

Virtualization Management Best Practices: Critical Aspects of Virtual System Backup and Recovery

A White Paper Sponsored by Veeam Software
May 2008



Table of Contents

Executive Summary	1
Key Requirements for Backup and Recovery	1
Specific Issues in Virtual System Backup	2
Virtual Is Not Physical.....	2
VMware is Not Enough.....	3
Best Practices in Virtual System Backup and Recovery.....	4
Data Aware Backup/Recovery.....	4
Policy-based Backup.....	4
Storage Considerations	4
Choice of Backup Type.....	5
Choice of Recovery Depth.....	5
Using External Data Storage.....	5
VMotion Aware Backup.....	5
Integration Capabilities.....	5
Comprehensive Backup.....	6
Backup Security.....	6
Virtualization Awareness	6
EMA Perspective.....	7
About Veeam Software	7

Executive Summary

EMA research shows that backup and recovery is a mature and mission-critical discipline, with significant and measurable benefits – including improved availability, faster mean time to repair (MTTR), reduced downtime and staff costs, and improved data security and compliance.

Physical and virtual backup and recovery share core requirements, such as ease of use, data assurance, automation, and data granularity. However, backup in a virtual environment introduces new and unique challenges.

Legacy physical backup mechanisms are not sufficient – they cannot generally cope with the highly dynamic nature of virtual systems, cannot backup the virtual hosts themselves, can cause data corruption, and add unnecessarily to the workload of virtual systems. Similarly, native virtual machine management mechanisms are not sufficient – they can increase restore times, do not provide granular restore, do not manage data based on business value, affect virtual system performance, and cannot back up all the required data.

Best practices in virtual system backup and recovery require backup mechanisms designed for virtual environments, that:

- are aware of the business value of granular data
- use policies to translate that value to backup and restore processes
- use a variety of storage, including tape and VTL, as well as SAN, NAS etc.
- provide a choice of backup types
- ensure comprehensive backup of data, configurations, and metadata
- enable a choice of recovery depths
- are aware of key virtual system capabilities such as VMotion and DRS
- provide security capabilities including authentication and encryption

This ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) White Paper explores these issues in more depth, and recommends a number of best practices for virtual system backup and recovery.

Sites with automated backup and recovery average almost a 30% faster mean time to repair (MTTR), and have almost 80 hours of additional uptime per year.

Key Requirements for Backup and Recovery

EMA research into Data Center Automation (DCA) from October 2007 highlights the importance of automated backup and recovery. This research shows that sites with automated backup and recovery average almost a 30% faster mean time to repair (MTTR), and have higher availability that results in almost 80 hours of additional uptime per year in a 24x7 operation. Staffing levels, on average, are 5% lower compared to sites without backup and recovery, resulting in direct salary cost savings. In addition, it reduces costs from downtime – which EMA estimates at an average of \$1,097 per minute, and as much as \$90,000 per minute, depending on the type, location, and impact of any data or system loss.

Effective backup and recovery needs to be easy maintain, to reduce operational expenditures. Purchase costs for software and hardware are typically lower than ongoing operational expenditures (especially staff costs). Even storage devices comprise a flat, one-time cost that is amortizable over time. However, operational costs for people, time, etc., are substantial, incremental, and perpetual. As a rule of thumb, EMA estimates data center operational expenditures to be around 6-8 times capital expenditures, and with average data center staff costs of \$65,000 or more, it is easy to see how this metric plays out.

Of course, in many senses backup is not hard; the real issue is recovery. Backup integrity is therefore another key requirement. Having a corrupt backup is possibly worse than having no backup at all, as it engenders a false sense of security. Mission-critical data needs more than just occasional snapshots; it needs rapid and reliable recovery facilities. This is not the same as data archiving, either. Where archives are maintained (e.g. off-site), they are mostly difficult to use for rapid recovery. Mission-critical systems need recoverability of data not just for disaster and business continuity, but also for accidental deletion and spot recovery.

Backup and recovery systems also need to be automated, ideally requiring as little human intervention as possible. EMA research consistently shows that automation drives substantial cost reduction, reduced error rates, better availability, and higher compliance. Automation also allows IT to establish rule-based policies, and capture sophisticated processes, allowing junior people to do more, and allowing senior people to be more strategic.

Backup and recovery should provide the greatest level of granularity possible – at least file level, if not block level. After all, in case of data loss it is important to restore as little data as necessary – just one or two files, instead of an entire drive – to reduce unnecessary impact on other systems and users. Image-level snapshots do have important use cases – e.g. a complete restore following a malware attack – but deeper (e.g. file-level) granularity is critical for other use cases (e.g. single file corruption in a multi-user application).

