Chapter 1

An Introduction to Data Protection Today

1.1 INTRODUCTION

As we start our discussion of the future of data protection, we would like to spend some time taking a look at data protection today and establishing some of the basic terminology that is commonly used. This will be a review for most, but it helps avoid confusion with some of the terms and usage. It also helps set the groundwork for looking to the future. We will start out this discussion by looking at the traditional backup and recovery.

1.2 TRADITIONAL BACKUP AND RECOVERY

When we talk about data protection today, we usually talk about the traditional backup and recovery, generally, the process of making secondary copies of production data onto tape medium. This discussion might also include some kind of vaulting process. This has been the standard for many years and to an extent continues to meet the foundational requirement of many organizations; that being an ability to recover data to a known-good point in time following a data outage, which may be caused by disaster, corruption, errant deletion or hardware failure. There are several books available that cover this form of data protection, including *UNIX Backup and Recovery* by W. Curtis

Digital Data Integrity David Little, Skip Farmer and Oussama El-Hilali

Preston (author), Gigi Estabrook (editor), published by O'Reilly and *Implementing Backup and Recovery: The Readiness Guide for the Enterprise* by David Little and David Chapa, published by John Wiley & Sons. To quote from the very first chapter in *Implementing Backup and Recovery: The Readiness Guide for the Enterprise*, 'A *backup* is a copy of a defined set of data, ideally as it exists at a point in time. It is central to any data protection architecture. In a well-run information services operation, backups are stored at a physical distance from operational data, usually on tape or other removable media, so that they can survive events that destroy or corrupt operational databases.'

The primary goals of the backup are to be able to do the following:

- Enable normal services to resume as quickly as is physically possible after any system component failure or application error.
- Enable data to be delivered to where it is needed, when it is needed.
- Meet the regulatory and business data retention requirements.
- Meet recovery goals, and in the event of a disaster, return the business to the required operational level.

To achieve these goals, the backup and recovery solution must be able to do the following:

- Make copies of all the data, regardless of the type or structure or platform upon which it is stored, or application from which it is born.
- Manage the media that contain these copies, and in the case of tape, track the media regardless of the number or location.
- Provide the ability to make additional copies of the data.
- Scale as the enterprise scales, so that the technology can remain cost effective.

At first glance this seems like a simple task. You just take a look at the data, determine what is critical, and decide on a schedule to back it up that will have minimal impact on production, install the backup application and start protecting the data. No problem, right? Well, the problem is in the details. Even the most obvious step, determining what is the most critical data can be a significant task. If you ask just about any application owner about the criticality of their data, they will usually say 'Mine is the most important to the organization.' What generally must happen is that you will be presented with various analysis summaries of the business units or own the task of interviewing the business unit managers yourself in order to have them determine the data, the window in which backup may run, and the retention level of the data once it is backed up. What you are doing is preparing a *business impact analysis* (BIA). We will discuss the BIA later in this chapter when we discuss disaster recovery (DR) planning. This planning should yield some results that are useful for the policy-making process. The results of these reports should also help define the recovery window, should a particular business unit suffer a disaster. The knowledge of these requirements may, in fact, change the budget structure for your backup environment, so it is imperative during the design and architecture phase that you have some understanding of what the business goals are with regard to recovery. This can help you avoid a common issue faced by the information technology (IT) staff when architecting a backup solution, paying too much attention to the backup portion of the solution and not giving enough thought to the recovery requirements. This issue can easily result in the data being protected but not available in a timely manner. This issue can be compounded by not having a clear understanding of the actual business requirements of the different kinds of data within an enterprise which will usually dictate the recovery requirements and therefore the best method for backing up the data. You should always remember that the primary reason to make a backup copy of any data is to be able to restore that data should the original copy be lost or damaged.

In many cases, this type of data protection is actually an afterthought, not a truly thought-out and architected solution. All too often when a data loss occurs, it is discovered that the backup architecture is flawed in that the data was either not being backed up at all or not being backed up often enough resulting in the recovery requirements not being met. This is what led us to start recommending that all backup solutions be architected based on the recovery requirements. As mentioned above, BIA will help you avoid this trap.

When you actually start architecting a backup and recovery solution as a part of the overall data protection scheme, you start looking at things such as

- Why is the data being backed up?
  - Business requirements.
  - Disaster recovery (DR).
  - Protection from application failures.

  - Protection from user errors.
  - Specific service level agreements (SLAs).
  - Legal requirements.
- What is the best backup strategy to meet the recovery requirements?
  - Backup frequency.
  - Backup type: full, differential incremental or cumulative incremental.
  - Data retention.
  - Off-site storage of images.

