

Backup & Recovery It's All About Data Protection

By W. Preston

Publisher: O'Reilly Media, Inc.

Pub Date: 2007-01-03

ISBN: 9780596102463

Table Of Contents

It's All About Data Protection.....	3
Business Reasons for Data Protection	3
Technical Reasons for Data Protection	6
Backup and Archive	9
What Needs to Be Backed Up?	10
What Needs to Be Archived?	11
Examples of Backup and Archive	12
Can Open-Source Backup Do the Job?	13
Disaster Recovery	15
Everything Starts with the Business	16
Storage Security	20
Conclusion	23
Index.....	24

Copyright (©) 2004, 2005, 2006, 2007 O'Reilly Media.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472

The copyrights in individual elements of this work are owned by their respective publishers, authors or others, as the case may be, and the prior written permission of the copyright owner is required for reuse in any form or medium of any individual element.

O'Reilly books may be purchased for educational, business or sales promotional use. Online editions are also available for most titles (<http://safari.oreilly.com/>). For more information, contact our corporate/institutional sales department: (800)998.9938 or corporate@oreilly.com.

Nutshell Handbook, the Nutshell Handbook logo, and the O'Reilly logo are registered trademarks of O'Reilly Media, Inc. Microsoft, the .NET logo, Virtual C#, Visual Basic, Visual Studio, and Windows are registered trademarks or trademarks of Microsoft Corporation. Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and O'Reilly Media, Inc. was aware of a trademark claim, the designations have been printed in caps or initial caps. The association between the image of the African crowned crane and the topic of C# is a trademark of O'Reilly Media, Inc.

While reasonable care has been exercised in the preparation of this book, the publisher and the authors assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

It's All About Data Protection

The chapters that have preceded talk about nothing but backup and recovery, but it's important to give you a little context before I close out the book. While backup and recovery are extremely important, they are but one part of the data protection landscape. *Data protection* is defined as all activities involved in protecting data from the various things that could happen to it. Data protection activities include backup, recovery, archive, storage security, and disaster recovery, and every company should be aware of all of these activities regardless of its size. While smaller companies may be less subject to some aspects of data protection, such as long-term archiving or compliance issues, they should regularly evaluate each aspect of data protection to determine how much it applies to them—and what they're doing about it. This is why this book, which focuses on backup and recovery, ends with a chapter on data protection—to make sure that you know that backup and recovery is just the beginning of your job.

Perhaps your company's archive and storage security requirements are easy to meet, allowing you to fulfill the rest of your company's data protection requirements with an open-source backup system and a simple off-site tape rotation scheme. Perhaps you've got sophisticated archive or storage security needs, and you're considering implementing an open-source backup and recovery system to leave room in the budget for some of the commercial solutions that address those other elements of data protection. Finally, it's also possible that you hadn't considered archiving or storage security prior to reading this chapter. Whichever description best fits you, I hope this overview of data protection gives you some food for thought.

We live in the information age. Even the smallest “mom-and-pop” businesses rely on some type of computer systems to store the information they generate. Perhaps it's a list of customer phone numbers, a log of business transactions, or even details about a new product. Whatever the information is, the business would be damaged if it was lost, deleted, destroyed, misplaced, or stolen. Therefore, a complete data protection system protects against all these risks.

A data protection system doesn't just restore data that has been destroyed or damaged. It also helps retrieve data that has been moved or retrieves data that is being requested in a manner different from the way it was stored. It prevents data from falling into the wrong hands and ensures companies are in compliance with regulations affecting their industries. And, of course, a complete data protection system ensures that the data can be restored in times of disaster.

There are both business and technical reasons for data protection. The next two sections provide an overview of those reasons.

Business Reasons for Data Protection

Like many other IT functions, the objective of any data protection strategy is to mitigate risks, reduce costs, or improve service levels.

Mitigating Risk

The primary job of a data protection system is to mitigate risk. In the IT world, risk mitigation is generally synonymous with data availability, internal/external security, and regulatory compliance.

Data availability

Many businesses today require that users and business applications have access to critical information 24 hours a day, 7 days a week, 365 days a year (24×7×365). Businesses that cannot access critical information may be unable to perform one or more key functions such as taking new orders or processing existing claims. Along these same lines, partners of these businesses can experience problems taking and processing orders if they don't have access to this information. Planned and unplanned outages of even a single system can therefore have serious ramifications that affect the business at hand, as well as the businesses of partners and customers.

The consequences of not being able to access data when needed can be serious and can include lost revenue (see the section “*Reducing Costs*,” later in this chapter). Worse still, public news of these types of events can have far-reaching consequences on all parties involved, affecting brand names and reputations.

Internal/external security

Chances are good that the information that drives one business is coveted by another. In fact, to the surprise of many businesses, large and small companies often go to extreme measures, including engaging in corporate espionage, simple mischief, and electronic terrorism, to get access to key competitor customer lists, development plans, and intellectual property.

Identity theft, or identity fraud to some, is another source of concern for IT departments as they carry out data protection strategies. A customer's name or other identifying information (such as address, birth date, and identification numbers like U.S. Social Security numbers) is about all someone needs to fraudulently empty bank accounts or gain access to credit.

Theft of or access to strategic business information can also have a number of serious business consequences, including loss of competitive advantage, loss of good corporate image, government-imposed fines, and even the loss of a business. All of these have happened to one or more major businesses in the last few years:

- Sadly, more than one company ceased to exist on September 11, 2001 when it was discovered that their hot-site was in the second tower.
- The dreaded “adverse inference”¹ instructions were given in more than one major U.S. lawsuit, resulting in a judgment for the plaintiff and hundreds of millions of dollars in fines.
- The reputation of several major companies was irreversibly damaged when it was revealed that they had not maintained control of the personal information of their customers.

Regulatory compliance

Regulatory compliance adds another layer of IT risk. Businesses today must contend with an increasing number of government and nongovernment regulations. For example, organizations

¹ Adverse inference is when a judge instructs a jury that the absence of a given piece of proof suggests that the claims the plaintiff is making are correct, even if the actual evidence is not present. It is an extreme measure used only when the judge feels the actual evidence was destroyed or covered up.

that store the medical information of U.S. citizens are subject to the Health Insurance Portability and Accountability Act (HIPAA). Financial organizations doing business in the U.S. must address the Securities and Exchange Commission (SEC) 17a-4 rule requirements, and industries of all types are accountable to Sarbanes-Oxley (SOX) stipulations. Anyone doing business with residents of the European Union will be very familiar with the Data Protection Directive of 1998, which governs the control and access of personal information, such as “an identification number or...one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.” Each major country in the world either has or is developing similar regulations. The risks of noncompliance are serious and can include fines that are often in the hundreds of millions of U.S. dollars, prosecution of key corporate officers, or a loss of business such that the organization is forced to close its doors.

