



*Backup Performance Gains  
with e-Space™ for UNIX and MPE*

---

<b>1. Introduction .....</b>	<b>1</b>
<b>2. e-Space Online File Archiving.....</b>	<b>2</b>
<b>3. e-Space Backup Performance .....</b>	<b>2</b>
<b>4. Disk Device Issues.....</b>	<b>3</b>
<b>5. CPU/Memory Issues.....</b>	<b>3</b>
<b>6. Tape Device Issues .....</b>	<b>4</b>
<b>7. Summary.....</b>	<b>4</b>

---

**1. Introduction**

As the data explosion continues to increase exponentially, concerns with data backup and storage grow. With the limited storage of disk devices, environments with hundreds of gigabytes, terabytes and more become overwhelmed. The data explosion is not new, and methods for managing huge data farms continue to change. Throughout the '90s, hardware solutions that used a disk subsystem and were controlled by a disk operating system became very popular.

However, while these hardware devices handled the data availability issues, they could not handle the backup, restore and disk-device installation issues. Data backup and restore continued to be large problems, and have actually increased with more and larger disk devices. Furthermore, adding new physical devices to server environments still required a shutdown of the server.

Storage devices and software continue to increase in their performance, but disk storage size grows faster. Therefore, methods for increasing backup and restores are more necessary than ever. Manufacturers continue to provide faster devices, such as the HP SureStore™ DLT-4. However, even this device only transfers data at 1.5 megabytes per second or 5.4 gigabytes per hour, or double with hardware compression.<sup>1</sup> Consequently, customers are wishing for faster speeds. At 10 gigabytes per hour and with five parallel devices, a backup will still require 20 hours per terabyte.

One of the most common thoughts in these environments is, "Why not use compression at the disk level?" When looking at the options available, however, customers quickly see the performance issues with the typical compression products. While the compression is solid and automatically accessible, configuration is by volume and affects all data on the volume. Hence, each I/O to and from the compressed volume requires a performance hit. In the case of a true OLTP situation, with thousands of users and hundreds of thousands of I/Os per day, negatively impacting OLTP I/O performance is not an option. Even a modest impact of 5 percent can cause OLTP environment performance nightmares.

Solution-Soft's e-Space solution, a module of the e-Storage product line, is the next step in data storage. With a unique combination between e-Space and the operating system, even large OLTP environments will not have their I/O impacted. The design of e-Space allows OLTP environments to take limited performance impacts, and yet continue to gain top backup performance and disk space usage.

---

<sup>1</sup> <http://www.hp.com/tape/datasheets/dlt40.html#specs>

## 2. e-Space Online File Archiving

A primary benefit of e-Space is its ability to achieve true online archiving of files. Typically, when thinking of archiving, two options exist. The first is to archive offline, such as to DLT; the second is to archive "nearline" using a storage device such as optical.

Disadvantages exist in both offline and nearline archiving. The major issue is the speed of access. While it is possible to implement an offline or nearline automated process, retrieving these files is generally hundreds of times slower than a direct access to disk. Therefore, when a user or application requests data from a file that is archived, while they are able to have access to the file, the wait time can be long and will surely be noticeable.

With e-Space the files continue to reside on the direct access disks. Therefore, the access time is greatly reduced. Since compression of archived files is common, the decompression time would be similar in all three methods -- online, nearline and offline archiving. The time required to decompress, usually less than 1 second per 5 megabytes is not of great concern. However, the time required to load up the correct optical disk platter or tape, and then to bring this data to the direct access disks, is of concern.

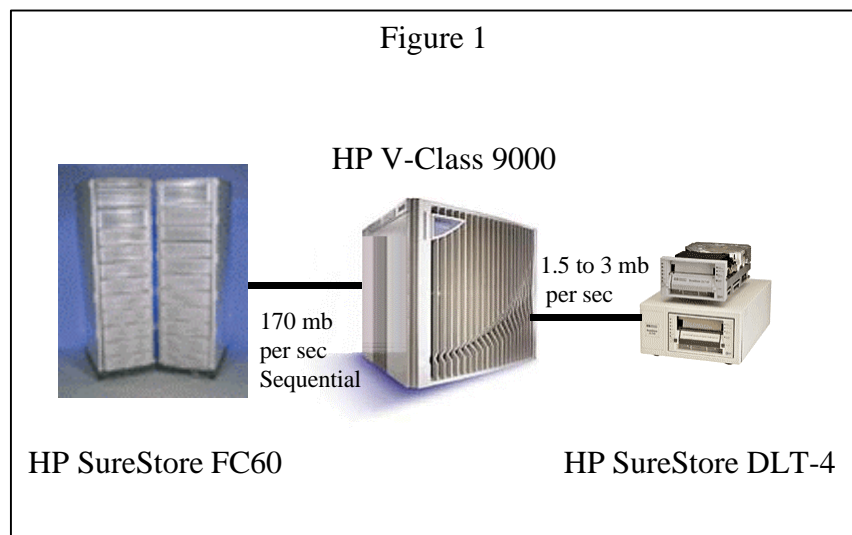
Because of the unique nature of the e-Space solution, in online archiving, customers gain the benefits of archiving, as well as the benefits of leaving data on direct access disks. This increases the availability of the data, while reducing the overall cost, including management cost of direct access disk.

