1

INTRODUCTION TO DATA PROTECTION

In This Chapter:

- What Does Data Protection Mean?
- A Model for Information, Data, and Storage
- Why Is Data Protection Important to the Enterprise?
- Data Loss and Business Risk
- Connectivity: The Risk Multiplier
- Business Continuity: The Importance of Data Availability to Business Operations
- The Changing Face of Data Protection
- Key Points

The explosion of corporate data in the 1990s, coupled with new data storage technology such as networked storage, has made the accumulation and management of large amounts of data a corporate priority. Corporations *try* to accumulate terabytes of data on increasingly large storage systems. Gathering customer data, vendor information, minute financial measurements, product data, retail sell-through data, and manufacturing metrics are now corporate goals. Even small to medium-size businesses (SMB) have begun to acquire terabytes of data. Management of storage systems, and the data held within them, is a cause of great concern within IT departments, corporate legal offices, and the executive suite.

With the advent of new regulations and the understanding of how incredibly valuable corporate data is, there is a new focus on protecting and accessing data. As companies received hard-earned lessons on what can happen when data is

destroyed, damaged, or unavailable, more focus has been placed on protecting mission-critical information than on simply accumulating it.

Typically, IT departments have tried to protect data by using high availability (HA) devices with redundant systems, backing up data regularly to tape, and data duplication techniques. Increasingly, more sophisticated methods of ensuring the *integrity and availability* of important corporate data are being used, including remote mirroring and remote copy, near-line backup, Data Lifecycle Management (DLM), and Information Lifecycle Management (ILM).

WHAT DOES DATA PROTECTION MEAN?

Data protection is just what it sounds like: protecting important data from damage, alteration, or loss. Although that sounds simple enough, data protection encompasses a host of technology, business processes, and best practices. Different techniques must be used for different aspects of data protection. For example, securing storage infrastructure is necessary to ensure that data is not altered or maliciously destroyed. To protect against inadvertent data loss or permanent corruption, a solid backup strategy with accompanying technology is needed.

The size of an enterprise determines which practices, processes, or technologies are used for data protection. It is not reasonable to assume that a small business can deploy expensive, high-end solutions to protect important data. On the other hand, backing up data to tape or disk is certainly something that any enterprise can do. A large enterprise will have both the resources and the motivation to use more advanced technology.

The goal is the same no matter what the size or makeup of the company. Data protection strives to minimize business losses due to the lack of *verifiable* data integrity and availability.

The practices and techniques to consider when developing a data protection strategy are:

- *Backup and recovery:* the safeguarding of data by making offline copies of the data to be restored in the event of disaster or data corruption.
- *Remote data movement:* the real-time or near-real-time moving of data to a location outside the primary storage system or to another facility to protect against physical damage to systems and buildings. The two most common forms of this technique are remote copy and replication. These techniques duplicate data from one system to another, in a different location.
- *Storage system security:* applying best practices and security technology to the storage system to augment server and network security measures.
- *Data Lifecycle Management (DLM):* the automated movement of critical data to online and offline storage. Important aspects of DLM are placing data con-

sidered to be in a final state into read-only storage, where it cannot be changed, and moving data to different types of storage depending on its age.

- *Information Lifecycle Management (ILM):* a comprehensive strategy for valuing, cataloging, and protecting information assets. It is tied to regulatory compliance as well. ILM, while similar to DLM, operates on information, not raw data. Decisions are driven by the content of the information, requiring policies to take into account the context of the information.

All these methods should be deployed together to form a proper data protection strategy.

A MODEL FOR INFORMATION, DATA, AND STORAGE

Traditionally, storage infrastructure was viewed differently from the data and information that was placed on it. A new, unified model has emerged that ties together hardware, management, applications, data, and information. As Figure 1–1 shows, the entire spectrum from devices through information can be thought of as a series of layers, each building upon the others and providing more advanced services at each layer

The model begins with the traditional world of storage: the hardware. The hardware or *device layer* includes all the hardware components that comprise a

