Using Automated Tape Libraries for Network Backup and Archive

---

by John W. Kranz

FEBRUARY 26, 1996

Doc. No. 99800010

Using Automated Tape Libraries for Network Backup and Archive

---

by John W. Kranz

INTRODUCTION

Network storage requirements have exploded in the last few years and the trend shows no signs of abating.

Several market trends and technologies have combined to fill servers' disks, seemingly as fast as they can be added on:

- RISC-based Computer Systems;
- Data intensive applications: CAD/CAM, Imaging, scanning, DTP;
- Downsizing: bringing tasks and data from mainframes to LANs, especially commercial database applications;
- Increased office automation and computer usage - more machines, more tasks, more applications.

While the requirements have certainly increased, the resources of system and network administration have not kept pace. Figure 1 shows the growing disparity between the amount of network storage and the resources to manage it.

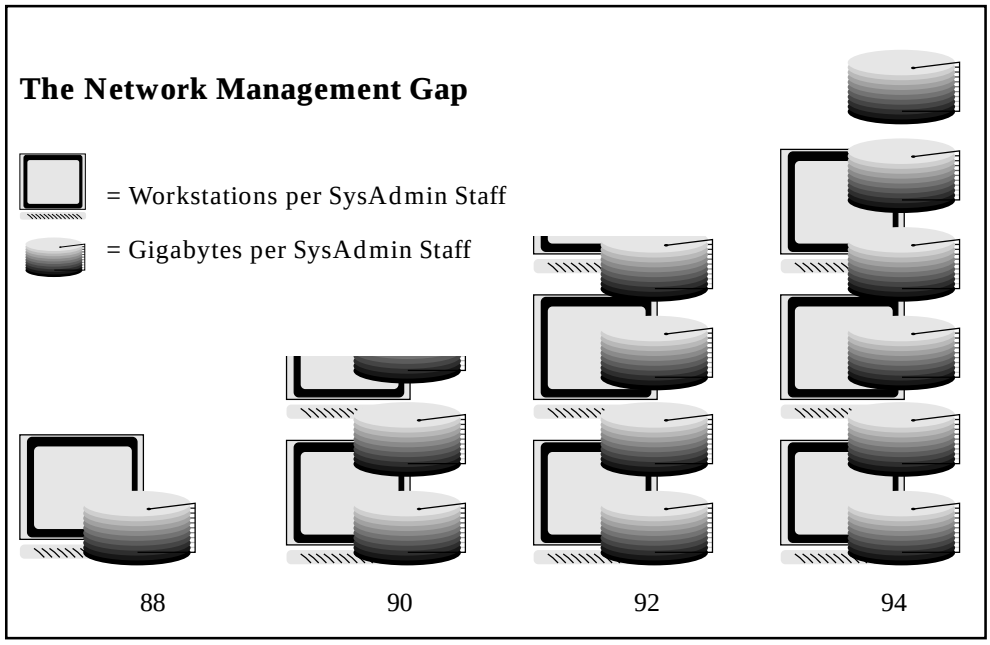

Figure 1 - Network Storage per System Administration Employee

The solution is to automate the automation: computerize computer backup.

At first this was done by increasing the capacity of backup media. More than a million helical scan tape drives have been sold -- many on the basis of backing up an entire network in one evening without an operator to change tapes. Timed-start backup software allowed the entire network to be backed up on one tape, overnight and unattended.

Now, more and more networks are exceeding the capacity of a single tape. Just as relevant to the rise in popularity of automated tape libraries, however, is the implementation of more sophisticated backup schemes and the need to manage and catalog each backup and the media on which it is stored.

Random vs. Sequential Access

If the sole requirement for backup automation is capacity, a sequential access device can meet this need at a lower cost -- and reduced software requirements--than a random access library.

A sequential access device (frequently called a “Stacker”) features a single tape drive and sufficient robotics to load the next tape in sequence on demand. There is usually no way of loading the previous tape in the stack and the only way to skip ahead more than one tape is to load and unload all the tapes in-between (see Figure 2.)

