

EAGLE Software has been in business since 1981 primarily providing system management and performance tools, as well as consulting and data recovery services to Data General MV customers. Over the last six years we've transitioned into the Unix market place and currently produce a file defragmentation utility which runs on at least six different Unix OS platforms.

We have a dedicated Systems Integration group at EAGLE and over the last four years, they have worked with providing integrated backup solutions to our clientele using Unix, Netware and Windows NT.

Today, we're covering network backup issues, primarily related to backups of multiple Unix computers on a network.

We will discuss software features that are considered to be important when choosing a network backup software solution. I'll also ask several open-ended questions designed help you think about potential issues of importance at your own site.

Next we'll look at some of the tape drive and library hardware options available on the market today.

We'll end up looking at network performance issues and I'll give you some guidelines which should help in determining whether or not your network bandwidth is adequate for accomplishing your backups across a network.

Let's start by looking at some of the factors driving the demand for backup products today.

This slide lists the items that are driving backup demand and contains a projection of disk storage that will be shipped through the year 2000. The items that drive backup demands are seemingly obvious to most of us.

Many of you likely have large Internet servers; Oracle, Informix, Sybase and other databases; or multi Tera-byte data warehouses.

The first multi-user computer system I worked on after college had a single 147MB disk drive. What is difficult for me to fathom now is the sheer volume of data that is 'out there' needing to be backed up. This is data at your sites. Data that is valuable to your business.

Now we have industry experts talking total disk storage in our industry in terms of Peta-bytes.

With each of us purchasing and installing so much disk storage, the next issue we have to deal with is backing up that disk storage. Because of the need for automation of multi-gigabyte backups, standalone tape drives are often not adequate. Therefore 'we' are purchasing tape libraries at record rates. Tape libraries are devices that hold many tapes and several tape drives. Those drives are typically 4mm, 8mm, DLT, or 1/2" tape devices.

What causes data loss? This slide gives several examples.

Our organization is often called on to perform data recovery. You would be surprised how often a combination of causes is involved in the most serious data loss situations.

For example, it may have been human error that incorrectly modified a backup script. It then isn't until a hardware malfunction weeks or months later that it is discovered a portion of the company's data hasn't been getting backed up.

Think about the 44% number. Only 44% of Data loss can be prevented with hardware redundancy, such as RAID technology or disk mirroring. You need backups to recover from the other 56% of the types of data loss.

At 32%, this source has one of the lowest percentages for the human error component of this chart. It's commonly as high as 72% in a LAN times article.

Let's look at one source's definition of database sizes. This should help you to classify your own site.

This is necessary because some backup solutions, whether we're talking the tape devices, backup software, or backup strategies, are better targeted at medium-sized database and file system backups while others may be best suited for large data warehouses or transaction processing environments

Let's look at some of the factors that you may want to consider as you choose backup solutions for your environment.

What type of performance do you need? Do you have a limited backup window and if so can your backup product perform within that window?

If you are doing backups while other processing is occurring, does your backup process consume too much of your system resources such as CPU cycles?

Can the backup solution grow with your business? For example, you might need a tape library solution that has room to add more tape drives as your backup needs grow.

Do you need support for multiple types of storage devices or multiple OS platforms? Most importantly, your backup software needs to support the type of devices which you've chosen to back up to, but it may also be necessary to support other devices for the ability to exchange tapes with other sites. Some software may be an NT only solution and you need NT and Unix backups. Some software may be limited to only one or two flavors of Unix: if your site has many flavors of Unix, this could be an unacceptable limitation. Choose a backup solution that has proven itself reliable and recoverable. Also look for vendor support that has displayed a successful track record of providing upgrades and customer support.

Another very important issue to consider is the backup of your company databases. If these are proprietary databases, you may have no choice but to close the database and perform backups while no processing is occurring. These are called 'cold' backups.

For many of the commercial database products, such as Oracle, it is quite common to have the ability to perform 'hot' backups. In other words some mechanism is available, either from the database vendor, or the backup software vendor, to perform backups while accesses and updates to the databases continue.

How many of your sites have computer operators 24 hours a day?

How many of your sites still use a night operator who stays until backups are done before leaving?

How many of you have sites that can still do a full system-wide backup to one piece of tape?

Human errors 32-72% of data loss.

Could your site benefit from automating the backup process, either by eliminating part or all of the computer operations night staff, or by eliminating human errors?

This about this: currently are your restores easy, or a nightmare?

Do you have to shutdown or bring users offline when you perform a backup? The answers to all of these questions may be pointing your site toward the need for greater backup automation.

Of the backup products available on the market today, there are several implementations of centralized versus distributed backup management and location of backup data and devices.

