Making Everything Easier!™

Quantum Special Edition

VM Data Protection

FOR DUMMIES®

Learn to:

- Identify your VM protection needs
- Ask the right questions of potential vendors
- Choose the VM protection solution that's right for you

Brought to you by

Quantum

Faithe Wempen

About Quantum

Quantum Corp. is a proven global expert in data protection and big data management. From small businesses to multinational enterprises, more than 50,000 customers trust Quantum to solve their data protection, retention, and management challenges. Quantum's best-of-breed, open systems solutions provide significant storage efficiencies and cost savings while minimizing risk and protecting prior investments. They include: DXi®-Series disk-based deduplication and replication systems for fast backup and restore, Scalar® tape automation products for disaster recovery and long-term data retention, StorNext® data management software and appliances for high-performance file sharing and archiving and vmPRO™ solutions for protecting virtual machine data.

Quantum Corp., 1650 Technology Drive, Suite 800,
San Jose, CA 95110, (408) 944-4000, www.quantum.com

For more information, visit www.quantum.com.

VM Data Protection FOR DUMMIES®

Quantum Special Edition

by Faithe Wempen

John Wiley & Sons, Inc.

VM Data Protection For Dummies®, Quantum Special Edition
Published by
John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030-5774
www.wiley.com

Published by John Wiley & Sons, Inc., Indianapolis, Indiana

ISBN: 978-1-118-19464-5
eISBN 978-1-118-19669-4

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents

Publisher's Acknowledgments

We're proud of this book and of the people who worked on it. For details on how to create a custom *For Dummies* book for your business or organization, contact `info@dummies.biz`. For details on licensing the *For Dummies* brand for products or services, contact `BrandedRights&Licenses@Wiley.com`.

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

Acquisitions, Editorial, and Media Development

Project Editor: Beth Taylor

Editorial Managers: Rev Mengle, Jodi Jensen

Acquisitions Editor: Kyle Looper

Business Development Representative: Karen Hattan

Custom Publishing Project Specialist: Michael Sullivan

Composition Services

Project Coordinator: Kristie Rees

Layout and Graphics: Claudia Bell

Proofreader: Melanie Hoffman

Publishing and Editorial for Technology Dummies

Richard Swadley, Vice President and Executive Group Publisher

Andy Cummings, Vice President and Publisher

Mary Bednarek, Executive Director, Acquisitions

Mary C. Corder, Editorial Director

Publishing and Editorial for Consumer Dummies

Diane Graves Steele, Vice President and Publisher, Consumer Dummies

Ensley Eikenburg, Associate Publisher, Travel

Composition Services

Debbie Stailey, Director of Composition Services

Business Development

Lisa Coleman, Director, New Market and Brand Development

Introduction

Virtualization is rapidly changing the way business IT operates, from small local businesses to multi-national corporations. If you are reading this, chances are good that your company is already taking advantage of virtualization's benefits.

Virtualization means that a single underlying piece of hardware, such as a server, runs multiple guest operating systems to create virtual machines, or VMs, with each of them being oblivious to the others. An administrative application, such as VMware, manages the sharing process, allocating hardware resources, memory, and CPU time to each VM as needed. And all applications look at this software construct exactly as if it were a real, physical server — even the VM thinks it's a real server!

Virtualization makes good financial sense. It enables a single server to offer multiple capabilities that otherwise would require separate servers. It includes native high availability features so you don't have to use any more complex clustering tools. This ability to combine capabilities means that you spend less money on server hardware and much less time providing IT support, systems administration and maintenance.

Some of the key benefits of virtualization include:

- Easier administration, because multiple server functions are combined in a single administrative interface
- Reduced hardware costs, because you need fewer physical servers
- More complete utilization of the servers you do have, with less waste of CPU time
- Reduced maintenance costs, because fewer servers are there to break down
- Better application availability and flexibility, because virtualization brings a cost-effective failover protection

against hardware failures and can allocate resources when and where you need them.

- Less energy usage, because you have fewer servers using electricity and generating heat
- Reduced IT support costs, because fewer IT support people are needed to support the hardware

However, the virtual environment can add to the complexity of your backup and other data protection processes, because existing data protection applications and appliances are not designed to work with VMs. Trying to use old protection solutions with the new technology of virtualization can result in data not being adequately protected, as well as considerable waste of IT resources and backup storage space.

What can you do about that? Find out by reading this book.

VM Data Protection For Dummies is designed to help IT decision-makers figure out how to handle virtualization data protection in the most cost-effective and efficient way. In this book, I explain in clear, concise language how virtualization changes an organization's backup needs, and how traditional backup solutions will — or won't — work with that technology. Then I tell you about a set of innovative solutions from Quantum that solve many of the problems inherent in virtual data protection.

How This Book Is Organized

As with other *For Dummies* books, this book doesn't assume that you'll begin on page one and read straight through to the end. Each chapter is written to stand alone, with enough contextual information provided so that you can understand the content.

Chapter 1: Protecting Virtual Data

Chapter 1 introduces you to the basic concepts of data protection. You find out how snapshots and backup software

have traditionally been used to protect virtual data — and you discover why they are not always optimal for that purpose. I also explain how VMs complicate data protection because of their dynamic nature and their high degree of data redundancy.

Chapter 2: Shopping for a Virtual Protection Solution

Chapter 2 covers the technical capabilities that you should look for when evaluating virtual data protection solutions. This field is rapidly changing, so it pays to ask detailed questions about what capabilities the various solutions offer. Here I explain the key qualities of an effective system, and help you ask the right questions to determine which solutions have those qualities.

Chapter 3: Making Virtual Protection Easy

Chapter 3 introduces the Quantum set of vmPRO protection solutions. This section describes four underlying technologies that form the foundation for all the vmPRO solutions, and it shows how they make protection easier, faster, and less expensive for small sites (remote offices, small businesses, and so on) or single application private clouds and for corporate data centers.

Chapter 4: Ten Frequently Asked VM Protection Questions (And Their Answers)

Chapter 4 provides concise answers to come of the most common concerns and questions that people have when considering virtual data protection.

Foolish Assumptions

This book assumes that you understand the basics of virtualization, and that you probably even have virtual machines running in your organization. However, the book *doesn't* assume that you know anything about the protection requirements and challenges of virtual environments because that's why you are reading this book. I define all the new terms you encounter, and I also make it a point to thoroughly explain complex topics so that you can understand how it all fits together.

Icons Used in This Book

The Tip icon points out helpful information.

The Remember icon marks important facts that are worth adding to your memory.

This icon gets technical and explains interesting but not required details.

Where to Go From Here

Just start reading! You can use the Table of Contents as a guide, or my description of the chapters in this Introduction. If you already understand the data protection challenges you're facing and want to skip straight to the solution, start with Chapter 3!

Chapter 1

Protecting Virtual Data

In This Chapter

- Determining what constitutes protection
- Understanding how VMs complicate the protection process
- Exploring methods of protecting virtual data

Virtualization is one of the most exciting technologies in IT today. More companies are deploying it every day from small mom-and-pop businesses to government departments to huge multinational corporations. Virtualization brings with it reduced hardware acquisition and support costs, reduced IT staffing needs, and many other bottom-line benefits. For many companies, adopting virtualization is a no-brainer.

