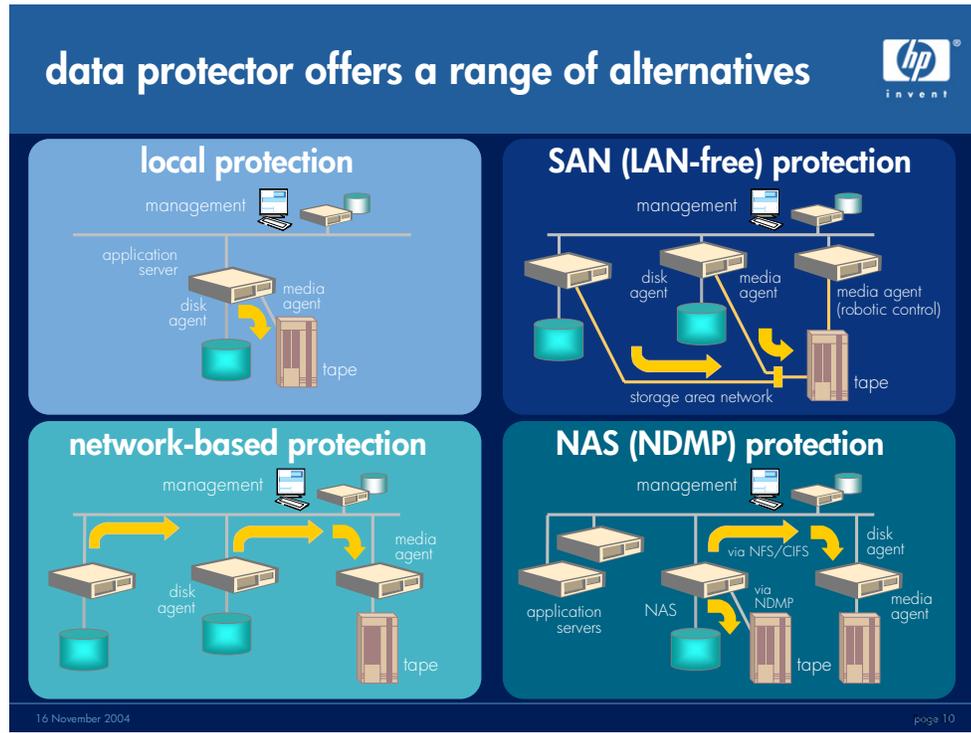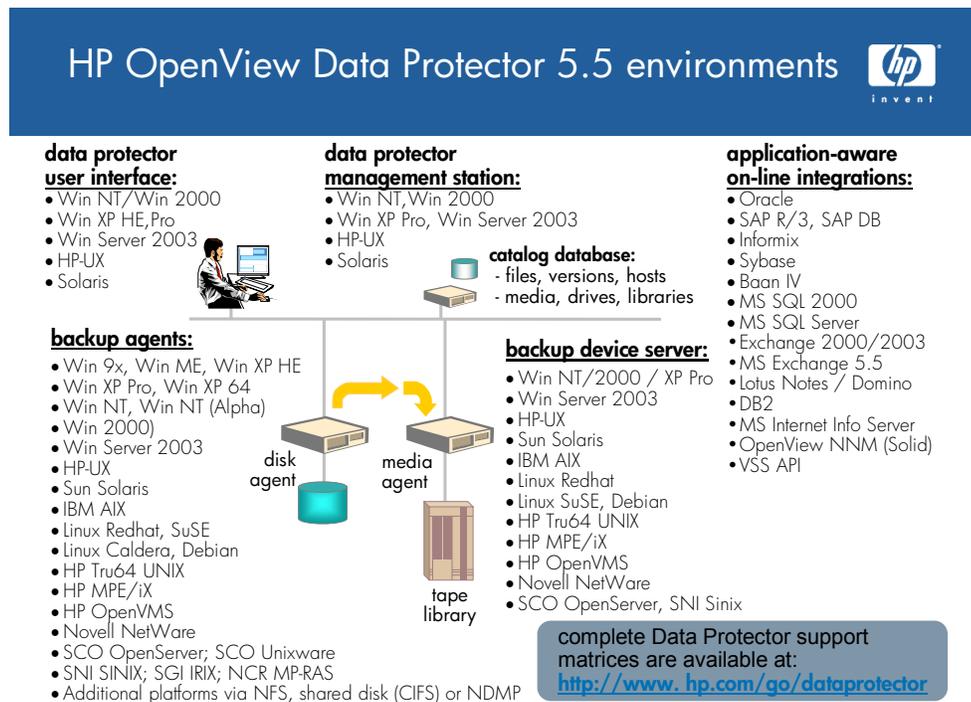
Figure 24.

Figure 25.

Figure 26.

HP OpenView Storage Data Protector 5.5

Enhancements from 5.1 (1/2)

enhanced

1. Advanced backup to disk
2. Advanced media management
3. Backup mirroring
4. Easier configuration and management of large SANs
5. Broader ecosystem coverage
6. Data Protector Media Operations becomes heterogeneous

1. Advanced backup to disk
  • Fast restore from disk and transparent access to data migrated to tape
  • Complements the Data Protector backup to disk technologies:
    zero-downtime-backup, instant recovery and virtual tape library (file library)
  • Fast & easy configuration and licensing

2. Advanced media management
  • Space efficient automated media copy
  • Tape types migration from disk to tape and between different tape types
  • Tape consolidation
  • Tape collocation/de-multiplexing
  • Improved media management by automatically generated label from barcode
    www.hp.com/go/dataprotector   support.openview.hp.com

November 16, 2004   4

Figure 27.

HP OpenView Storage Data Protector

Enhancements (2/2)

enhanced

3. Backup mirroring
  • Allows to mirror backups over unlimited distances to different physical destinations
  • Provides best mirroring performance in a SAN

4. Easier configuration and management of large SANs
  • The new multi-path device allows a simple view of all access paths from all
    systems to all devices in a large SANs
  • Automatic handling of changed SCSI addresses

5. Broader eco-system coverage
  • Extends the existing broad compatibility of operating systems, applications,
    drives, libraries and disk arrays
  • Data Protector zero downtime backup and instant recovery through the
    SMI-S industry standard
  • Launching of library web console
  • Updates the integrations with HP OpenView, supporting the latest releases

6. Data Protector Media Operations becomes heterogeneous
  • Apart from Data Protector, it can track and manage media outside the library for Veritas NetBackup

    www.hp.com/go/dataprotector   support.openview.hp.com

November 16, 2004   5

Future additions to this white paper

During the early part of 2005, HP will launch a range of automated tape library solutions using HP StorageWorks Ultrium 960 technology. As these products become available, they will be tested in realistic performance scenarios. These additional scenarios will involve multiple tape drives being used simultaneously and will involve the backup of complex applications such as SQL and Oracle® databases from very high-performance disk arrays such as the HP StorageWorks XP12000 Disk Array.

For more information

www.hp.com/go/tape

5982-9971EN, 12/2004