As you look at all these different elements that are used to make the architectural decisions, you should never loose sight of the fact that there is usually an application associated with the data being backed up and the total application must be protected and be recoverable. Never fear, the true measure of a backup and recovery system is the restorability of the data, applications and systems. If your backup and recovery solution allows the business units to meet or exceed their recovery SLAs, you will get the kind of attention we all desire.

Although a properly architected backup and recovery solution is still an important part of any data protection scheme, it is becoming apparent that the data requirements within the enterprise today require some changes to address these new requirements and challenges. Some of the changes are

- total amount of data;
- criticality of data;
- complexity of data, from databases, multi-tier applications as well as massive proliferation of unstructured data and rich media content;
- complexity of storage infrastructure, including storage area networks (SAN), network attached storage (NAS) and direct attached storage (DAS), with a lack of standards to enforce consistency in the management of the storage devices;
- heterogeneous server platforms, including the increased presence of Linux in the production server mix;
- recovery time objectives (RTO);
- recovery point objectives (RPO).

These requirements are starting to stress the traditional data protection methodology. The backup and recovery applications have been adding features to give the data owners more tools to help them address these issues. We will discuss some of these in the following chapters.

1.3 HIERARCHICAL STORAGE MIGRATION (HSM)

HSM is another method of data management/data protection that has been available for customers to use and is a separate function from tradition backup, but it does augment backup. With a properly implemented HSM product that works with the backup solution, you can greatly reduce the amount of data that must be managed and protected by the backup application. This is accomplished by the HSM product managing the file system and by migrating off at least one copy of inactive data to secondary storage. This makes more disk space available to the file system and also reduces the amount of data that will be backed up by the backup application. It is very important if implementing an HSM solution to ensure that the backup product and the HSM product work together so that the backup product will not cause migrated files to be recalled.

A properly implemented HSM application in conjunction with a backup application will reduce the amount of time required to do full backups and also have a similar effect on the full restore of a system. If the backup application knows that the data has been migrated and therefore only backs up the placeholder, then on a full restore only the placeholders need to be restored. The active files, normally the ones you are most concerned with, will be fully restored and restored faster as the restore does not have to worry with the migrated inactive data. Retrieving migrated data objects from nearline or offline storage when an application does access them can be more time consuming than accessing directly from online storage. HSM is thus essentially a trade-off between the benefits of migrating inactive data objects from online storage and the potentially longer response time to retrieve the objects when they are accessed. HSM software packages implement elaborate user-definable policies to give storage administrators control over which data objects may be migrated and the conditions under which they are moved.

There are several benefits of using an HSM solution. As previously stated, every system has some amount of inactive data. If you can determine what the realistic online requirements are for this data, then you can develop an HSM strategy to migrate the appropriate data to nearline or offline storage. This results in the following benefits:

- reduced requirements for online storage;
- reduced file system management;
- reduced costs of backup media;
- reduced management costs.

HSM solutions have not been widely accepted or implemented. This is mostly due to the complexity of the solutions. Most of these applications actually integrate with the operating system and actively manage the file systems. This increases the complexity of implementing the solution. It also tends to make people more nervous about implementing an HSM product. This is probably one of the least understood product of the traditional data protection and management products.

1.4 DISASTER RECOVERY

Another key ingredient of the traditional data protection scheme is DR. In the past, this was mostly dependent on a collection of backup tapes that were stored either at a remote location or with a vaulting vendor. In many instances, there was no formal planning or testing of the DR plan and procedures. As you might expect, many of these plans did not work as desired. Recently, more emphasis has been given to DR and more people are not only making formal plans but also conducting regular DR tests to ensure that they can accomplish the required service levels. We have always said that until your DR plan is tested and demonstrated to do what is needed, you do not have a plan at all.

As stated earlier in this chapter, do not succumb to the temptation to concentrate too much on the raw data and forget about the overall production environment that uses the data. If the critical data exists within a database environment, the data itself will not do you much good without the database also being recovered. The database is of only marginal value if all the input comes from another front-end application. As you put together a DR plan, you should always try to remember the big picture. Too often people concentrate on just recovering specific pieces without considering all the interdependences. By developing the BIA mentioned earlier you can avoid a lot of the potential pitfalls. One of the interesting results of gathering the proper data necessary to do the BIA can be a change in the overall way you architect backup and recovery for your enterprise. An example of this is a customer who discovered they were retaining too much data for too long a period of time due to lack of a business analysis of the data looking at both it's immediate value, the effects time had on the value of the data, and the potential liability of keeping too much data around too long. After doing the BIA the customer reworked their retention policy and actually experienced a sizeable cost savings by putting cartridges back into circulation.

