Choosing a Backup Environment

Jody Schivley Leber
Lehman Brothers, Inc.

Abstract

Backups are the most important task you can do to protect the vital data on your computer. The basic hardware, operating system, and application software can be replaced at a relatively small price. However, the data and programs you or your co-workers have generated cannot be reproduced easily. Hours, weeks, months, and years are embodied in this data. There is no possible price someone can put on that kind of time and effort. Proper backups are the only way you can guarantee restoration of your data subsequent to system failures.

Some people may consider it a mundane task. However, at many sites there are technical issues, political concerns, and integration challenges that make the assignment of implementing the corporate backup and recovery system(s) interesting and fun.

This paper provides a summary of the process involved with choosing a backup environment for Windows NT. The details are fully covered in the book, *Windows NT Backup and Recovery* that will be published by O'Reilly and Associates in the fall of 1997. Due to the scope of planning and implementing a successful backup system, not all the information necessary can be provided in this short paper. Where the information is too great to fit in this paper, you will be referred to the book.

1. Local Versus Networked Backups

The size of your site determines the type of backups you will implement. You either have local backups or networked backups. There is a variety of software and hardware available today to implement either type of backup system.

When determining which backup model to implement, there are interesting aspects that should be considered. Your site will do far more backups than restores. Due to this fact, the engineering must primarily focus on accomplishing the backups in a timely manner.

In contrast, the users are not concerned with how fast the backups are performed; they are concerned with how fast they can restore their data. The restores are what they are directly exposed to and use. As it relates to restores, the majority of restores performed are restoring a single file or a single directory. This is in contrast to restoring entire file systems or hard disk. However, losing a single file or directory generally does not cause as much corporate exposure and potential business loss as losing an entire hard disk.

To overcome this potential problem, it is important to implement backups with other, complementary technology. This technology includes:

- Uninterruptible Power Supply (UPS)
- Disk Mirroring
- Disk Duplexing
- Disk Striping

1.1 Local Backups

The local backup implementation is one in which backups are performed on individual systems. Each computer is connected to its own backup device and each computer is separately configured. These individual computers may or may not be connected to a network. The key is that the network or additional computers are not required for backups. This is a good way to perform backups in your home or at a small office.

If you choose this implementation, the backup configuration and tapes are co-located with the corresponding system. In the case of disasters, such as flood, fire, or burglary, the system and corresponding tapes may be damaged or destroyed. Therefore, you must move the tapes to another location to ensure you have the ability to recover after a system or site disaster. Additionally, local backups are not very scalable because they require systems to be individually configured and the tapes to be independently managed. As your site grows larger, it will require more and more manpower to adequately support.

The initial capital cost to implement local backups is relatively low. No network is required, inexpensive tape drives can be installed on each system, and you can choose to use free software. The Microsoft Tape Backup Utility (TBU) is packaged with Windows NT and does not have to be purchased separately.

1.2 Networked Backups

Networked backups are implemented by grouping computers. This implementation is generally referred to as departmental backups or centralized backups. The data is sent over the network to a designated computer where there is a shared tape drive or robotic tape library. The systems performing the backups are called backup servers and the other computers are called backup clients. You enter configuration details and manage your tapes for the backups on each backup server. The backup server(s) can be located in a secure and offsite area to reduce accessibility, tampering, or theft. Also, backup configuration and tape management is reduced because you have relatively few backup server systems to handle.

With departmental backups, computers are grouped by LANs, departments, floors, or some other logical grouping. With this implementation, if possible, the backup servers would be secured in a computer equipment room or closet.

With centralized backups, a larger number of backup clients send their data to a relatively small number of backup servers (this is contrasted with the departmental backups). Sometimes this strategy groups computers, but the groups are much larger and designated by business unit, building, or region.

With the centralized backup implementation, your backup server(s) can be located at a completely different site and placed in a machine room with other important systems. There are instances when a backup server located at site A, backs up the backup clients at site B and vice versa. This provides security and immediate off-site backups. With this implementation, your central operations group can handle the backup configuration and tape management for all backup servers.

Your initial capital investment for the networked backups is higher than for the local backups. Some of those costs include the following:

- The network may need to be expanded to handle the additional load (networked backups will place a significant load on your network infrastructure), even though your site may already have a large network infrastructure in place. You may choose to install a dedicated network for backups.
- TBU software cannot be used, since you are utilizing the network. So for this implementation, you have to purchase third party software.
- Purchasing a tape library is a definite requirement for centralized backups and desirable for departmental backups. The cost of robotic tape libraries varies greatly, but is definitely more expensive than individual tape drives.