Reducing Costs

When data is not protected properly, businesses can rack up a lengthy list of *hard costs* (such as fines levied on an organization in an electronic discovery suit) and *soft costs* (such as missed business opportunities or damaged reputation). An effective data protection strategy is able to minimize these costs by ensuring that data is available to authorized users who need it, when they need it, and according to business objectives.

Downtime costs

If the information that a company uses to generate revenue is unavailable, revenue is lost. However, just how much revenue a company ultimately loses depends on a number of factors including the type of business, the type of data that is unavailable, and how long the data is unavailable. The monetary cost can range from hundreds to millions of U.S. dollars per hour of downtime.

But downtime can also spell missed opportunity, which, though more difficult to quantify, may have equally lasting consequences for revenue. For example, customers who are unable to interact with a company because key information (such as data needed to process orders) is unavailable may choose to take their business elsewhere either temporarily or permanently. A long-time customer may be more willing to excuse downtime than a first-time customer.

Downtime also translates into a variety of other costs, including wages (to wit, paying employees to do a job they can't perform when the data is unavailable) and even additional storage costs in the form of extra backup copies. Employees, in particular those who do not believe their information is being protected properly, may keep extra copies of critical information on disk, tape, and other storage media. Depending on the amount of extra copies made and the type of storage media used, the added cost can be substantial.

Electronic discovery

Electronic discovery is a term used to refer to the practice of requesting information that has been digitally stored. For example, in a lawsuit a company may be ordered to provide all emails to and from a given employee or containing a certain set of keywords. The SEC may order a brokerage firm to forward all emails containing the words “promise” and “guarantee.” The EU may request proof of compliance to the Data Protection Directive. If a company is not set up to retrieve information in this fashion (that is, if this type of information is not immediately available), the costs of satisfying an electronic discovery request can quickly become quite large. While a defendant can petition the court to declare such costs unreasonable, a given judge may or may not grant their petition. Plenty of precedents in various state and federal cases show such requests being denied.

Security breaches

A security breach is the result of a type of unauthorized access to company information. As with downtime, the associated costs of a security breach can be difficult to quantify and can vary greatly among companies. Laws governing security breaches vary by location. Some U.S. states require companies to publicly disclose security breaches that can be financially devastating; others do not. Depending on the industry, a company may or may not be subject to a fine due to a breach. Companies doing business with persons who live in the European Union are subject to how the member country met the “judicial remedies, liability, and sanctions” chapter of the Data Protection Directive.

Data classification

Not all information has the same value within an organization, and it shouldn't be treated as if it does. Doing so can significantly increase data protection costs. For example, depending on the industry, an archived copy of a three-year-old order may not have the same value as data pertaining to a new product or service. However, many organizations mistakenly apply the same level of data protection to both types of information and store both on costly primary (or high-end) disk.

The key here is to classify information based on the age, regulatory status, and business importance of the data today and over time and then match storage systems and data protection levels accordingly. Companies that follow this type of plan can significantly reduce the total cost of ownership (TCO) of their storage resources.

Improving Service Levels

What does data protection have to do with business service levels? More specifically, how can better data protection help companies improve these levels? It boils down to the age-old dilemma of accessibility versus security: the more available information is to various applications, the greater the potential business benefit. However, the more users who have access to information, the greater the likelihood that someone without proper authorization can gain access to that information.

Companies that “lock down” mission-critical information are missing out on clear opportunities to increase productivity throughout the enterprise. The challenge is finding a way to strike a balance between accessibility and security.

Technical Reasons for Data Protection

As previously discussed, data protection is all about keeping company information safe (from accidental or intentional deletion, corruption, and mishandling), available (to authorized users as well as business departments and outside partners), and compliant (with various industry guidelines and governmental regulations), and that doing this is often quite challenging. From a technical standpoint, the job is equally challenging. Disk drive failures, worn tape cartridges, lost or stolen media, and the inherent security risks of network storage all complicate the IT task.

Device Issues

Many of the technical reasons for data protection stem from the characteristics of the many different devices that store data. Every new device increases the chances of failure or attack.

Disk failures

Physical disks (or disk volumes) can fail for a variety of reasons. While the average mean time between failure supplied by the vendor of an expensive disk drive is quite high, other factors such as usage, handling, and environmental conditions can affect the reliability of disk drives. Disk failures can also result from outside factors, such as fires, natural disasters, and acts of physical terrorism.

Tape media wear, stolen/misplaced tapes

Data is written to tape in a sequential fashion rather than in a random fashion like disk; consequently, both the tape drive and tape media can experience significant wear and tear during frequent read and write actions. Tape media is also highly susceptible to environmental factors such as humidity and heat.

Because data is typically written to tape in an unencrypted plain-text format (such as tar2), unauthorized users can rather easily retrieve information from the media. Therefore, stolen or misplaced tapes can result in a significant exposure to a company's intellectual property and the personal information of its customers. This is why it's important to have good physical security of your backup media.

Networked storage risks

Implementing storage networks, whether storage area networks (SANs) or network-attached storage (NAS) environments, is a two-edged sword in terms of pros and cons. On the one hand, storage networks can significantly improve data availability and manageability. On the other hand, they can open new security risks.

Prior to the advent of networked storage, ATA or SCSI disk drives were directly attached to host servers in a local area network (LAN) environment. (This is now referred to as direct-attached storage, or DAS.) With DAS, the only way to compromise the data on these drives was to compromise the security of each individual host. Because servers were "insulated" from widespread hacking, companies were able to set up different security levels in the LAN depending on the datatype.

In a networked storage environment, the situation is different. It is possible, depending on the configuration and the security level of a given storage network, to access multiple hosts' data from a compromised host without physically hacking into each host. If one host is compromised and is able to "see" the other hosts' disks, the hacker can gain access to the data on those hosts without physically compromising those servers, too. Access to communication between these servers is enabled over network protocols, such as Unix-based NFS, Windows-based Common Internet File System (CIFS), Fibre Channel, or IP-based iSCSI.

