# 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**

StorageIO

**June 2008**

**This Industry Trends and Perspectives Paper is Compliments of:**
Quantum®
**www.quantum.com**

## Introduction

Server virtualization continues to be a popular industry focus, particularly to address 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. There are many 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 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. Consequently, with the adoption of virtual server environments, having a sound data protection strategy is of magnified importance.

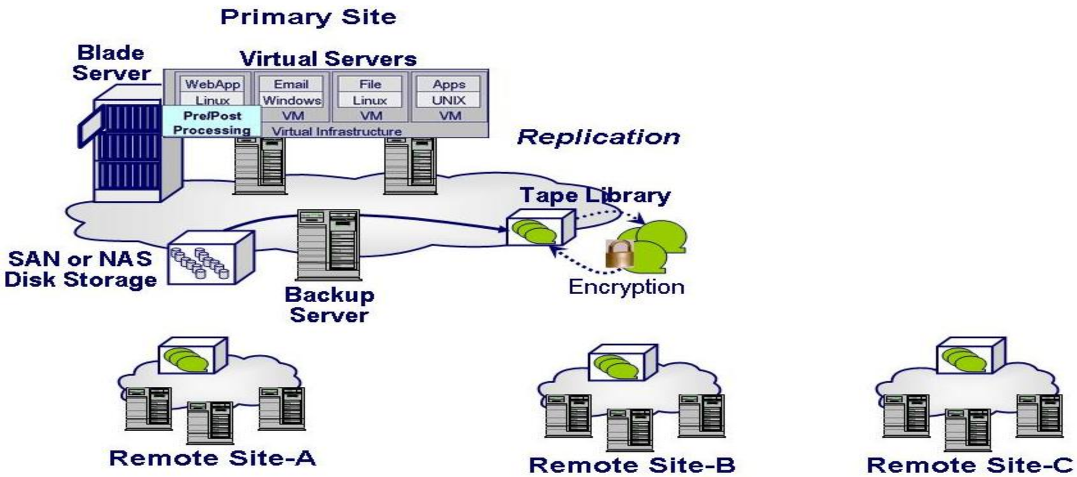


Figure-1: Providing High Availability and Protecting Local and Remote Virtual and Physical Servers

## Background and Issues

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 third wave of server virtualization on the horizon combining the tenets of the two previous scenarios with less of an emphasis on consolidation and centralization and more emphasis around utilizing virtualization to enable dynamic management of servers. For example, using virtualization to support redeployment of servers for workload changes and provide transparency for higher performance workloads that require more server compute power or to enhance and support application high availability (HA), disaster recovery (DR) and business continuance (BC).

There are currently several popular approaches and technologies to achieve server virtualization including Citrix/Xen, Microsoft, Virtual Iron 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

© 2008 StorageIO All Rights Reserved.

examples in this paper will be centered on VMware. The same issues and approaches apply to other virtualization technologies including Microsoft HyperV and Xen-based, such as those from Virtual Iron. 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 and Virtual Iron include SDKs and APIs to support pre- and post-processing actions for customization and integration with VMware Consolidated Backups (VCBs), VMotion or LiveMigration from Virtual Iron.

**Glossary of Common Virtual Terms**

| | |
|---|---|
| Agent | OS or VM software for backup |
| COS | Console Operating System |
| Guest OS | Guest operating system in a VM |
| IOV | I/O Virtualization |
| NPIV | N_Port ID Virtualization |
| OS | Operating System |
| Proxy | Backup Server |
| RDM | Raw Device Mapped Storage |
| RPO | Recovery Point Objective |
| RTO | Recovery Time Objective |
| VCB | VMware Consolidated Backup |
| VIO | Virtual I/O |
| VM | Virtual Machine |
| VMFS | VMware File System |
| VTL/VTS | Virtual tape library / system |

## 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. These solutions emulate various tape devices and tape libraries to co-exist with existing installed backup software and procedures. Magnetic tape remains one of, if not the most, efficient data storage mediums from a PCFE or green perspective for inactive or 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.

**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.

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 de-duplication 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.

© 2008 StorageIO All Rights Reserved.