On the right of this 'management' scale are software products that can allow system managers to create, schedule and monitor backup jobs for many computers on a network from one location. A single execution of the control software, can manage jobs for many systems and potentially many backup devices. The left side of the 'management' scale contains products that would require individual machine or device management through control software that resides on the machine where the data or backup device resides. On the top of the 'data' scale are products that have the ability to backup to devices that are distributed around the corporate network, versus the lower end of the data scale being products that are geared toward backing up all your data to one centralized location.

Obviously we don't have all backup products listed here. This is intended only as a guide and to help you think through which kind of backup and data management best suites your site.

Spectra Logic's Alexandria software represents a class of high performance backup software which is typically utilizing centralized administration of many UNIX machines on a network with multiple backup servers distributed around the network.

Alexandria focus exclusively on sites that intend to utilize automated libraries, commonly know as stackers or jukeboxes.

The price range and features of this software inherently orient it toward the high-end of the Open Systems marketplace.

Legato's Networker and Cheyenne's ARCserve represent a class of software that utilize distributed backup management to distributed backup devices. These products support a wide variety of stand-alone tape drives and libraries.

These products also have been providing for a mix of Unix and non-Unix servers and clients.

Networker supports a significant number of Unix platforms and seems to be entrenched in sites that backup Netware systems across to Unix backup servers.

ARCserve has a long and rich tradition as a Netware backup utility. Recently its Windows NT and Unix product lines seem to be catching up to the Netware product and all three OS products provide centralization backups of multiple OS environments to a central backup server.

Again, there are many software product out there that are similar to these. I reference these products because I personally have experience with them and feel qualified to comment on their capabilities.

Storage Tek users tend to have backgrounds in high end IBM systems. Typically these sites have many servers performing backups to a large central silo.

Storage Tek's Automated Cartridge System Library Software, or ACS/LS manages ACS library contents and controls ACS library hardware to mount and dismount cartridges on ACS cartridge drives.

ACS libraries include the 4400, Classic, Powderhorn, Extended Store, Wolfcreek and Timberwolf.

Storage Tek drives include the 4480, Silverton, Timberline and the Redwood drive which has a data transfer rate of 12 MB/second and each cartridge holds 20-100 GB.

Here is a client server backup model with a desirable configuration:

It would utilize centralized control, meaning you can control any of the backup servers or clients across a distributed network from a single point if you desire to.

It includes distributed fault tolerance: meaning one server, or one device, going down won't prevent any specific client backups from occurring.

This model also allows Peer-to-Peer backups to occur: meaning if the library on Server2 goes down, the backups of the data on Server2 could be allowed to go to Server1 or Server3.

And finally, implementation of this backup model should not include logical boundaries between backup servers. In other words Clients 1-4 for example should not be restricted to only using Server1. They should have the flexibility to send their data to any server based on server activity and device availability.

Here are some additional items that I think are highly desirable in your backup software:

Backup windows often times don't allow for full backups every night, so the ability to do full or incremental backups is critical.

Scheduling flexibility is essential for full and incremental backups. You may wish to have the jobs launch on the client or the server and have that same flexibility with where the databases are stored. Scheduling jobs should also be something that can be intelligently handled by a scheduling daemon, or initiated by another scheduling package or even manually performed by system operators.

Handling open file conditions is also important. Should the software skip open files, try repeatedly to back them up without them changing, or should it initiate an application or routine which would close the file or bring it to a non-active state?

ARCserve has a nice graphical user interface which makes it easy to select drives, directories or files; however it's filtering methods are not easy to implement.

Alexandria doesn't currently have the nice GUI, but it has lots of flexibility for specifying files by users, groups and filters (if you are proficient writing regular expressions).

Additionally, file restoration needs to have flexibility with scheduling and with who initiates the restores.

A critical piece in your ability to recover from your backup tapes, is the data in the backup package's database. I would look for a package that has some level of database fault tolerance.

This fault tolerance can be achieved simply by keeping the database in multiple locations. For example the obvious location in which to keep the database is on disk.

Since that database is basically another file, or set of files, that need to be backed up, there should be provisions in the software to backup its own database to tape. Some packages even backup their own executables each time it does a full database backup.

My suggestion would be to have the database periodically backed up to its own tape and that tape be periodically updated in vault or off-site storage.

Some software packages backup their own databases at the end of every backup, after the selected files are successfully backed up. Other packages put database information, about the files which were just backed up, after each tape file, or cluster. This is demonstrated on the next slide.

Database information pertaining to the files which were backed up in Cluster1 is written to tape in the DB cluster immediately after Cluster1. This allows recovery in a worst case scenario by reading the database clusters from the source tapes and repopulating a database on disk from that information.

