# Getting Hyper-V Backups Right

Storage Switzerland, LLC

# Table of Contents

# Chapter 1: It’s Important to get Hyper-V Backups Right

*Written by W. Curtis Preston*

It’s very easy to find yourself far down the path of virtualization and forget something so trivial as backup. This is not a problem new to virtualization, as backup is often on the sidelines in many people’s eyes. But with virtualization, it’s even more possible to forget backups since creating a new “server” is as simple as pressing a button. But there are a number of things you really must take into account when embarking on a Hyper-V virtualization plan.

It should go without saying, but without proper backups, there can be no restore. In this day of increased use of the cloud and increasingly reliable storage, sometimes we forget that hardware does fail. When it does we need a backup system that works. Without a good backup system, your new Hyper-V system could one day be nothing more than a paperweight.

The most obvious part of the backup setup is you need regularly scheduled backups that are based on a standardized configuration. No one should have to remember to do backups; they should just happen. As said earlier, backups are rarely at the forefront of most people’s minds, which means they will get forgotten if performing them relies on someone’s memory. IT should configure backups once with a schedule to run forever.

There should also be a retention period for backups able to deal with a number of different recovery scenarios. You cannot, for example, make a single backup each night that overwrites the backup from the previous night. The fact you are overwriting the previous backup means while you are conducting that backup, the system is deleting the previous backup. It is important to keep multiple versions of backups in order to recover from a number of scenarios.

Consider, for example, what happens when someone discovers that a very important database table or directory was deleted over a week ago. Storing backups for only a few days will not allow you to recover that table or directory, and that data is gone forever. You need to examine your environment to see the types of recoveries you need to recover from, including user error, data corruption, and outside attacks such as ransomware.

Configure your backup policies to be able to protect against them. Determine backup retention policies based on the business requirements of the company. But few companies can get by with anything less than 30 days.

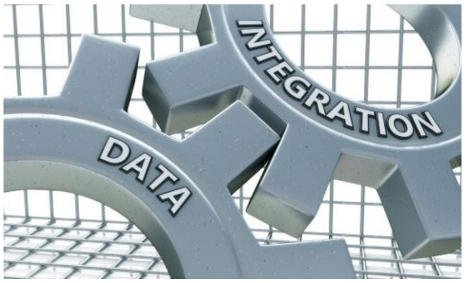
You also need to configure the backup system to interface with the virtualization system. It is a bad idea for many reasons to simply install a backup client inside each VM and back it up as if it were a physical machine. This uses too many I/O resources and typically costs more from a backup software licensing perspective. It’s better to install a client at the hypervisor level so it communicates with your VMs as VMs. This means it will also communicate with any applications that need to be put in any sort of special status prior to backup. This is typically done via the Volume Shadow Services, or VSS.

Backing up things at the hypervisor level will also ensure that you can recover any component of the system that goes bad. You should be able to restore individual files inside the VMs, although this may require the installation of a backup client in order to make that happen. It should also allow you to recover entire VMs or containers if someone accidentally deletes them or your storage fails.

Backing up Hyper-V is not hard, but it does require some forethought. Like anywhere else in IT, ignore proper backup design at your peril. Learn the proper methods to backup Hyper-V and use them, which includes regularly scheduled backups, enough retention to handle all recovery scenarios, and backing up at the Hyper-V level so that you can recover individual files, VMs, and containers. Once backups are configured and regularly scheduled, you can concentrate on building your Hyper-V environment.

# Chapter 2: The Value of NAS Integrated Backup in a Virtualized Environment

*Written by George Crump*



Network Attached Storage (NAS) servers are widely used for storing virtual machine backups. What's been lacking is a method to integrate the two workload types that many consider mutually exclusive: backup software and backup storage.

## The Advantage of On-Appliance Backup

If the data protection software is integrated onto the NAS device this opens up several possibilities. First, the NAS appliance now becomes a backup appliance with the ability to backup other systems in the environment. Most backup servers read data from the source datastore over the network and then sends that data one more time across the network to a backup storage. Integrating the backup software on the appliance eliminates one of these network hops.

If the NAS is being used as storage for VMware or Hyper-V virtual machines then a NAS with integrated backup software can backup those virtual machines to a separate volume on the same NAS. Ideally the backup software will include compression and deduplication so the backed up copies do not require as much storage capacity as the production environment.

It is important that the integration is intelligent enough to know that the virtual machine images and the backup images are on the same system. With this knowledge, the combined solution can perform these transfers while bypassing network protocols that will enhance overall performance.

