Virtual Machine Backup and Recovery: Five Critical Decisions

Mahmoud Magdy, Thomas Maurer, and Mikko Nykyri

Who should read this paper

This paper is intended for IT Architects, Backup and Recovery Administrators, and virtualization teams at medium size through Enterprise-scale organizations.

Content

Introduction

Because of the outstanding economy, flexibility, and service levels it offers, virtualization is transforming data centers at breakneck speed: by 2016, an estimated 80 percent of the world's x86 servers will be virtual machines (VMs).[1] But the speed of this transformation, along with the high resource utilization, ease of cloning, moving workloads, and other ways virtualization works its magic, raise challenges for "traditional" IT services and the teams that deliver them.

Nowhere is the complexity that virtualization creates for traditional IT services more apparent than in backup and recovery, which participants in a recent Symantec survey ranked among their least-successful IT initiatives.[2]

This paper addresses five critical decisions organizations must make when building a backup and recovery plan to:

- Maintain protection, visibility, and control of applications and data.
- Maximize utilization of established infrastructure, processes, staff, and budget.
- Use virtualization to improve backup and recovery processes.
- Create an efficient, scalable, future-prepared backup and recovery environment.

Each issue is presented first in general terms that apply across IT environments, and then add comments for specific platforms, applications, or industries based on our individual experience as VMware® vExperts and Microsoft® MVPs.

About the authors

Mahmoud Magdy is a Senior Architect with almost 10 years of experience, a Microsoft® Exchange Server MVP (Most Valuable Professional) since 2010, and a Symantec Backup Exec™ BExpert for 2012. He has extensive experience in Microsoft, Symantec, Citrix, Netapp, and EMC technologies, with deep knowledge of Unified Communications, Virtualization, System Center, and Data Center Storage and Blade Systems. Mahmoud tweets under the handle @_busbar.

Thomas Maurer is a Cloud Architect focused on Microsoft technologies, including Microsoft Private Cloud Solutions, System Center, and Virtualization. In 2012, Microsoft named him a Microsoft MVP for his expertise in Microsoft Hyper-V® Virtual Machines. Thomas works closely with Microsoft and its partners to promote Microsoft Technology as speaker for Microsoft at TechDays and other Technical events; his Twitter handle is @ThomasMaurer.

Mikko Nykyri is a Virtualization Product Manager for Backup Exec, with over 15 years of experience working in the IT industry with a focus on storage and networking solutions. He has spent the last four years in Backup Exec Product Management, specializing in data protection for virtual and mixed environments. Mikko is a frequent speaker at VMware User Group conferences, a 2012 VMware vExpert, and an overall technology enthusiast. Mikko's Twitter handle is @BackupMikko.

[1]Philip Dawson and Chris Wolf. Virtualization Key Initiative Overview. (Stamford, CT: Gartner, Inc. July 22, 2011).
[2]Symantec Corporation. Backup and Recovery Flash Poll. (Mountain View, CA: Symantec Corp. January 25, 2012).

Decision number one–Agent-based or agentless?

Agents on physical servers

The "old school" method of backing up physical servers is to install a backup agent (or client) to perform backups from inside the server (note that "backup agent" means software that performs full backup processing and I/O, not agents that only catalog, deduplicate, or provide context for off-host backup solutions). Because physical servers rarely run near full utilization, consumption of CPU cycles by the agents doesn't compromise availability of primary applications. And it's easy to schedule backups so they run off-peak.

Agents on VMs

Installing backup agents on VMs is a different matter. Virtual hosts are managed specifically to operate near their CPU and I/O limits, so adding CPU-intensive backup processes risks the availability of not just a single VM, but every VM on the host. This is true whether backup agents reside continuously on the VM, or are inserted at the start of a backup run and removed on completion.

> Backup agents inside VMs can't anticipate, monitor, or avoid outside competition for host resources.

Attempts to schedule agent-based backups around peak hours must balance the requirements of every business and backup application on the host, including new clones and workloads migrated in from other hosts. Backup agents inside VMs can't avoid resource competition, because they lack awareness of the overall virtual environment and therefore can't respond to higher-priority demands from other workloads. And in dynamic virtual environments, managing separate backup jobs for every workload is a management nightmare, and per-agent licensing costs can spiral quickly out of control.

