
# The Seven Wonders of Modern Data Protection

Brien M. Posey

One of the most significant trends in IT over the last several years has undoubtedly been server virtualization. The vast majority of the servers running in today's modern datacenters have been virtualized.

While the server virtualization trend is remarkable in its own right, it has also led to something of a renaissance with regard to data protection. Virtualization technology has made it possible to protect mission-critical servers in ways that were unthinkable just a short time ago. This paper is intended to familiarize you with seven of the most remarkable new data protection capabilities.

# Image-based backup and recovery

Virtual machine (VM) image-based backups have existed for several years now, and a number of backup applications are capable of creating image-based backups. Even so, many organizations still back up VMs in the same way that they back up physical servers. In fact, an August 2012 study by the Enterprise Strategy Group revealed that 46% of organizations still perform agent-based, file-level backups of their VMs.

Image-based backups provide a number of advantages over traditional file-level backups. For example, image-based backups are more efficient than traditional backups. With image-based backups, only changed data blocks need to be backed up, rather than entire files. And should a restore be necessary, image-based backups eliminate the time-consuming, tedious and error-prone task of rebuilding the VM. The result is better recovery point and recovery time objectives (RPOs and RTOs) for the business.

Another advantage of image-based backups is that they capture the VM in its entirety, including whatever operating system and applications happen to be inside. So while traditional file-level backups tend to only support specific operating systems and applications, these limitations go away with image-based backups.

*Note: Another issue with file-level backups is that they often fail to back up the VM configuration, which exists outside the VM in the host or virtual infrastructure management framework (for example, VMware vCenter or Microsoft System Center Virtual Machine Manager). The VM configuration contains important information about the VM, such as its unique identifier. This identifier is used for tracking performance history and setting up the security configuration for the VM on the host, and it's essential that this configuration information be backed up—and restored—with the VM.*

# Agentless backup and recovery

Many of the backup products on the market use agents to back up and restore the contents of virtual servers. While this approach can work, there are a number of disadvantages to using backup agents, including:

- Installing and managing backup agents is a considerable administrative burden
- Virtual servers that are powered off cannot be protected by agents
- Backup agents cannot protect hypervisor-level VM components, such as snapshots, virtual switch configurations and other configuration items
- Agents increase the VM footprint
- Agents increase VM overhead
- There may be security concerns related to using agents
- The presence of agents can make the process of troubleshooting a VM more complex

Hypervisor vendors such as VMware and Microsoft offer data protection APIs that vendors can use to build modern data protection products that offer comprehensive protection for VMs that don't require backup agents.

# Near-continuous data protection

Historically, Continuous Data Protection (CDP) and near-CDP products have been expensive because of the investment required in server hardware, storage and application-specific software. Today it is possible to reduce costs with near-CDP by replicating the contents of VMs.  This helps because VMs can be replicated to commodity storage and inexpensive servers. Furthermore, replication can be performed in a way that allows replicas to be reverted to a previous recovery point. This improves RPOs and provides protection in the event that a VM becomes corrupted and the corruption is replicated.

# Rapid VM recovery

One of the problems that has long plagued traditional disaster recovery solutions is that recovery operations take time (and a place) to complete. In the event that a full, bare metal recovery is required, the recovery time can take hours or even days to complete, depending on the backup product being used and the size of the virtual server being recovered. This is a big problem because outages mean lost productivity and lost revenue.

Rapid VM recovery works in a way that preserves the integrity of your backup image. When a recovery is performed, the VM is created with read-only links to the backup image. This allows the image to remain in a pristine state because all write operations are redirected to a differencing disk. Once the original VM is brought back online, all of the data from the differencing disk is merged into the recently recovered virtual server and then the user workload is redirected to the recovered VM. Assuming that the necessary hypervisor migration capabilities are in place, this process can be completely seamless.

This type of recovery mechanism can recover a VM very quickly. In recent benchmark tests by Veeam Software, VM recovery of a 16 GB Hyper-V VM completed in a mere seven seconds. Likewise, a test recovery of a 200 GB VMware VM completed in less than two minutes. By way of comparison, a standard VM recovery for the same VM using an image-based backup stored on disk took nearly two and a half hours to complete.

It is worth noting that some products offer features that sound similar, but do not provide the same results. For example, AppAssure offers a "Live Recovery" feature, but Live Recovery does not work with boot volumes. Instead, it only makes data volumes available and only after the VM has been restored.