As the saying goes, though, the devil is in the details. The benefits of a solid virtualization implementation are tempered, however, by the increased complexity and sophistication required to achieve effective consolidation and optimal performance. Modern virtualization platforms utilize a low-level software control system called a *hypervisor* in order to share access to the physical servers and their data. In the virtualized datacenter model, intelligent interaction with the hypervisor and associated tools form the new standard for system integration.

The tried-and-true data systems that people developed for backing up physical servers don't work efficiently in virtual ones—in some cases they cause decreased overall performance, and in other cases they don't work at all. Therefore, best practices for protection usually have to evolve for virtual environments and for mixtures of physical and virtual servers. This problem is not always obvious at the beginning and often

sneaks up on the IT staff. The reason is that many organizations start with just a few virtual machines (VMs) and extend their existing protection systems to the VM without changing IT practices. Although this may work for a while and for a few systems, taking this initial shortcut can create problems. The protection may be inadequate and inefficient, especially when the number of VMs and the data on them reaches a larger scale.

In this chapter, you look at some of the unique challenges in protecting VM environments as well as some of the methods that IT departments use to protect virtual data.

Determining What Constitutes Protection

Employees are a company's most important asset, and surely a company's IT assets are a very close second.

Most companies today rely on IT systems for almost every part of their business processes. When a company's IT system goes down, commerce grinds to a halt. Salespeople can't input orders. Warehouses can't print shipping labels. Executives can't make data-informed decisions. Accounting can't receive payments. The whole profit-making chain falls apart.

IT assets include computer hardware and peripherals, plus the operating systems they run, the data that they store, and the capabilities they enable. As an organization begins using virtual machines (VMs) as part of their IT systems, those VMs become part of that critical-to-protect pool of assets.

One of the reasons that VM protection issues can sneak up on companies is that in the early days of VMs, you could typically see them deployed in pockets of the organization development or test, for example, that aren't always directly in the time-critical path for operations. Sometimes those groups are responsible for their own backups outside the corporate processes. Initially, because there are fewer machines and they're on the edge of the enterprise, the protection problems aren't so visible. But as VMs grow up, and take on more core applications and data, the backup rock in the road can't be ignored any longer.

Covering your bases

That plan should include getting the data back, restoring the functionality, and doing it in a timeframe that prevents further losses.

Here are some example scenarios. Are you covered?

- A flood makes your entire building inaccessible for several weeks, and several of the physical servers on which your VMs are hosted are water-damaged beyond repair. Will you be able to completely and quickly restore your VMs on new hardware, in a new location?
- One of the employees has accidentally deleted a very important customer file stored in a VM. Without taking the VM out of service, can you get the file back quickly enough that the Sales department can continue to interact normally with the customer? Can you do it without having to restore a full copy of the VM?
- The VM that runs your website's online shopping cart crashes and won't reboot. A file check shows that one of the VM's files is corrupted. Can you quickly restore a recent backup of the entire VM?
- The physical hard disk on which several VMs are hosted has a mechanical failure, and is completely unreadable. Can you get the VM management software and each of the VMs up-and-running quickly on a new hard drive, with minimal data loss?

It's important that VMs and their contents are fully protected against every type of potential problem, from a single database record getting accidentally deleted to the entire corporate campus being destroyed by a natural disaster and everything in-between! So, I want to start by clarifying what it means to be *fully protected*.

Having full protection means that whatever IT loss scenario anyone can dream up, a plan is in place for dealing with it.

Generally speaking, total VM protection means that you can quickly, easily, and reliably do all of the following:

- Recover all IT data assets even when the entire hard drive, server, or building has been destroyed

- Retain files for as long as the organization requires them to be available whether that's a few hours, a few months, or many years
- Recover specific files, including older versions, without having to rebuild the entire VM from your backup files
- Restore entire VMs on new hardware rapidly
- Maintain a well-ordered backup system that meets any industry-specific laws or guidelines for retaining historical records
- Recover or replace corrupted, damaged, or missing VM files
- Minimize downtime and performance degradation of the primary applications while the data backup and restore operations are taking place
- Minimize administrative overhead

If your current IT protection system can't support VMs in all those ways, then maybe it's time to take a close and critical look at how it can be strengthened and improved.

Understanding How VMs Complicate the Protection Process

VMs have historically been used mostly in test or development environments, where the data they contained was not viewed as mission-critical for the company or at least it didn't slow down near-term production if there was a problem. If a VM crashed or lost its data . . . then it did. Groups recovered, someone might have to redo some work, no big deal. IT departments could afford to take the path of least resistance, letting the departments set their own backup policies, often employing the cheapest and most basic data protection solutions for its VMs.

But this scenario is no longer true. More and more companies are putting their important production and operations data into VMs, so VM backup has become a more critical issue, one

that has to be covered by corporate policies and practices. Companies that rely on VMs for their daily operations can't afford for VM data to be lost or for it to be unavailable for an extended period of time.

The problem is that when traditional backup systems and policies were designed, they didn't have VMs in mind. Yes, a VM environment can function like a regular server but only up to a point. It may seem like a regular server to the end-user interacting with it, but from a behind-the-scenes perspective, a VM is very different. And this fact makes finding an ideal data protection solution for a VM much more complicated than finding one for a physical machine.

They're dynamic

VM content is by its nature dynamic for both the VM itself and the data within it. All the things that happen on a physical server happen inside the VM, but VM content is much more than just the data files in the applications that run within it. A VM's content also consists of the operating system files, the logs, the services, and all of the files that allow the VM to fool all the applications into thinking that it's a server like every other. So you always have two layers in a VM — the whole VM itself and the individual application and data files that run on the virtual server.

By design, VMs are easy to set up. You can set up a new VM for a test project in just a few minutes, and in many companies, VMs spring up in multiple departments and locations every week wherever there's an IT person or other techie who needs to try something out before deploying it on a large scale. VMs are easy to transport from one physical location to another. For example, in VMWare, the vMotion tool makes VM movement from one physical server to another physical server a simple matter.

The dynamic qualities of VMs are part of what makes them useful, but those qualities also make them difficult to protect. VMs are so easy to set up that anyone can do it, even people who have no awareness of the importance of backups. With physical servers, the IT department can keep a close eye on what gets set up and where, but it frequently loses that tight level of control for VMs.

They're redundant

VM files can contain a lot of redundancy. A *lot.* For example, suppose that you have 50 VMs set up, all of which run the same basic OS, but with different applications. Therefore, you have 50 almost-identical copies of a lot of the OS system files, and each time you back up your VMs, you back up the exact same data 50 times. Not only is that a huge waste of backup media, but it's also a waste of the computing power required to do all those backups day in and day out.

VMs can also contain a lot of data that you don't need to back up. For example, each VM's OS generates temporary files as it operates, and these often don't get deleted as they should (because of crashes, bugs, and so on) and if they get deleted inside the file system of the guest VM, they often hang around for a much longer time in the file set that makes up the larger VM package.

