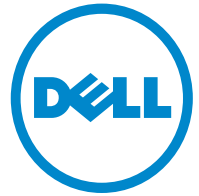
The CIO's guide to data protection in the virtual + physical data center

by Greg Shields, MS MVP & VMware vExpert



The power to do more

Chapter 1: Business risks associated with traditional data protection approaches

Executive summary

Your data center is constantly becoming more and more complicated. Today, the proliferation of virtualization—both for virtual servers as well as virtual desktop projects—has added an all-new layer of complexity to the picture. What used to be well-understood processes and technologies—such as data protection—have become bifurcated and chaotic. Now, on top of the tools and techniques you use to protect your physical data assets, you're being forced to adopt an entirely new set of solutions to protect your virtual data assets.

Making the situation worse is that data protection techniques have *never* done a superior job of actually meeting your organization's business goals. Unreliable backup media, awkward backup management processes, difficult-to-test disaster recovery techniques, and time-consuming recovery workflows have always presented business risks. Toss in the new challenges presented by virtualization, and you're looking at a recipe for business disaster.

This guide will examine major business risks associated with traditional data protection approaches, particularly with regard to their applicability for a hybrid, virtual + physical data center. It will redefine the business goals and drivers for data protection, and consider alternative approaches that better meet those goals.

Prologue: What is "traditional?"

Throughout this book, you'll see the term *traditional* used a lot. In this book, the term refers to the decades-old approaches to data protection that organizations still tend to rely upon. It refers primarily to tape-based backups made by copying files from a server onto tape. This traditional approach often relies on a locally-installed agent on each server; that agent is responsible for getting data off the server and into the backup solution, which then writes the data to tape. The agent may simply be reading files off disk, or it may be querying data from a database—such as an agent that connects to SQL Server or Exchange Server and requests data rather than attempts to grab the always-open database files themselves.

When applied to a virtual environment, traditional data protection approaches either treat the virtual server as physical—meaning the backup agent runs inside the virtual machine and accesses files, queries data, and so forth—or treat the virtual server uniquely, grabbing entire virtual hard disk files in order to back up the entire virtual machine. Either way, the backup data is ultimately written to familiar magnetic tape media, and in many cases, may also be written to disk as an intermediate, short-term storage strategy.

Inconsistent processes = Inconsistent results

In most cases, the traditional approaches that you use for physical data protection can be applied to data protection for your virtual assets. However, as this chapter will explore, you incur significant business risk that way.

That said, there is an equal amount of risk in using *different* approaches for physical and virtual. Any time you attempt to accomplish a given task using multiple approaches, you run a risk of inconsistent results. In other words, using different approaches for virtual and physical creates a significant risk that you will not achieve the same level of data protection for both. *That* is a significant threat to the business, particularly in organizations that are rapidly virtualizing physical assets.

For example, suppose you use a traditional backup and recovery technique to back up your physical servers. You're likely writing data off to tape and are prepared for the amount of time it takes to recover, say, an individual file from tape. Let's also suppose that you're using a virtualization-specific solution to back up your virtual machines. That solution provides you with yet another set of expectations with regard to recovery times, granularity, and so forth.

Now you migrate one of your physical servers to a virtual machine. Suddenly, the machine is no longer under your traditional, physical-focused backup methodology. The expectations users formerly held for recovery times on that machine are no longer valid. Now, recovery times are subject to the expectations of the virtual backup solution. Entirely different admins may now be responsible for recovery, meaning even your help desk's processes and expectations will have to be adjusted. Data may be recoverable with less granularity or with different service levels.

Why should virtual and physical have *any* differences? It's all just data, right? Why should users, your help desk, or your administrators have to follow any different process or have any different expectations? Isn't the whole point of the data center to present users and administrators with a unified environment of *services*? Why should anyone need to behave differently or expect different things just because a server is running virtually?

Lengthy time to recovery

When a failure of any kind occurs, whether an entire failed server or the loss of a single critical file, organizations always want to recover the data as rapidly as possible. Organizations typically define a Recovery Time Objective, or RTO, which states the maximum amount of time the organization wants to wait for the recovery to be completed.

Traditional data protection techniques tend to not lend themselves to a short RTO. Magnetic tape, while continually improving, remains painfully slow when comparedwith disk-based storage. Even backup solutions that incorporate disk-based storage still require moderately long RTO, often because such solutions commonly require entire servers, databases, and other large elements to be restored in order to access even a single piece of data. Why bring an entire file server back to life just to get to a single file?

This process can be especially painful when a traditional backup solution is used to grab entire virtual machines, because the *entire virtual machine* must be read back from tape, written to disk, and then started up in order to recover even a single piece of data. Because single-item recovery is a far more common occurrence than entire-server recovery, this additional overhead quickly multiplies into serious overhead.

The business risk here is straightforward: The traditional approaches simply require too much time. Users wait too long, being less productive. Expensive IT staffers spend too long, wasting money. Shorter RTO is always desirable, but traditional data protection techniques simply don't provide a great RTO. Sure, most organizations accept long RTO as a fact of life—although the fact is that far better RTO is possible. Achieving it simply requires re-thinking, and a recognition that traditional backup approaches are at an evolutionary dead-end.

Poor recovery granularity

As already stated, traditional backup approaches tend to have poor recovery granularity. Having to restore an entire mail server just to access a single lost mail message is overkill—but it's something many organizations accept as inevitable.

Of course, these traditional approaches recognize their granularity problem, and they often offer workarounds. For example, most tape backup solutions maintain an index of the files they back up. This index enables the solutions to wind through the magnetic tape to the location of a specific desired file so that just that file can be restored. This approach works fine for files but less well for complex data structures. Take Exchange Server as one example, where backup solutions almost inevitably read out *each individual e-mail message*, one at a time, so that they can recover single items. That's a fairly massive effort, and althoughthis approach increases granularity, it can also slow recovery operations simply because so much tape must be spun through in order to find an individual item. Further, this approach seriously hampers entire-server recovery, so in many cases, backup solutions will back up the entire database as a unit *and* back up individual messages. What a waste of space.

