

Brought to you by **Acronis**

*Acronis Special Edition*

# Backup

FOR  
**DUMMIES**<sup>®</sup>  
A Wiley Brand

## *Learn:*

- Answers to common questions about backup and recovery
- Ten tips for easier backup and recovery
- How to address the modern data protection challenges caused by virtualization, the cloud, and data growth

**Joel Berman**

*Acronis Fellow*



# About Acronis

Acronis sets the standard for New Generation Data Protection through its backup, disaster recovery, and secure access solutions. Powered by the AnyData Engine and set apart by its image technology, Acronis delivers easy, complete, and safe backups of all files, applications, and OS across any environment — virtual, physical, cloud, and mobile.

Founded in 2002, Acronis protects the data of over 5 million consumers and 300,000 businesses in over 130 countries. With its more than 100 patents, Acronis' products have been named best product of the year by Network Computing, TechTarget, and IT Professional and cover a range of features, including migration, cloning, and replication.

For additional information, please visit **[www.acronis.com](http://www.acronis.com)**. Follow Acronis on Twitter: **<http://twitter.com/acronis>**.

***Backup***

FOR  
**DUMMIES<sup>®</sup>**  
A Wiley Brand

***Acronis Special Edition***

**by Joel Berman,  
Acronis Fellow**

FOR  
**DUMMIES<sup>®</sup>**  
A Wiley Brand

## Backup For Dummies®, Acronis Special Edition

Published by  
**John Wiley & Sons, Inc.**  
111 River St.  
Hoboken, NJ 07030-5774  
www.wiley.com

Copyright © 2014 by John Wiley & Sons, Inc.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

**Trademarks:** Wiley, For Dummies, the Dummies Man logo, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. Acronis and the Acronis logo are registered trademarks of Acronis. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

**LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.**

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact [info@dummies.biz](mailto:info@dummies.biz), or visit [www.wiley.com/go/custompub](http://www.wiley.com/go/custompub). For information about licensing the *For Dummies* brand for products or services, contact [BrandedRights&Licenses@Wiley.com](mailto:BrandedRights&Licenses@Wiley.com).

ISBN: 978-1-118-95249-8 (pbk); ISBN: 978-1-118-95265-8 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

---

## Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

**Project Editor:** Carrie A. Johnson  
**Development Editor:** Kathy Simpson  
**Acquisitions Editor:** Katie Mohr  
**Editorial Manager:** Rev Mengle

**Business Development Representative:**  
Sue Blessing  
**Custom Publishing Project Specialist:**  
Michael Sullivan  
**Production Coordinator:** Melissa Cossell

# Table of Contents



<b>Introduction</b> .....	<b>1</b>
About This Book .....	1
Foolish Assumptions .....	2
Icons Used in This Book .....	2
Where to Go from Here .....	2
<b>Chapter 1: Data Protection 101</b> .....	<b>3</b>
Defining Data .....	3
Protecting Data .....	4
Making Backups Simple, Complete, and Safe .....	6
Working with simplicity .....	6
Understanding completeness .....	6
Relying on safety .....	7
<b>Chapter 2: Capturing Data for Backup</b> .....	<b>9</b>
Understanding Backup Types .....	9
File backups .....	10
Image backups .....	11
Backing Up According to Plan .....	11
Choosing full, differential, or incremental backups .....	12
Setting the RPO .....	13
Taking a snapshot .....	14
Backing Up with and without Agents .....	15
Considering Backup Products .....	16
Bare-metal restore .....	17
Single-pass backup .....	17
<b>Chapter 3: Storing Backups Safely</b> .....	<b>19</b>
Creating a Backup Policy .....	19
Backup plan .....	20
Retention policy .....	20
Selecting Backup Software .....	22
Picking Backup Media .....	23
Disk .....	24
Tape .....	25
Cloud .....	25

Choosing Offsite Storage Locations.....	27
Online network.....	27
Dark site.....	28
Cloud backup .....	28
Considering Compression and Deduplication .....	29
Calculating the Costs.....	30
<b>Chapter 4: Recovering Data . . . . .</b>	<b>31</b>
Recognizing Data Loss.....	32
Putting Your Recovery Plan in Motion.....	33
<b>Chapter 5: Managing Backup . . . . .</b>	<b>35</b>
Keeping Current on Backup Products.....	35
Setting the Backup Window.....	36
Creating and Checking a Backup Plan.....	37
Keeping it simple (or not).....	37
Setting backup windows .....	38
Checking execution .....	38
Monitoring the plan.....	38
<b>Chapter 6: Ten Things to Know about Backup . . . . .</b>	<b>41</b>
The Value of Your Data .....	41
The Cost of Downtime.....	42
Workload Priorities.....	42
Where Your Backups Are Stored .....	42
How Long to Keep Backups .....	43
What Recovery Tools to Use When .....	43
The Details of Your Backup Plan .....	43
What Data Is Excluded from Backup .....	44
How (And How Deeply) to Test Backups.....	44
How to Frame Backup Questions.....	44

# Introduction



**M**any people think that backup is simply about copying data and then copying it back if needed. The process probably was that simple in 1960 — as simple as punching another set of computer cards and keeping them in a safe place.

Technology has advanced rapidly in the past 50 years, of course, so backup processes have also had to advance. Backup covers a variety of use cases from simple file recovery to recovery of complex systems.

## *About This Book*

Backing up by simply copying a few files to another drive and then copying them back if needed satisfies only the most trivial cases. This book covers use cases from the simplest single server to large and complex systems.

Data protection (DP) is an umbrella term covering backup and recovery for every kind of information, including data, programs, and system software. The term is often confused with another meaning of data protection, which has to do with encrypting personal information such as credit card and healthcare data. This book is about backup and recovery. Along with DP you may see disaster recovery (DR). DR means recovering an entire system after some sort of physical damage such as a flood or a massive cyber-attack with various viruses being injected.

This book is your guide to putting these protection systems in place.

## Foolish Assumptions

When I wrote this book, I assumed the following things about you, the reader:

- ✔ You're familiar with information technology (IT), but you're not a data protection expert.
- ✔ You're somewhat experienced in administering systems but aren't a skilled sysadmin.
- ✔ Mostly, I assumed that you want to begin or improve your company's backup processes and are looking for some help on the foundational concepts, as well as guidance on choosing backup products and setting up a data protection solution.

## Icons Used in This Book

As in all *For Dummies* books, icons in the margins point out certain types of information.



Text marked with the Tip icon provides helpful hints on backup concepts or techniques.



The Remember icon flags important facts that you should keep in mind.



You don't have to read Technical Stuff text, but I hope you will, because it gives you a deeper understanding of backup.



Don't skip anything that has a Warning icon. Failing to heed the warning could cost you time, money, and/or data.

## Where to Go from Here

Like all *For Dummies* books, this one can be read in whatever order is most helpful to you. Start with Chapter 1 and go straight through, or skip around. It's up to you.



# Chapter 1

## Data Protection 101

---

### *In This Chapter*

- ▶ Understanding why data needs protection
  - ▶ Knowing what data to protect
  - ▶ Seeing what goes into a backup system
- 

Companies have protected their data ever since carbon copies were stored in bonded warehouses, but times have changed. The techniques for protecting information and recovering from disastrous data loss have also changed. This chapter brings you up to speed on modern data protection.



Throughout this book, I use the term *data protection* to mean storing data in such a way that it's easy to search and restore, no matter where it's located in the backup archives. I use the term *disaster recovery* to mean recovering a server, workstation, or entire data center quickly after a major problem.

## *Defining Data*

What is data? The question may seem to be very simple, but answering it isn't so easy. Data may mean simple text files, or it may mean a vast range of types of complex information. In a modern computer system — and for the purposes of this book — data means programs, files, and metadata.