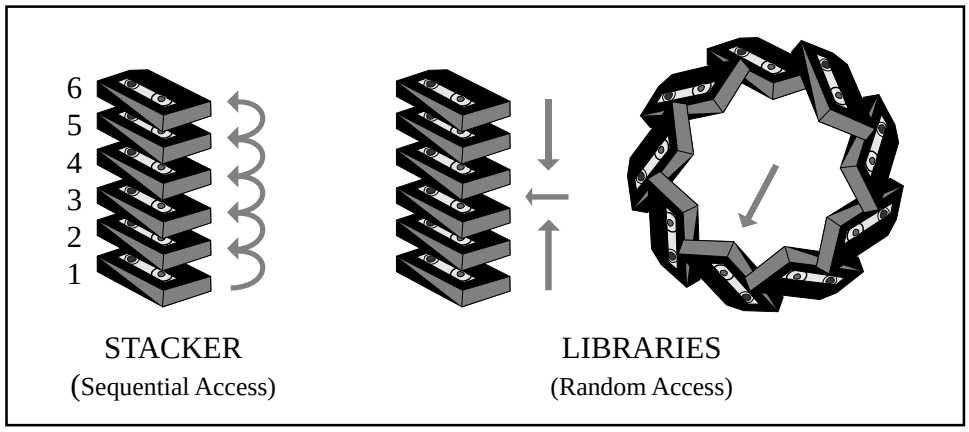

Figure 2 - Sequential vs. Random Access

A random access device (“Library”) conversely, can load any piece of media which is located in its domain. A library can also have more than one drive.

Need/Justification

If the cost justification for a stacker is capacity, the justification to acquire a random-access library (and software to utilize it) is to eliminate human intervention from the backup process.

A random access library of sufficient capacity, with the proper software, should be able to provide backup and restoration with little or no ongoing maintenance. Of course, backups should be brought off site and replaced with blank media, cleaning cartridges must be replaced, etc. Day to day operations, however, can be performed with media in the library.

By reducing human intervention, not only are labor needs decreased but also the opportunity for human error. Configured properly, the automated system will not likely use the wrong tape for a store nor fail because the right tape wasn’t available. Keep in mind that the main reason for performing backups has switched from protection from hardware failure to protection from human error. Also remember that the exact things you’re backing up to prevent can happen just as easily during the backup process.

Another benefit to automating backups with a library is centralized data control. Aggregating backup data and its media in libraries allows a system administrator to establish and enforce company policies and procedures for backup and security.

HARDWARE REQUIREMENTS

True automation can only be attained if the backup hardware meets the following requirements:

Random Access

As discussed before, a sequential-access device will increase the amount of data that can be backed up without intervention but only a random access library will allow automated restorations and automated media management.

Capacity Requirements

If the library does not have the capacity to hold all the data you want available, you will either have to scale down your requirements or be forced to continually swap media in and out of the library. Compromising your backup procedures and extensively manipulating media are exactly what automation is designed to avoid.

While there is no "magic formula" to determining the proper capacity library for a network, it should be able to hold:

1. All the data you want available (multiple copies of backups);
2. plus all the data you would want to take off-site at one time (before it is removed, it will require library capacity);
3. plus have room for a cleaning cartridge (helical-scan drives only);
4. plus have room for any special needs you or your backup software may have.

These are each discussed below:

1. On-line Data. NOTE: The term on-line is used in this document to describe data that is available without human intervention. For convenience and readability, this document will not distinguish between immediate on-line access (disk) and near-online (tape library) access.

How much data will be available for automated retrieval? The answer to this question begs two other questions: "How frequently will I backup?" and "How long do I want to keep each backup?"

If you wish to perform full stores every week and keep these stores available for six weeks, you will need six times the total storage of your network to hold full stores.