The business risk of poor granularity is somewhat subtle and easy to miss. Essentially, poor granularity results in increased recovery times, wasted backup storage space, or both. Both effects make your overall data protection strategy less efficient and less workable.

Excess administrative overhead

Traditional backup approaches are, despite years of evolutionary improvement, incredibly labor-intensive:monitoring backups to ensure they complete; shuffling tapes off-site, on-site, and so forth;not to mention the incredible amount of time that even a simple single-item recovery can consume—and the incredibly time-consuming process of a complete system recovery, should one become necessary.

In today's environment, where every single resource is expected to produce maximum value, traditional data protection approaches simply require too much time and effort to maintain. By having to dedicate so much administrative overhead to both daily backup and recovery operations, you simply risk running

out of resources. Projects that could be beneficial to the business end up unrealized because there aren't sufficient IT resources. Problems take longer to solve because resources are tied up with daily overhead.

Lack of recovery flexibility

When a system failure does occur—and you have to assume it will because that's the entire point of business continuity planning—traditional data protection approaches do little to offer you a large degree of flexibility.With the myriad technologies available today, organizations should have incredible flexibility in recovery: Recover a failed system to a standby server. Recover physical servers to a temporary virtual machine. Recover to a virtual machine in the cloud, or in another datacenter entirely. In fact, this flexibility would change the entire idea of disaster recovery, enabling more frequent and more successful disaster recovery testing and massively lowering disaster preparedness costs.

But traditional backup approaches don't really facilitate that elasticity. How are you supposed to recover a failed system to a cloud-based virtual machine, for example, when your backups are on magnetic tape? Traditionally, you haul your tapes to a dedicated recovery facility, where you start winding data off the tapes as quickly as possible to recover as many core systems as possible. Assuming the tapes don't let you down—and tape-based errors and corruption are hardly unknown, so there's no guarantee of not being let down—it can take *days* to recover even your most crucial systems. And you incur the costs and administrative overhead associated with having that recovery site.

Your organization deserves better. Your data center has evolved to offer greater flexibility, and you're using physical and virtual assets in combination to better achieve your business goals. Why shouldn't your data protection strategy also make better use of modern technologies such as virtualization so that you can—if needed—use a hybrid recovery approach to better achieve specific goals, mitigate disaster, and get the business back online as quickly as possible?

Massive amounts of at-risk data

This is perhaps the primary and most concerning risk with traditional data protection approaches: The sheer amount of data that is constantly atrisk for permanent loss.Because traditional approaches often require moreorless "offline" access to files and the backup process itself can be resource-intensive, backups are usually limited to evening, weekend, and other off-hours maintenance windows. That means all data created or changed between windows is at risk for permanent loss until the closing of the next window, when the backups have been completed.

A Recovery Point Objective, or RPO, is an organization's stated tolerance for data loss. Most organizations tolerate a loss of essentially a full day for many of their systems, and may only have more rigorous standards for their most critical data. Even *that* data may have an RPO of several hours. Organizations are accustomed to *congratulating* themselves for achieving an RPO of "just" several hours—but why in the world would any organization be happy about the potential to lose several hours worth of critical data?

Keep in mind that most high-availability measures *do not* improve RPO. Certainly, redundant disk storage can help keep a system online when a drive fails. Clustering can help keep services available when entire servers fail. But none of that protects against *data loss*, especially because most data loss in organizations is the result of accidental change or deletion rather than system or component failure.

How much data has your organization convinced itself it can stand to lose? If a file, database, e-mail message, or other data element was accidentally changed or deleted, how accepting would you be?

Why do organizations *ever* accept an RPO of anything less than *minutes*, meaning no data is atrisk for more than a few minutes after its creation or change? The answer to that question is fairly simple: Organizations accept poor RPO because, using their traditional data protection approaches, *that was the best they could do*. And that answer reveals the most insidious business risk associated with traditional data protection approaches: They've convinced us that they're *good enough*. They've made organizations complacent about inferiority, and made us stop trying to come up with a better solution.

Why is virtual + Physical so challenging?

It doesn't seem like data protection of virtual assets should be any more complicated than data protection for physical assets. So why is it so challenging? Why can't you effectively use the same decades-old protection approaches in the virtual + physical data center? The next chapter will explore specific challenges and look at why the virtual + physical data center presents a unique opportunity to rebuild our expectations for data protection.

Chapter 2: Challenges of data protection in the hybrid virtual + physical world

The previous chapter reviews the major business risks presented by traditional data protection solutions in a hybrid physical + virtual data center. Briefly, the major risks include the significant amount of data left constantly atrisk by traditional backup windows, the time-consuming and inflexible recovery processes, and the sheer amount of money-wasting administrative overhead involved.

But why, exactly, is virtual + physical data protection so challenging in the first place? If you're to create an effective, reliable approach for data protection in a modern data center, you need to fully understand the differences presented by the virtual + physical world. Eventually, you'll also need to re-state your business drivers for data protection.

The "Traditional" world

The previous chapter defined *traditional data protection approaches* as those technologies and techniques that have remained largely unchanged for decades. These approaches' hallmarks are scheduled backup windows, a reliance on magnetic tape media, and an architecture that was created well before the advent of virtualization.

Therein lies the major weakness in traditional approaches: They evolved when data centers consisted of just a few physical servers. In the earliest days, PC-based servers had tape drives directly attached to them. As the number of drives started to become unmanageable, organizations switched to tape libraries that were attached to a central backup server. As IT began moving from primarily relying on file servers to also relying on always-on databases,—think Exchange Server, SQL Server, and so on—the industry started creating technologies that let backup solutions extract static data from those databases during the backup window.

This process has really been a long game of workarounds. Too many tape drives? Switch to a single one. Tapes still too slow? Back up to disk first, and then dump to tape throughout the next day. Can't access open files? Create new technologies that can snapshot the open file for backup purposes or extract data directly from the database one bit at a time. We've never really taken an opportunity to re-think data protection from scratch, and as our environments continue to become more varied and sophisticated, the workarounds are starting to become less effective.