Because data takes many forms, it's easy to get confused about which data needs to be protected so it can be reproduced easily. In general, however, items that have value should be protected. Just as cavemen were careful to keep their children away from saber-toothed tigers, you're probably careful about where you leave an expensive watch or park your car. Why do you protect

your things? They have value to you. If these items are lost or stolen, you lose the value associated with them. Even insurance can compensate only with money; it can't restore the items that it covered.

Data is a bit different. If you take proper care of it, you should be able to recover it easily — and its value along with it.

## Protecting Data

Information technology (IT) experts protect data against loss and destruction, such as through theft, accidental deletion, or deliberate alteration. Fortunately, protecting data is easier than you may think. If you plan carefully and execute your plans well, it's possible to protect your data to any level you desire.

Consider protecting the following items:

✓ **Bootstrap data:** *Bootstrap data* is used to start a machine. It's the program that runs first when a system is powered on or restarted. Without a valid bootstrap, a system never becomes operational.

✓ **File-structure metadata:** *File-structure metadata* describes where all the files and folders are located, as well as where the bootstrap data, operating system, and drivers are. File-structure metadata records which blocks on the disk are being used and which are free, and it maps every directory and filename to specific locations on the disk drive. This type of data also contains permission and access lists that prevent unauthorized read or write operations by checking against those lists.

In some systems, special logs keep historical records as an audit trail of changes. These logs are used to recover from power failures and other abrupt halts. All this metadata is important when you need to restore an entire system.

✓ **Driver binaries:** *Driver binaries* control devices that read from discs, tape, or the network. Drivers are often part of the operating system (OS; see the next paragraph), and they must be compatible with the rest of the system.



They often come with devices that are purchased but not included with the OS.

- ✔ **Operating system:** OS code is often shipped on a physical disc. The manufacturer often issues updates over the network, so the initial disc becomes obsolete.

It's important to make backups of the OS after updates are applied so that you always have a current version for recovery purposes. If you don't save the updated version, you'll have to reapply all updates when you restore the system — an error-prone, time-consuming process.

- ✔ **Configuration files:** Configuration files are numerous. Some are as simple as a file containing the system's name or time zone; others are as complex as the Windows Registry, a large file containing thousands of pieces of information. Password files are configuration files, for example. Some software is so secure that if the password file is lost, no data can be recovered. Also, many applications have complex configuration files that store all sorts of information that's specific to the system.

- ✔ **Application programs:** Companies buy many of the applications they use from third parties, but some companies develop their own applications. One of the most tragic cases of lost data occurs when a program is developed in-house and then the original source code is lost. If a company loses these program files, it may be unable to use any of the data that has been saved. Having the data but not the programs that use that data is tantamount to losing the data itself. Businesses must protect applications with the same care that they protect the application data itself.

- ✔ **Data files:** Data files are highly portable files that are used by many programs, such as spreadsheets and PDFs. Other data files are unique to specific programs, such as log and database files. These files may be short, permanent, large, or rapidly changing, and any company's server could contain hundreds of thousands or even millions of them.

- ✔ **Databases:** While databases can be considered another application with data files, there are special considerations in protecting databases to ensure consistency. Don't assume that any backup process also protects databases.



# Making Backups Simple, Complete, and Safe

The mantra for backup is “simple, complete, and safe.” In this section, I break down these components.

## *Working with simplicity*

*Simplicity* is probably the most important factor, because it’s the quality that prevents mistakes.

Companies that have evolved their backup solutions over years may be using incompatible products and backup media, which can cause problems. If they have no standard recovery checklists or written procedures, they’re almost always asking for trouble. More times than I can think of, people have recovered data from the wrong backups or even overwritten good data with bad data because they were trying to follow a complex, poorly documented process under pressure and in a hurry.



Arguably more important than clear checklists and plans are easy-to-use interfaces, with minimal setup and single-click recovery. But in any case, make sure you consider usability in choosing an easy, complete, and safe backup product.

## *Understanding completeness*

*Completeness* means having not only all the data necessary to recover the system after any failure, but also the proper tools and processes. Your backup software may help you create a bootable CD or DVD, which you can use to recover when you have no running system at all. If your system doesn’t allow you to create such a disc, however, you have to reinstall the system — a lengthy process that requires configuration and OS updates (both discussed earlier in this chapter).

Likewise, your backup program should be able to reproduce disk partitions and formats instead of expecting you to do the work manually; this task is tricky and must be done perfectly.

## Any data, any environment, any location, any device

While virtualization has been around for over 30 years, it has exploded over the last ten years as VMware introduced a software hypervisor. Since then a number of hypervisors have appeared, and the hardware manufacturers have migrated many of the software functions into hardware for performance reasons. Additionally the emergence of multi-core processors and increasingly dense memory have enabled more and more workloads that had only been possible on mainframes and minicomputers. The result is that workloads are migrating from physical to virtual and back, and a number of hypervisor and cloud technologies are being used.

Because of these rapid advances, it's important to not consider virtual

machines as something separate, but to choose data protection/disaster recovery (DP/DR) software that works across physical and virtual and across the popular hypervisors. This allows the backup archives to be preserved and useful no matter where the workloads reside, and also exposes added value with the capability of migrating operating systems, applications and data among physical, virtual, and cloud environments using any hypervisor. Choosing a backup product that works with any data, any environment, any location, and any device gives you the greatest flexibility in taking advantage of the full range of computing offered today and tomorrow.

Finally, your backup software should print recovery instructions for you to follow, just as an airline captain follows a checklist to land an aircraft.

This can be easily summed up as “completeness means any data, any location, any environment.”

## *Relying on safety*

*Safety* has two components:

- ✓ **Reliability:** To be safe, backup data must be reliably captured. Obviously, the data must be readable, and if the backup depends on any other data, that data must also be readable. Backup software may use a variety of techniques to capture data safely, but if the data can't be

read, the system can't be recovered. One missing piece of data can make the entire system unrecoverable.

- ✓ **Security:** After data is captured, it needs to be protected against changes and theft. If your backup system isn't secure, a bad actor could sneak into it, steal data, and then do something to harm your main system with no one being the wiser.



The only cure is prevention. Keep your backup process simple and well documented. Start at the beginning by capturing the data properly (see Chapter 2). If you capture it incorrectly, all hope is lost.

## Chapter 2

---

# Capturing Data for Backup

.....

### *In This Chapter*

- ▶ Understanding image and file backups
  - ▶ Setting a backup plan
  - ▶ Using agents (or not)
  - ▶ Selecting backup products
- .....

**F**or most of us, backup means copying a few critical files to a separate disc or USB stick. But large systems and servers house many files, and the data in those files changes constantly. A systematic means of capturing system data, applications, and metadata even while files are open and being changed is mandatory if you want to recover from lost or damaged data and applications.

In this chapter, I walk you through the main types of backups, discuss the scope of backups, and introduce a couple of special backup situations.

## *Understanding Backup Types*

Excluding tape, most storage devices look like disk drives to the low-level software. CD/DVD devices, for example, have special burning requirements but still appear as disk drives — a sequence of blocks. The hardware presents a sequence of blocks to the operating system (OS), each block being separately readable and writeable. Within some of these blocks is *metadata*, which is used to hold directories, lists of used and free blocks, bootstraps, partition information, bad-block lists, and remapping information, as well as those blocks used for holding data.



By the way, the storage space used by the metadata takes away from the free space on a freshly formatted disk. That's why the free space you see on a disk is less than the raw capacity specified by the vendor.

Generally, you have two major ways to capture data for backup: image backups and file backups. I discuss both types in the following sections.



Although both methods allow you to search for specific files and data, the image backup is a superset of a file backup because it also contains system metadata.

