



Data Protection Options for Virtualized Servers

Demystifying Virtual Server Data Protection – Best Practices and Technologies

By Greg Schulz

Founder and Senior Analyst, the StorageIO Group



Version 2.0 August 20, 2009

This Industry Trends and Perspectives Paper is Compliments of:

Quantum[®]

www.quantum.com

Industry Trends and Technology Perspective White Paper
Data Protection Options for Virtualized Servers

Introduction

There is no such thing as a data or data protection recession, however, organizations of all size need to balance the need to support and protect an increasing amount of data in the same or smaller footprint. This has led to a focus around data center optimization and efficiency, including server virtualization. For example addressing IT data center power, cooling, floor space and environmental (PCFE) issues (commonly referred to as green computing) along with supporting next generation virtualized data center environments in order to sustain business and data growth.

Optimization and efficiency (e.g. Green IT) mean many things from energy avoidance and consolidation, to boosting performance, improving data protection and quality of service while enhancing overall IT productivity in the same or smaller footprint (power, cooling, floor space, management and cost) to enable business survivability and sustainability.

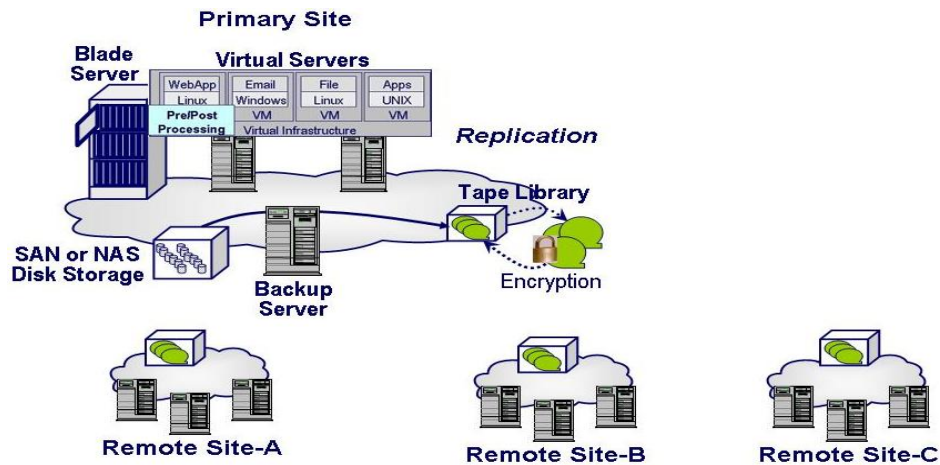


Figure-1: Providing High Availability and Protecting Local and Remote Virtual and Physical Servers

Background and Issues

There are many issues, challenges and options related to protecting data and applications in a virtual server environment (Figure-1). For example, in a non-virtualized server environment, the loss of a physical server would have an impact on the applications running on only that server.

However, in a highly aggregated or consolidated environment, the loss of a physical server supporting many virtual machines (VMs) would have a much more significant impact by affecting all the applications supported by the virtual servers. Another challenge is protecting the growing amount of structured and un-structured data both in primary data centers along with data in remote offices branch offices (ROBO), workgroups, field offices and other locations.

Yet other challenges facing IT organizations is determining what technologies, techniques and best practices, techniques

Industry Trend

There is no such thing as a data recession! While global economic challenges exist including a slowdown in IT spending, the reality is that more data continues to be generated, processed, storage and protected than in any time in history.

Thus, while IT budgets are being stretched to do more in an optimized and efficient manner, the result is more net storage and processing capabilities along with data protection requirements are being squeezed into the same or smaller footprint to meet demand and enable business sustainability.

Industry Trends and Technology Perspective White Paper Data Protection Options for Virtualized Servers

and technologies including tiered storage including disk vs. tape, virtual tape libraries (VTLs), deduplication, snapshots, backup and replication among others to leverage to address changing data protection and infrastructure resource management (IRM) requirements. Consequently, with the adoption of virtual server environments, having a sound data protection strategy is of magnified importance.

The majority of server virtualization currently being undertaken is for the consolidation of heterogeneous operating systems on underutilized servers. Another aspect is to address desktops and workstations, in part for consolidation, but also to simplify management, data protection and associated cost and complexity. StorageIO sees a next wave of server virtualization, which is life beyond consolidation on the horizon which is enabling agility and flexibility.

This next wave, (life beyond consolidation) combines the tenets of the two previous scenarios with an emphasis around utilizing virtualization to for agility and flexibility to enable dynamic management of servers and dynamic data protection.

For example, using virtualization to support redeployment of servers for workload changes and provide transparency. In this scenario, consolidation continues to be a driver, however there is also an emphasis on leveraging virtualization as a tool for applications, servers and storage that do not lend themselves to being consolidated, yet can benefit from business and IT IRM enabled agility including enhanced performance, high availability (HA), disaster recovery (DR) and business continuance (BC).

There are several approaches to achieve server virtualization including Citrix/Xen, Microsoft and VMware, as well as vendor-specific containers or partitions. Many of the data protection issues are consistent across different environments with specific terminology or nomenclature. Given the market adoption of VMware for server consolidation and, consequently, the growing spotlight on data protection issues associated with this particular approach, many of the examples in this paper will be centered on VMware. The same issues and approaches apply to other virtualization technologies including Microsoft Hyper-V and Xen-based.

Virtual server environments often provide tools to facilitate maintenance and basic data protection while lacking tools for complete data protection, BC or DR. Instead, virtual server vendors provide APIs, other tools, or solution/software development kits (SDKs) so that their eco-system partners can develop solutions for virtual and physical environments. For example, solutions from VMware, Citrix and Microsoft include SDKs and APIs to support pre- and post-processing actions for customization and integration with Site Recovery Manager (SRM), VMware Consolidated Backups (VCBs), VMotion or Microsoft Hyper-V Quick Migration.