The modern environment is a moving target

One specific challenge, which was touched upon in the previous chapter, is that data centers are changing so much. It used to be that the data center was the most static piece of technology in the business. Sure, it changed—servers would come and go from time to time—but we deliberately slowed the pace of change in order to provide better stability and reliability. We *wanted* the data center to stay relatively quiet.

Now, data centers are likely to exhibit more change, and more continual change, than the client computing base! We try to keep our client computers locked down for reliability and stability, and in many organizations, client computers continue to run a decade-old operating system. Our data centers, in contrast, are positively bustling. We support numerous versions of key platforms such as SQL Server and Windows itself. We're migrating physical machines to virtual ones on a regular basis. Virtual machines are moving between hosts to balance workload and ease maintenance processes.

This constant rate of change creates a significant challenge for data protection approaches that evolved in a relatively static world. For example, if you're using a backup solution that directly backs up virtual machine hard drive files, that solution might back up different data from each physical host on any given day, as virtual workloads shift around between hosts.

Some organizations quickly realize that their traditional backup solutions don't work well in an environment that's increasingly virtualized. However, they make the unfortunate mistake of acquiring another solution to deal specifically with their virtual machines. Doing so can make matters worse, as the organization then has to support *two* (or sometimes more) data protection solutions, each with different operational patterns, different capabilities, different weak points, and different recovery processes. When a physical server is migrated to a virtual one, it also migrates to a different data protection s`olution. What happens to

the old backups? Will users be able to get the same response times and service levels under the new data protection solution? Will there be more, or less, data atrisk on the virtualized server?

There's really only one way to address this situation: Organizations need to adopt a data protection approach that handles *all* of their servers, both physical and virtual. Although this approach might well treat virtual servers somewhat differently in terms of actual backup technology, the approach needs to manage everything consistently, from a single "pane of glass." It needs to respond to the continuous change in the data center by providing the exact same capabilities and expectations across the entire back-end environment.

Data is changing constantly

Traditional backup approaches came into existence in a world where we didn't actually create all that much data. Let's face it: In the beginning, our servers were mainly file servers. You came to work, you updated a few files, created a few more, and you went home. You had only a couple of servers to worry about, and it was easy enough to pop a fresh tape in the servers' tape drives before turning off the lights and going home in the evening.

Today, that's hardly the case. Even smaller organizations create massive amounts of data almost continuously, day and night. The idea of putting aside a backup window in which youcan't make any changes because the backups are running is becoming increasingly absurd. Windows are shrinking, and they're becoming harder to schedule. Businesses are becoming less tolerant of even having backup windows, instead seeking to run systems continuously to maximize business advantage.

Traditional data protection approaches have simply evolved more workarounds:

- Technologies such asWindows' Volume Shadow Copy Service (VSS) essentially create snapshots that backup solutions can read, while allowing the original data file to remain inuse. This operating system-level support removes the need for specialized "open file" agents on servers but stillleaves a significant amount of data atrisk. Essentially, any changes to data after the snapshot is created remain atrisk until the next snapshot is made. Further, these technologies don't generically work with every kind of file on disk. Platforms such as Exchange and SQL Server must be written specifically to work with this service (and those two have been), but applications that aren't VSS-aware remain challenging for traditional backup solutions.
- Disk-based solutions can be created to mirror entire disk arrays, ensuring that a spare copy of data is constantly available. The mirror can be broken and backed up, giving the traditional backup solution full access to that copy of the data. The mirror can then be re-established and brought back into sync. This setup still leaves a great deal of data atrisk, and can be extremely expensive in terms of disk devices.
- For decades, you've been able to create incremental and differential backups. These help reduce the amount of data you need to back up during that precious backup window, but they add an exponential level of complexity to the inevitable restores and disaster recoveries by giving you yet more tapes to spin through and more backup data to manage.

None of these workarounds really do a better job of meeting business needs—they simply ensure that decades-old backup technologies have a chance of still working properly. The fact is that old-school backup solutions are just at the limit of their capability—if not beyond.

Restoration is a daily task

Traditional backup approaches evolved *primarily* to deal with disaster recovery: You back up *everything* and hope you'll never have to use it. But the fact is that youspend far more time recovering single items—files, mail messages, documents, database tables, and so on—than you do performing disaster recovery for an entire server or data center. And those traditional data protection approaches are ill-suited for that daily restoration task.

Initially, recovering a single file often meant restoring an *entire server.* Recovering a mail message might entail recovering an entire data store, mounting it to a spare mail server, and then digging through—using the server's native tools—to find the desired message. This process is obviously incredibly time-consuming, and many organizations evolved additional workarounds to make things more palatable. Client-side recycle bins, for example, gave users a chance at recovering their own data without forcing the IT team to resort to tape. Of course, this method meant storing even more data in online storage, giving you *more* data to back up in that shrinking backup window, and giving you *more* data to manage in the backup archives.

Other workarounds evolved. Many traditional backup solutions can, these days, recover individual items in certain cases. Perhaps the solutions extract every mail message individually, enabling data storage on-tape in a way that the solution itself can read. Or, solutions provide native ability to mount certain data stores, so that you at least don't have to have a bunch of spare servers lying around.

But these workarounds remain *workarounds*. Patches, not fixes. The ultimate problem is that traditional backup approaches simply weren't designed with single-item recovery in mind, and everything organizations do to suit them to that purpose inevitably comes with downsides: Bigger data stores. Longer backup times. More complex backup processes. More complex recovery processes. Small wonder that many organizations rely on disk-based, solution-specific archive solutions—such as e-mail archives, an increasingly popular approach—simply to remove the burden from their outdated backup solution.

Disaster recovery can be more flexible

The one thing traditional data protection solutions *did* evolve specifically to support is disaster recovery. Sadly, even that function has fallen behind the times, as the options for disaster recovery have far outstripped solutions' ability to keep up.

The traditional disaster recovery scenario involves restoring an entire server, from tape, to its original—or closely matching—hardware. Extended to a true *disaster* might mean recovering multiple servers. Again, the presumption is usually that they're being restored to substantially similar hardware.

Of course, traditional backup solutions can obviously back up and recover virtual machines as well, and their ability to do so provides a great deal of additional flexibility to disaster recovery scenarios. Because virtual machines typically emulate a fixed, common hardware specification, any virtual machine can run on any virtual host. That means your backup tapes can be more easily transported to a recovery site and used to re-construct at least your most critical servers.