## *File backups*

The original type of backup was a file backup, which is still a popular method. A *file backup* copies all the files and folders from the current data to backup media. The process is similar to copying personal files to a USB stick, a USB drive, or another directory.

Because file systems keep track of when a file is created and modified, the file backup may only copy the files and folders that have changed since the last backup. It's simple for a backup program to copy files, because the OS provides all the necessary functions to look up and copy files. Copying, however, creates a lot of overhead for the backup system, because for every file, the system has to do the following:

- 1. Find the blocks where the folders are.**
- 2. Read the folders.**
- 3. Look for filenames.**
- 4. Determine where those files are located.**
- 5. Read and copy those blocks.**

This process can take a great deal of time. Also, if the system is operational at the time, there can be a lot of contention for resources, which increases backup time and/or reduces performance.



## Image backups

An *image backup* bypasses most of the OS file-lookup overhead by simply copying blocks to backup in order to the end, making a complete record of everything on the disk. Image-backup software is designed by people who have deep understanding of how data is placed on disks, so it's also able to determine which blocks have been changed since the last backup and copy only those blocks. If a 2GB file has only a small change, for example, the file backup has to copy the entire 2GB, whereas the image backup has to copy only the block that changed. This process results in extremely fast backups.



The fastest backup software captures data only from blocks that are in use. It won't copy bad, temporary, unchanged, or unused blocks unless specifically requested to do so.

Image backups allow you to inspect the image so you can determine where individual files are and do fine-grained file recovery. Often, image backups can be mounted as full disk drives so an administrator can recover or compare data from different periods.

Both image backups and file backups can be full, differential, or incremental, and both types can exclude specified folders and file types. I cover backup types in the next section.



Images used to be called *snapshots*, but that term has two different backup-related meanings, so it's better to say *image backup* than *snapshot backup*. I cover snapshots in "Taking a snapshot" later in this chapter. Also, both file and image backup should be able to exclude the copying of specific files so that temporary and other unnecessary files are not copied.

## Backing Up According to Plan

A backup plan describes the data to be backed up and the scope of the backup. In this section, I cover the basic decisions you need to make.

## Choosing full, differential, or incremental backups

Backups are divided into three types:

- ✓ **Full:** The first backup of a system, capturing everything in it.

The upside of a full backup is that it's self-contained. The downside is that it takes up a lot of space, can take a long time to complete, and can be almost identical to a previous full backup.

- ✓ **Differential:** A backup that captures only the differences between the current state and the last full backup. Recovering from a differential backup requires both the last full backup and the differential backup to be valid.

The upside of a differential backup is that it's much faster than a full backup. The downside is that it takes up more space than an incremental backup (see the next paragraph) and requires at least two backup files to be read for recovery.

- ✓ **Incremental:** A backup that captures only the differences between the current state and the last differential, incremental, or full backup.

The upside of an incremental backup is that it's very small and very fast. The downside is that recovering with an incremental backup requires all the data from the last full backup and every successive incremental backup until the recovery-point objective (see the next section) to be valid and read. As a result, recovery with this type of backup can be time-consuming.

Most backup software allows incremental backups to be consolidated offline, which greatly improves reliability and recovery time. A new backup type called *reversed incremental* automatically does the consolidation as the backup is captured. Another new type, called *always incremental*, has different meanings depending on the vendor.



For most companies, the size of a daily backup — and thus the size of an incremental backup file — is 3–5 percent of a full backup. This figure varies widely, however, depending on

the business. Also, right after upgrades or major changes, an incremental backup would be quite large. Most of the time if you're planning a major upgrade of systems, applications, or data, you should perform a full backup before the change and then another right afterward. If you intend to tinker with the system after the changes, you should make frequent backups.

## Setting the RPO

The *recovery-point objective* (RPO) defines how often backups must be made. The amount of time the system is allowed to be paused for backup is the *backup window*. The goal is to eliminate the backup window by using techniques that allow a running system with changing data to be backed up without having to disable and pause the system. Generally, the backup window and the RPO conflict because frequent recovery points are desired, which requires frequent pauses or slow-downs during the backup.

The RPO determines how often recovery points must be created. If the RPO is 30 minutes, a recovery point must be established every 30 minutes. The backup window defines how much time is available for the backup process. For a 30-minute RPO, for example, the backup window should be much less than 30 minutes. Likewise, if the backups can be done only between midnight and 2 a.m., it's impossible to have a RPO shorter than a day because if the backup completes at 2 a.m., the most recent recovery point would be the previous day at 2 a.m.



Many methods are available to shorten the backup window and allow more frequent recovery points, but they always involve more computing resources. The system must have enough power to run operations and complete a backup during the window. Good backup software allows you to limit the amount of resources the backup job consumes to protect response times of the running system, but this adds to the time a backup takes to complete.

## Applying facts and judgment to the RPO

Determining the RPO is a business question that requires facts and judgment. The *facts* are the cost of the downtime experienced, the cost of lost work, and the cost of providing that RPO. The *judgment* is weighing the risks against the

intangible losses, such as reputational loss. If your advertising and marketing campaigns position your company as being a low-cost provider, for example, you may be tempted to skimp on backup. But if you advertise as never going down and never losing data, even a few minutes of downtime — or the loss of even one document or transaction — could damage your business.

If you want to be able to restore to the time of any recovery point, you would have to retain an infinite number of backup files, so you may decide on a sliding RPO. For mission critical workloads, you may set the RPO as five minutes over the previous 24 hours, then every hour for the previous few days, daily for the previous month, and then monthly forever. For workloads that are less important or that don't change often, you may set the RPO daily for the first week and monthly thereafter. Remember that this is about deleting older recovery points to make space for new ones.

### ***Weighing costs, benefits, and risks***

Weighing the costs against the benefits, tempered by sound judgment of the risks, is difficult, but the process helps you understand what your RPO needs to be. It's easy to say that you can't afford to ever lose any data, but ensuring that you don't lose any data can be quite expensive. Backup doesn't provide fault-tolerant or zero-downtime operation, which requires second sites, redundant equipment, and specialized system design.

The time it takes to complete a backup job must fit into the backup window, which is also set by business requirements concerning the amount of planned downtime allowed. You have several ways to minimize the backup window without compromising the RPO, however. The easiest way to reduce the backup window is to use a snapshot (see the next section), but that process carries some risk.

## ***Taking a snapshot***

One method to speed up a backup is to minimize the data copied. The system is paused for an instant while a *snapshot* or copy of the metadata is made. This process takes a fraction of the time that copying the data would take. Then the backup is performed by using the metadata to locate the files. If changes

occur to the data during normal operations, the original meta-data is updated but the snapshot copy isn't, so the backup system won't back up any data added after the snapshot is taken. The snapshot has pointers to most of the data and only contains the actual data when that data has changed. The alternative would be to pause the system while all the data (full backup) or the data that has changed since the last backup (incremental backup) is recorded. This alternative is safer but takes much longer and will disrupt operations by pausing the running applications until the backup is complete.

Snapshots shorten the backup window considerably, and they're especially useful when you're doing a lot of updating because it's easy to revert the system to a snapshot. Also, snapshots are important in Storage Area Network (SAN) management because SANs are a widely shared resource that will disrupt much of the system if it's paused for more than a few seconds. Snapshots are safe only for short-term use, however, and managing snapshot deletion can take considerable resources.



A snapshot isn't a complete copy of the data. If the original disk is damaged, the snapshot is also damaged. Therefore, snapshots may be safe in the short term but are no replacements for backups.



Sometimes, an application can help the snapshot process shorten the backup window. VMware uses one such type, called Changed Block Tracking (CBT), to enable backup software to reduce the time to capture. Another type is Microsoft Volume Shadow Copy Service (VSS). Both the applications and the backup software must work with these technologies; otherwise, a complete backup can't be made.