## 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. 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** | • Move active or static VMs<br>• Facilitate load-balancing<br>• Provides pro-active failover or movement as opposed to data recovery and protection | • Vmotion, Xenmotion and LiveMigration among others<br>• May be physical processor architecture dependent<br>• Moves the running VMs memory from server to server<br>• Shared access storage required along with some other form of data protection for BC/DR or to prevent data loss |
| **Failover (HA) High Availability** | • Proactive movement of VMs<br>• Automatic failover for HA<br>• Local or remote HA<br>• Fault containment/isolation<br>• RAID disk storage | • Proactive move of a running VM to a different server<br>• Requires additional tools to insure all data is moved<br>• Low latency network bandwidth need for remote HA<br>• Replication of VM and application septic data<br>• Isolate from device failure for data availability |
| **Snapshots** | • Point-in-time (PIT) copies<br>• Copies of current VM state<br>• May be application aware<br>• Exists in different locations | • Facilitate rapid restart from crash or other incident<br>• Guest OS, VM, appliance or storage system based<br>• Combine with other forms of data protection for BC/DR or accidental file and data deletion or corruption |
| **Backup and Restore** | • Application based<br>• VM or guest OS based<br>• Console subsystem based<br>• Proxy server based<br>• Backup server or target resides as guest in a VM | • Full image, incremental, different or file level<br>• Operating system and application specific support<br>• Agent or agent-less backup running in different locations<br>• Backup over LAN to backup server and backup device<br>• Backup to local or SAN attached device (disk or tape)<br>• Proxy-based (VCB) for LAN and server free backup |
| **Replication** | • Application based<br>• VM or guest OS based<br>• Console subsystem based<br>• External appliance based<br>• Storage array based | • Application replication such as Oracle<br>• VM or guest OS or 3rd party software based<br>• Application aware snapshot integration for consistency<br>• Replication software running on an external appliance<br>• Storage system controller based replication |
| **Archiving** | • Document management<br>• Application based<br>• File system based<br>• Compliance or preservation | • Structured (database), semi-structured (email) and unstructured (files, attachments, PDFs, images, video)<br>• Compliance or regulatory along with IP and project data preservation for planned or possible future use |
| **Networking** | • NPIV for Fibre Channel<br>• Bandwidth services | • Move VMs independent of zoning or physical changes<br>• Remote mirroring, replication and backups |

Table-1: Data Protection Options for Virtual Server Environments

© 2008 StorageIO All Rights Reserved.

A 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 or Virtual Iron LiveMigration provide proactive movement for maintenance or other operational functions. These tools can be combined with third party data movers, including replication solutions, to 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 drivers or threat risks 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

**Industry Trend:**

**Virtualization, Beyond Consolidation**

Virtual servers are being used primarily for consolidation of underutilized physical servers to improve on power, cooling, floor-space, environmental issues (PCFE) along with simplified management and cost reduction.

Moving forward, the next wave of server virtualization will shift from resource consolidation to the enablement of dynamic virtual data centers to support ongoing maintenance of physical resources including the support of applications requiring more processing power than a single server can provide.

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

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 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.

© 2008 StorageIO All Rights Reserved.

www.storageio.com P.O. Box 2026 Stillwater, MN 55082 651-275-1563 info@storageio.com

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

- Continuous data protection or fine grained data snapshots and replication
- Storage I/O interfaces (internal dedicated, external shared SAS or iSCSI/Fibre Channel 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)
- File system or raw device type data protection and any associated issues
- Data footprint reduction techniques (archive, real-time or off-line compression and de-duplication)
- Data security, including logical and physical protection, authorization and encryption

**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.

**What is Storage VMotion?**

VMware added Storage VMotion to complement 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.

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?

**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).

© 2008 StorageIO All Rights Reserved.

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.

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 quiecesed (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

© 2008 StorageIO All Rights Reserved.

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.

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-based Backup**
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 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 maximizing 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 de-duplication to reduce data footprint of data on disk addressing PCFE concerns or requirements.