As an industry, we've applauded ourselves for that additional flexibility. No longer do recovery sites need to have specific hardware in place in order to support server rebuilding; you can just dump virtual machines backups onto any suitable host and you're back in business.

But what flexibility have you really gained? You *still* have to maintain a recovery site, which is a large expense. Yes, it can be less expensive thanks to virtualization—but backups still have a significant amount of data at-risk—everything that happens between backup windows can easily be lost. In a true data center–level disaster, you'll likely lose even more, forcing you to fall back to whatever set of tapes were last taken off-site for safekeeping. Being able to restore to virtual machines is all well and good, but the point-in-time focus and reliance on physical tapes still presents the major hurdle to better disaster recovery.

Given the available technologies currently on the market, you should be able to do better. You should be able to replicate backup data off-site almost continuously, ensuring that your most recent backups are safe from a total site disaster. You *should* be able to apply that backup data more flexibly, restoring to virtual machines in your data center, another data center, or even in the cloud, bringing your most crucial servers back online quickly, even without access to a disaster recovery site. But traditional data protection approaches simply don't support that level of flexibility.

Why protect data, anyway?

If there's been a recurring theme in this chapter, it's that traditional backup approaches do a poor job of meeting actual business needs. These approaches put too much data atrisk, limit recovery flexibility, and adapt slowly to ever-more-dynamic data centers. So why still use them?

Mainly because it's been too long since IT thought about what data protection goals *should* be. Organizations accept what the traditional backup technologies can provide without pushing them to provide more of what the business needs.

It's time to re-state the goals for data protection, from a business perspective. What do you want, without regard for what technology might be able to provide? If the sky was the limit, what would you ask for? From there, you can see if you need to back down your expectations a bit and accept something less, but it's been a long time since the invention of magnetic backup tape. It's possible that you *can* have everything your business really needs.

Chapter 3: Re-examining the business drivers for data protection

The first two chapters established that traditional data protection approaches——those relying primarily on tape media, on limited backup windows, and on generally slow and inflexible recovery strategies—create significant business risks for a modern organization. The book has also examined challenges presented by today's hybrid physical + virtual data centers, and how traditional data protection approaches have been ill-suited to meet those challenges.

In fact, upon closer examination, traditional data protection approaches have *never* done a good job of meeting business needs. Organizations have accepted these approaches' limitations as inevitable and insurmountable. But as traditional approaches' workarounds and shortcomings continue to pile up, and you realize that you need to consider new approaches, you should really start at square one, re-stating and re-defining what it is your *business* should be getting from data protection.

This chapter, then, will shoot for the sky, without regard to the abilities or availability of technologies to meet your goals. The chapter will define goals that meet the business' requirements, and *then* consider whether you need to back off on some of those goals based on what today's data protection technologies can deliver.

The shortest possible RTO

The Recovery Time Objective, or RTO, is an organization's stated tolerance for downtime while a restoration is in process. Most organizations will define several RTOs for different scenarios: Restoring a single file, for example, might have a shorter expected time than restoring an entire server, which might be shorter than restoring an entire data center.

Regardless of the operation, however, you always want the shortest *possible* RTO for that given operation. With traditional approaches, the RTO is determined primarily by "overhead" issues: The amount of time it takes to retrieve tapes, load the tapes, spin through the tapes to find the needed data, and so on. Only a fraction of the total restoration time is spent actually copying data from the backups.

Shortening the RTO, then, will require you to reduce or eliminate overhead as much as possible. For example:

- You might rely more heavily on on-line and near-line storage, such as disk arrays, for backup data. Disks are faster than tape, providing a shorter restoration time.
- You might utilize data protection solutions that provide direct access to the backed-up data *without* requiring it to be restored. For simple files and folders, that's pretty straightforward; for database-oriented data, such as Exchange messages or SQL Server tables, it might require a data protection solution that can natively mount those databases without requiring them to be restored to a live server.
- Maintaining a full, online index of backed-up data makes it faster to locate the data you need to restore. Because finding the desired data is often one of the most time-consuming parts of a restore, such an index could significantly lower your RTO.
- A data protection approach that embraces and leverages automation as much as possible will also help reduce the RTO. Manual labor is always slower, more error-prone, and less consistent; automation will always speed things up and reduce or eliminate wasted effort.

The more you can reduce the overhead associated with a restore, the faster an RTO you can set for your organization.

The shortest possible RPO

The Recovery Point Objective,or RPO, is essentially a business' statement of how much data they're willing to have atrisk. Many organizations routinely accept an entire day as their RPO, meaning that if a disaster struck, the *point* they could recover to would usually be the previous evening's backup.That's a pretty long RPO and, as outlined in previous chapters, it's largely based on what traditional data protection approaches can deliver, rather than what you're really comfortable with.

Frankly, your pie-in-the-sky answer for RPO should be "zero," or very close to it. In other words, your business doesn't want to lose *any* data, ever. You want to be able to restore your environment, or any given piece of data, to the way it was *immediately* before a failure or unintended change.

That isn't easy to achieve, but there are certainly approaches to consider:

- Continuous data protection is obviously going to be required, meaning you dispense with traditional backup windows and instead back up everything, all the time, as changes occur. You might still have to tolerate a few minutes' of at-risk data, but that's far better than an entire day or more.

- Backed-up data would ideally be time-stamped, allowing you to not only recover the most recent version but also recall any specific version of that data from the recent past.

- Because continuous data protection will necessarily involve on-line storage and some kind of server, which would *itself* be a potential failure point, you'll need to build in redundancy of some kind so that your backups themselves are protected.

Easy day-to-day restoration

As mentioned in the previous chapter, traditional data protection approaches tend to be built with *disaster* in mind, meaning they're often constructed primarily to recover entire systems. But day-to-day restoration of individual files, email messages, and other data is what most organizations deal with most of the time.