A hacker, if so motivated, could also attack hosts, as well as stage other attacks, from the vantage point of the storage. The hacker could do things within the storage environment that would wreak havoc in the computing environment, such as denying all hosts access to shared data. Suddenly all your data is unavailable, and you have no idea why.

Email Can Be Critical

I was the email administrator for a medical software company, and I warned our system administrators that we needed to back up certain email stores more often and to put the email

stores on a RAID-protected disk. Both of them blew me off. About a month after I became the email admin, the CEO, CIO, CFO, and “owner” email stores were corrupted due to disk issues (the physical disk was failing). It took some real creative work on all of our parts to get the data back so we didn’t lose the docs on a multimillion dollar deal. (We were a 50-person company, and this was a huge deal for us.) After that day, I put in a mirror drive and hosted the CXO’s and owner’s email stores. The backup admin started reading the daily reports on what successfully backed up and what didn’t. (He had been getting a daily report that was being sent automatically to his trash.)

Scott Boss

External Threats

In addition to risks introduced by the types of devices that are in use, there are threats introduced by the people who use or have access to the devices. These include viruses, worms, Trojan horses, and, of course, accidental or intentional deletion.

Viruses

Microsoft’s definition of a virus is a good place to start. It defines a virus as “a piece of computer code that attaches itself to a program or file so it can spread from computer to computer, infecting [software, hardware, and files] as it travels.”

Worms

A worm, like a virus, is designed to copy itself from one computer to another, but, unlike viruses, it does so automatically by taking control of features on the computer that can transport files or information. Once you have a worm in your system, it travels alone; it does not have to attach itself to a program or file to wreak havoc with your system or others.

Worms are also dangerous because they can replicate in great volume through email address books. The result can be devastating, causing heavy network traffic that can bring business networks to their knees and slow down Internet traffic considerably.

Trojan horses

As for Trojan horses, just as the mythological Trojan horse tricked the city of Troy into believing it was receiving a gift, today’s Trojan horses, which often come in the form of email attachments, trick users into believing they are receiving security updates or other important information. In reality, they are recipients of hidden viruses that attempt to disable antivirus and firewall software.

Whether the threat is a virus, worm, or Trojan horse, it is extremely important for companies to protect themselves against these types of attacks. Once they’re in, they can be extremely difficult to purge from your environment.

Accidental deletion

Whether companies admit it or not, data loss caused by human error is a common occurrence. It takes the form of accidental deletion of a single file, an entire filesystem, or a user from a network

configuration. The only way to recover from this type of internal threat is through a historical copy such as a backup or snapshot.

Intentional deletion

Some data is deleted because a malevolent person decides it should be deleted. There was a major incident a few years ago when a malicious employee deleted every file on every server and desktop at a major financial institution—all because he didn't get the raise he wanted. A good data protection system might notice that this has happened or even prevent it from happening. At a minimum, a good data protection system should be able to restore the data in question.

Backup and Archive

The first two elements of data protection are backup and archive. These are related but very different activities. Backup is copying data from one place to another in case the original is damaged. Archive is copying or moving data to long-term storage for quick retrieval of logical components for a specific business purpose. The following comparison between backup and archive clarifies the important distinction between these two activities:

Backups are the secondary copy of primary data.

The purpose of backups is to recover primary data if it's damaged, deleted, or corrupted. Therefore, a backup is a secondary copy of primary data.

Archives are the primary copy of secondary data.

In contrast, an archive is created to be the primary copy of data that is important, but not important enough to be placed on primary storage. This includes old files that are being moved to secondary storage to save space as well as second copies of primary data that are created for a secondary purpose, such as content searching.

Backups recover data that was damaged, deleted, or corrupted.

This is the only purpose for backups. If a file is deleted or damaged, you can restore it from backup copies. Note that in order to perform a restore, you usually need to know the server and filesystem the file resided in and possess the application that created it.

Archives retrieve data from secondary storage.

While the words restore, recover, and retrieve can be used as synonyms in certain contexts, they have different meanings in the context of backup and archive. Restore and recover refer to putting something back in its original condition whereas retrieval refers to obtaining it from an alternate storage location. This is why we *restore* or *recover* from backups, but we *retrieve* from archives, as in "I retrieved the folder from the file drawer."

Archives retrieve data in a manner other than that in which it was stored.

Many applications create numerous completely disconnected pieces of data that are stored in files, emails, and databases. Very few applications are able to search across information stored in multiple applications, filetypes, and files stored over different points in time, but modern archiving applications are meant to do just that. A dynamic archiving system, such as an email system, can monitor and archive all incoming and outgoing email and allow you to search for content across multiple emails, email servers, and points in time. Some filesystem products can do the same for files. In other words, archives can retrieve data in a manner other than which it was stored.

Backups are stored only long enough to cover the usage pattern of the data.

Backups need to be stored only long enough to be able to recover deleted files. How long they need to be stored is based on the usage pattern of the data. If some files are only accessed once a quarter, you should be keeping backups longer than a quarter. If a file is only accessed once a quarter, and you keep backups for only one month, you cannot restore a file that was deleted three months ago and not accessed until yesterday.

Archives can be stored for many years or decades.

Sometimes archives contain information that's been deleted from primary storage and archived in case the information is ever needed again, such as the design plans for a product a company no longer makes. Someone needs to decide when that information can be deleted from the archives. Sometimes archives contain information governed by regulations, and those regulations dictate when that information can be deleted. The result is that archives can be stored for years or even decades.

Document Your Changes

At a large oil company, prior to a company holiday, a backup operator decided to change the next full to an incremental because there would not be a large rate of data change. The admin promptly forgot about the change. Three months later when a restore was needed, it was discovered that they needed to pull from three months of incrementals to do the restore.

David Bregman

What Needs to Be Backed Up?

Just about everything in a corporate environment needs to be backed up. The more important question to ask is what needs to be restored and how quickly does it need to be restored. As discussed in the disaster recovery section, you must define recovery requirements for every piece of data in your company. The primary reason for performing backups is to provide continuous access to a corporation's data in the event the primary copy of data is unavailable. Companies need to back up three types of data:

Intellectual property

This is the information about a company's core competency. In the case of a biotech firm, it is access to data captured in a discovery process; for a market research firm, it is access to database records.

Customer data

Examples range from scanned copies of patient x-rays to market research information and records about the buying patterns of particular market segments. It can also include information that can be used to conduct identity theft, such as a customer's address, birth date, or identification number.