## 3-2-1 Backup Still Required!

The age old philosophy still holds; three copies of your data, on two different types of media, with at least one copy off-site. While the NAS integrated backups are fast and efficient, they are somewhat vulnerable. But, let's be fair, most NAS systems are very reliable and include redundant hardware, RAID protection, snapshots and now a backup copy to a second volume, also with its own protection. The only remaining step is to get the data secured to an off-site location.

## Enter Cloud Enabled NAS

If the NAS system can replicate its data to either another NAS or to a public cloud provider then administrators resolve vulnerability concerns of NAS integrated backup. The public cloud, especially for small to medium sized data centers is particularity attractive because it not only provides a disaster recovery capability, some NAS systems can integrate with public cloud services to provide file sync and share. That solves yet another common IT problem.

Alternatively, or in addition, the backup software could directly interface with a public cloud provider and replicate the backup data there. The challenge with some backup solutions is supporting the cloud as a backup destination is mostly an afterthought and getting that integration working correctly can be challenging. More modern solutions integrate directly with cloud providers like Amazon, enabling a fast and simple use of the resource.

Most small to medium sized data centers will leverage a NAS as their primary storage system, including using it to host their Hyper-V or VMware environment. They may also use NAS as a backup storage target. It makes sense then to select a backup software solution that can integrate directly with the NAS so backups of those environments can be fast and efficient.

# Chapter 3: Why do Hyper-V Backups Waste Space?

*Written by W. Curtis Preston*

There are too many traditional backup products trying to tackle a nontraditional backup problem. As a result, they continue to use some of the old world backup techniques in the very new world of IT. These techniques waste storage space, network bandwidth and computing power. Let's take a look at the two main ways they do that.

## The Full Backup - A Holdover from Days Gone By

Whether you are backing up Hyper-V at the Hyper-V-level or at the VM level, you are most likely still performing backups that have a feature leftover from the tape era – the occasional full backup. Most backup software products were designed during the days when backups were sent to tape. In fact, many of these products were written during a time when very few customers used tape libraries. During the tape days it was very important to limit the number of tapes needed for an individual restore. This was even more true if a customer was not using automation, because every tape had to be manually put into the tape drive. I remember the days of literally sitting in front of a tape drive being forced to swap tape after tape. What a pain.

The best way to limit the number of tapes needed for a restore was to perform a regular full backup, typically once a week. Most people did this on the weekend when fewer people were using the servers, because a full backup impacted the performance of the server and the network. The weekly full backup became a standard which continues to this day.

In the last 10 years or so, backup to tape is not so common, basically non-existent in the mid-market. Environments like Hyper-V are typically backed up to disk. So one would think that the idea of a full backup has been done away with. Yet vendors still build backup software packages around this idea of an occasional full backup. This is because they build software around the idea of first restoring a full backup and then restoring the associated incremental backups – like we did back in the day. If one never performed a full backup, the software would eventually not be able to perform a restore during a reasonable amount of time.

One idea that some backup software products have started using is a *synthetic full backup*, where a full backup is synthetically created from incremental backups. A backup system performing incremental backups already knows all of the files that are still present and any files that are no longer there. This means it has all of the knowledge it needs in order to synthetically create a full backup from all of the incremental backups that it already has. Depending on its implementation, a synthetic full backup system can save storage space, computing power, and networking resources.

## Full File Incremental Backups - Another Holdover

The second offender from a space-saving perspective is full-file incremental backups, defined as a backup that backs up an entire file if any part of that file changes. This has always been a space waster, but it's a particular problem with applications like Hyper-V.

On one level, Hyper-V is just another Windows application. From a backup perspective, it's a lot like a database application, because it uses a number of very large files (e.g. VDK files), all of which change every day. Just like a database application, when anything happens on a Hyper-V VM, it's going to change the modification time and archive bit of the VDK files underneath that VM. This means that any application backing up a Hyper-V server without proper integration be performing the equivalent of a full backup every single time. This is why backup products must integrate with Hyper-V and perform block-level incremental backups.

Full backups must go, at least in the way we know them. They waste space, network bandwidth, and computing power. Administrators can use synthetic full backups without such negative impacts, if they're designed in a way that minimizes these things. Ask your vendor about how their synthetic full backup saves backup storage space. If they say it doesn't, then they only focused on the network and CPU savings. Look for a backup product that saves storage space as well.