An ultimate goal – which is difficult to achieve – is to implement continuous data protection (CDP). This provides not just backups at specific points in time, but continually backs up data as it changes, so that there is a recovery point for every single data change point. This is an ideal scenario, but is not universally available, and requires a significant change in attitude, technology, and cost.

EMA research consistently shows that automation drives substantial cost reduction, reduced error rates, better availability, and higher compliance.

Specific Issues in Virtual System Backup

Virtual Is Not Physical

Investigating backup and recovery in a virtual environment builds on these requirements, but it is very important to realize that virtual system backup is fundamentally different from physical system backup. It is simply not enough to run legacy backup agents in a virtual environment. For example:

- Even though application data may be virtualized – such as on a SAN or NAS – backing up system files from a physical device with a legacy backup process relies on the location of that system. However, virtual systems are highly dy-

dynamic, and move from one physical location to another. Backup systems need to accommodate specific systems, regardless of their locations.

- Legacy backup agents installed in the virtual management platform (e.g. the VMware ESX service console) will not always be available (such as when a guest is moved from one ESX server to another) and with limited access and integration with other management tools, are difficult to manage and to secure.
- Agent-based backup in a virtual environment adds to the management stack, and increases requirements for storage and CPU cycles in executing Virtual Machines, potentially impacting service levels.
- Legacy agents running in the virtual machines themselves simply cannot back up a VMware ESX service console or configuration data. They are, in fact, not even aware that the ESX host exists.
- Using legacy agents, whether in the service console or externally, to back up active virtual machines risks corrupting open and cached files, damaging the integrity of the backup.
- Legacy agents running on top of a guest OS are unable to perform granular backup of migrated, offline, or paused images, as they cannot see at file level of a dormant virtual hard drive. All they can do is take a full disk backup – not granular enough for small-scale, low-impact data recovery.

VMware is Not Enough

Similarly, unsophisticated backup/recovery methods that are common in simple virtual environments (especially VM snapshot, golden image recovery, or build-from-scratch) are not good enough for mission-critical data. For example:

Unsophisticated backup and recovery methods that are common in simple virtual environments are not good enough for mission-critical data

- Golden images are by definition a very static baseline, with little if any application-specific customization. Active systems and applications, however, usually need some changes that are not part of the golden image. Using golden images as backups then adds to the recovery time (and jeopardizes its success) as changes need to be manually reapplied after the image is restored.
 - Full image snapshots are important, but have no granularity, so the resulting recovery process is limited to either a full image restore, or no restore at all. Both of these are unacceptable except when a full disk restore is required in case of catastrophic or global failure.
- Backup processes must be aware of the potential impact they have on the systems they are backing up. Simply scheduling VM snapshots, however, provides none of this visibility, and can drag production machines to their knees as they steal cycles from production in order to provide backup.
 - Similarly, backup processes need to avoid impacting the virtualization management layer – especially where processes like live migration (VMotion) and high availability are being applied in mission-critical production environments. To avoid impacting the response time of these processes, backup systems should avoid or at least minimize demands on the VirtualCenter or ESX server itself as

much as possible.

- Having a virtual host system back up itself is simply not logical, or effective. A system cannot restore itself when it fails, and it cannot even back up critical files that are in use. A virtual environment needs an external backup system to overcome these issues.

Best Practices in Virtual System Backup and Recovery

Organizations looking for best practices in virtual system backup and recovery need to first look at best practices in the physical arena – then modify, adjust, and expand these based on the unique needs of virtual environments.

Key practices that EMA recommends for virtual system backup and recovery include:

Data Aware Backup/Recovery

Organizations need to understand their recovery point objectives – the specific goals of recovery in terms of timeliness, cost, granularity, and importance of data – and bake them into service level objectives. Then, establish a backup regime around these objectives. For example, if a system has personal MP3s or files for a fantasy football pool, there may be no need to waste cycles restoring it all, whereas financial analysis spreadsheets or product research data needs immediate, file-level restore. Once an organization understands the value of data, it can make appropriate decisions on backup methodologies and policies.

Organizations need to understand their recovery point objectives – the specific goals of recovery in terms of timeliness, cost, granularity, and importance of data – and bake them into service level objectives.