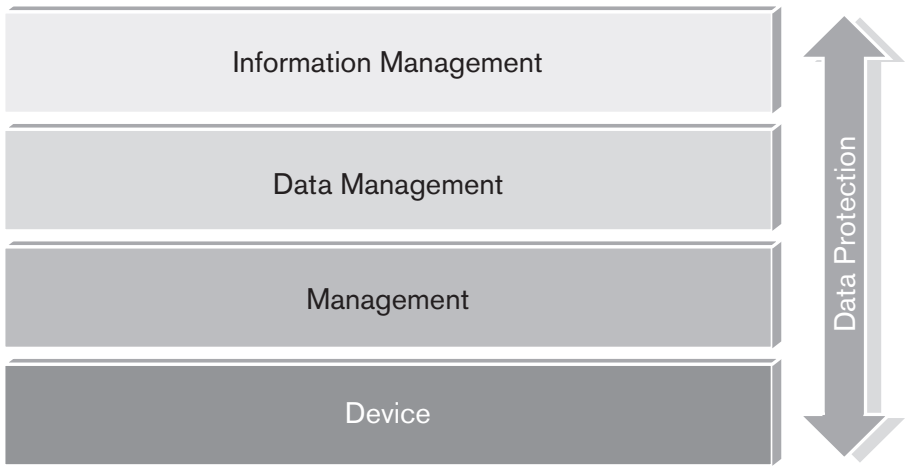

FIGURE 1–1 A MODEL FOR INFORMATION, DATA, AND STORAGE

storage system, including disks and tapes up to entire Storage Area Networks (SAN).

Next is the *management layer.* This layer is comprised of all the tools for managing the hardware resources. Some typical functions of this layer include device and network management, resource management, network analysis, and provisioning.

The *data management layer* consists of tools and techniques to manage data. Some typical functions within this layer are backup and recovery, remote copy, replication, and Data Lifecycle Management practices.

The final piece of the model, and the uppermost layer, is the *information management layer.* This layer addresses the difference between information and data: context. Business practices such as Information Lifecycle Management look at what a collection of data means and manages it accordingly.

Data protection cuts across all levels of the model. A successful data protection strategy will take into account the hardware, especially its security and configuration. The management layer is less pronounced in the data protection strategy, because it mainly serves the hardware. The data management layer is heavily involved, and the information management portion ties many aspects of data protection together while filling in significant gaps.

While reading the rest of this book, keep this model in mind. It will help provide a framework for thinking about data protection.

WHY IS DATA PROTECTION IMPORTANT TO THE ENTERPRISE?

There are several reasons for spending money, time, and effort on data protection. The primary one is minimizing financial loss, followed by compliance with regulatory requirements, maintaining high levels of productivity, and meeting customer expectations. As computers have become more and more integral to business operations, data requirements from regulators such as the U.S. Securities and Exchange Commission (SEC), as well as from customers, have been imposed on businesses. There is a clear expectation that important data be available 24 hours a day, 7 days a week, 365 days a year. Without a working data protection strategy, that isn't possible.

The single most important reason to implement data protection strategies is fear of financial loss. Data is recognized as an important corporate asset that needs to be safeguarded. Loss of information can lead to direct financial losses, such as lost sales, fines, or monetary judgments. It can also cause indirect losses from the effects of a drop in investor confidence or customers fleeing to competitors. Worse yet, stolen or altered data can result in financial effects that are not known to the company until much later. At that point, less can be done about it, magnifying the negative results.

Another important business driver for data protection is the recent spate of regulations. Governments throughout the world have begun imposing new regulations on electronic communications and stored data. Businesses face dire consequences for noncompliance. Some countries hold company executives criminally liable for failure to comply with laws regarding electronic communications and documents. These regulations often define what information must be retained, for how long, and under what conditions. Other laws are designed to ensure the privacy of the information contained in documents, files, and databases. Loss of critical communications can be construed as a violation of these regulations and may subject the corporation to fines and the managers to legal action.

A third driver, which does not get the attention of the press but is important to organizations nonetheless, is productivity. Loss of important data lowers overall productivity, as employees have to deal with time-consuming customer issues without the aid of computer databases. Data loss also results in application failures and similar system problems, making it difficult for people to do their jobs. A poor data protection strategy may leave people waiting for long periods of time for systems to be restored after a failure. During that time, employees may be idle or able to work only in a reduced capacity, further diminishing productivity.