VMs also typically reserve more data space for use than they actually need. This over-provisioning takes place to minimize admin time as the system grows, but until growth happens, you've got extra blank space dedicated to the VM. The host PC on which the VM is running treats that claimed-but-not-used space as being already in use, and usually has to back it all up because it shows up as part of the entire package that makes up the whole VM.

VMware has developed various methods of shrinking down the amount of redundancy within VMs, but there are still lots of over-provisioning and lots of expired, out of date, junk data. When backing up a VM, if you don't use a backup system that intelligently looks for and excludes that kind of empty data, it ends up in your backup sets, increasing the size of the backup data and the amount of time it takes to execute.

Looking at Traditional Methods of Protecting Virtual Data

Most companies don't jump into wide deployment of VMs right from the start; they gradually ease into relying on VMs, adding

new VMs and new storage locations for them whenever it makes sense to do so. Therefore, if a company's VM usage has recently grown, the VM protection system that was perfect yesterday may not be adequate today.

Here's a look at some of the most common traditional approaches to protecting virtual data. Chances are good that you already have one or more of these in place. (If not, what are you waiting for? Put down this book and go run a backup — stat!)

Taking snapshots

Many IT departments, especially those that have only a few VMs in what they think are peripheral applications, stop at using a snapshot system to protect their VMs. As the name implies, a *snapshot* captures a moment in time for a VM, including not only all the software on it (OS, applications, and data) but also the content of the VM's memory at the time of the snapshot. For example, if there are open applications and running processes and services, those states will be captured as part of the snapshot. If one or more of the VM's files becomes lost or corrupted, a system administrator can restore the entire VM by using a stored snapshot that is earlier than the problem. Snapshot files can also be used to retrieve historical data, such as to go back to yesterday's version of a database by reverting to an earlier snapshot. Snapshots usually start as frozen images on the same storage system as the original data — which makes rolling back in time fast and easy.

Snapshots for VMs are most often created by the software that manages the VM, such as VMware, and the features and options available are controlled by that management software. In VMware, for example, a snapshot is comprised of several files, including files for each virtual disk connected to that VM when the snapshot is taken.

Some companies, especially those that use VMs for their non-core uses, rely exclusively on snapshots as the only form of data protection. Some of the benefits of snapshotting are:

- Snapshot-taking is inexpensive (often built into the VM management software).
- Quick recovery from fault conditions by restoring from the snapshot.
- Snapshots are easy to manage, and the process of creating them is easy to automate.

Relying on snapshots alone, though, can be problematic for several reasons.

One of the most important problems is that snapshots do not offer adequate protection against all types of data loss. Because the snapshot is often stored on the same physical machine as the VM itself, snapshots offer no protection against server hard disk failure, certain kinds of malware, or complete site loss (such as with a fire or flood). It is possible to arrange to have snapshots placed on a different physical machine to offer stronger protection, but unless that other machine is off-site, the data is still vulnerable in the case of a disaster that affects the entire local facility.

Storing snapshots off-site makes the snapshots retrievable in the case of a wider variety of events, but doing this has drawbacks, too. If you choose to store snapshots off-site, you may run into issues of network bandwidth when creating, managing, or restoring snapshots because of the large size of snapshot files and the variable speeds of networks due to contention for bandwidth.

Another problem with snapshots is that they don't normally provide long-term or even medium-term protection. Although you can create and retain multiple snapshots of a VM, it is unwieldy and storage-intensive to do so for any significant time period, so typically only a few snapshots of a VM are retained at any given time.

In addition, retrieving individual items from snapshots can be difficult. For example, if you're searching for a particular file to restore, and you aren't sure when it was created or deleted, you may have to load, search through, and unload several snapshots to find the file you want. A snapshot has no list of backed-up files and no externally accessible directory, so you can't easily browse the contents of the disks within a snapshot from outside.

Backing up files

The next step up in data protection is to employ some type of backup software — which usually starts, incidentally, by making a copy of a snapshot.

A *backup application* is an application that stores a backup of the files you specify, in a location of your choice. You can back up to the same physical disk (although that's not optimal), to a different local disk or other media, to a network location on your LAN, or to a completely separate off-site location using backup software.

Backup applications allow for more flexible backup data sets than snapshots do, because you can pick and choose which files to back up. For example, if you have the installation files you would need to reinstall the operating system, you might decide that you don't need to back up the operating system inside the VM — or at least not frequently. You might confine the backup operation to just the user data.

Backup sets also let you have multiple versions, and you can hold onto those versions for a long period of time. You can go back several months — or years — in the past to pick up an earlier version that has an important file, or a specific version of a file that someone urgently needs to recover. Backups are usually either of the whole data set (a full backup) or some combination of partial backups that cover changed data (a differential or incremental backup), and most provide a cataloging system to help users find old files more efficiently.

Typically, a company uses backup software to create a recovery data set on a location different from the primary storage. For site loss protection, users need to move the backup to a second site or create a copy at a remote location through a process, such as replication.

The data backed up by a traditional backup application is usually not ready for immediate reuse, because backup applications typically encode the data using a specific format that is proprietary to the backup application. This means that the restore process with traditional backup apps usually takes place through the backup application — and there's a separate stage before data can be used.

Traditional backup software

One option for VM backup is to treat each VM as if it really were what it pretends to be — a physical server. Backup using traditional backup software is a well-understood process for IT professionals, so little or no additional training is required. It also provides one unified set of processes, and backs up to a single easily accessible location. Backup software can be configured to run automatically, so after the initial setup, human intervention is required only if a problem occurs.

Backing up VMs as if they are physical servers, however, has some significant drawbacks. One problem is that running backup software on a VM can bog down system performance significantly, to the point where the other VMs on the same physical hardware suffer. This effect is compounded when more than one VM tries to run backup software at the same time and that happens all the time because multiple backup jobs typically kick off on different physical servers all at once. With VMs, though, 10 or 20 of those "servers" may really be housed on one physical machine. So a backup operation that should take only an hour on a physical server might end up taking eight hours or more per VM if several are competing for CPU time and memory.

The second big issue is that traditional backup software only backs up the data within the VM, not the VM itself. For full protection, you need to make a copy of the VM itself as well as the data that is inside the virtual server — and that's something that traditional backup apps that only treat the VM like a real server just can't do.

Remember the extra junk data in VMs? Traditional backup apps can't find the redundant copies and don't understand which files need to be backed up and which files can be ignored, so they often back up many files and lots of junk data that are unnecessary to include in the backup set.

Cost can also be an issue. Whether or not using traditional backup software is cost effective depends on the way the software is licensed. If your company has a site license, or a license that charges you per physical server, it can be a great bargain, because you can put several VMs on the same physical server and pay for backup of only one machine. However,

if the license is per-server, the costs can quickly escalate, because each VM is considered to be a separate server.