The BIA is basically a methodology that helps to identify the impact of losing access to a particular system or application to your organization. This actually is a process that is primarily information-gathering. In the end, you will take away several key components for each of the business units you have worked with, some of which we have listed here:

1. Determine the criticality a particular system or application has to the organization.
2. Learn how quickly the system or application must be recovered in order to minimize the company's risk of exposure.
3. Determine how current the data must be at the time of recovery.

This information is essential to your DR and backup plans, as it describes the business requirements for backup and recovery. If you base your architecture on this information and use it as the basis for your DR plan, your probability of success is much greater. Another by-product of the BIA and the DR plan is developing a much better working relationship between the business units, application owners and the IT staff.

With the growing emphasis on DR and high availability, we begin seeing the mingling of data protection and data management techniques. Users started clustering local applications and replicating data both locally and remotely. We will discuss these in detail in a later chapter. RTO and RPO requirements are two key elements to consider when making the decision on which technique to use for DR, as seen in Figure 1.1.

As history has shown us, there are many different kinds of disasters, and a proper DR plan should address them. The requirements can be very different for the different scenarios. There is an excellent book that can be very helpful in preparing a good DR plan. It is *The Resilient Enterprise: Recovering Enterprise Information Services from Disasters* from VERITAS Software publishing.

1.5 VAULTING

Any discussions that concern DR should also include a discussion about the vaulting process. In very basic terms, a vaulting process is

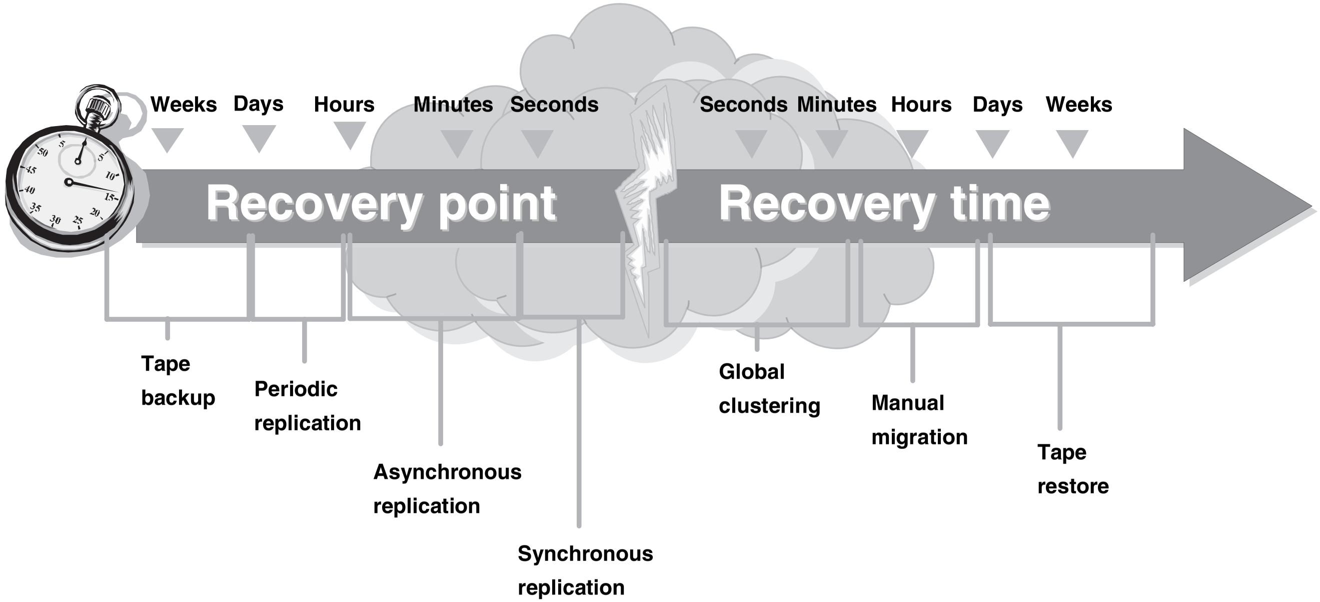

Figure 1.1 RPO/RTO

the process that allows you to manage and accomplish any or all of the following steps:

- Create a duplicate of the backup image for storage off-site.
- Automate ejecting the media that need to be taken off-site.
- Provide reports that allow you to track the location of all backup media.
- Manage recalling media from the off-site location, either for active restores or for recycling after all the data on the media has expired.

