Realtime
publishers

The Shortcut Guide™ To

Availability, Continuity, and Disaster Recovery

sponsored by

ARCserve.com™

Dan Sullivan

Introduction to Realtime Publishers

by Don Jones, Series Editor

For several years now, Realtime has produced dozens and dozens of high-quality books that just happen to be delivered in electronic format—at no cost to you, the reader. We've made this unique publishing model work through the generous support and cooperation of our sponsors, who agree to bear each book's production expenses for the benefit of our readers.

Although we've always offered our publications to you for free, don't think for a moment that quality is anything less than our top priority. My job is to make sure that our books are as good as—and in most cases better than—any printed book that would cost you $40 or more. Our electronic publishing model offers several advantages over printed books: You receive chapters literally as fast as our authors produce them (hence the "realtime" aspect of our model), and we can update chapters to reflect the latest changes in technology.

I want to point out that our books are by no means paid advertisements or white papers. We're an independent publishing company, and an important aspect of my job is to make sure that our authors are free to voice their expertise and opinions without reservation or restriction. We maintain complete editorial control of our publications, and I'm proud that we've produced so many quality books over the past years.

I want to extend an invitation to visit us at http://nexus.realtimepublishers.com, especially if you've received this publication from a friend or colleague. We have a wide variety of additional books on a range of topics, and you're sure to find something that's of interest to you—and it won't cost you a thing. We hope you'll continue to come to Realtime for your educational needs far into the future.

Until then, enjoy.

Don Jones

Copyright Statement

Chapter 1: The Business Case for Recovery Management

Professionals can develop their businesses with effective strategies, stay ahead of the competition by analyzing dynamic market conditions, and build brand loyalty with exceptional customer service—and it could all turn out to be for nothing. That is, if the IT systems that store, manage, and distribute our information fail and there is no recovery management process. It would be as if we had never done our work in the first place.

Data loss is an all-too-common problem. We lose information on the small scale with damaged laptops and misplaced flash drives. We lose information on the large scale with natural disasters that destroy entire data centers. Sometimes human error is at fault and sometimes applications fail with unfortunate consequences. Regardless of the cause of the initial data loss, the ripple effects can result in redundant work to reproduce the lost data, or in the worse case, to legal liabilities, brand damage, and business disruptions.

Recovery management is a framework to mitigate the risk of lost data and lost IT systems. It includes practices such as making backup copies of essential data, maintaining stand-by systems in case primary systems fail, and establishing policies and procedures to cost-effectively protect business assets, applying appropriate procedures based on the value of the data. Of course, no business practice will eliminate all risk or guarantee we can recover from calamitous events. We can implement cost-effective measures that allow a business to continue to operate at, or near, normal operating levels in spite of adverse events.

The purpose of *The Shortcut Guide to Availability, Continuity, and Disaster Recovery* is to provide you with the information you need to understand the business drivers behind recovery management, the technical aspects of recovery management, some of the operational challenges you might face, and best practices for implementing recovery management.

The four chapters of this guide cover the following topics:

- Chapter 1 discusses the obvious and sometimes not-so-obvious business drivers behind recovery management. The chapter also describes how to develop a recovery management strategy, including assessing threats and risks and outlining policies and applications needed to implement an effective recovery management strategy.
- Chapter 2 examines the challenges posed by the increasing complexity of IT environments, including virtualization, application-specific backup requirements, and remote office protection.
- Chapter 3 delves into common issues in recovery management, including scheduling and monitoring, media options, controlling costs, the growing volumes of data, and recovering in case of disaster.
- Chapter 4 digs into the details of how different types of organizations frame their recovery management strategies. The chapter concludes with a discussion of best practices for availability, continuity, and disaster recovery.

We start our examination with the business requirements that drive the need for recovery management.

Keeping the Business Running: The Obvious and Not-So-Obvious Business Requirements for Recovery Management

Recovery management addresses the threats of different kinds of losses, from hardware failures and software bugs to stolen laptops and malicious acts. One of the surprising aspects of recovery management is the number of different situations that benefit from having a sound plan in place. Some of these are obvious, but many are not.

The Obvious Business Requirements for Recovery Management

The “obvious” drivers behind recovery management are the reasons that come to mind first when we think of file backups and stand-by servers:

- Isolated software or hardware failures
- Natural disasters and catastrophic failures
- Compliance with regulatory or internal policies

It is easy to imagine a what-if scenario in these areas, especially if you have ever had a hard drive fail or tried to recover a system after fire or water damage. A brief meeting with an auditor can dispel any lax attitudes toward maintaining the integrity and availability of essential corporate data.

Isolated Failures

Isolated failures are limited in scope affecting few people or business processes. A typical example is the accidently-deleted file. Someone may decide they no longer need a file and delete it. Fortunately, popular end user operating systems (OSs) frequently have a staging area for deleted files (for example, the Windows Recycling Bin and the Mac OS trash can) so that removed files can often be recovered by end users. Once the staging area of deleted files has been purged, restoring a backup copy of the deleted file is the best way to recover it.

Undelete Programs Helpful But Not Enough

Utility programs are available for recovering files even after they have been purged from a staging area, like the Recycling Bin. These programs work by reclaiming the data blocks on the disk that contain the contents of the deleted file before another application overwrites those blocks. Once the contents of a block have been written over, there is no easy way of recovering it, at least by conventional standards.

Application errors present another type of isolated error. A bug in a database application may incorrectly update data. As with OSs that provide a staging area for deleted files just in case there is an error, databases often store recovery information at least for short periods of time. If an error is caught in time, the database application can help recover the correct data. After that, restoring data from a backup copy is often the preferred method for correcting the mistake.

In addition to our own applications, we need to consider the risk of malicious applications, generally known as malware. Sometimes these programs are designed to corrupt files on compromised devices. If the corruption is found in time, backups can be used to restore files to their original states. Unfortunately, not all data loss incidents are so easily remedied.

Natural Disaster and Catastrophic Failures

Many of us only think of natural disasters when we are paying our insurance premiums. Like insurance, though, we will be glad to have backups and disaster recovery plans if disaster ever occurs.

Large-scale disasters, such as Hurricane Katrina in 2005 and the Northridge California earthquake in 1994, are infrequent, but fires, flooding, and other local events are common enough to warrant disaster recovery planning. Some of the key elements one needs to consider in a recovery management strategy relate to these catastrophic failures. When formulating a recovery management strategy and defining requirements, consider questions such as:

- If the data center with production servers were unavailable, what business operations would be affected?
- If a data center was destroyed by fire and all local backups destroyed, how would we recover?

- If the data center were unavailable, where would we house a temporary data center?
- How long would it take to restore business operations?
- How should we prioritize the restoration efforts? Which business process are the most critical?
- What level of degraded performance can be tolerated in disaster recovery mode?
- What is the cost per hour when an application or data center is unavailable?

Stakeholders in a business depend on IT professionals to protect information assets of the enterprise from the worst consequences of disasters. Regulations and internal policies define collective expectations for protection. Ensuring compliance is another obvious driver behind recovery management.

Compliance

There are many dimensions of governance and one of them is ensuring that business can continue to function under a range of circumstances, including the failure of key processes and systems. Recovery management plays an important role here. In the event of technical failure, human error, or natural disaster, business has a means to recover and re-establish a normal operating mode.

Compliance often entails more than just having a “Plan B” in the event of disaster. We need to demonstrate we have that capability in place and test it periodically to ensure our recovery management policies and procedures continue to meet the changing needs of the business.

Hardware failures, natural disasters, and compliance are obvious drivers behind the adoption of recovery management practices. They are not, however, the only aspects of business operations that should drive, and benefit from, recovery management.

The Not-So-Obvious Requirements

It is easy to think about backups and disaster recovery in the most basic terms: Make copies of important data so that you can restore in case of an adverse event. This is certainly sound reasoning but it does not capture everything we need to consider about recovery management. The problem with this line of reasoning is that it focuses only on data and not on other business aspects that drive the creation and use of that data in the first place.

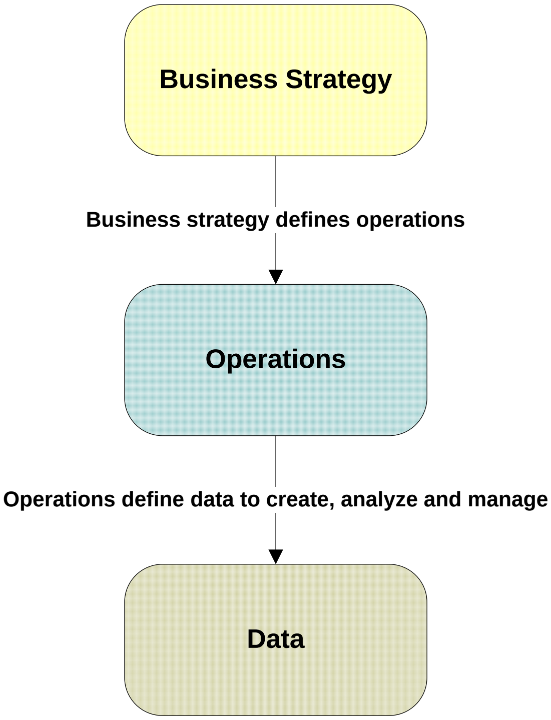
Business Strategy
Business strategy defines operations
Operations
Operations define data to create, analyze and manage
Data

Figure 1.1: Additional requirements for recovery management become clear when we consider the business strategy and operations that drive the creation, analysis, and management of business data.

If we examine why we create, analyze, and manage the particular types of data we have, we will find that the tasks are tied to some operational process. For example, we keep customer data for order fulfillment and sales operations. Human resources data is kept to track employees' performance history, benefits, and skill sets. These operations are in turn created in order to execute a business strategy, such as increasing market share, improving customer service, and retaining top talent.

If we consider each level of this three-tier model as a source of influence on recovery management, we can ask two broad questions. First, how does each tier shape requirements for recovery management? Second, does recovery management enable new capabilities that allow us to expand or improve each tier? To answer these questions, we will start at the bottom and work our way up with:

- Data-driven recovery management requirements
- Operations-level opportunities and constraints
- Business strategy and its demands for recovery management

These levels all include a combination of business and technology issues but with varying emphasis. Data-driven requirements are dominated by technical considerations while business strategy is, not surprisingly, subject primarily business considerations.

Data-Driven Recovery Management Requirements

Rule number one of data-driven recovery management is that not all data is of equal value. Before we can define recovery management procedures, we need to understand how data falls into different groups based on:

- How long we have to recover data once an adverse event occurs before the business suffers
- How much data can be lost because it was not backed up before an adverse event
- How fast the volume of data is growing

Sometimes it is more important to recover all data than it is to get it back quickly. A company's financial database may be down for several hours without significant impact on the business, but if even a single entry in the general ledger were missing, the integrity of financial reports is lost. In other cases, the time it takes to recover data is the most important factor. For example, as long as a company's product catalog is unavailable for online purchases, online revenues stop and purchases are potentially lost to competitors.

Recovery Time Objectives and Recovery Point Objectives

The duration between a data loss event and the point at which the data should be available again is known as the recovery time objective (RTO). The point in time from which we should be able to recover lost data is known as the recovery point objective (RPO). RTO specifies how long we can tolerate being without our data; RPO specifies how much lost data (in terms of time windows) we can tolerate.

RPOs are based on how much data we are willing to lose to a data loss event. Figure 1.2 depicts a basic backup strategy employing nightly backups. At any point in time, we can recover all the data from the previous day, but any data created or modified during the day a data loss event occurred would not have been backed up. This may be sufficient for applications with a low number of transactions during a day, such as an HR database tracking changes to employees' 401(k) funds. If data is lost, it is neither difficult nor expensive to recreate it. Applications with high levels of transactions or those for which recreating data would be difficult require more robust recovery management strategies, such as continuous data protection.

Cross-Reference

We'll talk more about continuous data protection later in this chapter in the Expectations for Continuous Availability section.

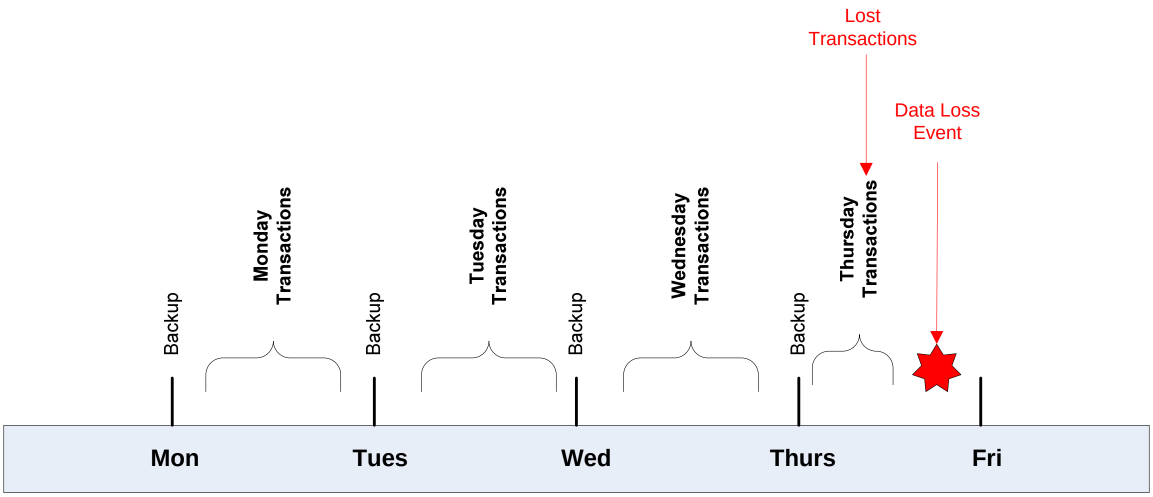

Figure 1.2: In the case of a simple example, backups are performed nightly. This implements the previous day's close of business as the RPO. In this example, the business is willing to risk the need to recreate up to a full day's worth of transactions.

In addition to deciding on an RPO, we must decide how long we are willing to be without our data. Some categories of data can have relatively long RTOs. Again, an HR application may be down for a day without severe adverse consequences. Sales and customer support applications and data, however, may require near continuous availability. In the event of data loss, the business operations that depend on these systems may not tolerate the time it would take a systems administrator to find the proper backup tape, select the lost data, and restore it to the application. In this way, our RTOs and RPOs constrain our options for implementing backup and recovery.

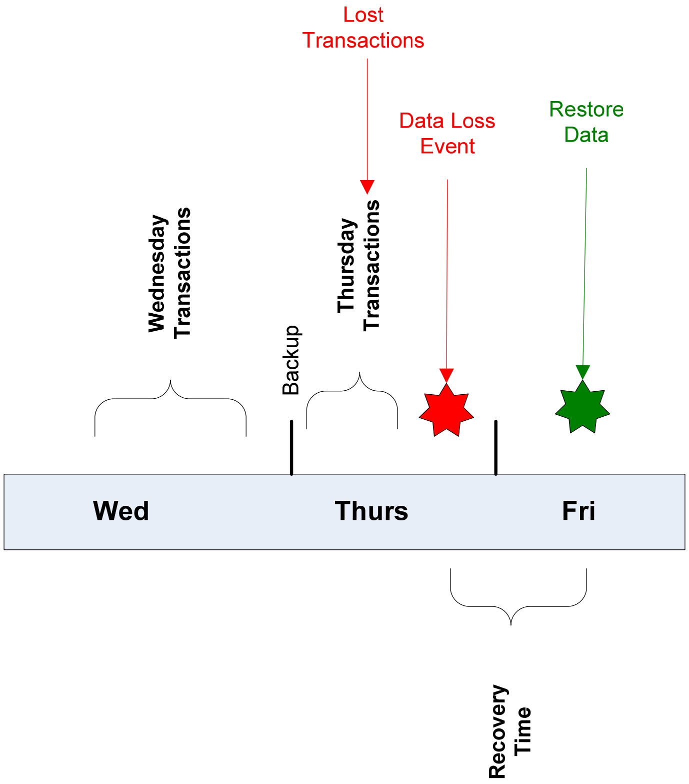

Figure 1.3: RTOs are defined by the amount of time that can pass between a data loss event and the restoration of data before there are adverse consequences for business operations.

Data Growth and Its Impact on Recovery Management

Another constraint on how we implement backups and disaster recovery procedures is the rate at which data volumes grow. There are many sources for increasing volumes of data:

- More automated contacts with customers, such as through self-service systems and online account management
- Increased use of business intelligence techniques to analyze sales, marketing, and operations data
- Detailed account and auditing of transactions in support of compliance and security management efforts
- Innovations in products and services offered to customers that generate new data
- Increased use of email and other collaboration systems

The rapid growth in data volumes is driving the adoption of better data management techniques, such as more efficient management of network storage devices and the use of deduplication in backup systems.