Glossary of Common Virtual Terms

Agent	OS or VM software for backup
COS	Console Operating System
Dedupe	Eliminate duplicate data
DLP	Data Loss (or leak) Protection
DPM	Data protection management
FCoE	Fibre Channel over Ethernet
Green IT	IT efficiency and Optimization
Guest OS	Guest operating system in a VM
IOV	I/O Virtualization
IRM	Infrastructure Resource Mgmt.
NPIV	N_Port ID Virtualization
OS	Operating System
Proxy	Backup Server
RDM	Raw Device Mapped Storage
RPO	Recovery Point Objective
RTO	Recovery Time Objective
SRM	Storage resource management
SRM	VMware Site recovery manager
Source	Where data comes from
Target	Where data is sent too
VCB	VMware Consolidated Backup
VIO	Virtual I/O
VM	Virtual Machine
VMFS	VMware File System
VTL/VTS	Virtual tape library / system

Industry Trends and Technology Perspective White Paper Data Protection Options for Virtualized Servers

Time to Re-Architect and Upgrade Data Protection

A good time to rethink data protection and archiving strategies of applications and systems data is when server consolidation is undertaken. Instead of simply moving the operating system and associated applications from a “tin” wrapped physical server to a “software” wrapped virtual server, consider how new techniques and technologies can be leveraged to improve performance, availability, and data protection.

For example, an existing server with agent-based backup software installed sends data to a backup server over the LAN for data protection. However, when moved to a virtual server, the backup can be transitioned to a LAN-free and server-free backup model server. In this case LAN and other performance bottlenecks can be avoided.

From a historical data protection perspective, magnetic tape has been a popular, cost-effective, and the preferred data storage medium for retaining data to meet backup and recovery, BC, DR and data preservation or archiving requirements.

Recently, many organizations are leveraging storage virtualization in the form of transparent access of disk-based backup and recovery solutions including VTLs. VTLs continue to evolve to support traditional disk to tape (D2T), disk to disk (D2D), and disk to disk to tape (D2D2T) on a local and wide area or cloud basis leveraging various technologies including compression, policy based deduplication, replication and tiered storage. These storage solutions emulate various tape devices and tape libraries to co-exist with existing installed backup software and procedures using virtual to bridge the future (e.g. disk) to the past (e.g. existing backup or data protection software and tape) for maxim investment protection and ROI.

Magnetic tape remains one of, if not the most, efficient (Green) data storage mediums from a PCFE perspective for inactive and long-term archived data. Disk to disk (D2D) based snapshots; backups and replication have become popular options for near-term and real-time data protection to meet RTO and RPO requirements.

With a continued industry trend towards using D2D for more frequent and timely data protection, tape is finding a renewed role in larger, more infrequent backups for large scale DR. Tape is finding a renewed role supporting long-term archiving and data preservation of project data and compliance data. For example, D2D, combined with compression and deduplication disk-based solutions, is used for local, daily and recurring backups. Meanwhile weekly or monthly full backups are sent to tape to free disk space as well as address PCFE concerns.

Various Technologies and Techniques – Virtual Server Data Protection Options

Just as there are many approaches and technologies to achieve server virtualization, there are many approaches for addressing data protection in a virtualized server environment.

Industry Trend

Efficiency and Optimization extend from improved utilization of hardware and software resources to general Infrastructure Resource Management (IRM). This means that boosting IT efficiency includes optimizing how data protection which falls under the IRM umbrella is accomplished.

By optimizing data protection, more data can be made safe, secure, consistent and accessible to sustain business now and strategically position organizations for growth in the future in an efficient manner.

Data Protection Tip:

If your data is important enough to be backed-up or replicated. Or if you need an archive to preserve data for planned or possible future use, then the data is important enough to make multiple copies - including on different media types - at different locations to meet your applicable requirements.

Industry Trends and Technology Perspective White Paper Data Protection Options for Virtualized Servers

Table-1 provides an overview of data protection capabilities and characteristics to address various aspects of data protection in a virtualized server environment.

Capability	Characteristics	Description and Examples
Virtual Machine Migration	<ul style="list-style-type: none"> • Move active or static VMs • Facilitate load-balancing • Pro-active failover or movement vs. recovery 	<ul style="list-style-type: none"> • Vmotion and Xenmotion among others • May be physical processor architecture dependent • Moves the running VMs memory from server to server • Shared access storage for BC/DR
Failover (HA) High Availability	<ul style="list-style-type: none"> • Proactive VM movement • Automatic failover for HA • Fault containment/isolation • RAID disk storage 	<ul style="list-style-type: none"> • Proactive move of a running VM to a different server • Requires additional tools to insure all data is moved • Low latency network bandwidth need for remote HA • Replication of VM and application specific data
Snapshots	<ul style="list-style-type: none"> • Point-in-time (PIT) copies • Copies of current VM state • May be application aware • Exists in different locations 	<ul style="list-style-type: none"> • Facilitate rapid restart from crash or other incident • Guest OS, VM, appliance or storage system based • Combine with other forms of data protection for BC/DR or accidental file and data deletion or corruption
Backup and Restore	<ul style="list-style-type: none"> • Application based • VM or guest OS based • Console subsystem based • Proxy server based • Backup server or target resides as guest in a VM 	<ul style="list-style-type: none"> • Full image, incremental, different or file level • Operating system and application specific support • Agent or agent-less backup running in different locations • Backup over LAN to backup server and backup device • Backup to local or SAN attached device (disk or tape) • Proxy-based (VCB) for LAN and server free backup
Replication	<ul style="list-style-type: none"> • Application based • VM or guest OS based • Console subsystem based • External appliance based • Storage array based 	<ul style="list-style-type: none"> • Application replication such as Oracle • VM or guest OS or 3rd party software based • Application aware snapshot integration for consistency • Replication software running on an external appliance • Storage system controller based replication
Archiving	<ul style="list-style-type: none"> • Document management • Application based • File system based • Long term preservation 	<ul style="list-style-type: none"> • Structured (database), semi-structured (email) and unstructured (files, attachments, PDFs, images, video) • Compliance or regulatory along with IP and project data preservation for planned or possible future use
Networking	<ul style="list-style-type: none"> • NPIV for Fibre Channel • Bandwidth services • Converged networks • I/O Virtualization • Wide area data services 	<ul style="list-style-type: none"> • Move VMs independent of zoning or physical changes • Remote mirroring, replication and backup optimization • Fibre Channel over Ethernet (CEE and DCE) • Enable network and I/O abstraction transparency • Bandwidth, data, file and application optimization
Data Protection Management	<ul style="list-style-type: none"> • Data protection tools • Analysis and correlation • Backup and replication 	<ul style="list-style-type: none"> • VMware Site Recovery Manager • Data protection advisory and analysis tools • Manage various aspects of IRM and data protection