Policy-based Backup

Backup and recovery should be policy-based, to achieve business objectives. For example, policy directives that understand the value of data enable that data to be backed up (and recovered) in a cost-effective way that caters to the importance of that data to the business. If backup and recovery is to be automated – which it absolutely should be – then it simply must be policy-based, otherwise automation systems cannot make backup and recovery decisions based on business priorities. Policies should then define the value of data, map that value to backup types (e.g. incremental, full, synthetic), granularity (image-level, file-level), and schedules (e.g. continuous, hourly, weekly) as appropriate to the business importance of the data.

Storage Considerations

Organizations should look at tiered storage for backups – such as fast disk, slow disk, virtual tape libraries (VTL), physical tape, etc. – based on the value of specific data. A best practice is to establish policy-based (i.e. not arbitrary) use of tiered backup services (e.g. with various levels of backup type, granularity, or frequency) and tiered backup hardware (with various levels of device cost) based on business priorities for cost and data recovery. Note that this is not the same as hierarchical storage management (HSM). HSM is more about just reducing cost, and less about assuring the value of data. As a result, HSM eventually puts everything on tape because it is cheaper, which does not allow for faster recovery of high-value data.

Choice of Backup Type

It is important to balance the time and resources for backup against the desired recovery time, according to the value of the data. Organizations should therefore choose their backup level based on the business needs for recovery. A full backup takes a copy of all data on the system. This is a time- and resource-intensive process, but provides a point in time from which all files can be rapidly restored. An incremental backup, by contrast, only takes copies of data that has changed. This is much faster and uses fewer resources, but recovering data can require it to be rebuilt from multiple backups, which takes much longer than restoring from a recent full backup. Synthetic backup dynamically reconstructs a full backup by adding each new incremental backup to the previous incremental backup as it is taken. This allows data to be recovered without being rebuilt, yet does not use all the time and resources of a full backup.

Choice of Recovery Depth

Regardless of which backup method is used, it must provide recovery to multiple points in time, so it must provide depth of backup generations, and the ability to identify and recover files from a specific point in time – regardless of whether that point in time is contained in a full, incremental, or synthetic backup. Similarly, it must be able to process backups sequentially and apply incremental rollback to restore an entire system to a known good state (e.g. in case of complete system failure, disk corruption, software failure, or virus).

Using External Data Storage

It is important to use external storage (SAN, NAS etc.) as much possible for active storage of production data (as opposed to system files) so it can be backed up (e.g. to VTL, then to tape) at any time without impacting production cycles. Backup files themselves should also be stored off-host, to protect against critical disk-failure on the virtual host itself. Storing backups on the same disk as the files that are being backed up is a recipe for downtime.

VMotion Aware Backup

Especially in a VMotion, High Availability, or Distributed Resource Scheduler environment, backup storage should be accessible from any possible restore candidate for the virtual system – not just the physical systems that happened to be hosting the image when it was backed up. Restore mechanisms must therefore understand the location (and resulting policy-based variations) of virtual systems, not just physical servers. Using VMware Consolidated Backup (VCB) will help to ensure awareness of the VMware DRS and VMware HA features in VMware Infrastructure, to allow restore onto a different system.

Integration Capabilities

Backup systems should have visibility and integration across both physical and virtual systems. Physical backup that is not aware of the contents of virtual machines can damage active operational instances, e.g. by locking and copying active files. Virtual backup that is not aware of the physical environment can duplicate data, e.g. by backing up an entire virtual environment whether individual VMs are active or dor-

Backup systems should have visibility and integration across both physical and virtual systems.

mant. Ideally, virtual backup should also integrate with other management technologies – e.g. performance and activity monitoring, workload automation – so that backup and recovery activity can be aware of, and cater to, the activity in the broader environment. Virtual backup in a VMware environment should specifically be compatible with the VMware compatibility matrix.

Comprehensive Backup

In most environments, backup and recovery systems need to support multiple file systems – e.g. NTFS, FAT, VFS, Linux FS, etc. They also need to support not just file or data backup, but also backup for metadata, including virtual host and guest properties, file properties, security settings, sharing settings, etc. Without this additional metadata, recovered files may be unreadable, no longer shared, or even worse, may be unexpectedly exposed to unauthorized users. In addition, backup should be able to back up not just image and data files, but also host, guest, and network configuration data. Manually recreating such configurations is difficult and error-prone, introduces potential security and compliance risks, and may extend potentially mission-critical downtime.