Operational data

This last category includes every other kind of data in the organization. It can include data about where organizations purchase supplies to build products to information about who is

responsible for the delivery of products to customers. It includes payroll and accounting information and any other type of information that isn't intellectual property or the personal information of customers.

What Needs to Be Archived?

Beyond backing up data, organizations must also develop a strategy for archiving data. Both are integral components of an effective data protection strategy and necessitate a clear understanding of the business value of data (archiving maybe even more so than backup).

When executed correctly, archiving not only can save organizations money but it can also be a lifesaver, especially for those requiring access to historical information for regulatory compliance or audit purposes. Conversely, when archiving is performed incorrectly, it can cost a company dearly in terms of lost revenue, fines, and other penalties.

The problem is that many organizations mistakenly think of backup as archiving, and vice versa. The confusion regarding archiving often comes from some backup vendors that claim that their products also have archiving capabilities. Frequently, these capabilities equate to nothing more than backing up a data set and then deleting that dataset from primary storage. This is not archiving.

Vendor products offer different levels of “archiving” capabilities. At one end of the spectrum, some vendors treat archiving as simply a backup followed by a deletion of the data from primary storage. This type of “archiving” is really intended to assist organizations in removing old data that is cluttering up servers—a problem that is better addressed with storage resource management (SRM) or hierarchical storage management (HSM) tools.

So, what is archiving and how does it fit into the data protection landscape? Archiving is the long-term storage of information for the retrieval of logical components for a specific business purpose. In comparison, backups are intended to protect against short-term data loss, such as accidental deletion, device failure, and data corruption.

Archival data candidates include periodic corporate financial information that needs to be retained for auditing purposes, medical patient information that must be retained for HIPAA compliance purposes, and clinical trial data for a new drug that is winding through the Food and Drug Administration drug approval process. Other examples include email, check images, and other types of electronic communication that could be requested in an audit.

The long-term nature of archived data presents a number of new problems:

Backward compatibility

Because tape and optical drives typically can't read media that is more than a generation or two old, organizations must give some thought to the long-term recovery of data that is archived to tape. Data can be migrated to new tape or optical platforms, but migration can present validation and authentication issues in some regulated industries.

Media longevity

If data is to be maintained for a long time on tape or optical media, steps must be taken to ensure media integrity. This includes maintaining proper environmental control and refreshing volumes as needed.

Readability/usability of the data after a restore

The archived data must be “portable.” The archived data cannot depend on an obsolete version of an application or operating platform in order to be restored.

For both archiving and backup, it is critical to develop an understanding of the corporate value of the data to be protected. Deciding what data should be archived, when it should be archived, and how long it should be stored is central to the storage management process. A system of data classification like this can lead to intelligent policy management of both primary (disk) and secondary (backup and archive) storage resources.

Examples of Backup and Archive

The following two examples illustrate the differences between information that is archived and information that is backed up. Keeping backup copies for several years does not inherently make them archive copies. They are simply backups that have been kept for an extended period of time.

WingsRUs, a fictitious aircraft manufacturer, has detailed information about its latest plane, the WingsRUs 563. The data includes critical marketing information, detailed design specifications, invoices for materials, testing plans and results, FAA inspection and approval information, and customer information. The data is stored on different databases on different servers at its various locations. WingsRUs backs up this data regularly using traditional backup applications.

When the company begins manufacturing the next new plane, the WingsRUs 565, it determines it no longer needs to back up data regarding the 563; instead it needs to make room for the new 565 data, which is now the more valuable data to the company.

In the event that there are issues with the plane, the company decides to retain much of the 563 data in case the FAA should ask for detailed design specifications, testing plans, or other related material in the future. WingsRUs migrates its 563 backup data to a reference archive system, which provides appropriate access to this information when needed at a price point that is cost-effective. If the data is properly archived, the company should be able to retrieve the old plans by simply asking for the 563 without needing to know (or have access to) the hostnames, filesystems, or applications that stored the data in the first place.

A U.S. financial trading firm communicates with its customers primarily via email. After receiving a series of complaints that this firm is reportedly “promising” a specific rate of return on certain investments, the Securities and Exchange Commission begins an investigation. The investigation begins with an *electronic discovery request* to see all email written in the last two years that contain the words “promise” and “guarantee.” A discovery request is a legal term for a request for background information on a particular subject. An electronic discovery request is a discovery request that requires you to obtain the information from your computer systems.

Without an email archiving system, this type of request is difficult or even impossible to meet. If the firm had performed daily backups of its email system for the last two years, it would need to restore 730 versions (365 backup copies×2 years) of its Exchange Server and then search each version for the requested words—at best a daunting task. However, if the emails had been archived to an email archiving system, the company could easily search all emails sent in the last two years for the words “promise” and “guarantee,” and they would be immediately presented with a copy of all such messages.

Can Open-Source Backup Do the Job?

Electronic information is stored in a number of different ways, some of which require special treatment during the backup process. Ignoring these differences can have a number of negative side effects, including:

- A significant decrease in backup performance
- An even larger decrease in restore performance
- An inability to recover the data or system in question

How information is treated depends on how the information is stored. The standard, or most common, way that information is stored is as a file in a filesystem. The most common example of a filesystem is the “C:\” drive on Windows desktops and the “/” on Unix-based systems. Most backup products handle ordinary static files without issue, but some filesystems and datatypes can cause problems and often need to be treated specially. They include:

- Very active filesystems
- Filesystems with large (more than 1 TB) volumes of data
- Filesystems with millions files
- Information stored inside databases
- Metadata not stored in a filesystem or database
- Link file structures and device files
- Information stored on NAS-based filesystems
- Information stored on SAN-based filesystems

Most commercial backup products have additional features to handle these datatypes, usually at an additional price. This section considers whether the open-source products covered in this book can handle the same challenges.

Very Active Filesystems

The basic assumption of traditional backup software is that a file is not changing while it is being backed up. In the past, IT managers put systems into single-user mode prior to performing the backup to ensure that files were static, or unchanging, during the backup process. IT managers often do not have the luxury of doing this today, so backup application vendors have developed techniques for backing up these types of files.

Constantly changing files present a special challenge to backup and recovery software applications—even commercial products. In addition, some operating systems and applications can lock files for exclusive use, preventing even backup applications from accessing them. If a file is too active or locked during backup, commercial backup systems use snapshot techniques to ensure that the files are protected. Snapshot technologies present a static view of the filesystem to the backup application. If this particular challenge is present in your Windows environment, you may want to investigate how your open-source product integrates with Windows snapshot services. If this challenge is present in your Unix or Macintosh environment, you may find that you’re not able to solve it without moving to a commercial product.

