Backup and recovery best practices for the HP EVA array in a VMware environment using HP Data Protector

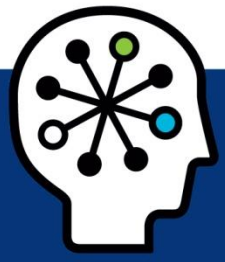


Table of contents

Executive summary

The deployment of HP BladeSystems that contain multicore server technologies deployed with virtualization is a growing trend—one that is rapidly transforming conventional IT infrastructures. These virtualized server environments provide significant opportunities to realize many benefits associated with consolidation, but also pose new challenges ranging from predeployment, management, and high-availability considerations, including backup and restore. In direct response to these challenges, the purpose of this project is to develop a set of technical backup and restore best practices for VMware environments using HP SAN, blade servers, Enterprise Virtual Array (EVA), and Virtual Tape Library (VTL) devices. This project uses several methods for SAN-based backup and restore of both Virtual Machine (VM) and user data by using VMware Consolidated Backup in conjunction with HP Data Protector. By following the recommendations outlined in this white paper, administrators can realize a VMware ESX backup solution that surpasses the 1 TB/h performance threshold, while minimizing storage resource overhead. For large VMware datacenters that require even better backup performance, this white paper illustrates how to design a backup solution with performance that is limited only by the number of backup targets (VTLs).

The recommendations outlined in this white paper focus on backup and recovery best practices that guide administrators through the planning, design, implementation, and management phases of backup and restore operations in VMware environments. You will come to understand backup workflow and solution dependencies, how to help administrators understand potential solution bottlenecks, and, more importantly, how to avoid them. This white paper presents extensive test-proven recommendations combined with a referenced architecture that can serve as an administrative roadmap, ensuring the successful deployment of backup solutions with Data Protector and VMware.

Target audience: This white paper is intended for people who are proposing solutions, who are providing installation services or consulting, and who might be assisting in deploying VMware ESX backup solutions with HP ProLiant servers and HP StorageWorks storage technologies. It is also of interest to IT professionals who might be deploying or managing VMware ESX solutions. This white paper focuses primarily on guidelines developed during extensive testing performed in HP labs.

HP recommends that you use this information in conjunction with product solution documentation and resources referenced throughout this white paper.

This white paper describes testing performed during July through September 2009.

Deploying HP Data Protector with VMware ESX 3.5U4

A VM is a complete representation of the physical server resources stored in a file structure. Administrators who are migrating to VMware environments now have file stores that represent physical servers, operating systems (OSes), and application data. Moreover, in some cases, critical application data stores might actually be embedded in the files that comprise the VM OS. Administrators face significant challenges when designing efficient backup and recovery strategies in virtualized environments. Many administrators have come to realize that carrying over existing backup paradigms from the physical server environment is not adequate and can be highly inefficient, negatively impacting VMware ESX and VM server resources. Instead, administrators require highly adaptable and efficient backup solutions that can be uniquely tailored to individual business requirements in virtualized environments. HP Data Protector meets and exceeds these requirements, providing a one-stop VMware backup and recovery solution.

Data Protector with VMware—a basic overview

Data Protector 6.1x offers a complete VMware backup solution. Data Protector with VMware integration offers administrators a flexible, reliable, and performance-driven solution with an unsurpassed selection of backup options. Administrators can expect the following benefits:

- Backup windows eliminated

  Administrators face complex challenges backing up constantly growing data stores. With Data Protector and VMware ESX integration, backup windows can be reduced or eliminated.

- Resources offloaded

  Backup solutions can be heavy resource consumers. Typically, many backup solutions require backup agents to be installed on every target machine, drawing on ESX host and VM resources. Data Protector offers virtualized environments a shift from this backup model, offloading ESX host, VM, and to a certain extent storage resources (with Data Protector snapshot only), providing uninterrupted operations and scheduling flexibility.

- Backups and restores centrally managed

  Complex datacenters need not translate to complex backup solution management. With Data Protector, administrators can manage backup and restore operations from a single user interface, without scripting and additional add-on requirements.

- Reliable and seamless restores

  Administrators understand that more important than performing backups is having the confidence that backups can be successfully restored. With Data Protector, administrators can rest assured that restore operations are reliable and seamlessly integrated into the user interface. When returning to normal operations is critical, Data Protector with VMware integration provides a straightforward restore solution.

- Backup choices

  Data Protector offers unsurpassed backup flexibility, matching each datacenter's unique requirements. Data Protector provides seamless integration with VMware Consolidated Backup (VCB) image and file functionality. In addition to these VCB standards, Data Protector offers an additional image backup (Data Protector snapshot) and a highly effective application (Zero Downtime Backup (ZDB) and Instant Recovery (IR)) solution. Table 1 provides an overview of features with all of the backup options.

Option overview

Table 1. Backup option overview with Data Protector

DP 6.1x	VCB proxy	Compatibility	Central interface	Storage overhead	ESX host offloaded	Incremental differential	LAN-free option
VCB image	Yes	Most OSes	Yes	High	Yes	No	Yes
VCB file	Yes	Microsoft® Windows® only	Yes	Low	Yes	Yes	Yes
DP snapshot	No	Most OSes	Yes	Low	No	Yes	Yes[1]
DP ZDB/IR	No	Major applications	Yes	Low	N/A	N/A	Yes

[1] With Data Protector integration agents and backup media device installed directly on ESX host server

Figure 1 and Table 2 outline the integration components and associated high-level steps required for each backup option with Data Protector.

Figure 1. Data Protector integration components

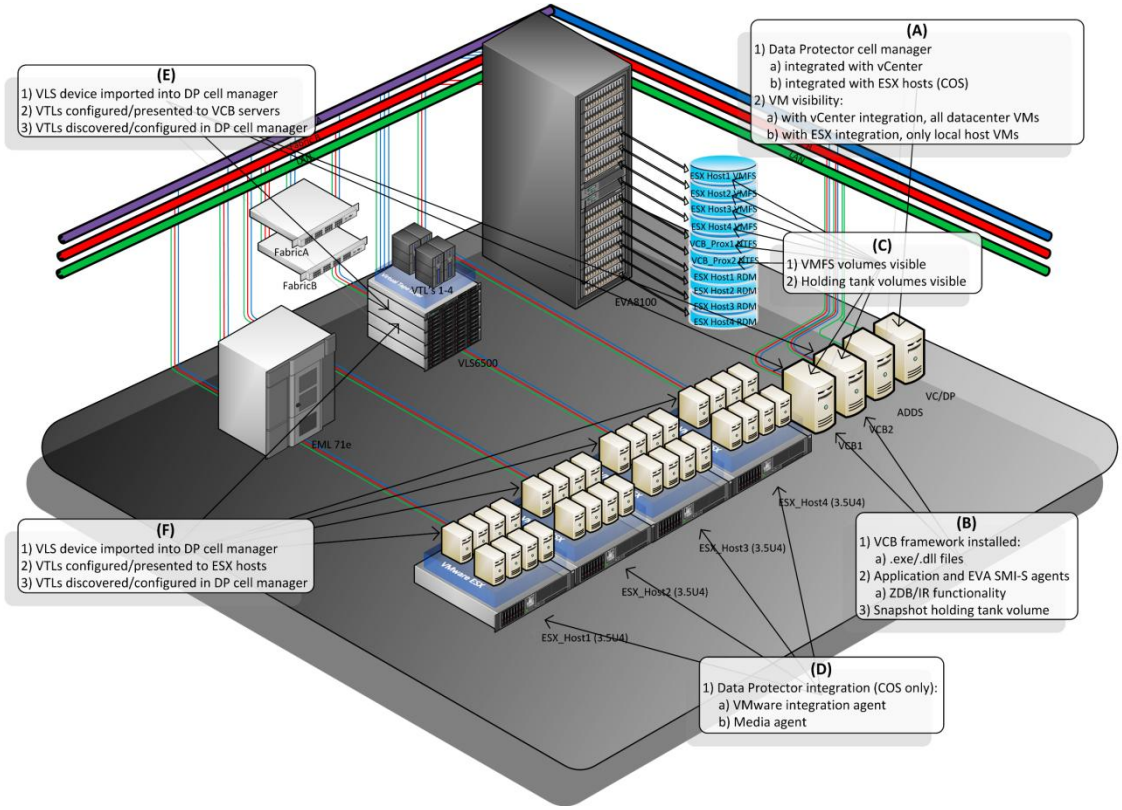


Table 2. Option processes

	VCB image	VCB file	DP snap	ZDB/IR
A	Initiates job with vCenter quiesce, snap, and temp virtual machine disk file (vmdk); and releases original vmdk for export.	Initiates job, and scans file system (via snapshot).	Initiates job with ESX host quiesce, snap, and temp vmdk; and releases original vmdk for export.	Initiates job.
B	VCB framework begins disk export to proxy holding tank volume.	N/A	N/A	Starts agents, identifies objects, places DB in backup mode, and creates replica.
C	VCB proxy discovers VMware File System (VMFS) storage path, and remounts VM on local holding tank volume.	VCB proxy discovers VMFS storage path, and mounts file system.	N/A	With disk-to-disk-to-tape (D2D2T), exports replica to holding tank volume. With IR, replica on array.
D	N/A	N/A	ESX host accesses VM files directly on VMFS volume, and executes commands.	N/A
E	Completes disk export, writes files to VTL, and merges vmdk files on completion.	Immediately writes files to VTL, and merges vmdk files on completion.	N/A	Completes disk export, writes replica to VTL, and replays logs.
F	N/A	N/A	Immediately writes disk export to backup media, and merges vmdk files on completion.	N/A

Things to consider when using Data Protector and VMware integration

SAN

A Fibre Channel (FC) SAN provides the backbone for the test solution and offers exceptional backup performance. VCB backup operations require exceptional SAN throughput performance. The potential for reducing backup windows is partially driven by available bandwidth on the SAN. While adequate performance can be achieved with legacy 2 Gb FC topologies, testing demonstrates that VCB operations consistently benefit from a 4 Gb backbone by filling the entire interconnect bandwidth with concurrent backup streams. For this reason, administrators who are planning large-scale VMware deployments should note that backup performance is highly dependent on I/O throughput and should implement solutions accordingly, even giving consideration to the 8 Gb/s interconnect speeds that are now widely available for the datacenter.

Disk IOPS (I/O operations per second)

Disk resources can be a potential bottleneck in a VCB solution if not properly sized and configured on the backend storage array. Administrators must make sure that spare disk I/O is provisioned for backup operations that occur during production hours. If it is not available, the potential for both workloads to be in contention for disk resources is likely, negatively impacting ongoing production VM operations. Appendix B provides disk sizing information (disk quantity, disk speeds, RAID, and application latency) that can be leveraged with VMware ESX and outlines the referenced configuration workload characteristics.

Array caching

VMware ESX servers benefit from controller cache algorithms. Specifically, write cache is beneficial, but must be monitored for saturation. Read cache is effective for sequential I/O activities, which are essential for both VCB backup operations and creation of VMs from templates. Beyond these scenarios, VMware testing indicates that read cache is not effective with SAN-shared VMFS volumes across ESX host resources. VMware states that this behavior is due to the random rate and timing of incoming read hits from VM resources, nullifying controller read cache algorithms.

Single or dual fabric

Smaller VMware environments might not require or be able to afford the redundancy that a dual fabric provides. However, for environments that require no single points of failure (SPOF), dual fabrics can be leveraged for an additional benefit with backup operations. Administrators can design and distribute backup loads across redundant fabric resources, eliminating potential bottlenecks on the SAN. The test environment, which takes advantage of this design, is described in Referenced configuration.

Host bus adapters

Optimal VCB performance is driven by efficient I/O across interconnected storage, ESX host, VCB proxy, and backup media devices. To achieve optimal performance, administrators must understand workflow processes across host bus adapter (HBA) resources. For more information, see VCB image workflow. By ensuring that HBA resources are properly configured, administrators can eliminate potential bottlenecks during backup operations and reduce backup windows.

Storage provisioning

Administrators have two storage planning methods available with VMware ESX. The context of these two planning methods is outlined in the following sections with an emphasis on backup operations.

Note

For more information about VM threshold levels in both one-to-many and one-to-one volume/ESX host configurations, see Appendix A.

Adaptive storage design

Adaptive storage designs suggest that administrators create a single large volume with a singular storage characteristic (that is, disk protocol/speed and RAID). In an iterative manner, administrators then add four or five VMs and test for acceptable disk/application performance on a single VMFS volume. This design method allows for identification of volume limitations progressively based on VM application workload demand. After limitations are identified, administrators can add or modify storage resources to achieve the best VM performance. An adaptive storage design generally creates fewer, yet larger volumes and has the following characteristics:

- VM provisioning flexibility with fewer storage allocation requests
- Fewer stores to manage
- Potential for volume locking in SAN-shared environments

Note
The adaptive method can potentially incur SCSI reservation contention with snapshot operations (see Volume locking) and large volumes shared across many ESX hosts. For this reason, adaptive storage designs are not the preferred design solution in large-scale backup environments.

Predictive storage design

Predictive storage designs suggest that administrators create several volumes with different storage characteristics (that is, disk protocol/speed and RAID). Administrators can then create VMFS datastores within each volume type that are named according to volume characteristics, and then install applications in the appropriate datastore. A predictive storage design generally creates many smaller volumes and has the following characteristics:

- Less disk resource contention due to locking and SCSI reservation issues
- More efficient disk usage
- Better flexibility with path selection (preferred and multipath) and disk share management
- More flexible backup policies with disk-based snapshots
- More stores to manage

Best Practice
The predictive storage design model is leveraged in the referenced storage configuration and for purposes of this testing. Many datacenters benefit from this design method and its efficient use of storage resources and reduced volume locking with VMware ESX.

Volume locking

Administrators must be aware of potential performance issues with clustered volume resources in a VMware environment that can be the result of backup operations. The VMFS file structure allows the distributed locking that is required for VM isolation and concurrent VM operations on a single clustered volume resource. However, some administrative operations require brief periods of exclusive volume control, which can result in SCSI reservation conflicts across ESX hosts. Namely, the following administrative operations require volume file locks:

- Creating and deleting virtual disks
- Extending VMFS volumes
- Creating and deleting snapshots

For the purposes of HP testing, snapshot operations and their impact on clustered volume performance is considered. Outside the scope of HP testing with backup operations, administrators should understand the effect of these administrative tasks in an ESX clustered volume solution. When an administrator initiates any of these administrative operations against an ESX host, brief periods of exclusive volume control are required for metadata update operations on disk.

While an ESX host gains exclusive volume control, other ESX hosts and their hosted VMs that share the common clustered volume experience SCSI reservation timeouts. VMware testing reveals that during these timeouts, VM throughput levels are reduced (–6.9%) while latency increases (+7.2%) during volume lock intervals. For this reason, VMware recommends that major administrative tasks, including snapshot operations, be performed during off-peak hours whenever possible so they do not impact VM performance.

To be clear, the impact of SCSI reservations is only an issue when a single clustered volume hosts VMs that are registered with more than one ESX host. However, administrators have an alternative for VMs that run mission-critical applications, those that cannot experience any performance degradation whatsoever. The storage design method in which this is accomplished is to register groups of VMs to a single ESX host with a dedicated cluster volume. This configuration does not prevent essential Distributed Resource Scheduling (DRS) and High Availability (HA) operations because the cluster volume is still presented to all participating ESX hosts in the cluster. VMware testing with dedicated ESX host/VM/volume groupings demonstrates that throughput (–1.6%) and latency (+1.8%) are still affected by administrative operations, yet at a much lower rate. Administrators should note that this design method increases the time it takes for administrative tasks to complete by approximately one order of magnitude, but is still measured in mere seconds. For more information about SCSI reservations, file locks, and their impact in ESX clustered solutions, see the VMware publication *Scalable Storage Performance*.

Best Practice
Register the mission-critical VMs that cannot experience the performance degradation that is typically experienced in clustered ESX volumes in groups on a single ESX host with a dedicated volume resource.

Best Practice
Make sure that concurrent snapshot operations that run during production hours do not interfere with production VMs registered across more than one ESX host on clustered volume resources.