VM-specific backup software

As an alternative to a standard backup program, you can choose to use specialized backup software designed specifically for VM protection. There are several possible sources for this kind of product: the virtualization supplier (VMware, for example), the traditional backup software vendor (Symantec, for example), or companies and products that have emerged to address only the problem of VM backup (like Quest). Using specialized VM backup applications has many benefits because they all understand at some level that a VM is different from a physical server, and they all have a way of protecting the entire file set that represents the entire VM—not just the files inside the guest machine. Because VMs and the structure of their data have evolved rapidly, many different approaches for backing them up are available. Most of the traditional backup applications have the user load VM-specific agents, often on each VM, in order to let the application treat the files in the right way. Sometimes agents aren't needed if the user always wants to treat the VM like a complete entity, but they might be needed if the admin wants to be able to recover a single file instead of restoring the entire VM. Some of the emerging backup applications only support disk as a target. If tape is needed, they recover a backed up VM and present it to a traditional backup application to write to tape in a kind of second backup.

Whatever system they use, virtually all of the VM-specific backup applications, just like the traditional apps, encode the backup data in their own format so that whatever recovery takes place goes back through the backup app before the data can be made available to use.

As the variations suggest, the drawbacks of VM-specific backup software apps center around cost and complexity. They're full, separate backup applications that run independently of your traditional backup software, so the IT department must manage two separate backup systems. They may cost more because you have to buy additional software, in many cases buy and provision a proxy server outside of the

virtual environment to manage the backup activities, and possibly add extra dedicated storage to hold the backups. And what about site-loss protection? Traditional applications can write to tape — a removable media that can provide DR protection — but most of the VM-specific backup systems only write to disk and often can't support replication directly. For that function, they usually rely on cooperating with another product, like a deduplication appliance.

Chapter 2

Shopping for a Virtual Data Protection Solution

In This Chapter

- Looking for a virtual-centric solution
- Finding out how it functions
- Evaluating ease of integration
- Considering the costs

Protecting and managing virtual environments and their data is important, as I discuss in Chapter 1, and methods designed for non-virtual environments aren't always the best fit. So, where do you go from there? You go shopping for a solution that's right for your company and its needs.

Before you can evaluate products and solutions, you should understand what it is that you're shopping for. In this chapter, you find out about some of the most important qualities that a virtual protection solution should have and some specific questions you should ask when meeting with potential vendors.

Keep in mind as you compare solutions that virtual data protection is a new and rapidly changing field, so the answers you get to your questions today may be different when asked six months or a year from now.

Looking for a Virtual-Centric Solution

First and most critically, is the system you are looking at designed specifically to work effectively with the virtual environment? Don't settle for one that isn't. Such a system is more efficient and effective than a traditional protection method applied to a virtual environment because it is built from the ground up specifically to take advantage of the efficiencies — and quirks — of the virtual environment.

When talking with vendors, don't just stop at a simple "Yes, it's designed to work with VMs" answer when interviewing vendors and shopping for products. The following sections provide you with specific questions you should ask.

How well does the backup application work with the hypervisor interface?

The hypervisor is the VM environment management utility on each of the physical servers that host VMs. A backup solution designed for the VMware environment will be able to work directly with the hypervisor on each server. It should also be able to interface seamlessly and directly with the vSphere API and utilities, such as vCenter, the VMware management console.

Because VMs are dynamic and highly mobile, any data protection system has to be able to find them wherever they are, see them whether they are active or not, and make effective use of hypervisor tools, such as vCenter to manage them.

The ideal system would tie into the basic hypervisor API so that all the key data about the VM environment is immediately available to the backup application. For example, suppose that your IT department decides to consolidate several VMs that were previously on different servers, in different branches, into a single server in the central office. Your VM protection software should be able to automatically detect the move and update its records.

How does the backup application handle the extra data inside the VM?

Easily half of the data in a VM can be extraneous, and need not be backed up. Besides the files in the VM's file system, there is data created by the VM's operational overhead, empty space reserved for future use (overprovisioning), redo logs, and many more types of data that get created inside the VM as part of its existence.

It makes sense to ask how that data is handled. When a protection system looks at a VM, can it tell the difference between the active data in the guest file system and these extraneous files? Can it exclude the data that doesn't need to be backed up, without excluding anything critical to full data recovery? If it can, it can dramatically decrease the backup storage and time required for backup, as well as cut down on network bandwidth that the backup uses.

You should also look for the capability to support third-party products that reduce redundancy in backups through data deduplication. Deduplication appliances speed up backup, streamline disaster recovery protection, reduce admin overhead and save users money, so it's important that any VM protection system can work with the well-established deduplication appliances on the market to get both maximum data reduction and effective replication for DR protection. If the only option is deduplication that happens in the backup software, it's probably too limited — that approach to dedupe is usually not very effective at finding redundancy, which means that backups take up too much space and require too much bandwidth for replication. (Look at the latest edition of *Data Deduplication For Dummies*, available from Quantum, for lots more details).

Is a physical server required?

All the same benefits of VMs that caused your company to start using them in the first place also apply to your virtual protection solution as well. Put simply, it's cheaper, easier, and provides more administrative flexibility to host the virtual protection function within the virtual environment than it is to put it on a physical server.

Some backup applications designed to protect VMs require separate physical servers or recommend them highly for maximum performance. That kind of requirement both raises costs and adds to complexity of the system. The ideal VM backup system would leverage virtualization to a high degree and would operate well when installed on a platform, such as an ESX server in a VMware environment. In fact, the best system would be one that installed as a virtual appliance inside the VM environment — virtual appliances provide maximum flexibility and lower management overhead.

Finding Out How It Functions

Theory is all well and good, but how does the product work in practice? Something that sounds like a great idea on paper could turn out to be less than ideal when implemented.

For example, suppose a vendor tells you, "Sure, you can restore whole VMs or single files with our product, no problem!" But then when you get down to doing so, you find out that it's an onerous process that requires you to first restore the whole VM every time, and then extract the right file from the restored VM. It's impossible to tell from just looking at a company's marketing literature whether a particular capability is easy or difficult to take advantage of. It's important to ask detailed questions about the specifics of the product's functionality, and to see a demo running in a situation that's similar to yours.

The following sections give you some important questions to ask as you are shopping.

What's the backup process like?

Ask to see a backup operation in action, and ask what it takes to set it up and manage it. Some backup operations are complicated to administer; others are simple and can be set up with a few clicks. Ask to see a demo!

You should also find out about the options for various backup scenarios that you are likely to really need. You need to be able to define policies for backing up VMs in a particular resource pool — and that includes not backing some up or just using a

snapshot when short-term protection will do. You need to be able to choose between full backups and incremental backups, too.

Watch out for features that might get piled on that make deployment harder and management more complex that you don't really need.

Try to think of all the real uses that the backup application will see and understand what it takes to get it done. Resist the impulse to make the process more complex than you really need.

Are agents required?

An agent is software that you put in an environment that allows coordination with outside software. Some protection systems require you to install an agent on each of the virtual servers you'll be backing up. That's not optimal, of course, because it's another software layer, an extra level of complexity, it may require more hours from the sys admin team, and it limits the system's agility. When new VMs and their physical server hosts are brought online, they might not be protected until someone installs the needed agent — and, in most cases, those agents are something for which you pay extra.

What management tools are provided?