Since backup databases can potentially grow unchecked, this tape format also can allow you to purge detail data from the 'on disk' database but preserves the detail DB information until the point at which the tape itself is overwritten.

This type of cluster format has several other advantages:

1. It can provide for very quick restore performance by taking advantage of a tape drive's ability to fast-forward to a specific tape file or cluster.
2. It allows backups to span multiple media elements.
3. And it facilitates worst case disaster recovery operations on machines other than the one that created the backup tape.

Other fault tolerance issues to consider in your backup software:

Can it automatically recover from media, drive, library or system failures? Some packages terminate the backup job if there is a tape error while others will continue the backup to a new tape.

Another consideration; should the process continue to a new backup hardware device if there is a hardware failure?

What about system crashes or failures? Do the databases automatically recover or does it take intervention from the Sysadmin to recover and restart the backup processing?

Does the software provide for tracking multiple copies of the user data? For example it may be necessary to keep copies of your backup both on- and off-site. My preference would be that the backup software assist in my tracking of the off-site media as well as the media which is retained in a robotic stacker or library.

No Unix backup software package should be considered adequate unless it maintains the *user*, *group* and *other* read, write and execute permission masks during backups and restores.

Security within the software itself is another issue which can be overlooked. Do you need to restrict access to certain functions within the backup package by *user* and *group*? How about access to certain backup devices on the network?

While not at first obvious, utilizing a proprietary media format can provide a unique form of security because it limits restoration of data from the tapes to only systems with the specific backup software.

And again to emphasize the critical issues of performance and reliability. Look for independent evaluations and benchmarks. Look for reviews in trade magazines or independent journals.

Often these reviews will evaluate performance issues such as CPU overhead and data throughput. They also consider issues such as the reliability of the store and restore processes and procedures.

Don't forget that performance is not only measured in data transfer rate. Other factors include media selection time, find or filtering procedures, media manipulation and database updates. For example database updates are typically very comparable among software products at small or mid-sized installations. But products such as Alexandria tend to outperform the others in Large and Very Large DB installations.

Other items to consider in your software selection include:

The user interface--Do you need a particular front end such as X11/Motif or HTML? Do you need command line access for remote dial-in or text-only based terminals.

The tape format. Often times a backup vendors proprietary format will provide additional features not supported with industry standard formats such as tar or cpio.

Bar Code support--Tracking tapes is much easier if they are barcoded, plus backup software is typically more efficient if it can inventory robotic libraries by scanning barcode labels. Consider taking the inventory of a 60 tape library if it takes 2.5 minutes per tape without barcode labels. You'll be waiting 2.5 hours.

Device Management--Does the software provide flexibility in selecting drives or libraries? Can you write to more than one device at a time? Especially with drives that use Metal Particle media, regular cleaning of the tape heads is essential. Will the software automate the drive cleaning schedule? My personal opinion is that it should perform drive cleaning based on the number of hours of usage on the drive. Some software that I've worked with will only let you schedule drive cleaning based on a specific time of the day or day of the week. Other software doesn't even automate the tape drive cleaning process!

Lastly, if you have a problem with drive errors, can you optionally take the drive off-line or does the software continue to try to use it?

Let's shift gears now and talk about backup hardware. This information includes performance, capacity, reliability and cost information on several current technology tape drives.

This graph is a quick review of the types of SCSI controllers and devices available. Installation and performance must be considered when selecting a SCSI type to best fit your requirements.

Backup device types should also be considered as through put and capacity differ.

AIT, 8mm and DLT encompass two basic types of media: MP (Metal Particle) technology is used in DLT-4000 & DLT-7000, and AME (Advanced Metal Evaporated) technology is used in Mammoth and Sony AIT.

Both media types contain a base film and a recording layer of magnetic (metal) material. The key difference being the MP recording layer is composed of about 45% magnetic material mixed with a binder and other additives, while the AME media recording layer is 80% to 100% magnetic material which allows higher recording densities.

The AME media also contains a smooth, protective diamond-like carbon coating on the recording surface, a lubricant, and a protective back coating for greater durability.

The unique features of the AME media give it an advantage over MP media in terms of the number of passes it can withstand without degradation.

DLT-4000 and DLT-7000 use the same MP tape cartridge and are read compatible with each other. The AME media use and compatibility is as follows:

- There is one AME tape cartridge specifically designed for the Sony AIT drive, and there is another for Exabyte's Mammoth drive.
- The Sony AIT drive accepts only AME media made for the AIT drive and will reject all other 8mm tape cartridges.
- AME tapes written by a Sony AIT drive are write protected from other 8mm drives including Mammoth.
- Exabyte's Mammoth product will read both AME and MP media but will not write to the MP media.