VMware file systems

For this white paper, HP testers limit testing to VMFS and Raw Device Mapping (RDM) file systems. VMware's proprietary VMFS file system offers advanced clustering, distributed file lock, and efficient file management at the ESX kernel. These features are required for VMs to coexist on a single volume and for VMware's essential reactive (HA) and proactive (DRS) VM operations.

The VMFS file system is generally the preferred implementation choice for most VM solutions. However, to implement the advanced functionality of Data Protector's ZDB/IR solution, target VM application stores must be configured with RDM storage resources in physical compatibility mode.

VMware implements RDM disk resources to benefit from most of the functionality found with VMFS file systems. Namely, RDM devices can benefit from distributed file locks and snapshot functionality (virtual mode) through disk resource mapping handled by the ESX kernel. However, for the purposes of ZDB/IR with Data Protector, integration requirements (EVA storage array replica functionality)

dictate nonclustered volume presentations and direct SCSI communications with the backend EVA storage array, a feature that is not possible with either VMFS file systems or RDM virtual mode layers.

Deployment best practices

Deployment strategy

Administrators have many choices when designing and implementing backup and recovery environments with VMware ESX. Along with these choices, administrators must carefully consider the hardware and software components to achieve expected backup and recovery performance. The following sections provide important points of consideration when designing a backup environment with VMware ESX and Data Protector. Additionally, Referenced configuration describes many of the best practice recommendations implemented in the test environment.

Best practices for configuring Data Protector and VMware integration

Storage

An EVA storage array provides an exceptional storage solution for demanding VMware production and backup workloads. Even with the EVA's high-performing characteristics, careful storage planning and preproduction testing is always recommended before you implement any storage solution in a live environment. The following sections outline the storage best practice recommendations for VCB workloads and the HP EVA storage array.

Disk group sizing

VMware ESX benefits from the EVA storage array's advanced virtualization algorithms that aggregate large pools of physical disks into logical disk groups. Based on HP testing, VCB operations especially benefit from large pools of disk resources for the holding tank (HT) volume. The HT volume is used as a staging area with VCB image operations before writing the backup to tape. The EVA storage array leverages disk virtualization technology by absorbing and spreading volume I/O that is generated by VCB operations across the underlying pool of disk resources, achieving faster response times with improved throughput transfer rates.

Best Practice
VCB image operations perform best with ten (10K) or eight (15K) FC disk group spindles per VCB disk export backup stream.

Volume RAID type

Protecting against data loss begins with RAID technologies implemented at the storage system controller. For both production VM and backup I/O workloads, administrators must factor in RAID technology overhead. Read performance with either type of workload with the EVA storage array is very comparable between Vraid5 and Vraid1 protection levels. However, Vraid5 write performance does not perform at the same levels as Vraid1 because of the disk and controller parity overhead. With VCB operations, write I/O performance is critical to the HT volume on the backup proxy. For this reason, HP recommends that you configure the HT volume on the fastest storage available in the environment.

Best Practice
Configure the HT volume on Vraid1 storage for the best performance with VCB image disk export operations. For more information, see Figure 16.

Volume ownership

With backup operations, it is essential that volume workflow is balanced across controller resources. Testing with concurrent snapshot (eight streams) operations processed through a single HT volume pushes write I/O activity beyond the EVA controller specification maximums. For more information, see Figure 11.

It is imperative that administrators monitor the impact that backup operations have on backend controller resources, especially with concurrent VCB image backup streams. Even more important, if both VCB and production VM I/O activity is expected to be absorbed by controller resources concurrently, administrators must carefully monitor controller performance counters for potential resource saturation. In the event of saturation, production VMs with latency-sensitive applications are negatively impacted.

Best Practice
Limit VCB image disk export operations to no more than four concurrent streams per controller at any given time to avoid EVA controller write I/O saturation.

VCB proxy

VCB operations require a coordinated flow of data between several initiator (server) and target (array/backup device) platforms. Careful resource planning and a solid understanding of VCB I/O patterns are required if efficient and optimized data transfers are to be achieved. The following sections outline how this objective is realized in the Data Protector and VMware test environment.

Important
Before presenting any volumes to the VCB proxy, make sure that you disable the auto drive assignment (automount disable and automount scrub with diskpart) as outlined in *VMware Consolidated Backup: Best Practices and Deployment Considerations for SAN Environments*.

LAN backups are not the focus of this testing. Instead, a backup environment with FC SAN-attached VTLs is evaluated. For this backup solution, the following significant benefits are realized:

- Backup traffic is kept off the LAN (on slower GigE backbone connections in many datacenters).
- VMware host and VM resources are offloaded from the backup process.
- Backup performance is increased through widely implemented high-speed FC networks.

The following best practices are based on extensive testing and supplement those outlined in *VMware Consolidated Backup: Best Practices and Deployment Considerations for SAN Environments*.

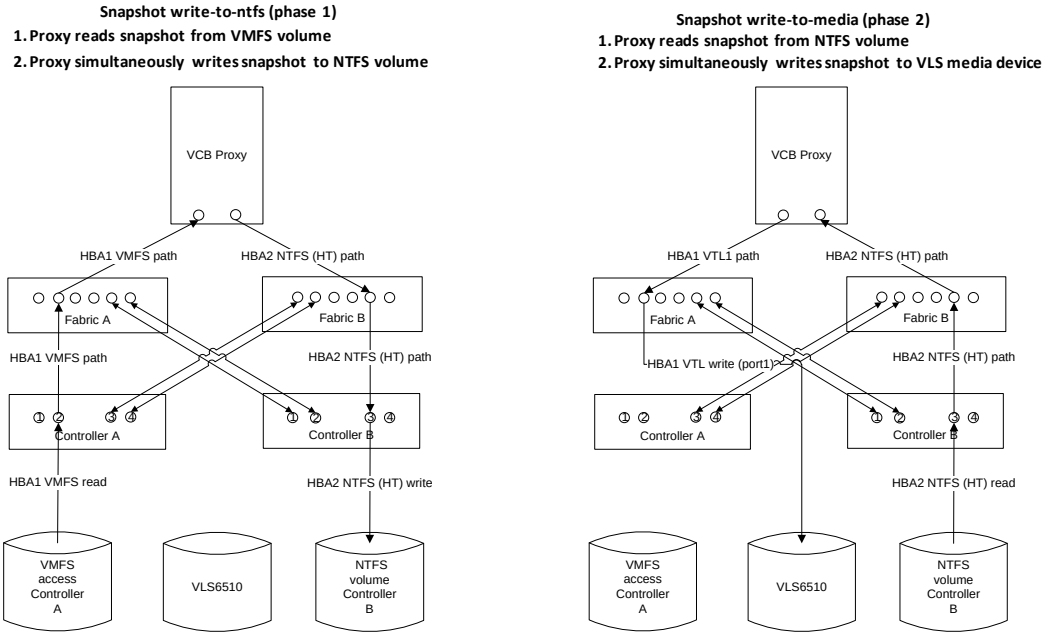
- Use a dual proxy (for information on the referenced solution, see Test hardware and software configuration) or a two HT volume (on a single VCB proxy with separate owning controllers) configuration for balanced storage resource usage with concurrent VCB image/file operations.
- Present VMFS volumes with matching LUN IDs to all ESX hosts and to the VCB proxy (required for proxy path discovery with multiple VMFS volumes).
- Configure the VCB proxy with a minimum of two FC HBAs for best performance.
- Make sure that the HT volume is on the fastest storage available in the datacenter for the best performance (use the default allocation unit size).
- Manage VCB FC HBA paths (zone correctly, see Figure 27) to and from all volumes.

- Configure the VCB proxy with a minimum of two CPUs (four or more CPUs if you are planning concurrent backup operations).
- Install VCB framework 1.5 update 1 (release 150805) or later to reduce SCSI timeout errors with concurrent snapshot operations.
- Consider using NTFS disk compression (Test results and analysis outlines this recommendation) on the HT volume to offload storage array overhead.
- Install the supported HP MPIO DSM (Multipath I/O Device Specific Module) if you are designing a fault tolerant VCB proxy configuration.
- Configure the VCB proxy with a minimum of 2 GB of RAM.
- Disconnect mapped network drives on VMs before running snapshot operations. Mapped drives cannot be part of the backup operation.

VCB image workflow

HP testing demonstrates that managing I/O workflow is imperative with VCB image operations, especially with concurrent backup streams. In the test environment, this is accomplished by configuring two FC HBAs on the VCB proxy and carefully managing volume mapping and zoning presentations across fabric resources. For more information, see Test hardware and software configuration. Figure 2 shows VCB image workflow characteristics in a single proxy configuration.

Figure 2. VCB image workflow



VLS = Virtual Library Systems

Observe the following important points with VCB image workflow from Figure 2:

- Each VCB image operation requires two read and two write operations. For this reason, VCB image is classified as a heavy SAN resource consumer.
- The volume workflow is balanced across both storage controllers (VMFS and NTFS volumes managed at the storage port level with preferred path settings).

- The volume workflow is balanced across fabric resources. VMFS, NTFS, and VTL workloads are all balanced.
- The simultaneous HBA read and write I/O (with two HBAs) on the VCB proxy is isolated on separate HBA resources.

The benefits realized by carefully managing I/O workflow outlined in Figure 2 result in reduced resource contention and increased backup performance with VCB operations.

Note
Limited HBA resources in the test environment prevent configuring a redundant backup solution across SAN fabric resources. However, administrators who require a backup solution with no SPOF could easily mirror existing VCB proxy, ESX host, and VTL media zone configurations (zones dedicated to backup as shown in Figure 25) across the alternate fabric with the inclusion of standby zone member resources.

Best Practice
With VCB image disk export operations, you must thoroughly understand workflow characteristics and manage all resources for the best performance.

VCB file

Administrators have the following choices with Data Protector and file backups:

- VCB file (preferred solution)
- Data Protector disk agent (installed on each VM)
- Storage array volume replicas (not tested)

Testing demonstrates that VCB file is a best practice recommendation with Data Protector for the following reasons:

- No disk agent installs are required on VMs (less management overhead)
- ESX host and VM resources are offloaded (Data Protector agent backups have significant resource overhead)
- SAN backups are possible (Data Protector disk agents are LAN-based backups)
- VCB file backups provide efficient write-to-media performance (disk export to proxy not required)
- VCB file provides backup granularity (pick and choose files to back up)

Unlike VCB image, VCB file is a very efficient solution with flat file backups. However, it is still a best practice to implement two HBAs and/or dual port adapters and manage zoning presentations with VCB file operations. Figure 3 shows the workload differences between the two VCB operations.

Figure 3. VCB file workflow

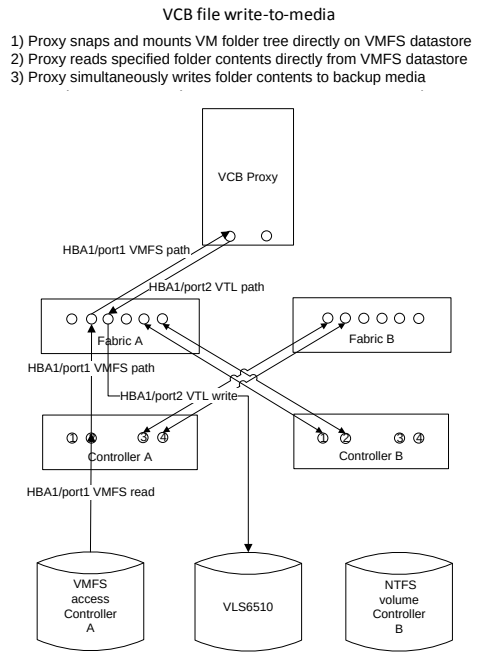


Figure 3 shows the observed VCB file workflow:

- One read and one write operation is required for each backup operation. For this reason, VCB file is classified as a low SAN resource consumer.
- For best performance, isolate read/write I/O on separate HBAs or HBA ports.[2]

Best Practice
Implement VCB file for the best performance and offloading characteristics with both ESX host and VM resources when compared to individual disk agent installations.

Understanding the importance of proxy HBA resources

HP testing evaluates the effects of HBA storage adapters on VCB image disk export operations. Testing demonstrates a significant 16% throughput increase with the addition of a properly zoned second HBA adapter (or port). For more information, see Figure 17.

Best Practice
Configure a minimum of two HBA adapters (or ports) per VCB proxy installation and make sure that read I/O is routed independently from write I/O at all times. For more information, see the referenced zoning configuration shown in Figure 25.

[2] Figure 3 illustrates an ideal configuration using dual port HBAs.

Understanding proxy storage resource consumption

VCB image disk export operations are heavy consumers of storage and proxy server resources. Backup loads (large block sequential write operations) draw heavily on controller resources with concurrent backup streams. The first stage of VCB image, the disk export stage, places a high write I/O demand on the EVA storage array controller. By using two VCB proxies (dual proxy configuration with two HT volumes or two HT volumes presented to a single proxy) as shown in Figure 27, you can reduce the impact on controller resources by spreading the workload across both controller resources.

Best Practice
With two or more HT volumes, make sure that each volume's ownership state is alternated between both controllers for the best storage performance.

Further reductions in storage controller and disk overhead can be achieved with NTFS file compression on the VCB proxy HT volume. This implementation method can be leveraged to enhance VCB's resource offloading attributes (ESX host and VM) at the storage resource layer.

Best Practice
It is imperative that the VCB proxy server have ample CPU resources to achieve results similar to those in the referenced dual proxy configuration with an NTFS compressed HT volume. Without adequate CPU resources, disk export performance can in fact degrade.

In the referenced configuration with four concurrent streams per VCB proxy and NTFS file compression, the following performance characteristics and benefits are recorded:

- A significant offloading of array resources occurs; per controller CPU is reduced from 43% (noncompressed volume) to approximately 8% (compressed volume).
- Storage resource consumption is eased for backup operations outside of scheduled windows (parallel production and backup workloads is potentially absorbed by storage resources).
- A 60% reduction in the required HT volume size (with 50% disk usage of backup source) occurs.
- There is 48% faster backup performance (310 GB/h > 460 GB/h) in a dual proxy configuration with eight concurrent streams.

For backup administrators who are pressed for storage resources and require the best possible performance with VCB image, these benefits can be significant.

Best Practice
Implement HT volume compression for reduced storage overhead and the best possible performance with Data Protector and VCB image operations in a dual proxy configuration.

ESX host

ESX hypervisor requires minimal modification when paired with the Data Protector solution. It is through this solution that scaled performance and efficient image backups with Data Protector can be achieved. Before running Data Protector snapshot operations from the ESX host console operating system, (COS) administrators should note the following best practices:

- Implement SAN-attached VTL media devices on the ESX COS for the best backup performance with Data Protector and VMware integration.
- When implementing Data Protector on the COS, factor an additional 10% ESX host CPU overhead.
- When implementing Data Protector on the COS, increase the ESX host default swap partition size to the maximum setting (1,600 MB).
- When implementing Data Protector on the COS, increase the ESX host default COS memory size to the maximum setting (800 MB).
- Always implement separate Ethernet connections for COS, VM, and restore activity. All image and file restore operations occur over the LAN.

Important
FC SAN-attached VTL media on the ESX console OS is now a recognized solution with VMware ESX 3.5U4. For more information, see *Configuring HP StorageWorks D2D & VLS virtual tape libraries on ESX 3.x (Partner Support)*. As of this writing, ESX 3.5U5 and 4.x are not supported. To request VTL support on these versions, contact VMware at http://www.vmware.com/support/policies/howto.html.

VCB image performs well in the referenced configuration. Data Protector snapshot implemented through the ESX host COS delivers the most scalable, efficient, and high-performing solution available.

Best Practice
Make sure that you configure Data Protector snapshot with the ESX host as a client in the Data Protector management interface. If you configure it with the vCenter instance as a client, backup operations occur over the LAN.