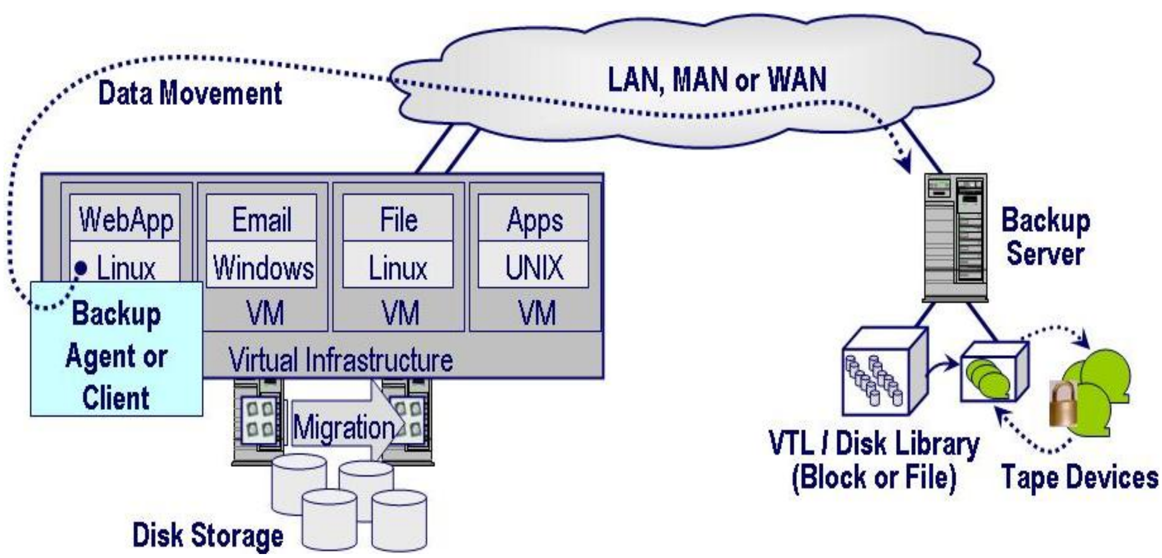


**Figure-2: Agent-Based Backup over a LAN**

© 2008 StorageIO All Rights Reserved.

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 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?
- What data does the agent backup?
- 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 de-duplication, 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 de-duplication enhance backup performance?
- Can the agents integrate with VM snapshots or quiecese 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?

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 performance conflicts and bottlenecks.

**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

© 2008 StorageIO All Rights Reserved.

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-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.

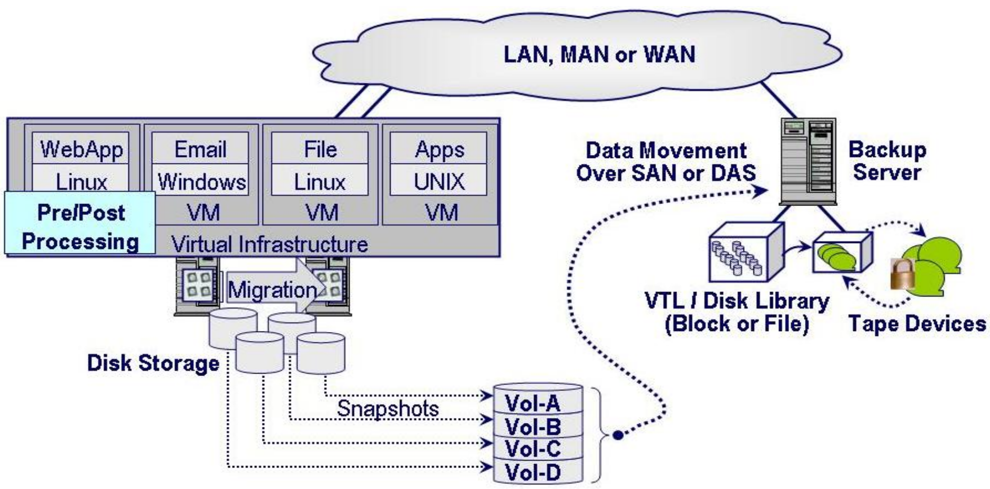


**Figure-3: Proxy-Based Backup Such as VMware VCB**

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 de-duplication 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 de-duplicated.

> **Industry Trend:**
> **Tin vs. Software Wrapped Servers**
> In the quest to consolidate physical servers using virtualization to drive up utilization, the by-product can be negative impact on application and IT infrastructure resource management functions including data protection. Exercise caution when placing backup servers, VTL software or other data protection targets or destination functionality into a VM to avoid contention and performance induced instability.

In addition to off-loading the physical server 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 SAN or NAS, depending on specific VM implementation. How VCB in a VMware VM environment works is rather straight forward, following the progression of the data protection techniques up to now.

© 2008 StorageIO All Rights Reserved.

First, the proxy server running third party backup and data protection software tells a VM to prepare for a snapshot 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.

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?