Data is easily duplicated. Database records, email messages, and multiple versions of a document can all be data structures that result in redundant data. An obvious question is: Why backup up and store redundant copies? Why not backup up one copy and track references to where the data is re-used? This is exactly what data deduplication does.

Deduplication processes operate either at the source system being backed up or at the target system receiving the copy of the backup. As each block of data is processed, the deduplication process determines whether a block of data with the same content has already been backed up. If it has, the system stores a reference to the copy that was made earlier instead of making another copy of the block.

Expectations for Continuous Availability

As our expectations for continuous availability grows, acceptable RTOs shrink. It is difficult to find maintenance windows to update applications, patch OSs, and perform other routine maintenance because customers are coming to expect 24 hour a day, 7 day a week access to applications. Again, the answer depends on the type of data and its level of criticality for business operations, but it is safe to say, for many customer-facing applications, the tolerance for downtime is close to zero. Businesses look to continuous data protection to ensure continuous availability. If data is so critical that we cannot tolerate virtually any downtime, data replication is probably the appropriate strategy.

With replication, as data is written to a primary system, it is copied to a stand-by system that maintains a close to real time copy of data from the primary system. If the primary system fails, operations switch to the stand-by system and continue as normal. When the primary system is restored, data that had been updated on the stand-by server is copied to the primary server and then operations can be shifted back to the primary system.

Key considerations include:

- Time required to update the stand-by system once a change has been made to the primary system
- Time to switch from a failed primary server to the stand-by server; should it be done automatically or can it be done manually and still meet RTO objectives?
- Tolerance for degraded performance with the stand-by server. For example, could the stand-by server be a virtual server sharing a physical server with several other stand-by virtual machines?
- Can the stand-by server be used for read-only operations, such as loading a data warehouse or generating reports, to reduce load on the primary server?
- Cost of additional hardware, and possibly software licenses, to perform replication

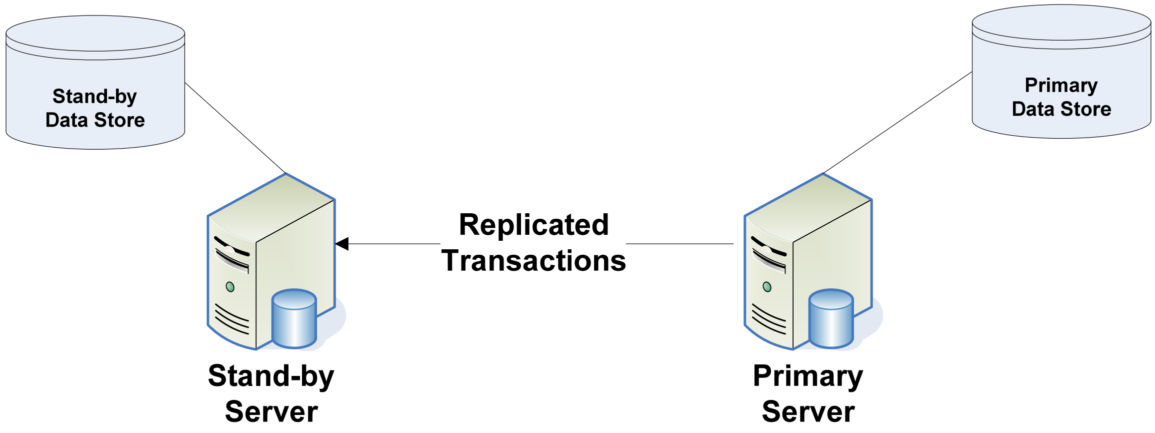

Figure 1.4: Replication duplicates all transactions on a stand-by server. In the event of a server or storage failure on the primary devices, the stand-by devices can be rapidly deployed.

Replication supports disaster recovery as well as continuous availability. Stand-by servers may be located in different offices or data centers from the primary site. This helps to mitigate the risk of site-specific threats, such as fire and flooding, to the primary site.

Data-Driven Requirements

The type and volume of data we have drives some recovery management requirements. Factors such as how long we can function without certain types of data and how long we can wait before data is restored have long been fundamental issues. The increasing volumes of data are also driving the need for more cost-effective storage strategies such as data deduplication. Expectations for continuous availability and the needs of disaster recovery are well met by replication technologies. In addition to these data-drive requirements, the day-to-day operations required to maintain an IT infrastructure are also the source of recovery management requirements.

Understanding Operations-Driven Requirements

Operations-driven requirements focus on the implementation aspects of recovery management. These are the issues that systems administrators and IT managers have to consider when formulating the best way to implement the data-driven and business strategy–driven requirements. Three commonly-encountered types of requirements are:

- Effective and efficient operational management
- Opportunities and constraints with virtual environments
- Application-specific backup requirements

We will delve into the technical details of these and other operation issues in Chapters 2 and 3, so we will just introduce some of the most salient elements of these issues here.

Effective and Efficient Operational Management

In an ideal world, recovery management procedures require minimal manual intervention, especially with backups, replication, and other ongoing tasks. Even in our less-than-ideal world, backup procedures and disaster recovery preparations should be as automated as possible for two reasons. The more automated the procedure, the less opportunity for human error. Backups can be scheduled for automatic execution. Logs can be generated for future reference. Errors can be flagged and generate alerts to notify systems administrators.

Restoring data and services can be automated as well. When replication is used, automatic failover to a stand-by server can be enabled, at least with some replication systems. A drawback of this type of rapid failover is a potentially more complex configuration. For example, an additional service may be required to detect the failure of the primary server and automatically redirect service to the stand-by server. Alternatively, if automatic failover is not used, a systems administrator could manually update a local domain name server to map a domain name to the stand-by server instead of the primary server.

Opportunities and Constraints with Virtual Environment

Virtualization can significantly increase server utilization and reduce costs but it also introduces new variables into the recovery management equation. The most basic question is how should we back up our virtual machines? There are several options:

- Treating virtual servers the same way we treat physical servers
- Shutting down the virtual machine and copying the machine's files to backup media
- Making a snapshot copy of the virtual machine while it is running and backing up that copy

Treating virtual servers as physical servers can simplify backup procedures, but it requires installing a backup client on each virtual machine. The second option eliminates the need for installing a client in exchange for shutting down the virtual machine. This may be acceptable depending on the function of the server. Snapshot copies require a staging area to store the snapshot and can briefly degrade performance of the virtual machine while the snapshot is made. The best option will depend on the specific business requirements of the virtual machine.

Application-Specific Backup Requirements

Backup and restore operations become more complex when we are working with files that are used with applications such as email servers and database servers. Consider some of the characteristics of database servers, for example. Databases typically use a small number of large files to store data about a large number of transactions. This has a number of implications for backup and recovery operations:

- If a single record in the database must be restored, will the entire contents of the storage file have to be restored?
- A single transaction may cause updates to multiple storage files. What is the optimal method for organizing those files to reduce the number of files that are updated by a single transaction?
- Can backups be performed while the database is active or must it be offline to ensure consistency across all transactions?

As we can see, the way applications use file storage to implement services, such as data management and email, can have an impact on the way systems administrators implement recovery management operations. Recovery management requirements are shaped in part by operational considerations, such as the efficiency of day-to-day procedures, the increasing use of server virtualization and the implications for backup operations, and application-specific constraints on the way we perform backups for databases and email systems.

Understanding Business Strategy–Driven Requirements

Unlike data- and operations-driven requirements, business strategies are as varied as businesses themselves, so there is no universal set of requirements we can all adopt as our own. Instead, in this section, we will consider two broad strategies that can provide examples to help elucidate the types of business strategy–driven requirements in your own business.

Improving Customer Service

A business may decide that providing online access to detailed, historical account data is crucial to improving customer service. Implementing this strategy will require increasing amounts of storage to support the customer service application, but it will also increase the demands on recovery management services. These increased demands include additional backup storage and increased throughput to continue to meet RTO and RPO with larger volumes of data. Meeting these demands can be done with a combination of additional hardware and network services as well as improved backup techniques, such as deduplication.

Maintaining Continuous Access to Business Services

Availability is a fundamental attribute of online services. It would be hard to imagine running a factory without a steady supply of power; it is equally difficult to imagine running a modern business without continuous access to the applications and data that provide business services. To mitigate the risk of lost services, businesses can implement redundancy at multiple levels:

- Redundant disk arrays
- Replicated data
- Stand-by servers
- Multiple points of access to the Internet
- Redundant power sources in data centers
- Physically distributed servers
- Well-defined backup and restore procedures

We must remember that there are many ways a business service can fail, so there will be multiple techniques required to mitigate that risk. Both traditional backup operations and replication services should be considered when there is a need to maintain continuous access to business services.

By considering recovery management from the perspective of data, operational, and business strategy requirements, we can identify essential aspects of business processes that need protection. The next logical step is to develop a plan for addressing those needs.

Developing a Recovery Management Strategy

The first stage in developing a recovery management strategy is assessing threats and risks to services. This is followed by assigning RPOs and RTOs as well as defining the policies and applications needed to address those threats and ultimately implement the recovery management strategy.

Assessing Risks and Threats

Risks are adverse outcomes that we typically want to protect against, such as data loss, system failure, or security breaches. Threats are ways in which a risk can be realized, for example, a data loss (the risk) could occur if a poorly-developed application inadvertently deleted files from a server (the threat). Although there are many types of threats, we will consider several with obvious impact on recovery management.

Threats to Data and IT Operations

Threats that disrupt the functioning of IT services fall into several categories, all of which must be addressed in a recovery management strategy:

- Hardware and software failures
- Malware and other security threats
- Natural disaster
- Human error
- Power failure

Hardware and Software Failures

Hardware failures are better understood than software failures. Consider the fact that when we buy hard drives, we can get estimated mean time between failures. This metric does not tell us when a hard drive will fail, but it at least gives us some idea of how long we can expect the device to function, at least on average. Software, including OSs, is more complex and diverse as well as developed under widely varying levels of quality control. There are no well-established metrics comparable to mean time between failures for measuring the reliability of software. From a recovery management perspective, it is safe to assume that both will fail and could corrupt relatively isolated sets of data or damage entire disks of data; given that assumption, we backup appropriately.

Malware and Other Security Threats

Security threats can pose significant threats to information systems. Threats such as viruses, worms, Trojan horses, and blended threats (multiple attack vectors in a single package) can all be used to destroy or tamper with data. Data breaches that result in large numbers of disclosed records are well documented in the popular press. Security threats to the integrity and availability of data are less frequent topics of discussion but still dangerous to businesses. Reliable and timely backups can make a significant difference in the cost and time it takes to recover from a security breach. Of course, if files on backups are corrupted by malware or other security threat, this is not an option,. Often the best strategy is to have a security management strategy in place to mitigate the risk from malware and other security threats. One way to mitigate malware risks is to use backup software that contains antivirus software, which can scan files during both backup and restore operations.

Natural Disasters

Natural disasters need little explanation. The key questions we need to answer about disaster recovery include where to store backup copies of data and stand-by servers, how long are we willing to tolerate service disruption, and what procedures need to be in place to ensure services can be started at a disaster recovery site. In addition, what are the steps to resuming normal operations once the primary site is up and functional?

Human Error

Human error will always be with us, so we must design systems in ways to minimize the potential impact of error. Programmatic techniques, such as validating input and prompting for verification of destructive operations, are one way. Organizational techniques, such as separation of duties and requiring authorizations from multiple individuals are another way to mitigate the risk of human error.

Power Failures

Disruption caused by power failures can be mitigated with multiple power supplies. Large data centers may employ a redundant source of primary power, including on-site generators, which may not be practical for smaller facilities. Facilities of any size should consider uninterruptable power supplies (UPS) for temporary power. A UPS can provide power during brief outages and allow time for a controlled shutdown of systems in the event of long outages. These types of risks are just some of the ways risks to business can be realized.

Risks to Business

When systems are down and information is unavailable, businesses are adversely affected. Some of the most immediate concerns we have about loss of business continuity are:

- Lost productivity and backlog of work
- Loss of revenue because sales cannot be completed, pre-sales information cannot be provided, orders cannot be processed
- Loss of customer confidence and brand damage that can arise from the inability to access systems and account information or execute transactions; there is also the potential for lost confidence in the business to provide reliable, robust services
- Cost to restore operations to a normal state; without proper planning and disaster recovery management, the task of recreating data and reinstalling systems under tight deadlines can be costly
- Cost of fines for compliance or e-discovery violations resulting from being unable to produce data as required.

Businesses face a host of risks to their operations and many risks can be realized by multiple types of threats. It is prudent, and cost effective, to plan ahead and develop a recovery management strategy before an adverse event occurs.

Elements of a Recovery Management Strategy

A sound recovery management strategy is a combination of (1) policies that address the various data, operations, and business requirements with respect to the risks and threats a business faces and (2) applications and technologies that enable the implementation of those policies.

Recovery Management Policies

The purpose of recovery management policies is to document and put into practice methods for mitigating the risks facing businesses. Five types of policies are essential:

- Backup policies
- Continuity and failover policies
- Disaster recovery policies
- Testing policies
- Security policies

Policies should define the scope of what should be done to mitigate risks; technical implementation details are defined after policies are formulated. They are codified as procedures that are executed by systems administrators and other IT professionals responsible for day-to-day operations.

Backup Policies

Backup policies specify what types of data and applications should be backed up, the RPO for each type of data, and the RTO for each as well. For example, and HR database may have an RPO of the previous business day and an RTO of 4 hours. Procedures for this policy may call for a combination of weekly full backups plus incremental nightly backups.

Continuity and Failover

Continuity and failover policies focus on critical data and applications. The purpose of these policies is to ensure that systems that should be available at all times are protected with high-availability techniques. For example, a sales database may have an RPO of the last 10 minutes and an RTO of 10 minutes as well. These demanding constraints warrant a replication-based solution.

Disaster Recovery

Disaster recovery policies specify what disaster recovery procedures should accomplish and who should be involved. These policies specify criteria for establishing disaster recovery sites or services, such as location in separate buildings or different localities depending on the criticality of the data and services protected. They should also specify the RPOs and RTOs of different categories of services. The policy should also include some description of when disaster recovery procedures are implemented, typically when service infrastructure is so compromised that normal services cannot be maintained.

Testing

Disaster recovery policies should also indicate the need to test disaster recovery procedures and systems at regular intervals. Modifications to procedures should be tested when they are implemented and then tested again during regularly scheduled test operations.

Security

Security policies must take into account much more than recovery management but should include directives on the appropriate use of the Internet and restrictions on installing non-authorized software on company devices.

In addition to policies defining what is required of disaster recovery procedures, we need applications to meet those needs.

Applications

Disaster recovery depends on two types of systems: backup and restore applications and high-availability solutions. Backup and restore applications give us the means to recover from a wide array of adverse events, from hardware failures that lose data and software bugs that corrupt the integrity of data to natural disasters that destroy entire data centers. It is important to consider backup and recovery operations when deploying new systems and implementing new business services. We must be able to back up and restore critical data with the time ranges allotted to us by the business. Growing volumes of data make this more difficult; however, techniques like deduplication can help us keep pace with the growth in data volumes.

High-availability solutions allow us to replicate services and data on stand-by servers and keep them up to date. These solutions are essential when we must maintain 24 × 7 systems and allow for extremely short RTOs.

Summary

A recovery management strategy should take into account a variety of requirements. Some of these requirements are a function of the criticality of the data we have, some are dictated by operational and efficiency considerations, and others are derived from business strategy. Regardless of the source of the requirements, a sound recovery management strategy starts with codifying those requirements in policies that can be used to develop operational procedures to protect business services and data. Applications such as backup and restore systems and high-availability solutions play a critical role in implementing those policies and procedures. We will turn our attention to those implementation issues in the next chapters.

Chapter 2: Breaking Through Technical Barriers to Effective Recovery Management

Information technologies are constantly advancing in ways that enable businesses to execute their strategies more efficiently and effectively than was possible previously. Virtualization improves server utilization, relational database provides high-performance data services, and email offers what has become a dominant form of business communication. With these advances come new levels of complexity, some of which have a direct impact on recovery management practices.

In addition to technical advances, there are organizational structures that create technical challenges to effective recovery management. The physical distribution of offices, for example, affects how we implement recovery management practices. If a company has multiple sites, it may not be practical to have skilled IT support in each office. Centralized IT support is often more economical; however, it raises the question of how to remotely provide recovery management protection. What starts as an organizational issue quickly leads to technical issues.

This chapter will examine several technical barriers commonly encountered when implementing recovery management services. These common challenges include:

- Protecting virtual environments
- Meeting the specialized backup and recovery requirements of databases and content management systems
- Solving remote office backup and recovery challenges
- Ensuring continuity in disaster recovery operations

Throughout this chapter, we will see examples of the need to adapt recovery management techniques to application-specific requirements and systems-implementation–specific requirements. These examples show that recovery management is much more than simply a matter of backing up files.