Therefore, a modern data protection approach should make those day-to-day restorations as easy as possible. For some organizations, that might even extend to end-user self-service restoration mechanisms, but most organizations will prefer to retain control over data restoration. Self-service restoration does, after all, raise potential operational, data integrity, and security concerns, so having the IT staff handle restores helps to centralize the effort and avoid human error.

That said, the IT staff needs to be able to *easily* and *quickly* perform day-to-day restores. Many of the approaches that can help shorten RTO can also ease these day-to-day restore tasks:

- A data protection solution that provides direct access to the backed-up data without requiring it to be restored. Again, this setup might simply mean mounting the backed-up data as a browse-able disk volume without needing to actually restore the data. In other words, accessing the data *directly from the backup archive*, even if the data is in a database file such as an Exchange mail store or SQL Server database.

- An index of the backed-up data would, again, make it faster to locate the data to be restored.

The more that can be done to reduce the overhead and time of day-to-day restores, the better.

Easy, fast, testable disaster recovery

Having easier day-to-day single-item recovery doesn't mean that whole-system disaster recovery is any less important. This task also needs to be made as easy and as automated as possible.

Most important, whole-system disaster recoveryneeds to be easily *testable*, something that traditional disaster recovery processes rarely offer. You shouldn't have to send half the IT team off-site to test your disaster recovery process; you should be able to push a couple of buttons and bring up selected servers in a test environment—possibly a virtual one—right within your own data center. Yes, testing off-site recovery is important if it's a component of your overall recovery plan, but you should be able to *frequently* test your ability to bring critical systems back online *without* having to wait for the once-a-year trip to the off-site recovery facility.

Organizationsshould put parameters around "easy," too, because different people definitely have different definitions. Ideally, recovering an entire system to its most recent state should be as straightforward as selecting the system in the data protection solution, specifying a restoration target (such as a virtual machine in a test environment), and clicking "OK." The data protection system should take over entirely at that point, bringing the selected system back online on the designated target.

And of course disaster recovery should be fast—it's one of the main things you should define a specific RTO for. Most businesses, accustomed to the limitations of traditional data protection approaches, might accept "several hours" as a reasonable RTO for full-system recovery. They shouldn't: It should be possible to have a system back online in *minutes*.

For example, a recovery solution that was able to directly mount backed-up data as a usable volume could simply provide such a volume to a virtual machine, enabling the virtual machine to come online almost instantly, albeit possibly with a somewhat degraded level of performance. Other techniques might enable a data protection solution to bring a system back online *while the data was still being streamed* to the recovery target. Smart algorithms could allow the solution to prioritize data that users were attempting to access, getting users productive even while the recovery is still technically underway. *That* is a short RTO, and is something that business decision makers should push for.

Flexible disaster recovery

You should demand much more flexibility in your recovery options, too. Today, recovery approaches tend to focus on restoring a system to its original system, whether physical or virtual. You should push to break down that wall and create a variety of scenarios to choose from:

- Physical server restored to virtual machine
- Physical server restored to dissimilar physical hardware
- Virtual machine restored to physical server
- Physical server restored to similar physical hardware
- Virtual machine restored to virtual machine—running under the same, or even a different, hypervisor technology

There's no reason to settle for anything less; the technologies to support these scenarios all exist. They're just rarely incorporated in traditional recovery solutions and processes.

Think of the business flexibility these scenarios gives you, though. If a critical server dies, you can quickly bring it back up in a virtual machine while the hardware is repaired. If your entire site becomes inaccessible or unusable, you could spin up critical servers in someone else's virtualized data center—potentially even in a cloud environment. With this kind of flexibility, your disaster recovery plan can include numerous options that best fit a variety of circumstances, giving your IT team and management a plethora of options to choose from based on the specific situation at hand.

Minimize backup storage and management

Many of the goals outlined so far in this chapter will likely involve disk-based backup storage, which is not always inexpensive. That means you also have to set goals for controlling storage costs and reducing the potential administrative overhead related to that storage; for example:

- Minimize storage consumption by automatically collecting incremental and differential backup data into a periodic full backup image.
- Further minimize storage consumption by de-duplicating data on a per-byte basis.
- *Further* minimize storage consumption by compressing data using a variety of advanced compression algorithms.
- Reduce management overhead by automatically cycling older backup data out of the system, based on configurable top-down policies.

By aggressively minimizing the storage used by the data protection solution, and by automating the removal of older backup data, you should be able to make a disk-based backup system both affordable and easily manageable over the long term.

Can your data protection goals be met?

With your business goals expanded and clearly stated, it's time to determine whether you can actually meet those goals using current data protection technologies. It's fairly obvious that traditional approaches *won't* meet these goals, so you'll need to consider newer approaches. That discussion is coming in the next chapter.

Chapter 4: Modern approaches for data protection in the virtual + physical data center

Having outlined business goals for a more modern data protection approach, it's time to see whether such an approach actually exists or can be created. To briefly summarize those goals, you want a solution that:

- Minimizes the Recovery Time Objective (RTO), getting you back online as quickly as possible
- Minimizes the Recovery Point Objective (RPO), putting the least amount of data atrisk as possible
- Supports both day-to-day single-item restoration as well as complete disaster recovery
- Makes disaster recovery testable
- Supports flexible disaster recovery options, including P2V, V2V, V2P, P2P, and even cloud-based x2V
- Minimizes backup storage requirements and day-to-day management overhead

Youknow that traditional data protection approaches won't work. They're too focused on grabbing data snapshots during backup windows, putting too much data atrisk. They're inefficient at both single-item recovery and whole-system recovery. So you need something different.

Going the opposite direction on data protection

There are several reasons that traditional data protection approaches don't really do a good job of meeting real business needs.First, traditional approaches depend entirely on grabbing snapshots of data during backup windows, often just once a day—or sometimes even less often. Next, consider the exact opposite approach: Grabbing data as *it changes*, in the smallest increment possible. That would result in *continual* backups, which would definitely do a better job of meeting many of your business goals.

Traditional evening backups become problematic simply because of the massive amount of data they have to grab in a relatively short period of time. It's not unusual for an incremental or differential daily backup to copy gigabytes' worth of data, and that just takes a lot of time. Many organizations can't tolerate too many differentials or incrementals—the differentials get bigger and bigger every night, while more incrementals slows down the recovery process—so they still have to grab full backups on occasion. Whether those full backups take place weekly or monthly, it can mean grabbing *terabytes* of data during a shrinking backup window. It's just impractical.