**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.

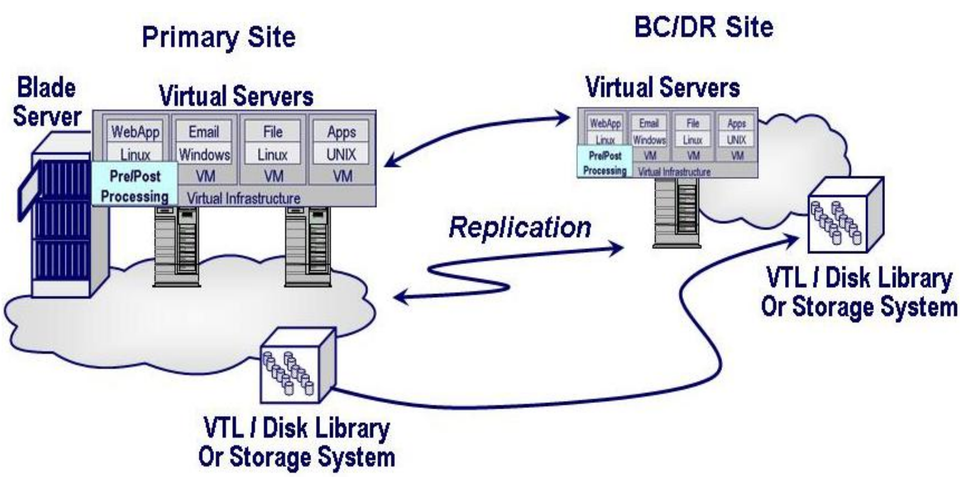


**Figure-4: Data Replication for HA, BC and DR Data Protection**

© 2008 StorageIO All Rights Reserved.

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 "shims", drivers, agents, path managers or other software is required?
- 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 de-duplication or bandwidth optimization are supported?

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. 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.

| | 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 3rd 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 processing overhead and management across multiple operating systems environments |

Table-2: Local and Remote Data Replication Options

**Archiving and HSM**

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

© 2008 StorageIO All Rights Reserved.

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.

### 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 Virtualization[1] or virtual adapters for storage and networking connectivity, Fibre Channel adapters can be shared in virtual server environments across the various VMs, however, the various VMs share a common world wide node name (WWNN) and world wide port name (WWPN) address of the physical HBA. The issue with a shared WWNN and WWPN across multiple VMs is that, from a data security and integrity perspective, volume or LUN mapping and masking have to be performed on a coarse basis.

A by-product of the fine grained and unique WWPN is that a LUN can be moved and accessed via proxy backup servers, such as VMware VCB, when properly mapped and zoned. NPIV is supported by adapter vendors including Emulex and Qlogic along with switch vendors Brocade, Cisco and Qlogic and applicable support from virtual infrastructure such as VMware ESX 3.5. Learn more about local, metropolitan and wide area storage networking, interfaces, protocols and technology tips in Chapters 3, 4 5 and 6 of "Resilient Storage Networks" (Elsevier)[2].

## 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, 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 de-duplication, along with compression, integrate with various leading independent third party data protection management applications.

Quantum's StorNext software technology enables heterogonous 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. The flexibility of StorNext can be used to deploy a transparent tiered storage or hierarchical storage management solution across different operating systems and types of storage from different vendors and applications. 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-

---

1 See "I/O,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.

© 2008 StorageIO All Rights Reserved.

term retention and archiving as well as disk-based backup VTL leveraging data de-duplication with multi-site replication to protect remote locations.

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.

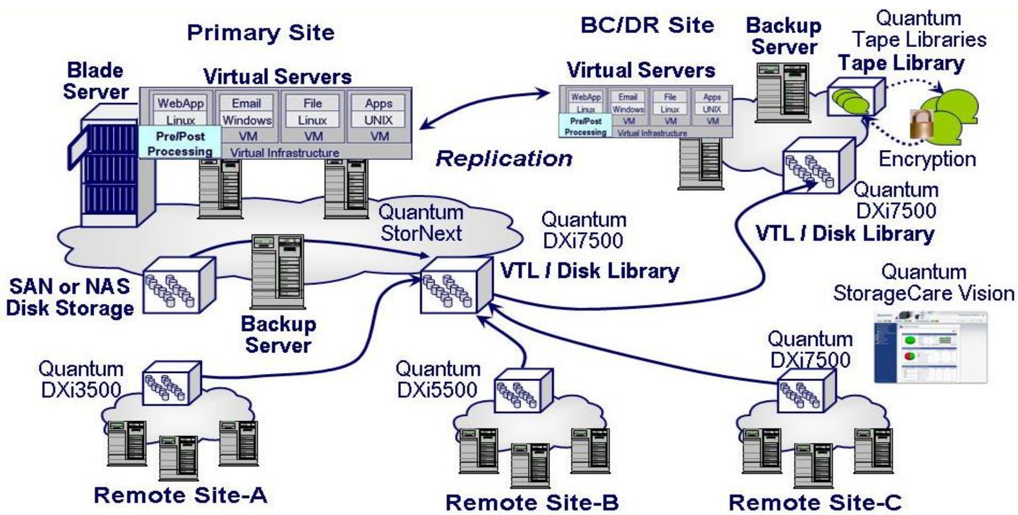


**Figure-5: HA, BC and DR Solution for Virtual Servers and Remote Sites**

## 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. After all, you can not go forward if you can not go back!

## About the author

Greg Schulz is founder of the StorageIO Group (www.storageio.com) and author of the book *Resilient Storage Networks — Designing Flexible Scalable Data Infrastructures* (Elsevier).

*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, inconsequential, 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.*