Protecting Virtual Environments: Challenges and Solutions

Server virtualization is widely adopted because it allows us to utilize computing resources more efficiently. With multi-core and multi-CPU servers, we can run compute-intensive jobs faster and on fewer servers. If we have a steady stream of these CPU-hungry applications, we can keep a server utilized. This is often not the case, though.

Typical business workflows have periods of heavy demand, such as when generating business intelligence reports for department heads and line managers, followed by periods of low demand. If we were to dedicate a single server to a single application, we would find under-utilization during off-peak periods. However, running multiple applications on the same instance of an operating system (OS) can lead to problems with incompatibilities. One application may require the latest set of link libraries while another older application is incompatible with those. Security requirements, like access to the root or administrator account, may be incompatible. The maintenance requirements of one application might require OS reboots that would unnecessarily shut down the other application. Virtualization avoids these problems.

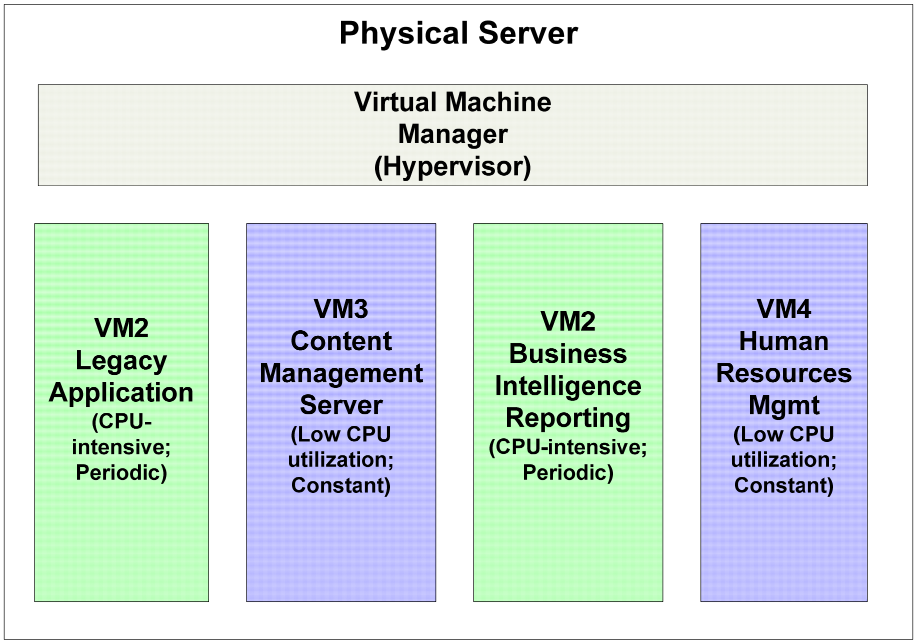
Physical Server
Virtual Machine Manager (Hypervisor)
VM2 Legacy Application (CPU-intensive; Periodic)
VM3 Content Management Server (Low CPU utilization; Constant)
VM2 Business Intelligence Reporting (CPU-intensive; Periodic)
VM4 Human Resources Mgmt (Low CPU utilization; Constant)

Figure 2.1: Virtualization allows for efficient mixed-workload combinations on a single physical server. Incompatible dependencies, security requirements, and maintenance requirements can make mixed workloads on a single OS instance impractical.

Of course, each virtual machine will have its own backup and recovery requirements. Meeting each set of requirements while minimizing the impact on other virtual machines is a source of new challenges we have to address.

Challenges Introduced by Virtual Environments

The types of challenges introduced by virtual environments fall into three broad categories:

- Performance issues
- Granularity of backup and restore operations
- Management issues

These challenges will strongly influence how we structure and schedule our backup operations.

Performance Issues

Backing up virtual machines can be compute-intensive, especially if deduplication and compression are done on the source system. In a virtualized environment, the guest OSs share the same physical resources, such as memory and bandwidth. If all the virtual machines were to run backup operations at the same time, there would likely be contention for these shared physical resources. Similarly, if backup operations on one virtual machine were schedule during the peak demand period of another virtual machine, such as a business intelligence reporting system, there could be contention for resources. In both scenarios, backups may not finish in the time allotted to perform them.

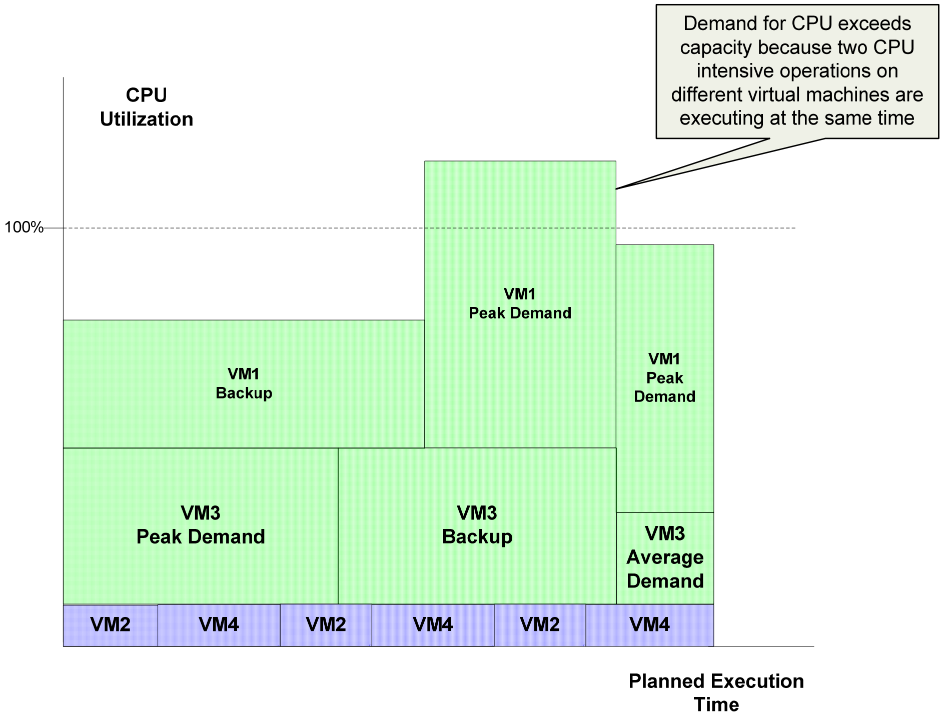

Figure 2.2: One of the challenges in managing virtual environments is understanding the distribution of workloads across virtual machines on a single server. Backup operations that execute at the same time as other CPU-demanding applications can exceed the available capacity and cause operations to run for longer times than planned.

Deduplication is an especially effective way to improve backup performance. Deduplication performed on the backup target relieves the source system of the CPU load associated with the process. Virtual machines tend to create a high level of duplicate data, so deduplication will result in cost savings as well.

Granularity of Backup and Restore Operations

A virtual machine can be backed up as a virtual machine image (and related configuration files) or as a set of files. Backing up a virtual machine as a single image can simplify backup procedures. All components of the virtual machine are backed up under the same criteria and a single backup image contains the entire virtual machine. This approach would be of limited value, though, if the backups could only be restored as a full image; often, restore operations are targeted to a single file or relatively small set of files.

Management Issues

There are a few management issues with regards to backing up virtual environments. Virtual sprawl, or the rapid deployment of virtual machines (sometimes outside of standard operating procedures), can cause headaches for the systems managers left to back up the growing numbers of virtual machines. Unless proper controls are in place to control provisioning and deprovisioning, the number of virtual machines active on a server can change quickly with new virtual machines adding to already complex backup schedules. Sometimes images of virtual machines that are no longer in use are left intact and included in backup operations, unnecessarily using compute, storage, and network resources.

Another management issue arises with regards to disaster recovery. Backups for disaster recovery purposes ideally could be restored to a bare-metal server in a different configuration of virtual machines. For example, a developer's server may host development and test virtual machines under normal operating conditions. Under disaster recovery conditions, the testing virtual machine may not be deployed in order to make resources available for mission-critical applications.

Virtual machines are started and shut down as dictated by business requirements. If a virtual machine is shut down during the backup window for its host, it should still be backed up. As virtual machines persist as images on disks, they should not have to be started to perform the backup. However, if the backup software used does not allow for image backups with file-level restores, systems administrators may opt to restart the virtual machine rather than forfeit the flexibility of restoring at the file level.

Options for Backup and Recovery in Virtual Environments

Systems administrators have a number of options for backing up virtual environments:

- Traditional file-level backups
- Virtual machine image backups
- Mixed-mode backups
- Backup by proxy

These options are not always mutually exclusive. For example, virtual machine image backups may or may not be performed using a backup proxy. It should be noted that this discussion is focused on how backup applications used in physical server environments can be leveraged in virtualized environment. In addition to these approaches, in a virtual environment, one may also be able to use storage area network (SAN)-specific techniques, such as creating snapshots of allocated storage units.

Traditional File-Level Backups

Virtual machines can be treated as equivalent to physical servers. Under this scenario, systems administrators would install a backup application and run whatever combination of full, incremental, and differential backup that is required for the virtual machine. The virtual machine will need to write data to a backup server, so sufficient network bandwidth will need to be available. Care must be taken to schedule backups at times that do not adversely affect other virtual machines on the same host (See Figure 2.2).

One of the advantages of the traditional file-level backups is the ability to perform fine-grained restore operations. Individual files are easily restored with this method. At the same time, however, this may not be a reliable method for full virtual machine restores.

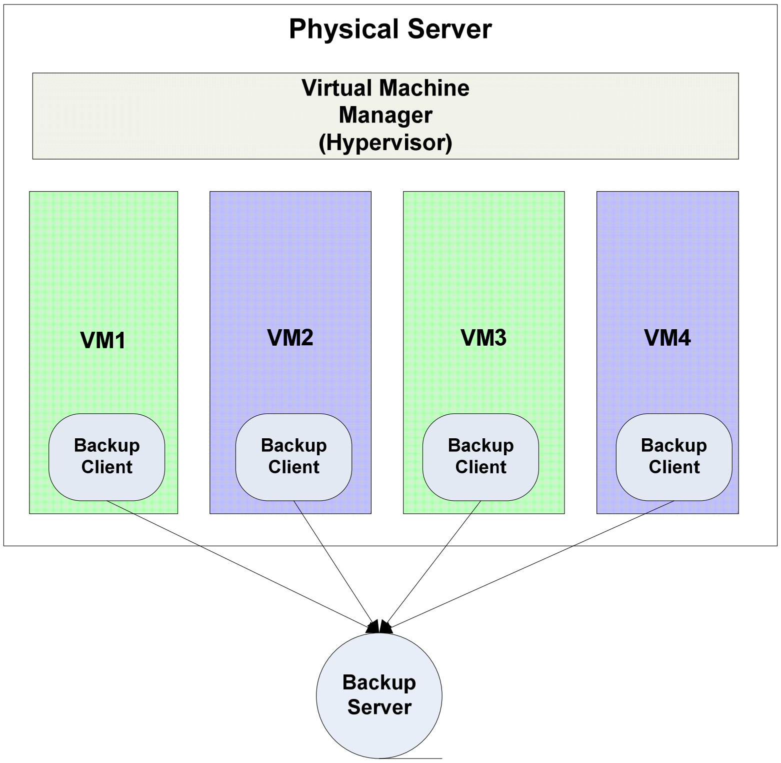

Figure 2.3: In a traditional backup model, each virtual machine runs its own backup client.

Virtual Machine Image Backups

A second approach is to install a backup client in the host machine and back up virtual machine images.

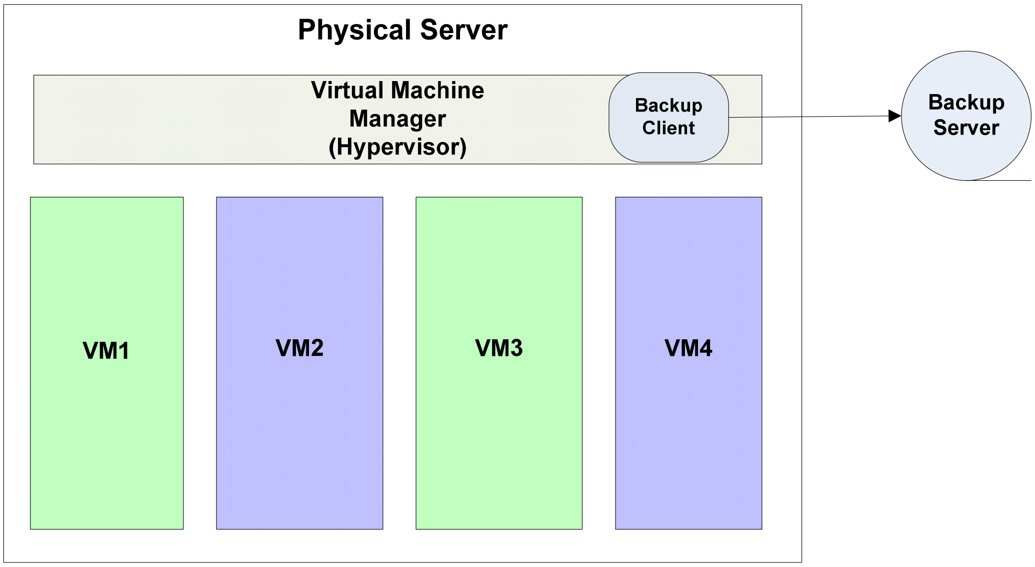

Figure 2.4: Backup clients can also run in the service console of the virtual machine manager to provide image backups to all virtual machines on the host.

Images need to be in a consistent state throughout the entire backup operation; otherwise, we could encounter a situation where one part of the image is backed up, there is a change to the state of the virtual machine, and the rest of the image is backed up. In this case, the first and second parts of the backup may be out of sync.

One way to avoid this problem is to perform image backups only when the virtual machine is shut down. For images with frequent down time—for example, a business intelligence reporting system that generates nightly management reports and is then shut down—this model is sufficient. For virtual machines with high-availability requirements, a snapshot-based backup is a better option. With this method, a virtual machine is kept in a consistent state for a brief period to time, just long enough to make a snapshot copy. The image copy is then backed up without adversely affecting the running instance of the virtual machine.

Mixed-Mode Backups

Sometimes a combination of file-level and image-level backup is the optimal combination to meet recovery management requirements. For example, a weekly image backup followed by daily incremental file backups, has advantages over just file-level or just image-level backups. The weekly image backup provides the ability to rapidly restore a fully functional virtual machine. It does, however, often require more storage space to keep the entire image when compared with incremental file backups. If a relatively small percentage of files in a virtual machine change each day, incremental backups will capture those changes without unnecessarily duplicating data that is unchanged since the weekly image backup.

Backup by Proxy

Proxy backup servers relieve servers by taking on the load imposed by backup operations. Snapshot images of virtual machines can be copied to a proxy server where they are backed up. An advantage of the proxy model is that it allows for centralized management of backups. This method can also alleviate some performance issues by performing snapshots during non-business hours and then creating backups from the snapshots during business hours. Also, jobs can be scheduled to accommodate the particular requirements of each virtual machine and systems administrators can use a single management console to monitor and administer backup operations.

> **Note**
>
> It should be noted that although virtual machine vendors provide backup applications, advanced performance and management features may only be available with third-party backup applications designed to support virtual machines.

Virtualization Vendor's APIs and Services in Enterprise Backup Strategies

Advances in virtualization backup progress on two fronts: with virtualization vendors and with backup application vendors. (It appears that the age-old economic principle on the specialization of labor applies to new technologies such as virtualization and backups as well.)

Benefits of Virtualization APIs

Virtualization vendors are improving the performance and functionality of their hypervisor platforms. In addition to improving speed and reliability, vendors are providing application programming interfaces (APIs) that allow third parties to programmatically access key functionality. Enterprise application vendors, including recovery management vendors, can take advantage of these APIs to extend the functionality of their products to include support for virtual machines. For example, recovery management vendors will provide more support for advanced management features than one would expect from a virtualization vendor more focused on implementation details of their hypervisor.

Virtualization vendor's APIs are a critical linchpin that enable third parties to provide the same kinds of recovery management features provided for physical servers as well as to offer specialized functionality needed only in virtualized environments. Consider, for example, the challenge of recovering individual files from a virtual machine image backup.

Virtualization APIs and File Restoration

Restoring a single file from a virtual machine image typically requires a number of steps:

1. Restoring the backup virtual machine image to a physical server
2. Restoring the file from the restored virtual machine to temporary storage
3. Copying the file to target location
4. Shutting down the instance of the virtual machine from which the file was restored

Ideally, we should be able to restore individual files from an image backup. Although the functionality is not typically part of virtualization vendor offerings, this type of advanced functionality can be incorporated by third-party providers if the appropriate APIs are made available.

Protecting virtual environments pose plenty of challenges with regards to performance, granularity of backup and restore operations, and manageability. There is no single best way to back up virtual environments. By combining different modes of backups (for example, file vs. image), taking advantage of virtualization-specific techniques, such as backup by proxy, and exploiting extended functionality enabled by virtualization APIs, businesses can choose from a variety of options to find the best backup model for their needs.