It is interesting to note that Exabyte's Mammoth drive will actually read both AME and MP media. Exabyte designed Mammoth to read both media types so that it would be read compatible with existing MP 8mm tapes.

This compatibility goal forced special read/write head material design requirements and necessitates special cleaning practices by the user.

For example, if an MP tape is read by the Mammoth drive, the drive will not accept another tape until a cleaning cartridge has been used. Cleaning is required because the MP media binder chemistry is more prone to leave debris on the head and in the tape path.

This raises a question as to how many times a Mammoth drive can read an MP tape without suffering permanent damage. The implications here are even more unappealing with respect to tape libraries in which tapes are read many times. It is perhaps more realistic for Mammoth users to transition to the AME media and avoid the implications of using MP media.

This slide is an overview of some of the issue one should consider before purchasing a tape library.

The issue of tape drive technology is listed at the top because it is one of the most important considerations when selecting a tape library. Your backups can only be as good as your media and drive allow them to be. The library design is also important, and should be of simple design (KISS), and scalable to your needs. Here is a list of some tape libraries. It's not complete but can be used as a reference.

Sizing of the library is also very important. Listed are some of the factors that need to be considered when selecting the right size library.

One should also keep in mind the software being used to manage this library. Questions like, does the software allow for multiple backup to exist on one tape. A larger tape capacity can be wasted if only one backup per tape is allowed by the controlling software.

Like so many things in this world, you get what you pay for. Bandwidth is a way to measure the speed of the network your computers use to communicate. Higher bandwidth, or more speed, costs more but can potentially reduce the amount of time needed to perform your backups.

This slide shows, best case, the amount of data that can be backed up across a 10 baseT network.

This slide shows, again best case, how much data can be backed up across a 100 BaseT network.

This slide shows capacities of comparable 4-drive tape libraries with less than 60 slots available on the market today.

- The Spectra 10000S uses the Sony AIT (SDX-300C) drive
- The EXB-440 uses the Exabyte Mammoth (EXB-8900).
- Breece Hill, Odetics, ADIC and StorageTek use the DLT-4000 (the DLT-7000 is not yet available).

- Sony AIT per tape capacity is 25 Gig native and 65 Gig with ALDC 2.6:1 compression.
- DLT-4000 per tape capacity is 20 Gig native and 40 Gig with DLZ 2:1 compression.
- Mammoth per tape capacity is 20 Gig native and 40 Gig with IDRC 2:1 compression.

Library native capacities vary due to the number of cartridges used in each library.

This slide shows capacities of comparable 2-drive tape libraries available on the market today.

- The Spectra 10000S uses the Sony AIT (SDX-300C) drive

- The EXB-220 uses the Exabyte Mammoth (EXB-8900).
- Breece Hill, Odetics, and ADIC use the DLT-4000 (the DLT-7000 is not yet available).

- Sony AIT per tape capacity is 25 Gig native and 65 Gig with ALDC 2.6:1 compression.
- DLT-4000 per tape capacity is 20 Gig native and 40 Gig with DLZ 2:1 compression.
- Mammoth per tape capacity is 20 Gig native and 40 Gig with IDRC 2:1 compression.

Library native capacities vary due to the number of cartridges used in each library.

This slide shows the performance of comparable 4-drive tape libraries available on the market today.

- The Spectra 10000S uses the Sony AIT (SDX-300C) drive
- The EXB-440 uses the Exabyte Mammoth (EXB-8900).
- Breece Hill, Odetics, ADIC and StorageTek use the DLT-4000 (the DLT-7000 is not yet available).

- Sony AIT data transfer rate is 3MB/sec native and 7.8MB/sec with ALDC 2.6:1 compression.
- DLT-4000 data transfer rate is 1.5MB/sec native and 3MB/sec with LZW 2:1 compression.

Mammoth data transfer rate is 3MB/sec native and 6MB/sec with IDRC 2:1 compression.

This slide shows the performance of comparable 2-drive tape libraries available on the market today.

- The Spectra 10000S uses the Sony AIT (SDX-300C) drive
- The EXB-220 uses the Exabyte Mammoth (EXB-8900).
- Breece Hill, Odetics, and ADIC use the DLT-4000 (the DLT-7000 is not yet available).

- Sony AIT data transfer rate is 3MB/sec native and 7.8MB/sec with ALDC 2.6:1 compression.
- DLT-4000 data transfer rate is 1.5MB/sec native and 3MB/sec with LZW 2:1 compression.
- Mammoth data transfer rate is 3MB/sec native and 6MB/sec with IDRC 2:1 compression.

Once again comparisons were made to 4 drive tape libraries AVAILABLE TODAY. Domestic list price divided by total native capacity was used to obtain the figure.