But if you could grab data as it changes, then you'd really only be backing up a few *kilobytes* at a time, constantly, throughout the day. Think about it: The smallest increment of storage in a Windows disk system is the disk block, which is the smallest unit of storage that can be written to disk. Disk blocks are usually just a kilobyte or two, so even a massive new file can be broken down into kilobyte-sized chunks. Grabbing those wouldn't be a big deal, and needn't present a lot of overhead.

Of course, how you store and use those kilobyte-sized chunks will determine how well this approach actually meets your business needs. So, with that general "continually grab small chunks" technique in mind, let's put some more specifics around the overall approach.

Block-based backups for servers

The approach you're considering here is called *block-based backup*. Essentially, it involves installing a small, low-overhead *volume-level filter,* sometimes called a *shim*, on your servers. Connected to a traditional software agent, this filter is registered directly with the server file system and gets to "preview" any changes that are being written to disk. By grabbing those changes as they're being written to disk, the software is able to grab each and every change to the server, as it happens, and transmit those to a data protection archive.

This approach *completely* changes how you think about backups. Every piece of data that matters on a server is eventually written to disk—including the data in Exchange mail stores, SQL Server databases, and so forth. You don't need any specialized knowledge of an application in order to back up its data, because that data will always end up on disk eventually. And once it does,your little agent would see it.

You no longer need worry about open files such as databases because you're not working with files. You're working directly with changes to the low-level disk storage, as those changes are received by the operating system and being written to disk. You've essentially crawled "under the hood" of every application ever written, down to their lowest common storage denominator. It doesn't matter how data gets to the disk, it's going to get there eventually, and you'll be able to back it up.

Those individual disk blocks can also be time-stamped as they're being backed up, giving you an incredible amount of granularity in the data protection archive. In theory, you could not only reconstruct entire files, mail messages, and so forth by reassembling their most recently changed disk blocks but also reconstruct those items to *any point in time* by simply restoring the disk blocks associated with that point in time.

This new approach to backups is powerful. It lets you treat physical and virtual machines absolutely the same, regardless of the hypervisor technology in use. After all, *inside* those virtual machines is just a normal operating system, which thinks it's writing disk blocks to a physical disk device. You simply run your data protection agent *inside* that virtual machine, capturing disk block changes exactly as you would from a physical server—and getting all of the benefits outlined.

This approach helps you meet another, overriding business requirement: Treating physical and virtual machines exactly the same, backing them up from a single mechanism. Migrating a physical server to a virtual machine? No problem: The data protection solution continues to operate *exactly the same in every way.* You get the exact same management processes. Users get the exact same recovery expectations. You get the exact same, low overhead per server. *Nothing changes.* Migrating a virtual machine back to physical? Nothing changes.

Because you've gotten down to what is literally the lowest possible common denominator in storage, you can back up anything that hits that storage—no matter *how* it writes its data to disk. Applications don't need to be aware of any proprietary technologies such as Volume Shadow CopyService (VSS); so long as they're writing data to disk, you've got them backed up all the time, throughout the day.

Backup windows go away. Systems can remain operational for longer because you don't need a backup window anymore. Instead, scheduled maintenance windows can reflect your *maintenance* needs, such as patching or other updating, rather than focusing on data protection. Data protection just becomes a quiet, continuous thing that always happens, with little or no thought on your part.

Live data protection archive

All of those disk blocks will be transmitted to a backup server, which will store them in some kind of archive. The construction and capabilities of that archive is what really determines the value of this data protection approach. After all, many of the shortcomings of traditional data protection revolve around its poor abilities to *recover* data quickly, so this archive is going to have to lend itself to those activities.

Let's start by proposing an archive that can maintain itself. In other words, it would periodically collect all of those backed-up disk blocks into a single, coalesced point-in-time image of the disk volume. Doing so removes a level of granularity for your recovery, meaning you could recover only to the point in time represented by that image. That's roughly equivalent to the traditional approach of discarding incremental and differential backups after getting a more recent full backup—but in the new approach, the process happens

automatically on a schedule you set. Your backups are *only* ever incremental or differential; the solution *constructs* the "full backup" from those on a periodic basis. That helps to reduce storage consumption and lower administrative overhead.

The archive could also utilize compression and block-level de-duplication to further reduce consumption. Other solutions that utilize those techniques can often achieve storage savings of up to 80%, which is significant, and would enable you to keep a longer period of more-granular backups, if desired.

Because it consists of backed-up disk blocks, such an archive should be able to quickly present an image that *looks like a disk volume*. That could easily be made into a mountable image, meaning you could treat it like a disk drive, browsing for whatever files and folders you needed to recover. Because the backed-up disk blocks are time-stamped, you could have the solution create a volume that looked like a snapshot of *any given point in time* in recent history, letting you recover files to a single, specific point in time, not just the most recent point in time.

This approach would help drastically reduce the overhead and time consumption of those day-to-day single-item recoveries. Just pick a point in time, push a button, and you've got a disk volume that you can browse and copy files off of.

But what about complex storage, such as Exchange mail stores and SQL Server databases? Would those still have to be copied—that is, restored—to a live Exchange or SQL Server system in order to be read? Not necessarily. The formats of those files are documented and wellunderstood; it would be straightforward to construct the data protection archive itself with the ability to open those files and let you browse or search for individual items. Again, doing so would drastically reduce the time necessary to recover single items from those data stores.

And that's just one possible approach. Given that you're considering an archive that can present backed-up data as a live disk volume, it might be possible in some cases to simply attach databases—such as a SQL Server database—to a live instance of SQL Server. After all, SQL Server doesn't much care about what kind of disk storage you're using, so long as it *looks* like a Windows disk volume. Accessing the backup data in that way—by using native technologies that see the backup archive as plain disk storage—would enable you to use the products' native capabilities to extract data and perform other tasks. You could, using that approach, even use something such as SQL Server to *compare* backed-up data to live data. This method might work in instances where a data restoration wasn't actually necessary but you needed to see what had changed.