The demands of a 21st-century business are such that customers expect the business to operate at all times. In an increasingly global economy, downtime is not tolerated by customers, who can readily take their business elsewhere. The inability of a business to operate because of a data loss, even a temporary one, is driving many businesses to deploy extensive data protection schemes. It is not only the e-commerce world that experiences this situation. All types of businesses—including health care, financial, manufacturing, and service—operate around the clock, or at least their computer systems do. Even when no humans are around, computers are available to take and place orders, send orders to the warehouse, and manage financial transactions. Data protection strategies need to take into account these 24/7 expectations.

DATA LOSS AND BUSINESS RISK

Risk is a measure of potential economic loss, lack of return on an investment or asset, or material injury. Another way to state this is that risk is a measure of exposure to harm. Some common risks are material loss (for example, damaged equipment, facilities, or products), risk to sales and revenue, lawsuits, project failure, and market risk. Risk is associated not only with hard assets, such as building or machinery, but also with revenue, customer loyalty, and investments in projects.

How risk is measured depends on the assets deemed to be at risk. In computer security circles, risk is usually a measure of threats (the capability and willingness for malicious behavior), vulnerability (the holes in the system that can be exploited), and harm (the damage that could be done by a threat exploiting a vul-

nerability). No matter how you measure risk, the most important component is harm. Without harm, there is no risk.

Insurance, locked cabinets, background checks, and currency hedges are ways that companies seek to minimize harm to their assets and the profitability of the business. If one thinks of information as being a corporate asset, protecting the underlying data is necessary to ensure the value of the asset and prevent its loss. Ultimately, data protection is about mitigating business risk by reducing the ability of some threat to do harm to mission-critical data.

The Effect of Lost Data on Business Operations

Companies recognize that data loss represents a business risk. Even if a monetary value is not assigned to the data, the negative effects on operations can be significant. In many cases, corporate operations can be so adversely affected that companies feel the need to mention the risk in regulatory filings and shareholder reports.

Three types of damage may occur because of data loss. First, data may be unrecoverable. In this case, important business records may be lost forever or available only in hard-copy form. Any business process that is dependent on that data will now be considerably hindered. This is the worst form of damage that can occur.

Next, data may be recoverable but may require considerable time to restore. This scenario—the most likely—assumes that data is backed up in some other place, separate from the primary source. This is a better situation than irrecoverable loss, but the data will be unavailable while recovery operations take place. In some cases, not all the data may be recovered. This is a common problem with data restored from nightly backups. Any data created during the day when the primary data was lost is not on the backup tapes and is lost forever.

Finally, while data is unavailable, either permanently or temporarily, applications not directly related to lost data may fail. This is especially true of relational databases that reference other databases. Loss of a central database of customer information, for example, may cause problems with the sales system because it references customer information. A loss of this type can result in cascade failures, in which several applications fail because of their dependence on another application's data.

RISK TO SALES A company may suffer measurable harm when data loss makes it impossible for it to interact with customers. The result is that the company will not realize sales and revenue.

E-mail has become a primary form of corporate communication. Losing an important e-mail or attachment may mean that a customer may not be serviced correctly; thus, sales are lost. This is especially true of companies that sell capital equipment to other companies. A hard drive crash on the e-mail server may cause an important bid to go undelivered. The salesperson may not even know that the

bid was not received by the customers (because it is sitting in the Sent folder stored on a local hard drive) until the sale is lost.

As large companies have become more dependent on call centers, they have become equally dependent on the customer relationship management (CRM) systems that help them track customer issues and orders. This represents a risk to sales, revenue, and profitability. If this risk is realized—if the worst-case scenario comes true—the harm done to the business may be severe enough to propel it into bankruptcy.

Even Mother Nature Fears Data Loss

In the quarter ending March 31, 2000, Mothernature.com, an Internet-based retailer of health and beauty products, saw fit to mention the following in its U.S. Securities and Exchange Commission (the U.S. regulatory body for public companies and markets) Form 10-Q filing:

"If our existing technical and operational systems fail, we could experience interruptions or delays in our service or *data loss,* and could be unable to accept and fulfill customer orders."[1]

In the paragraph that followed, the company outlined how the risk of data loss could make it impossible for it to meet customer expectations and fill orders. Clearly, inability to ship an order represents a major risk to a catalog or Internet reseller.