Agentless solutions

Agentless backup applications take snapshot images of hosts and perform backup operations on an external server or appliance. This approach avoids the host CPU burdens, application availability risks, and management complexity of agent-based backups on each VM. It also reduces network and storage I/O because it uses off-host processing for deduplication, encryption, and other backup-related tasks. Virtualization-aware backup applications communicate directly with virtualization platforms through Application Programming Interfaces (APIs) to use their block optimization, changed-block tracking, and other features—further reducing backup time, CPU cycles, I/O, and storage requirements.

Agentless backups work well "as is" to restore VMs for disaster recovery—files and folders can be restored by mounting images on a new or spare host. But application-level backups involving active Microsoft SQL®, Active Directory®, Exchange, or other databases with internal structure require special handling. Here are three alternatives:

- **Agent + agentless**—Use agentless backups for VMs and files, but onboard backup agents for specialized data sets. Both types of backup run fast, but the need to mount VMs to install recovered data delays recovery.
- **"Context" extensions**—Optional extensions such as Volume Shadow Copy Service (VSS) or VM tools give backup servers or appliances an application-aware, quiescent image of the VM.
- **"Agent-assisted" backups**—In a similar way, lightweight agents (not full backup agents) bundled with agentless backup solutions enumerate and map applications and databases within the VM and send it to the external backup solution, locating items inside the backup image for granular recovery from a single-pass backup.

Special considerations:

Mikko—Specialized operations such as sub-VM backup and recovery, Physical Compatibility Mode Raw Device Mapping (RDM), VM backup, and source deduplication can be performed using agent-based backups.

Mahmoud—Most enterprise networks can handle the increased traffic from agentless backups. But firewalls or IPS protection between workloads and backup servers or appliances—common at service providers—can cause congestion.

Decision number two–Purpose built virtual-domain backup tools, or unified data protection?

Mature IT organizations rarely use point solutions or purpose-built backup and recovery tools, preferring single, full-featured solutions that can protect their computing environments from end-to-end through one user interface.

But when infrastructure split into physical and virtual environments, backup and recovery responsibilities often split along the same lines, with virtualization teams showing a preference for "virtual-only" backup solutions. These are often inexpensive and quick to cover the latest hypervisor features, and an appealing option when they meet an organization's standards for scalability, security, and long-term support. But dual solutions add cost and complexity from duplicate license and maintenance fees, separate storage pools—including redundant storage for heavily duplicated system files—and separate administrative teams.

Dual environments also introduce inefficiencies and risks, including:

- Duplicate alerting, troubleshooting, reporting, compliance, and disaster-recovery processes, with delays and errors coordinating policies across domains.
- Discontinuous visibility across physical and virtual backup domains, with the potential for coverage gaps and duplication.
- Inefficient communications such as ticketing processes or weekly meetings that can't keep up with dynamic virtual environments, risking unprotected VMs.

> "Virtual-only" backup tools don't protect physical IT environments—including Active Directory servers and the backup solutions themselves.

Finally, purpose-built virtual machine backup solutions lack capabilities that experienced backup and recovery teams expect. First, of course, is their inability to protect physical or mixed environments. Most organizations will maintain mixed environments for many years, and even organizations aiming for 100 percent virtualization should maintain at least one physical Active Directory server to avoid time synchronization and other issues.[3] Second, organizations must make sure that virtual-only solutions offer enterprise-class deduplication and automation capabilities, and sufficient scalability to protect large numbers of production VMs. Third, few virtual-only solutions can make use of priority-based tiered storage or tape off-site backup, raising storage costs. And last, all-virtual solutions often lack the ability to protect themselves, requiring reinstallation and configuration on a physical server in case of a disaster.

Mature, unified data protection solutions operate smoothly across physical/virtual boundaries and support all major operating and hypervisor platforms. They provide end-to-end visibility and control over all of IT with unified monitoring and alerts, unified management of backup and recovery, improved economy, coordination, and efficiency, and a single team, solution, toolkit, and management interface.

Special considerations:

Mahmoud—Virtual-only solutions may be appropriate when a development team or single department is responsible for its own backups. But teams responsible for organization-wide backup and recovery should use solutions that give them a single point of visibility and control across the entire IT environment.

Thomas—Backup solutions make and move copies of legally-protected information like customer records. Before committing to any solution, be sure it meets data-protection standards in every jurisdiction in which your organization operates and in which the data may be physically stored.

