

Using Tape Drives & Libraries For Data Backup and Storage

Roger Pozak
Boulder, CO
(303) 786-7970
roger.pozak@am.sony.com

Updated January 2000

Why Backup?

Today we are witnessing an explosion in critical data stored on desktop computers, servers and networks. According to International Data Corporation typical data storage is growing at 60% to 100% per year for most organizations. This growth is driven by applications such as spreadsheets, word-processing documents, computer-aided engineering, databases, the internet, and the fact that standard disk capacity sold in the typical computer doubles every 18 months.

What will an individual or company do if this data is lost? Without access to critical data, many businesses would suffer loss of sales, profits and customers. In fact, according to a recent study 50% of businesses that incur significant data loss go out of business within 5 years. With 32% of data loss occurring due to human error, it doesn't take a typhoon or a flood to wipe out critical data, and getting the data back from a failed hard drive can take several weeks and cost between \$20,000 and \$100,000 for just 20MB*. Data loss happens every day in every company using computers, and backing up that data ensures some or all can be recovered within a few minutes or hours.

This paper focuses on the hardware solutions for tape backup with a few words about software considerations.

Why Use Tape Drives & Tape Libraries?

Backing up data on computer disks to a single tape drive is easy and relatively inexpensive at only a few pennies per gigabyte. Tape is still the lowest cost backup media available today, and since it is a *removable media*, tape allows unlimited capacity and off-site storage options.

The total cost of tape storage changes when a single backup grows beyond the capacity of one tape cartridge. Manually managing a single tape drive and many cartridges gets very expensive in a very short time in today's computing environments. According to industry analyst Strategic Research, it costs about \$80,000 per year to manually back up and manage only 10GB of data. As a solution to this problem, manufacturers have brought to market *Tape Libraries* employing one or more tape drives and multiple tape cartridges. These robotic systems reduce the cost of manually backing up and managing files by 50% on average compared to single tape drives.

In summary, A single tape drive may save thousands of dollars in costs of reconstructing lost data. Automated tape libraries further minimize human errors, minimize the risk of lost data, and reduce the labor required to manage large amounts of data.

Basic Tape Backup Terms and Definitions

Full Backup and Incremental Backup - A full backup backs up all files on the disk(s) and usually contains a copy of the operating system, applications and other system data. An incremental backup backs up only the files that have been changed since the last backup. Most users employ the incremental backups nightly during the week and perform the larger full backup over the weekend. This saves significant time and money.

Backup Window – This is the amount of time you’re willing to let your system or network be tied up during the backup process. For a single computer this may be a couple of hours, but for a network this may be overnight or over an entire weekend.

Hot Backup - This is performed during normal work hours on systems that must operate 7x24. This type of backup application does not require a backup window, but the tape devices, processors, software and network must be tuned for minimal system overhead impact so that network speed does not become bogged down.

Restore - is the process of retrieving data, applications and even operating systems from tape and placing back on the hard drive either after a disk failure, a disaster, or simply to restore a file that was accidentally deleted.

Archiving and Hierarchical Storage Management (HSM) – These tape storage applications differ from backup in that the data is placed on tape permanently for future use.

Determining Your Site Requirements

There are a few things you must estimate before moving ahead to choosing a technology for backup, archiving or HSM tape storage. You must estimate the total amount of data (in gigabytes) being backed up during both incremental and full backups. You must also estimate the backup window (hours) for both incremental backups and full backups to ensure the tape drive or library speed meets both time frames. Finally, you must establish how fast you restore must be accomplished.

Network administrators typically focus on getting the backup done in a short amount of time, but the user only really cares about restore time. It is very important to consider restore time since this unfortunately becomes the primary job of the tape device after disaster or other data loss problem. For HSM and data archiving applications, file retrieval time, restore time, and media durability become primary criteria for selecting tape hardware and software.

Backup window and restore time is primarily affected by network speed, tape device speed, software, and level of automation (tape library). In a *local backup* the tape drive or library is connected directly to one computer or server typically via a SCSI cables, so the maximum speed is gated only by the device or SCSI specification. In a *network backup* several computers or servers are backed up over a network backbone such as FDDI, Fast Ethernet, ATM or FC-AL. Network backup speeds are typically bottlenecked by the speed of the network backbone itself rather than the tape technology or software used. Look at the *maximum sustained data rate* specifications on the tape drive, type of SCSI connections used, and/or network backbone used to determine where your bottlenecks are.

Which Tape Drive Technology to Use?

The lucrative tape drive business has attracted many world class manufacturers including Sony, Quantum, Seagate, IBM, StorageTek, Hewlett Packard, and others. Tape drives cost from under \$500 to over \$20,000, and this cost is based primarily on capacity per cartridge and data

transfer speed. Choosing the right tape drive technology is difficult, but here are a few tips to make the process less painful.