INABILITY TO OPERATE Extreme data loss such as loss of an entire database, even temporarily, has been known to cause organizations to fail. A company may not be able to fulfill orders, update employee records, produce financial reports, manufacture goods, or provide services. It may not even have an operating phone system. Computer technology and the data associated with it are integrated into all aspects of an organization's operations. Because of this dependence on information technology, there is a *clear risk* that data loss can make it impossible for an organization to perform properly.

Even partial data loss can disrupt business operations and produce negative effects. Employees may be idled for long periods of time while data is re-created or recovered, reducing productivity. Applications may fail unexpectedly when referencing data that is no longer available. Essential reporting may be incomplete because component data is not available.

Loss of data also makes it difficult for managers to measure company operations. Most modern businesses rely on financial, market, and manufacturing metrics. Without the ability to gather and report on key business indicators,

1. Mothernature.com Inc. Form 10-Q, filing date May 15, 2000.

managers are running blind as to the health of the business. Destroyed, damaged, or altered data skews metrics and disrupts decision-making. The overall effect of this type of disruption is reduced revenue and higher expenses, leading to loss of profitability.

LAWSUITS AND FINES There is potential for lawsuits and fines when a company experiences data loss. With shareholder lawsuits fairly common, failure to protect data could easily lead to litigation, especially if data loss can be tied to a negative change in the share price of the company's stock. A more likely scenario is that data loss will affect operations and sales, causing the business to underperform. This can then trigger shareholder suits.

Other types of legal action can result in adverse judgments for companies. Companies may be sued for failure to perform duties outlined in contracts or the inability to produce goods and services that have been paid for. A lost order record may result in a customer's suing for direct and collateral damages.

Regulators now have the power to impose data retention requirements on companies. *Data retention requirements* tell a company what data must be kept and for how long. Fines can be levied when these requirements are not met.

It is not enough simply to have good policies; the policies have to be followed up with good practices. In 1997, Prudential Insurance was fined heavily because it did not properly *implement* existing electronic document retention policies. This led to the destruction of electronic documents needed as evidence. There was no indication that employees willfully destroyed evidence—only that the company did not take *sufficient action* to ensure that it was preserved. Though Prudential had a good electronic document retention policy in place, its inability to implement it properly cost the company $1 million in fines.[2]

Damaging legal situations can occur when data loss causes financial information to be released late. Regulators, markets, and shareholders expect certain reporting to occur at previously announced intervals. When a company fails to meet these expectations, that failure often leads to fines, lawsuits, drops in price of the company's stock, or even delisting from financial markets.

All these situations represent financial harm to the business. As such, steps need to be taken to protect the business against the risk of lawsuits and fines.

THEFT OF INFORMATION Another type of harm that requires data protection is theft of corporate information. This may take the form of theft of secrets or a violation of private data. Theft of secrets happens when a thief is able to access internal company information vital to current and future operations. Some examples of this these secrets are product plans, product designs, and computer source code. The economic impact of theft of secrets is difficult to ascertain, because the harm is indirect and manifests itself over long periods of time.

2. *The National Law Journal,* November 3, 2003.

Theft of private information, such as customer information, may have three effects:

- Lawsuits may arise when it is known that this information has been stolen. Customers may sue for damages that result from the use of this confidential information.
- Regulators in some countries may be empowered to take criminal and civil action against a company that suffers such a breach. The European Union, for example, requires that "Member States shall provide that the controller must implement appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing."[3]

  Other political entities have similar laws that require the safeguarding of information from destruction or breach.
- Customers may refuse to do business with a company that allows such a theft of private information. It is reasonable to assume that a customer would not want to continue to do business with a company that has not taken adequate care to safeguard private information.

Reasons for Data Loss

As one might expect, there are many reasons why a corporation might lose important data. Broadly, they can be broken into the following categories:

- Disasters
- Security breaches
- Accidents or unintended user action
- System failure

Some data protection techniques can be applied to all these causes of data loss; others are better used for specific categories.

DISASTERS Disasters are the classic data-loss scenario. Floods, earthquakes, hurricanes, and terrorists can destroy computer systems (and the data housed on them) while destroying the facilities they are kept in. All disasters are unpredictable and may not behave as forecast. The goal of data protection is to create an

3. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995, Section VIII, Article 17.

environment that shields against all types of disasters. What makes this difficult is that it is hard to predict what type of disaster to guard against, and it is too costly to guard against all of them. Companies guard against the disasters most likely to occur, though that is not always good enough. Until just a few years ago, most U.S. companies did not take into account terrorism when planning for disasters.