[3]See, for example, Brad Bird. Active Directory virtualization best practices (blog post). (San Francisco: CBS Interactive, Inc./TechRepublic. January 6, 2010). http://www.techrepublic.com/blog/networking/active-directory-virtualization-best-practices/2433.

Decision number three–Managing deduplication: why, where, and how?

Much of the growth in storage volume that accompanies virtualization consists of duplicate files, especially system files, maintained in multiple VM, operating system (OS), and application instances, and across multiple data centers, physical and virtual domains, and office locations. Backup and recovery solutions use deduplication to dramatically reduce storage requirements, server network and storage I/O, and network congestion on-premises and across wide-area networks to and from remote offices and disaster-recovery sites.

Decisions about where to deduplicate depend on multiple factors:

- **At the client**—This solution minimizes host I/O and network traffic, since only deduplicated data leaves the client. But it's suboptimal when deduplication contends with critical applications for CPU resources.
- **At a local backup server**—This intermediate solution minimizes loads on remote-office servers and wide area network (WAN) traffic.

> Deduplication tactics will differ according to workload requirements, but the key question is not whether to deduplicate, but where.

- **At a central backup server**—Deduplication at a global external server simplifies global deduplication and backup from multiple locations or across multiple storage pools, minimizing storage requirements.
- **At a purpose-built backup appliance**—All-in-one backup appliances can be used as central or local backup servers when speed of deployment and ease of use are key decision criteria. Combining the backup server and software, deduplication, and storage, they are optimized for and directly integrate with virtual environments, and offer very high performance.

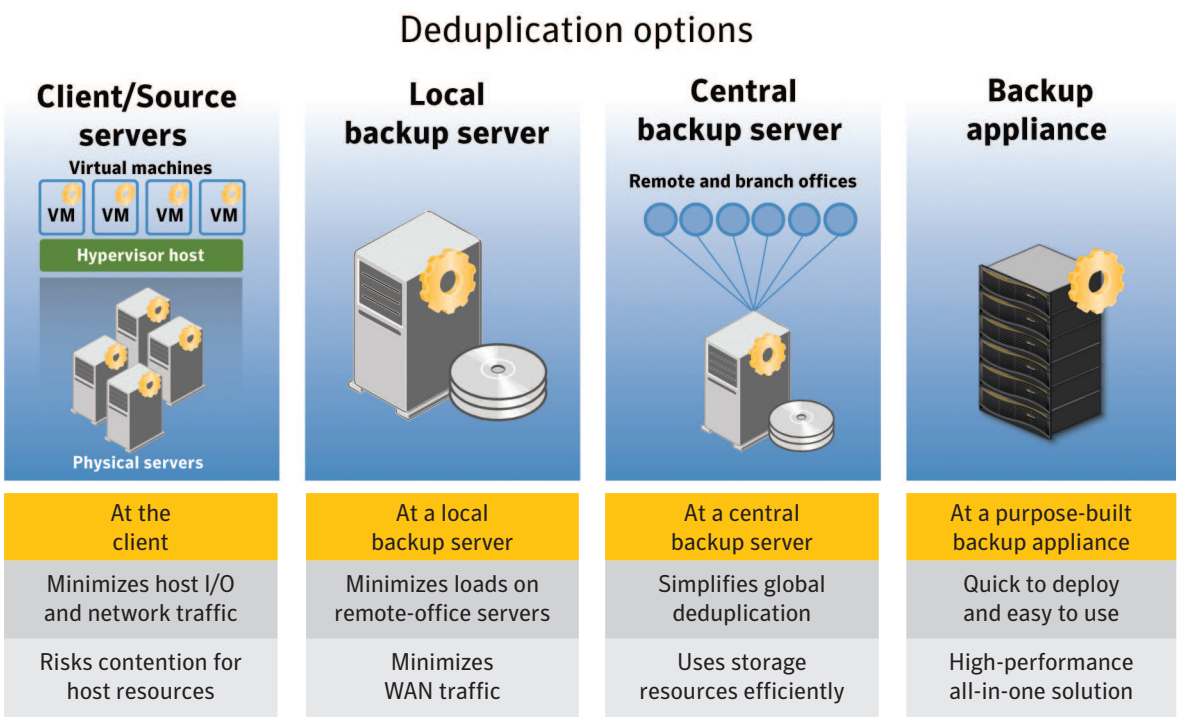

Figure 1. The choice of where to deduplicate depends on many factors, including workload complexity and priority, server capacity and utilization, local area network (LAN) and WAN bandwidth, availability of backup and recovery expertise, local data-privacy regulations, and many more. Look for a backup and recovery solution that supports all the options you are likely to need.