## Backing Up with and without Agents

Backup programs may access data on your system in two ways:

- ✓ **With an agent:** A small backup program called an *agent* is installed on every physical and virtual machine (VM).

✔ **Without an agent:** In cloud and virtual environments, the number of VMs can grow quite large, so *agentless backup* comes into play.

Agentless backup still uses agents, but only a small number of them, so the process is easier to manage. Typically, one agent is installed for each virtual host, usually running in a VM itself. This agent can communicate with the host and can back up every VM on that host. Most systems have multiple hosts, and VMs can migrate among hosts, so the backup and capture systems must be aware of where each VM is at all times.

Agentless is a good way to operate. In special cases, when the host/hypervisor is unable to back up all the objects connected to the VM, you should install an agent to back up that machine directly. In most cases, however, the agent in the host takes care of everything.



You need to ensure that you install the right number of agents, update them regularly, and maintain a license database if you're required to do so.

## Considering Backup Products

Image backups used to be done by one application while file backups were performed by another. Today, good backup software is capable of doing both types of backups. Backup products can perform image backups in full, differential, and incremental styles (discussed earlier in this chapter), as well as use snapshots to reduce backup windows (also discussed earlier in this chapter). In addition, they can capture data at the block level and then recover files from that image.



For some network drives, as well as for some SAN and Network Attached Storage (NAS) storage systems, it's impossible to perform an image backup because the backup agents don't have programmable access to the metadata. As a rule of thumb, though, create image backups if possible, and create file-level backups only when you have a good reason not to perform an image backup.

A full-featured backup application can recover the entire image to a new system. It can even adjust to different-size disks and possibly inject drivers, as well as change bootstraps to allow

recovery to a different model of hardware with different device controllers — even different CPU types and different storage and memory configurations on physical machines and VMs.



The fastest backup software captures data only from blocks that are in use; it doesn't copy bad or unused blocks.

## *Bare-metal restore*

*Bare-metal* restore is a system with absolutely no software installed in it. Because file backups, unlike image backups, don't have all the system metadata and bootstrapping, they can't restore to bare metal. One major benefit of an image backup is that it provides both the ability to recover files to a running system and a complete restore to bare metal even when the backup wasn't made on an identical system. File backup can only recover files to a running system, but image backup is able to both recover individual files to a running system and recover everything to bare metal. Additionally, the backup software should be able to convert a physical image to a virtual image, and vice versa, and the virtual image should be exportable to any common virtualization system. Ask your vendor whether it has a universal backup format that can be recovered to physical machines and virtual machines (VM).

## *Single-pass backup*

*Single-pass backup* means that only one pass through the data is required to capture and store the backup, and only one pass is required to recover the data. Single-pass backups are faster than multi-pass backups and, therefore, afford more frequent recovery points and a shorter backup window.

If image and application backups are combined in the same product, all the data required for complete recovery can be captured in a single pass. If, however, you have an image-backup product, a file-backup product, and an application-backup product, even if they're all single-pass products, you must run three passes. Data is stored in separate archives for each product and managed separately, creating additional complexity and opportunity to fail at recovery time.



## Data-capture requirements

Here are some things to think about in deciding your data-capture requirements:

- ✓ RPO by subsystem and application
- ✓ Backup window or amount of downtime you're willing to accept for backup
- ✓ Type of backup (image, file, or both)
- ✓ Which applications are covered
- ✓ How many backup processes can be administered safely

If your image, physical, virtual, cloud, database, email, and user backups require different programs and management processes, the

resulting backup files will be incompatible, and your data may be in danger because of this multiplicity and complexity. The rapid pace of innovation in computing means that significant value can be gained from adopting new architectures or new types of hardware. Often though, the difficulty, time, and cost of migrating all your data and applications to the new systems holds you back. A backup system that can capture systems, programs, and data from physical and virtual machines and any hypervisor prevents lock-in to a particular style. Choosing a backup system that has a long life and works with any data and any style of computing should always be a major consideration.



## Chapter 3

# Storing Backups Safely

### *In This Chapter*

- ▶ Setting a storage policy
- ▶ Choosing software, hardware, and sites
- ▶ Understanding compression and deduplication
- ▶ Estimating storage costs

**T**he age-old maxim for backup has been the *3-2-1 rule*: Keep three copies of your data on two media types, with one copy stored at a separate location.

The media types used to be tape and disk, but cloud is replacing tape in businesses. Data size and the cost of networking make cloud storage economical because it's both a different media type and a separate location (and for smaller datasets eliminates much of the cost of a local backup archive). This isn't to say that tape is going away — only that cloud is being used more and more frequently as an alternative to tape. If you're starting off with small amounts of data you don't need to make large capital investments to use cloud backup.

This chapter discusses the safe storage of your backed-up data.

## *Creating a Backup Policy*

Your business may already have a backup policy, based on existing systems and practices. But if you're considering adding new technologies, you may need to update your old practices. This section gives you some pointers on setting an effective backup policy and plan.

## Backup plan

A *backup plan* is the record of what data is to be backed up on what schedule. It can be as simple or complex as you like:

- ✔ A simple backup plan may be to perform a full image backup at midnight every day.
- ✔ A more complex backup plan could be to perform a full backup weekly, with differentials every night and incrementals every four hours. The backups on machines A and B could start at 10 p.m., with a random delay of up to two hours; the backups on machines C and D could start at midnight, with a two-hour random delay; and the backups of machines E and F could start at 2 a.m., also with a two-hour random delay. (See Chapter 2 for more on full, differential, and incremental backups.)
- ✔ An even more complex plan might back up the system image weekly, the Windows Exchange system continuously, the Microsoft SharePoint system nightly, user files every other day on a random schedule, the system configuration data weekly, the virtual hosts weekly, and the Active Directory data every eight hours.

A large number of parameters should be optional, such as bandwidth limits, whether systems that have been shut down to conserve power should be awakened on local-area network (LAN) activity, and what actions are being performed on errors in the log. (If you start to get disk errors in the Windows log, for example, you could trigger an emergency backup start.)



It's important to have considerable flexibility in a backup plan, but I recommend that when you consider backup products, you give priority to those that let you set up the plan you want to use. If you're creating full image backups, for example, you should use software that allows you to create virtual machines (VMs) from image backups. (See the "Going virtual" sidebar later in this chapter.)

## Retention policy

A key part of every backup plan is the *retention policy*. Unless you have infinite space for backups, you eventually have to delete recovery points to make room for new ones. The retention policy determines which recovery points you delete next.

The most primitive retention policy is to monitor space and delete some of the older backups to make room when available space runs low. The tricky part is determining which ones to delete.

In the following sections, I discuss two common types of retention policies.

### **GFS**

Imagine that every day, you make a backup. At the end of a week, you have seven backups and are running low on space, so you take one of the daily backups, rename it as a weekly backup, and start making daily backups again. At the end of the second week, you take the latest daily backup, rename it as a weekly backup, and continue making daily backups. At any point in time, you have an entire week's worth of backups, plus weekly backups. You're still running out of space, however, so every four weeks, you rename the weekly backup as a monthly backup and reuse the weekly backups.

This type of policy is called *Grandfather-Father-Son (GFS)*. The daily backups are the son, the weekly backups are the father, and the monthly backups are the grandfather.

Follow this policy long enough, however, and you eventually run out of space, so you must delete something. What do you delete? If you can't decide, consider the following policy.