There are two classes of disasters: natural and manmade. Natural disasters are often large in scope, affecting entire regions. Earthquakes and hurricanes, with their ability to do widespread damage to infrastructure, are especially worrisome; they rarely provide enough time to develop a plan for data protection if one is not already in place. After the disaster begins, it is too late to try to save data.

Manmade disasters are often more localized and generally create much less damage. Fires are the most common manmade disaster, although many other manmade incidents can cause data loss, too. The worse manmade disaster resulting in widespread loss of data (and life) was the September 11, 2001, terrorist attack on the World Trade Center in New York City. The destruction of key computer systems and the harm that it wrought to the economy of the United States led the U.S. Securities and Exchange Commission and the Comptroller of the Currency to jointly issue policies[4] requiring that data be adequately protected against regional disasters.

SECURITY BREACHES When an intruder breaches the network, server, or storage defenses of a company, he usually has one of three goals: to look at information he shouldn't look at, to deny the company the use of its data, or to damage and destroy data. Because the harm is intentional, an intruder can do more selective damage aimed at long-term harm.

Intruders come in two types: insiders and outsiders. The press tends to accentuate the problem of outsiders, yet insiders are as big a problem. Insiders can do more damage because they already have access to vital systems (and don't have to work as hard to get at important data) and know what type of damage can do the most harm. Insiders also have the advantage of less scrutiny. Most IT departments have sophisticated methods of detecting outsiders trying to break in. Fewer companies monitor activity inside their network. For this reason, insiders can go undetected until they do damage, whereas outsiders are often stopped at the network perimeter.

Security concerns affect data protection strategies in two ways. First, it is important to keep backups or copies of data, in case a security breach results in damage or destruction of critical data. Second, part of the data protection strategy needs to be securing vital data and information assets against harm. Although network and server security is well formed and understood by IT professionals,

4. SEC Policy Statement [Release No. 34-48545; File No. S7-17-03].

storage system security is much less mature, in terms of both technology and best practices.

ACCIDENTAL DATA LOSS Accidental loss represents one of the most common data loss scenarios. End-users are often the culprits; they delete, overwrite, and misplace critical files or e-mails, often without knowing they've done so.

In the 1980s and early 1990s, it was not at all unusual for the help desk to get frantic calls from end-users who had reformatted their hard drives. Fortunately, changes in desktop operating systems have made accidental reformatting of a hard drive much more difficult, and it is now a rare event. Damaged or reformatted floppy or Zip drives are still a common problem, though this usually destroys only archive data. As other forms of mobile media, such as solid state memory devices, are used by more people, the likelihood of loss of data on these devices grows. And yes, people sometimes drop their smart media cards in their coffee.

Though IT personnel may feel frustrated by the silly errors end-users make that result in data loss, they are responsible for quite a few errors themselves. Botched data migrations, hastily performed database reconfigurations, and accidentally deleted system files are everyday occurrences in the IT world. One of the most common and most damaging IT errors occurs when a backup tape is overwritten. Not only is the previous data destroyed, but there is no good way to recover much of it. Also, quite a few backups are damaged due to sloppy storage practices.

The risk that the end-user represents is usually a recoverable one. Although it's a hassle to dig out backups and pull off individual files, it is still something that can be done if the data in question is important enough. Good habits, such as backing up files to file servers or automated backups and volume shadow copying (now part of the Windows operating system), can alleviate many of the effects of end-user data loss.

IT mistakes represent much greater risk. The effects of an IT accident are not limited to individuals; instead, they affect entire applications and systems, many of which are mission critical. Strict policies and controls are necessary to prevent these types of errors.

SYSTEM FAILURE System failures often cause data loss. The most famous type of failure is a hard drive crash. Although hard drives don't fail with the frequency that they used to, failures are still a major problem for many system administrators. This is especially true of drives in high-use servers, in which drive failure is inevitable. Data can also be corrupted or destroyed because of spurious errors with disk array hardware, Fibre Channel and SCSI host bus adapters (HBAs), and network interface cards (NICs). Fluctuations in electricity, sudden power outages, and vibration and shock can damage disks and the data stored on them.