Unlike VCB image, which requires disk export to the HT volume before write-to-media operations, Data Protector snapshot offers the following solution benefits:

- An immediate export of snapshot files to backup media (similar to VCB file operations)
- An additional reduction in storage resource overhead (–53%) from the referenced VCB image configuration with NTFS compressed HT volumes
- An additional performance increase (+147% and surpassing 1.1 TB/h in the referenced configuration) from the referenced VCB image configuration with NTFS compressed HT volumes
- Flexibility with full image, incremental, differential, and mixed-mode image backups
- Scalable performance limited only by the backup media (disk count, number of host ports, and host interconnect speeds) in the environment

Best Practice
Implement Data Protector snapshot on the ESX COS for the best possible backup performance, flexibility in backup choices, and minimal storage and ESX host overhead.

Backup media

Data archival methods are driven by unique business requirements. Many organizations require unique recovery point and time objectives that are driven by media solutions (D2D, D2T, and D2D2T), backup methods (incremental, differential, and mixed-modes), and backup types (image, disk, and file snapshots). As a result of these unique and complex requirements, detailed coverage of these topics falls outside the scope of this white paper.

For the purposes of this testing, evaluating the high-level benefits that administrators can achieve with SAN-attached VTL media devices in a VMware ESX environment is outlined. VTL concepts expand on the foundation of virtualization technology, but at the tape device layer. VTL devices offer administrators the same benefits that are realized with server virtualization (that is, backup media isolation, encapsulation, compatibility, and less power consumption in the datacenter). HP testing uses the HP StorageWorks VLS6510 System as the backup media target.

Note

For a more comprehensive review of VTL concepts, see *HP Data Protector Software Advanced Backup to Disk Integration with Virtual Tape Libraries*.

The following best practice recommendations deliver the best performance when designing a VTL solution with Data Protector and VMware ESX:

- Configure a single host port per VTL device node and zone—isolate active backup operations at all times.
- Present SAN-shared VTL nodes to ESX hosts (for Data Protector snapshots) and VCB proxy (for VCB file operations) for the best resource usage. Do not schedule backup operations concurrently.
- Use multiplexing up to a 4:1 ratio (VM to media drive) for the best backup performance with a moderate impact on restore times (–25% with 4:1 versus 1:1 ratio).
- Use the VLS built-in Echo Copy[3] feature to seamlessly transfer backups to a tape library. This requires storage port configuration.
- Use the Data Protector Smart Copy feature to seamlessly transfer backups to tape. This requires storage port configuration and Storage Management Initiative Standards (SMI-S) integration on Data Protector cell manager.
- Disable VTL media drive compression for increased backup performance (+60%).
- Set the VTL media advanced tape block setting to 256 for the best performance (+50% over the default setting).

Best Practice

Because of the impact that multiplexing can have on restoration times, do not exceed a 4:1 ratio whenever possible.

From an implementation perspective, VTL nodes are configured and presented to the VCB proxy as any normal tape device. However, presenting VTL nodes directly to the ESX COS must follow VMware multitarget requirements. When properly configured, both media changer and tape drive will be assigned a unique SCSI ID. For example, to accomplish this requirement on the VLS6510, the media changer could be configured on FC port 0 and the tape drive on FC port 1.

[3] The Data Protector product services do not support the Echo Copy feature.

Best Practice
When creating and presenting VTL nodes to ESX hosts, follow the VMware multitarget requirements.

VTL solutions provide administrators with flexibility when designing backup solutions with VMware ESX. Administrators can create any number of VTL nodes on the SAN and provision them as required in the datacenter. In the test environment, three VTL nodes are provisioned with dedicated host ports and presented to independent ESX hosts. When parallel backup jobs with Data Protector snapshot run concurrently in this configuration, backup performance averages 1.1 TB/h. It is important to note that to achieve these results, each VTL node is provisioned with a single dedicated FC host port interconnect to limit collision domains and completely fill the entire disk bandwidth of the VLS target.

Best Practice
Testing is limited to a VLS6510 with a 2 Gb/s host interconnect. A 4 Gb/s host port solution is available and is the recommended configuration. With a 4 Gb/s interconnect, two VTL nodes that run concurrently maintain the theoretical disk transfer limit of 1.4 TB/h with a single VLS node.

The three VTL solutions effectively achieve 100% of the theoretical disk throughput performance of a four enclosure (48 disk) VLS6510 solution for 77% of the recorded elapsed time. Two of the three VTL nodes record 150 MB/s each, while the third is limited to 105 MB/s (increases to 150 MB/s on completion of other two nodes). The recorded throughput (405 MB/s) saturates the 1.4 TB/h disk I/O maximum of the VLS6510. HP offers a rich VLS product mix. Administrators can scale beyond the referenced VLS configuration's disk bottleneck with multinode solutions, solving even the largest VMware ESX datacenter backup requirements.

Note
For more information about HP VTL product offerings, contact your sales representative.

Data Protector

Installing Data Protector with VMware integration requires steps outside the scope of this white paper. For detailed installation guidelines, see the *HP Data Protector A.06.11 Integration guide for VMware Virtual Infrastructure, Sybase, Network Node Manager, and Network Data Management Protocol Server*. For detailed licensing information, see the *HP Data Protector A.06.11 Installation and licensing guide*.

After Data Protector is integrated (V6.1x) with VMware ESX, administrators find that operations are streamlined from earlier releases. Full integration of all backup and restore operations are administered through a single management interface, simplifying backup administration complexities. The following are the Data Protector best practices recommendations with VMware ESX:

- Implement SAN-attached backup media devices for scaled performance with concurrent and parallel backup operations.
- When scheduling concurrent snapshot operations, schedule no more than eight VM snapshots (VMware best practice recommendation) on any given VMFS at any given time (per job).

- When scheduling concurrent snapshot operations, pool similar sized VMs together whenever possible to maintain optimal transfer rates throughout the backup operation. Streaming transfer speeds remain high and backup jobs complete together.
- When scheduling concurrent snapshot operations, be aware that VM disk and backup I/O are in contention for storage array resources.
- When scheduling backups during peak hours, make sure that spare disk I/O bandwidth is available on storage resources.
- When restoring VM images, make sure that the original VM being restored is first removed from inventory on the vCenter installation.
- When restoring VM images, plan to have at least one ESX Data Protector COS installation available for restoring VMs directly to ESXi hosts. This feature was added with Data Protector 6.11.
- For datacenters with ESXi and/or 4.0 installations, consider installing ESX 3.5U4 COS backup nodes (with scheduled VM migration to and from for backup operations) to achieve the benefits outlined in this white paper with Data Protector snapshot and VTL solutions.
- Plan to implement a dedicated physical node for file-level restores (can be VCB proxy and/or Data Protector Cell Manager), reducing the overhead for installing, configuring, and maintaining backup disk agents on every VM.
- When implementing dedicated physical restore nodes, use CIFS/NFS shares to quickly distribute recovered files to VMs.

Best Practice
Datacenters with primarily ESXi and/or 4.x implementations should consider installing ESX 3.5U4 COS backup nodes to achieve the benefits of Data Protector snapshot operations. To request VTL support on ESX 4.x, contact VMware at http://www.vmware.com/support/policies/howto.html.

ZDB/IR

Separating Data Protector from other solutions, ZDB/IR provides administrators with the flexibility of implementing VM application-consistent backups and restores. ZDB/IR is tightly integrated with storage array and VM application agents, offloading both ESX host and VM resources from the backup process. ZDB/IR leverages the EVA storage array snapshot and cloning functionality through integration with the SMI-S V1.3. The SMI-S standard focuses on heterogeneous out-of-band (LAN-based) storage management. Data Protector leverages this standard to coordinate a sequence of events, which is shown in Figure 4 with Instant Recovery.

Figure 4. ZDB/IR SMI-S integration

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Starting agent on sqltwo.dp.esx.local.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Resolving objects for Instant Recovery.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" This StorageWorks EVA SMI-S provider has access to the following StorageWorks EVA unit:

Array Name: ESX-BUR-BP
Array WWN: 50001FE1500A90D0
Array Status: Good
Status Description: initialized_attention
Firmware Version: 6200

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" The resolve of StorageWorks EVA units has completed.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" A StorageWorks EVA unit has been successfully located for the storage volume:

Storage volume name: 6005-08B4-0006-8139-0000-D000-0034-0000
StorageWorks EVA name: ESX-BUR-BP

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Beginning the resolve of storage volumes.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" The resolve of this storage volume has succeeded. The information returned is:

Storage volume name: 50001FE1500A90D0\\Virtual Disks\sql\Host3 SQL1\ACTIVE
Storage volume WWN: 6005-08B4-0006-8139-0000-D000-0034-0000
Storage volume UUID: 6005-08B4-0006-8139-0000-D000-0034-0000

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" The resolve of this storage volume has succeeded.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" The resolve of storage volumes has completed.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Beginning the resolve of remote replication relationship for the storage volumes.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Checking for presentations of target volumes.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Objects for Instant Recovery successfully resolved.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Disabling the application system.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" The application system was successfully disabled.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Starting Instant Recovery.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Resuming the application system.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Resumption of the application system completed.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" Instant recovery successfully completed.

[Normal] From: SMISA@sqltwo.dp.esx.local "SMISA" COMPLETED SMIS-Application agent on sqltwo.dp.esx.local.

==========
Session completed successfully!
==========

The benefits of ZDB/IR include the near elimination of databases being placed into hot backup mode and the availability of the powerful instant recovery (IR) feature. With IR, previously saved volume replicas are immediately restored, providing backup administrators with the flexibility to recover application databases in mere seconds through completely automated processes.

Best Practice
For immediate recovery of application databases with VMware ESX virtual machines, implement Data Protector ZDB/IR.

vCenter

Backup operations with Data Protector and vCenter integration have few notable performance-related considerations. However, HP testing demonstrates improved VCB snapshot completion rates with concurrent operations after upgrading the default Microsoft SQL Server Express database installation to an Enterprise edition. Additionally, in this test environment, the user interaction through the VI (Virtual Infrastructure) client while managing the vCenter instance is improved with a locally installed database versus one installed on the network. In addition to these recommendations, implement the following VMware vCenter best practice recommendations specific to backup operations:

- Install the latest version of VMware tools on the VMs. The latest sync driver is required for quiescing operations before backup operations.
- Create a defined vcbuser account with predefined VMware consolidated backup privileges.

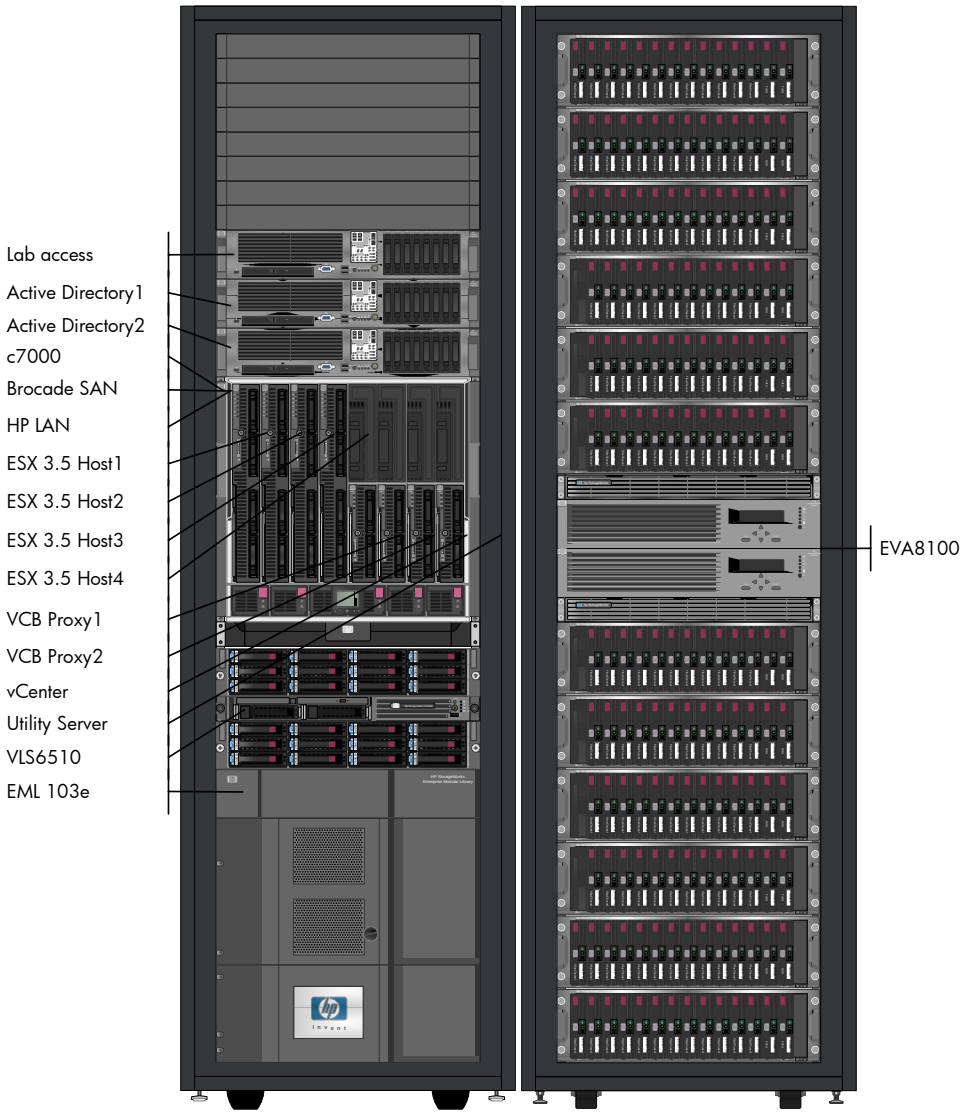
- Configure the VCB proxy server administrator and vcbuser as the same account. This protects consolidated backup privileges from being compromised.
- With multiple VCB proxy installations, implement unique vcbuser accounts and passwords.
- Rotate vcbuser passwords.

Performance tests

Test hardware and software configuration

The following sections provide detailed information outlining the test configuration and tests conducted. Many of the recommendations throughout this white paper are based on these test results. Figure 5 shows the physical configuration of the referenced solution.

Figure 5. Solution hardware and software configuration



EVA8100

- Disk quantity: 144
- Disk type: 300 GB 10K FC
- Host interface: 4 Gb FC x 4 per controller
- Vraid levels tested: 0/1/5
- Control cache: 2,048 MB per controller
- Read cache: 1,024 MB per controller
- Write cache: 512 MB per controller
- Mirror cache: 512 MB per controller
- Firmware version: 6200

BladeSystem enclosure

- c7000 (V2.41)
- Brocade 4/24 SAN Switch x 4 (V6.0.2d)
- HP 1/10Gb Virtual Connect Ethernet Module x 2 (V1.2)
- ROM version: I14/I15

ESX Host servers

- HP ProLiant BL480c G1
- Dual-Core Intel® Xeon®, 3.0 GHz x 2
- 28 GB RAM
- QLogic QMH2462 4 Gb FC HBA x 2
- NC326m Dual Port 1 Gb NIC x 1
- HP Integrated Lights-Out (iLO) version: 1.70

VCB/vCenter servers

- ProLiant BL460c G1
- Quad-Core Intel Xeon, 3.0 GHz
- 4 GB RAM

Infrastructure servers

- DL385
- Dual-Core AMD Opteron™, 2.6 GHz
- 4 GB RAM

Software

- VMware ESX 3.5U4
- VMware VirtualCenter 2.5U4
- VMware Infrastructure Client 2.5
- VMware Consolidated Backup 1.5U1
- HP Data Protector 6.1x
- HP StorageWorks Command View TL/Command View VLS 2.3
- HP StorageWorks Command View EVA Software 8.0.2
- HP ProLiant Support Pack v8.2

- Microsoft Windows Server 2003 (VM and infrastructure servers)
- Microsoft SQL Server 2008 (VirtualCenter and ZDB/IR nodes)
- SMI-S 1.3 (ZDB/IR integration)

VLS6510

- 2 GB RAM
- Disk quantity: 48
- Disk type: 250 GB 7.2K SATA
- Host interface: 2 Gb FC x 4
- RAID level tested: 5
- Media tested: ESL E-Series LTO-3
- Firmware version: 2.3.0

EML 103e

- Tape drives: LTO-3 x 4
- Host interface: 2 Gb FC x 2
- Firmware version: 1070

Test storage layout

The EVA storage array and VMware ESX solution performs best with large disk group configurations as discussed previously. For this reason, the referenced solution is created with a single online disk group (DG) with 144 10K FC spindles as shown in Table 3.