### **TOH**

My recommended retention plan is *Towers of Hanoi (TOH)*, which is a child's game where discs must be moved from one tower to another, but only one disc can be moved at a time, and no disc can be placed on a smaller one. The sequence you use to solve this puzzle is a binary pattern. This retention plan when applied to backup allows you to reuse space and place backups on various media. If all your daily, weekly, and monthly backups are on one disk drive, and both the main system and that drive fail, you lose everything. For that reason, you must rotate your use of storage media so that you're not putting all your eggs in one basket.

TOH isn't used as frequently as it should be because it's complex and hard to manage. (The details, in fact, are too complex to discuss in this small book.) Some backup products do

automate this retention plan, however, and I recommend that you look for a product that does.

Some products just delete the oldest recovery points. Although this method is simple, it isn't optimal. You should retain at least a few older backups in case of an undetected problem existing months or even years before.

## Selecting Backup Software



When you choose your backup software, look for the following high-level points:

- ✓ **Sufficient recovery features:** If you back up a SQL database, for example, you should be able to mount the backup file as a SQL database and start using it immediately. If you suspect that you backed up a disk image containing virus code, your software should allow you to mount that image as a disk and run an antivirus scan on it first.
- ✓ **No reinstallation requirements:** Vendors of backup products that can't back up and recover VMware, Windows, or Linux virtualization hosts may tell you that it's very easy to install the hypervisors from scratch. Well, the process is easy if you have all the necessary software and skills, good guidelines, and all the configuration parameters you need. Easier still is performing image backups of the virtual host so you don't need to reinstall anything; just load and go.
- ✓ **Compatibility with your existing hardware systems:** Servers used to be quite finicky about operating-system (OS) configuration; if the configuration was wrong, they wouldn't start. A lot more standardization is common today, and modern OSes do make some adjustments for hardware differences.



If you're buying hardware from different vendors and/or of different generations, you should ensure that your recovery software allows you to restore to dissimilar hardware.

- ✓ **Compatibility with your virtualization systems:** If you're performing image backups or merely backing up the virtual disks, there's no reason why you can't recover to either virtual or physical machines.



Look for backup software that, after completing a backup operation, can export a VM file and insert it into the virtual management program of your hypervisor. That way, if you need to restore to an earlier recovery point, the VM will already be there, ready to start.

## Picking Backup Media

To choose which backup media you need, first you need to determine how much storage you need for backup. Here are some guidelines to get you going. The storage needed is a function of how many copies you want to keep, how far back you need to maintain copies (this could be controlled by regulation in some industries), and how rapidly your data grows and changes. To start off, try 3 to 5 times your current data size, watch it carefully over a few months, and begin to project future needs.

If you want to be more precise, measure your own data to see what value compression and deduplication gives you. You have to watch how much new data is backed up per day for a few weeks, and consider how many backups and how old they need to be.

After you figure out how much storage you need, decide which backup media to use. Your choices are disk, tape, and cloud, each of which has pros and cons:

- ✔ Disk is fast but expensive.
- ✔ Tape is also fast but more complex than disk to track and manage.

Tape is arguably less reliable than disk because it can be damaged easily when handled. Tape, however, is by far the lowest-cost option, especially when larger amounts of data are involved.

- ✔ Cloud is great for remote end-points and for small servers.

When backup is done locally and then staged to the cloud, it's an effective solution. The local backup is used for recovery of anything other than a major disaster, and in the event of a disaster, the backup is safe in a remote location. This dual-protection of local and cloud is consistent with the 3-2-1 rules and very economical.





Make sure your cloud provider allows initial seeding where you can send hard drives to the cloud location to be copied and also can send drives to you for a large scale restore.

Which type of storage media do you choose? Often, the answer is “All of the above.” All three media types are viable in terms of cost and space.



The best practice is to use two media types, per the 3-2-1 rule (see the chapter introduction). Cloud is considered to be one media type. So unless you have petabytes’ (PB) worth of data or a very limited budget, you may find local disk and remote cloud to be a good solution. For large amounts of data, local storage on tape may be more cost-effective than disk.

In the following sections, I discuss all three media types in detail.

## Disk

Disk drives have many advantages as backup devices:

- ✓ They’re reliable.
- ✓ They’re nonvolatile, so they retain data even when the power is off.
- ✓ They’re fast, so they give you the shortest recovery time of all three media types.

### How much is 6TB?

Probably the biggest change in data has been driven by the falling cost of computing. In 1980, a single gigabyte (GB) of disk storage cost \$200,000, whereas today, 1GB of disk storage costs 4 cents. As a result, all the computing power used during World War II could fit into the sound chip in a musical birthday card, and the

total computing power used for an Apollo moon shot would easily fit into a smartphone.

So how much data can today’s largest disk drive — 6TB — really hold? A gigabyte easily holds 100,000 e-mails, and 6TB = 6,000 GB, so 6TB = 6,000,000,000 e-mails.

As disk drives increase in capacity to terabytes (TB), with the largest available today holding 6TB (see the nearby sidebar), performance may become an issue. This problem is rare with drives used solely for backup, however, so it's best to purchase drives that deliver the best cost per gigabyte. See the "Retention policy" section earlier in this chapter for tips on determining how much disk media you need.



If you're going to be transporting disk drives or tape cartridges — say, to offsite storage — it's best to make two copies of each, in case one is lost or damaged during transport.



Solid-state drives (SSDs) are emerging as backup devices because they're even faster than disk drives and much more durable. These drives, however, have some life-cycle issues that the backup software needs to manage. These issues are related to the way that data is written to the devices. Ask backup vendors whether their products take special care when writing to SSD. If not, the reliability of those products may be less than optimal.

## *Tape*

Every year, conventional wisdom claims that tape is on the way out, yet tape vendors manage to keep tapes reliable and cost per gigabyte low by increasing capacity. The current tape version, LTO-6 (Linear Tape Open), stores 5.6TB per tape. LTO-7 (16TB) and LTO-8 (32TB) are on the way. Because you can keep 560 slots of mounted tape in one standard 19-inch rack, 17PB of online tape per rack is more capacity than all but the largest data centers need.



If you're storing tapes offsite, you should make a copy of each tape you transport, due to the risk of damaging tapes during transport.

## *Cloud*

The greatest advantage of cloud storage is convenience. You don't have to concern yourself with transporting media and making multiple copies in case the disks or tapes are damaged in transport. But cloud services also have a few disadvantages:

✔ **Security:** Make sure that you know exactly how safe each cloud facility that you're considering is. Ask vendors the following questions:

- Is the facility fireproof?
- Does the facility have emergency power generators and redundant network attachment points?
- Who performs the actual backups?
- Is stored data encrypted?
- Who has access to data within the storage center?
- Is the facility staffed around the clock, or is it fully automated?

✔ **Price:** The pricing of cloud storage is tricky. In addition to charging you for the storage you use, many providers charge for data transfers.

Also, cloud backup may not eliminate your need for staff, so you may not realize a significant savings. You still need to employ people to administer your storage, plan capacity upgrades, set backup schedules, monitor backup completions, and track data use.

✔ **Network bandwidth:** You should determine the network bandwidth you need to meet your recovery-time objective (RTO; see “Recovery in the cloud” later in this chapter). Your daily backups will be smaller than a full-scale recovery.

Along with the seeming negatives, cloud and network have many positives:

- ✔ The cloud provider is going to have excellent networking connectivity, so you can access your data from anywhere.
- ✔ Using the network is very convenient.
- ✔ You don't need to worry about storing and testing media; the cloud provider ensures reliability for you and often makes redundant copies (although some companies charge more for highly available cloud storage).
- ✔ The cloud company allows you to increase and decrease capacity very quickly.



## Choosing Offsite Storage Locations

The 3-2-1 plan (see the introduction of this chapter) advises you to keep three copies of your data: the running system, local backups, and offsite backups. You can physically transport disk or tapes to an offsite location (or use a vendor that picks up the media for transport), or you can transfer data wirelessly to be written on a networked storage device. In this section, I cover a few options.