Table-1: Data Protection Options for Virtual Server Environments

Complete and comprehensive data protection architecture should combine multiple techniques and technologies to meet various RTO and RPO requirements. For example, VM movement or migration tools such as VMware VMotion provide proactive movement for maintenance or other operational functions. These tools can be combined with third party data movers, including replication solutions, to

Industry Trends and Technology Perspective White Paper **Data Protection Options for Virtualized Servers**

enable VM crash restart and recovery or basic availability. Such combinations assume that there are no issues with dissimilar physical hardware architectures in the virtualized environment. It is important to be aware of the motivators and drivers for data protection of a virtual server environment when creating the architecture.

Examples of threat risks drivers to protect against include:

- Accidental or intentional deletion or corruption
- Operating system, application, server or storage failure
- Loss of access to site, servers or storage
- Site, campus, local, metro or regional disaster or event
- Business or regulatory compliance requirements
- Data loss prevention and information privacy concerns
- More data being generated, stored and used remotely

Additional items to consider, including applications and virtual server requirements:

- RTO and RPO requirements per application, VM/guest or physical server
- How much data changes per day along with fine grained application aware data protection
- The performance and application service level objectives per application and VM
- The distance over which the data and applications need to be protected
- The granularity of recovery needed (file, application, VM/guest, server, site)
- Data retention as well as short term and longer term preservation (archive) needs
- Data usage and access patterns or requirements to meet business needs
- Focus on doing more with less, or, doing more with what you have

Another consideration when comparing data protection techniques, technologies and implementations is application aware data protection. Application aware data protection approaches ensure that all data associated with an application, including software, configuration settings, data and current state of the data or transactions, is preserved. To achieve true application aware and comprehensive data protection, all data, including memory resident buffers and caches pertaining to the current state of the application, needs to be written to disk. At a minimum, application aware data protection involves quiescing of file systems and open files data to be written to disk prior to a snapshot, backup or replication operation. Most VM environments provide tools and APIs to integrate with data protection tasks including pre-freeze (pre-processing) and post-thaw (post processing) for application integration and customization.

Additional attributes to consider for various types of data protection techniques include:

- Continuous data protection or fine grained data snapshots and replication
- Application aware ensuring that all data is completely and consistently copied for data integrity
- I/O interfaces (internal dedicated external shared) SAS, iSCSI, Fibre Channel/FCoE SAN or NAS
- Movable (re-locatable) snapshots to facilitate data protection using snapshots to perform backup
- Physical to virtual, virtual to virtual, virtual to physical and related topology issues
- Agent or agent-less backup, location of the agent (in the guest OS or in the console system)
- Data footprint reduction techniques (archive, real-time or off-line compression and deduplication)
- Data security, including logical and physical protection, authorization and encryption
- Tiered storage: tier0 (FLASH SSD), tier1 (fast disks), tier2 (capacity disks), tier3 (tape)

Industry Trends and Technology Perspective White Paper Data Protection Options for Virtualized Servers

Virtual Machine Movement

Often mistaken, or perhaps even positioned as data protection tools and facilities, virtual machine movement or migratory tools are targeted and designed for maintenance and proactive management. The primary focus of tools such as VMware VMotion or LiveMigraiton from Virtual Iron is to be able to proactively move a running or active VM to a different physical server without disruption that has shared access to the storage that supports the VM.

For example, VMotion can be used to maintain availability during planned server maintenance or upgrades or to shift workload to different servers based on expected activity or other events. The caveat with such migration facilities is that, while a running VM can be moved, those VMs still rely on being able to access their virtual and physical data stores.

This means that data files must also be relocated. It is important to consider how a VM movement or migration facility interacts with other data protection tools including snapshots, backup and replication, along with other data movers to enable data protection.

In general, considerations pertaining to live movement facilities for VMs include:

- How does the VM mover support dissimilar hardware architectures (e.g. Intel and AMD)?
- Is the feature a conversion tool (e.g. physical to virtual) or does it perform live movement of VMs?
- Can the migratory or movement tool work on both a local and wide area basis?
- How does the tool interact with other data protection tools to ensure data is moved with the VM?
- What are the ramifications of moving a VM and changes to Fibre Channel zoning and addressing?
- How many concurrent moves or migrations can take place at the same time?
- Is the movement limited to virtual file system based VMs or does it include raw devices?
- What 3rd party data movers via hardware, software or network services are available or required?