Table 3. EVA8100 disk group configuration

Disk group	Disk number	Disk type	Disk size (GB)	Total available storage (TB)	Protection level
Online DG	144	FC 10K	300	43.2	1
Nearline DG	24	FATA 7.2K	500	12.0	2

Note
Appendix B provides IOPS comparison tables with both 15K and 10K disk group configurations.

Table 4 lists the provisioned test environment with dedicated VMFS resources per ESX host server for reduced SCSI contention with administrative routines.

Table 4. EVA8100 online virtual disk configuration

Vdisk contents	Vdisk name	Vdisk type	Vdisk quantity	Vdisk size	Vraid	Backup options
VMs	Host1-VM	VMFS	1	500 GB	1	File/image
VMs	Host2-VM	VMFS	1	500 GB	1	File/image
VMs	Host3-VM	VMFS	1	500 GB	1	File/image
VMs	Host4-VM	VMFS	1	500 GB	1	File/image
Flat files	Host1-FileServer	VMFS	1	1,000 GB	5	File/image
Flat files	Host2-FileServer	VMFS	1	1,000 GB	5	File/image
SQL DB	Host3-SQL1	RDM	1	1,000 GB	1	ZDB/IR
SQL DB	Host4-SQL2	RDM	1	1,000 GB	1	ZDB/IR
Holding tank	Prox1-HT	N/A	1	500 GB	0	N/A
Holding tank	Prox2-HT	N/A	1	500 GB	0	N/A

Best Practice
With file server implementations, make sure that you configure the VMFS block size to 1/2/4/8 MB for provisioning 256 GB/512 GB/1 TB/2 TB file stores respectively.

Best Practice
For large file server installs, configure a dedicated VMFS volume for reduced SCSI contention and best performance.

VMware ESX with Data Protector benefits from optimized partition configurations. These adjustments facilitate improved sequential streaming operations with Data Protector and VMware ESX. Additional partition modifications are made as listed in Table 5.

Table 5. ESX host optimized partition configuration

Mount point	Approximate size	Type	Use
/boot	100 MB	Ext3	Service Console
/	5 GB	Ext3	
/tmp	2 GB	Ext3	
/usr	5 GB	Ext3	
/opt	2 GB	Ext3	
(none)	1.6 GB	swap	
/var/log	5 GB	Ext3	Log files
/vmfs/volumes	~60 GB	VMFS	Local VM files
(none)	100 MB	vmkcore	Core dump

Best Practice
Increase the ESX default swap size to the maximum setting of 1.6 GB. Additionally, increase the ESX host default COS memory size to 800 MB for the best backup performance with Data Protector.

Key performance metrics

The following key performance metrics are used to monitor and record performance in the referenced solution and are leading indicators of backup performance.

EVA storage array

HP EVA Host Connection

HP EVA Host Port Statistics

HP EVA Physical Disk Group

HP EVA Storage Array

HP EVA Storage Controller

HP EVA Virtual Disk

VCB proxy servers

LogicalDisk

Memory

Paging File

PhysicalDisk

Process

Processor

Workload

The workload is limited to concurrent streaming operations with VCB image, VCB file, and Data Protector snapshot operations. Each specific test result outlines the backup workload per test routine. All read activity averages approximately 256 KB transfer sizes. All write activity averages approximately 64 KB transfer sizes.

Test results and analysis

The following tests and analysis are based on extensive HP testing with Data Protector and VMware ESX and form the basis for many of the best practice recommendations throughout this white paper.

EVA storage array testing

The following sections describe the testing and analysis of backup operations on EVA storage array resources.

EVA disk group MB/s comparison (VCB image)

Figure 6 through Figure 9 show the VCB image backup tests (eight concurrent VCB image backup operations) run on four different sized disk groups. These tests illustrate the importance of proper disk group sizing with concurrent backup streams with VMware ESX.

Figure 6. Disk group size comparison (MB/s)

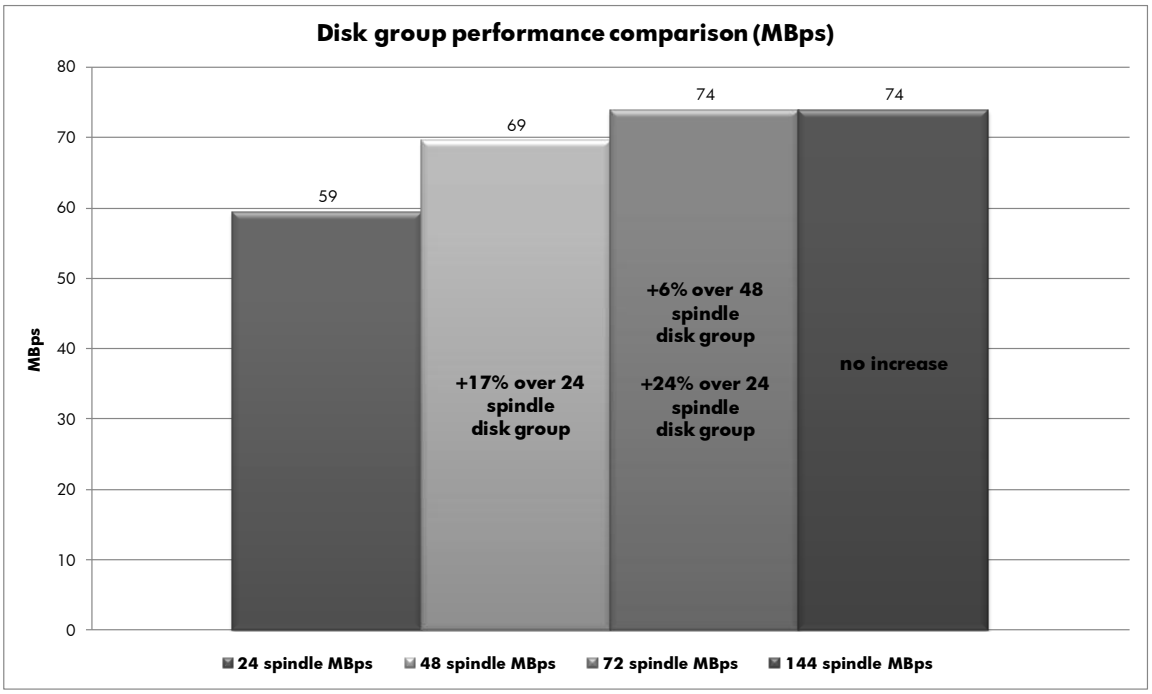


Figure 6 shows a moderate, yet progressive throughput increase with VCB image operations up to the 72 spindle disk group configuration. At this point, MB/s performance remains flat when compared with the larger 144 spindle disk group. A preliminary evaluation of Figure 6 might lead to the conclusion that VCB image backup operations are only moderately impacted by smaller disk group configurations. However, other factors must be evaluated before you form this conclusion and are illustrated further by understanding VCB IOPS demands as shown in Figure 7.

Figure 7. Disk group size comparison (IOPS)

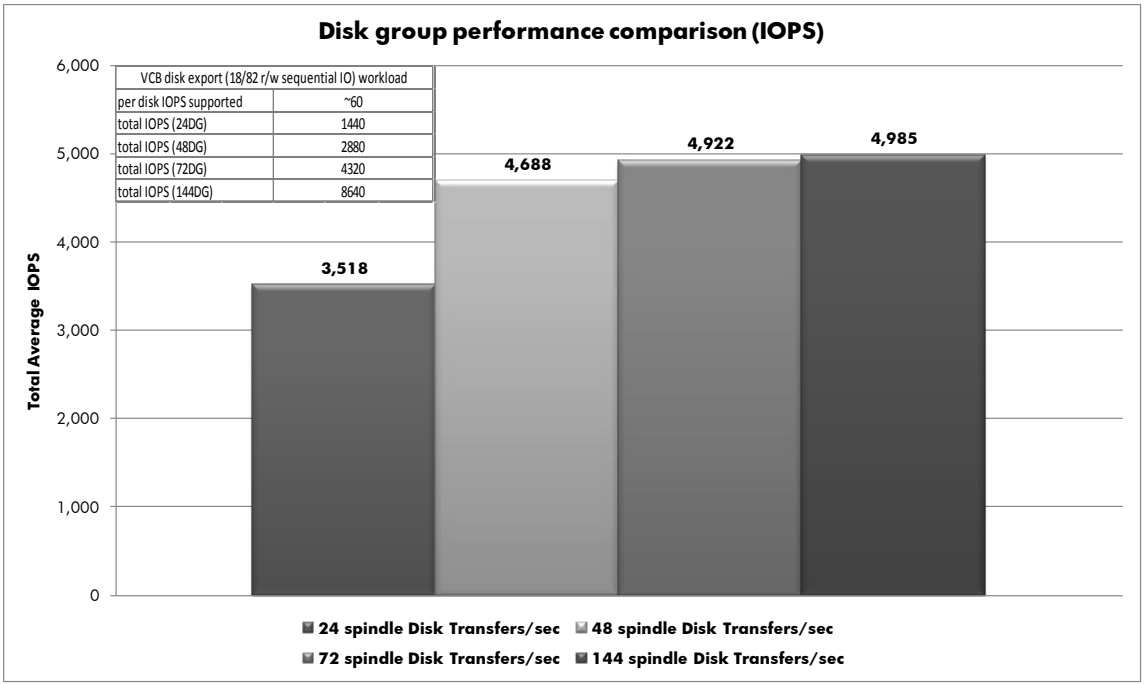


Figure 7 shows the following test results:

- The table in Figure 7 provides the expected IOPS maximums for the tested storage configuration (10K FC disks) and a large block sequential workload.
- Supported disk IOPS (approximately 60) is calculated based on an 18% read and 82% write sequential workload as shown in Figure 30.
- The 24 spindle disk group can effectively service 1,440 IOPS for the given backup workload, yet is achieving 3,518 IOPS (surpassing the expected disk IOPS by 144%).
- The 48 spindle disk group can effectively service 2,880 IOPS for the given backup workload, yet is achieving 4,688 IOPS (surpassing the expected disk IOPS by 63%).
- The 72 spindle disk group can effectively service 4,320 IOPS for the given backup workload, yet is achieving 4,922 IOPS (surpassing the expected disk IOPS by 14%).
- The 144 spindle disk group can effectively service 8,640 IOPS and the effective backup workload is achieving 4,985 IOPS (achieving IOPS that are 73% of the expected total).
- With IOPS overhead available, the only disk group that is capable of effectively absorbing both production VM and backup workloads simultaneously is the 144 spindle configuration.
- The VCB disk export workload requires ten 10K (eight if using 15K) spindles per backup stream for optimal performance ((Disk Transfers/sec ÷ backup streams) ÷ expected disk IOPS).

$$10 = \frac{4985 \div 8}{60}$$

Based on IOPS analysis in Figure 7, both the 24 and 48 spindle disk groups are undersized for the given workload. As a result, throughput and IOPS performance is impacted, although at a somewhat lessened degree respectively. The 72 spindle disk group levels out (closely sized for the backup

workload) while the 144 spindle configuration maintains a 43% disk overhead. The effects of placing an oversized backup workload on undersized disk resources become more evident when evaluating disk queue depth and latency counters on the EVA storage system in the following figures.

EVA disk group queue depth comparison (VCB image)

Figure 8. Disk group size comparison (queue depth)

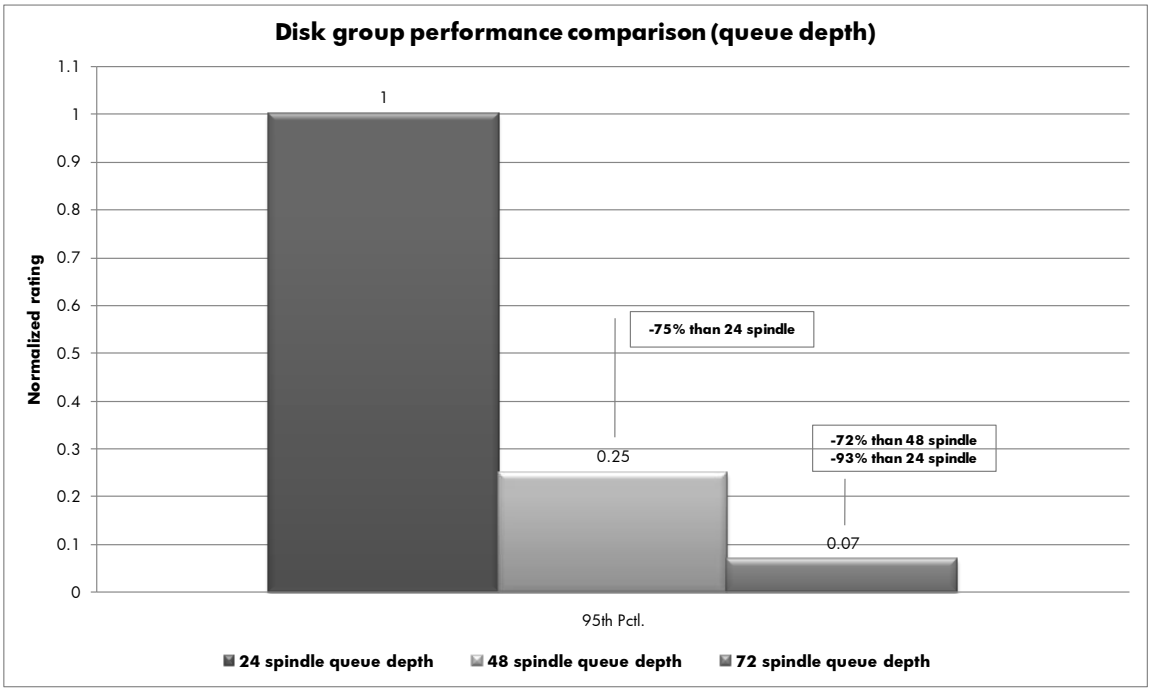


Figure 8 shows the following test results:

- The 144 spindle disk group is excluded from the comparison with a minimal recorded change from the 72 spindle configuration.
- A 95th percentile (top 5% of recorded samples omitted) value is recorded.
- The 24 spindle test results provide a normalized baseline of comparison.
- The undersized disk groups maintain significantly higher queue depth levels than the closely sized (per backup workload only) 72 spindle configuration.
- Recorded queue depths are less than 5 on the 72 spindle configuration.
- Recorded queue depths are greater than 15 on the 48 spindle configuration.
- Recorded queue depths are greater than 50 on the 24 spindle configuration.

In addition to increased queue depths on the undersized storage configurations, significantly higher latencies are recorded as seen in Figure 9.