New approaches for whole-system disaster recovery

Whole-system recovery will change with this data protection approach, as well. For starters, *you'd always have a complete disk volume available at the touch of a button*. You could, in theory, spin up a virtual machine and have it use the disk volume directly from the data protection archive, meaning you could get back up and running in just seconds, at least to some degree.

Disk blocks could also be streamed to any recovery target—physical, virtual, cloud-based, you name it—giving you a great deal of recovery flexibility. The backed-up disk blocks could also be streamed off-site, giving you incredibly flexible disaster recovery options. For example, streaming backup data to a cloud-based provider would enable that same provider to spin up virtual machines from your most recent backups in just minutes. So even if your entire facility suffered a disaster, you could still get critical servers up and running almost immediately, with very little data loss.

Again, it's all about *flexibility* and *speed*. Because the backup archive already has your backup data available as a full-disk volume, there's essentially no restore process—just attach to the data and start using it. Or begin immediately streaming it over a high-speed network connection to a target server, either virtual or physical.

There are potentially a lot of clever capabilities that could be built into such a recovery system. For example, a server under restoration could have data streamed to it based on user data requests. In this case, the data needed by users *right then* could be prioritized and restored first, getting the server back online—albeit at a reduced performance level—much faster than having to wait for the entire server to be restored.

The real value in disaster recovery these days is virtualization. After all, in a disaster situation,organizations are often willing to accept lesser performance in exchange for functionality that would otherwise be lost. Thus, even stacking more-than-usual numbers of virtual machines onto virtualization hosts is acceptable. The data protection approach this chapter is outlining

is ideal for disaster recovery because this approach lets you spin up virtual machines almost instantly, because the backup data is already live and accessible. If such a solution could also convert its volume snapshots to virtual disks—in any hypervisor format—then it'd get you up and running even faster with whatever tools came to hand.

It's about *removing restrictions*. Your disaster recovery plan needn't be a single, one-size-fits-all plan. Instead, it can be a range of options to fit a variety of circumstances. Your IT team needn't be treated like a bunch of trained monkeys just executing a single-path plan; instead, they can apply their experience and knowledge of your business to leverage the right option for the scenario. Need to recover your VMware virtual machines to a Xen-based virtualization cloud? Fine, you can do that. Need to spin up a dozen formerly physical servers as virtual machines at a recovery site? Fine. The point is to give you and your team options so that your *exact* of-the-moment business priorities can be met. Your data protection solution should be an enabler, not a bottleneck.

Does this new approach meet your business goals?

What you've looked at in this chapter is definitely a new approach to data protection. But *new* isn't necessarily better; what's important—really, all that matters—is whether or not this new approach aligns well with your stated business goals for data protection. The next chapter will look at that in quite a bit of detail, analyzing how this approach's specific capabilities enable—or not—the needs you've identified for your business' data protection.

Chapter 5: How better data protection approaches meet business goals and mitigate business risks

The previous chapter examines a new potential approach for data protection. Relying on block-based backups for servers (and technically, client computers as well), this approach promises near-real-time continuous backup. Coupled with a well-designed backup data store, this approach should enable rapid recovery, lower overhead, and flexible restoration options.

But all of that means nothing if the approach can't meet your business goals, which were clearly stated in the third chapter. This chapter will return to those objectives to see how this new, block-based data protection approach stacks up.

Recovery time objective

does this new approach offer a shorter Recovery Time Objective (RTO) than traditional data protection approaches? As earlier chapters outlined, organizations can't have just one RTO. Just as there are different restoration scenarios, there are going to be different RTOs:

- It's tough to imagine a faster possible RTO for day-to-day single-item recoveries than that offered by this new approach. Given that the new approach has the ability to mount the backup store as a live, readable volume, the actual restore time is technically near-zero. It may take administrators some time to *locate* the item they want to restore, and in the cases of items such as a SQL Server database, a less-skilled administrator may need more time simply due to the overhead of the native technology. But the backup solution isn't adding any time to the process.
- In terms of full-system recovery, the RTO will depend a bit on the exact recovery scenario you choose. Restoring to a virtual machine, for example, can be incredibly quick because the backup solution is capable of presenting the backed-up system's entire disk volume as a virtual hard drive, ready to be used by a running virtual machine. That's an RTO of just a few minutes, depending on how quickly your administrator can get the process started.
- A full datacenter recovery—or at least the recovery of critical systems—can also be incredibly fast via the new approach. If you've been replicating backup data offsite—something the proposed approach would be capable of—then you can use the offsite copy to quickly spin up virtual machines. Again, your RTO would be measured in minutes, with much of that time attributed to overhead not related to the backup system, such as provisioning offsite virtual machines.

So this approach definitely meets the goal of the "shortest possible RTO." It isn't always zero, but it's logarithmically less than traditional tape-based backups, and even faster than traditional, point-in-time disk-based backups that still require a lengthy restoration process.

Recovery point objective

Does this new approach help to minimize the amount of at-risk data? Absolutely. By continuously backing up individual disk blocks as they change, there's essentially no data atrisk. Obviously, reaching an Recovery Point Objective (RPO) of zero isn't technically feasible in most cases, but this new approach should be able to push the RPO as close to zero as possible.

With a modern network available to it, this new approach should be able to reach capacities in the neighborhood of 8GB of data *per minute*. That's a continuous rate of more than 3TB of changes data per workday—an incredible capability. With that much speed, such a solution should be able to quickly capture whatever changes your organization is piling onto a server, and capture them *fast*. Your RPO should be measurable in single-digit minutes, if that's what you need for a particular server.

This *continuous backup* capability is what really "sells" this new, block-based approach. Think about it: Your backup windows are *gone*. There are no more full backups, no more incremental backups, no more differentials. Backups are no longer a *task* that gets performed; backups are just something that happens automatically and autonomously. Virtually no data in your organization is *ever* at risk, at any time of day or night. As soon as data is created or changed, it's safely backed up.