These deduplication alternatives are components of mature, full-featured backup and recovery solutions, not separate solutions that need to be strung together. Most enterprise environments will use multiple deduplication tactics depending on workload requirements. The key point is that it's rarely a question of "whether" to deduplicate, but where deduplication is most effectively accomplished. Using one or more of the solutions above, backup and recovery teams can address complex or changing requirements with deduplication processes that deliver data protection, economy, and performance optimized to meet all IT and business requirements.

A few exceptions:

Deduplication saves so much time, storage, and network bandwidth for so little cost in processor overhead that it should be the default option. But there are some cases in which it adds little value:

Mikko—In backup to tape, where speed and simplicity of recovery are paramount and storage costs are low, deduplication is not preferred, since it would delay the recovery processes.

Mahmoud—In complex, fast-changing databases that are backed up incrementally and often—Exchange and SharePoint are typical examples—deduplication has little impact, and complicates recovery.

Decision number four–Managing the recovery trade-off

Granular recovery from a single-pass backup

Backup exists solely for the purpose of recovery, so recovery time, resources, and risks are the core of any backup and recovery strategy. Recovery strategies need to balance:

- Business-critical but infrequent recovery of lost servers, applications, or entire data centers.
- Lower-impact but very frequent recovery of individual files, folders, and application objects that make up the majority of restore requests at many organizations.

Image-based disaster-recovery backups are simple and fast, but the need to process backups at restore-time complicates recovery of granular items like files or emails. Agent-based file and folder backups address routine restore requests, but delay disaster recovery as new VMs are provisioned, deployed, and powered up. Backing up both ways meets both sets of requirements, but doubles the time, I/O, storage, and management effort consumed by backup processes.

An alternative is to use solutions that index and catalog application information during an "agent-assisted" backup process. This approach leaves the full images available for fast disaster recovery, and locates files, folders, and application objects within the images for routine restores. Single-pass backup with cataloging saves almost half the time needed for a two-pass backup, and still allows granular recovery directly from tape or disk, with no restaging, down to the individual file or application-object level. Cataloged backups also simplify meeting specialized recovery use cases, and the unique requirements of such key stakeholders as Help Desk staff, VM administrators, and application owners.

Recovering multihypervisor environments

Even if your virtual environment relies on multiple hypervisor platforms, be sure your disaster recovery processes are unified and consistent, so recovery steps prioritize business value, not technical convenience. The "bare-metal restore" capabilities included with many backup solutions can accelerate recovery in surprising ways, for example by recovering the backup of a physical machine to a virtual environment, virtualizing older applications and data sets formerly run on physical hardware, or recovering virtual machines from the VMware hypervisor platform into Hyper-V or the reverse.

> Make sure your recovery processes prioritize business value, not just speed or technical convenience.

New options for testing

Unless it is thoroughly tested and proven, a disaster recovery plan is little more than a disaster recovery hope. Yet the time and resources needed for comprehensive tests of recovery processes keep many organizations from conducting them as often as they should. Virtualization lets backup and recovery teams recover systems, applications, and data to a virtual environment for an end-to-end test of their disaster recovery plan, with minimal impact on operations. Testing is a critical component of any sound disaster recovery process, if you don't want your first disaster recovery attempt to be in the aftermath of a real disaster hits.

Special considerations:

Thomas—Consider a tiered backup and recovery strategy that prioritizes recovery-time and recovery-point objectives, data change rates, lifecycles, and storage costs as well as retention time. Business customers like to say, "Everything is Tier 1," but budgets mandate a more conservative approach.

Mahmoud—Test carefully to be sure full applications run properly after full-image restore—internal organization and object "watermarks" may compromise image-based backups of Exchange, Active Directory, and other key applications.

Mikko—Be sure to test not only basic file and folder restore capabilities but complex restores of application objects like Active Directory user restores or Exchange email restores. This is where many products have hidden complexity and performance issues.