It is possible to develop all the tools and procedures to accomplish all of these tasks, but it can be a tedious and potentially risky endeavour. Some of the backup applications offer a vaulting option that is fully integrated with the backup solution, which is much easier to use and more reliable. Figure 1.2 shows the basic vaulting process flow.

There are at least three options for creating a backup image that will be taken off-site for secure storage:

- Take the media containing the original backup images off-site.
- Create multiple copies of the backup image during the initial backup.
- Have the vaulting process duplicate the appropriate backup images.

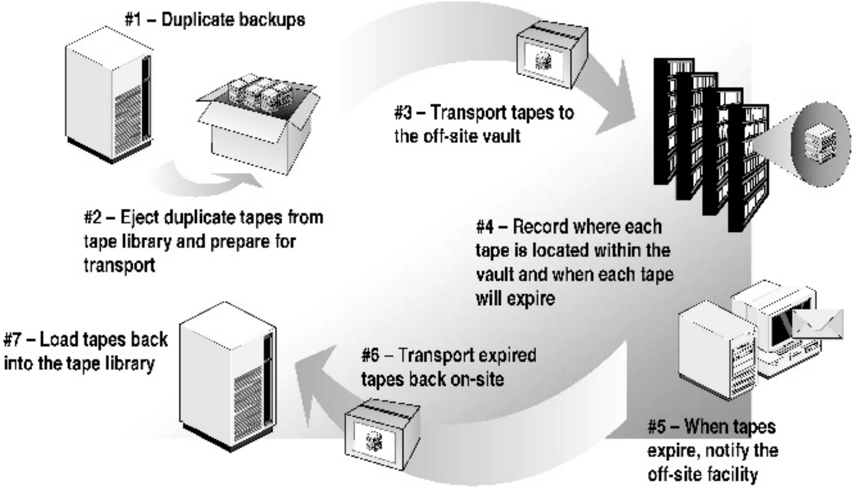
#1 – Duplicate backups
#3 – Transport tapes to the off-site vault
#2 – Eject duplicate tapes from tape library and prepare for transport
#4 – Record where each tape is located within the vault and when each tape will expire
#7 – Load tapes back into the tape library
#6 – Transport expired tapes back on-site
#5 – When tapes expire, notify the off-site facility

Figure 1.2 Basic vaulting flow

1.5.1 Offsiting Original Backup

If you select this method of selecting which medium is stord offsite in a secure storage facility you must be prepared to accept a potential delay in restore requests. Any request to restore data will require the original media be recalled from the storage facility. This obviously not only will affect the time to restore but also has the real possibility of causing the backup media to be handled more, which can reduce the life of the media. It also puts you at a greater risk of loosing media as it is being transferred more often.

1.5.2 Create Multiple Copies of the Backup

Some of the backup applications have the ability to create more than one copy of the backup data during the initial backup process. By doing this, you can have your vault process move one of the copies off-site. This removes the problem of always having to recall the off-site media to fulfil all restore requests. It also makes the off-site copy available as soon as the backup is completed.

1.5.3 Duplicate the Original Backup

This has been the more common method of creating the off-site copy of the backup. After the initial backup is complete, the vault process will create copies of any backups that need to have an off-site copy.

After the backups are duplicated, one of the copies is moved off-site. After you have the images on media that are ready to be taken off-site, the vaulting process should create a list that includes all the media IDs for all the media destined to be taken off-site or vaulted. A good vaulting application will actually perform the ejection of the media, so the operator or administrator can physically remove the media.

The vaulting process should be capable of creating reports that show what images need to be moved and the inventory of all media that are currently off-site. It should also create a report that can be shared with the off-site storage company that shows all the media that need to be returned on any given day. These are generally the media on which all the backup images have expired. These media will be recalled and reintroduced into the local backup environment, usually going back into an available media set.

A good vaulting application will also manage the backup and off-site storage of the data that makes up the backup application's internal catalogue. It will also track this information, which is very important if you need to recover the backup server itself.

The off-site storage companies have warehouses that are especially built for providing the highest possible protection against disasters – natural and otherwise. These companies offer services to physically transport the tapes to and from the warehouse. Some advanced vaulting applications provide reports and data formats that make it easy to integrate with the vault vendor's own data management systems. It is important to remember that backup is a sticky application. Users should carefully evaluate the potential off-site storage vendor for their staying power. Backup is also a critical application, so the user should look at what support the vendor is able to provide. You want to be comfortable that the backup vendor and the off-site storage company are going to be around for the long haul. Otherwise, all those backup images that the user has been saving for 7 years might be of little use.