When you keep one set of backups with your running machines and a second set at an offsite location, consider how far from your site the offsite location needs to be. Most people want to be able to drive to the offsite location and back within eight hours, so that's a good maximum distance. If your area is at high risk for natural disasters — earthquakes, floods, tornadoes, or hurricanes, for example — you may want to keep your data even farther away.

### Online network

If you're considering using an online network as your second location, ask providers about data security and network bandwidth. Also look at how much of your data is changing. The industry norm is 5 percent, but your volume could be higher or lower. Databases change frequently, whereas application code doesn't.

Speed is yet another consideration. Suppose that you have a small business with 10 users at 5GB each and 2 servers at 10GB each, or a total 70GB per day and 1.4TB for the initial backup. Also suppose that your network speed is 100Mbps. Your daily incremental backup takes 90 minutes plus any network overhead and delays in routing.



Compression can make a big difference. If the data is compressed to half its size, only half the data has to travel on the wire, yielding a major improvement in speed. (I discuss compression in more detail in “Considering Compression and Deduplication” later in this chapter.)

## *Dark site*

If you don't need or can't afford to have two running sites, the next option is a *dark site*: a remote computer room, usually not operational, with a minimal amount of equipment. Generally, all backups are sent over the network to storage devices running there. Periodically, a company's dark site is lit up and run to make sure that the disaster-recovery plan actually works.

## *Cloud backup*

Cloud backup is quickly gaining popularity among companies that want to have onsite data centers and are willing to use a cloud for temporary needs and disaster recovery. It's becoming easier to use the cloud as a second site, storing backups in the cloud, and recovering them to servers in that cloud.

### *Recovery in the cloud*

If you decide to use a network/cloud solution, ensure that you have enough bandwidth available to meet your RTO goals. The first backup to the cloud can take quite a while because everything is being sent on the wire, so your cloud facility should let you ship initial seeding disks to reduce the time for the first backup. That way, if you ever need a full recovery, your cloud-storage vendor can ship your data back on disk drives.

RTO specifies how long you're willing to be down. If your recovery-point objective (RPO; see Chapter 2) is four hours and your RTO is two hours, you'll be up and running two hours after the failure, but you may have lost as much as four hours' worth of data, depending on the age of your last recovery point. If you want to be operational and up to date in two hours, your RPO and RTO together must add up to two hours. RTO generally applies to a catastrophic failure of a subsystem or the entire system. Your order-intake department may have an RTO of 5 minutes, for example, but payroll may have an RTO of two days.

### *Public or private cloud*

One big decision is whether to use a public cloud facility. Many backup vendors offer cloud storage that's optimized for backups, and major public clouds also offer storage online. A

big advantage is that capacity can be increased easily when needed.

The fee structure for public cloud services can be confusing. Some services charge a fee based on capacity per year; others charge for capacity per month, plus a fee for data transferred from storage to the Internet. Cloud companies may have additional charges, such as data-deletion fees. Finally, make sure that you check out the security features.

## Considering Compression and Deduplication

*Compression* — the process of making files smaller by using various algorithms to substitute abbreviations for repeated information — is useful for local copies of backups, and most storage specs that you see assume compression. Compression works when the data contains a lot of predictability. If your company name is Acronis, for example, and that name appears often in text, compression notices that fact and creates an abbreviation for the name.



Some data, such as pictures and video, is already compressed, so it can't be further compressed by backup. Encrypted data by its very nature shouldn't be predictable, so you can't compress it either. In fact, if you *are* able to compress encrypted data, you should be very suspicious of your encryption system.

*Deduplication* (also called *dedupe*) is similar to compression. If you back up 1,000 system images of standard corporate laptops, for example, you find the same OS files over and over. High amounts of duplication require a lot of space that can be saved through the use of deduplication. Dedupe works by keeping one copy of the original data and inserting pointers into each set of backup data that contains the duplicate data. If deduped data contains thousands of characters and the pointer is only 20 characters long, the savings can be huge.

De-duplication works remarkably well when you have a lot of duplicate data. You can use multiple methods — one being de-duplicating on the fly, which reduces the data that needs to

be written to the backup archive. The other is a post process that de-duplicates the archives after they've been created. You need to consider a number of performance and space tradeoffs so if you are counting on de-duplication to save a lot of time and space, make sure you test it well and compare the various options. It may be better to just reduce the amount of duplicate data you're creating.

## Calculating the Costs

It's very difficult to calculate an accurate cost for backup storage, but a few rules of thumb may help:

- ✓ At the raw storage level, the cost of disk is 3 or 4 cents per 1GB, and tape costs roughly 1 cent per 1GB.
- ✓ Cloud vendors charge a monthly price but usually have a data-transfer cost as well. At this writing, for example, Amazon.com charges 2 to 3 cents per month for storage and 12 cents per 1GB of network transfer per month for data going from Amazon to the Internet.
- ✓ Cloud prices include the costs of space, air conditioning, controllers, racks, and power. When you add those items yourself for local storage, the total cost can be 5 to 20 times the raw storage cost, depending on your location and how much storage you have.



Over a three-year period, tape is the cheapest backup method, followed by disk, and cloud is the most expensive. But from a cash-outlay perspective, you can start using the cloud very inexpensively and pay as you go, whereas with disk and tape, you need to purchase most of the infrastructure right away. My best advice is to ensure that your backup software works with disk, tape, and cloud, as well as with physical, virtual, and cloud systems, and then choose the most economical system for you while obeying the 3-2-1 rule. As long as your backup works with all the popular storage media, including the cloud, you can adjust media strategy when it makes business sense to do so.

# Chapter 4

## Recovering Data

---

### *In This Chapter*

- ▶ Knowing when you've lost data
  - ▶ Putting your recovery plan to work
  - ▶ Striving for simplicity
- 

**D**ata needs to be recovered for many reasons. The most common is human error, such as accidentally deleting an e-mail or copying an old version of a file over the new version. Other reasons include viruses and malware; sabotage by disgruntled employees; hardware issues in disk drives, controllers, and networks; and software bugs in applications and operating systems (OSes). Sometimes, data is lost simply because no one can remember where it was filed.

*Data recovery* is how you get your data back, whether the lost data is as small as a phone number or as large as a complete business destroyed by a flood.

In this chapter, I show you how to create a data recovery plan that will serve you well if the worst happens. With proper backups, you can be up and running again quickly.



If you've come to this chapter because you lost everything and don't have a backup, I can give you little advice. Some companies can try to pull the data from your disks if the loss is due to an electronic failure inside the drive and if the disk and heads aren't physically damaged by rain, heat, or dirt. Also, some programs available on the web may be able to rescue your data if you deleted it accidentally or reformatted a disk before backing up. But if you securely erased the disk, if the disk is encrypted and you don't know the password, or if you wrote on the disk quite some time after the accidental erasure occurred, all I can offer is my sympathy.

## Recognizing Data Loss

Data loss isn't always obvious. Sometimes, it's confused with hardware failures, software bugs, low memory, or insufficient storage. Here are a couple of scenarios:

- ✓ **The system boots after a crash, but the applications crash.** You can look at the applications' logs and error messages, and maybe you can consult system monitoring tools, but you probably have a deadline for getting the system back up. If you're having problems with several applications, full recovery may take time. So you need to determine whether the problem is, say, a recent patch that caused your applications to malfunction or whether the data that your applications are using is corrupt.

Often, it's easier to recover the system into a virtual machine (VM) — or to mount the backup as a virtual disk and do some quick compares — than it is to diagnose the problem. But after everything is running, you should try to diagnose what happened to prevent a recurrence.