Figure 9. Disk group size comparison (read/write latency)

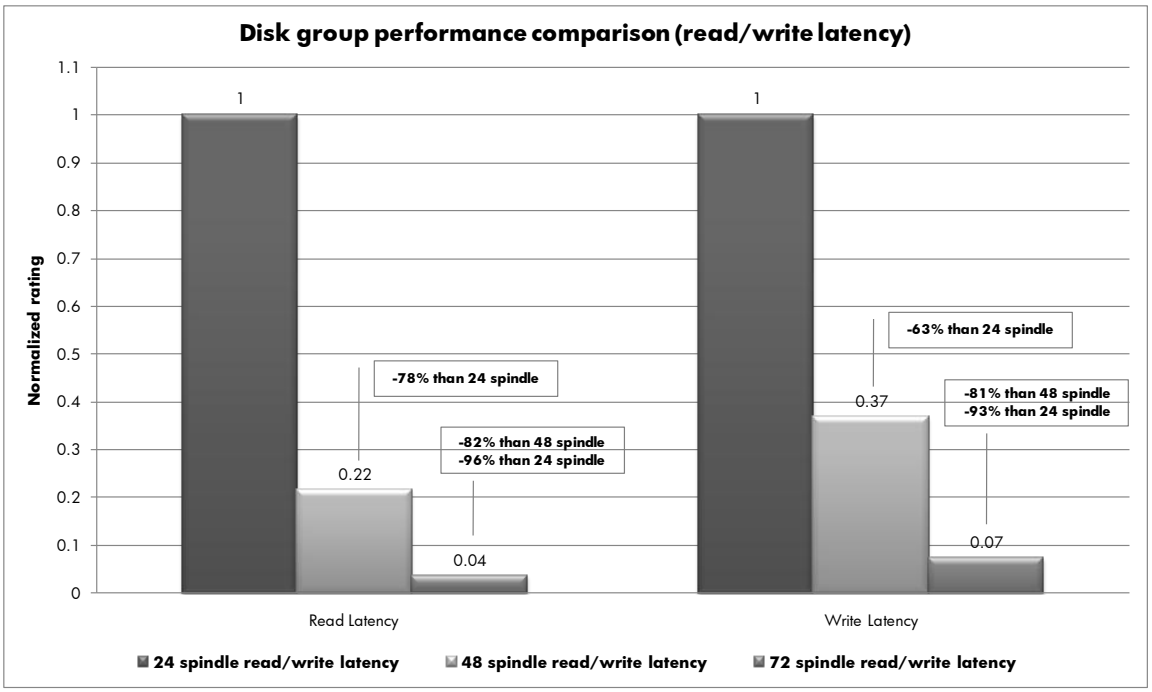


Figure 9 shows the following test results:

- The 144 spindle disk group is excluded from the comparison with a minimal recorded change from the 72 spindle configuration.
- Average disk read/write latency values are used.
- The 24 spindle test results provide a normalized baseline of comparison.
- The undersized disk groups maintain significantly higher latency levels than the closely sized (per backup workload only) 72 spindle configuration.
- Both read and write latency recorded less than 30 ms on the 72 spindle configuration.
- Both read and write latency recorded more than 50 ms on the 48 spindle configuration.
- Both read and write latency recorded more than 250 ms on the 24 spindle configuration.
- Both undersized disk groups are being saturated with the tested backup workload.

The preceding disk group performance test results (Figure 6 through Figure 9) illustrate the impact of configuring concurrent backup streams on undersized storage systems with VCB image operations. The net effect for storage systems that are this grossly underprovisioned is poor performing solutions prone to potential disk failures. Finally, testing demonstrates the importance of proper disk sizing and the need to properly provision storage for the expected system workload, especially with parallel production and backup operations.

Controller CPU cost—MB/s and IOPS (VCB disk export only)

Figure 10 shows the controller CPU cost per MB/s and IOPS with VCB image disk exports with and without compressed HT volumes and shows significant offloading of storage resources with compressed NTFS volumes on the VCB proxy server.

Figure 10. EVA controller CPU cost comparison

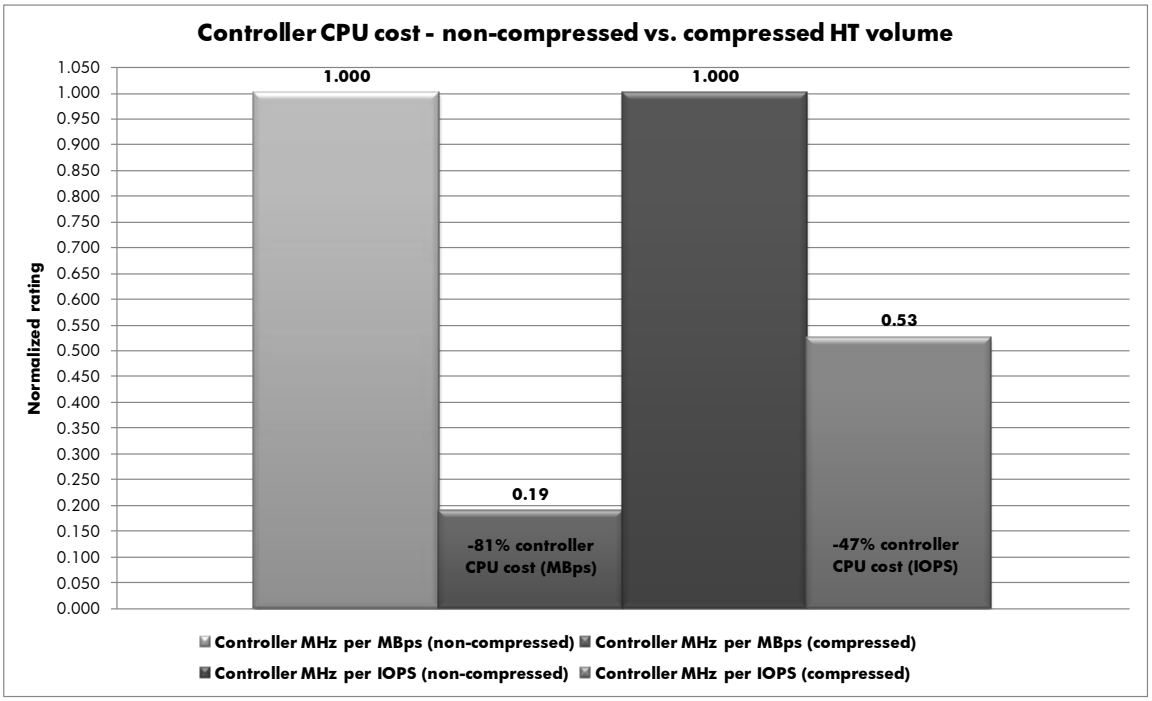


Figure 10 shows the following test results:

- The tests compare the proxy HT volumes with and without NTFS compression enabled.
- The tests represent the disk export stage only to the VCB proxy HT volume (applicable with VCB image and ZDB D2D2T operations only).
- The tests are configured with two VCB proxies, each with four concurrent VCB disk exports (eight VCB disk exports total).
- The noncompressed storage volume serves as a normalized baseline of comparison.
- Each storage controller realizes an 81% reduction in CPU cost per MB/s transfer with an NTFS compressed volume.
- Each storage controller realizes a 47% reduction in CPU cost per IOPS transfer with an NTFS compressed volume.

A compressed HT volume reduces the disk write activity on storage resources with VCB disk export operations. This in turn reduces the controller cost per CPU for both MB/s and IOPS activities.

Controller write I/O usage (VCB disk export only)

Figure 11 shows the impact on controller write resources when running sequential backup streams and forms the basis for the recommended maximum of four concurrent VCB backup streams on a noncompressed NTFS HT volume.

Figure 11. EVA controller write I/O comparison

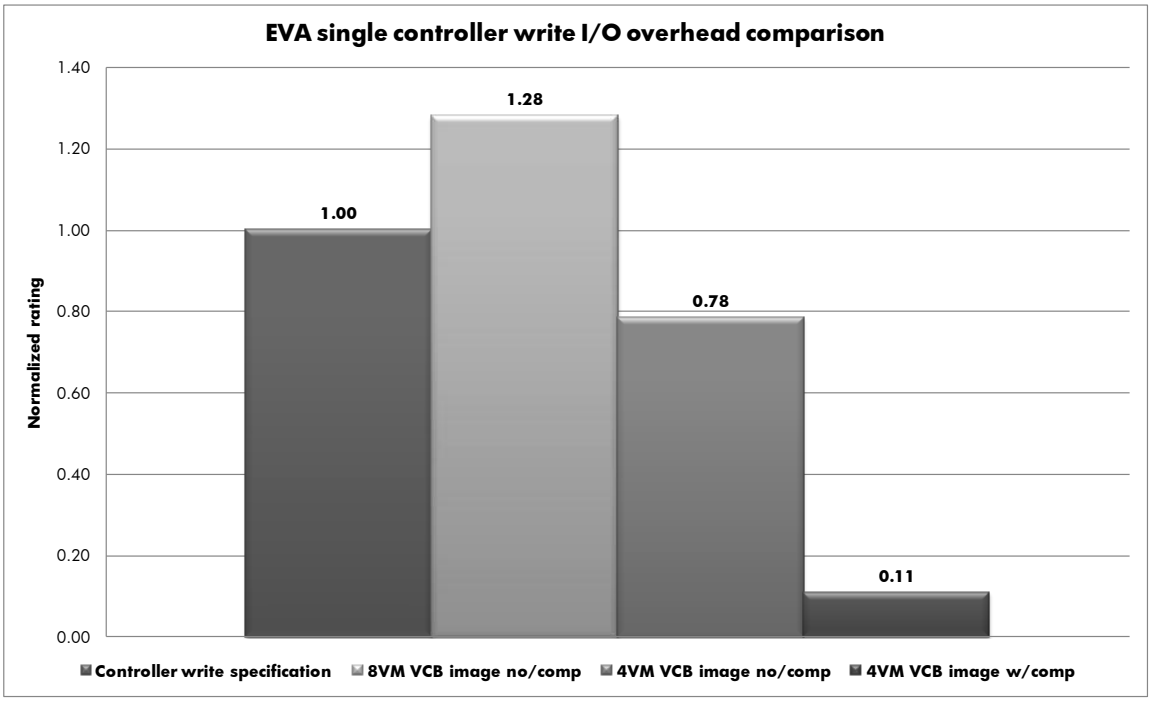


Figure 11 shows the following test results:

- The tests compare the controller write I/O overhead with and without NTFS compression enabled.
- The tests represent the disk export stage only to the VCB proxy HT volume (applicable with VCB image and ZDB D2D2T operations only).
- The tests are configured with eight concurrent VCB disk exports processed through a single controller, four concurrent VCB disk exports processed through a single controller without NTFS compression, and four concurrent VCB disk exports processed through a single controller with NTFS compression.
- The EVA controller write specification maximum serves as a normalized baseline of comparison.
- The 8VM disk export results in EVA controller write processes that exceed the specification maximums by 28%.
- The 4VM disk export on a noncompressed HT volume records 78% of the EVA controller write specification maximum.
- The 4VM disk export on a compressed HT volume records 11% of the EVA controller write specification maximum.

Note
The EVA8x00 is capable of approximately 500MB/s writes (approximately 250 MB/s per controller) at 30 ms latencies, limited only by memory bus architectures. These specifications presume a balanced workload across both controllers. In certain cases, a single controller can be pushed beyond 250 MB/s by leveraging unused memory resources that are typically dedicated for mirror port traffic (proxy reads).

Controller CPU usage (VCB disk export only vs. Data Protector snapshot)

Figure 12 shows the controller CPU usage of all image operations tested: VCB image with and without NTFS compressed HT volumes and Data Protector snapshot operations. These tests clearly characterize controller CPU workload and the benefits of both compressed NTFS volumes and Data Protector snapshot operations.

Figure 12. Controller CPU overhead comparison

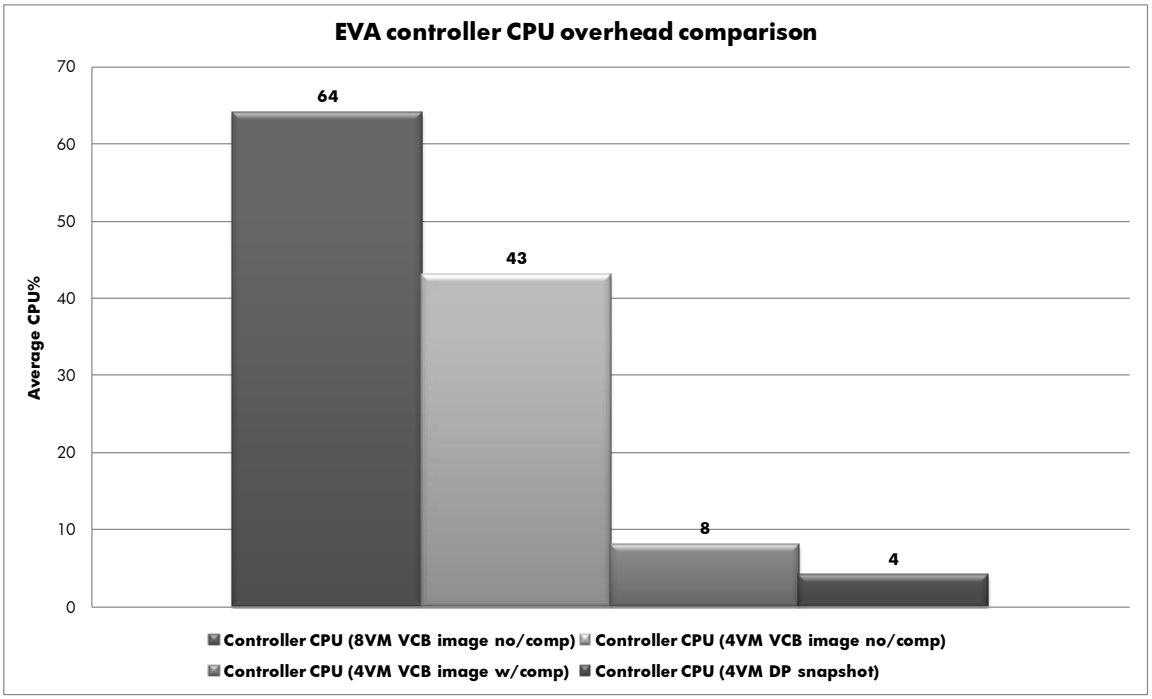


Figure 12 shows the following test results:

- The tests compare the controller CPU usage of VCB image disk export with and without NTFS compressed HT volumes vs. Data Protector snapshot operations.
- The VCB image tests represent the disk export stage only to the VCB proxy HT volume (applicable with VCB image and ZDB D2D2T operations only).
- The tests are configured with eight concurrent VCB disk exports processed through a single controller without NTFS compression, four concurrent VCB disk exports processed through a single controller without NTFS compression, four concurrent VCB disk exports processed through a single controller with NTFS compression, and four concurrent Data Protector snapshot operations with direct write-to-media operations.
- The 8VM disk export without NTFS compression records a 64% hit on controller CPU.

- The 4VM disk export without NTFS compression records a 43% hit on controller CPU (33% less than the 8VM test).
- The 4VM disk export with NTFS compression records an 8% hit on controller CPU (87% less than the 8VM and 81% less than the 4VM tests).
- The 4VM Data Protector snapshot records a 4% hit on controller CPU (94%, 91%, and 50% less than the preceding tests respectively).

Best Practice
Implement Data Protector snapshot, which offers administrators the most storage efficient image backup performance of all tested solutions.

Controller CPU cost—MB/s (VCB disk export only vs. Data Protector snapshot)

Data Protector snapshot records significant reductions in both controller CPU cost per MB/s and CPU cost per IOPS when compared to the VCB proxy solution as shown in Figure 13 and Figure 14. These tests continue to characterize the controller CPU workload, building on the advantages Data Protector snapshot offers over the VCB image with an NTFS compressed volume (previously the most efficient storage solution).