Ask to see the management tools in action, if the system provides any. Are they easy to understand and use? Do they have all the capabilities you want? One important feature to look for is the ability to look inside the guest file system from outside the VM, for example. Seeing a demo is critical here. Does this management system look like something you and your staff could easily use, and does it do everything you want it to?

What kind of restores can you do, and how long does it take?

When you have a situation in which you need to restore from a backup, you are already having a bad day; you don't need to

add to that by struggling with a complicated restore process that isn't able to accomplish what you want.

If you back up the whole VM, can you get to individual files inside it when you restore? Most products claim that they can, but it makes sense to find out what that takes. In some cases, you have to do something different at the front end to make it happen, like add an agent on each VM. In other cases, you may have a multi-stage process at the back end, such as having to restore the whole VM first. The ideal solution would back up the whole image of the VM, so the backup is fast, but it would allow people to restore the whole system if they wanted or find individual files inside the VM and get to those directly.

Speed of restoration is also an important factor to consider. You need to be able to very quickly bring critical VMs back up after a disaster wipes them out, not only to restore availability of important data, but to keep servers up-and-running that are required to interact with customers, process sales, manage inventory, and so on. Look for a protection system that's flexible enough to be able to restore backups quickly from and to any location, so you're not tied to specific hardware being available.

What's the backup format?

It makes sense to ask what format the backup copy is stored in, and to understand what implications that has for storage, use and restoration.

Most backup systems use a proprietary backup format that only the application that wrote the backup can read. This isn't optimal because you might not always have immediate access to the original backup program when you need to grab something out of a backup file.

Some backup solutions, however, use an open backup format that many different applications can read in a standard file system view. That's much better, because even if you aren't running a copy of the backup application, you can boot up the VM and look at the data inside it. This results in a much more direct and usable system. For example, suppose you need to restore one particular Microsoft Word file that a manager urgently needs from a backup set. With an open-format backup,

you can browse the backup set with a standard file system browser and access that file directly, without having to restore the backup or even use the backup application at all.

What kind of support can I expect?

Whether you pick a specialized backup product for VMs or an option from a legacy supplier, you want to know what it will take to get the system up and running and how much ongoing care it will need. If you need an agent on every VM, for example, you need to think about what that means as your system grows and changes. You also want to be sure to think about what kind of support you can get. Especially with startups or companies below a certain critical mass, it's common to have the service and support arm lag behind the product development and sales side of the house. If you have the admin resources to be independent this may not be a concern, but it is important to understand what kind of help you can expect if you need it.

Evaluating Ease of Integration

Virtual data protection is only part of the company's full data protection plan. There is probably also a great deal of data stored in more conventional, physical-server environments that also must be backed up on a regular basis. It's important when integrating virtual backup into the larger environment to understand exactly how — and how well — the systems will work together. That way you know that all the organizational protection mandates are being met — short-term backup, off-site protection for DR purposes, and long-term retention for compliance — and that the VM protection will be supported by the backup, replication, and tape creation processes you're using for your other data.

Don't just look at a checklist. Look at the system in action and get a chance to see what it really takes to implement and to work with.

Find out how the virtual backup solution you are considering will integrate with the company's existing backup system. Does the virtual system work in isolation, or is there a way of having the two work in cooperation with each other? The ideal VM protec-

tion system would have an interface that allows it to work synergistically with the other applications to provide specific features or fit easily into existing corporate protection processes.

Make sure that you ask whether the system you are considering requires a multistage process to move backup and restore data from conventional machines via the virtual protection system. Products that require multiple transfers of data to integrate the backup systems are less efficient and more costly and time-consuming to use.

Some companies already have data deduplication systems in place for their conventional and virtual servers; others may plan to implement it in the future. Deduplication appliances are an important part of an increasingly large number of backup environments. They store more data in less space, and most provide very effective replication that only moves new data in each backup, so users get off-site protection over modest networks.

Backup applications that can work well with deduplication appliances can provide extra value and make the entire backup and DR process more effective and easier to administer, so it's important to understand how the application interfaces with appliances. As you shop for VM protection, ask whether the system does its own deduplication, and how well it works with other deduplication systems. Find out specifically how that integration works. For example, is there a link between the kinds of data reduction that the backup process performs on the VM and the kinds of data reduction done via deduplication? Are the two synergistic, or at odds with each other? What specifically do they do that makes each other better or worse?

Considering Costs

Different virtual backup solutions can be more or less expensive to implement. Be sure to look beyond just the initial setup costs, and ask about all the additional items that might be required — both at the initial deployment and over time. The cost is not only the licensing charges, but also any ongoing support changes and additional hardware that might be required. The ideal system keeps total costs as low as possible, not just the initial acquisition cost, so it's important to

go through the whole list of all the items that may be involved in the deployment of a system and make sure you understand the full cost of all of them.

Here are some points to consider involving cost:

- **Licensing:** When you are comparing the costs involved in different systems, make sure that you are comparing apples-to-apples in terms of licensing and that the VM protection's approach is compatible with the way that the hypervisor structures its licenses. Some products may be licensed to the entire site, with unlimited VMs and physical servers supported; others may allow you to have only a limited number of one or both, and may charge you extra later as your organization expands. Ask if you can have a second copy of the backup application running without incurring extra license fees, for example, and find out if it costs extra to use different operating systems or to take advantage of the product's most advanced features.
- **Physical servers:** Some VM protection solutions require separate physical servers; others don't. Make sure that you know the costs associated with whatever servers the solution requires.
- **Storage media:** The largest hardware cost involved in data protection is often storage media — that is, the disk system on which you store the backed-up data. Make sure you understand the costs of the recommended storage media, and what the costs may be in the future if your needs increase.
- **Technical support:** Is product setup, initial IT staff training, and ongoing technical support included in the product you are considering, or does it cost extra? Support contracts can add significantly to the cost of a product that does not provide it as part of the package.

Chapter 3

Making Virtual Data Protection Easy

In This Chapter

- Simplifying management
- Understanding the four core technologies
- Helping overburdened hosts
- Introducing vmPRO 4000 and vmPRO Software

Now that you know the challenges of protecting virtual environments, and you know what to look for in a protection solution, you are probably ready to start looking at some specific products that may be the right fit for your company.

Complete virtual data protection and management can involve dozens of different tasks, executed separately on each physical server, each with its own hypervisor. Quantum offers an innovative set of protection technologies aimed at VMware environments: vmPRO 4000 for smaller or remote sites and vmPRO Software for datacenters.

Simplifying Management

The Quantum vmPRO technology that underlies both the vmPRO 4000 and vmPRO software is based on software that runs on a virtual appliance inside a VMware environment. This minimizes the burden of virtual storage management and unifies all your VM-based storage in a single namespace. Accessing virtual storage becomes as simple as accessing

a network drive. The solution is managed and monitored through VMware vCenter Server, an interface with which your IT staff is probably already familiar.

All of the vmPRO solutions are simple to implement, too. They install easily in the virtual environment and integrate directly into the VMware management tools.