Meeting Specialized Backup and Recovery Requirements of Databases and Content Management Systems

In its most basic form, backup operations are about making copies of files. Files many of us work with on a day-to-day basis, such as word processing documents, spreadsheets, and presentations, are easy to back up. These kinds of files are self contained and, unless they are open for update operations, are not going to change during the backup process. Not surprisingly, this simple model of file usage begins to break down as we consider more complex applications:

- Email
- Databases
- Content management systems

Challenges begin to arise in applications such as these because they utilize multiple files with different functions and different rates of change and varying levels of dependencies between them. It is not just the underlying design or technology that makes backups of these systems more difficult; sometimes, it is the way we use these systems and the organizational requirements we impose on them that stymie simplistic backup models.

Email Backups

Although we still use the term email for applications such as Microsoft Exchange and Lotus Notes, the name does not capture the extent of collaboration and communication functions provided by these applications. Even a basic email system today is likely to include:

- Email services
- Calendar
- Contact management
- Task lists
- Notes

Backup applications must be able to capture these different types of data and restore them to a consistent state. Although a disaster recovery situation may require a complete restore of an email system, a more common task may be restoring a small set of deleted messages, users' folders, or other data structures. In such cases, the ability to rapidly identify and restore selected data is essential to meeting recovery time objectives (RTOs). Backup applications can take advantage of metadata about the structure of user messages and other data to provide email-specific functions, such as single user restore. The requirements for email backup are now much more than basic copy and restore operations.

Archiving and e-discovery have grown in importance as email has become more central to the communications of business. E-discovery is the process in civil litigation in which electronic information is reviewed for details of relevance to the case at hand. We only need to look to the now-famous case of Qualcomm v. Broadcom Corp. for the importance of e-discovery. In that case, Qualcomm was severely sanctioned (at a cost of more than $9.25 million) for failure to produce emails relevant to the case. The cost of backing up and archiving email can pale in comparison to the cost of insufficient e-discovery.

Key requirements for email backup with regards to e-discovery include the ability to:

- Create and maintain a comprehensive archive of all email messages
- Search messages for particular terms
- Search based on message metadata, such as date of messages, sender names, recipients, and so on

Email archiving programs may support these features natively in ways sufficient for small and midsize businesses. Large organizations may require specialized e-discovery software that imports email messages and other content into a content repository for specialized analysis and document classification.

Email and e-discovery demonstrate how organizational requirements can spur the development of advanced backup functionality. Databases are good examples of applications that provide plenty of technical drivers to backup innovation.

Database Backups

Databases are pervasive in today's software environment. Relational databases, in particular, have become the persistence mechanism of choice for many kinds of developers. A combination of features drives the adoption of relational databases:

- Ease of development with relational database
- A standard query language, SQL
- Broadly supported programming interfaces, such as ODBC and JDBC
- Reliability
- Scalability
- Flexibility

To realize these features, especially the last three, relational database systems depend on a complex storage management system that ultimately depends on low-level file system support. (Actually, in some cases, file systems can be bypassed in favor of "raw" disk access, but that is a much less common implementation option).

Databases Require Multiple Types of Files

The details of the storage management system will vary from one database management system to another, but the following list offers typical elements that affect how we perform backups:

- Files for storing data and indexes; they are sometimes called tablespaces
- Files for keeping a temporary copy of records as they are updated so that processes consistently read the same data, even if another processes is updating it; these files are sometimes called redo/transaction logs
- Files for auditing operations on data; these files may include username, times, client software information, and the IP address of the process making the changes
- Parameter files with configuration information for the database
- Error logs
- Temporary file space for sort operations

As with email systems, databases require specialized files to implement the full range of features and non-functional characteristics, such as scalability, that we have come to expect. Figure 2.5 shows a simplified version of a single update operation can result in changes to several types of files underlying the database system.

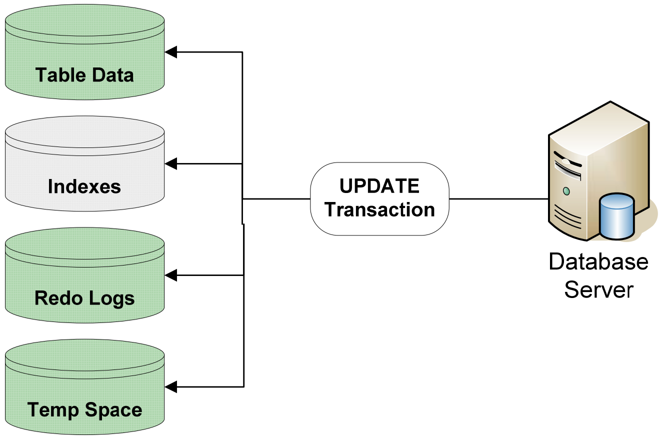

Figure 2.5: A single database transaction can update multiple files (green) as well as read from multiple files (blue). To capture a consistent state of the database, backups must be performed (a) when no changes are made to the database or (b) by a backup application that can track dependencies between the different database components and ensure a consistent database state is captured in the backup image.

The complexity of the underlying storage system creates difficulties for some types of backup and restores operations.

Restoring a Single Logical Record

Let’s consider how we would restore a customer record that was accidently deleted from a database. The customer record might include identifying information, such as name and address, purchase history, credit rating information, and account balance details. If all this information were stored in a single file, restoring it would be trivially easy. Relational databases efficiently manage large volumes of data, in part, by distributing the data in a single logical record across a number of tables. Names and addresses, for example, may be in one table, while purchase history is kept in several tables, including tables with order summary, order items, product codes, and other data. When a logical record is deleted, information may be deleted from a number of tables.

From a backup perspective, the critical question in logical record backup is, What rows from which tables need to be restored? The answer is highly application specific. Often, the answer can only be reliably determined by understanding the code that manipulates the database. Alternatively, rather than trying to restore a single logical record, one could restore to a particular point in time. The assumption here is that the database is always in a consistent state, so backing up to a point prior the time the logical record was deleted will allow you to restore to a consistent state in which the logical record is in place. A drawback is that any changes made since the restore point will be lost.

Online vs. Offline Backups

When databases are online and actively updated, there is the potential to make a backup copy of database files that are in inconsistent states. One method to avoid this is to take the database offline before performing the backup. This may not be practical when the database needs to be constantly available. Another approach is to export data using a database-specific utility that can capture a consistent state of the database using a combination of table data and redo log information. The export files can then be backed up using standard file-based methods. Content management systems, like databases, have application-specific backup and recovery issues we must take into account.

Content Management System Backups

Content management systems, such as Microsoft SharePoint and a variety of open source wikis, are increasingly used to manage unstructured data in analogous ways to which databases are used to manage structured content. (In fact, much content and configuration data is now stored in databases.) One of the advantages of content management systems is that they allow content owners to define groups of users with varying privileges to content. For example, a content creator may grant read and update privileges on a document to members of her department but only read access to other employees. Backing up content in these systems requires that we capture this type of access control information as well as the content itself.

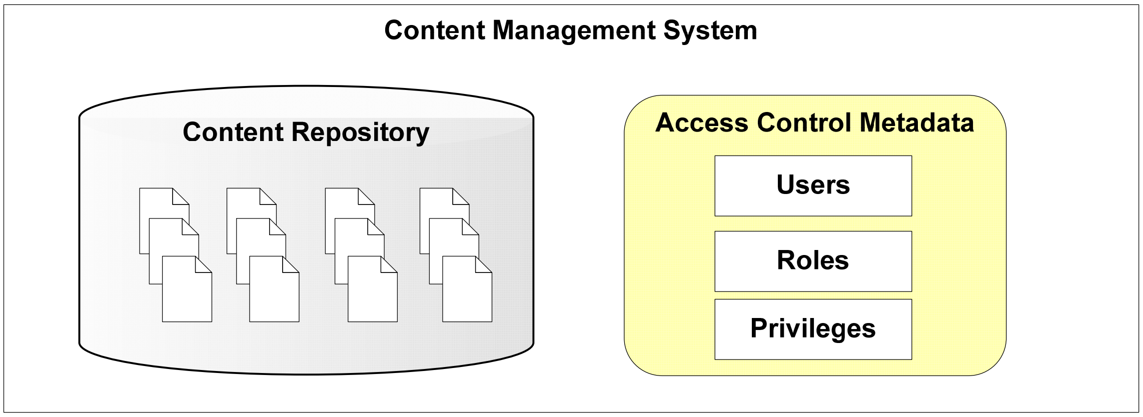

Figure 2.6: Content management systems consist of both unstructured data in content repositories and structured metadata, including access controls. Both types of data must be captured by backup programs and must be consistent with each other.

Here again, we have an example where a logical unit of business information is stored across multiple files within a complex application.

Options for Addressing Specialized Backup and Recovery Requirements

The options for dealing with application-specific backup and recovery requirements fall into three broad categories:

- Shutting down applications and backing up all files used by the application. This ensures a consistent copy of data but at the cost of system availability.
- Using a specialized application to backup application data or export data to a file or files that are then backed up using a general backup application. This approach requires additional storage space for the exported files(s) until they are copied to a backup device.
- Replicating data to a standby system. The standby system could be taken offline to allow for backup without interrupting availability of the primary system. This approach also provides for rapid recovery in the event the primary system fails. This feature requires additional hardware and possibly software licenses.

The best option will depend on a combination of factors, including cost and RTOs.

Solving the Remote Office Backup and Recovery Challenge

One of the challenges with managing IT infrastructure in remote offices is the lack of onsite technical support. Often, there is no justifiable case for having full-time IT staff in each remote office. At the same time, we cannot expect the non-IT staff in those offices to suddenly become an on-call reserve for taking care of IT operations. Remote offices should have the appropriate level of recovery management services as required by their business operations; this is the same standard that should be applied to central offices. The question is, How do we deliver those services without busting the IT budget?

Local Backup Option

One option is to maintain remote office resources for backing up and restoring data. This could entail having a backup server in each remote office as well as sufficient disk space for all backups. For disaster recovery purposes, replicated data will have to be maintained on some type of off-site storage. In cases where there are few remote sites or sufficient network bandwidth is not available, this may be a reasonable option. As the number of remote sites grows, the economies of scale inherent in backup infrastructure come into play and a centralized approach would be more economical.

Centralized Backup Option

With a centralized backup system, data is copied from remote offices to a central location where it is backed up and stored. Centralized backups have a number of advantages over local backup options. First, the marginal cost of adding sites is low compared with a local backup option. Additional licenses for backup agents are required and sufficient bandwidth must be in place between the remote office and the central location. Additional storage and backup hardware is not required onsite. Second, by consolidating backup services, the central site can share resources among multiple offices. The chance of having unused storage capacity is reduced. Servers are more likely to be utilized because they can manage backups for multiple sites. Finally, systems administrators are on site and can quickly respond to hardware failures, networking problems, or other issues that require a knowledgeable person on site to correct.

Essential Features of Remote Office Backup Solution

When evaluating remote office backup options, look for features that increase manageability and limit demands on network resources:

- Centralized management console that allows administrators to schedule backup jobs, examine logs, receive alerts, and generate reports on backup operations
- Ability to back up remote servers efficiently through the use of deduplication and other methods to minimize demand on bandwidth
- Allow for centralized control of installation and updates to backup agents running on remote servers
- Ability to back up to either a central facility or to local storage at the remote site

Also, be sure to consider how the options under consideration affect your ability to meet RTOs and recovery point objectives (RPOs). Depending on requirements, you may have very short RTOs that demand off-site backup services.

Ensuring Continuity in the Event of a Disaster

An essential part of recovery management is preparing for disaster—that is, the loss of compute, storage, and network services which prevents the ability to deliver essential IT services. Disaster can be isolated to a single business, such as a fire that destroys a data center; regional, such as hurricane or earthquake damage; or widespread, like the Northeast Blackout of 2003 that caused widespread power failure throughout the US northeast and parts of Canada. When disaster disrupts IT services, by what means can they be restored?

Backups can be used to restore data and applications assuming servers and network devices are in place. Reconfiguring servers, installing OSs, and restoring backups can be a time-consuming operation subject to human, and technical, error. An alternative to waiting until there is a need for disaster recovery infrastructure is to maintain standby servers and keep them up to date with a continuous data protection process.

The Need for Replication

Replication services are used to keep standby servers up to date with primary servers. Consider an example of how these systems work: An order fulfillment database is continually updated during the day as new orders arrive. Orders are processed, customer credit is verified, inventory is checked, and shipments are readied from this system. The system is capturing business-critical transactions, so each time the database is updated, a copy of the update is sent to a standby server located in a separate data center. The standby server is running the order fulfillment system as well and is ready to take over for the primary server should it fail.

Replication systems can reduce their impact on production systems by minimizing additional computation or I/O on the production server. For example, rather than implementing a custom procedure to copy every transaction as it is executed in the application, low-level I/O operations can be duplicated instead. There is no need to execute full extent of the programming logic required to calculate the final results; duplicating the results is sufficient.

Another consideration with regards to performance is the demand on LAN and WAN resources. Replication technologies that minimize the amount of data replicated between primary and standby servers reduce overhead on the network.

Even with the additional overhead on production systems, the benefits of replication and continuous data protection can outweigh the costs. Replication reduces the time to recover by eliminating the time required to restore data from backup storage. By configuring the standby server prior to a disaster, there is more time to correct errors in configuration and diagnose other problems that may arise when setting up the standby server. The last thing any systems administrator wants when recovering from a disaster is debugging an unanticipated configuration error.

Replication also supports stringent RPOs. Replication processes can be configured to commit changes to the standby server on a frequent basis. This reduces the amount of data lost when the primary system fails. Only the data generated since the last update to the standby server would be lost, and that is typically far less data than the amount generated since the previous night's backup.

Replication Failover Options

With data replicated to standby servers, systems administrators have a number of options for failing over to the standby server. High-availability options for failover include:

- Server monitoring with a high-availability application—If a failure is detected, the system automatically sends traffic to the standby server.
- Manually redirecting traffic to backup servers—This can be done by systems administrators, for example, by updating local domain name services entries to map server domain names to the secondary server.
- Another option is to replicate data only and not have a standby server in place. This option reduces the delay in recovering data but does not tie up a standby server for a specialized purpose. This option could be used, for example, if a virtual machine image needs to be started to act as a standby server.

A general rule of thumb with failover options is that the faster and more automated the failover, the more complex and costly the solution. Such guidelines have to be considered with respect to business requirements. When rapid recovery is needed, high-availability options with continuous data protection are a sound option.

Summary

If recovery management were just a matter of backing up and restoring files, our professional lives would be much simpler. Complex IT systems have complex recovery requirements. Virtual machines are a boon to improving server utilization, but they introduce several challenges for backup and recovery operations, especially with regards to capturing a consistent state of the virtual machine without adversely affecting availability. Applications, as diverse as email, databases, and content management systems, make efficient use of file systems but in ways that challenge simple backup strategies. Remote offices need recovery management protection but cannot afford onsite, dedicated IT staff. Disaster recovery and the need for constant availability are driving the adoption of high-availability solutions. In all of these cases, we need to adapt recovery management practices and backup applications, including replication services, to these challenges. As the examples in this chapter show, a combination of recovery management practices and the most advanced backup and replication applications can be combined to meet these demanding challenges.

Chapter 3: Top-5 Operational Challenges in Recovery Management and How to Solve Them

Maintaining effective recovery management procedures is not a trivial task. From making sure processes are running correctly to controlling the costs of operations, there is no shortage of challenges. In this chapter of the *Shortcut Guide to Availability, Continuity, and Disaster Recovery,* we will examine five of the top operational challenges we commonly face in recovery management:

- Scheduling and monitoring
- Choosing the right storage media option
- Controlling the costs of off-site storage
- Keeping up with growing data volumes
- Recovering when disaster strikes

These five challenges are interrelated. For example, dealing with growing data volumes is directly related to controlling costs. Monitoring is less directly related to recovery operations but is just as important—we do not want to find out about a failed backup operation when we try to restore critical data after a disaster. As we address each of these five challenges, we will consider both the fundamentals of the individual challenges as well as how the challenges influence each other.

Challenge 1: Scheduling and Monitoring

Scheduling and monitoring go hand in hand. It is the combination of deciding what types of backups are required and making sure they are performed correctly. As with most IT operations, recovery management is driven by business requirements, so in the course of this discussion, we will have to refer back to those requirements when deciding on the proper schedule of backup types.

Scheduling: Various Types of Backups and Their Uses

In Chapter 1, we looked at the business case for management recovery. That discussion included some obvious requirements, such as restoring from isolated failures and compliance, and some not-so-obvious requirements, including varying backup policies based on the business value of data. Meeting these requirements with operational procedures entails implementing the right combination of backup types at the right times.