Figure 13. Controller cost comparison VCB proxy vs. DP COS (MB/s)

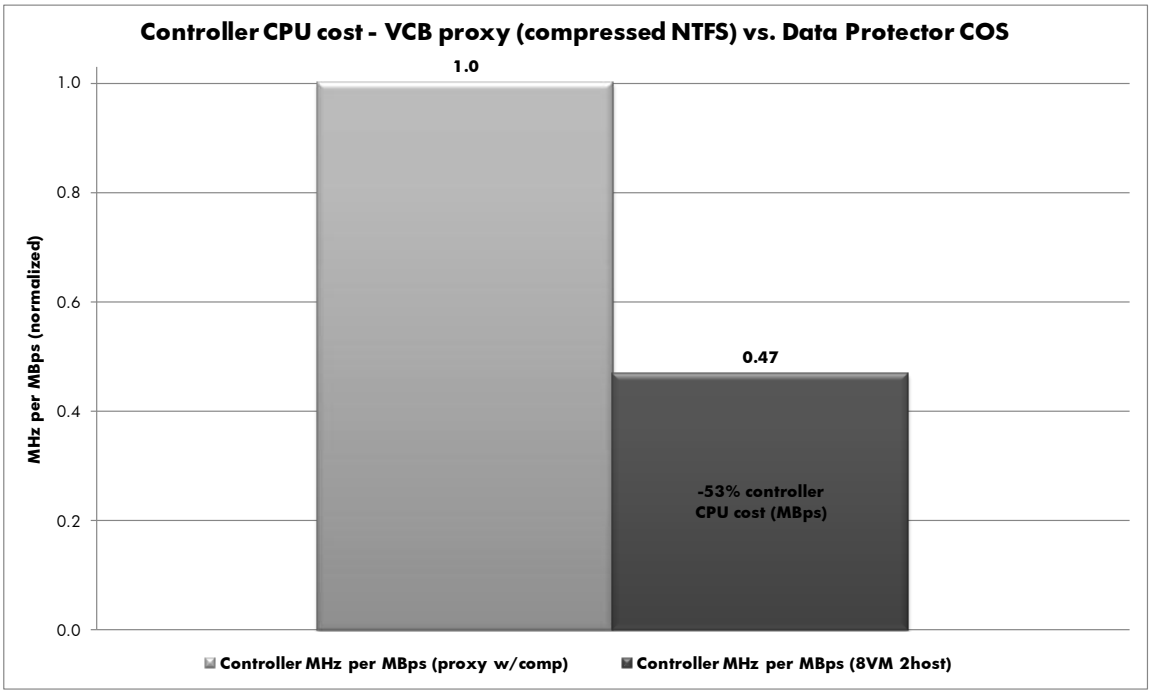


Figure 13 shows the following test results:

- The tests compare the storage controller costs of the VCB image/proxy server and Data Protector snapshot/ESX COS solutions.
- The illustration represents the disk export stage only to the VCB proxy NTFS volume (applicable with VCB image and ZDB D2D2T operations only) and the disk export to backup medium with Data Protector snapshot.

- The normalized baseline for comparison (Figure 10) is configured with two VCB proxies; each with four concurrent VCB disk exports (eight VCB exports total) with NTFS compressed volumes (the most storage efficient method with VCB image disk exports).
- The normalized baseline test earlier achieves an 81% reduction in MHz per MB/s on the EVA controller when compared to a noncompressed NTFS volume (shown in Figure 10).
- Data Protector snapshot records an additional 53% reduction in MHz per MB/s when configured with two ESX hosts, each with four concurrent snapshot streams and when compared to the baseline test.

Controller CPU cost—IOPS (VCB disk export only vs. Data Protector snapshot)

Figure 14. Controller cost comparison VCB proxy vs. DP COS (IOPS)

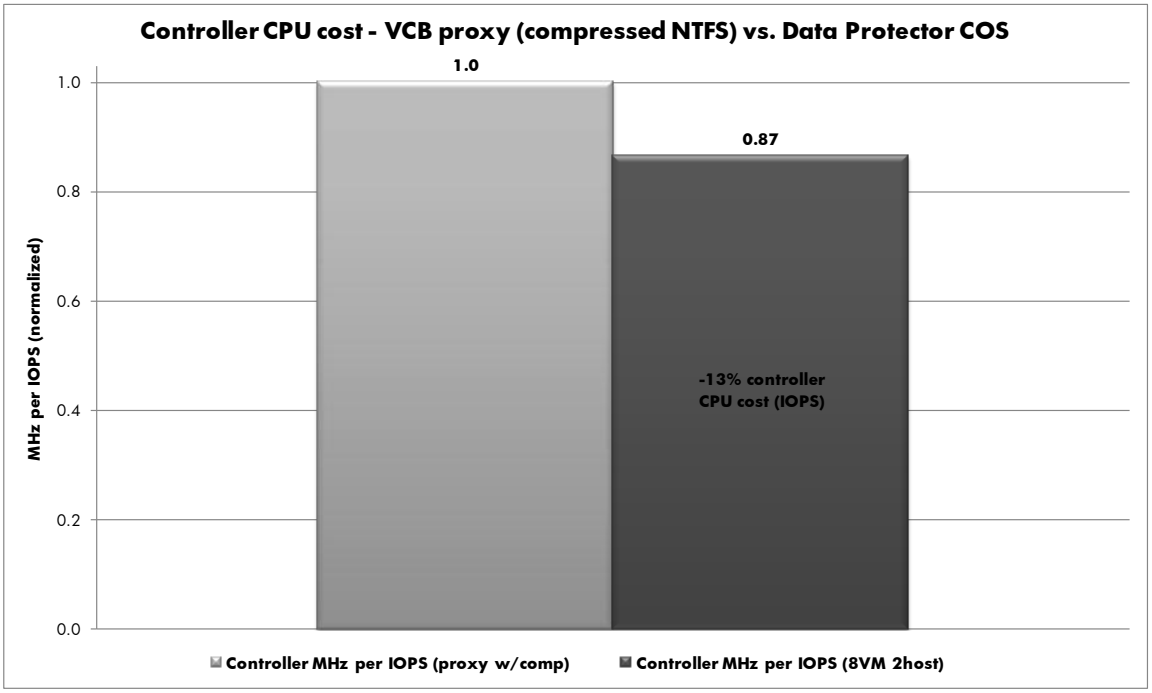


Figure 14 shows the following test results:

- The tests compare the storage controller costs of the VCB image/proxy server and Data Protector snapshot/ESX COS solutions.
- The illustration represents the disk export stage only to the VCB proxy NTFS volume (applicable with VCB image and ZDB D2D2T operations only) and the disk export to backup medium with Data Protector snapshot.
- The normalized baseline for comparison (Figure 10) is configured with two VCB proxies; each with four concurrent VCB disk exports (eight VCB exports total) with NTFS compressed volumes (the most storage efficient method with VCB image disk exports).
- The normalized baseline achieves a 48% reduction in MHz per IOPS on the EVA controller when compared to a noncompressed NTFS volume (shown in Figure 10).
- Data Protector snapshot records an additional 13% reduction in MHz per MB/s when configured with two ESX hosts, each with four concurrent snapshot streams and when compared to the baseline test.

Controller read cache performance

Figure 15 shows the effectiveness of the EVA controller read cache algorithm with snapshot operations and validates the enabling of read cache on the EVA storage array with VMware ESX.

Figure 15. EVA cache hit backup performance

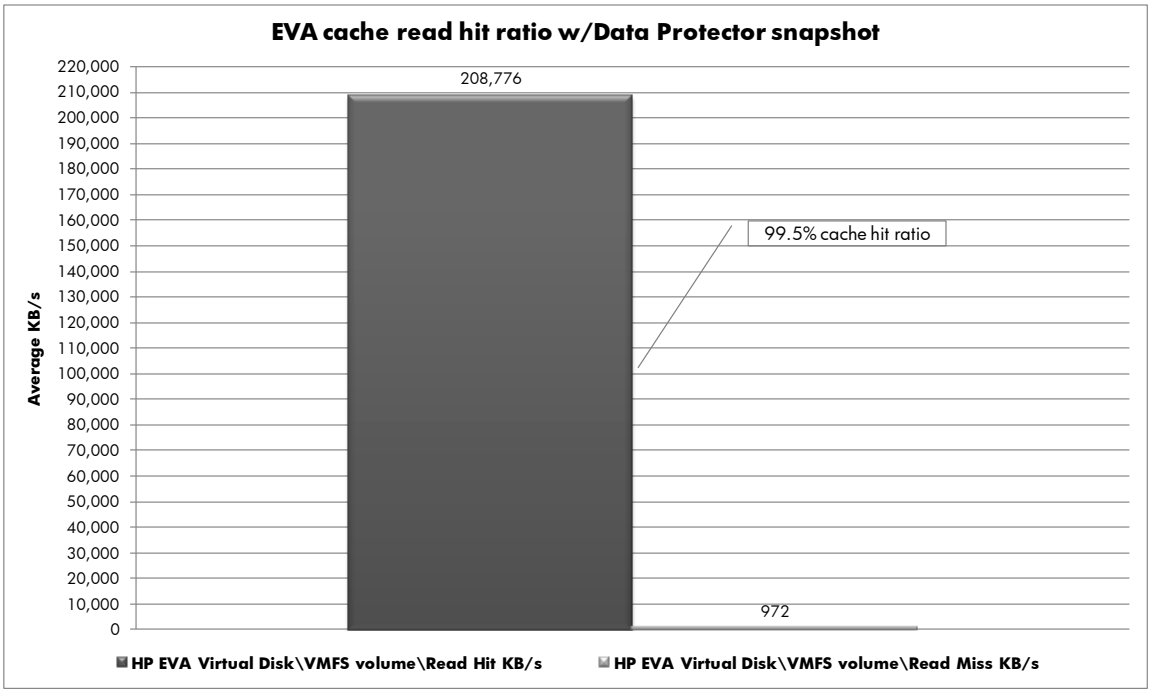


Figure 15 shows that sequential backup operations benefit from the EVA storage array's advanced controller read cache algorithms, achieving a 99.5% hit ratio, and should be enabled at all times.

ESX host and VCB proxy testing

The following tests demonstrate storage and server resource consumption with backup operations on both the VCB proxy and ESX host.

VCB proxy volume performance

Figure 16 shows recorded performance differences between Vraid levels with the proxy HT volume and forms the basis for the recommended usage of Vraid1 with the proxy HT volume. In all tests recorded, the HT volumes are carved from a 144 disk group configuration with 10K FC spindles.

Figure 16. VCB proxy HT Vraid performance comparison

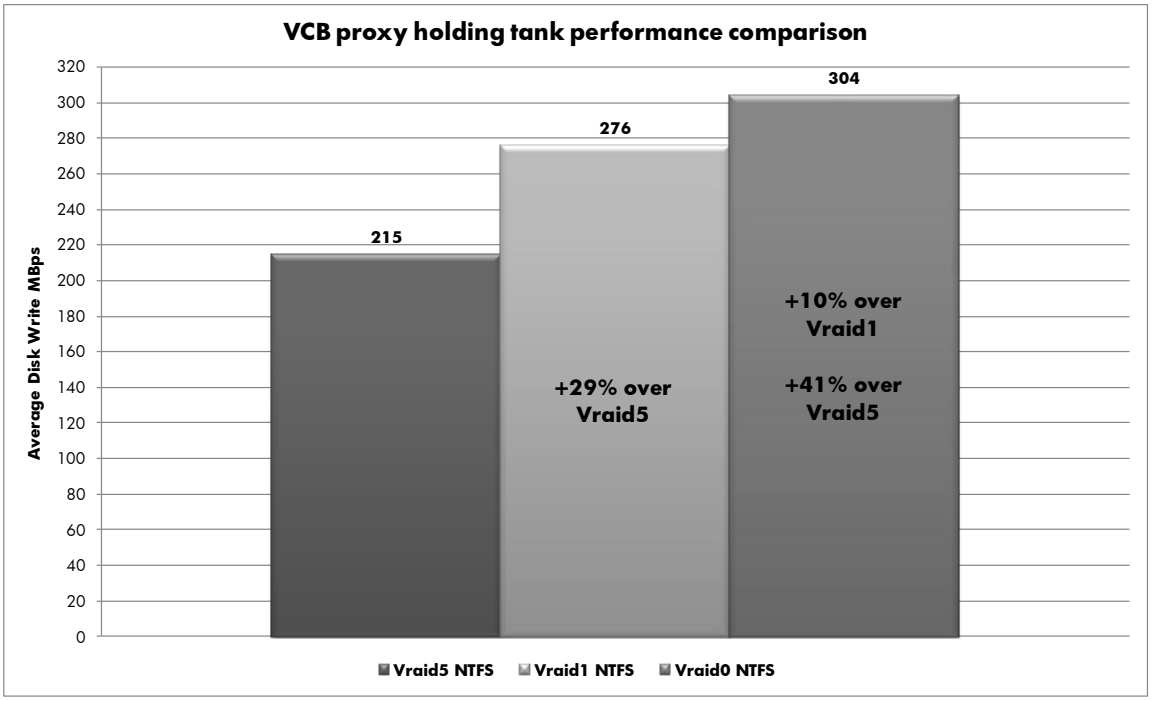


Figure 16 shows that in all cases where backup performance is critical, administrators should be aware of the impact of Vraid5 (–22% and –29% compared with Vraid1/Vraid0 respectively) with VCB disk export operations. The purpose of the HT volume is a temporary staging area for snapshots only. Because no persistent data is stored on this volume, environments that require the best possible backup performance and those concerned with the allocation of a space-consuming Vraid1 volume might consider Vraid0 as a viable option. Testing with Vraid0 records an additional 10% throughput increase over Vraid1 with concurrent disk export operations.

Important

Administrators should be aware that any disk failure in the disk group from which the HT volume is provisioned causes Vraid0 volume failure. After a failed disk is replaced, the affected volume is available again.

VCB proxy HBA performance

Figure 17 represents the recorded performance gain that is realized by a second VCB HBA adapter and forms the basis for the recommendation of a second HBA adapter with VCB proxy backup operations.

Figure 17. VCB proxy performance comparison (single vs. dual HBA)

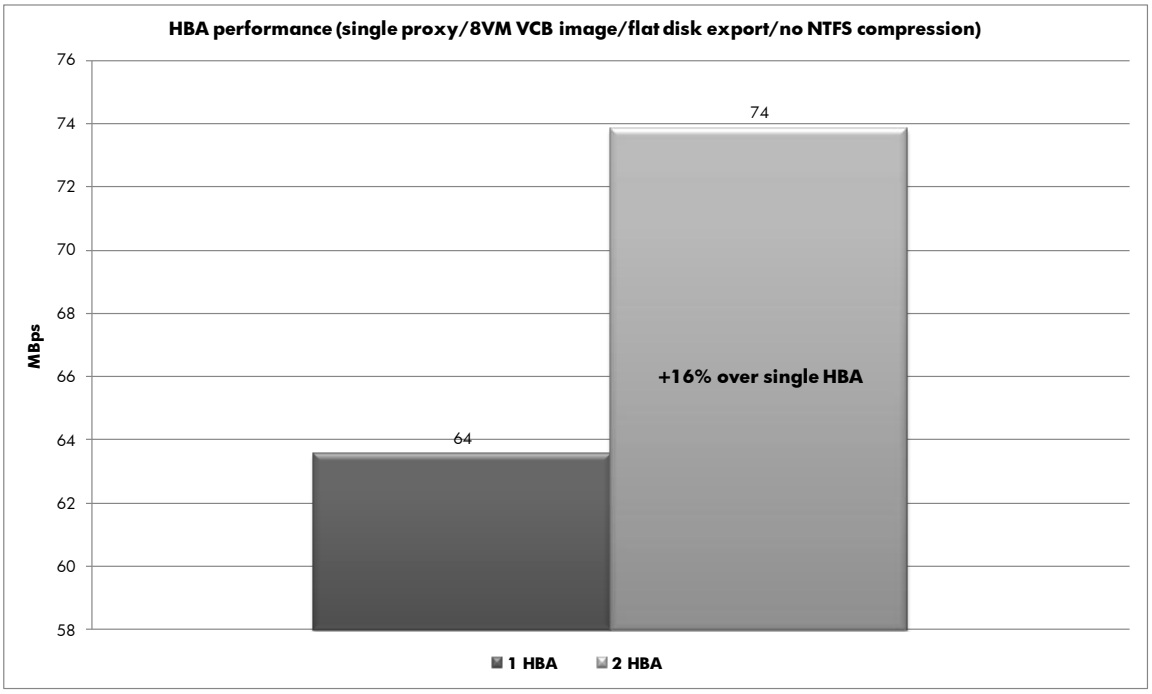


Figure 17 shows the following test results:

- Testing is performed on the disk export stage only to the VCB proxy HT volume (applicable with VCB image and ZDB D2D2T operations only).
- Both one and two HBA configurations are tested.
- All tests are configured with Vraid1 NTFS proxy volumes.
- Each configuration is tested with eight concurrent VCB image snapshots.
- Testing reveals that VCB proxy HBA queue depths never exceed 30 in either test scenario.
- Testing results in a significant increase (16%) with the addition of a properly zoned second HBA adapter.
- Isolating concurrent read and write activity on separate adapters achieves the best performance.