There are two fundamental methods of recording data on tape; helical scan and linear serpentine. Helical scan drives include 4mm DAT (also called DDS), Exabyte 8mm, Ecrix VXA and Sony AIT. Linear drives include Quantum DLT, Benchmark DLT1, TRAVAN, IBM Magstar, IBM 3490/3590 and Tandberg MLR/SLR. In helical scan drives the tape is partially wrapped around a rotating drum/head assembly and data tracks are written at an angle or *helical* pattern with respect to the edge of tape. In linear devices the tape passes over a head assembly that moves incrementally up and down and writes linear *serpentine* tracks parallel to the edge of the tape. Each has its advantages and disadvantages:

- Mid-range linear devices such as DLT are larger, require more power, have slow tape load & file access time, and are more expensive, but the reliability is relatively high and there is a large market acceptance.
- Mid-range helical scan devices such as AIT have fast tape load and file access time, are smaller, use less power, are very reliable, and are less expensive, but these technologies are relatively new and may require changing over from an existing technology.

The buyer can narrow their choice of tape technology by first determining the cartridge capacity and data transfer speed needed:

Cartridge Capacity = Typical Full Backup (in GB) X 3

Data Transfer Speed = Cartridge Capacity (in GB) ÷ Backup/Restore Window (in hrs)

For example, if my full backup is 10 GB and I have a 3 hour backup window, I need a tape drive with at least 3 X 10 = 30 GB and a transfer rate of 30 GB/3 hr = 10 GB/hr. Note that the backup window for both incremental backups and full backups, and the restore window, must be considered to ensure proper tape transfer rate speed.

Mid-range tape drive reliability has improved greatly since the early days of tape. Users should be wary of reliability specifications such as MTBF which often carry many disclaimers and fine print. It is best to compare all specifications including head life, media passes and duty cycle. Fig 1 shows several leading tape drives with their respective specifications.

	<u>Native Capacity</u>	<u>Native Transfer Rate</u>	<u>File Access Time</u>	<u>MTBF</u>	<u>Head Life</u>	<u>Annual Failure Rate</u>
DDS-3	12 GB	1.2 MB/s	45 sec	200,000 hrs	N/A	11%
Exabyte Mammoth	20 GB	3 MB/s	55 sec	250,000 hrs	35,000 hrs	3.1%
Quantum DLT-4000	20 GB	1.5 MB/s	68 sec	80,000 hrs	10,000 hrs	4.5%
Quantum DLT-7000	35 GB	5 MB/s	60 sec	300,000 hrs	30,000 hrs	4.5%
Sony AIT-1	35 GB	3 MB/s	27 sec	250,000 hrs	50,000 hrs	2.5%
Sony AIT-2	50 GB	6 MB/s	30 sec	250,000 hrs	50,000 hrs	N/A

Fig. 1 - Leading Tape Drive Specifications

A lower cost 4mm DAT drive will not hold up to the punishment of high duty cycle (high daily usage) as will a more expensive DLT or AIT drive. Fig 2 shows daily backup ratings for several tape drives based on the manufacturer's specified duty cycle and data transfer rates. Running tape drives beyond the specified duty cycle (backing up more per day than shown in Fig 2) may result in excessive failures.

	<u>Native Transfer Rate</u>	<u>Duty Cycle Spec.</u>	<u>Daily Backup Rate</u>
DDS-3	4.3 GB/hr	20%	20.6 GB/day
Exabyte Mammoth	10.8 GB/hr	10%	25.9 GB/day
Quantum DLT-4000	5.4 GB/hr	100%	129.6 GB/day
Quantum DLT-7000	18 GB/hr	100%	432 GB/day
Sony AIT-1	10.8 GB/hr	40%	103 GB/day
Sony AIT-2	21.6 GB/hr	40%	207 GB/day

Fig. 2 – Backup Ratings of Leading Tape Drives

Note that *native* capacity and transfer rates are shown here. All of the tape drives listed offer *data compression* that doubles the capacity and transfer rates, but actual compression ratios are highly data dependent. For example, images do not compress well at all, but repeating numbers and text is highly compressible. Two white papers that go into detail regarding capacities, data transfer speeds, reliability and cost among several leading tape technologies are *Tape Drive Technology Comparisons* and *Rating Tape Drive Technologies*. The reader can get free email copies by calling the author at (303) 786-8947. Budget considerations and getting references from other users with similar applications will help narrow the choice down to one or two.

Which Library to Use?