This assumes your library and software can efficiently pack data onto media: to properly expire those tapes would require a minimum of six tapes. Even if your software can append data onto those tapes, count no less than six times the capacity of a single media element.

You might also want to keep quarterly backups around for a year. Add four tapes for a total of ten. If you can fit a full store on one tape, multiply the single cartridge capacity by ten. Else, multiply ten times the number of cartridges required for a full store.

Next, consider incrementals. If you perform a full store every week, you may wish to perform an incremental store (backup everything that has changed since the last backup) at least once a week and possibly every day.

Assuming that two incrementals are made per week and that they are kept only up until the next full backup, two more tapes are required.

This backup schedule requires an absolute minimum of 12 tape cartridges to automate: 22 if a full store spans two tapes. This also assumes that incrementals will not span multiple tapes.

There are still no provisions for off-site backups, cleaning cartridges or other requirements. Notice, however, that the library has reached steady-state: until the capacity requirements or backup schedules change, cartridges will expire as fast as new ones are needed.

2. Off-site Data. Off-site backups should fit in the library together with the steady-state load. This allows them to be automatically generated and migrated to the vault at a convenient time. If the library were full of system backups, an off-site backup would have to be created by hand-loading and removing media. Add the number of tapes required for a full off-site store.

3. Cleaning Cartridge. If you are using helical-scan tape drives, a library slot should be designated (or at least available) for a cleaning cartridge. Infrequent cleaning jeopardizes data integrity, hardware longevity and voids the warranty on the tape drive. Exabyte, manufacturer of all 8mm tape drives, recommends head cleaning every 30 hours of use for its tape drives. Performing this manually is not only labor intensive, it is too risky to be put off, forgotten or missed due to a vacation.

4. Other Requirements. While data management is being centralized, perhaps it makes sense to back up other machines onto the library. Sharing a library across multiple platforms (UNIX, VAX/VMS, Novell, Macintosh) can be complicated but rewarding. Think carefully about the ability of the software to interact. Spectra Logic's Alexandria Backup Librarian (for UNIX networks) can be configured to partition the library, essentially freeing up portions of the library for other packages to use.

Alexandria also keeps one or two copies of its database in the library at all times. This allows the application, configuration and file catalog information to be restored in case of failure or corruption. If you're using a package which does this, allocate tape slots to account for this feature.

Assuming no other platforms will be backed up to this library, the scenario described so far would require the following to automate on a small scale (<50 GB) network:

If this library is using Exabyte EXB-8500 tape drives (5.0 GB capacity), 15 slots will backup to 5 GB, 26 will back up as much as 10 GB, 37 will backup to 15 GB, etc.

Although this example is hypothetical, it illustrates two things:

- Backup schedules determine library needs more than total network capacity.
- The number of slots available in a library must be considered as much as the total capacity.

Reliability

The third hardware consideration to true automation is reliability. Reliability made this short list of essential requirements rather than the longer list of "Other Hardware Considerations" because hardware failure jeopardizes the backup schedule as well as data integrity.

A failure to load the tape may not require service on the hardware but it could easily preclude a whole night's backups. Intervention then becomes necessary not only to fix the device but also to fix the backup schedules. If the available backup window is scheduled tightly, which it usually is, it could be days before continuity is restored.

Murphy's law (surely Murphy is the "patron-saint" of backup) dictates that catastrophe will strike on the day after the failed backup--and probably to the exact disk and sector where the backup stopped!

OTHER HARDWARE CONSIDERATIONS

Robotics control vs. Data transfer path

For compatibility requirements, consider the hardware interface to the library. Consider both the data path as well as the control path: in other words, the interface to the drive and the interface to the robotics.