# Chapter 4: Why you Should Back Up Hyper-V to the Cloud

*Written by W. Curtis Preston*

If data in your Hyper-V system is worth backing up, it's worth copying and storing off-site. Too many things can go wrong to a data center that can damage both the servers and backups – if those backups are stored in the same location. If they're also located somewhere else, however, you can survive just about any disaster.

In fact, having a copy of the data off-site is just one step in what is called the 3-2-1 rule of backups. Keep at least three copies of your data on at least two different storage types and keep at least one of those copies off-site.

Keeping three copies is easy enough to understand, so let's talk about having them on two different storage types. The way to do this is to either store a copy on local storage and another with a cloud provider, or to store a copy with different cloud providers. You can also do both of these things, creating a local copy and two cloud copies stored with different providers. The idea is to not allow a single type of outage – such as an outage of your cloud provider – take out all your backups.

Most backup products have the ability to create a local backup before copying it elsewhere. The advantage of having a local backup is that it makes for a much faster recovery. But as mentioned previously, having only that copy is problematic. Therefore, we need to copy the data to at least one other location – two would be even better.

In the old days, we would accomplish this by simply copying the tape and handing it to an off-site storage vendor. That practice hasn't been considered a best practice for a long time due to the incompatibility of tape and the typical backup process. The modern-day equivalent is to electronically copy it to another location. This could be another data center that you manage, or a cloud provider. This second copy needs to be in a place that is highly available and is capable of supplying the data at a high rate of speed for a large restore.

One challenge with storing copies of all backups at a cloud vendor is that the type of storage they use to provide high-availability and high-bandwidth often costs more than the alternatives, such as infrequent access storage or cold storage. But you can have the best of both worlds by keeping the recent copies on the highly available, high-bandwidth, costly storage, and keeping older copies on less expensive cold storage.

If you combine these three methods, you have a very strong system. First, create a local backup used for operational recoveries and large recoveries. Keep enough local storage to keep a week's worth of backups. Second, replicate each backup to a cloud provider that can provide the data with high-bandwidth for a disaster that takes out your entire data center. Finally, replicate some of those backups to an even less expensive cold storage cloud provider for longer-term storage. (For reasons outside of the scope of this article, though, you should not store your backups in the cloud forever. That is what archive software is for.)

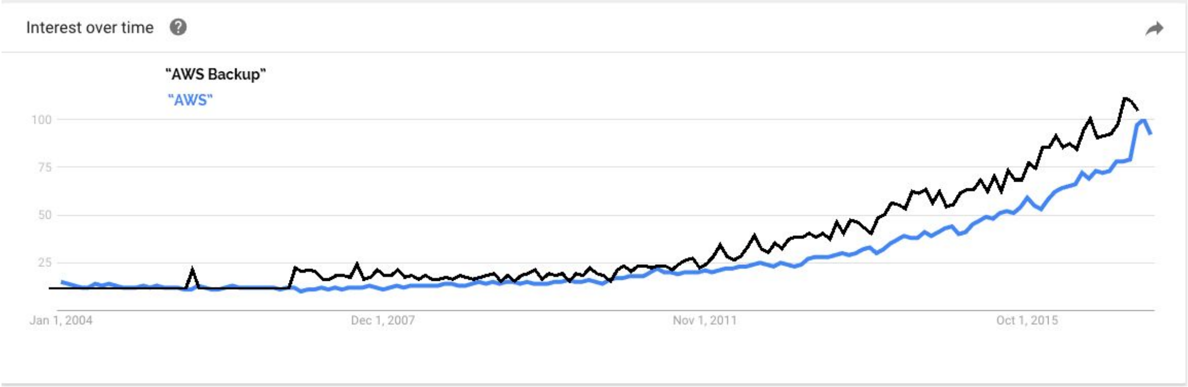
It is possible to backup your Hyper-V system using modern technology and still follow the old 3-2-1 rule of backups. Rules of thumb like this are just as valuable – if not more valuable – as they were back when we were making tapes. Cloud providers fail, storage arrays fail, and fires and terrorism still exist. The 3-2-1 rule still applies. It's just how we apply that rule that is different.

# Chapter 5: Do AWS VMs Need to be Backed Up?

*Written by W. Curtis Preston*

I'm amazed at how often the question of backing up virtual machines running within the Amazon Web Services (AWS) infrastructure comes up. The short answer is, "absolutely!" Thinking that your VMs in AWS do not need backups is a surefire way to lose data. Of course, the next question is about how you accomplish AWS VM backup. That question takes a little longer to answer.