Let's start with a quick review of the various types of backups, then consider how we combine them to achieve the level of data protection we need.

Types of Backups

There are a few different types of backups because we need to balance competing requirements in recovery management. Ideally, we would have comprehensive coverage of all data at all points in time and we would be able to restore any or all of that data rapidly. Add to that minimizing cost and resources required to perform backups, and we would have the ideal solution. We will not be getting our ideal solution anytime soon, so we will have to settle for the optimal realizable solution.

Pragmatic recovery management solutions build on three types of backups to find that optimal solution:

- Full backups
- Incremental backups
- Differential backups

These methods have different advantages and disadvantages and so tend to complement each other as part of a recovery management strategy.

Full Backups

Full backups, as the name implies, make a complete copy of data that might later need to be restored. It has a number of advantages. A full backup is a completely self-contained backup, so a complete restore operation can be performed using a single backup. A key disadvantage of full backups is the time and space required to create and store them. The size of full backups is proportional to the amount of data to be protected. (The size is proportional to, not equal to, the size of the source data because of compression).

Incremental Backups

Incremental backups reduce the time and space required for full backups by copying only data that has changed since the previous backup. Consider a simplified example to see how significant the storage reduction can be.

If we assume that only 10% of data changes, then the amount of storage required for backup can be significantly reduced. For example, assume we have to backup about 5TB of data. Assuming a 30% compression rate by the backup software and a 10% rate of changes, an incremental backup of 5TB of data can be stored with as little as 350GB of storage (see Table 3.1).

Total Data Size (GB)	Full Backup Size (GB) (30% Compression)	Changed Data Size (GB) (10% Rate of Change)	Incremental Backup Size (GB) (30% Compression)
500	350	50	35
1000	700	100	70
2000	1400	200	140
5000	3500	500	350

Table 3.1: Incremental backups yield significant savings in storage when compared with full backups.

Clearly the savings in storage is substantial; however, in IT as in economics, there are no free lunches. What we gain in reduced storage costs and time to perform backups is accompanied by a disadvantage during restore operations.

In its simplest form, restoring from incremental backups requires restoring from a full backup and then restoring each incremental backup performed since the last full backup. Depending on backup software, it is possible to create a synthetic backup, which is a backup that results from merging a full backup and some number of incremental backups. The advantage of synthetic backups is that they combine the advantages of both full and incremental backups. As Figure 3.1 depicts, a synthetic backup is a full backup plus all changes since the backup was made.

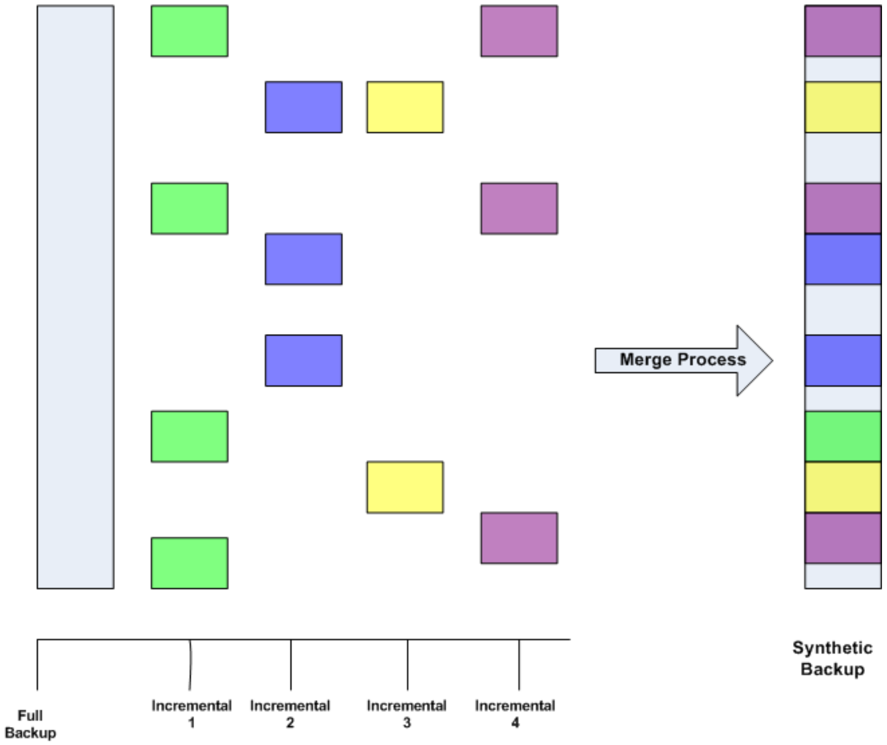

Figure 3.1: Synthetic backups merge full backups and incremental backups to allow for more efficient restore operations.

Differential Backups

An alternative method to incremental backups is the differential backup. Like incremental backups, differential backups start with a full backup and then back up only changes. The difference is that the differential backup captures all changes since the last full backup, not since the last backup whether it was full or incremental. An advantage of differential backup is that it only requires the full backup plus the latest differential backup to perform a complete restore. A disadvantage, relative to incremental backups, is that additional storage is needed for the differential backup.

Given the various types of backups, what is the best combination of these backup methods to provide adequate data protection while controlling costs?

Backup Scheduling

There a number of ways to schedule backups that are based on the idea of a rotation. The idea started when tapes were virtually the only option for most businesses, so we sometimes hear these methods referred to as “tape rotation” schemes. Rotation schemes balance the desire to re-use backup storage while maintaining a reasonable number of recovery points.

One of the simplest schemes is the round-robin method. A business has a fixed amount of backup storage, either tapes or disk, but we will describe the process in terms of tapes for simplicity. Under the round-robin scheme, a tape is used for each backup and once all tapes are used, the tape containing the earliest backup is re-used. A small business, for example, may be satisfied with using five tapes and performing a full backup each night of the business week. Every Monday, the previous Monday’s tape is overwritten, and so on for each day. This method is simple to manage but leaves the business without a recovery point earlier than the prior week. This method makes it difficult to maintain off-site backups while preserving the ability to rapidly recover in case of an isolated failure.

A common rotation scheme is based on a combination of monthly, weekly, and daily backups. This is sometimes referred to as the grandfather-father-son (GFS) method. The basic idea is that a full backup is done each week (the father) and differential backups are performed each day (the son). At some point in the month, the latest father tape is promoted to “grandfather” status and removed from the rotation. This backup can be stored for long periods in off-site storage. The remaining weekly backups are maintained, typically on-site, and overwritten after 1 month. The daily backups are overwritten on a weekly basis. There are a number of variations on the GFS scheme, but the key points are that it supports long-term off-site backups as well as weekly and daily backups.

The GFS scheme provides a good balance of efficient use of tapes or disk with the ability to maintain multiple recovery points.

Monitoring

Monitoring backups is a two-part process. We want to monitor backup operations to ensure they execute as expected and we want to verify that the backups performed are valid and usable.

Monitoring Backup Operations

As the number of servers that requires backup grows, so does the complexity of monitoring and managing backup operations. The complexity stems from a number of factors:

- Not all data is equally valuable to the business. Different types of data require different levels of protection, and that means different backup schedules and retention policies.
- As the number of servers and data volumes grows, so does the likelihood of a failure somewhere in the overall operation. For example, if a failure is likely to occur 1 in 1000 backup operations, and we run 10 backup operations at a time, the chances of having a failure somewhere in the enterprise is 1 in 100.
- Data is geographically dispersed. Small remote offices will require data protection services but will likely not have a dedicated IT staff on-site placing additional demands on central IT professionals.
- There are simply more devices and processes to manage as data volumes and the number of servers grows.

As a baseline, backup administers should capture and monitor several metrics for all backup operations:

- The start and end time of backup operations—Duration information can help identify trends in changes to the amount of time required to perform the backup.
- The amount of storage used for backups—As with time, this can be used to detect long-term trends, but it can also help optimize storage use in the short term. For example, backups with modest storage requirements may be combined to make more efficient use of storage devices.
- Performance metrics on backup servers, particularly memory use and CPU utilization—Although backup and restore operations are obviously I/O intensive, compression and encryption can put significant load on the CPUs of the backup server.

In addition to those metrics, it can be important to track errors and failures. Significant errors, such as a failed backup, should be addressed as soon as possible in most situations. Minor failures, such as the detection of a bad block on a disk, should be monitored over time in case the failure is not an isolated incident but part of a larger systemic problem.

Verifying Backups

When you need to restore a file from a backup, it is the wrong time to find out the backup is corrupted. There are few ways to verify backups, although details will vary with backup software.

One way to verify backup is to have the backup software perform a verification pass immediately following the backup operation. This operation compares the contents of the original files with the files on the backup medium. The ability to perform this type of verification will depend on your backup software. It will also extend the time required to perform a backup, which may not be an option if you are already working within a tight window of time.

Compression features may also provide the ability to check the integrity of the compressed files without performing a block-by-block comparison. This could be faster than other full comparison operations but again depends on the features of your backup software.

A tried-and-true method is to randomly select a backup, restore it, and compare it with the original file. This is helpful but does have some drawbacks. First, unless a backup administrator is well versed in statistics and probability, she might not adequately sample the set of backups at least from a formal quality control perspective. Second, data is constantly changing, so the original files may not be available to compare against. (After all, one of the reasons we make backups is to be able to restore a previous state). Even with these limitations, it is useful to perform some level of verification using features offered by backup software and to randomly select backups for manual verification.

Scheduling the right combination of full, incremental, and differential backups helps to ensure a business will be able to recover to a number of points in the case of isolated or catastrophic failure. The verification process is an extra but worthwhile step to help ensure the integrity of backups.

Costs are always a factor that must be taken into consideration when managing risks, and backups are no different. An optimal backup schedule will balance costs with the ability to recover to different points in time. Verification processes may be less than ideal because of other business constraints, including costs. Nonetheless, proper scheduling and monitoring of backup operations contributes to data protection and overall risk mitigation related to data loss.

Challenge 2: Choosing the Right Storage Media Options

Magnetic tapes have a long history in information technology, and they have proven themselves a reliable and cost-effective storage media. As volumes of data have grown along with the seeming never-ending desire for faster backup and restore operations, the information technology industry adopted disk storage as an alternative. There are advantages and disadvantages to both, so choosing between them is not a trivial task. The key to finding the right choice, or the right combination of tape and disk, will depend on a sound understanding of operational requirements, budget constraints, and your business' data protection strategy.

Tapes: Advantages and Disadvantages

Tape backups have a number of features that make them a clear choice for business backups:

- Tapes are cost effective—The industry Linear Tape Open (LTO) standard is widely adopted, so businesses have a non-proprietary option with LTO, which often translates into competitive pricing. (LTO products often use the term Ultrium as well). The cost per gigabyte of storage is generally lower for tapes than for other media, including disks.
- Tapes are easy to physically transport—This is especially important with regards to archiving and offsite storage.
- Tapes have a long history as the backup medium of choice, so tapes are compatible with many backup applications.

Now, let's consider for the disadvantages:

- Tapes use a sequential access method, which results in relatively slow restore operations when compared with disk-based restores.
- Tapes are more vulnerable to environmental factors, such as heat and humidity, than disks. Environmental factors can adversely affect the durability and reliability of tapes.
- Tapes are routinely physically handled, which opens opportunities for human error, such as lost tapes, and unintentional damage.

The fact that tapes are easy to handle is an advantage as well as a disadvantage shows the kinds of challenges and lack of clear-cut choices we have with backup media.

Often, the advantages of tapes outweigh their drawbacks. Advances in disk technology created a viable alternative to tapes.

Disks: Advantages and Disadvantages

Disk storage has a number of advantages for data protection. Some of these, not surprisingly, correlate with disadvantages of tapes. The advantages of disks as a backup media include:

- Disk performance is faster than tape—This is due to both the random access nature of disks versus the sequential access pattern of tapes and to the speed with which data can be written to disks. Disk drives used for backups often use fibre channels, which have higher throughput than tape drives.
- Disks are more protected from environmental factors than tapes, which are typically moved from tape drives to short- and long-term storage locations.
- Disk capacity can grow relatively easily when using a storage area networks (SAN).

Disk storage does have its downsides:

- Disk are generally more expensive than tapes when comparing the cost of storing a gigabyte of data
- Disks may not be suitable for archiving, in which case, mobility of the storage media is especially important

So how should one balance the pros and cons of both options to come up with the optimal mix?

Identifying the Best Option for Your Business

The best option for your business is determined by balancing often-competing requirements. For example, maintaining a low-cost solution is an obvious concern, but the total cost of data protection includes more than the cost of tapes and a tape drive. If there are significant opportunity costs associated with long recovery times, backing up to disk may actually cost the business less in the long run than using tapes. However, backing up archived user directories to tape may provide the required data protection at a lower cost than using disks because rapid restoration is not essential.

Step 1: Classifying Data by Protection Requirements

The first step to identifying the best option for your business is to classify data according to the level of protection required for each type of business data. In this case, the level of protection includes:

- Recovery point objectives (RPOs)
- Recovery time objectives (RTOs)
- Archiving requirements
- Other compliance-related requirements

Some data will require near-time recovery points and rapid recovery, such as an order management system. For these data, the additional cost of disks can be justified. For other types of data, the reliability of the storage media may be especially important. Although email may appear at first glance to be a moderate concern, that can easily change in the event of litigation. If e-discovery is required in response to litigation, a business should be able to produce required documents from online systems or archives.

> **Resource**
> For more information about e-discovery, see the Electronic Discovery Reference Model at http://edrm.net/.

Also, consider the need for data life cycle management. The amount of data stored cannot grow forever without cost; at some point, the value of retaining archives will be outweighed by the cost of maintaining those archives. The difficult process is determining when one reaches that point.

Step 2: Mapping Recovery Objectives to Backup Options

For each category of data, determine the volume of data that must be processed in a given time period to meet RTOs. Also, determine the amount of storage required to meet RPOs. These two pieces of information are somewhat dependent on each other. For example, the amount of data that needs to be processed to meet recovery point and time objectives will depend on which type of backup is used. You might want to consider a few different scenarios with different combinations of full, incremental, and differential backups to understand the trade-offs between time to backup, time to restore, and the volume of storage needed for backups.

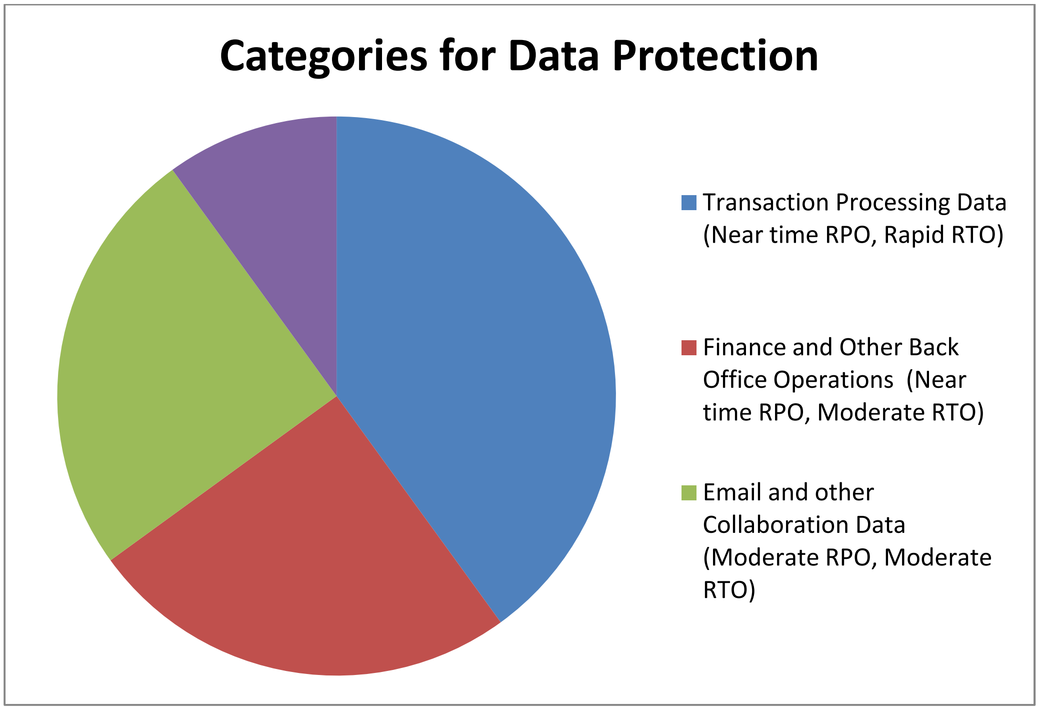

Figure 3.2: Data should be categorized according to recovery and archiving objectives to facilitate determining the optimal use of disk and tape storage.

Step 3: Determine the Most Cost-Effective Ways to Meet Recovery Objectives

The last step in the process is selecting the most cost-effective storage option that meets your requirements. If an option does not meet the needed recovery point or time objectives, the question of cost is irrelevant.