Failures in software are also a source of data loss. Updated drivers and firmware are notorious for having bugs that cause data to be erased or corrupted. The same can happen with new versions of application or database software. The failure of IT to properly back up and *verify the integrity* of a backup before installing new software is an age-old problem leading to irrecoverable data loss.

System failures cannot be completely prevented, but steps can be taken to reduce the likelihood of losing data when they occur. One of the most common steps is to buy high availability (HA) devices for mission-critical applications. HA units offer better protection against shock, flaky electricity, and link failures that can corrupt data. They also have software protection that ensure that I/O is complete and that bad blocks do not get written to disks. Good backup and archive procedures are also important parts of a plan to protect against system failure.

CONNECTIVITY: THE RISK MULTIPLIER

When networking was introduced, the risks associated with it were relatively low. Most networks were small, with only a handful of computers linked. The Internet started as a network of only four mainframes. Local-area networks (LANs) did not become widely deployed until the late 1980s. Access to these networks was very limited, and the number of assets involved was low.

As the networks grew, both in size and complexity, security problems became more prevalent, and the risk involved in using a network became higher. There were more devices of different types, with many more access points. Whereas in the past, disasters or hackers could be contained to one computer, networking allowed problems to spread throughout a large number of machines. There is now network access to more computers than at any time before. Many homes now have several linked computers and network devices, and have become susceptible to the same security and network problems that have plagued the corporate world for years.

Network Attached Storage and Storage Area Network technology have had a similar effect on storage. Data storage devices have traditionally been isolated behind a server. Secure the server, and you secure the storage as well. That is no longer the case, and storage devices are experiencing many of the same problems that other network devices do. Some people would argue that the ability to get unauthorized access to a Fibre Channel SAN is low. However, if a malicious hacker does get through system defenses, he or she now has a greater number of devices to wreak havoc on. *Connectivity increases risk* because it gives more access to more resources that can be damaged.

Because risk is outcome based, the outcome of a successful intrusion or data corruption in a networked storage environment can be much more devastating than with an equal number of isolated, directly connected storage devices.

Even when system security is not the issue, connectivity can magnify other problems. Previously, one server could access only a small number of storage

devices. If something went wrong, and the server caused data to become corrupted, it could do so to only a small amount of data held on its local resources. Servers can now potentially access hundreds or even thousands of storage devices and can corrupt data on a scale that was not possible before.

Networked storage also has increased the complexity of the storage system, which can introduce more problems. The complexity of the storage infrastructure has increased dramatically, with switches, hubs, cables, appliances, management software, and very complicated switch-based disk array controllers. The opportunity to introduce errors into the data stream and corrupt or destroy it is much higher with so many devices in the mix.

In the networked storage environment, there are many servers and many storage devices. More servers can damage or provide unauthorized access to data. Even a single server can affect many data storage devices. The potential harm is multiplied by the high degree of connectivity that a modern storage infrastructure allows for.

BUSINESS CONTINUITY: THE IMPORTANCE OF DATA AVAILABILITY TO BUSINESS OPERATIONS

Business continuity is the ability of a business to continue to operate in the face of disaster. All functional departments within a company are involved in business continuity. Facilities management needs to be able to provide alternative buildings for workers. Manufacturing needs to develop ways of shifting work to outsourcers, partners, or other factories to make up for lost capacity. Planning and execution of a business continuity plan is an executive-level function that takes into account all aspects of business operations.

Information technology plays a key role in maintaining operations when disaster strikes. For most modern companies to function properly, communications must be restored quickly. Phone systems and e-mail are especially important, because they are primary communications media and usually are brought online first. After that, different systems are restored, depending on the needs of the business.

Protecting data and the access to it is a primary component of business continuity strategies. Restoring systems whose data has been destroyed is useless. What is the point of restoring the financial system if all the financial data has disappeared? IT, like other departments, needs to ensure that the data entrusted to it survives. In many cases, it is less important that the hardware systems themselves survive, so long as critical data does. If the data is still intact, new hardware can be purchased, applications reloaded, and operations restored. It might be a slow process, and there will be financial ramifications, but at least the business will eventually return to normalcy. Without the data, that will never happen.