When I was researching for this column, I used Google Analytics to see how often the phrase "AWS backup" was searched over the last 10 years or so. Would it surprise you to learn it is searched more often than "AWS?" Take a look at the graph below. The blue graph shows how often people searched on the phrase AWS, and the black line shows how often they searched on AWS backup.



First, I have to say I am encouraged by this graph. I am not the only one in the world who thinks data protection is important! The fact it was searched more often suggests people were interested in migrating to AWS, but only if they were able to protect their data. It also shows a lot of people aren't quite sure about the topic.

## Why You Need to Backup AWS

There is nothing built-in to Amazon Web Services that protects your VMs or the data within them. While this may seem surprising to some, there is no more or less data protection built into in Amazon VMs than there is in any server that you would purchase from your favorite vendor.

The same could be said of your favorite server virtualization vendor. They provide virtual machines and data storage for said machines. The protection of the data on those machines is your responsibility. Yes, some services and offerings bundle (an often rudimentary) backup application, but you have to enable it.

Anyone who disagrees with the above paragraph should immediately Google codespaces.com. This company offered a website where you could store your code. Its service was entirely within Amazon, as were any backups that they had made. Its account was hacked, passwords changed, and because administrators had not previously activated two factor authentication and had not done off-cloud backup, their entire company ceased to exist with a few keystrokes. They, of course, reached out to Amazon for help. Amazon responded with the statement that backups of AWS VMs are the customer's responsibility.

## How to Backup AWS

There are a number of ways to backup Amazon VMs now. Unfortunately, there is no ability yet to back up at the hypervisor level like you can with VMware and Hyper-V. All backup solutions will either run an agent in the VM or connect to a service already running in the VM, such as *sshd*. While each organization will have its own requirements generally you should look for the ability to regularly schedule backups that will only transfer the blocks that changed since the last backup. This is crucial to making Internet-based backups feasible and for reducing your AWS bandwidth bill. But the most crucial feature should be the ability to automatically get your backups off some infrastructure different than the one you’re running on, as suggested in the 3-2-1 rule. Have three backups on at least two types of media, one of which should be off-site. Storing all your backups on the same Amazon account you’re backing up breaks two of those rules.

Some options are available in the Amazon Marketplace, and others sell their services externally. While existence in the Amazon Marketplace does help you to ensure that the solution you are considering is already certified with AWS, lack of presence in the Marketplace should not be equated with lack of quality in a product. The decision to offer a product in the Amazon Marketplace is more a business decision than a technical one.

You need to protect your VMs running in AWS. In addition, make sure you enable two factor authentication on your AWS instance. And if at all possible, the backups of your AWS VMs should also be located on a different infrastructure. Even if you back them up to S3, for example, you should figure out how to replicate S3 to another system like Google cloud. That way a single hack cannot take out your company the way it did codespaces.com.

# Chapter 6: Why is Operational Recovery Needed?

*Written by W. Curtis Preston*

The reason backup strategies exist is to ensure operational continuity after someone makes a mistake, purposefully corrupts data, or hardware or software component runs afoul. Operational recovery is the day-to-day reason you backup data. It is the recovery of specific parts of the IT infrastructure, a file, an email message, a component of Active Directory or now even a virtual machine.

## Types of Operational Recovery

Operational recovery of accidentally damaged data is the most common thing anyone responsible for backups will find themselves doing. Someone will accidentally delete a file, or someone else will accidentally delete an entire directory. The biggest restore I ever participated in was due to a shell script coding failure that deleted thousands of home directories.

One must also consider how important email has become, as many people use their email systems like a file system – storing important files by how they were sent. They might not remember where they stored a file but they do remember they sent that file to a particular person. Unfortunately, the space available to many people in their email system causes them to do the same kinds of things that caused them to delete files. Therefore, many people find themselves looking for an email they swore they wrote but cannot find – because they deleted it for one reason or another.

Then, of course, there is Active Directory. Those that must administer Active Directory are under constant pressure to ensure that only accurate data is represented there. Only authorized users should be in the appropriate groups, and only active users should be represented at all. For security reasons, many environments adopt a "delete first and ask questions later" policy. They'd rather have an inconvenienced employee than have an employee or ex-employee given access to the wrong data. This policy unfortunately creates the need to restore Active Directory as well. But you can't just restore all of Active Directory; you need to be able to restore select portions that have been accidentally changed or deleted.