High Availability (HA)

Virtual machine environments differ in their specific supported features for HA, ranging from the ability to failover or restart a VM on a different physical server, Other differences include to the ability to move a running VM from one physical server to another physical server (as discussed in the previous section). Other elements of HA for physical and virtual environments include eliminating single points of failure to isolate and contain faults. For example, using multiple network adapters (such as NICs), redundant storage I/O host bus adapters, and clustered servers.

A common approach for HA data accessibility is RAID enabled disk storage to protect against data loss in the event of a disk drive failure. For added data protection, RAID data protection can be complemented with local and remote data mirroring or replication to protect against loss of data access due to a device, storage system or disk drive failure. RAID and mirroring, however, are not a substitute for backup, snapshots or other point-in-time based discrete copy operations that establish a recovery point.

What is Storage VMotion?

VMware Storage VMotion is a component of vSphere that compliments the VM migration tool VMotion. On the surface Storage VMotion sounds like a data protection tool, however, similar to VMotion, it is a maintenance and management or migration tool that can be used for pro-active movement of VM storage as of a point in time.

Storage VMotion by itself is not a data protection tool. Instead it relies on a combination of snapshots, backup and/or data replication capabilities powered by 3rd party hardware, software and services solution Providers.

Industry Trends and Technology Perspective White Paper **Data Protection Options for Virtualized Servers**

RAID provides protection in the event of disk drive failures; RAID does not protect data by itself in the event that an entire storage system is damaged. While replication and mirroring can protect data in the event that a storage system is destroyed or lost at one location, if data is deleted or corrupted at one location that action will be replicated or mirrored to the alternative copy. Consequently, some form of time interval based data protection, such as a snapshot or backup, needs to be combined with RAID and replication for a comprehensive and complete data protection solution.

Snapshots

Point-in-time (pit) copies, commonly known as snapshots, are a popular approach to reducing downtime or disruptions associated with traditional data protection approaches such as backup. Snapshots vary in their implementation and location with some being full copies while others are delta-based. For example an initial full copy is made with deltas or changes recorded, similar to a transaction or redo log, with each snapshot being a new delta or point in time view of the data being protected. Another way snapshot implementations can vary is in where and how the snapshot data is stored on the same storage system or the ability to replicate a snapshot to a separate storage system.

Because snapshots can take place very quickly, an application, operating system or VM can be quiesced (suspended), a quick snapshot taken of the current state at that point in time, and then resume with normal processing. Snapshots work well for reducing downtime as well as speeding up backups. Snapshots reduce the performance impact of traditional backups by only copying changed data, similar to an incremental or differential backup but on a much more granular basis. Snapshots can be made available to other servers in a shared storage environment to further off-load data protection. An example is using a proxy or backup server to mount and read the snapshots to construct an off-line backup.

For virtual environments, snapshots can be taken at the VM or operating system layer with specific feature and functionality varying by vendor implementation. Another location for snapshots to occur is in storage systems that have integration with the guest operating system, applications or VM. Snapshots can also take place in network or fabric based appliances that intercept I/O data streams between servers and storage devices. One of the key points is to make sure that when a snapshot is taken, the data that is captured is the data that was expected to be recorded.

For example, if data is still in memory or buffers, that data may not be flushed to disk files and captured. Thus, with fine grained snapshots, also known as near or coarse continuous data protection (CDP), as well as with real-time fine grained CDP and replication, 100% of the data on disk may be captured. But if a key piece of information is still in memory and not yet written to disk, critical data to ensure and maintain application state coherency and transaction integrity is not preserved. While snapshots enable rapid backup of data as of a point in time (RPO), snapshots do not provide protection by themselves in the event of a storage system failure, thus, snapshots need to be backed up to another device.

Site Recovery Manager (SRM)

Not to be confused with storage resource management which is focused on resource usage reporting, VMware SRM is focused on data protection management including orchestrating or facilitating various aspect of BC and DR operations includes failover.

VMware SRM is a data protection management (DPM) framework tool that leverages 3rd party hardware, software and services to enable a robust and resilient virtualized server and data protection environment.

Industry Trends and Technology Perspective White Paper Data Protection Options for Virtualized Servers

Items to consider about snapshots for virtual environments include:

- How many concurrent snapshots can take place, and how many snapshots can be retained?
- Where is the snapshot performed (guest OS, VM, appliance or storage) and what is captured?
- What API or integration tools exist for application aware snapshots and synchronization?
- Are there facilities for pre- and post-processing functions to wrap around snapshots?
- Do the snapshots apply to virtual disks or physical disks?
- What is the performance impact when snapshots are running?
- How do the snapshots integrate with third party tools including backup or replication?
- What are the licensing and eminence costs for the snapshot software features?

Agent-/client-based Backup

Client or agent-based backup, also known as LAN-based backup, is a common means of backing up physical servers over a LAN. The term agent-based backup comes from the fact that a backup agent (backup software) is installed on a server with the backup data being sent over a LAN to a backup server or to a locally attached tape or disk backup device.

Given the familiarity and established existing procedures for using client-, LAN- and agent-based backup, a first step for data protection in a virtual server environment can be to simply leverage agent-based backup while re-architecting virtual server data protection.

Quantum Technology Example:

Quantum DXi disk-based backup solutions can be attached to physical servers for agent-based backup off-loading LAN data traffic to boost performance. In the theme of optimizing IT resources associated with many server virtualization initiatives, the DXi system can be segmented for use as both an agent-based and proxy backup target using data policy enabled data deduplication to reduce data footprint of data on disk addressing PCFE concerns or requirements.