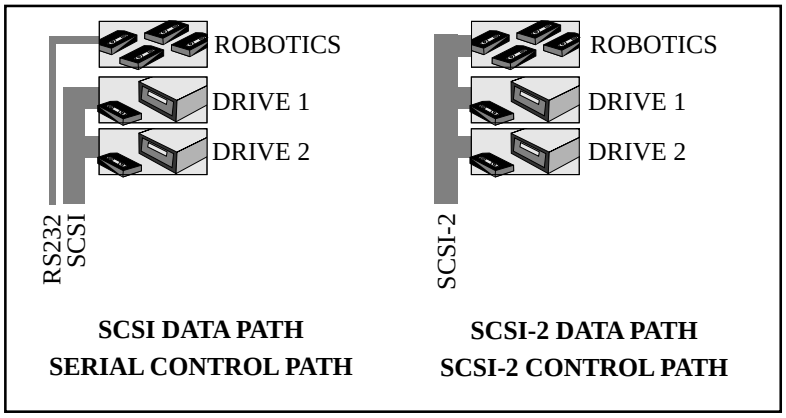

Figure 3 - Control Paths & Data Paths

The original SCSI language specification does not include any commands to manipulate media in libraries. This is the most popular interface for the class of system for which a small to midrange library is suitable. Therefore, most manufacturers have included a serial (RS232) control path. This serial port controls the robotics but, once a tape is loaded in the drive, the data is sent via the SCSI bus.

The SCSI-2 language specification does include commands to manipulate media in libraries. While this should simplify library integration and save that extra serial port for another purpose, SCSI-2 controllers and device drivers have not found their way into the mainstream of UNIX workstations nor PCs at this time.

Many library vendors provide both a serial and SCSI-2 interface. This is a good compromise in that it supports a variety of present systems and has the capacity to accept SCSI-2 as it becomes more prominent, which is quite likely.

Whatever interface you choose, the important thing is that it is supported by your hardware and software. Serial library interfaces are not standard--ascertain that the hardware you want is supported by the software you want.

Number of Drives

More than one drive in a library provides two advantages:

1. Multiple drives can be accessed in parallel to increase throughput;

Before expecting incredible transfer rates from multiple drives, however, make sure that parallel writes are supported by your software and that the host hardware has the I/O bandwidth to support it. Helical-scan tape drives perform best when they can be kept streaming; keeping four drives waiting will probably provide less throughput than keeping one drive streaming.

2. In the event of drive failure, another drive is available.

Drive redundancy is highly desirable, both to ensure that a backup window is not missed because of drive failure and to enable backups and restores until the drive can be repaired or replaced.

Compression

Compression tape drives can boost capacity and throughput figures substantially, yet not always as substantially as some vendors' claims. When comparing different products, compare uncompressed specifications -- rates of compression will be based less on the hardware than the data you're sending it.

For typical, mixed-use, network backup a 2:1 compression ratio should be achievable. This should allow twice the data to fit on each tape at twice the transfer rate (if your hardware can support that).

Compression drives do not usually cost that much more; they represent a good buy for many sites. Keep in mind, however, that twice the capacity is not a fair trade for twice the tape slots. Compression will not double your cleaning cartridge and, unless your backups are spanning more than one tape, may not use any fewer tapes.

If your full backups span more than one tape now or in the future, compression will prevent or delay needing a larger library.

Barcode Readers

From the media expiration schedules discussed in the capacity section, it's obvious that your media pool can quickly grow very large. While the library eventually reaches a steady state, archives and off-site backups accumulate as time progresses. Managing these tapes becomes challenging for both you and your software.

Add to this that a library may need to be re-synchronized after a power failure or catastrophe. If each tape must be mounted in a drive to be identified, a 40, 60 or 120 element library will take hours. By contrast, a Spectra Logic library can read all barcodes and re-sync in less than a minute.

SPECTRA LOGIC'S TAPE LIBRARIES

Automated libraries are available in many different media formats, sizes and capacities, from Hewlett-Packard's six-cartridge DAT, to massive 1/2" (3490) libraries by Storage Tek.