With the centralized backups, there are a few other costs and considerations involved and these are listed below:

- Your backup server system(s) should be used exclusively for backups.
- Depending on your site’s size, the centralized backup implementation may require dedicated staff to effectively operate. Although, the hidden cost of many people working part time on backups may actually be more than the dedicated staff, both in terms of cost and productivity.
- In some organizations, it may be difficult to achieve central backups because you are crossing organizational boundaries. Experience has shown that a mandate or visible support from your CIO-level management or equivalent is sometimes the only way centralized backups can be successfully implemented.

2. Phases

Properly performing backups, as well as properly planned backups, is the key to successful restores. A methodical approach is also how the backup implementation can be successful. This approach should include specific phases you should accomplish for the backup and recovery project at your site. These phases are briefly described below. There may be portions of the phases not appropriate for your site initially, but sites do change and grow, so you should consider all of them carefully.

As you will see, the different phases described below all work together. Each phase feeds vital information into other phases. Understanding the global picture of what needs to be done will make the details easier to comprehend.

2.1 Phase 1: Requirements Definition

There are many requirements that have to be considered to properly choose the backup software and hardware to be used. These items include what type of data and how much data you think is important, how much time you have to do backups, and your restore expectations. The *Windows NT Backup and Recovery* book will provide a checklist of items to help you step through your site to determine what is currently in place or planned to be implemented. This site survey helps you ask the appropriate questions. As they apply to backups and recovers, and why the questions are pertinent.

When completed, the results of the site survey will guide you in appropriately sizing the backup system(s) required for your site and assist you in determining if your network infrastructure is adequate to handle the additional load. The survey also assists in determining what software and hardware to use for your backup server as well as helping you to perform testing.

It is important to keep one concept in mind as you do the site survey and sizing. Rarely in the history of computing has the amount of data at a particular site been reduced. The disk capacity grows and the data grows. Disk drive sizes are rising and prices are falling. Some sites have reported that the distributed environment is growing at over 100% per year in data storage. With this type of growth, you need to plan for it and not implement a backup solution that you outgrow immediately. You don't want the amount of data to back up, the number of systems to backup, or the size of the backup server to be obsolete in the first 18 months. Of course, you need to use common sense, but definitely don't minimize your requirements.

2.2 Phase 2: Policy and Politics

Backups and restores are not strictly technical in nature. The non-technical aspects include writing a backup policy, planning corporate education, and possibly implementing a charge-back policy. Backup projects fail or are ineffective, more often than not, due to political and organizational issues rather than technical problems. Taking these into consideration in the early phases of planning will assist in having your project more widely accepted and more successful

A backup policy statement checklist will be provided in the *Windows NT Backup and Recovery* book. It lists the items that should be in your backup policy. The information obtained in the requirement definition phase assists in writing your policy. When implementing a backup solution for your corporation, you should also plan educational sessions appropriate for your users and your site.

2.3 Phase 3: Software

Once you have defined your requirements and policies, you need to examine the various software options available. This phase covers the backup and restore software aspects.

The Windows NT TBU software is shipped with the base operating system, but it has a limited feature set and is not scalable. There is a wide range of third party software that you can purchase. To find the package that bests suites your site, you have to take the time to understand the design characteristics and features of the different third party software. The feature and functionality of the backup software varies greatly. The *Windows NT Backup and Recovery* book will provide a checklist of features to consider to help you sort through them. It is important to primarily base your choice of software vendor on one who most closely meets those features and functionality that are most significant to your site.

2.4 Phase 4: Hardware

Like the software, the backup hardware should be customized to your site. The amount of data to be backed up, your backup window, and tape management are important factors in purchasing hardware that you will not outgrow too fast.

Backups can be performed using different media. You can backup data to floppy diskettes, CDs, hard disks, or magnetic tape. By far, the most popular media for backups is magnetic tape. This is predominately because of the cost of the media, the ease of implementation, and the convenience of physically storing the tapes.

With local backup implementations, every computer backs up itself and therefore the backup server hardware is not a concern. Additionally, there is no need for tape libraries at small sites. Generally, there are individual tape drives attached to each system and no tape libraries. So, for local backups, the tape drive hardware may be the only item of consideration. Different size tapes and different speed tape drives may be utilized for different systems, depending on how much data has to be backed up.