Backups are as easy as breathing. Take a moment and think about how that will change your IT team's daily lives. No longer will you need to juggle applications and data, worrying about which ones support such-and-such a backup technology or which ones will be problematic. No longer will you have to worry about operating system-level support for backups—support that changes with new version releases. That's all gone. Simply make sure that critical data is being saved to disk—and where else would an application save it?—and it's backed up, immediately and without effort.

Day-to-day restoration

Does this new approach make the tedious, day-to-day single-item recovery processes any easier? Without a doubt, yes.

This approach is capturing disk blocks, making it very straightforward to simply present those disk blocks as a complete disk volume. It's basically like having a mirror image of your servers' storage systems, capable of being mounted as a readable volume with the click of a button. From there, just find the item you're after, and copy it to wherever you'd like it. Need to restore something other than the most recently backed-up version of a file? No problem: Just pick your point in time, and let the solution re-assemble the volume using the disk blocks from that point in time.

Once again, this approach completely changes the daily lives of your entire IT team. Without needing to shuffle tapes, scan for files, or perform any of that other onerous recovery overhead, they can simply reach in and grab whatever data they need. Mount databases to their native technologies without restoring them, and copy out whatever data is required.

This approach is such a radical improvement over traditional approaches, especially tape-based ones, that it'll actually take you and your team some time to realize just how much easier things have gotten. It's like taking a complex route through back roads to get to and from work, and then one day realizing that there's an empty freeway that takes you practically doortodoor. The time you'll save, the effort you'll save, the frustration you'll *eliminate*—you almost have to experience it to fully appreciate it.

Disaster recovery

Disaster recovery is the task organizations tend to focus on when they think of backups and data protection, so does this new approach make things better?The short answer: Immeasurably so. Think about all the pitfalls and downsides of traditional disaster recovery, and how this new approach helps to eliminate them:

- Reliability. Tape media, the traditional backbone of disaster recovery, remains notoriously unreliable. You often don't realize that there's a bad tape sector until you're actually in the middle of a recovery—which is hardly a good time. This new approach eliminates tapes as well as self-tests the data it backs up for accuracy.

- Speed. Disaster recovery involves copying a lot of data, and that takes time. Unless, of course, you have a new approach that can stream data to target systems, prioritizing the data based on user need and getting systems back online in minutes, not hours. That's exactly what this new approach offers.

- Ease. Disaster recovery is usually a convoluted, complicated process. Restore the full backup. Then the incrementals. Then the differentials. Restart the computer. Resolve any hardware-related problems caused by the restore target being different hardware than what was backed up. Cross fingers. Hope for the best. This new approach eliminates all of that: Choose a server. Click recover. Choose a target. Tell the users to get back to work.

- Testability. Testing traditional disaster recovery plans is painful, timeconsuming, and expensive—and so you don't do it nearly as often as you know you should. With this new approach, you can test disaster recovery—easily—whenever you like. Simply recover whatever servers you want to a virtualized disaster recovery test environment. Do it in the middle of the day, with your users none the wiser. Test a recovery every month to make sure your processes are working. It's easy.

There's a massive amount of flexibility in this new approach as well. Restore any server to any target: physical, virtual, cloud-based, whatever. Simply by clicking a couple of buttons, you can have entire systems back up and running in just minutes. Easy enough that anyone on the IT team can do it, and fast enough that users will think someone just accidentally rebooted the server. With this new approach, disasters aren't so disastrous anymore. Whether you need to bring a single server back to life or recover an entire data center, this new approach makes it fast and simple.

Minimizing storage utilization

With all of those disk blocks being backed up all the time, this new approach certainly seems capable of using unprecedented amounts of storage. Fortunately, it can take advantage of modern technologies to compress and de-duplicate data—reducing storage consumption by up to 80%. So even with that continuous, no-data-at-risk backup process, you'll probably *still* use less storage space than your traditional data protection approaches.

Storage utilization can be further optimized by managing the store. Essentially, each captured disk block represents a tiny incremental backup. By saving each time-stamped block, the solution lets you restore data to any given point in time. However, it's unlikely that you'll need to retain that level of recovery granularity far into the past. So you define top-level policies that dictate exactly how much recovery granularity you need. Once the data store passes that point, the backup solution can automatically merge those incrementals into a full snapshot. Unlike traditional approaches, *you* never have to worry about incrementals or differentials; it's done for you. By simply defining your business goals for granularity, you can enable the solution to use as little or as much disk space as you want.

Minimizing administrative overhead

Traditional data protection solutions tend to take a lot of overhead. First, you have to constantly check to make sure the backups *worked*. There are tapes to shuffle on- and off-site. When you actually need to recover data, you have another whole process to contend with.

The new approach essentially does it all for you. It's looking at disk blocks—tiny little chunks of data—so it's easy to verify them as they're backed up. If there's a problem, the solution re-captures the data and notifies you with a proactive alert. There are no tapes to shuffle—although if you need Tdata off-site for safekeeping, built-in data replication simply handles *that* for you as well. As outlined earlier, even the backup store can be self-maintaining, merging older incremental disk blocks into a point-in-time full snapshot on whatever basis you define.

Set it and forget it. That's pretty lowoverhead.

It's time for better data protection

It seems clear that traditional data protection approaches, bound in backup windows, slow media, and high overhead, have *never* really met true business needs. You've accepted it because it's what was possible at the time—but times have changed. Newer technologies and better approaches are available: With block-based server backup and an intelligently designed backup archive, you can finally meet the business objectives you've had all along: no compromises, more efficiency, and better recovery flexibility. It's all within reach.

About the Author

Greg Shields is a Senior Partner with Concentrated Technology. With fifteen years of IT experience, Greg is one of the world's leading experts on virtualization, cloud, and systems management technologies. He is a Contributing Editor and columnist for Microsoft TechNet Magazine and Redmond Magazine, has written over fourteen books, and contributes regularly to online publications like TechTarget and MCPMag.com. He is also a highly sought-after and top-ranked speaker for both live and recorded events, and is seen regularly at conferences like TechMentor, the Microsoft Management Summit, Microsoft Tech Ed, VMworld, Connections, among others. Greg is a multiple-year recipient of Microsoft's Most Valuable Professional (MVP) award and VMware's vExpert award.