However, another key benefit of the e-Space solution, which is often overlooked, is the performance gain during normal backup procedures. Many companies do not use nearline or offline archiving because of the performance issues mentioned previously, and those that do, often don't archive files that have been accessed during the previous six months or year, and in some cases even longer. Therefore, during normal backup situations, files that should be archived -- but are not -- are continually stored. These files can greatly influence the storage time, and definitely will influence any restoration time. With e-Space, files are archived on the direct access disks, reducing storage and restoration times.

## 3. e-Space Solution Backup Performance

Figure 1 shows a simple "environment where the I/O bottleneck is the tape system. Even at 3 megabytes with hardware compression, it would require 85 DLT-4 devices running in parallel to keep up with the disk device system. In the best scenario, this is hardly acceptable.

Since the obvious bottleneck in the backup process is the tape device, most companies attempt to solve this through compressions during the backup process. This can be with hardware compression on the tape device itself, or through software compression provided by the backup software. However, several other areas exist within the environment that can have bottlenecks causing longer backup times.



#### ***4. Disk Device Issues***

In our simple example, it appears that no bottlenecks can occur on the disk device side; however, there can be one possible bottleneck. If there are enough tape devices, memory and CPU power, it is possible for a backup process to "outrun" the disks. On the other hand, in an EMC Symmetrix™ or HP SureStore™ high-speed environment, it is unlikely. In a standard SCSI environment, and especially in slower, older technology, it's not too difficult. An ultra-wide SCSI runs approximately 3 to 6 megabytes per second. With 10 DLTs, against a bank of disk devices on one connection, the bottleneck can occur at the disk access level.

A subtle issue that may cause a bottleneck is the connection of disk devices over a network. Generally, this is done when there are several computers in an environment and they need connections to a shared disk system. In this case, if there are many I/Os occurring from all the machines, the machine doing the backup will not have full access to the network bandwidth. It is possible that a bottleneck will occur simply because of access from other machines that are only related to the backup machine in that they share a disk system or disk network.

With e-Space speed will always increase at the disk device level - even with the simple example. Since the data is already compressed on the disk device, the megabyte-per-second speed will increase by the amount of the compression.

If we assume 60 percent of the data is archived online, and the average reduction in space is only 50 percent, the data transfer speed between the computer and the disk device will be increased by 30 percent. (Forty percent of the data is not compressed and will require 100 percent of its normal access speed. Sixty percent is compressed by 50 percent, resulting in a 30 percent increase in data transfer speed. Therefore, 40 percent plus 30 percent equals 70 percent and 100 percent minus 70 percent results in a 30 percent savings.)

#### ***5. CPU/Memory Issues***

It is easy to think the speed of the CPU and memory will have no affect on a backup, and in the case of a standard, single process, non-compressed backup this may be true. On the other hand, if multiple backup processes and software compression are occurring, it is possible to outrun the CPU and memory.

Standard queuing theory tells us that a machine's performance can begin to degrade once it becomes 80 percent busy. There can be many reasons why the "busyness" can occur. If the memory is overloaded -- a good possibility in many multi-processor backups -- the CPU begins to spend more and more time managing the memory and it's overlay candidates. Likewise, if there are other processes running, as is the case in an "online" backup, the CPU could be busy just managing the processes that are running on the system.

Software compression adds additional overhead. Any additional computation, such as compression, will increase the use of the CPU. In some environments, non-software compressed backups will not overload the CPU and memory, but once software compression is added the overload occurs.

Again, the e-Space solution will increase the amount of data processed within any given time period. Where backup software compression is not used, the e-Space application will not have a direct impact, either good or bad, on the use of the CPU and memory. However, it will process more data with the same amount of time. This will decrease the overall usage of the CPU and memory because backups will complete faster, allowing for other uses of the CPU and memory.

Where backup software compression is used, the e-Space solution will have a positive impact on CPU and memory usage. Since the files are already archived, the backup software will not be using additional CPU or memory to compress the data. This will result in a decrease in the amount of CPU and memory used. In addition, since the archiving occurs once, the positive impact of e-Space will be realized each time a backup occurs.

## ***6. Tape Device Issues***

Tape devices cause the most serious and obvious bottleneck in a backup procedure. While tape devices have increased in performance greatly, and it is likely they will continue to increase, the increase in performance of disk devices, CPU and memory continues to outpace the tape devices. In other words, unless older disk devices and machines are used, the tape devices will never "catch up".

One of the most common options is to add several tape devices to the machine. With this approach, the amount of throughput to the tape devices can increase, as long as you do not overload the CPU, memory or disk devices. An overloading in these areas, especially the CPU and memory, can occur with as few as six DLT devices. With older, "slow" devices such as DDS-1, the speed drops to less than 10 kilobytes per second, so the resulting bottleneck will occur more quickly.

Again, however, it is easy to recognize the benefits of online archiving through e-Space. If a file is compressed 98 percent, the bandwidth to the tape device has "increased" 98 percent. If the file is compressed 60 percent, the bandwidth is "increased" 60 percent.

## ***7. Summary***

The performance benefits of e-Storage's e-Space module during backup are more than might be expected. There are many possible areas for bottlenecks during a backup process, and e-Space provides benefits in each area. Generally, the e-Space solution will increase the overall speed of a backup, even when compared to a backup with compression, resulting in fewer bottlenecks, decreased backup times and increased throughput.

Additionally, a primary benefit of e-Space is the enabling of true online archiving of files. Since the archived files are located on direct access disk devices, the access speeds are greatly reduced compared to nearline and offline devices.

With the e-Space solution the benefits of online archiving and increased performance during backup and restore are achievable.