Physical servers, when running backups, have to stay within prescribed backup-up windows while avoiding performance contention with other applications on that server along with avoiding network LAN traffic contention. In a consolidated virtual server environment, it is likely that multiple competing backup jobs may also vie for the same backup window and server resources including CPU, memory, and I/O and network bandwidth. Care needs to be exercised when consolidating servers into a virtual environment to avoid introducing or aggravating performance conflicts and bottlenecks.

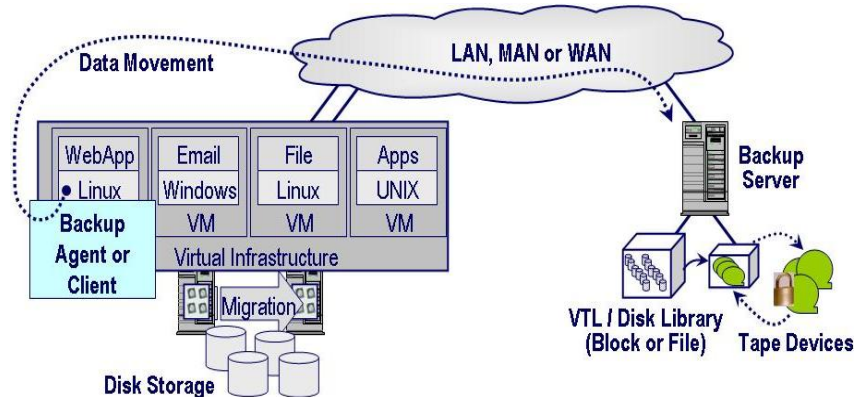


Figure-2: Client-/Agent-Based Backup over a LAN

Agent-based backups shown in Figure-2 are relatively easy to deploy, as they may be in use for backing up the servers being migrated to a virtual environment. The main drawback to agent-based backup is that

Industry Trends and Technology Perspective White Paper **Data Protection Options for Virtualized Servers**

they consume physical memory, CPU and I/O resources causing contention for LAN traffic and impacting other VMs and guests on the same virtualized server.

Backup client or agent software can also have extensions to support specific applications such as Exchange, Oracle, SQL or other structured data applications as well as handling open files or synchronizing with snapshots. One of the considerations regarding agent-based backups is what support exists for backup devices or targets. For example, are locally attached devices (including internal or external, SAS, iSCSI or Fibre Channel SAN or NAS disk, tape and VTL) supported from an agent, and how can data be moved to a backup server over a network in a LAN friendly and efficient manner?

Additional considerations with regard to agent-based backup include:

- Where does the agent exist, in the VM, on a guest OS or in a console subsystem?
- Can the agent reduce the data footprint with deduplication and compression?
- Where does data get backed-up too (local fixed or removable media, remote cloud)?
- What data does the agent backup, can a full bare metal restore be performed, or, only user files?
- Does the agent backup VM file system or virtual disks as well as raw devices?
- How is the software licensed, and what additional drivers or software are needed?
- How does the agent handle data compression or deduplication, and what is the server overhead?
- What are the data security capabilities including encryption and key management?
- How can disk based backup targets with built-in deduplication enhance backup performance?
- Can the agents integrate with VM snapshots or quiesce to minimize downtime?
- What is the performance impact on the physical server of running multiple VM agent backups?
- What scripting or customization is required to support the agent-based backup?
- What level or granularity of backups are provided, full image, differential or incremental, file based?
- Is the software really an agent or, light weight client software able to backup individual servers?
- Does the client or agent software provide significant value to off-set IRM maintenance concerns?
- Does the client or agent software require a dedicated backup or proxy server for operation?

Proxy based backup

Agent- or client-based backups running on guest operating systems consume physical resources, including CPU, memory and I/O, resulting in performance challenges for the server and LAN network during backup (assuming a LAN backup). Similarly, an agent-based backup to a locally attached disk, tape or VTL would still consume server resources resulting in performance contention with other VMs or other concurrently running backups.

In a regular backup, the client or agent backup software, when requested, reads data to be backed up and transmits the data to the target backup server or storage device along with performing associated management and record keeping tasks. Similarly, on restore operations the backup client or agent software works with the backup server to retrieve data based on the specific request. Consequently, the backup operation places a demand burden on the physical processor (CPU) of the server while consuming memory and I/O bandwidth. These competing demands can and need to be managed if multiple backups are running on the same guest OS and VM or on different VMs.

An approach to addressing consolidated backup contention is to leverage a backup server and configure it as a proxy (shown in Figure-3) to perform the data movement and backup functions. Proxy backups work by integrating with snapshot along with application and guest operating system tools for pre- and post-

Industry Trends and Technology Perspective White Paper Data Protection Options for Virtualized Servers

processing. As an example, VMware Consolidated Backup (VCB) is a set of tools and interfaces that enable a VM, its guest operating system, applications and data to be backed up by a proxy while reducing the CPU, memory, and I/O resource consumption of the physical server compared to a traditional backup.