Tape library manufacturers include ADIC, Breece Hill, ATL, Overland Data, Spectra Logic, Hewlett Packard, IBM, StorageTek, Sony, Digital Equipment, EMASS, Qualstar, Tandberg and others. Libraries range from small autoloaders employing 1 drive and a few tape cartridges to very large systems employing 6 or more drives and over 120 cartridges. Here's how to choose one for your needs:

Determining the number of cartridges

- 1) Determine the size in gigabytes of one full backup of your data.
- 2) Multiply your full backup by 3. This will allow room for two full (weekly) backups and a number of incremental (daily) backups. If your data is growing rapidly and is not compressible, multiply by 4 or 5.
- 3) Take the value from 2 above and divide it by the native (uncompressed) capacity of the tape cartridge of the tape drive technology of your choice. This will tell you how many cartridges you need in your library.

Determining the number of drives

- 1) Determine your backup windows for both incremental and full backups.
- 2) Multiply your full backup and incremental backup capacities by at least 3.
- 3) Divide the values from 2 above by your incremental and full backup windows respectively. This larger number gives the GB/hr required.

- 4) Take the value from 3 and divide it by the your tape drive's transfer rate in GB/hr. This gives the minimum number of tape drives you'll need in your library assuming no network speed bottlenecks.

Should I Buy a Barcode Reader With the Tape Library?

The answer to this question is almost always yes. A barcode reader will provide faster cartridge inventory assessment upon power-up, will facilitate off-line media/cartridge management, and will take full advantage of backup software capabilities. This option usually costs a few hundred dollars, but it is well worth it in added data protection and ease of use.

What About Connectivity?

It is very important to specify the connector to match your server, network or disk system and to match your data transfer needs. Today, most midrange tape drives and libraries use fast/wide SCSI. Fibre channel is getting a lot of attention, and this may be worth looking into for very high speed, high-end networks with many servers located far apart across a campus.

What Software is Needed?

Leading backup software manufacturers include Veritas BackupExec, Computer Associates ARCserve, Legato Networker, HP Omniback, IBM ADSM, Spectra Logic Alexandria, Novastor, Stac and others. There are many features available, but not all features are available from some packages and you may not need certain features anyway. Although this paper focuses on hardware, here are several considerations when choosing backup software:

- 1) The backup software must be compatible with the host system, tape library and tape drives. The backup software must also work well with these various products given your site requirements and future plans for new operating systems such as Windows NT. Check the compatibility matrix carefully and do not trust the vendor's claim that "it will work" even when they never tested it.
- 2) Consider whether you will need to accommodate multiple clients, multiple servers and if you want to back up multiple servers to multiple tape drives and libraries. CPU overhead can become a real problem if you overload the system during backups, and the various software packages handle this differently.
- 3) Choose a backup software package that has the scheduling flexibility you need. Calendar based scheduling is a necessity in many sites.
- 4) Determine if you require *failover* to an alternate server if the backup server fails. Not all packages can perform failover. Also check the software's monitoring and failure alert capability. You may want to monitor the backup process, the tape drive and/or the tape library.
- 5) Carefully consider your media management and how you want to handle off-line media (cartridge) storage and tracking for archival or disaster recovery applications. The software must be able to track tapes offline based on barcode labels.
- 6) Consider whether you will need to do only "cold" backups when the network is inaccessible or if you will do "hot" backups while users access the network.
- 7) Choose a backup software package that is compatible and optimized for your database such as Oracle, Sybase, Informix, SAP, etc. .

8) Consider ease of use, ease of installation, and customer support and service.

A good book on backup software and implementation is *Windows NT Backup & Restore* by Jody Leber from O'Reilly press. Although focused on Windows NT environments, this is a comprehensive backup software reference regardless of the operating system.

Summary

Data is critical to businesses today, so backing up this data is now a standard requirement for most organizations. Tape drives are the most cost effective devices for backing up data. Tape libraries can save businesses a great deal of money in terms of automated data management and data protection. There are many types of tape technologies and tape libraries to choose from today. When choosing a tape drive or library, quantify your needs in terms of total data storage capacity, backup window and tolerance for failures. Choose your tape drive technology wisely and put an equal amount of thought into the software package that supports the library.

About the author: Roger Pozak has worked in the tape drive industry for the past 10 years. He has held positions in engineering, marketing and sales at companies including Philips LMSI, Exabyte, Spectra Logic and Sony Electronics and holds degrees in engineering and business from Michigan State University and the University of Colorado. Today he works and lives in Boulder Colorado. This and other white papers written by Roger Pozak including "Rating Tape Drive Technologies" are available by calling (303) 786-8947. Electronic copies are available.

All trademarks are owned by the respective companies. This white paper does not necessarily reflect the views of Sony, Exabyte, Quantum or any other company. This paper is updated twice each year. Copyright © 1997, 1998, 1999, 2000 Roger Pozak, all rights reserved.