Likewise, VMware vSphere Data Protection offers Changed Block Tracking (CBT) as a mechanism for expediting a recovery operation. However, CBT only works under certain circumstances. Expedited recovery becomes impossible if the VM has been deleted or the underlying storage LUN has been lost.

CBT-based recovery is a commonly used technique; however, it is known to be problematic in situations in which a virtual server needs to be recovered due to disk corruption. vSphere keeps track of the blocks that have and have not been modified, and CBT-based recovery operations are designed to expedite the recovery process by restoring only the blocks that have been modified. The risk with using this as an exclusive approach is that if disk corruption has occurred, then some of the "unmodified" blocks could have been impacted by the corruption. Because vSphere indicates that these blocks are unmodified, a CBT-based recovery will not attempt to recover these blocks. The end result (in the event that disk corruption has occurred) is that the virtual server still contains corruption, even after the restoration has completed.

# Instant item recovery

Although many different solutions exist for performing granular recovery of files and folders that are stored on a VM, instant recovery of application data has historically proven to be much more elusive. Today's modern data protection solutions make the recovery of granular application data fast and relatively easy.

Because many modern data protection products back up VMs as a VM image, it is possible to "spin up" a backup image as a VM copy in an isolated environment. Once the VM copy is running, it becomes possible to retrieve granular application data, such as individual email messages or rows and tables that were accidentally deleted from an Oracle database.

The nice thing about this approach is that the backed up VM image can be started and data can be extracted without the complexities and time consumption of a traditional restoration and without affecting the production environment.

Some backup vendors have built on this approach by also offering utilities to simplify the data selection and extraction process. Suppose, for instance, that an administrator needs to recover a SharePoint object. Mounting a point-in-time SharePoint VM is only the first step. The administrator still has to know how to find and extract the necessary data. That being the case, some backup vendors have begun offering application-specific interfaces that allow an administrator to easily browse application data and select the items they want to recover.

The modern approach for instant item-level recovery theoretically works for any application, and this is in stark contrast to legacy backup solutions. While some legacy products offer object-level recovery for application data, these solutions are application-specific and work with only a few of the most common applications. Recovery of objects almost always requires application-specific agents (which usually cost extra), and the recovery process tends to be very resource intensive. This can be especially problematic in virtualized environments where multiple workloads share a common set of hardware resources.

# Backup verification

At first the concept of backup verification probably seems undeserving of a spot on a list of the seven wonders of modern data protection. After all, primitive forms of backup verification have existed since at least the 1980s. However, modern backup verification is far different from the legacy verification mechanisms from long ago.

The old methods of verifying a backup tape's readability or comparing a backup's content to the backup source are inadequate. For one thing, these forms of backup verification are time consuming. More important, legacy backup verifications can lead to a false sense of security.

Imagine, for example, that the NTOSKRNL.EXE file has been deleted from a Windows server. The server would continue running normally because the file is only used during the boot process. If an administrator were to back up the

server and then verify the backup, the verification would report a successful backup because everything on the server was backed up. The backup and the corresponding verification process have no way of knowing that a critical system file is missing. The next time that the server is rebooted, the boot process will fail because of the missing file. Restoring the backup will not correct the problem because the system file was deleted before the backup was created. In essence, the backup is bad even though the verification process correctly reported the backup as being good.

Modern data protection products do not compare the backup's contents to the backup source like legacy backup solutions do. Instead, they can automatically verify the recoverability of the backup by creating an isolated lab environment and then starting the VM directly from the backup. In doing so, the backup software verifies that the virtual server's operating system boots correctly and that applications within the VM are also able to start.

Of course many applications have a number of external dependencies. Microsoft Exchange Server, for example, depends on the Active Directory, which in turn depends on domain controllers and DNS servers. Because such dependencies exist in the real world, some modern backup solutions can be configured to group multiple VMs together for testing purposes. By doing so, the software is able to verify not only the integrity of each virtual server backup, but also that the backups were created in a way that respects the underlying dependencies.

The really nice thing about modern backup verification is that it can happen automatically. Administrators do not have to worry about manually testing each backup. The backup software can be scheduled to automatically create a lab environment, test the backup, notify the administrator, and then clean up after itself by deleting the lab environment after the testing process has been completed.