Backup Security

Security of backups is always critical, including user authorization, content encryption, and physical security of portable backup media (tapes etc.). This is perhaps even more important in a virtual environment than a physical one, because virtual backups tend to contain entire operating environments – guest systems, applications, and data – that can be recovered into any virtual container, in any location, and with ESX security authorization, giving an intruder the highest levels of authority to intrude and extract confidential data. However, some protections can increase management effort and cost – for example, passwords and encryption keys need to be managed and stored independently of the data they protect. Again, the cost of backup practices must be weighed against the value of specific data.

Security of backups is always critical, including user authorization, content encryption, and physical security of portable backup media (tapes etc.).

Virtualization Awareness

Virtual backup systems and processes must consider the impact of backup processing on production applications and end users. On a physical machine, CPU and I/O cycles are often available for non-production tasks like backup. In a virtual machine – especially in a typical use case where multiple workloads are consolidated onto a single physical server to maximize resource utilization – processing cycles are much more precious. Using the built-in off-host proxy capabilities of VCB is one important virtualization-aware option that will reduce resource consumption and impact on running systems, but even that will not be aware of the effect concurrent I/O operations (like backup) can have in saturating network and storage resources. Virtualization awareness also should extend to a choice of traditional backup (i.e. image level in proprietary format, which uses less storage, is compressed and encrypted, etc.) and replication (native ESX format, which is more resource-expensive because it uses ESX server cycles, but is ready to be restored directly to any ESX server). It should also be able to back up virtual systems whether they are cold (shutdown), warm (paused), or active.

EMA Perspective

Like insurance, backup and recovery systems provide a service that you hope you will never need – but without it, you are exposed to the risks of losing critical business systems, applications, and most importantly, data. Business continuity and availability therefore depends on the ability to provide automatic backup and recovery of systems, applications, and data. Every enterprise needs data backup and recovery to guard against catastrophe in case of continuity events – natural disasters like flood, earthquake or fire that disable an entire data center; or small-scale disruptions like security breaches, disk failures, or application errors that disrupt access to or validity of a limited set of data.

Backup and recovery systems provide a service that you hope you will never need – but without it, you are exposed.

Virtual systems need backup and recovery no less than physical systems do. Indeed, with the growing reliance on virtualization to provide high availability, disaster recovery, and business continuity, backup and recovery for virtual systems can be even more important. However, virtualization introduces a range of new layers, and a range of new difficulties for backup and recovery, and as a result need new, purpose-built backup and recovery technologies. Legacy backup is not enough, nor is manual processing, not are bundled backup mechanisms. Specific systems that accommodate the unique requirements of a virtual environment are essential.

However, technology is not the complete answer either. Storage administrators and virtualization administrators must work together to establish their own best practices, using these recommendations as a guide. This will enable them to define a common structure, process, and toolset for automated backup and recovery for virtual systems that enhances, rather than detracts from, the business value of data.

About Veeam Software

Veeam Software, a VMware Technology Alliance Partner best known for its [FastSCP](#) product, provides practical, innovative solutions for managing virtual server environments. Veeam Software is led by Ratmir Timashev and Andrei Baronov, the founders of Aelita Software. Today the company offers products including [Veeam Backup](#), the 2-in-1 backup and replication solution; [Veeam Reporter](#), an award-winning automated way to discover and document VMware virtual server environments; [Veeam Configurator](#), a simple way to configure and provision ESX servers; and [Veeam Monitor](#), for a bird's-eye view of key performance metrics across the entire VMware infrastructure. Learn more about Veeam Software by visiting www.veeam.com.

About Enterprise Management Associates, Inc.

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst and consulting firm dedicated to the IT management market. The firm provides IT vendors and enterprise IT professionals with objective insight into the real-world business value of long-established and emerging technologies, ranging from security, storage and IT Service Management (ITSM) to the Configuration Management Database (CMDB), virtualization and service-oriented architecture (SOA). Even with its rapid growth, EMA has never lost sight of the client, and continues to offer personalized support and convenient access to its analysts. For more information on the firm's extensive library of IT management research, free online IT Management Solutions Center and IT consulting offerings, visit www.enterprisemanagement.com.

This report in whole or in part may not be duplicated, reproduced, stored in a retrieval system or retransmitted without prior written permission of Enterprise Management Associates, Inc. All opinions and estimates herein constitute our judgement as of this date and are subject to change without notice. Product names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

©2008 Enterprise Management Associates, Inc. All Rights Reserved. EMATM, ENTERPRISE MANAGEMENT ASSOCIATES®, and the mobius symbol are registered trademarks or common-law trademarks of Enterprise Management Associates, Inc.

Corporate Headquarters:
5777 Central Avenue, Suite 105
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com