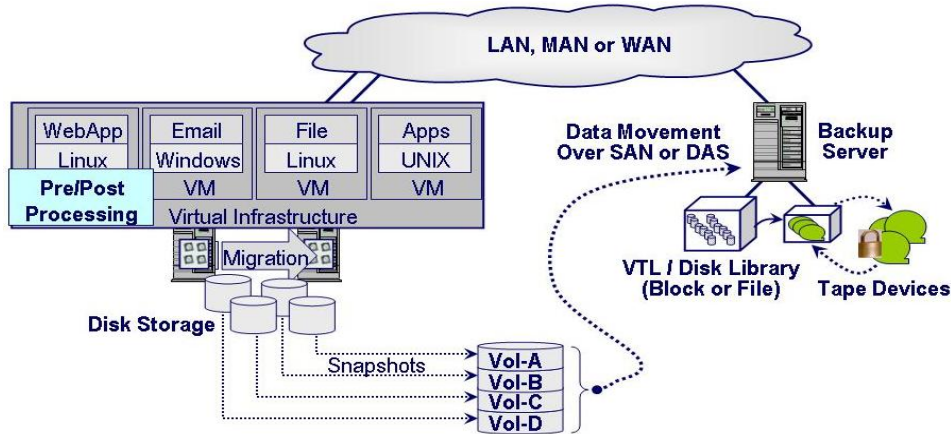


Figure-3: Proxy-Based Backup Such as VMware VCB

VMware VCB is not a backup package. Rather, it is an interface to VMware tools and enables third party backup and data protection products to work. To provide data protection using VCB, third party backup tools are required to provide scheduling, media and backup management. Third party tools also manage the creation of data copies or redirecting data to other storage devices, such as VTLs and disk libraries, equipped with compression and data deduplication to reduce data footprint. Virtual machine virtual disk images are sparse or hollow, meaning that there is large amount of empty or blank space with many similar files that lend themselves to being compressed and deduplicated.

In addition to off-loading physical servers during the proxy backup, LAN traffic is not impacted as data can be moved or accessed via a shared storage interconnect, including direct attached shared SAS storage, iSCSI and Fibre Channel/FCoE SAN or NAS, depending on specific VM implementation. How VCB in a VMware environment works is straight forward, following the progression of the data protection techniques up to now.

First, the backup proxy server running third party backup data protection software tells a VM and guest OS to prepare for a snapshot. The snapshot request which includes performing any pre-snapshot work such as quiescence (suspend, freeze or pause) of file systems, application integration to flush or capture data buffers in memory and commit to disk. Then, the snapshot occurs, followed by any post snapshot processing and un-freeze or un-quiesce of the file system and VM to resume normal processing. The proxy server is then able to access the storage volume to open and read the snapshot to build the backup.

The actual processing or CPU time and I/O impact associated with reading the data to be backed up occurs on the proxy server without the need to move data over a LAN network and off-load the physical server that is hosting the VMs. The proxy server, depending on third party backup software, can then write data to an attached (directly to backup server or via a SAN) disk, tape or VTL backup system, including the creation of multiple copies such as a disk-based backup and copy on tape, to be sent off-site. Third party backup and data protection software on the proxy can also perform other tasks, including replicating the data to another location, keeping a local copy of the backup on disk-based media with a copy at the remote site on disk as well as on a remote off-line tape if needed.



Industry Trends and Technology Perspective White Paper Data Protection Options for Virtualized Servers

Industry Trend

Data deduplication continues to evolve in terms technology maturity along with customer adoption and deployments. While data deduplication debates continue around immediate (aka inband) vs. deferred or scheduled (post processing), discussions are shifting towards what approach or mode to use when, where and why with the availability of solutions enabling policy based data deduplication.

Policy based data deduplication solutions support both immediate and deferred modes of operation in addition to being able to disable deduplication functionality depending on specific policies to meet various service level commitment requirements.

For example, immediate deduplication mode can be used to provide space savings (e.g. reduction ratios) where capacity is a concern for remote or workgroup environments. Deferred or scheduled mode can be used where the primary focus is on performance (data transfer rates) for larger environments that must meet data protection or backup windows.

A hybrid approach has emerged that combines the best of client or agent along with proxy based backup and data protection techniques while countering the issues associated with each of those models. By enabling software to be lightweight in terms of IRM operational and maintenance concerns, while enabling agility and flexibility to use different tiered and virtual storage mediums alleviates many traditional concerns with agent or client based backup software.

Similarly, enhanced software that can perform both sourced side data protection while off-loading server and common issues of proxy based backup solutions are becoming available. For example, a common issue with proxy based backup is limitations on what operating system guests and their file systems that can be supported without loss to operational or data protection functionality. Another common issue is the amount and complexity of customized scripting or other interfacing that must be done to “glue” various 3rd party enabling solutions into virtualization frameworks.

General questions to consider regarding proxy-based backup include:

- Does the proxy-based backup work with virtual disks or raw physical volumes?
- What I/O interfaces are supported for proxy-based backup direct attached, SAS, SAN or NAS?
- Can full image and file level backup and restore be performed and for what OS?
- What is the performance impact of hosting proxy in a VM on the same physical server as guest VMs?
- What licenses fees are required along with application integration for data consistency?
- How are LUNs and volume mapping, masking and zoning changes handled for the proxy?
- How many concurrent backups can run on a proxy, how many concurrent snapshots?
- What guest operating systems and/or file systems can be restored on an individual file basis?

Data Replication (Local and Remote)

There are many approaches to data replication and mirroring, shown generically in Figure-4, for local and remote implementations to address different needs, requirements and preferences. Data mirroring or replication can be performed by various 3rd party data movers either application, layered software or middleware, guest operating system or file system, network or appliance software, or, storage system based. Storage system based replication includes both primary, secondary along with compression and deduplication enabled VTLs.