Very Large Filesystems

Very large filesystems present another set of unique backup challenges. Because traditional backup software applications are filesystem-based, each filesystem or drive is backed up separately. While

this method works fine with small to moderate-sized filesystems that are dozens or hundreds of gigabytes in size, it does not perform well with filesystems that are greater than 1 TB in size.

The problem is that the speed of the fastest tape drives can push data at a rate of about 200 MB/s (at this writing). If you could supply a stream of data fast enough, you could back up a 1 TB filesystem in approximately two hours. The problem is that you probably couldn't keep up with the 200 MB/s tape drive, and it would take significantly longer than that.

If this particular challenge is present in your world, you may consider near-continuous data protection. A near-continuous data protection backup system uses replication techniques to maintain a copy of the data for backup purposes. Since it never has to do a full backup, it needs a lot fewer resources to run successfully. Since it's disk-based, it's also going to be able to back up and recover as fast (or as slow) as the filesystem you're trying to back up. See *Chapter 7* for a discussion of three open-source near-CDP products.

Filesystems with Too Many Files

Filesystems with lots of files also present a set of unique challenges to IT managers. In fact, a 1 GB filesystem with five million files is actually as challenging to back up as a 1 TB system with a few thousand files. Why? Because of the number of operations that must be performed during the backup and restore process.

The problem is even worse on the recovery side, which requires even more steps to complete. For example, the backup application must first tell the operating system that it needs to create a file; it then needs to open that file for writing and transfer the data into that file. After the restore, a series of checks is performed to verify that the data written to the filesystem is the same as the data that was backed up.

Again, each of these operations takes time, and because they are performed in sequence, they can actually bring the restore process to a screeching halt. The speed of the backup device is irrelevant. The fastest disk drive or tape drive in the world is still going to have to sit and wait while each of these operations is performed for each file.

An alternative approach to traditional backup and restore processes is *image-based backup*, which bypasses the filesystem (and files) and backs up data at the block level. Unfortunately, any open-source project designed to address this challenge probably requires the drive to be unmounted during backup. If this limitation is unacceptable, you have to switch to a commercial product.

Information Stored in Databases

Information stored in databases can be difficult to back up because of the way it is stored, the changing nature of the datafiles, and demanding recovery point objectives, or RPOs.

Databases generally store databases in files in the filesystem, but some databases store “raw” data, or datafiles, directly on disk. While storing data on raw disk can improve performance, it can make the backup environment significantly more complex.

Datafiles change constantly; therefore, the challenge is to create a consistent image of the datafile during the backup process. A variety of techniques, including cold backups, scripted hot backups, and database backup agents, can help IT managers create these images.

A *cold backup* is the backup of datafiles after a database has been shut down. A *scripted hot backup* places the database in some type of backup mode before backing up its datafiles using a

regular backup program. Either method works well with most of the backup utilities covered in this book. A *database backup agent* interfaces directly with the database for backup purposes. At this writing, none of the open-source backup products support backing up any database using its agent. However, many of the database agents do support backing up to disk without a commercial backup product. For example, Oracle's `rman`, Sybase and SQL Server's `dump` database, and `ntbackup`'s Exchange plug-in all support backing up to disk while the database is active. You could therefore create a disk-based backup that is then backed up by the open-source backup system you chose.

Databases generally have more demanding RPOs. While it may be acceptable to restore a word processing file to last night's backup, it is generally not acceptable to restore a database file from a backup copy that is several hours old. Databases provide *transaction logs* that track changes in between backups. A proper backup of a production database includes a system for backing up the transaction log during the day.

Information Stored on Shared Storage

Information stored on shared storage can also create extra backup requirements. Shared storage comes in two main flavors: SAN and NAS. SANs are based on the SCSI protocol and allow several systems to have block access to shared disk or tape drives. SANs typically run either SCSI over Fibre Channel or IP (iSCSI). NAS is based on NFS or CIFS protocols, which allow multiple servers to share files across an IP network.

SAN-based filesystems

The low-cost backup products covered in this book are not going to treat SAN-based filesystems any differently than a locally attached filesystem. If you need a product that performs SAN-based backups, you need a commercial product.

NAS-based filesystems

Although the snapshot and off-site replication software offered by some NAS vendors has great recovery features, NAS filers must still be backed up at some point. All of the open-source products discussed in this book are going to back up NAS-based filesystems via a share (NFS or CIFS).

Disaster Recovery

Devising a good disaster recovery (DR) plan is a bit like how many cities paint their bridges. The painters start at one end of the bridge and paint until they reach the end. Then they go to the other side of the bridge and start painting it until they reach the other end. Then they start all over again; they basically never stop painting the bridge.

So should it be when you're building your DR plan. You have to build it from the ground up, and it can take months or even years to perfect, at which point you have to go back to the beginning and start all over. Since computer environments are changing constantly, you continually have to change and test your plan to make sure it still works.

This section is not meant to be a comprehensive guide to disaster recovery planning. There are books dedicated to just that topic, and before you attempt to design your own disaster recovery plan, I strongly advise you to research this topic further. This part of the chapter gives you an

overview of the steps necessary to complete such a plan as well as a few details that are typically left out of other books.

Everything Starts with the Business

The design of any disaster recovery system should be driven by the ability to make available to the business the critical systems and information systems required to conduct normal production activities, without making those systems and information available to the wrong people. We are not protecting this data because it is a school project or an interesting hobby. We are protecting the data because if the data is lost, the ability to conduct business operations is at risk. Thus, it all starts with the business.

Define the Core Competency of the Organization

When looking at data protection, the first question to ask is, “What are the core products and/or services that this organization offers?” followed by “What is the information required to provide that product or service, and what applications are required to effectively use the information?” The answer to the second question is what defines your organization’s intellectual property (IP). Without these information systems, the organization would not be able to function.

Many types of information qualify as IP. For example, it could be your version of KFC’s 11 herbs and spices—the “secret recipe” that makes your company’s product or service different from everyone else’s product or service. Of course, without customers that secret sauce is not worth much. All of your customers’ locations, names, and contact information are also part of your IP, as are the names of potential customers. Any plans that your company has for doing something different, reaching a different market, or selling to a new group of people are also part of your IP. If it is information that you do not want in the hands of your competitor, it is part of your IP. This broad definition can include many types of information.