In the case of vmPRO software, which is designed to work with third-party backup application, users find that no radical changes are required to transition from your current methods. All job schedules, policies for backup, security scans, replication, and so on continue to be managed by your existing backup application. No new policies need to be created, and no manual or scripted coordination is needed. Therefore, your current protection processes don't have to change — they just get better. They become more compact, more efficient, and easier to manage.

Understanding the Four vmPRO Core Technologies

Quantum VM protection solutions are based on four core technologies, each of which works seamlessly with the others to form a complete protection and management system. The following sections provide an overview of the technologies.

SmartView

A core component of all vmPRO solutions, SmartView is software that works with the VMware management interface to enable you to see all the files in a virtual system in native file system mode. That way, you can browse the content of a VM without being inside its interface, and you can also see the files that comprise the VM environment.

SmartView provides a single access point that spans all your SAN, NFS, and DAS storage across all VMware hypervisors. Any application that can use a CIFS share or NFS volume can use this access point. Because vmPRO is working directly with vCenter Server, new virtual machines are automatically detected and included in SmartView, so you never have to

worry that you've overlooked important changes. Figure 3-1 summarizes the key features of SmartView.

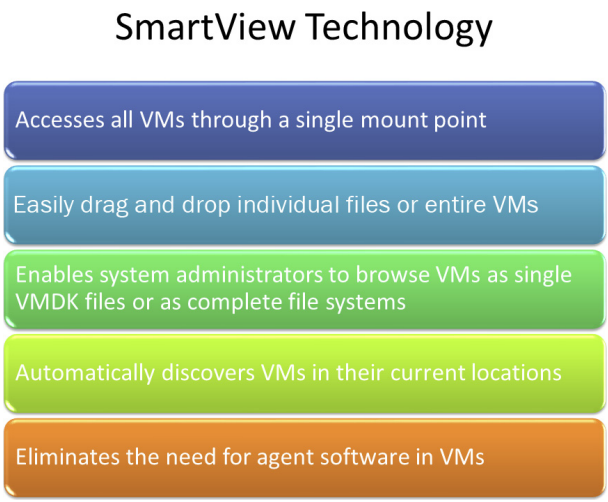

Figure 3-1: Understanding SmartView features

Progressive Optimization

Part of the challenge of protecting VM data, as I explain in Chapter 1, is that it can be difficult to distinguish the important files from the unimportant ones. Most VMs are vastly over-provisioned and contain quite a bit of redundancy and blank space. Progressive Optimization is core technology inside all the vmPRO solutions that can look inside a VM's file system and distinguish what files are active or used and what files and what data are transient, expired, unused, or simply filling up space.

Progressive Optimization analyzes the metadata in file systems, looking at allocation tables and file locations to determine what data is active and what is not. This analysis reduces the amount of data that needs to be protected by up to 75 percent or more.

VMware includes a technology called change block tracking that lets people see which blocks have changed since the last protection event. Progressive Optimization takes that general idea to

the next level by looking at *why* certain blocks have changed. For example, blocks that are deleted or expired also appear as "changed" as far as VMware is concerned, so they still get backed up by most applications. Progressive Optimization looks deeper than that, and marks for backup only the blocks that still contain valuable data. As much as 30 percent of data that change block tracking *thinks* should be backed up actually does not need to be. Reducing the size of the backup set not only means a reduction in storage needs, but also reduced CPU usage and backup time. Progressive Optimization relieves I/O on the entire hypervisor, which is very important because I/O is critical to virtualization environments. Figure 3-2 summarizes the key features of Progressive Optimization technology.

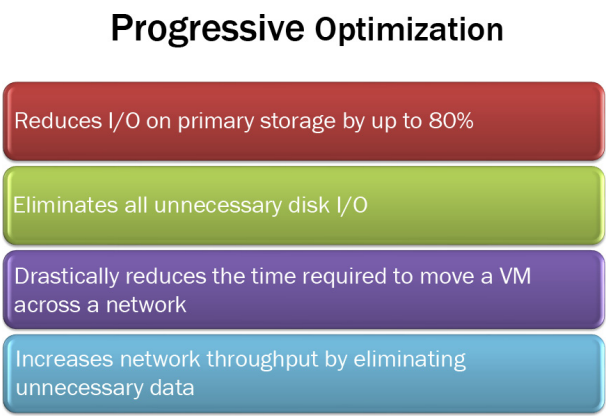

Figure 3-2: Understanding Progressive Optimization technology

SmartMotion for fast, agile backup and restore

Agility is an important asset to have in any VM protection solution. You want to maintain maximum flexibility in the management of your VMs, including the ability to move them between physical hosts. And, when problems occur, you want to be able to get back up and running as quickly as possible.

Most backup applications, as I discuss in Chapter 1, sacrifice immediate usability in the name of storage efficiency. In order

to control the backup data, they store it in a format that is unreadable from outside the backup application. To use it, you must go through the restore function of the backup application. That can take quite a bit of time and can be inconvenient, too.

The Quantum vmPRO solutions include SmartMotion, a simple backup and movement technology that enables you to create native format copies of VMs on a second disk target, so that the copy is the full file system view and *usable as-is*. No complicated restore operations are needed for you to be able to access the backed-up data. That way, if something happens to the original, you can get an extremely fast restore from the copy of the VM, which is all there in file system view. You can boot it up and make it an active VM whenever you want, wherever the VM may be now located. This restore capability can have a number of uses — not only to restore a backup, but to try something out in a test environment that mirrors the production server.

Recovering individual files is easy, too. Because the backup is available for management in native file system view, you can browse the disk associated with it (the guest file systems) to find the exact files you want on that disk and retrieve them. You can also retrieve earlier versions of individual files in cases where the original file was changed inappropriately or corrupted.

With Quantum, you don't have to choose between storage efficiency and restore convenience. Progressive Optimization technology is built into the backup system, which screens out unused space, temporary files, and other files that aren't important for the VM's operation. As a result, you have a backup that, even though it is still completely browsable, takes up much less disk space than the original VM, and is fast to read and restore from.

This technology, along with Progressive Optimization and SmartView, is installed directly into the virtual environment and runs as a virtual client inside the hypervisor. No agent is necessary, and no physical server is needed. That makes it extremely economical to implement.

Data deduplication

The Quantum VM-oriented data protection solutions set also includes Quantum's unique data deduplication technology, an extremely effective method of reducing the amount of redundancy in a storage system.

Many methods are available for compressing or reducing data, but none of them offers the flexibility and granularity of Quantum's method. A base technology in the data reduction field is *data compression*, which compresses individual files by searching for repeated strings within each file and replacing them with codes for reconstructing them when the file is restored. Compression typically reduces data by around 2:1, although it depends on the data set. Newer technologies, including single-instance storage and deduplication, take data reduction much further — and they usually include compression in their own technology stack as well.

Single-instance storage systems look for repeated entire files. For example, if two servers contain the same Windows system files, and the two files are identical (same name, size, and date), a simple storage reduction system might back up only one of them and include a pointer to the first copy whenever another instance shows up. Single-instance storage is efficient when files are identical, but it is not very useful when there are small changes. If anything in the file changes (sometimes even the name!), the single instance storage system has to store an entirely new copy of it.