VCB proxy CPU cost

Figure 18 demonstrates the impact on proxy CPU resources with NTFS file compression. This test clearly illustrates the demand that NTFS compression places on CPU resources on the VCB proxy server and the need for ample CPU overhead for the best performance.

Figure 18. VCB NTFS CPU compression cost

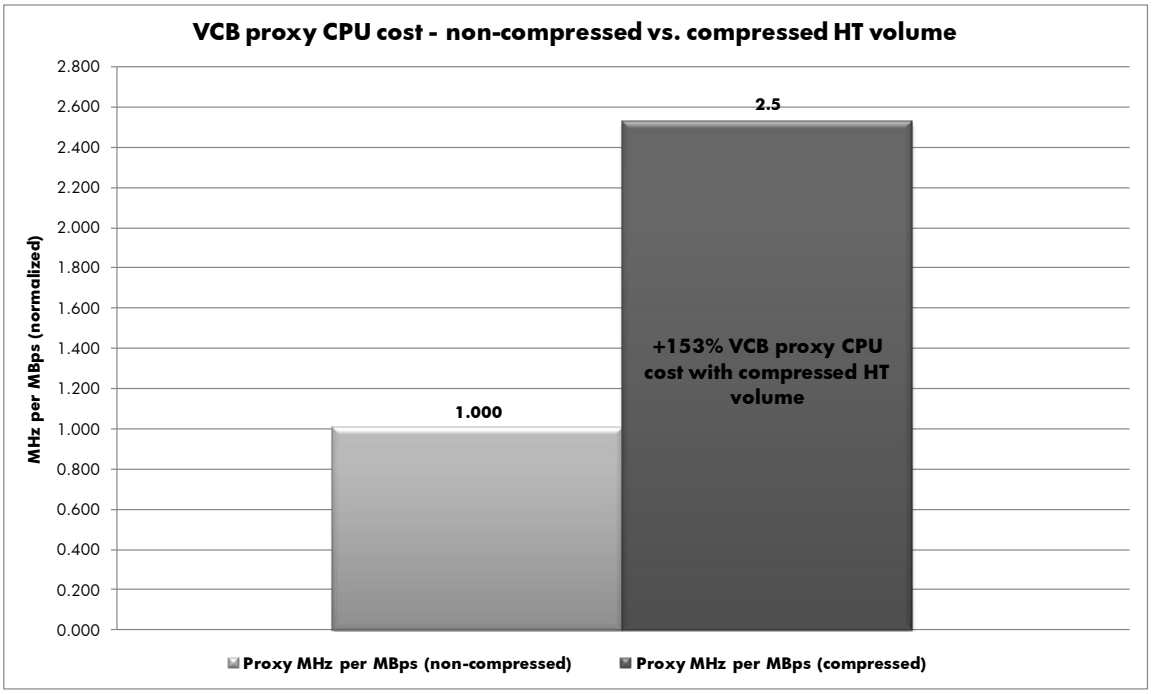


Figure 18 shows the following test results:

- The tests represent the disk export stage only to the VCB proxy HT volume (applicable with VCB image and ZDB D2D2T operations only).
- The test is configured with eight concurrent VCB image exports in a two VCB proxy configuration (four VCB image exports per proxy).
- The noncompressed proxy volume serves as a normalized baseline of comparison.
- The VCB proxy realizes a 153% increase in CPU cost per MB/s transferred (based on the raw transfer size and before compression) when using NTFS compression.
- The VCB proxy CPU resources become saturated and disk export throughput begins to degrade (four sockets at 3 GHz) beyond four concurrent streams per proxy.

Best Practice

Before implementing NTFS compression on VCB proxy resources, administrators must make sure that CPU resources are adequately sized or I/O throughput performance can degrade. Additionally, NTFS compressed HT volumes provide the best performance with two or more concurrent streams and up to four in the referenced configuration.

Data Protector COS overhead

Data Protector moderately impacts ESX host resources when implemented on the COS. Many VMware HA clustered solutions can readily absorb the recorded overhead due to being overprovisioned for failover operations. With the sheer performance and efficiencies offered by Data Protector snapshot, administrators should consider this solution to be a highly viable option. For more information, see Figure 19.

Figure 19. COS overhead with Data Protector snapshot

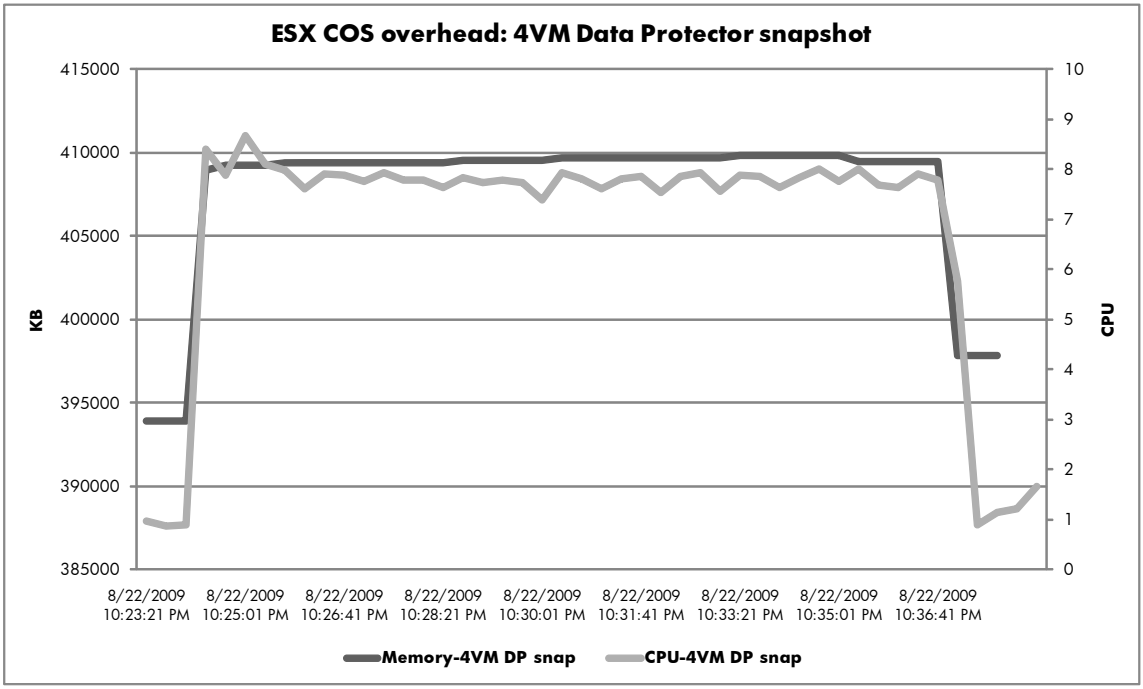


Figure 19 shows the following test results:

- CPU and memory overhead with four concurrent Data Protector/ESX COS snapshot streams on a single ESX host
- A direct write-to-media operation with Data Protector snapshot method
- A minimal 4% increase in average ESX host memory usage
- A minimal 8% increase in average ESX host CPU usage

Backup and restore performance testing

VCB image performance

Figure 20 provides a comparison of a complete VCB image cycle (including backup media write operations). This test illustrates the advantages achieved through backup concurrency, dual proxy, and NTFS compression with VCB image operations.

Figure 20. VCB image performance comparison

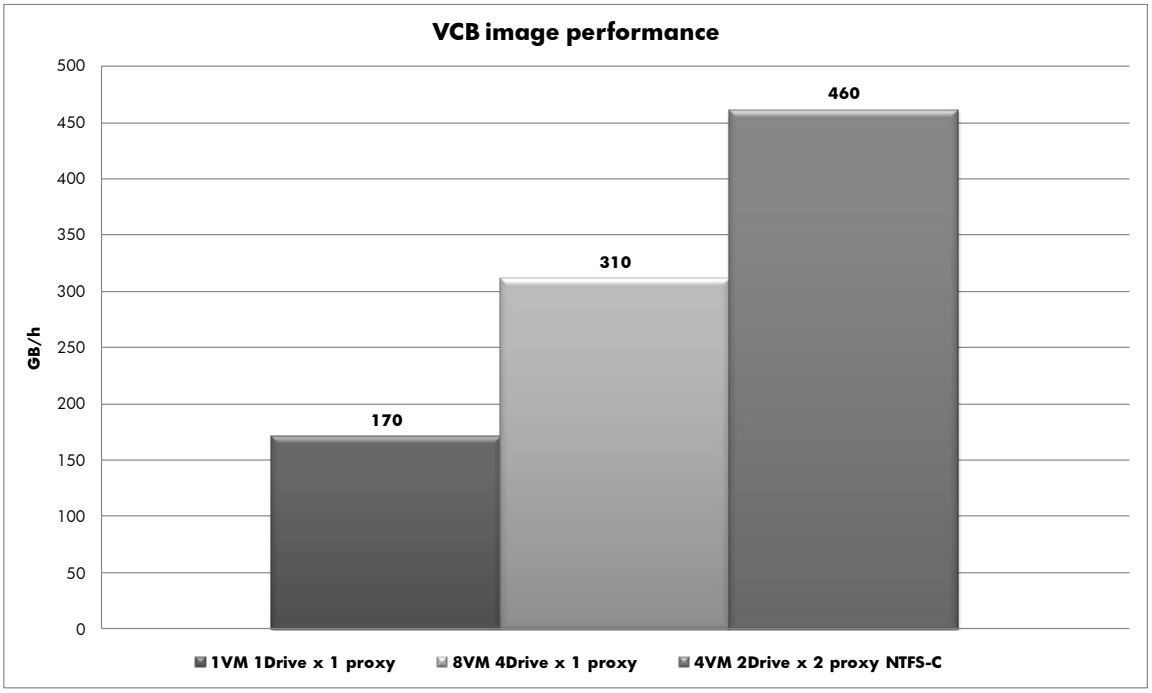


Figure 20 shows the following test results:

- Testing records VCB image performance through write-to-media operations (VTL).
- Testing is configured with one and eight concurrent VCB image operations.
- Both one and two VCB proxy configurations are tested.
- All tests are configured with Vraid0 NTFS proxy volumes for the best possible performance.
- Both noncompressed and compressed NTFS volumes are tested (only with concurrent streams).
- Multiplexing ratios (VM snapshot to media drive) perform best at 2:1 in all test scenarios.
- VCB image in the referenced dual proxy configuration nearly achieves ½ TB/h performance.

Data Protector snapshot performance

Figure 21 shows the performance scaling potential with three VTL nodes presented to three ESX COS installations while running parallel backup jobs. Clearly, parallel backup operations and concurrency when combined with multiple VTL nodes offer administrators a backup solution limited only by the backup media's capabilities.

Figure 21. Data Protector snapshot performance on the COS

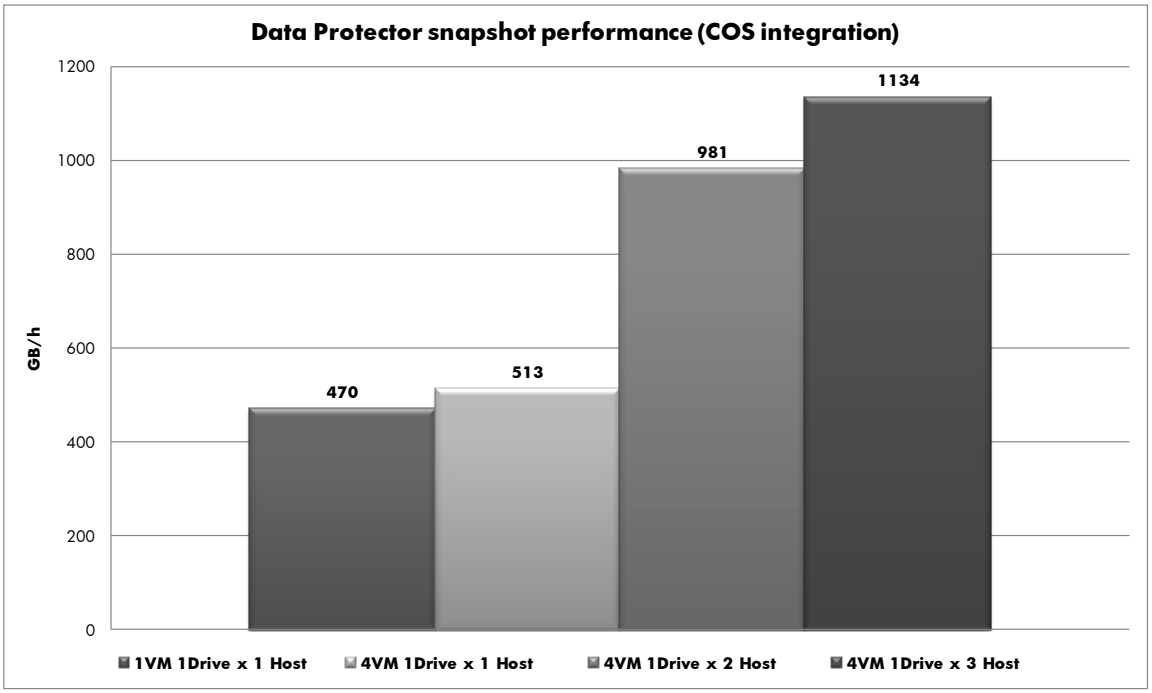


Figure 21 shows the following test results:

- Data Protector snapshot performance (with immediate write-to-media).
- One Data Protector snapshot stream on a single ESX host.
- Four concurrent Data Protector snapshot streams on a single ESX host.
- Eight concurrent Data Protector snapshot streams configured on two ESX hosts (four snapshot streams on each host).
- Twelve concurrent Data Protector snapshot streams configured on three ESX hosts (four snapshot streams on each host).
- Multiplexing ratios (VM snapshot to media drive) 1:1 and 4:1.
- 1VM records a 176% performance increase compared with the 1VM VCB image stream shown in Figure 20.
- 8VM records a 113% performance increase compared with the dual proxy 8VM VCB image stream shown in Figure 20.
- 12VM records a 147% performance increase compared with the dual proxy 8VM VCB image stream shown in Figure 20.

VCB file performance

VCB file performs well in the test configuration as shown in Figure 22. Because of the efficiency of VCB file with direct-to-media operations, VCB file offers administrators considerable value over other file backup options.

Figure 22. VCB file backup performance

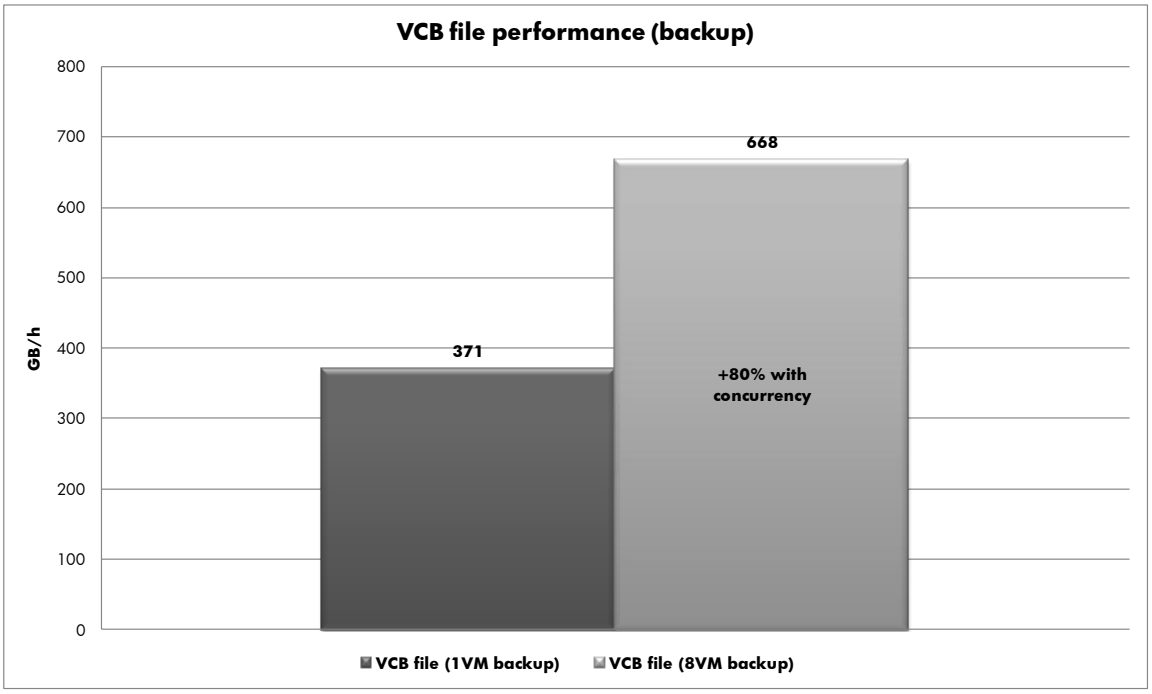


Figure 22 shows the following test results:

- Testing records VCB file performance through write-to-media operations (VTL).
- Testing is configured with one and eight concurrent VCB file operations (10 GB each).
- Only a one VCB proxy configuration is tested.
- Multiplexing ratios (file snapshot to media drive) tested are 1:1 and 4:1.
- The eight stream VCB file operation consumes the entire 2 Gb FC host interconnect of the VTL backup medium.
- VCB file performance scales out and improves similar to Data Protector snapshot performance with concurrent streams balanced across a two VCB proxy/VTL solution (with separate FC host interconnects).