Industry Trends and Technology Perspective White Paper
Data Protection Options for Virtualized Servers

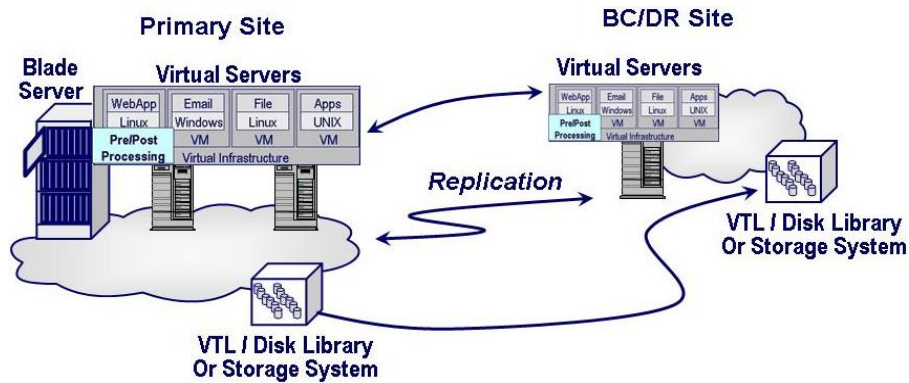


Figure-4: Data Replication for HA, BC and DR Data Protection

Replication should be combined with snapshots and other point-in-time discrete backup data protection to insure that data can be recovered or restored to a specific RPO. For example, if data is corrupted or deleted on a primary storage device, replication will replicate the corruption or deletion to alternate sites, hence the importance of being able to recover to specific time intervals for rollback.

Table-2 presents a summary of various data replication options, benefits and caveats. A general caveat is that replication by itself does not provide complete data protection; replication is primarily for data availability and accessibility in the event of a component, device, system or site loss.

	Benefits	Caveats	When To Use
Application Based	Tight integration with application for consistency, storage system agnostic	Software licensing costs, server performance impact, additional data protection required	Use in combination with other data protection for application data integrity
Guest OS Based Built-in or 3rd party software	May be more economical for smaller environments Storage system and network agnostic with various topology and configuration options	May not be supported by all guest OS or VM environments, server performance impact, software or agents on each of the VMs or guest operating systems and license fees	Multiple vendor (heterogeneous) storage exists, no storage system based replication exists, leverage application specific features
Console System Based	Shifts replication from guest OS or VM to the console, storage agnostic	Not supported by all VM environments, performance impact to the VM environment	Storage system or application or OS based replication not supported
Appliance or Network Based	Off-load server and VMs performance impact, shifts management and costs to 3 rd party appliance, storage system agnostic	May require host software or agents for application aware, introduces another point of management, potential performance bottleneck	Off-load server processing overhead and replicate across multiple vendors storage, no storage based solution exists
Storage System Based	Off-load server performance, less software to manage, heterogeneous storage system support	Potentially higher cost depending on solutions, not all storage supports native replication	Off-load server overhead and management across operating systems. Storage system or VTL to VTL replication

Table-2: Local and Remote Data Replication Options



Industry Trends and Technology Perspective White Paper Data Protection Options for Virtualized Servers

In general, regarding data replication for virtual environments, the following should be considered:

- What are the management and cost implications including software licensing and maintenance?
- What application integration and support exists (e.g. Sharepoint, Exchange, SQL, Oracle, etc)?
- What data transfer modes are supported (e.g. real-time synchronous, time delayed asynchronous)?
- How does the replication integrate with snapshots and application aware data protection?
- What topology options exist (i.e. one to many, many to one, many too many servers)?
- What capabilities exist for virtual to virtual, virtual to physical and physical to virtual server modes?
- What compression, data deduplication or bandwidth optimization are supported?
- Does deduplicated data need to be “re-inflated” in order to be replicated?

Archiving and HSM and ILM

Data preservation or archiving of structured (database), semi-structured (email and attachments) data along with unstructured (file oriented) data is an effective means to reduce data footprint and associated PCFE, backup/recovery, BC, DR and compliance issues. Given the current focus on addressing PCFE and other “green” associated issues, and the growing awareness to preserve data off-line or near-line to meet regulatory compliance and non-compliance requirements, magnetic tape is an effective complementary technology to D2D backups. Magnetic tape continues to be a strong solution for long term cost and performance, effective “green” off-line data preservation and to reduce the data footprint and associated storage management costs including backup. See the Quantum white paper document number WP00130 leveraging tape in its evolving role for supporting long term data preservation.

Local I/O and Networking connectivity

N_Port ID Virtualization (NPIV), an ANSI T11 Fibre Channel standard enables a physical HBA and switch to support multiple logical World Wide Node Names (WWNN) and World Wide Port Names (WWPN) per adapter for shared access purposes and should not be confused with the emerging category of I/O Virtualization¹ or virtual adapters for storage and networking connectivity, Fibre Channel adapters can be shared in virtual server environments across the various VMs. A by-product of the fine grained and unique WWPN is that LUNs can be moved and accessed via proxy backup servers when properly mapped and zoned. Learn more about local, metropolitan and wide area storage networking, interfaces, protocols and technology tips in Chapters 3, 4 5 and 6 in the book “Resilient Storage Networks” (Elsevier)² along with Fibre Channel over Ethernet (FCoE) converged networks and I/O virtualization (IOV) in chapter 9 of “The Green and Virtual Data Center” (CRC).

Putting it All Together - Building a Solution with Quantum’s Help

Quantum’s data protection solutions can be used to address numerous virtual and physical server data protection needs. For example, (Figure-5) Quantum solutions can be attached either directly to virtualized servers or to backup servers for timely backup. Similarly, Quantum tape and DXi disk-based solutions can be attached to backup servers to support proxy or consolidated backups either directly to disk or tape or as part of a disk-to-disk backup. Quantum DXi disk-based solutions leveraging dynamic data deduplication, along with compression, integrate with various leading independent third party data protection management applications.

¹ See “[I/O, its off to Virtual Work We Go](#)” – December 2007 Enterprise Storage Forum

² Chapters 3 – “Networking with your Storage (DAS, NAS and SAN)”; Chapter 4 – “Storage and I/O Networks (LAN and SAN)”; Chapter 5 – “Fiber Optic Essentials” and Chapter 6 “Metropolitan and Wide Area Storage Networking (MAN and WAN)” found in “Resilient Storage Networking – Designing Flexible Scalable Data Infrastructure” (Elsevier Books) ISBN 1555583113 by Greg Schulz.