- ✓ **Everything is running after a crash, but you're getting data-corruption error messages.** You may not be sure whether you have data loss and need to recover. Because the system is running, you can perform some queries to see what happens. If you notice that the corruptions fall within a certain date range or concern only one type of operation, for example, your database may have a corrupted table. Again, it's often easiest to just recover everything as long as you have an up-to-date backup.

One common mistake is trying to repair the damage, thereby using up a lot of time that could have been better used in recovering to the last known good point.

You will suffer data loss. It's not an "if" but a "when" kind of thing. While many cases of data loss are minor and contained, they occur more frequently than most people realize. Smaller companies report needing to recover data once or twice a week. Maybe some user accidentally deleted an e-mail and needs it back, or another user can't find a presentation from a few months ago. If you can't recover the data, then you need to redo work that can take hours. Or maybe decisions get made without all the data.



Recovery is more complicated if an entire system becomes corrupted during an update, but if you have a tested recovery plan in place, the task of recovery is far from impossible. And unfortunately, there are more problems with updates than you realize.

Also, although most of your recovery actions are related to user mistakes that involve only a file or two, you need to be ready to call for a full system restore when doing so makes sense. It's better to have good backup procedures that allow you to recover easily than to spend hours trying — and failing — to get nonexistent data back from the system. Albert Einstein defined insanity as “doing the same thing over and over again and expecting different results.” We've all been in denial that data is lost and have tried over and over to restart the system.



The best advice for recovery is to practice recovery procedures frequently. Practice proves that you know how to recover, that you have the materials and software you need, and that your backup plans are working. Practice also gives you the confidence you need to recover quickly and effectively.



Be careful that you don't hurt your backups. Sometimes, when a system is corrupting disk drives, users claim that they don't have time to restore, so they boot from a backup image. A few minutes later, however, they discover that their backup image becomes corrupted by the same problem. In other cases, users mount backups and decide to reformat the system drive before restoring, just to make sure, but they're so nervous that they accidentally reformat the backup drive.

## *Putting Your Recovery Plan in Motion*

I hope that you have a written set of recovery procedures that you practice on a regular basis. This plan will make it easy to recover in the case of data loss.



## Keep things simple

I want to warn you about complexity. The simplest thing to do is choose one vendor that can supply backup and recovery for physical, virtual, and cloud storage, as well as for Windows, Linux, and granular application recovery, and do it all with image backup. You'll be better off than if you choose one product to recover Microsoft Exchange, another for SQL, a third for bare-metal recovery, yet another for

file backup and recovery, and so on. Working with multiple vendors and systems can get confusing and can provide less interoperability, and you'll have to update and change your procedures much more frequently.

Try to keep your backup environment comprehensive and complete but as simple as possible by using a minimal number of vendors.



Some backup software can print a recovery procedure for you. It's very easy to forget which tapes or disks contain which files, especially when you're rushed and under pressure, and printed backup instructions help keep you on track.

If you lose data, recovering that data is a fairly simple process when you have a recovery plan. Follow these steps:

- 1. Boot up the host system that you want to place in operation with the backup program.**
- 2. Recover the hypervisor and possibly the VMs to the disk on the host.**
- 3. Boot the host.**
- 4. Start the VMs, or recover them from another backup set and then start them.**

If you've practiced these steps, they shouldn't take long to complete and should be successful.



You should appoint a recovery leader to take charge of any recovery efforts. If too many people are trying to help and taking conflicting steps, the recovery is likely to fail. You may also choose a vendor that provides an active restore capability that allows the system to run as soon as enough of the data has been recovered. Then the recovery completes while the system is running. This is a very useful feature.



# Chapter 5

---

# Managing Backup

.....

## *In This Chapter*

- ▶ Staying up to date on technology
  - ▶ Knowing what and when to back up
  - ▶ Planning your work and working your plan
- .....

**T**he secrets to successful backup and recovery are developing the proper habits and being meticulous in your work. When things fall apart, you're the last resort. Accordingly, you train and practice for a day when everything fails at the same time.

Your job is simple and twofold:

- ✓ Ensure that a great backup plan is available and track its execution.
- ✓ Ensure that a great recovery plan exists and verify its efficacy.

In this chapter, I show you how.

## *Keeping Current on Backup Products*

The first task is staying up to date with the backup products you use, as well as with the hardware and software being used in your company. This task itself is a lot of work, not only because the products themselves gather new features as new versions are released, but also because the way that information technology (IT) is used changes.

Virtualization, for example, has grown from a small part of a system that allowed lightly loaded servers to be consolidated to a major (and growing) component of data centers that allows much greater resource sharing and use. The growth of virtualization alone has led to many new features and capabilities in backup products. Things to watch for are the expanding use of software-defined storage, the Internet of Things, and the increasing merging of development and operations (DevOps). These emerging technologies demand new capabilities from backup software.



**TIP** You should have a long-term backup strategy that's congruent with your company's IT strategy. Increased use of software as a service (SaaS), increased use of virtualization, or expansion to multiple locations may be the basis of an IT strategy, and the backup paths chosen should be able to move to new IT structure as well.

## Setting the Backup Window

The next task — and perhaps the hardest one — is determining the recovery-point objective (RPO), fitting it into the backup window, and meeting the recovery-time objective (RTO). (I cover RPO in Chapter 2 and RTO in Chapter 3.) To do this, you need to determine the following:

- ✓ The major applications you're protecting
- ✓ The amount of data (current and projected) associated with those applications
- ✓ The applications' RPO and RTO

This information allows you to determine how much time you can afford to give backup. Unfortunately, some short pauses may occur during backup, and applications may run more slowly. You may decide to offset these problems by pushing backup to less-busy times, but the busy times are when the data changes most frequently and needs protection most.



**REMEMBER** There's no single answer for the amount of time backup should take. Simply ensure that you have enough capacity and computing resources to run normal operations and backup simultaneously.

## Creating and Checking a Backup Plan



If you want short recovery times, frequent recovery points, and long retention, try to be clever. Here are a few ideas:

- ✓ Hybrid local and cloud storage (see Chapter 3) can help minimize the data that has to be moved to a second location.
- ✓ A Towers of Hanoi (TOH) retention plan (see Chapter 3) can ensure that you have frequent and recent recovery points, as well as some older recovery points in case you discover silent corruption that may have been in the system for a while.
- ✓ Deduplication and compression can minimize storage requirements (see Chapter 3).
- ✓ Image backups, along with appropriate use of incremental backups and consolidation, can reduce storage requirements without sacrificing RTO (see Chapter 2).

Also, you should have one backup plan. Your backup vendor should offer a single management console that can take a master backup plan and customize it for each system, install it on the system, monitor its progress, and check for errors.

### *Keeping it simple (or not)*

A backup plan can be as simple as this: Take a full backup of everything every night at midnight. It can also be as complex as the following:

For order intake, take a full backup weekly and an incremental backup every hour. For inventory control, take a backup of just the database every 15 minutes. For the manufacturing controllers, take a full image backup every four hours. For all the user endpoints, take a full backup monthly, dedupe at the source, and take incremental backups with encryption and compression of each user directory every 12 hours, but randomize the times so they don't all occur at once.



The more complex a company's RPO and RTO are, the more complicated the backup plan is, so it's important to use the smallest possible number of backup vendors.

## *Setting backup windows*

The *backup window* is the period of time when the system can be down or degraded to allow backups to complete. If your manufacturing company runs two shifts, for example, an eight-hour period can be considered to be the backup window. If the company operates around the clock, however, all backups may have to be done while the system is running.



Backup technology is improving, but enacting perfect zero-backup-window processes is difficult. Sometimes, using several short backup windows for different workloads can solve the problem. At other times, a strong centralized management console can automate and optimize backup windows.