Intellectual property is a wonderful thing, but it is not the only important information in your company. Circling around the creation and delivery of your product or service are a number of other systems, such as procurement, payroll, accounts receivable and accounts payable, sales, and customer support. Each system is also critical to your business, and each needs a particular set of information to perform its function.

Prioritize the Business Functions Necessary to Continue the Core Competency

Once you identify all intellectual property and supporting applications and systems, you must prioritize the business functions necessary to continue providing your company’s core products or services. This phase is not just about importance; you must also consider urgency or criticality as well.

In order to establish what IP needs to be protected, you must understand the organization’s core competencies. Next, you need to prioritize the protection and recovery of these systems should they become unavailable. If the core competency relies on the manufacturing of a product, the systems to continue the process are vital to the continuation of the business. Other systems, such as email, may not be vital to the core competency and do not require the same level of protection. However, systems supporting customer communications may be critical and thus perhaps these email systems should be treated as critical applications as well. It is important to understand what

is vital and critical to the organizations supporting the business's core competency and not just protect whatever data happens to reside on your servers.

Correlate Each System to a Business Function, and Prioritize

Let us consider a power company as an example. If it did not deposit customer payments for a few weeks, some people might notice and many would not care, but the company's creditors would notice and care. Some overly conscientious customers might notice that their checks had not cleared yet, but it would not bother most of them. The company's creditors would only notice if it failed to make payment on a payable account. Even then, the company could probably explain to its creditors that it is in the middle of some sort of emergency, and the creditor would probably hold off the firing squad. However, what would happen if it stopped delivering electricity or gas for just a few minutes? It would be on the evening news, all its business customers would be angry at the impact to them, all the residential customers would have to reprogram their DVD players and microwaves, and the company could potentially cause a rolling blackout, similar to what happened in the U.S. Northeast in the early 2000s. (This happens in some parts of the world on a regular basis.) This means that the company's ability to deliver power is the most critical business function it has—its core competency.

Once you figure out what your IP and supporting systems are, and which ones are critical, you need to figure out where they reside and all of the resources required to use them. Is the information stored in a database? Is it stored in files on a filesystem somewhere? In most cases, the data is going to be stored on some type of computer system. Every computer and storage system must be assigned to a business function based on that business unit's level of criticality, thereby giving that system the same recovery priority as the business function to which it belongs.

A great example of this type of prioritization can be found in a publication of the U.S. Federal Communications Commission. It shows the FCC's different types of data and its criticality, and it is published at <http://www.fcc.gov/webinventory/>. Interestingly enough, its most critical systems are those required by law or presidential decree. It lists mission critical as the next level of criticality, followed by frequently requested data, and other data. For reference purposes, most companies use the term "mission critical" to describe their most important systems. In this case, the FCC has acknowledged that without governmental decree, it would have no mission. Therefore, it has another level higher than mission critical. The important thing to learn here is that each industry and company is different, and you must perform this prioritization of business functions specifically for your organization.

Define RPO and RTO for Each Critical System

Your recovery time objective, or RTO, is how quickly you want the system to be recovered. RTOs can range from zero seconds to many days, or even weeks. Each application serves a business function, so the question is how long you can live without that function. If the answer is that you cannot live without it for one second, then you have an RTO of zero seconds. If the answer is that you can live without it for two weeks, you have an RTO of two weeks.

The recovery point objective, or RPO, defines the point in time that is reflected once you have recovered a system, also referred to as how much data you can afford to lose. Consider two examples: customer orders and system logs. If you lose one customer order, the company loses significant revenue. Therefore, many companies determine that they cannot lose any customer orders. That means they have an RPO of zero seconds for customer orders. On the other hand, system logs might be useful only when troubleshooting problems or when auditing systems. If you

lose several days of them, it is a problem only if you need to troubleshoot a problem or audit the logs from that time period. However, if you acknowledge that the time period is lost anyway (due to a disaster), it is more important to just get the order system running immediately; the logging system is not as time-sensitive as the order system. Therefore, you can lose one week of system logs without really losing anything critical to the business. That means there is a one-week RPO for system logs.

Once you have established a priority for each system and determined the various outages that you are going to protect against, you must create an RTO and RPO for each system that is to be protected. Most of your customers really do not care what causes an outage or delay, so the RTO and RPO should be the same in all but the most extreme events, like a catastrophic earthquake affecting an entire region. Depending on their level of criticality, most systems have the same RTO and RPO for each disaster type. For some systems, however, you may find that a longer RPO and RTO is acceptable or unavoidable for major disasters.

Create Consistency Groups

It is often necessary to recover several systems to the same point in time. This is primarily caused by applications that pass data to one another. Consider a manufacturing company with the business processes of sales and custom manufacturing. There are possibly several different computer systems involved in this process, including the customer, orders, procurement, and manufacturing databases. If this business has a customer expecting a product, hopefully all four systems know that. What would happen, for example, if it was manufacturing a custom product and lost the original order, or it had the order, but didn't know which customer it went to? What would happen if it took the order, but the manufacturing database became corrupted, and the company did not know it was supposed to be making a product? This would represent a serious integrity problem.

Therefore, if your company has several systems that perform related business processes, those systems need to be in the same consistency group. In addition to determining an RTO and RPO, you must identify those systems that are related to each other because they need to be recovered to the same point in time. It is also important to identify a consistency window, or a window of time during which not all affected systems are changing.

If your consistency window is larger than (or the same as) your backup window, it's relatively easy to meet the consistency requirement for a consistency group. For example, a 5 p.m. to 8 a.m. window is a 15-hour consistency window. If all systems are down between 5:00 p.m. and 8:00 a.m. (no new data is being created), and backups start and finish sometime between 5 p.m. and 8 a.m., it does not matter if System A is backed up at 10:00 p.m. and System B is backed up at 2:00 a.m. They will still be in sync.

However, if your consistency window is too short to back up all systems in a consistency group, you need to do one of two things. One option is to create a custom backup window for those systems and ensure that they back up within that window. This option is certainly preferable because it does not significantly complicate your backup system. If your consistency window is too short for this approach, the second option is to augment your backup system with snapshots or business continuance volumes (BCVs). They allow you to quickly create a "virtual backup" on disk of several systems within a few seconds and then later convert the virtual backup to a physical backup (that is, back up the snapshot to tape or virtual tape).