For the networked backup implementations, the performance of the backup server and the tape drive hardware is an important consideration. The appropriate backup throughput is what will allow you to meet the backup window and the required restore speeds will allow you to quickly recover lost data. For ease of tape management, sites with networked backups should consider an implementation using tape library technology.

The *Windows NT Backup and Recovery* book will provide a matrix of different tape drive technologies as well as a checklist of features to consider when purchasing a tape library.

2.5 Phase 5: Testing

Once the hardware platform is chosen and the field of commercial software is narrowed down to three or fewer vendors, the testing can commence. You should not skip the testing phase because you will uncover some specific issues that you cannot uncover any other way. The vendors perform as much testing as possible, but remember your site configuration is not exactly the same as the vendor's quality assurance test lab. Actually, most vendors have a limited test capability compared to all of the technical options available today. There are thousands of different combinations of hardware and software and no two companies are identical. Each site has different PC hardware platforms, different network components integrated together, and different combinations of installed software. Therefore, testing at your site is important to the success of your backup and restore implementation.

During the test phase, you should evaluate the features of the application software as well as the performance of the overall system. The performance results of the on site testing should be compared to the server and network sizing goals obtained in Phase 1, Requirements Definition. The comparison of the live testing results to these performance goals will help determine if the system, as a whole, meets your throughput requirements. The feature, performance, and pricing information will all help you make the best choice of backup and restore software to purchase for your site.

Testing will take time, but you will learn a substantial amount about the product. After completing the testing process, you will better understand what is involved with installation and configuration of the product and you will understand what has to be done in the event of an emergency situation (such as a loss of a hard drive in which you experience catastrophic data loss).

The steps involved with recovering data after a crisis situation are not always well defined. The testing process allows you to perform and refine the steps involved with restores from catastrophic data loss. This will make you more comfortable with this type of situation. Inevitably, you will sustain a complete system failure. Knowing the restore procedures ahead of time can make a stressful situation a little bit easier to handle.

The *Windows NT Backup and Recovery* book will step you through the basic functionality and performance testing procedures that should minimally be done at your site.

2.6 Phase 6: Integration

After testing, it is time to integrate the backup software and backup hardware chosen. Integration may be viewed as a one-time occurrence; however, this may not necessarily be true. If you are successful with the backup solution you implement, you may be asked to implement your solution at other sites within your company. Also, as your organization grows you will have to expand your backup implementation to support the growth. Proper integration is important and must consider many different aspects of your organization.

With some projects, testing and integration are run together into one phase. Don't let this happen. Testing and integration are distinctly different phases and should be treated as such:

- Testing is geared towards evaluating product features, library features, and product performance and comparing different products. It also allows you become familiar with the installation, configuration, and operation of the software and hardware. The goal of testing is to determine the hardware and software best suited for your environment.

- Integration consists of steps required to get the backups operational in a your production environment. You ultimately want the backup software and hardware in production and then into an administrative maintenance mode. The goal of integration is to perform the installation and configuration of the backup environment with production in mind.

There are particular guidelines that can be given to you to help you initially bring up the backup server based backups. The details of the integration are very dependent on the hardware and software you have chosen. The *Windows NT Backup and Recovery* book will provide basic generic guidelines to help you integrate the hardware and software at your site.

2.7 Phase 7: Administration

Administrative maintenance is required for the life of your backup environment, on an ongoing basis. This maintenance can be substantial work, depending on the size of your site. Obviously, the larger the site, the larger the job.

These are the tasks that must be done occasionally and other tasks that must be done on a daily basis. The occasional tasks are commonly overlooked, but are just as important as the daily tasks. You cannot just put the backup solution in place and ignore it; otherwise you will be caught by surprise down the road. The *Windows NT Backup and Recovery* book will provide an initial list of both types of tasks that should be considered for your site.

3. Disaster Recovery

Disaster recovery and backup and recovery are related topics. Disaster recovery is the preparation for, and procedures to take in the event of, a disaster. Disaster recovery is a huge topic in and of itself. Backup and recovery are just the data recovery components of disaster recovery.

During the planning of the backup implementation, the group in your organization responsible for disaster recover should be involved with your work. They may have requirements that you can easily implement with planning up front. This is in contrast to waiting until after your implementation and having their requirements difficult, or perhaps impossible, to integrate.

Summary

Backups are an essential part of any enterprise computing strategy. Careful analysis, planning, testing and implementation all contribute to the overall success of any backup project. Attempts to shortcut any of the phases outlined here have the potential to lead to an ineffective implementation at best or project failure at worst.