Decision number five–Performance, reliability, and manageability at "future scale"

Growth in data volume and the business role of IT will continue to be massive, global, and relentless—and today's small and medium businesses will soon face enterprise-level requirements for security, availability, compliance, and data protection. Backup and recovery planning must anticipate these requirements, and choose solutions that can scale to meet them.

Technically, this is the domain of optimization and tuning—crafting backup and recovery processes adapted to the priorities of one organization's business goals and IT infrastructure. There are no "one size fits all" solutions to these challenges—but here are some useful avenues to explore:

- **Multiple backup streams**—Backup solutions and applications capable of configuring multiple parallel independent backup streams run faster, spend less time waiting for I/O, and reduce backup windows.
- **Load balancing**—Another way to avoid waiting for applications to release I/O is to activate the load-balancing features of advanced backup solutions across multiple hosts. Especially when combined with multiple backup streams, load balancing maximizes the time backup solutions actually spend on backup, and decreases the likelihood of failed or missing backups.
- **Integrated accelerators**—Backup solutions that know the structure and context of backup stream components can optimize their performance instead of relying on worst-case assumptions. Look for accelerators and stream handlers that use information from the hypervisor platform—about changed blocks, for example—to boost performance.
- **Deduplication, again**—While it has been already discussed, importance for deduplication can't be overemphasized. Lower data volume means faster backups, faster recoveries, and reduced storage requirements. Use the latest deduplication technologies, and have configured them to balance performance and storage consumption everywhere possible.
- **Optimized storage**—Storage I/O constraints can create bottlenecks for backups, but not every backup merits Tier 1 storage. A multitiered storage solution can help you balance performance, capacity, and cost—and plan for growth. The best practice is to map retention periods onto the cost-effectiveness of storage media, for example using tape for records that must be retained for years.

Maintaining performance, reliability, and manageability at future scale demands technically sophisticated backup and recovery. But just as important, it requires a backup and recovery solution provider with strength, global reach, commitment to research and development, and proven relationships with virtualization solution providers.

Future-proofing your backups:

Thomas—When building private-cloud solutions that let customers order VMs through a portal, make sure backups, along with SLA options and chargebacks, are included—and that everything is automated using a framework like Windows® PowerShell.

Mahmoud—As your backup sets grow and you optimize backup operations for performance, you will eventually start to encounter network constraints. Anticipate and avoid them by isolating the network used for backups, and making it very fast.

Mikko—Solutions that use VMware vStorage™ APIs offer different tuning levels. Some solutions even let you modify the size of transfer buffers, which can as much as double backup performance in some environments.

Why Symantec?

Symantec offers two complete backup and recovery solutions, optimized for organizations with different requirements. Symantec NetBackup™ unifies backup, deduplication, replication, snapshots, and appliances for global enterprise organizations running heterogeneous infrastructure. Symantec Backup Exec™ provides a simple, easy-to-use interface for the Windows-centric small to medium businesses.

Both solutions protect physical, Hyper-V, and VMware environments, integrating at the API level across both hypervisors and associated management and storage solutions. NetBackup and Backup Exec share a long history of development, years of proven performance in the world's most demanding IT environments, compatibility with a broad range of storage solutions, market leadership, and from technology partners, industry analysts, and publications, including winning the Best of VMworld (2007, 2008, 2010, and 2012) and the Best of Microsoft Tech Ed (2012).

Other Symantec solutions for virtual environments include Symantec™ ApplicationHA,to maintain High Availability of mission-critical applications that span physical and virtual environments, and multiple operating platforms, and Veritas™ Dynamic Multi-Pathing by Symantec improves storage I/O performance and availability across heterogeneous server and storage platforms.

To speak with a Product Specialist in the U.S.

Call toll-free 1 (800) 745 6054

To speak with a Product Specialist outside the U.S.

For specific country offices and contact numbers, please visit our website.

More Information

Visit our website: **http://enterprise.symantec.com**

About Symantec

Symantec protects the world's information and is a global leader in security, backup, and availability solutions. Our innovative products and services protect people and information in any environment—from the smallest mobile device, to the enterprise data center, to cloud-based systems. Our world renowned expertise in protecting data, identities, and interactions gives our customers confidence in a connected world. More information is available at **www.symantec.com** or by connecting with Symantec at **http://go.symantec.com/socialmedia**.

02/13 21283144