The best vendors offer a management console that helps you put the plan together and schedule it so backups don't occur at the same time and consume too many resources. When a new server or workload is added, you can quickly add it to the backup plan. When you define the backup plans, you allocate the right number of storage vaults and archives to take advantage of deduplication and ensure that the backup writes are spread across enough units to share the load.

## *Checking execution*

Tracking the execution of a backup plan involves checking for errors to ensure that no backups failed or determining the cause of any failure. Running out of disk space is the most common cause of backup failure, with network-connectivity issues running close behind.

## *Monitoring the plan*

When the backup plan is created, daily backups are working, capacity management is in place, and network performance is established, you have only two more key tasks:

- ✔ **Watch for changes.** Today's flexible virtual and cloud data centers allow users to split workloads easily, and new virtual machines (VMs) may pop up at any time. Backing up those VMs and understanding any file or synchronization constraints on them are major responsibilities of backup management.
- ✔ **Maintain a census.** Keep a census of systems, disks, and archives to ensure that everything that needs to be backed up is backed up.



## Chapter 6

# Ten Things to Know about Backup

---

### *In This Chapter*

- ▶ Understanding costs
  - ▶ Setting priorities
  - ▶ Knowing what to do when
- 

**W**hen your backups are consistent and verified, and you've practiced recovery procedures, even a 2 a.m. call shouldn't throw you for a loop. You'll be prepared and confident. This chapter lists ten things you should know to make backup and recovery easier.

## *The Value of Your Data*

Your company has many forms of data. Some data changes slowly; some changes rapidly. Some is tied to a sale; some is tied to a product or service; some is tied to financial reporting, marketing, or human resources. Knowing how important each type of data is and how often it changes helps you determine the recovery-point objective (RPO) for that data.



Usually, companies enumerate the RPO by workload or application. For more information on the RPO, see Chapter 2.

## The Cost of Downtime

Sometimes, this calculation is easy. If a manufacturing system isn't running, you can compute the cost of the idle workers and the value of the products that aren't being made. The story is different in the case of a failed airline-reservation system, however. If the system can't sell seats, the airline may lose existing customers, its reputation, and potential customers, all of which represent value.

Downtime costs are business-dependent, of course, but in every business, data has value, and so does uptime. That's the reason to have backup processes in place.



A recent IDC *Disaster Recovery Survey* (May 2014) sponsored by Acronis shows that over 90 percent of companies report downtime costs in excess of \$20,000/hour, while close to half report costs above \$60,000/hour.

## Workload Priorities

In the event of a total loss, you should be able to prioritize recovery. Consider the following items:

- ✓ The order in which workloads should be brought up
- ✓ Which workloads should have redundancy and failover
- ✓ Which workloads can wait a few days and which can't
- ✓ Which workloads must be stopped to give their capacity to workloads that have failed

## Where Your Backups Are Stored

The best practice for storage locations is the 3-2-1 rule: three copies, two media types, and one copy stored remotely (see Chapter 3). Ideally, you should have the running system, an offline but local copy, and a copy in a remote location. Be sure to consider safety and security.



## *How Long to Keep Backups*

Storage space is limited, so at some point, you'll need to delete backups according to the company's stated retention policy (see Chapter 3). Consider three things:

- ✔ **Legal requirements:** Some companies are required to maintain certain records for a certain period. These requirements may be legislative (in regulated industries) or contractual (between the company and its customers).
- ✔ **How often files are needed:** Your business may typically exchange e-mail and files with a customer for three months; then the job is over. In that case, you may want to keep backups for only four to six months.
- ✔ **Versioning:** Files may go through many revisions, and it may not be important to keep each revision. When the project is over, the final versions are kept in backup, but not all the intermediate versions are needed.

## *What Recovery Tools to Use When*

To recover, you need a combination of hardware; operating-system (OS) backups, patches, and updates; applications; configuration data; and (of course) the data needed by the application. If you're performing full image backups, all this data is in the image, but if you're using file backups, you must collect and update them.

## *The Details of Your Backup Plan*

You need documented recovery procedures for several reasons:

- ✔ If the CEO calls in the middle of the night because he accidentally deleted a file, he needs to know who to call.
- ✔ The person whom the CEO calls needs to know how to recover the file.
- ✔ In the event of a more extensive failure, the technical staff needs to know where to find the backup and what servers need to be recovered, as well as how.

## *What Data Is Excluded from Backup*

When space gets tight, administrators get clever with excluding unnecessary data. Generally, you don't find out until too late.

It's fine to exclude files that can be re-created easily, but if your list of files excluded from backup is very long, make sure that you're not saving a few dollars of disk space by setting yourself up for days of installation work.

## *How (And How Deeply) to Test Backups*

If your company has a good test plan for verifying performance and patches, you can use a similar test to verify that backups have been restored properly. If you have spare servers or spare capacity on a virtual host, it's a good idea to automate the testing of backups. At minimum, you should recover the backups, run disk verification routines, and compare file sizes.

## *How to Frame Backup Questions*

To answer any question about backup, turn it into a question about recovery. Instead of asking yourself what media to use or which retention plan makes the most sense, for example, you should ask one of these questions:

- ✔ "Which media allows me to recover fastest?"
- ✔ "Which media recovers most reliably?"
- ✔ "Will a GFS retention scheme allow me to recover from old backups?" (I cover Grandfather-Father-Son policies in Chapter 3.)
- ✔ "Which retention plan affords the fastest recovery?"



There's no point in optimizing backup if doing so will cause problems with recovery.

# Acronis

## Business Backup & Storage Management

### Products and solutions for any business environments

Acronis delivers the best new generation data protection for virtual, physical, mobile and cloud environments

Powered by **the Acronis AnyData Engine**, Acronis offers the easiest to use and most complete solutions for backup, disaster recovery and secure access. As a world leader in data protection and system management, we offer a wide range of technological advantages:



#### **Award-winning backup technology**

that captures *everything* in one simple step.



#### **Ultra-fast disaster recovery**

that will get you back up and running in minutes



#### **Flexible data recovery**

that can restore anything—from a single file to an entire server



#### **Multi-destination support**

that can store backup data anywhere (including the Acronis Cloud)



#### **One-step system deployment tools**

that make system maintenance fast and cost-effective



#### **Smart disk management utilities**

that streamline IT tasks and boost system performance



#### **Software designed for exceptional ease of use**

that requires almost zero learning curve



#### **Centralized management**

and easy reporting via a single, easy-to-use console

Download the Acronis Backup and Recovery Software Free Trial today.  
Visit [www.acronis.com/en-us/backup/free-trials](http://www.acronis.com/en-us/backup/free-trials)

Acronis

## Never lose a file again!

Times are changing with regards to how you protect your data. Today's data is stored on physical servers, desktops, laptops, virtual machines, and the cloud. As these new technologies appeared, the techniques for protecting information and recovering from disastrous data loss have had to change, too.

- *Modern backup and recovery — understanding new generation data protection*
- *Creating a backup plan — key points and best practices*
- *Virtual, physical, and cloud — protect any environment*
- *Recover any data — files, applications, and entire systems*

**Joel Berman** has worked in IT for over 40 years. He has worked with the world's largest financial institutions and telecom companies, ensuring the reliability and integrity of their information technology.



**Open the book and find:**

- **A data protection primer**
- **How to capture data for backup**
- **How to store your backups safely**
- **Ways to recover your data**

**Go to [Dummies.com](http://Dummies.com)**  
for videos, step-by-step examples,  
how-to articles, or to shop!

# **WILEY END USER LICENSE AGREEMENT**

Go to [www.wiley.com/go/eula](http://www.wiley.com/go/eula) to access Wiley's ebook EULA.