Quantum's *data deduplication* takes data reduction a step further by looking at blocks of data rather than files, so it stores changed blocks instead of whole files. Quantum's deduplication segments a dataset into variable-length blocks and then checks for duplicates. When it finds a block it has seen before, instead of storing it again, it stores a pointer to the original. The sequence of pointers makes sure that all blocks are accessed in the correct order when the file is read, ensuring that the deduplication process is invisible to the end-user. (You might want to check out *Data Deduplication For Dummies* for more information. It's available from Quantum.)

Because data deduplication recognizes differences on a block-by-block basis both within and between files, it's the most efficient data reduction technique on the market, and allows for the greatest savings in storage costs. How much savings are we talking about here? Get ready to be amazed. You can store *10 to 40 times more data* in the same disk space than if you were not deduplicating. The effects of that benefit can be dramatic. Because so much less data has to be moved, deduplication can make it possible to replicate data over low-bandwidth networks, so you can store a secure off-site copy

of a data set while actually transferring only a very small percentage of the data.

The Quantum vmPRO technology works with Quantum deduplication to make it more effective. The vmPRO Progressive Optimization performs as a kind of preconditioning that actually improves the rate of deduplication. Both the vmPRO 4000 and the vmPRO software provide integration with the Quantum deduplication technology to take VM protection to a higher level.

Introducing Quantum vmPRO Solutions

Quantum includes its core VM protection technologies in two different product offerings. For small and medium-sized businesses, the vmPRO 4000 delivers a complete turnkey solution that includes both backup software and deduplication-enabled storage. For larger enterprises that may already have robust backup solutions in place, the vmPRO Software integrates seamlessly with Quantum midrange or Enterprise deduplication appliances to protect VMs as part of the company's overall data protection system.

vmPRO 4000

Quantum's vmPRO 4000, for small- to medium-sized businesses, is an all-in-one, simple technology appliance for protecting VMs in a smaller environment or remote office. It includes all four of the core technologies discussed earlier in this chapter, and protects any small environment with a single data set of up to 8 to 12 terabytes. Its high-speed backup utility writes data directly to disk and uses Quantum's industry-leading deduplication system for long-term retention. For DR protection, the vmPRO 4000 automatically replicates deduplicated data to other vmPRO systems or to DXi appliances in central sites.

vmPRO 4000 is economical to implement, because it doesn't require an extra server and it includes a deduplicating storage target as part of the solution as well as backup software. All the software runs inside the virtual environment, so no extra hardware is needed. One affordable price gives you

everything you need for protecting all your VMs, with no hidden costs or extra charges. The application runs on a self-contained VM appliance, so no agents or dedicated servers are necessary.

The vmPRO 4000 system can act as a target for other backup applications, too, such as those protecting physical servers, with that data being sent to the same deduplication-enabled storage system as the virtual data — and together they can be replicated to other vmPRO 4000 or DXi systems for automated, consolidated site-loss protection. As a result, you have one consolidated storage system for all your backups.

The vmPRO 4000 deployment process is fast and easy, with automated VM discovery, simple scheduling options, secure remote administration for multiple units, and capacity-on-demand scalability, so you can grow your system at any time without disrupting existing functionality.

vmPRO software

Quantum's vmPRO software for data centers is designed for larger companies that want to integrate VM protection into their established data protection system — the one they use for protecting and retaining data from their physical servers as well. It is designed to work with any of Quantum's mid-range or Enterprise DXi appliances, making it useful for even large, diversified data protection needs. The vmPRO Software includes all the VM protection technologies described in the previous section, including Progressive Optimization and the software utility for creating backup copies, SmartMotion. It works with DXi appliances to bring the technology package to larger sites.

vmPRO software is a light-weight application that runs as a virtual appliance in the VM environment. As in the vmPRO 4000 package, it can auto-discover VMs and present a file system view that enables you to back up VMs or files within VMs without adding VM-specific agents. The software filters out inactive data, reducing backup volumes by up to 75 percent and boosting deduplication rates. To support fast recovery, vmPRO software supplements traditional backup with a

simple VM backup utility that creates native-format VM copies on a secondary disk, allowing you to restore entire VMs or single files with ease.

The benefit of using the software-only approach in larger environments is integration with an existing backup application. This seamless tie-in with third-party backup software enables you to continue to utilize your existing infrastructure while adding the enhanced capabilities of the vmPRO protection technology. It supports much higher capacities on deduplication appliances, a larger number of VMs, and integration with complex existing backup systems.

Chapter 4

Ten Frequently Asked VM Protection Questions (And Their Answers)

In This Chapter

- Finding out why you need a backup process designed specifically for VMs
- Determining what features and formats are most important
- Figuring the real cost of a VM backup solution

In this chapter, I answer the ten questions most often asked about protecting VM environments.

Do I Need a Separate Backup Process If I'm Using Snapshots to Protect My VM?

If the data truly needs protection, you need something beyond snapshots. Generally speaking, a snapshot is a point-in-time image of a VM created by making copies of changed sectors on the same storage medium as the primary VM. Snapshots are terrific for quickly rolling back to a known good state in the event of a problem, but they don't do the job of backup and DR protection. Backups, which often start with a snapshot as their source, create a copy of the entire VM on a different medium

that is isolated from the original. With backup software and backup targets, data can be stored longer, it is protected from hardware faults, malware, and user errors, and it can be used as part of a disaster recovery process that protects data against site loss or damage.

Why Can't I Use the Same Backup Process That I Apply to Physical Servers On My Virtual Ones?

In theory, there's nothing wrong with that idea at all, and most people do just that when they first start trying to protect their VMs. The problems start to show up pretty fast for most companies, though, especially when the number of VMs grows.

The first issue is cost and complexity. The price structure of most backup apps is based on licensing agents for each server, so license costs tend to rise fast, there are lots of agents to manage, and delays may occur in protecting new VMs.

Then there's the performance problem. Primary applications that get virtualized tend to have variable duty cycles, so multiple virtual servers can often coexist on the same physical servers without negatively impacting performance. However, backups usually have to be scheduled during the same backup window, so when multiple backups are running at once, there's more contention for the physical resources and backup speeds drop dramatically. That's why most system administrators quickly gravitate to backup apps that are built specifically for VMs.

Why Are So Many Backup Applications Designed to Work for VMs?

There are several reasons, but one of them is that innovation tends to come from people who are free to look at new problems in innovative ways. It can be faster to solve tough problems if you start from scratch rather than fiddling with

older technology and making it work for a new situation. VMs definitely create a new situation that is substantially different from the legacy server systems that traditional backup applications were originally aimed at.

Another is that the architecture of VMs has also evolved quickly. The first-generation backup approaches (some of them provided by the VM providers themselves) quickly disappeared, and the internal structure of VM data and file systems has changed several times. It's hard for large, mature software products to adjust quickly in that kind of environment.

Why Do VMs Seem to Hold More Data Than Physical Servers Doing the Same Job?