Finally, the other thing that tends to happen in an operational environment is accidental deletion of an entire VM. This didn't happen so much in the physical server days, as "deleting" a physical server required more than a mouse click. However, the pressures of resource management cause virtualization administrators to have the same reasons to delete VMs as everyone else does to delete data. Of course, sometimes they delete a VM that they should not have deleted and it needs to be recovered.

## Operational Recovery Requires Object Level Recovery

Given that these are all the reasons a backup system might need to perform a recovery, you would think backup systems are designed to make these quicker; however, this is often not the case. For example, consider emails that are stored inside Exchange that are stored inside a VM. Many backup systems would require that you first perform a full recovery of the VM, followed by a recovery of Exchange, followed by recovering an individual email from Exchange. That's three steps to get a single email, and that's too many.

Backup systems should allow for instant recovery of files, emails, and Active Directory objects no matter how they are backed up. They also should allow for instant boot of a VM from backups, as that can allow you to quickly recover from one of the worst things that can happen operationally – a deleted VM. Backup systems that make these types of recoveries very fast honor the primary purpose of a backup system – operational recovery.

Disaster recovery is also an important component of an entire data protection system, but they also rarely happen. What happens almost daily at many companies is something that requires an operational recovery, and backup systems should make those happen as quickly as possible before they begin to enhance other parts of their system.

# Chapter 7: How To Design a Hyper-V Disaster Recovery Plan

*Written by George Crump*

Server failure, storage system failure and data center failure are all forms of disaster that will impact the Hyper-V environment. Now, IT planners should add ransomware to that list. How does the Hyper-V Administrator design a disaster recovery plan?

## Disaster Recovery Basics

Most IT professionals think of natural disasters as the primary threat. But events over the recent weeks are proving cyber-attacks may be the greater concern since they can hit any organization, anywhere and at any time. Add to that, the ever present danger of a server failure or storage system failure and it’s easy to see IT has its work cut out.

The first step in any recovery effort is to make sure backup captures a secure copy of data frequently enough so applications or workloads can return to operation quickly to meet the organization's expectations.

The second step is to make sure that protected data is available in a remote location. For most organizations, preparation of the remote site is done by replicating the backup copy to another data center owned by the organization, a managed service provider or the public cloud. IT should not consider the backup process complete until the remote location has a copy of the data.

The final step is test. The purpose of testing is twofold. First, a test is the ultimate verification that the data at the remote site is valid. Second, a test gives IT the experience it needs to execute the disaster plan flawlessly in the event of an actual disaster.

## Recovery At Remote Site

If the organization decides to count on its own remote site for disaster recovery then it needs to first make sure the site is far enough away from the primary site so the same natural disaster will not impact it.

Second, the organization needs to have a hardware acquisition plan. Part of this will likely include some standby servers that are on-premises and waiting for the recovery effort to begin. There, more than likely, will also be applications that can wait for servers to be ordered in prior to being recovered. IT needs to make sure it knows the recovery times for each server and it needs to resist the temptation to provide a high level of recovery service to all.

One of the values of Hyper-V is that multiple applications can run on a single server but the organization needs to be careful not to recover so many applications on the physical server that performance during the disaster is unacceptable.

### Recovery In The Cloud

The other option is to use a backup application that can replicate data to the cloud. For Hyper-V, Azure is an ideal destination. But Amazon AWS will also work. IT planners need to understand what the data protection application's cloud capabilities are. Does the data need to be restored out of the backup application's format into the cloud's data store? And does the virtual machines need to be converted into something that the cloud's hypervisor can work with.

### Reducing DR RTO/RPO

IT is constantly under pressure to meet increasingly strict recovery time and recovery point objectives. A recovery time objective is essentially the time it takes to enable users to login to an application. RTO windows are narrowed by pre-position data, so less has to be transferred at the point of a disaster or by enabling the backup storage target to also act temporarily as a primary storage device also known as boot from backup.

RPO windows are narrowed by capturing copies of changed data more frequently. More frequent copies require thinner backups so the application is interrupted for a shorter period of time and less data has to transfer over the network. Technologies like source side deduplication, change block backups and replication enable thin and more frequent backups.

Disaster Recovery for Microsoft Hyper-V requires a backup and replication product that is Hyper-V aware so it can properly interface with the hypervisor and perform frequent backups as well as replicate those backups to an off-site location, either owned by the organization or a cloud provider. The final step is testing to make sure that IT understands DR from an execution standpoint.