1.6 ENCRYPTION

There is a rapidly growing requirement that all data that is moved off-site be encrypted. The data protection application vendors are hurriedly working on updating the existing encryption solutions to allow for more selective use. The entire subject of encryption is detailed in Chapter 6, but we can highlight some of the requirements and options that are currently available:

- Client-side encryption.
- Media server encryption.
- Encryption appliance.

1.6.1 Client side encryption

With client-side encryption, all of the data that is moved from the client is encrypted before being sent off the client. This involves using the client central processing unit (CPU) to actually perform the encryption. This can have a performance impact on the client, depending on how much of the CPU is available and therefore can have an impact on the backup performance.

1.6.2 Media server encryption

This method of encryption allows you to encrypt only backups that are being sent off-site or just those being created by the vault process. This still uses a CPU to perform the encryption, but now it is the media server CPU that is being used. The basic work of the media server is that of a data mover and generally there is not as high a demand on its CPU. You also have more control on when this is being done so you can pick a more idle time. The downside here is that the data is moving across the network from the client without being encrypted.

1.6.3 Encryption appliance

This method involves purchasing a specialized hardware appliance that is installed in the data stream. This appliance can encrypt data as it passes through. It removes the CPU load of the other two methods, but does require the purchase of the special hardware with its own software/firmware.

As we will see in Chapter 6, the process of encrypting the data is only a piece of the puzzle. Generally when you elect to encrypt data there are keys involved that must be managed. Without the proper keys the data becomes very secure. No one can read it, not even the owner. The key management is different for each of the options.

1.7 MANAGEMENT AND REPORTING

In the traditional backup and recovery data protection scheme, there is generally a silo approach to management with each group doing its own management and reporting. This duty usually falls on the administrators, not the people who actually need the information. This just becomes another task for administrators who have plenty of other responsibilities. In many cases, they do not actually know the SLAs that they are reporting on.

Reports are typically generated by scraping the application logs and presenting either the raw data or some basic compilation of the data being collected. The resulting reports often do not have enough details or the correct details to facilitate the type of management that is truly required to ensure that all the SLAs are being met. The fact that we often have the wrong people trying to manage the data protection scheme with inadequate reporting has made overall data protection too often not properly implemented and managed.

This is further compounded by the fact that reports concerning storage are generally done by the storage administrators, reports concerning systems by the system administrators and reports about the network by the network administrators. It is very difficult for any one person or group to know exactly how well the enterprise is being managed and protected with this widely diverse method of management and reporting.

1.7.1 Service Level Management

Increasingly, storage services, including backup and recovery, are offered to business unit 'customers' based on established service levels. The business units are then charged back based on their consumption of the resource, bringing a measure of accountability into IT resource

Table 1.1 Service levels

	Availability shelf life	Recovery time	Recovery point	Backup window	Underlying technologies
Platinum	Forever	5 min	Zero data loss	No impact 24 × 7	Data replication, snapshots and off-host backup to tape on replica data off-site
Gold	7 years	10 min for first 30 days, 1 h for next 11 months, 1 day after that	30 min	No impact 24 × 7	Rolling snapshots every 30 min with 24 h retention, 1 year worth of nearline tape capacity in library, on shelf for remainder
Silver	2 years	1 h for the first year, 1 day for the second year	1 h	No impact during production day, midnight to 6 a.m.	
Bronze	30 days	24 h	24 h	Backup can occur anytime	Daily backup to tape

consumption. Service levels can generally be established into a small number of narrowly defined offerings, based upon the metrics by which a business unit has recoverability. The metrics are not communicated in IT terms, such as server platform, tape or disk technology, SAN/NAS and so on, but rather in simple terms that quantify the expectations for data recovery. For example, one could establish a simple four-tier hierarchy, which offers platinum, gold, silver and bronze services. An example of service levels is shown in Table 1.1.

By establishing clear SLAs and monitoring delivery against these commitments, the operation can be properly funded by more responsible business unit owners. Also, the underlying technology infrastructure can be better managed and upgraded as needed to allow the storage group to deliver on its commitments to the business units.

1.8 SUMMARY

As we have seen, historically data protection has been accomplished by traditional backup and recovery with some mingling of HMS solutions. This was coupled with DR schemes that were also mostly based on the same backup and recovery techniques and included a vaulting process. The silo approach to reporting did little to assist in moving beyond this methodology. We are starting to see service levels also becoming a part of the management process.

In the following chapters, we will see the move that has already started to augment these traditional data protection techniques with the more tradition data management tools. In later chapters, we will follow some of the more advanced integration of these tools and techniques and then look beyond these to the totally new approaches being developed to meet the data protection needs of today and tomorrow.