Be sure to consider a combination of tape and disk options to meet all requirements. For example, it may be best to use disk-to-disk backup to meet recovery objectives and then use tapes for archiving. When the archives are created from the backups, the process is called "disk to disk to tape." One of the advantages of this approach is that the time to backup up from a production system is reduced, but you still retain the long-term cost advantage of using tapes. In addition to cost advantages, this method can reduce the load on servers being backed up by allowing archives to be made from a copy of production data. Another cost consideration with regards to archives and disaster recovery is the cost of offsite storage.

Challenge 3: Controlling the Costs of Offsite Storage

Offsite storage is an essential part of comprehensive data protection strategies. It provides a level of protection in the event of a disaster at a primary facility. Businesses will vary in their use of offsite storage and the types of facilities used. For example, for a small business, offsite could be a safe deposit box in a local bank. Larger businesses might use different offices to store each other's backup tapes or they may use a third-party service provider. Regardless of what form of offsite storage is used, there are a number of considerations to controlling the cost associated with this part of data protection.

Time to Move Data to Offsite Storage

The time required to move data to offsite storage, and the associated cost, will depend on the type of storage media. Moving data to offsite disk storage, such as a backup service that hosts disk arrays for clients, will have relatively low labor costs. Depending on the volume of data transmitted, there may, however, be a marginal cost for necessary network bandwidth. The time required to move the data will depend on the volume of data and the network speed. Here again, RPOs and RTOs will help determine the necessary capacity.

The cost of moving tapes to offsite storage will be typically dominated by labor costs. Small businesses may be able to use informal practices, such as having an employee leave the office early to drop off tapes at another office on the way home at the end of the day. Other businesses will require more established procedures using staff or third-party providers to transport tapes.

Risks Associated with Offsite Tape Rotation

Any time tapes are moved offsite, there is a risk of losing or damaging the tapes. A lost tape could be a minor inconvenience or a significant problem, depending what is on the tape. If the tape contains sensitive information, such as personally identifying information, protected health care data, or financial information, it should be encrypted to prevent unauthorized disclosure. Consider compliance requirements when transporting tapes offsite to ensure information is adequately protected when it leaves the business.

Cost of Offsite Storage

The cost of offsite storage will vary with the volume of tapes stored offsite, the length of time they are stored, and possibly the number of times the storage facility is accessed. In addition, the quality of the storage facility with regard to data protection will influence cost. Renting a self-service storage locker from a local vendor might be the least-expensive option possible, but we can forget about any type of environmental controls, fire suppression equipment, or other necessary controls. At the other end of the spectrum, storing tapes inside a mountain may protect your tapes as well as one can expect, but the costs for such a service may only be justified for the most sensitive corporate information.

Cloud Storage as an Offsite Storage Option

Cloud storage is an emerging option for offsite storage. Public cloud providers typically charge by the volume of data stored and the length of time the data is stored. This type of "pay as you go" model may help reduce the need for additional hardware as the volume of data grows. As with other third-party providers, we must consider the long-term viability of the cloud storage provider along with the reliability of its service. Also, consider using strong encryption when storing confidential, sensitive, and private data in the cloud. Remember that the definition of strong encryption changes with time. Encryption algorithms and key lengths that were once consider sufficient for protecting information can be compromised using today's hardware and cryptanalysis techniques. The cost of offsite storage depends on a number of factors, including the volume of data we store, what type of media is used, the level of service provided by the storage facility, and how the data is transported to the facility.

Challenge 4: Keeping Up with Growing Volumes of Data

Businesses are faced with growing volumes of data. When we look into the source of this phenomenon, we find there is no single source of additional data; instead, it is driven by a wide range of business initiatives and requirements:

- Additional applications designed to take advantage of new opportunities that generate additional data
- Increasing use of collaboration tools, including email, messaging, and document repositories
- Compliance requirements that specify data retention requirements
- Improved analytics that drive the motivation to collect more data in order to more efficiently and effectively target key markets and customer segments
- The ease with which we can all download documents, presentations, videos, and other business-related content from the Internet

Much of this data will fall under the umbrella of enterprise data protection strategies, which means more data to back up. It does not necessarily mean more time to back it up or longer recovery times or even more budget to cover the cost of additional hardware and storage media. About the only option left is to improve backup technology. A significant advance in this arena has been in the area of data deduplication.

Deduplication Options

Deduplication takes advantage of the fact that the contents of data blocks on disk are often duplicated across multiple data blocks. Duplication of data creeps into even the most storage-conscious organizations. In part, this is due to the fact that it is often easier to produce and manage information if we duplicate information. Consider a few examples: Multiple employees save the same copy of a report. Software developers save a number of versions of similar application code. Sales reports with common corporate and department data are sent to sales staff along with their personalized reports. This type of duplication creates opportunities for more efficient storage on backup media.

Deduplication works at the data block level. During a backup operation, a data block is read and compared with all other data blocks that have been read before it in the same backup operation. This step is faster than it may sound at first. Instead of explicitly comparing blocks with each other, a value (known as the value of a hash function) is calculated for each block. If a prior block has the same value, the backup program only needs to store a pointer to the previous block instead of storing another copy of the block. This can result in significant savings in storage.

There are two of options when performing the deduplication process, both of which have advantages and disadvantages:

- Source-side deduplication, in which the deduplication process runs on the server hosting the data being backed up
- Target-side deduplication, in which the deduplication process runs on the backup server

With source-side deduplication, the network traffic is reduced because duplicate blocks are not sent to the backup servers—only references to an already copied block are sent. A potential drawback is that the source server's performance is adversely affected by the additional load. This can be mitigated by scheduling backups during non-peak demand periods. However, if the source server is a virtual machine, you should consider the load on the physical server by other virtual machines during backup. When backing up virtual servers, be sure to consider the particular needs of virtual environments.

Cross Reference

See Chapter 2 for more information on this topic.

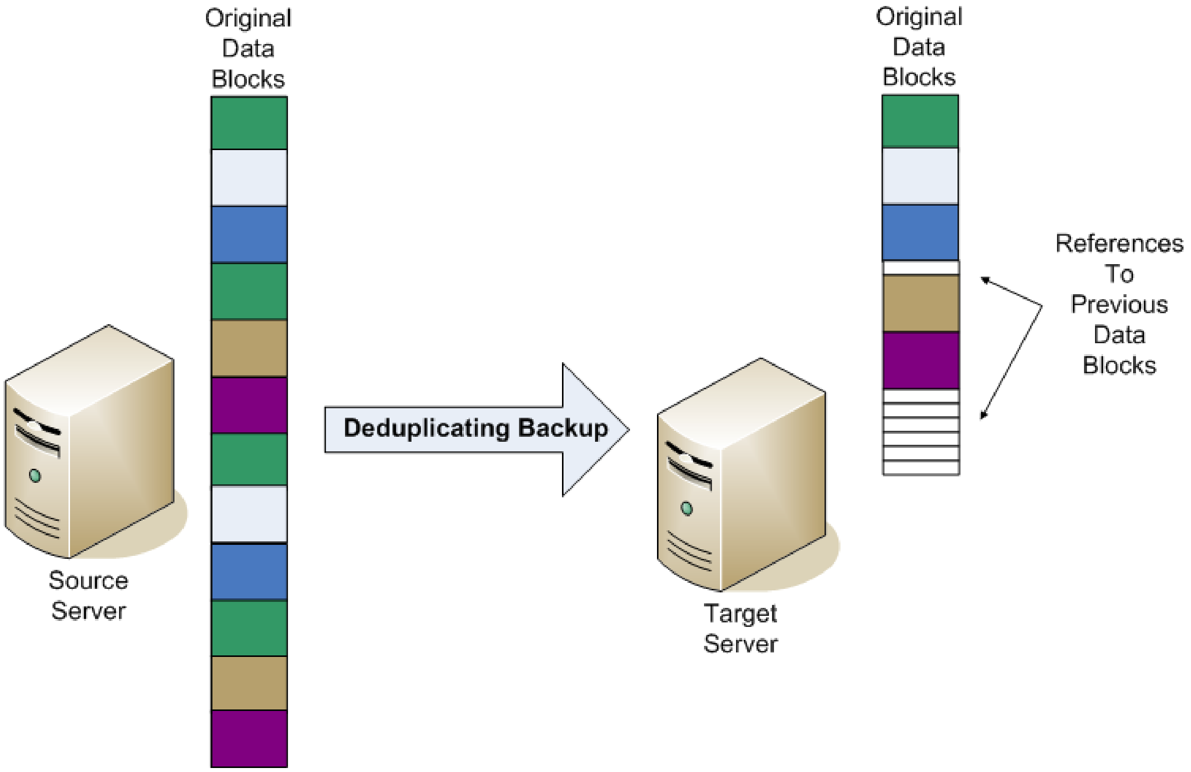

Figure 3.3: Deduplication reduces the number of data blocks stored in backups by substituting pointers to previously backed up data blocks.

Target-side deduplication performs the deduplication operations on the backup server. This approach can offer greater throughput in highly distributed environments with dedicated backup servers. A tradeoff is that network traffic is not reduced because duplicated blocks are copied from the source system to the backup server.

Deduplication is a key technique for accommodating growing data volumes in data protection strategies. With deduplication, existing backup infrastructure can keep up with growing data volumes without requiring an equivalent increase in storage media.

Challenge 5: Recovering When Disaster Strikes: Continuity and Failover

The last of the five key operational challenges of recovery management we will examine is disaster recovery. A business' disaster recovery plan answers the question "How would we keep the business running if there was a catastrophic loss at a key business site?" It is important to understand how this differs from other reasons we perform backups. If someone accidently deletes an important file, someone in tech support can restore the file from a recent backup. The staff is in place, the backups are available, the backup applications and servers are up and running, and, perhaps most importantly, it is an isolated incident. It may be an important file, but chances are its loss would not adversely affect ongoing operations. Disaster scenarios are different.

Disaster recovery plans are implemented when normal business operations cannot continue as normal. This can be caused by a range of factors:

- A catastrophic physical loss, such as a fire or natural disaster that destroys buildings housing the business
- An extended loss of power, including exhausting backup power supplies
- Regional storms that prevent employees from reaching offices and data centers

In cases such as these, normal backup and restore procedures are not enough and the chances of maintaining normal business operations will depend heavily on how well the business planned prior to the disaster. That planning should take into account both physical requirements and application requirements.

Physical Requirements

The physical requirements for disaster recovery include both information technology infrastructure and a physical space to house equipment and employees. At minimum, this includes:

- Backup servers for running essential applications—Business processes should be prioritized to identify which services should be restored first and which can wait. Also, consider the level of performance required in a disaster situation. If a high-priority process cannot tolerate degraded performance, a backup server equivalent to the primary server should be maintained. In other cases where a lower level of performance is acceptable, for example with an email system or a management reporting system, the applications could be run on virtual servers hosted on a small number of physical servers that are shared with other applications.
- Sufficient storage to restore essential data to continue operations and to accommodate new data generated while in disaster recovery mode—Once again, it is essential to prioritize. Not all business data is equally valuable. For example archived data from department-level data warehouses or operational data stores is probably not necessary until normal operations are fully restored. Rather than incur the cost of maintaining disaster recovery storage for all possible business data, maintain only as much as is needed for necessary operations.
- Office and data center space—This includes physical space for employees and equipment as well as critical infrastructure, especially power.

Proper planning will help identify essential business operations, levels of service required, and the minimal amount of physical space required in a disaster situation. This planning in turn will help control the cost of maintaining the physical requirements for disaster recovery.

Application Requirements

With the physical requirements attended to, the next step is planning on ensuring data and applications are up to date and ready to carry on business operations in the event of disaster. A basic question that must be answered is, How long can operations be down before there are significant, adverse consequences? Can your business, for example, continue to operate if data on backup tapes has to be restored to standby servers before the disaster recovery site becomes operational? Can the business recreate data that was created after the last offsite backup and before the disaster struck? If the answer is no to either question, replication and high-availability services should be considered.

Replication systems are used to copy data from primary servers to standby servers on an ongoing basis. Replication systems can efficiently capture changes to disks, copy the data to standby servers, and ensure data is duplicated offsite in a reasonably short period of time. This approach does require continuously operating hardware at both the primary and disaster recovery sites as well as sufficient bandwidth to keep the standby server up to date.

Replication services maintain the data but do not control the process of switching operations from the primary to the standby servers. High-availability systems perform this function. If rapid failover is needed, high-availability systems can be configured to monitor primary servers and automatically switch to the standby servers as soon as a failure is detected. As a general rule, the more speed and automation required in the failover process, the greater the cost. In situations where disaster recovery budgets are constrained and slower failovers are tolerable, a manual failover process can be used.

Disaster Recovery Service Providers

An alternative to maintaining a dedicated disaster recovery site is to use a third-party disaster recovery service. The business model of such operations is based on the idea of pooled risk. It is not likely that all customers would need to use standby servers at the same time, so they can more efficiently provide disaster recovery services than a typical business.

Maintaining continuous business operations in the event of disaster is one of the most difficult challenges in recovery management. Planning is a key to success. By prioritizing business operations and identifying operations that can tolerate degraded performance, business can find a balance between maintaining services and controlling the cost of disaster recovery services.

Summary

Recovery management entails a number of technical challenges. Scheduling and monitoring backups, choosing the right storage media, controlling the cost of offsite storage, keeping up with growing data volumes, and planning for disaster recovery all present technical issues. Understanding business drivers behind recovery management is necessary to properly prioritize business operations which in turn is essential to making the appropriate choices when confronting each of these challenges.

Chapter 4: Putting It All Together—Recovery Management Scenarios for Small Businesses to Emerging Enterprises

Throughout *The Shortcut Guide to Availability, Continuity, and Disaster Recovery,* we have explored how to address the business and technical requirements of data protection. Some of the requirements are obvious and apply to all organizations: restoring from isolated failures, for example, accidently deleting a file, and recovering from catastrophic failure, such as a natural disaster that destroys a data center. There are also less obvious technical and business needs. For example, server virtualization is widely adopted for its ability to improve server utilization and help control costs, but it introduces additional technical challenges with regards to backup and recovery. Business strategies can also influence recovery management objectives. A move to improve customer service by providing longer periods of access to online data directly affects the cost and required resources of recovery services.

These and other considerations have been woven into both the business strategy discussions and the technical assessments documented in earlier chapters. In this chapter, we take a different approach and consolidate key recovery management issues according to business types and the special case of failover recovery. We will consider five scenarios. Each scenario delves into typical business and technical issues faced by particular types of businesses or technology use cases; in particular, we will consider:

- Small business backup and recovery
- Midsize business and remote office protection
- Operational management and enterprise backup
- Backup and recovery with virtual machines
- Continuity and failover recovery

These scenarios are not mutually exclusive. Some of the discussion of small business backup and recovery services may be relevant to midsize businesses, especially those with remote offices. Similarly, virtual machine recovery management may be relevant to all types of businesses, regardless of size. Continuity and failover is such an important topic that we address it separately, although we will touch on failover in other sections when relevant. We will conclude this guide with a summary of best practices in availability, continuity, and disaster recovery.

Different Business Requirements Drive Different Solutions

When it comes to recovery management, one size does not fit all. Business requirements will vary by industry and company size. Consider how different the needs may be in the following examples that highlight different industries:

- A financial services company may need 24×7 availability of recent transactional data as well as more historical reports. If systems are down or data is unavailable, basic operations could come to a standstill. A credit union customer could not walk into her branch office and make a deposit, for example. When core systems are down in financial services, you are essentially closed for business.
- A manufacturing company that loses access to its inventory management system may be able to continue to function at a lower level of productivity. Using a combination of phone calls to other parts of the plant and having runners check and update paper backup systems, the company can keep some level of operations in place for a short period of time.
- A real estate management firm might be able to operate at somewhat normal levels for a day or even two if their primary management systems are down. Office phones, personal smart phones, and a stop at the closest coffee shop for WiFi access will keep real estate professionals in the loop, at least with clients and colleagues.

These examples show the range of needs with regard to recovery management. Perhaps the most telling aspect of these scenarios is the one common theme: *ad hoc* solutions may work for some period of time but sooner or later core systems must be recovered. Recovery management is not an option; it is a requirement in business. What type of recovery management solution you implement will vary according to your needs.