Determine for Each Critical System What to Protect from

Once the business functions are prioritized, and each system is assigned to a business function, it is time to identify the things that can happen that trigger a recovery scenario. “Disasters” come in many forms. First, create a list of the different levels and types of disasters that are likely for your area and type of business.



The Disaster Recovery Institute states that each company should define its own levels of disasters. I’ve listed the way I define them, which starts with a loss of a single system.

Level 1 disasters are those that take out an entire application or server:

- Disk or disk array outage
- Internal corporate sabotage
- Electronic terrorism (denial-of-service attack)
- Disgruntled employee attack

Level 2 disasters are those that take out an entire data center:

- Building fire or flood
- Natural disasters (hurricane, tornado, earthquake)
- Building condemnation (chlorine gas leak)
- Physical terrorism (bomb)
- *Really* disgruntled employee
- Loss of all network connectivity
- Loss of all electrical power

Level 3 disasters take out an entire campus, city, or metropolitan area:

- Large-scale natural disasters
- Physical terrorism (bombing of power plant)
- Act of war

Determine the Costs of an Outage

Once you have determined all of the different types of disasters and their associated probability, you must assign a cost to each type of disaster, for each type of system. For example, if a fire took out your test server for a week, your cost may be nothing. However, if a fire took out a server that you deemed in the previous exercise to be mission critical, a loss of only a few minutes may cost you millions, depending on the level of criticality and the business you are in.

Such costs can come from a number of areas, starting with the loss of business. While your systems are down, you are not taking orders, making your product, or delivering your service. Another cost is a loss of reputation, which can result in the loss of future business. No company wants to be on CNN because it lost its data. Labor costs must also be added to the equation, and there are two kinds of labor costs. The first labor cost is when data was created and then lost; work has to be redone. The second type of labor cost is the opportunity cost of labor for workers who are not doing anything useful because the system is down. Depending on the type of business you are in, there may also be the loss of source materials used to create your product. For example, if you are a food maker of some sort, and the automation control systems are down, your ingredients may expire before the product is completed.

Another concept to think about when calculating the cost of an outage is that outage costs are logarithmic. Some costs can be avoided if the outage is minimal. A five-minute outage, for example, can be overlooked as a nice break for your employees in the middle of a busy workday. However, as the outage gets longer and longer, other contingency plans go into effect and the costs of the outage start increasing. Before you know it, companies that count on you for your product start looking elsewhere, and then things get really out of control.

Plan for All Types of Disasters

Many companies attempt to perform risk assessments of all possible disaster situations, determining for each the likelihood that it will happen to any particular data center. For example, coastal regions usually prepare for hurricanes or tsunamis. In the U.S., parts of Texas, Oklahoma, Kansas, and Nebraska have had so many tornadoes that they call it “Tornado Alley.” Other parts of the world are more susceptible to earthquakes. Do not dismiss any particular type of disaster with a “that will never happen” type statement. Murphy’s Law will find you. The disaster you do not prepare for is the one that will strike you.

Whereas natural disaster or malicious actions seem to be the most obvious cause of outage, accidental causes are probably most common. While people frequently accidentally delete files or misplace them, complete data loss has been caused by power outages when construction workers in the area cut power lines to the data center. Other common occurrences include water damage from broken water pipes and software upgrades gone awry. You should undertake some research to consider all of the possible types of disasters and make sure your disaster recovery plans take each of them into account.

Prepare for Cost Justification

Once you begin the process of selecting data protection systems, you need to justify the cost of each purchase. To be successful in doing so, you must have completed the steps mentioned previously in this section:

1. Define your RTOs, RPOs, backup windows, and consistency group requirements.
2. Determine for each critical system what to protect.
3. Determine the cost of each outage.
4. Plan for all types of disasters.

Once you have accomplished this, justifying the cost of each data protection system should be a relatively easy thing to do. You simply need to state your required RTOs and RPOs, what you’re protecting against, and what a system to protect against those things costs. If any part of the system is turned down, you simply need to explain how that affects your ability to meet these requirements.

Storage Security

Although the details of storage security are really beyond the scope of this book, it is important to understand that security is a very important part of data protection. This section gives you an overview of the vulnerabilities in storage systems.

Plain-Text Communication

In a storage network, we refer to communications within the network, such as a host requesting data from a storage device, as *in-band*. Historically, all of this communication has been in plain text. If someone can view in-band traffic, she might be able to read data she's not supposed to, or to learn something that might assist in an attack. We refer to communications outside the network—perhaps someone managing a storage device via its IP management port—as *out-of-band*. If someone can view out-of-band management information, they could take control over the storage network and give themselves access to information, or conduct denial-of-service (DOS) attacks.

The key to solving both of these problems is encryption. For out-of-band communication, more and more storage vendors are supporting secure communication protocols, such as ssh or https, on their management ports. For in-band support, there are host-based encryption systems and hardware encryption appliances. Only host-based encryption can encrypt data from the point of departure, but encryption software has typically been very CPU-intensive, slowing down the transfer of data by as much as 50 percent. The other in-band choice is an encryption appliance that can go in the storage network and encrypt data as it's stored on the device, preventing readability even if a hacker is able to gain physical access.

Poor Authentication and Authorization Systems

Unix's NFS and Windows' CIFS allow the sharing of files between multiple servers. This is collectively referred to as network-attached storage, or NAS. A major challenge with NFS and CIFS is their simple host-based authentication mechanisms. If your IP address resolves to the appropriate hostname, you are given access to the shared directory. In addition, much of the authentication mechanisms are also sent in plain text, telling a hacker exactly what addresses he needs to spoof. A hacker could easily spoof the appropriate IP address and be given access to the wrong information.

Fibre Channel SANs have authentication and authorization issues as well. Two very insecure, but very common, practices are the use of World Wide Name-based (WWN-based) zoning and soft zoning. (A *zone* is the Fibre Channel equivalent of a VLAN, with some differences.) Let's first take a look at the authentication issue, then we'll look at the authorization issue.

The authentication issue with Fibre Channel is the common use of WWN-based zones, where zone membership is determined by a host's WWN, which is equivalent to a MAC address. The problem with using WWNs for authentication is that they are easily spoofed. The ability to change the WWN is built right into the driver.

A much more secure, albeit slightly harder to manage, authentication method would be to specify zone membership using the switch port a given host is plugged into. Port binding, a recent advancement in Fibre Channel switches, can also improve WWN-based authentication. Using this authentication method, a WWN is bound to a particular port and is granted access only if it is seen at that port.