There are a couple of important reasons that VMs take up more space than it seems like they should based on their file systems. One reason is over-provisioning. Many IT departments for sound operational reasons assign more capacity to a VM upfront than it needs at first. Doing so makes it easier to scale the systems later on when needs grow. When most backup applications look at a VM, they see everything in the VM package, including all that empty space. When they back up the VM, they back up all that space, too.

The second reason is that there's an imperfect link between the part of the guest file system inside the VM that holds the applications and user files and the larger package of files that make up the whole VM (including all the stuff that's needed to make the VM look like a physical server). The link means that some of the data that will eventually get discarded hangs around in the larger VM package and gets backed up by most applications.

Fortunately, some applications solve both the over-provisioning and the extra data problem and back up only the real data. Quantum vmPRO Software, for example, actually reduces the amount of data that needs to be backed up in some VMs by up to 75 percent by screening out that extra data!

Why Can't I Access Individual Files From Some Types of VM Backups?

Maddening, isn't it? This occurs because of that difference between the file system inside the VM and the larger VM package that we talked about in the preceding question. It's easy to just back up the whole image of the VM, without knowledge of what's inside the guest file system, and so that's what most people do. So when you go to restore the VM, you have to "recover" it, that is, bring back the entire virtual machine and then look inside it to find files.

You can get around this problem by choosing the right protection solution. Some VM backup applications can capture the guest file system information when they do the backup, so you can get to individual files. When shopping for a protection system, though, make sure that you understand what hoops you might have to jump through to get the backup app to do that. For example, some products require an extra agent to enable you to look inside a VM for individual files, or require an extra step at the recovery. The best kind of VM backup provides both the whole VM and the guest file system so that you can recover the whole machine or individual flies directly. It's even better if the files are presented in native format.

In Which Formats Are Backup Files Presented, and Why Should I Care?

Most backup applications create files that are encoded in a proprietary format that only they can read. Therefore, you have to run a backup data set through the backup application that created it before you can do anything useful with the data in it—like open an individual document or presentation file.

This limitation made sense back in the days when backups were mostly written to tape, a linear storage medium that

didn't allow random access. However, with today's all-digital disk-based backup media, it's possible to view and access backup data in much more flexible and usable ways — provided, of course, that the backup software is capable of moving beyond "the way it's always been done before."

Some of the best and newest backup applications back up files in their original native format. That's important because it means faster and more flexible restoration. Data can be restored by the original applications that created the data files, without using a backup application at all. That means data can be restored even at sites where there's no copy of the backup application! That makes it easy to use backup copies for non-backup purposes — as archives of usable files, or as files to use for test purposes, both things that are really hard with data in proprietary, backup-only formats.

Do I Have to Choose between One of the New VM-only Backup Applications and My Legacy Backup Software?

Quite a few companies run two different backup applications — one for physical servers and another for VMs. If these are completely separate systems, however, doing so can create extra costs and more complexity. It can also leave some data vulnerable, because some of the VM-only products haven't yet evolved to provide a good system of DR protection.

The most promising approach we've seen lately is one in which a VM backup product can work directly with traditional applications in a kind of synergistic way. Quantum's vmPRO software is an example. It can see the native file system structure inside the larger VM package, and it can enable traditional backup applications to see the same view too. This arrangement lets the legacy application back up all the data in the VM without requiring any extra licenses or agents. It also gives the legacy application the benefit of vmPRO's filtering process — the one that screens out extra data and over-provisioned storage — so the legacy application gets the

same high levels of data reduction (up to 75 percent). This gives IT departments access to innovative VM-specific backup code, but also ties directly in to the corporate data protection process that has been built around the mature, large-code backup app and all of its retention and management capability.

How Does VM Backup Work with Deduplication Appliances?

It depends entirely on the backup application. Some VM systems have some kind of deduplication of their own, and may even make it hard for users to disable it. That's a problem because the deduplication built into independent appliances is usually a lot better, and supports mature, proven replication products for DR protection — something that is rare among the VM backup software.

The best VM protection system is one that can support existing deduplication appliances. Some products, such as Quantum's vmPRO software, go so far as to provide a kind of "pre-duplication." In this case, the VM backup preconditions the data being read out of the VM, substantially enhancing overall deduplication rate for the backup set once it reaches the appliance. This synergy of technologies is one of the reasons that Quantum, one of the leaders in deduplication, added the vmPRO technology to its family of technologies in mid-2011.

How Can I Estimate the Real Cost of a VM Backup Solution?

The best way to estimate your overall cost of ownership is to have someone knowledgeable, with input from the vendors you are considering, show you all the elements that will be needed. This total must include initial license costs and support costs, of course. Be sure to find out if there are any additional agents that might be required. (Some software requires them if you want to be able to recover single files and believe me, *you do*.) You also must consider the costs of any additional hardware that might be needed. You might have to

ask this a couple of times to get a straight answer from some vendors. Some of their apps don't "require" separate physical servers, but "recommend" them strongly in order to get the kind of performance you need. And don't forget to look at how costs will change as your system grows. Next year, you will have more data and you will have more VMs.

What Are the Most Important Management Features I Should Look For in VM Backup Software?

If you are adding a new application for protecting VMs, you want to be sure that you are bringing something in that minimizes complexity, cuts costs, and saves time. It would be ideal to have a product that took advantage of the VM environment and could run as a virtual appliance, for example.

It's also important to find a solution that sets up easily, has clear policies for backing systems up (or not, depending on your needs), and that integrates easily with the management tools provided by the VM environment. If you use VMware, for example, the product should be able to work with vCenter and should have a vSphere API. Look for products that can auto-discover VMs, automate backup, provide native file system view of the VMs and the backup copies, and can work effectively both with other backup applications and with deduplication products.

Find out the best way to protect virtual data

More and more companies are saving big money on hardware, software, and IT labor by virtualizing many of their most mission-critical server operations. But regular backup systems aren't designed for virtual environments, and trying to back up virtual data using traditional methods can be wasteful of both time and disk space.

- ***How VMs complicate data protection* — and how to choose a backup solution that takes full advantage of the VM environment, not just works around it**
- ***What the traditional methods of protecting virtual data are* — and why none of them are optimal**
- ***Why snapshot backups are useful* — and why, for mission-critical data, they are not nearly enough**
- ***What a next-generation VM protection product brings to the mix* — and what specific features you should look for**
- ***How Quantum's vmPRO solutions stack up* — in terms of flexibility, efficiency, and cost savings**

Open the book and find:

- **What specific backup challenges the VM environment presents**
- **Why traditional backup methods are inadequate for virtual machines**
- **What specific features to look for in a VM protection suite**
- **What questions to ask potential vendors about their products**

Go to Dummies.com®
for videos, step-by-step examples, how-to articles, or to shop!

Faithe Wempen, M.A., is the author of over 130 books on computer hardware and software technologies in her 20+ years in the computer publishing industry, including *Microsoft Office 2010 For Dummies eLearning Kit*. She is also an adjunct instructor of Computer Information Technology at Indiana University/Purdue University at Indianapolis, specializing in hardware and software architecture.

For Dummies®
A Branded Imprint of
WILEY

ISBN: 978-1-118-19464-5
Book not for resale