Spectra Logic elected to pursue the midrange library market with two 8mm carousel tape libraries and three 4mm (DDS-2) libraries. The following will briefly describe how the Spectra libraries compare with the parameters discussed in this document.

Random access

All of Spectra Logic's tape libraries are random access.

Capacity

Spectra Logic libraries are available in the following configurations:

Spectra Logic Tape Library Configurations

Model	Format	Number of cartridges	Number of drives	Uncompressed Capacity	Compressed Capacity
Spectra 4000/20	DDS-2	20	1 - 4	80 GB	200 GB
Spectra 4000/40	DDS-2	40	1 - 4	160 GB	400 GB
Spectra 4000/60	DDS-2	60	1 - 4	240 GB	600 GB
Spectra 9000/20	8mm	20	1 - 4	140 GB	280 GB
Spectra 9000/40	8mm	40	1 - 4	280 GB	560 GB

The Spectra 4000/20 and 4000/40 can be field upgraded to a 4000/60 and the Spectra 9000/20 can be field upgraded to a 9000/40.

Reliability

Spectra Logic tape libraries have been extensively engineered and tested for reliability. The single greatest factor in their reliability is simplicity. A Spectra Tape Library utilizes only three motors. Fewer motors mean fewer moving parts, and less complex coordination.

Additional reliability comes from a fully closed loop design. The controller board's firmware knows at all times the exact position of the carousel, pickers, entry/exit door, drives and tapes. Software developed for these libraries need not initiate retry operations. Rather, when the Spectra library receives a move command, it positions a cartridge into the drive, then issues SCSI inquiries to that drive to confirm a successful load. Only after a load has been confirmed will the unit issue command completion status to the host.

The advantages of a closed loop system extend beyond normal usage. The unit even has feedback sensors to detect and report operator induced problems, such as leaving the carousel unlocked, the access lid open, cartridges loaded backward or the entry / exit door unclosed. These features not only contribute to reliability, but also to operator safety. Spectra libraries have been certified by UL, CSA, FCC & VDE.

Great emphasis has also been placed on wear avoidance. For the cartridge guides, Spectra Logic selected an engineered plastic to reduce cartridge wear. In extensive life testing, individual cartridges were loaded and unloaded over 300,000 times, from a single slot, without showing excessive wear.

Potential jams have been avoided through software-driven retries. In the event of a re-seek or jam, the libraries' firmware will make slight corrections and retry. The result is units that will typically run their entire lifetimes without jams. Spectra 4000 / 9000 libraries have been developed to provide over 1,000,000 load / unload sequences without failure.

Robotics Control Path/Data Path

All Spectra libraries include both an RS-232 interface and SCSI-2 interface to library robotics.

Number Of Drives, Compression

Spectra 4000 and 9000 tape libraries are available with one to four tape drives. Users can easily add additional drives in the field or upgrade as newer-technology drives become available. DDS-2 and EXB-8500XL tape drives all have built in compression.

Barcode Capabilities

All Spectra libraries are available with or without a barcode reader. Non-barcode units can be field upgraded to include barcode support.

SUMMARY

For many networks, automating backup can easily pay for itself in labor savings, freedom from human intervention and error and piece of mind. Important factors to consider include:

- Is the hardware compatible with the backup software I want to use?
- How fault tolerant is the complete package: does the hardware/software recover gracefully from failures?
- Will this hardware and software be sufficient to fully automate backup? What happens if I am gone one day? One week?
- Will the hardware and software meet my future storage? If not, can they be upgraded?

Procedures and Testing

Statistics abound as to the expense of reclaiming lost data and the importance of system backup.

Whatever hardware and software you select, remember the final part of a network-wide or enterprise-wide backup strategy is to constantly ascertain that all files are backed up regularly and that backup procedures are routinely tested. An occasional test restore of a directory or filesystem -- while the pressure's off -- will pay big dividends when the pressure is on. A "fire-drill" can show flaws in strategy and systems before off-line data is on-the-line.