Industry Trends and Technology Perspective White Paper
Data Protection Options for Virtualized Servers

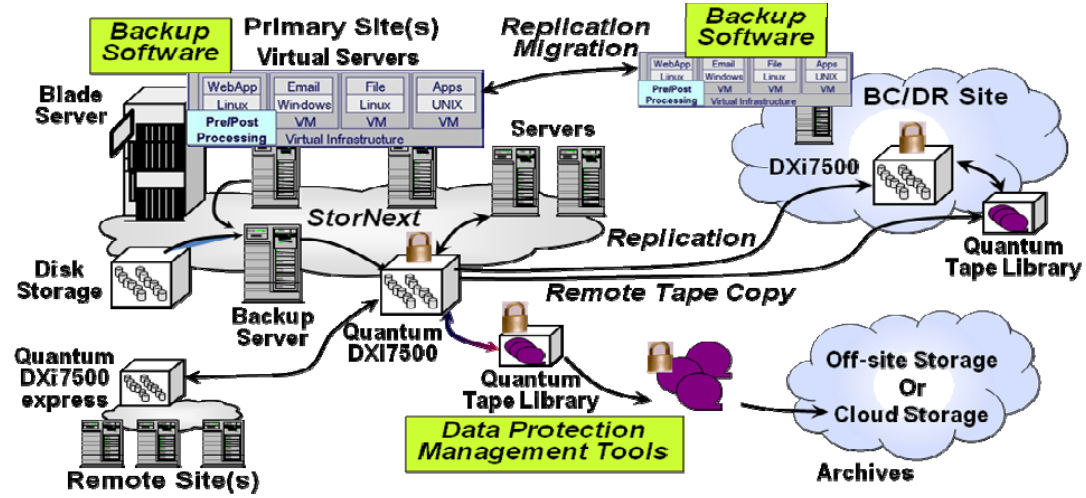


Figure-5: HA, BC and DR Solution for Virtual Servers and Remote Sites

Quantum's StorNext software technology enables heterogenous operating systems on virtual servers and non-consolidated physical servers to share storage and share data. Instead of using NFS and CIFS to meet specific application performance for file system constraints, StorNext data sharing can be used to share data with local and remote clients without the need to copy data across different storage systems and operating systems. Figure-5 shows an example of different techniques, including disk-based backup devices, presenting themselves to backup servers or backup agents as either a block-based VTL or NFS NAS appliances.

Also shown are tape libraries with compression and encryption for long-term retention and archiving as well as disk-based backup VTL leveraging data deduplication with multi-site replication to protect remote locations. Quantum DXi disk based storage systems providing VTL functionality also support policy based data deduplication. Policy based data deduplication enables IT organizations to determine the best mode of deduplication operation to meet specific application service and performance requirements. For example, selecting immediate mode reduces data footprint on the fly to reduce storage capacity requirements in smaller or edge environments including remote offices branch offices, workgroups, call centers or satellite offices where a focus is on data reduction ratios. For larger or more performance sensitive environments where the focus is on meeting or improving on data protection including backup windows, the emphasis is on speed and data transfer rates where deferred or scheduled data deduplication modes can be used.

A tiered storage environment, aligning the most applicable technology and mode of operation to the application, data and service level requirements, a combination of disk, tape, local and remote replication, compression, policy-based deduplication along with data protection management enable an efficient and optimized data protection environment. Quantum target based data protection techniques work with and compliment various 3rd party backup and data protection solutions. In addition, Quantum also provides backup software with their DXi based solutions as part of a turnkey, easy to use out-of-the box data protection for virtualized environments offering. For data protection management Quantum's StorageCare Vision data protection management software is shown enabling simplified management of tiered storage systems. StorNext software may also be used as a hierarchal data management tool. Learn more about Quantum data protection solutions for virtual and physical environments along with related topics for virtual servers at www.quantum.com.

Industry Trends and Technology Perspective White Paper Data Protection Options for Virtualized Servers

Conclusion

The benefits of server virtualization for consolidation as well as management transparency are becoming well understood as are the issues associated with protecting data in virtualized server environments. There are many options to meet different RTO and RPO requirements. Virtualized server environments or infrastructures have varying functionalities and interfaces for application aware integration to enable complete and comprehensive data protection with data and transactional integrity.

A combination of tape and disk-based data protection, including archiving for data preservation, coupled with a data footprint reduction strategy can help to address PCFE or “green” while meeting other needs and issues. There is no time like the present to re-assess, re-architect and re-configure your data protection environment particularly if are planning on, or have already initiated a server virtualization initiative. The bottom line is that virtual server environments require real and physical data protection.

About the author

Greg Schulz is founder of Server and StorageIO, an IT industry analyst consultancy firm and author of the books *The Green and Virtual Data Center* (CRC) and *Resilient Storage Network* (Elsevier). Learn more at www.serverandstorageio.com or on twitter @storageio.

All trademarks are the property of their respective companies and owners. The StorageIO Group makes no expressed or implied warranties in this document relating to the use or operation of the products and techniques described herein. The StorageIO Group in no event shall be liable for any indirect, consequential, special, incidental or other damages arising out of or associated with any aspect of this document, its use, reliance upon the information, recommendations, or inadvertent errors contained herein. Information, opinions and recommendations made by the StorageIO Group are based upon public information believed to be accurate, reliable, and subject to change. This industry trends and perspective white paper is compliments of Quantum Corporation (www.quantum.com).