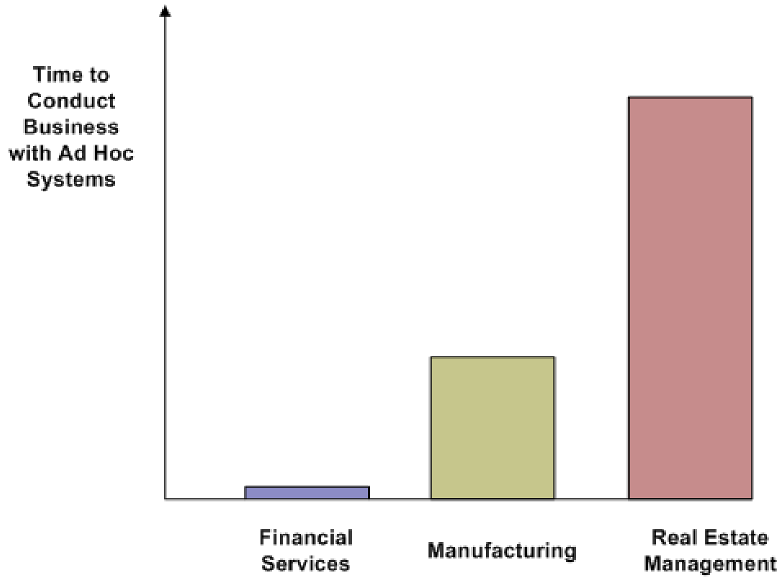

Figure 4.1: When data is lost or systems fail, it is only a matter of time before ad hoc solutions fail and business operations suffer significantly.

The notion of how long your business can operate without key systems is a helpful first step to understanding your recovery management needs. Thought experiments like this help get you into the right ballpark. To get the kinds of precise information you need to make business decisions, you have to delve into more detailed questions. In the next several sections, we will consider different business and technical scenarios and see how they influence a recovery management strategy. Across these scenarios, we will consider several key dimensions of recovery management, including:

- How long should it take to restore data and systems? These are commonly referred to as recovery time objectives (RTOs).
- What state of operations can we recover? This is known as a recovery point objective (RPO).
- Do regulatory compliance requirements impose recovery management requirements?
- How much data will have to be archived? How long will we keep it? These questions and others fall into the area of data life cycle management.
- How will we monitor backup operations and check the integrity of backups? These are operational issues within recovery management.
- What would happen if there were a catastrophic failure caused by a natural disaster? How would data be protected? Where would your business run critical operations? These questions fall into the area of disaster recovery management.

How you answer these questions dictates the type of recovery management solutions you will put in place. There is no one set of answers for particular-size businesses or for particular industries. A small health care provider may be subject to the same regulations as a large network of hospitals. A midsize construction company will likely have different recovery management needs than a similar-size retailer. In the following scenarios, we consider a range of examples that show how to understand the dimensions of recovery management, ask relevant questions, and discern the information about one’s business that will help you understand the requirements of your business.

Scenario 1: Small Business Backup and Recovery

Mention the term “small business,” and you are likely to conjure up images of aspiring entrepreneurs filling niche markets while creating more jobs than their larger enterprise counterparts. Ask a small business owner what it is like running one of those companies and you’ll probably get some of that positive imagery along with a healthily dose of reality. That reality includes a significant amount of time spent with accountants, lawyers, and bankers on top of the time spent on “real” work. Now add to the list the need to be the resident IT professional, and you can understand that recovery management might not get the attention it deserves.

Easy to Use Backups

One of the first requirements for small business recovery management is that it should not require a significant amount of time to manage backups. Small businesses, by definition, have limited staff. Automation and ease of use is a key factor. Backup software that provides agents for desktops, laptops, and servers can reduce the systems management overhead for small businesses.

In addition to easy-to-use software, small businesses should have easy-to-follow practices. If you have 10 desktop PCs and are installing backup agents on all of them, it is probably best to keep all 10 on the same backup schedule. That schedule should fit the needs of the most critical data. For example, the CFO's desktop should be backed up every night (at least). A good case could be made for not backing up as frequently the shared desktop used by part-time interns. After all, the data on that device is less important than the company financial data. The problem is that this strategy optimizes for storage space but increases the complexity of maintaining backup scripts and procedures. You now have two scripts to maintain rather than one. The savings in storage space is hardly likely to offset the extra management overhead.

In small businesses, a single, standardized backup schedule is better than many different ones. Larger businesses should classify their data and establish recovery management strategies based on different levels of criticality. For the small business, it is safe to assume all data is critical. (The one exception to this rule regards archiving, more on that follows).

RPOs and RTOs

Small businesses should define their RTOs and RPOs. If core business operations cannot continue without particular systems, those systems require short RTOs; financial services and retailers are examples. For other types of businesses, an RTO of several hours may be acceptable with next-day recovery conceivably sufficient for some low-demand systems.

RPOs address the question of how much work and data can be lost without adversely affecting the business. Weekly backups may be sufficient for businesses with relatively slow rates of change and those with comprehensive paper trails. In those cases, employees could re-enter transactions lost since the last backup. With advances in backup software, including ease of use and data deduplication technologies, it is probably a better idea to implement incremental nightly backups. There is not likely any marginal increase in the cost of backup software, additional storage costs are marginal, and the reduced risk is often worth those small, additional costs.

Disaster Recovery

A fire, flood, or other natural disaster can wipe out small business. If buildings are damaged, inventory destroyed, and paper files are lost, a small business could still recover. New office and industrial space can be rented and insurance can help cover inventory losses. Offsite backups may be the only way to recover important business records and transaction data.

Disaster recovery for a small business does not have to entail a complex set of data replication services and high-availability servers (although those are important for larger midsize and enterprise businesses). A simple strategy of keeping a full backup of all business data in a bank safe deposit box or other secure, offsite location can mean the difference between a disaster that sets back a business and one that breaks it. Cloud computing should also be considered. Cloud providers offer offsite storage that does not require you to transport, store, and manage physical media; as cloud services are available from any where with Internet access, your backups are readily accessible from virtually any location.

Archiving and Data Life Cycle Management

When it comes to keeping copies of data for long periods of time, you have to ask yourself what data is worth keeping. In larger businesses, the cost of data management warrants categorizing and prioritizing different types of data in the business. The high-priority and more critical the data, the more protection it receives. We noted earlier that small business should treat all business data as equally business critical in order to reduce management overhead. This keep-it-simple strategy can run into problems when we archive data for long periods of time.

The difference is that when we backup up data, it is written to tapes or disks on site. Both can be reused according to some schedule that meets data protection requirements. For example, tapes or disk space used for incremental backups can be reused once a full backup is made (assuming an RPO does not require the ability to restore to that incremental point). The goal of archiving is to keep a long-term copy of data, preferably in an offsite location in case of disaster. A practical consideration is how much storage media can be kept in an archive location, like a safe deposit box. Another question is, What is worth keeping long term? This is where the "treat all data as critical" rule of thumb breaks down for small businesses. Legal and financial professionals should be consulted on this question.

Archiving is a recovery management concern where we start to see some of the issues that midsize businesses have to address. Figure 4.2 shows a chart depicting the relative importance of various recovery management requirements to a typical small business. Of course, small businesses will vary in which topics they find important and this graphic is not meant to categorically describe all small business; however, it is useful for understanding the differences small businesses face when compared with midsize or emerging enterprises.

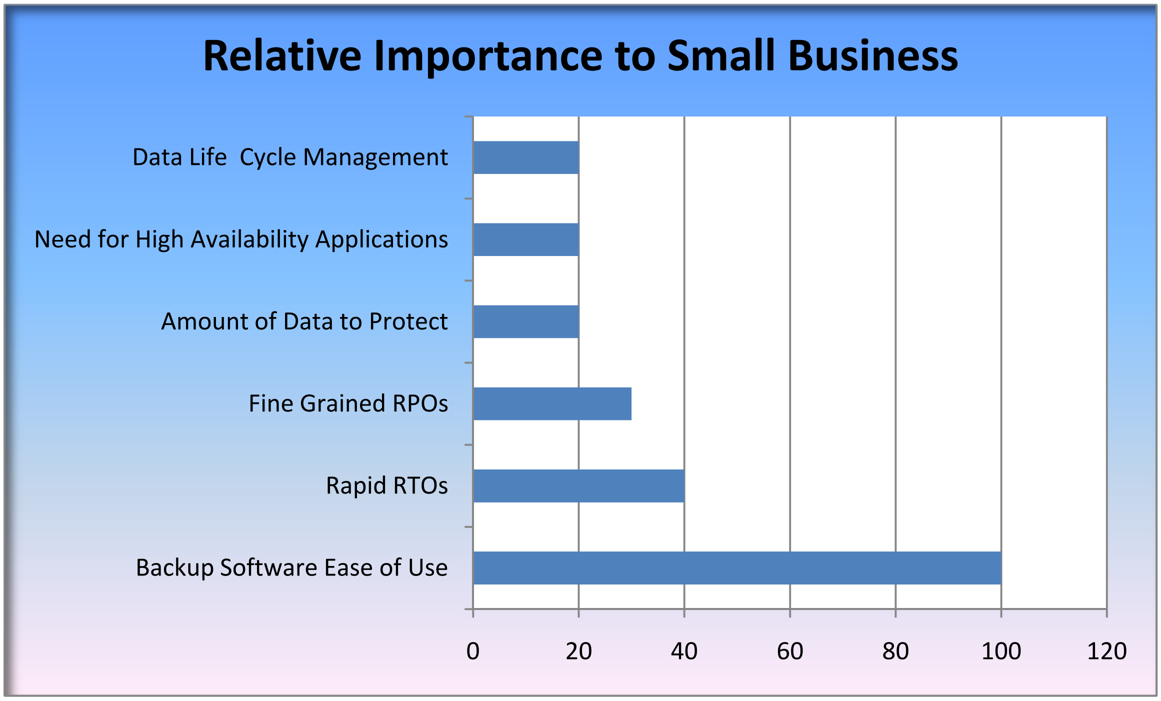

Figure 4.2: Top priorities for small businesses are ease-of-use considerations and the ability to provide basic data protection features.

Scenario 2: Midsize Business and Remote Office Protection

As companies grow in size, additional requirements move from the “nice to have” category to the “essential for business” category. Growing from small to midsize business tends to bring with it more demands on recovery management and data protection.

Increasing Data Protection Needs

In that way, a midsize business is more like an emerging enterprise with more risks to manage. In other ways, a midsize company may be more like small businesses in that they cannot support a substantial internal IT staff with deep and broad experience in a range of IT management and technical issues. This results in a typical set of requirements that is a mix of both small business and emerging enterprise requirements.

Ease of use, the ability to meet somewhat more stringent RTOs and RPOs, and increasing need for disaster recovery and more formal life cycle management procedures will be found in midsize companies. Another area of concern that midsize and larger businesses face is the need to protect data in remote office locations.

Protecting Remote Offices

Midsize and lager businesses often face the challenge of protecting data in remote office locations. Each of these remote offices can, in some ways, be considered like a small business. There are multiple devices requiring backup services but no dedicated staff to ensure data is properly protected. For the midsize business, there are a number of considerations with regard to remote office backups, including:

- Consistency—Businesses should have a recovery management strategy in place and it should be applied consistently across offices.
- Cost control—Duplicating backup servers across multiple offices can be inefficient. Remote offices may not warrant a dedicated backup server onsite; however, several remote offices sharing a single backup server can be cost effective.
- Management—Backup administrators need to have information about the status of backup operations. If a backup fails, for example because of insufficient disk space on the backup device, an administrator should be alerted. Management reports are also needed to monitor trends in time required to perform backups and the growth in backup storage.
- Service delivery—Restore operations should be done in ways that meet RPOs and RTPs. They must also be done in a cost-effective way, which means not requiring a systems administrator onsite at the remote location. Restores, like other management operations, must be done remotely. This requirement is especially problematic in those remote offices that have no IT staff, so such operations are passed on to someone else in the office.

All of these considerations can be met with a backup system that allows for remotely managed backups. Backup agents can be installed on remote office desktops and servers. The agents then function with backup servers in a central office to perform backup, restore, and management reporting operations. There is minimal need for onsite tasks, except for steps such as powering on devices. (Although this is less of a problem with remote management software that takes advantage of hardware that allows for network-based power-on and other basic management tasks).

The scenario depicted in Figure 4.3 assumes fixed, remote offices. In addition to these, midsize businesses should consider data protection for mobile users. We are far less tethered to our offices than we were even 10 years ago. Businesses do not need a dedicated office in a region to have a presence there. The US map in Figure 4.3 could easily become a global map if we were to include regional representatives for a midsize business who might be located in Asia, Europe, Oceania, or other regions of the world.

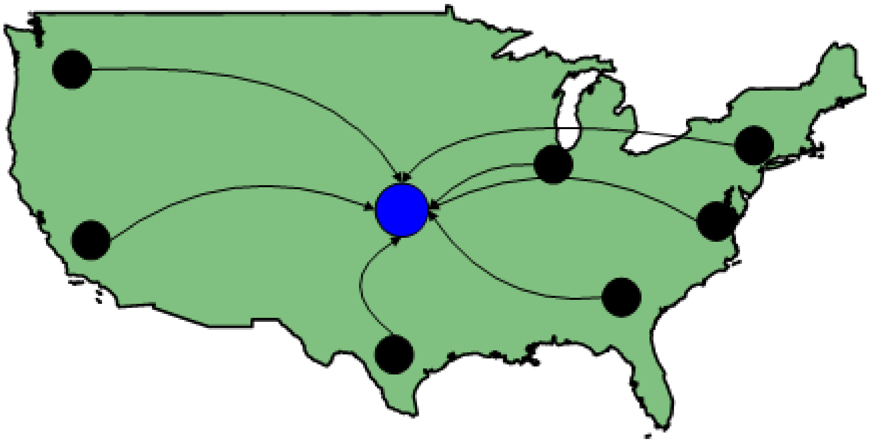

Figure 4.3: Remote offices should back up to centralized backup servers for efficiency, reliability, and manageability.

Note, when selecting backup systems to support laptop users, consider how often those users will have slow or unreliable Internet connections. Ideally, backup software will be robust enough to reliably perform backup and restore operations in spite of less than ideal connectivity.

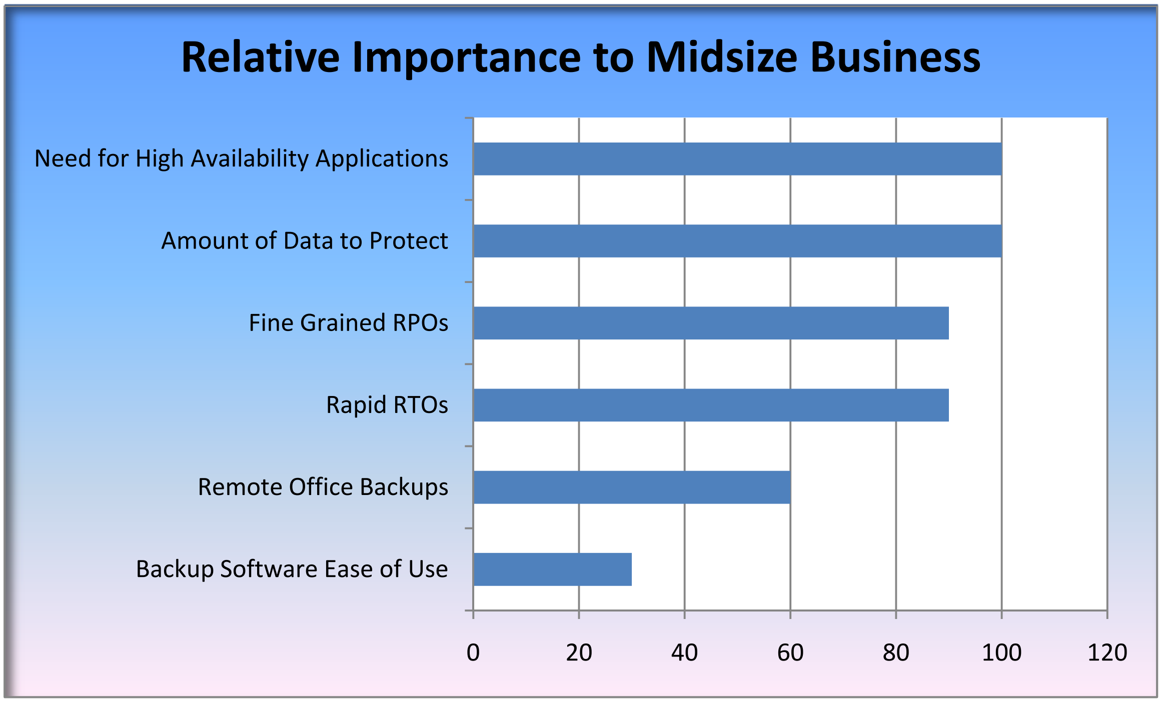

Figure 4.4: Top priorities for midsize businesses are similar to small business priorities but there is increasing emphasis on RPOs, RTOs, and amount of data to protect. Remote office backups become a concern for midsize companies.