There are also authorization problems in Fibre Channel SANs, especially when using *soft zoning*. With soft zoning, you won't be able to query the name server to get members of a zone if you're not in that zone, but you can still communicate with a device if you have its WWN, which is relatively easy to determine. The opposite of soft zoning is *hard zoning*. With hard zoning, only members of a zone can access the devices in that zone.



Many people believe that soft zoning and WWN-based zoning are the same, and that hard zoning and port zoning are the same. This comes from the long-standing practice of offering them together. Nonetheless, they are two very different concepts. WWN-based and port-based zoning specify zone membership. Hard zoning and soft zoning specify whether or not zone membership is required to communicate with a member of a zone.

While the solution of using only hard zoning seems simple, it hasn't been that easy. Historically, soft zoning went hand in hand with WWN-based authentication, and many people use WWN-based authentication to make changes more easily. Today's switches are beginning to let you independently choose which authentication and zoning methods you want to use. The most secure combination, of course, would be hard zoning with port-binding-based authentication.

Backup Flaws

Backup systems' most obvious security flaw is the plain-text backup tape. There are many new encryption options for protecting this media. They include host-based filesystem and application encryption, encryption in the backup software, and a number of appliances that sit in the hardware data path and encrypt the data as it is written to tape. (Some of these appliances are now available inside the tape library or tape drive.) These hardware appliances are the most expensive, but they are much easier to implement and maintain than the other options. In addition to encrypting at line speed and providing superior key management, they also support compression. Since encrypted data can't be compressed, some have a compression chip that compresses the data before it's encrypted. This gives these appliances a major advantage over the other solutions, such as application encryption and backup encryption, because their encrypted data is not compressed by the tape drive.

The next security issue with backup systems is that they have typically used hostname-based authentication to authenticate the backup server and client to each other. A hacker with a spoofed IP address could do two things to exploit this vulnerability. First, she could create a rogue backup client and ask the server to restore data for the real client, thus stealing the information. A rogue client could also populate the backup server with bogus versions of backed-up files. A malicious hacker could also create a rogue backup server and back up any client that the server is authorized to back up. This, of course, would be a perfect way to steal or corrupt all kinds of data. Some backup products, including some of the open-source products discussed in this book, have addressed this serious vulnerability with additional levels of authentication beyond the hostname. Unfortunately, the added complexity of such authentication systems has made them less than attractive to backup administrators.

Most backup systems have taken an "all or nothing" approach to administrative authorization. This means that someone can do everything or nothing at all within the backup system. For example, by giving a new administrator the ability to eject tapes from the library, you also give them the ability to delete or change every backup policy, delete all backup history, and overwrite every tape you own with garbage. This presents the possibility of a novice administrator pushing the wrong button, accidentally erasing all the tapes in your tape library. A healthcare company actually had that happen a few years ago. Some backup software products have begun resolving this problem by introducing role-based administration, so you can give each person only the capabilities he needs to do his job.

The introduction of role-based administration in backup software, along with other new functionality to secure stored data, shows that storage vendors are waking up to the importance

of security. If your products don't support this kind of secure functionality, put pressure on your vendors so they understand it's critical for the safety of your most precious data.

Conclusion

Companies of all sizes must ensure that they are taking care of all parts of the data protection landscape that pertain to them. Unfortunately, the open-source projects that have enabled this book to exist have not yet embraced the archive, security, and disaster recovery elements of data protection in the same way they have embraced backup and recovery. While some aspects of archiving, security, and disaster recovery have been addressed in some open-source projects, we don't yet have open-source answers to the commercial products that are addressing today's archiving, security, and disaster recovery needs. Once we have open-source email archiving products, line-speed encryption products, or full-scale block-level replication products, it will be time for another edition of this book. For now, I hope this book has been helpful.



BackupCentral.com has a wiki page for every chapter in this book. Read or contribute updated information about this chapter at <http://www.backupcentral.com>.

Index

C

- cold backups, 14
- cost
 - backups, 20

D

- data protection, 3
 - backup and archive, 9
 - business continuance volumes (BCVs), 18
 - business issues, 3, 16
 - cold backups, 14
 - cost justification, 20
 - database backup agents, 15
 - device issues, 6
 - disaster recovery (DR), 15
 - email, 8
 - external threats, 8

- improving service, 6
- mitigating risk, 4
- open-source, 13
- outage costs, 19
- poor authentication and authorization, 21
- recovery point objective (RPO), 17
- recovery time objective (RTO), 17
- reducing costs, 5
- regulatory compliance, 4
- scripted hot backups, 14
- shared storage, 15
- soft versus hard zoning, 21
- storage security, 20
- synchronization requirements, 18
- systemic flaws, 22
- very active filesystems, 13
- very large filesystems, 14

Check out the Resources from our Sponsors

data domain

"Buyers Beware: All Data Deduplication is NOT Created Equal" - StorageSwitzerland

Data Domain. #1 in Deduplication storage. Highest in use dedupe vendor...

Top Deduplication FAQs: What Everyone's Asking

EQUALLOGIC[®] SIMPLIFYING NETWORKED STORAGE™

PS SERIES BEST PRACTICES Deploying Microsoft® System Center Data Protection Manager 2007 in an iSCSI SAN

PS Series Groups Backup and Recovery Overview

Delivering Holistic Business Continuity with Auto-Replication

FalconStor Software

Korea Telecom (KT) Case Study

Demystifying Data De-duplication: Choosing the right solution

Accelerating Backup/Restore with the Virtual Tape Library Configuration that Fits your Environment

Overland STORAGE[®]

Business Continuity planning resource page

REO SERIES is the world's best-selling disk-based backup and recovery appliance

Calculate your Cost of Downtime now!

Quantum

Experiencing Data De-Duplication: Improving Efficiency and Reducing Capacity Requirements

The Power of Disk-Based Backup: Advanced Data De-Duplication Article

Taneja Group Opinion - Quantum Establishes Itself as a Formidable Player in Data De-duplication

SEPATON[®]

Gartner Report: Virtual tape library overview, including user recommendations

ESG Lab Validation Report: SEPATON S2100 Virtual Tape Library with Deduplication

TCO Report: Comparing enterprise data protection costs of tape, VTL and deduplication solutions

O'REILLY[®]

Purchase your own copy of Backup & Recovery from O'Reilly Media and get 35% off the purchase price.

Simply enter discount code D7BRTT at checkout to take advantage of this limited time offer.