# On-demand sandbox

Modern data protection's most impressive capability could be that of creating an on-demand sandbox. For decades, it has been difficult for administrators to know what impact a software upgrade will have. Administrators often build lab environments that they can use for software upgrade testing or patch management testing.

While this technique sounds promising in theory, it does have a few shortcomings. First, it can be expensive to build a lab environment. Even if the lab is based on the use of virtualization, there are still costs associated with host server hardware.

Another problem is that building a lab environment can be a very tedious and time-consuming process. And even after all of the work that goes into building a lab has been done, the lab environment may still not truly represent the production environment. As such, tests that are run in the lab environment may not have the same outcome they will have in the production environment.

On-demand sandbox features address these problems by allowing administrators to instantly create isolated virtual lab environments. The labs run directly from the backup environment, so there is no additional hardware to purchase (aside from any required host servers) and no storage to provision. Labs are isolated from the production environment and created in a way that maintains integrity  by preventing modification of backups. Best of all, because the labs use your backups, it means that the lab environment is essentially a mirror image of your production environment. This helps to ensure that the testing you perform in a lab environment will be relevant to your production environment as well.

# Conclusion

Server virtualization has allowed data protection products to offer features and capabilities that were unimaginable even a few years ago. Modern data protection products can help organizations to better prepare for disaster through automated testing, while also reducing overall complexity by eliminating the dependency on agents. More important, modern data protection products make it possible to restore VMs without having to wait for hours or days for a traditional restoration to complete.

# About the Author

**Brien Posey** is a freelance technical writer who has received Microsoft's MVP award 9 times for his work with Exchange Server, Windows Server, IIS, and File Systems Storage.

Brien has written or contributed to about three dozen books, and has written well over 4,000 technical articles and white papers for a variety of printed publications and Web sites.

In addition to his writing, Brien routinely speaks at IT conferences and is involved in a wide variety of other technology related projects.

# About Veeam Software

Veeam® is Modern Data Protection™. We believe today's IT requirements have changed and that "3C" legacy backup problems—high costs, increased complexity and missing capabilities—are no longer acceptable for any organization. Veeam provides powerful, easy-to-use and affordable solutions that are Built for Virtualization™ and the cloud—a perfect fit for the modern datacenter.

Veeam Backup & Replication™ is VMware backup, Hyper-V backup, recovery and replication. This #1 VM Backup™ solution helps organizations meet RPOs and RTOs, save time, eliminate risks and dramatically reduce capital and operational costs. Veeam Backup Management Suite™ combines Veeam Backup & Replication and Veeam ONE™ in a single integrated solution to protect virtualization investments, increase administrator productivity and help mitigate daily management risks. Veeam Management Pack™ (MP) extends enterprise monitoring to VMware through Microsoft System Center. Veeam also provides free tools for the virtualization community.

Learn more by visiting http://www.veeam.com.

VEEAM
Modern Data Protection

vmware PARTNER
TECHNOLOGY ALLIANCE

Microsoft Partner
Gold Application Development
Gold Management and Virtualization

# Modern Data Protection
## Built for Virtualization

Powerful | Easy-to-Use | Affordable

## Veeam Backup & Replication

## #1 VM Backup for VMware and Hyper-V

Virtualization changes everything – especially backup. If you've virtualized on **VMware or Hyper-V**, now is the time to move up to the data protection solution Built for Virtualization: **Veeam Backup & Replication**.

Unlike traditional backup that suffers from the **"3C" problem** (missing capabilities, complexity and cost), Veeam is:

- **Powerful:** Restore an entire virtual machine (VM) or an individual file, email or database record in 2 minutes
- **Easy-to-Use:** It just works!
- **Affordable:** No agents to license or maintain, works with your existing storage, and includes deduplication, VM replication, Microsoft Exchange recovery, and more!

Join the 58,000 organizations who have already modernized their data protection with Veeam. **Download Veeam Backup & Replication** today!

BEST OF vmworld 2010 GOLD AWARD NEW TECHNOLOGY

BEST OF vmworld 2011 GOLD AWARD NEW TECHNOLOGY

InfoWorld.com ★2013 AWARDS★ TECHNOLOGY OF THE YEAR

Windows IT Pro 2012 Community Choice GOLD

Windows IT Pro 2012 Community Choice SILVER

WINNER 2011 VIRTUALIZATION REVIEW

To learn more, visit **http://www.veeam.com/backup**