Scenario 3: Operational Management and Enterprise Backup

Moving to larger businesses and emerging enterprises brings with it more complexities in recovery management and more diversity in requirements. Rather than try to capture a typical large business, we will consider three broad but common areas of concern:

- Operations management
- Data protection in virtualized environments
- Continuity and failover

Operations management in larger organizations includes virtually all the requirements found in small and midsize business but with different levels of importance. For example, there is less importance on ease of use when selecting backup software. This is not to say that ease of use is not important, it is; however, in large organizations, essential functionality must be in place even if it might not come with an easy-to-use interface. There are also features that become increasingly important as we move from small to midsize and emerging enterprises, such as:

- Encryption
- Deduplication
- Centralized management

These additional features are driven by a range of business requirements, including compliance, cost control, and the ability to meet service level agreements (SLAs).

Encryption

Encryption usually comes up in discussions about security. Discussions about transmitting sensitive business data over the Internet will quickly focus on encryption technologies to protect the confidentiality of that data. When businesses need to ensure confidential information is protected on laptops, even if they are lost or stolen, full disk encryption should be considered. Backups can share important similarities with these use cases and for those reasons, encryption can be an important feature of a backup solution.

Backups are often moved or transmitted to locations other than where the data originated. Offsite storage mitigates the risk of losing both a server and a backup to damage to a facility. Backup tapes may be lost or stolen while in transit. Security procedures at an offsite facility may be more lax than one would expect. In both cases, you have confidential data outside the protection of a business' normal access controls and physical security. Encrypting data on backup tapes and disks can provide an additional level of protection for ensuring the confidentiality of private and sensitive data. As a general rule, all business data that is stored offsite or in the cloud should be encrypted.

Deduplication

Reducing the amount of storage required for backups and the time required to generate backups translate into cost savings. Deduplication will be useful for any size business, but the larger the organization, the greater the benefit.

Deduplication can occur on either the source or the target system. Deduplication on the target side has the advantage of reducing demand on the source device's CPU. Backup agents do not have to support deduplication functionality on the client side when the operation is performed on the target. This can reduce problems with increasing the footprint of the backup agent on the source system. In addition, the target backup server can be sized appropriately to handle the computing requirements for deduplicating data from multiple source systems.

Centralized Management

Management features are important for any size business and any type of business. At minimum, you need to know that your backup processes run, your restores are successful, and the storage you have allocated for backups is sufficient. Management complexities increase quickly with the size of an organization, leading to several additional needs:

- Information about different types of backup processes, such as full, incremental, and differential
- The ability to define multiple policies for different classifications of data that require different backup schedules
- Reports showing trends in disk and tape usage as well as CPU utilization on backup servers
- Effective deduplication rates
- The ability to restore with minimal manual intervention, for example, not physically visiting a remote site to restore files

Centralized management affects how you perform backup operations, how you collect data about those operations, and how efficiently you can perform these operations. As Figure 4.5 shows, centralized management becomes one of the most important features of recovery management systems as the size of the organization grows.

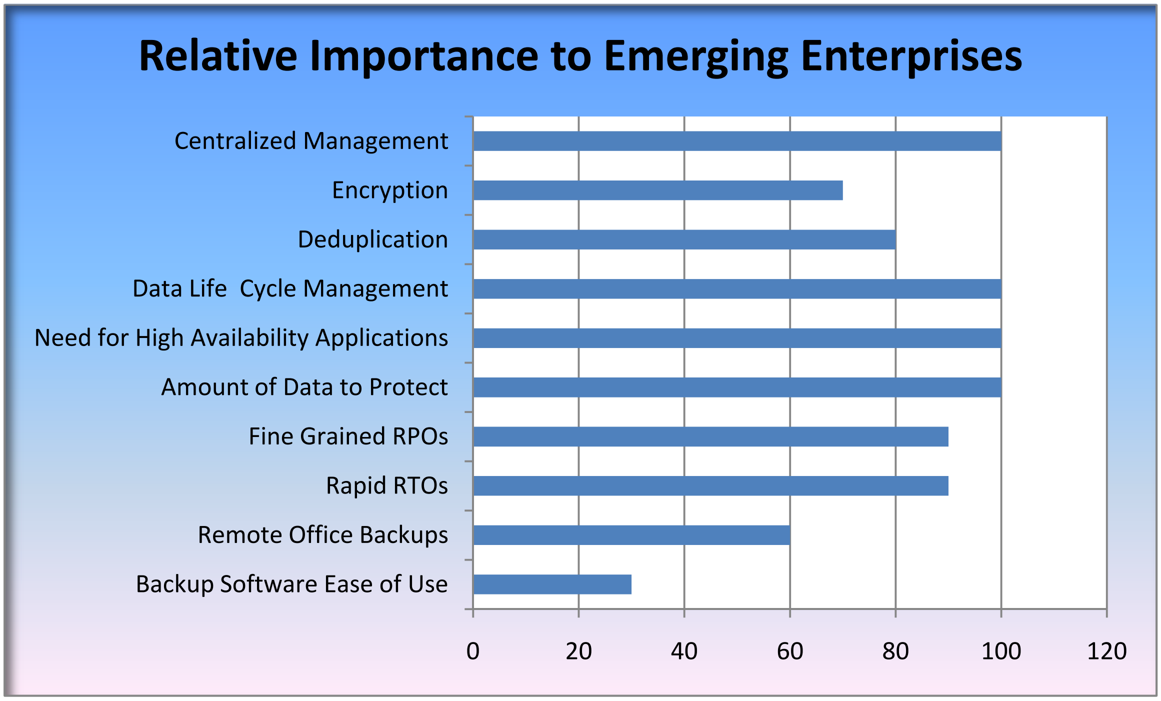

Figure 4.5: Top priorities for emerging enterprises expand beyond the set found for smaller business. Encryption, deduplication, and centralized management are important functions for larger organizations.

Scenario 4: Data Protection in Virtualized Environments

Server virtualization can significantly increase server utilization and help control hardware costs. There are, however, some recovery management considerations that need to be taken into account, especially for midsize and larger companies that deploy large numbers of virtual servers. Some of the most important issues are:

- Options for restoring virtual servers and files
- Demand on CPUs
- Integration between virtual machine services and recovery management software

Implementation details are important to understand when planning the use of backup software with virtual servers. Suppose you need to restore a file from a virtual server backup. Does the backup software support individual file restores from a virtual machine backup? If not, you will need to plan for additional time and storage space to accommodate the recovery operation. One way to handle the lack of selective file restores is to restore the entire image to a staging server and then copy the needed files to their target location. This process will increase recovery time as well as require additional storage space.

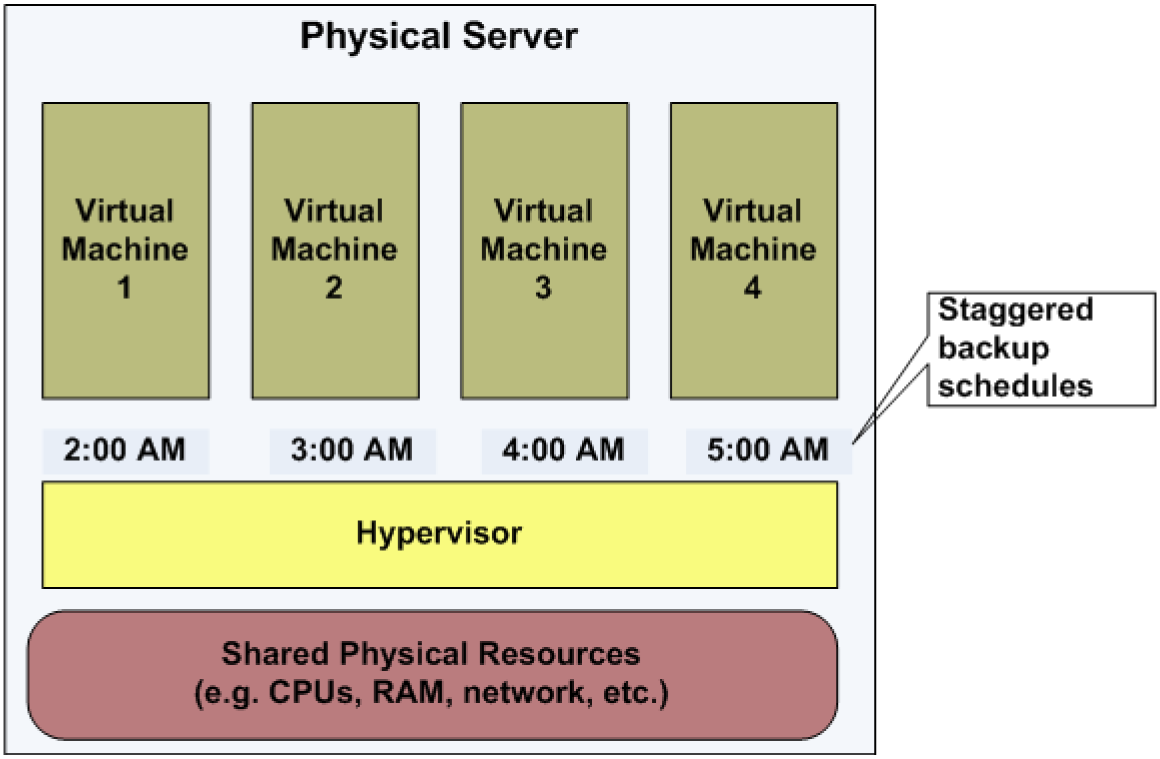

Figure 4.6: Backups should be staggered on virtual machines to minimize competing demands for shared resources.

Another implementation detail you need to watch carefully is the demand on CPUs during backup operations. Backups should not be scheduled to compete with each other or with other peak-demand periods on other virtual machines sharing the same physical server. Virtual machines isolate many management concerns to a single virtual machine, but backups are one of the areas where due consideration has to be made for global operations on the server.

Another issue to consider is how to maximize backup performance and available features by leveraging the features offered by the virtual machine vendor and recovery management software vendors. Each has specialty expertise. The virtual machine vendors can optimize low-level operations in the hypervisor, typically those close to the hardware, to increase performance. Recovery management software vendors are likely to offer more management functionality and provide a common feature set that hides some of the implementation details when dealing with virtual machines. Virtual machines are staples of day-to-day operations in midsize and larger organizations, but they can also play a crucial role in continuity and failover protection.

Scenario 5: Continuity and Failover

This chapter opened with a discussion of how different business requirements will drive different recovery management strategies. The question of how long and how well a business can function with systems down is a telling example of how wide ranging requirements can be. The ability to maintain continuous IT services increases in importance with the size and complexity of business operations.

Let's consider how a hypothetical midsize business might tackle the problem of disaster recovery. Our fictitious firm is based in the US with a headquarters in Chicago and regional offices in Atlanta, Boston, Houston, Denver, and Los Angeles. The executives in the business have determined that the company needs to be able to provide core business services even if the headquarters is hit by a natural disaster. The CIO and IT staff know five things they need to do to ensure a sound disaster recovery strategy and practice:

- Identify mission-critical applications and servers
- Define RTOs and RPOs
- Design a failover architecture
- Implement and manage recovery procedures
- Test the disaster recovery systems and procedures

Small businesses can reasonably assume that all data in the business is equally valuable and should be protected at the same levels. This is not necessarily true, but it is a useful fiction because it reduces the management overhead. As data volumes grow, the advantages of simplified management no longer outweigh the cost of maintaining unnecessary RTOs and RPOs.

From a disaster recovery perspective, mission-critical applications must have their data and servers protected in such a way that recovery times are short and recovery points are close to the time of failure. In our example business, management determines that the sales processing, financials, and customer relationship management (CRM) systems are critical. Other back-office applications, such as human resources applications, inventory management, and decision support are designated second-tier systems. They can be unavailable for up to 48 hours.

The sales processing system is accessible to customers through a self-service Web application, so management does not want customers to experience degraded service; that application will run on a dedicated server as well. Management also determines that they can tolerate some drop in performance in disaster recovery situations for other applications. The systems designers take advantage of this by running the financials and CRM systems, which normally each run on their own dedicated servers, in virtual machines hosted on a single physical server. The second-tier applications will run on a single physical server.

Designers decide to use the Atlanta and Denver offices as disaster recovery sites. The Atlanta office, like headquarters, has a dedicated IT staff, so it is chosen to host critical applications. Two servers are installed and dedicated to disaster recovery; one for the sales processing system and the other to host virtual machines for the other critical back-office applications. Denver does not have a dedicated IT staff, but they do have excess server capacity that can accommodate virtual machines running the second-tier applications.

To implement the disaster recovery plan, a high-availability application is installed in Chicago and Atlanta. Data is continuously replicated from Chicago to Atlanta to ensure RPOs and RTPs are met. High availability is not needed for secondary applications, so backups of those systems are copied from Chicago to Denver on a daily basis. In the event of a disaster, the Chicago IT staff will work with Denver staff to start virtual machine instances and restore the second-tier applications from backups.

The design seems sound. RPOs and RTOs can be met, business objectives are accounted for, and staff is in place to implement the plan as needed. Good disaster management strategies, no matter how well designed, should be tested. Things are bound to go wrong. Disasters tend to be, well, disasters, so you need to ensure that plans cover as many decision points as possible. Businesses should at least test at the unit levels that data is replicated from primary to standby servers, backups are reliable, and staff understand the procedures to follow in the event of a disaster.

Continuity and failover services are important parts of a recovery management strategy. Backup software and well-defined recovery procedures provide for basic disaster recovery services; larger businesses with more complex information management needs should consider high-availability and data replication applications as well.

Best Practices in Availability, Continuity, and Disaster Recovery

The *Shortcut Guide to Availability, Continuity, and Disaster Recovery* has covered a lot of ground in backup, recovery, and high availability. Sometimes we have delved into the technical challenges, other times we've looked at recovery management from a business perspective, and in some cases, we've tried to bridge the technical and business perspectives. As we conclude this shortcut guide, it is time to consolidate some of the topics we've examined and compile a summary of best practices in availability, continuity, and disaster recovery.

Best Practices: Making the Business Case for Recovery Management

- **Do not forget to address the obvious requirements.** Some of us in IT can get captivated by new applications, hardware, and methodologies to the point we risk losing site of basic business requirements. Remember, recovery management must be able to protect against basic risks: lost files, application errors, corrupted data, and catastrophic failures.
- **Account for expected growth in data volumes.** New business initiatives will create new demands for backup and disaster recovery. Make sure those needs are considered when assessing the feasibility and worthiness of new initiatives.
- **Not all data is created equal; not all applications are equally critical.** Do not buy more than you need when it comes to recovery management. Businesses face many types of risks without the need to eliminate all of them. Distinguish the value of different applications and types of data. Protect each according to their value to the business. Use risk management practices to inform your recovery management strategy.
- **Use data protection strategies to enable new business initiatives.** Recovery management, like other IT services, is not only a cost to businesses but an investment that enables innovative business operations. Give customers access to more data because you can keep it online for longer periods of time; open up analytic tools to customers to better understand their buying patterns because you can provide a reliable service thanks to replication and high-availability servers.

Best Practices: Overcoming Technical Challenges

- **Protect virtual servers and their data**. Backup and restore operations on virtual servers introduce new constraints not seen when dealing with physical servers. Ensure your backup software accommodates virtual servers, ideally providing the same features for both physical and virtual servers.
- **Understand application-specific requirements.** Relational databases, content management systems, and email systems all introduce challenges with restoring fine-grained data structures (for example, tables in databases and individual messages in email systems).
- **Support remote office backup and recovery.** Centralized recovery management can help control costs by minimizing backup infrastructure and reducing the need for IT support in remote sites. Also, accommodate laptop and other mobile devices that may not be continually connected to a corporate network.
- **Replicate critical data to a disaster recovery site.** Restoring from backups takes time. When continuous service is needed, use replication to keep a standby server up to date and ready to take over in the event of data loss or other failure. In cases where immediate recovery is needed, consider the use of high-availability applications that can monitor the primary server and switch to the standby server when needed.

Best Practices: Recovery Management Practices

- **Use the different types of backups, such as full, incremental, and differential to maximize protection while reducing storage costs.** Using incremental and differential backups can also reduce the time required to perform backups.
- **Use disk storage for performance; consider tapes when cost and portability are top priorities.** Disk-to-disk backups have speed advantages over tapes but tapes can be less expensive and are easy to transport to offsite storage facilities.
- **Understand data life cycle requirements.** Just as not all data needs equal backup protection, not all data needs the same level of archiving and preservation. Store backups and archives only as long as they serve a business need.

Summary

From small businesses to emerging enterprise, businesses are facing technical and business challenges to protect data and maintain continuous business services. Although there are not "one size fits all" solutions, there are sound practices for determining your business' particular needs. Backup and recovery software has continued to evolve with technology. Backup solutions have adapted to virtual environments and take advantage of low-cost disk storage and changing business needs. These solutions provide centralized management consoles, consolidated reporting, and other tools to help IT staff keep up with increasing demands for data protection. Business practices have also matured as businesses leverage techniques such as remote office backups and practices such as managing data life cycle requirements to control costs without sacrificing data protection. Technology and business practices will no doubt continue to advance.