# Chapter 8: Ransomware Protection Requires Different Protocols

*Written by George Crump*

The core server component of most modern backup applications run on Windows. Unfortunately, Windows is also the primary target of ransomware attacks. While there are incidents of attacks on Linux or Macs, ransomware creators are going to go for the largest population possible. That means Windows and backup data is at risk. Obviously, dumping Windows from the data center isn't an option either. A simple change in backup software configuration, assuming the software supports it, should provide the protection the enterprise needs.

## How Does Ransomware Work?

Most organizations are taking great steps to protect themselves from a ransomware attack. They have software to prevent the intrusion and are patching servers to plug vulnerabilities. The problem is that ransomware exploits the organization's biggest weakness – users.

A user can easily defeat even the best security policies by clicking on a suspicious link, opening an email attachment or plugging in a found USB drive. Once the ransomware is inside the organization it can access any file the user account can access, and some ransomware can even promote their credentials. As the malware gets into those files, they are encrypted and can only be accessed via a key, which the ransomware author will sell you; typically through BitCoin. The speed at which ransomware can attack is stunning, corrupting thousands of files every minute.

Ransomware is a new type of a disaster. But ransomware does not attack just user home directories and laptops. Critical applications like MS-SQL create and store data as files, so they are also vulnerable. But most ransomware payments are to unlock typical user file data more so than databases. The reason is most organizations focus on protecting mission critical databases instead of files. But these files contain contracts, proposals, invoices, financial spreadsheet, etc. And in many cases they can't be reproduced and are never physically saved (printed). In short, IT needs to up its game in terms of protecting user files and other types of unstructured data.

## Protecting Against Ransomware

IT should assume ransomware will compromise the organization at some point. To protect against ransomware they should perform backups of data multiple times per day. Since they are occurring more frequently, these backups need to be efficient. Only data changed since the last backup should transfer across the network to the backup storage target. That efficiency enables the backup to complete quickly with less impact on the server being protected and the network over which it is transferred. The less impact the data protection event has, the more frequently it can occur.

Backup data is also at risk, so the organization should make copies of backup data. But the organization also has to be more careful than ever on deciding where to store the copies of the backup. Copying data to another Windows server leaves that copy vulnerable to a ransomware attack.

Instead, the organizations should send the copy to a different protocol. For example, if it copies the backup data to a Linux based NFS share or to a remote Amazon facility via S3 the backup data will be far more secure. At this point, there are no known ransomware attacks that have penetrated both Windows and Linux platforms in a single attack.

Short of disconnecting the backup copy from the network, copying the backup data to two distinctly different platforms (NFS or the Cloud) greatly minimizes the chances the ransomware malware will encrypt of both copies of data.

File data is a primary target of ransomware, yet IT spends most of its time making sure that the protection of application data is covered. And for the first time, ransomware is specifically targeting backup applications to try to make it harder for IT to recover. To protect the organization IT needs to frequently backup all data types and it needs to secure backup data on a different platform than the original copy.

# About Storage Switzerland

Storage Switzerland is an analyst firm focused on the storage, virtualization and cloud marketplaces. Our goal is to educate IT Professionals on the various technologies and techniques available to help their applications scale further, perform better and be better protected. The results of this research can be found in the articles, videos, webinars, product analysis and case studies on our website **storageswiss.com**

**Written by George Crump, Chief Steward**

*George Crump is President and Founder of Storage Switzerland. With over 25 years of experience designing storage solutions for data centers across the US, he has seen the birth of such technologies as RAID, NAS and SAN. Prior to founding Storage Switzerland he was CTO at one the nation's largest storage integrators where he was in charge of technology testing, integration and product selection.*

**Curtis Preston, Lead Analyst**

*W. Curtis Preston (aka Mr. Backup) is an expert in backup & recovery systems; a space he has been working in since 1993. He has written three books on the subject, Backup & Recovery, Using SANs and NAS, and Unix Backup & Recovery. Mr. Preston is a writer and has spoken at hundreds of seminars and conferences around the world. Preston's mission is to arm today's IT managers with truly unbiased information about today's storage industry and its products.*

***Sponsored by NAKIVO***

NAKIVO

***About NAKIVO***

NAKIVO is a US corporation that develops a fast, reliable, and affordable data protection solution for Hyper-V, VMware, and AWS environments. NAKIVO Backup & Replication v7 native Hyper-V backup and replication. VM backups can be easily copied offsite or to AWS/Azure clouds by backup copy jobs. Over 10,000 companies are using NAKIVO Backup & Replication to protect and recover their data more efficiently and cost effectively. Visit www.nakivo.com to learn more.