THE CHANGING FACE OF DATA PROTECTION

In the past, data protection meant tape backups. Some online protection could be obtained by using RAID (which is explained in Chapter 2) to keep data intact and available in the event of a hard drive failure. Most system administrators relied on copying data to tape and then moving some of those tapes offsite. This is still the most common form of data protection, but only part of a whole suite of techniques available for safeguarding data.

Remote Data Movement and Copy

It was natural to extend the paradigm of duplicating important data on another disk (RAID) to duplicating it to another storage system, perhaps located in a different place. In this process, called remote copy, exact copies of individual blocks of data are made to a remote system. This system might be right next door or hundreds of miles away. Remote copy allows a second storage system to act as a hot backup or to be placed out of harm's way and available for the disaster-recovery site to use. At present, remote-copy systems tend to be expensive. The telecommunications needed to support them present the IT manager with a high recurring expense. The costs involved with remote copy have tended to relegate its use to high-end applications and very large companies.

Disk-Based Backup

Typically, backups consist of copying data from a disk system to a magnetic tape. Tape is, unfortunately, slow to write to, lacks the capacity that modern disks have, can be difficult to manage, and is very slow to recover data from. Because the purpose of a backup, as opposed to an archive, is to produce a copy of the data that can be restored if the primary data source is lost, slow recovery is a problem.

Because of these limitations, disk-based backups are gaining in popularity. Originally positioned as a replacement to tape, this method is seen as being part of a more sophisticated backup strategy. With disk-based backups, similar software and techniques are used as with tape, except that the target is a disk system. This technique has the advantage of being very fast relative to tape, especially for recovery. The disadvantages are that disk drives generally are not removable, and the data cannot be sent off-site the way a tape can.

Networked Storage

The biggest changes in data protection come courtesy of networked storage. In the past, storage was closely tied to individual servers. Now storage is more distributed, with many clients or servers able to access many storage units. This has

been both positive and negative for data protection. On the one hand, networked storage makes certain techniques—such as remote copy, disk-based backup, and distributed data stores—much easier to implement and manage. The ability to share certain resources, such as tape libraries, allows for data protection schemes that do not disrupt operations.

However, the networked storage environment is much more complex to manage. There tend to be many more devices and paths to the data. Because one of the key advantages of networked storage is scalability, these systems tend to grow quickly. This growth can be difficult to manage, and the sheer number of devices in the storage system can be as daunting as other types of corporate networks.

Networked storage allows for multiple paths between the server or client and the data storage devices. Multiple paths work to enhance business continuity by making link failures less of a problem. There is less chance that a broken cable will cause applications and backups to fail. Overall, networked storage is more resilient. It produces an environment in which safeguarding data and recovering from failure are performed more quickly and efficiently.

Information Lifecycle Management

The future direction of data protection is in a recent concept called Information Lifecycle Management (ILM). ILM is less concerned about the underlying data than about the upper-level information. Information is data with context; that context is provided by *metadata,* or data about the data. ILM guides data protection by determining what type of protection should be applied to data, based on the value of the information it supports. It makes sense to spend a lot of money on remote copy for very valuable information. Other information may not be worth protecting at all. ILM helps determine which path to take in making those decisions.

KEY POINTS

- Data protection is the safeguarding of important data from destruction, alteration, or loss. It is achieved through a combination of technology, business processes, and best practices. Core components of a data protection strategy are backup and recovery, remote data movement, storage system security, and Information Lifecycle Management.
- Business drivers influencing data-loss strategies are fear of financial loss, the need to comply with regulations, attempts to maintain high productivity, and the need to meet ever-increasing customer demands.
- Risk is potential exposure to harm. An organization can be harmed when data is unrecoverable, applications fail as a result of data loss, or the time to

recover data is unacceptable. The risk to a business manifests itself in lost sales, the inability to operate properly for some time, lawsuits and fines, and the effects of theft of critical information.

- Increased connectivity creates a risk multiplier. The more resources that can be affected by an event, the greater the potential harm.
- Business continuity strategies strive to keep businesses operating in the event of disaster. Data protection is a key component of business continuity.
- Data protection is changing. New practices and technologies include disk-based backup, networked storage, ILM, and remote data movement. These practices and technologies are providing system architects more options for protecting data.