VCB image restore performance

Figure 23 shows the image restore performance over the LAN. Administrators should note the effects that multiplexing has on restore operations.

Figure 23. VCB image performance (restore operations)

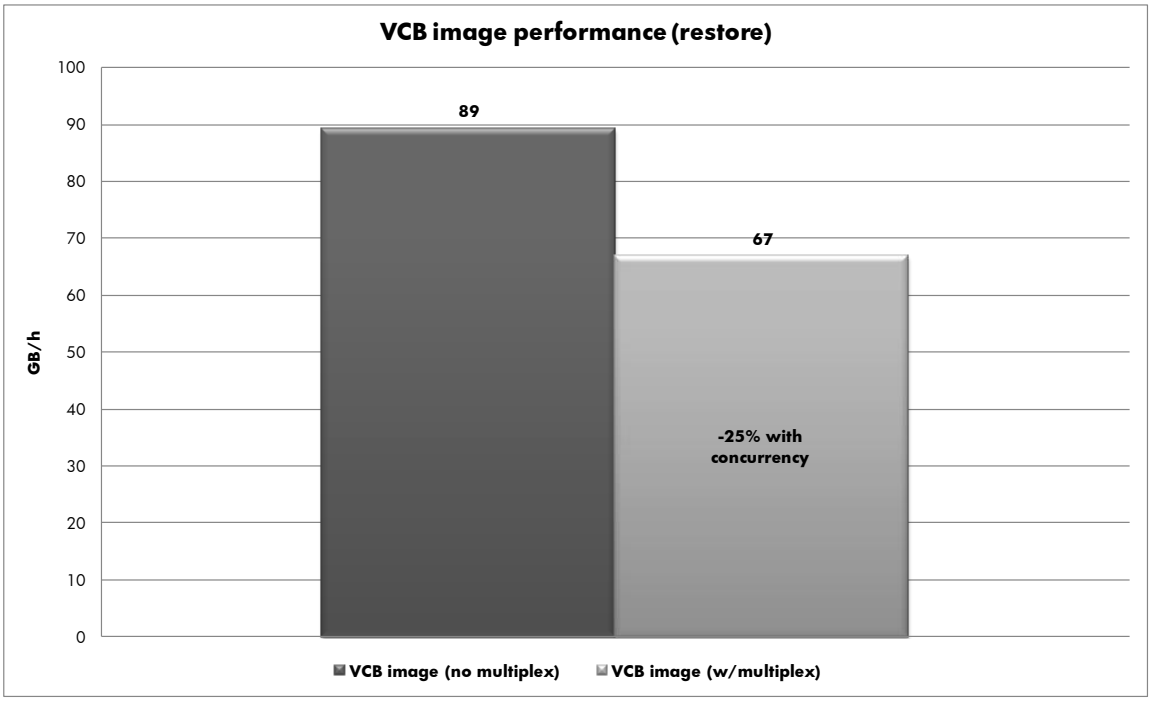


Figure 23 shows the following test results:

- Testing records VCB image restore performance (from VTL device and over the LAN).
- Testing records multiplexing ratios of 1:1 and 4:1.
- VM image restore time (throughput measured in GB/h) is reduced by 25% with a multiplexed backup.

Note

Data Protector snapshot restorations are not recorded, but are expected to be similar to those recorded in Figure 23. Administrators must recognize that image restorations place a significant demand on ESX host resources, primarily because of the Ethernet processing requirements.

VCB file restore performance

Figure 24 shows the effects of multiplexing on VCB file operations. Administrators should note the effects that multiplexing has on restore operations.

Figure 24. VCB file performance (restore operations)

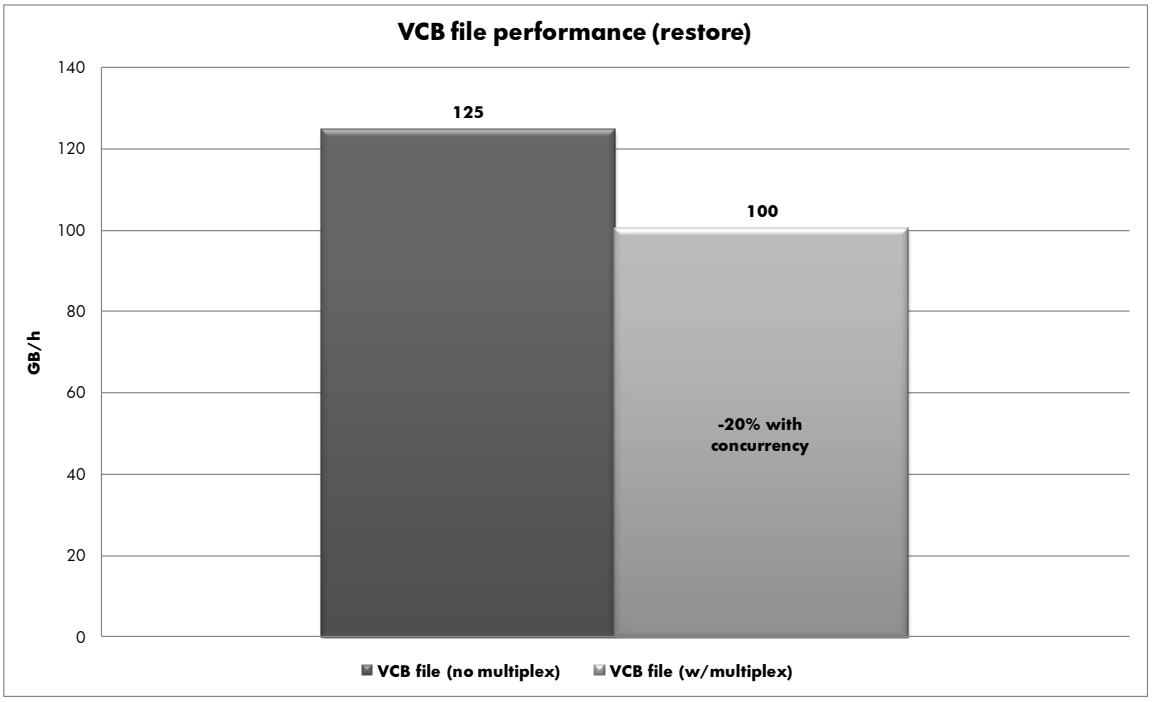


Figure 24 shows the following test results:

- Testing records VCB file restore performance (from VTL and over the LAN).
- Testing records multiplexing ratios of 1:1 and 4:1.
- VM file restore time (throughput measured in GB/h) is reduced by 20% with multiplexed backups.

Note

As seen with image restorations, multiplexing affects restoration times with file backups, although to lesser extent. Additionally, as is the case of image restorations, administrators must be aware of the overhead required for file restoration. However, in the case of VCB file, this can be managed by dedicating a system as a restore node, offloading this activity from the VM and ESX host resources.

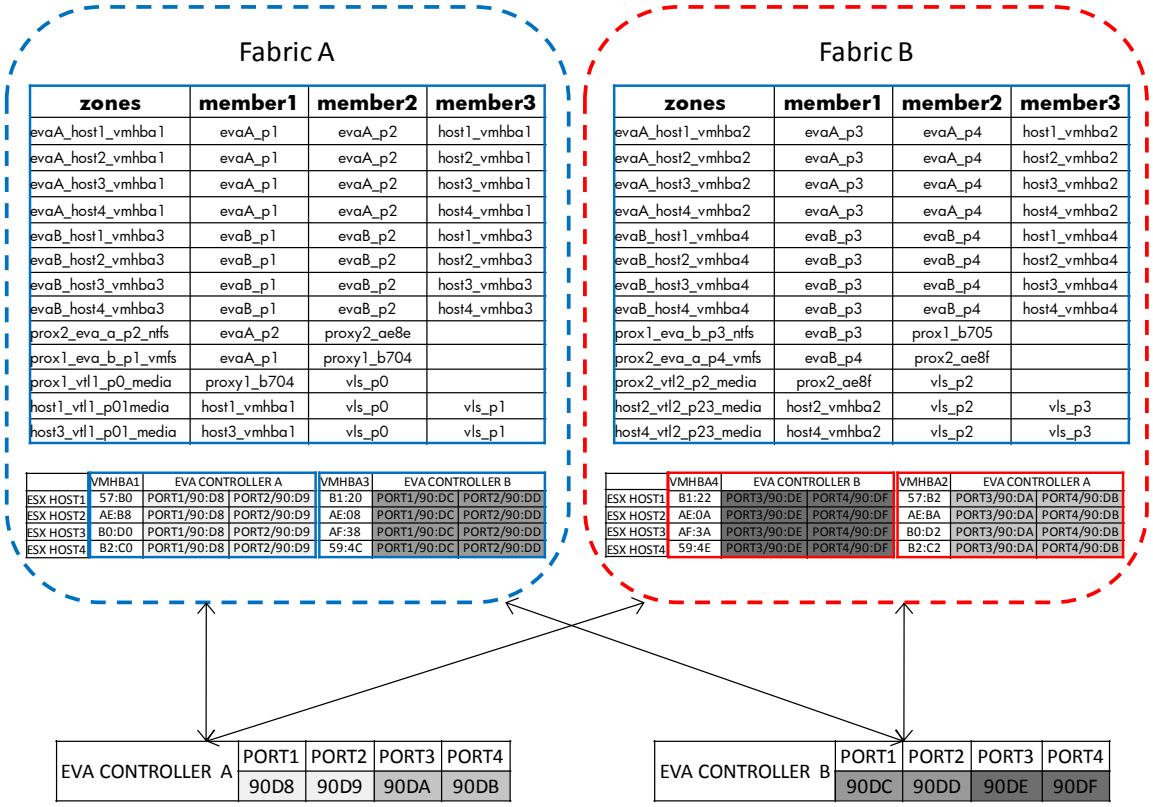
Referenced configuration

Figure 25 through Figure 28 outline in detail the test environment and illustrate many of the best practice recommendations throughout this white paper.

Zoning

Data transfer rates can be enhanced by reducing collision domains (fabric zoning) and isolating disk resources (array masking and mapping) on the SAN. VCB backup operations are heavy SAN I/O resource consumers. Testing reveals that a properly zoned VCB environment is the difference between a poor performing and a finely tuned backup solution. For an illustration of the test environment zoning implementation, see Figure 25.

Figure 25. Test environment zoning



zones	member1	member2	member3
evaA_host1_vmhba1	evaA_p1	evaA_p2	host1_vmhba1
evaA_host2_vmhba1	evaA_p1	evaA_p2	host2_vmhba1
evaA_host3_vmhba1	evaA_p1	evaA_p2	host3_vmhba1
evaA_host4_vmhba1	evaA_p1	evaA_p2	host4_vmhba1
evaB_host1_vmhba3	evaB_p1	evaB_p2	host1_vmhba3
evaB_host2_vmhba3	evaB_p1	evaB_p2	host2_vmhba3
evaB_host3_vmhba3	evaB_p1	evaB_p2	host3_vmhba3
evaB_host4_vmhba3	evaB_p1	evaB_p2	host4_vmhba3
prox2_eva_a_p2_ntfs	evaA_p2	proxy2_ae8e	
prox1_eva_b_p1_vmfs	evaA_p1	proxy1_b704	
prox1_vtl1_p0_media	proxy1_b704	vls_p0	
host1_vtl1_p01media	host1_vmhba1	vls_p0	vls_p1
host3_vtl1_p01_media	host3_vmhba1	vls_p0	vls_p1

	VMHBA1	EVA CONTROLLER A		VMHBA3	EVA CONTROLLER B	
ESX HOST1	57:B0	PORT1/90:D8	PORT2/90:D9	B1:20	PORT1/90:DC	PORT2/90:DD
ESX HOST2	AE:B8	PORT1/90:D8	PORT2/90:D9	AE:08	PORT1/90:DC	PORT2/90:DD
ESX HOST3	B0:D0	PORT1/90:D8	PORT2/90:D9	AF:38	PORT1/90:DC	PORT2/90:DD
ESX HOST4	B2:C0	PORT1/90:D8	PORT2/90:D9	59:4C	PORT1/90:DC	PORT2/90:DD

zones	member1	member2	member3
evaA_host1_vmhba2	evaA_p3	evaA_p4	host1_vmhba2
evaA_host2_vmhba2	evaA_p3	evaA_p4	host2_vmhba2
evaA_host3_vmhba2	evaA_p3	evaA_p4	host3_vmhba2
evaA_host4_vmhba2	evaA_p3	evaA_p4	host4_vmhba2
evaB_host1_vmhba4	evaB_p3	evaB_p4	host1_vmhba4
evaB_host2_vmhba4	evaB_p3	evaB_p4	host2_vmhba4
evaB_host3_vmhba4	evaB_p3	evaB_p4	host3_vmhba4
evaB_host4_vmhba4	evaB_p3	evaB_p4	host4_vmhba4
prox1_eva_b_p3_ntfs	evaB_p3	prox1_b705	
prox2_eva_a_p4_vmfs	evaB_p4	prox2_ae8f	
prox2_vtl2_p2_media	prox2_ae8f	vls_p2	
host2_vtl2_p23_media	host2_vmhba2	vls_p2	vls_p3
host4_vtl2_p23_media	host4_vmhba2	vls_p2	vls_p3

	VMHBA4	EVA CONTROLLER B		VMHBA2	EVA CONTROLLER A	
ESX HOST1	B1:22	PORT3/90:DE	PORT4/90:DF	57:B2	PORT3/90:DA	PORT4/90:DB
ESX HOST2	AE:0A	PORT3/90:DE	PORT4/90:DF	AE:BA	PORT3/90:DA	PORT4/90:DB
ESX HOST3	AF:3A	PORT3/90:DE	PORT4/90:DF	B0:D2	PORT3/90:DA	PORT4/90:DB
ESX HOST4	59:4E	PORT3/90:DE	PORT4/90:DF	B2:C2	PORT3/90:DA	PORT4/90:DB

The VCB referenced configuration in Figure 25 illustrates the following best practices:

- Careful SAN mapping/zoning results in balanced resource usage
  (esx_host <--> esx_host_vmhba# <--> fc_switches <--> eva_controllers <--> controller_host_ports)
- Solution redundancy across fabrics
- Solution redundancy across storage controllers
- Solution redundancy across storage controllers host ports
- Solution redundancy across ESX storage adapters (vmhbas)

- Solution bottlenecks reduced or eliminated
- FC collision domains reduced or eliminated with initiator-to-target zoning practices
- Improved performance through isolation of backup traffic (controller A storage ports 1 and 2, and controller B storage ports 3 and 4) and VM production traffic (controller A storage ports 3 and 4, and controller B storage ports 1 and 2) on controller resources

Controller host ports

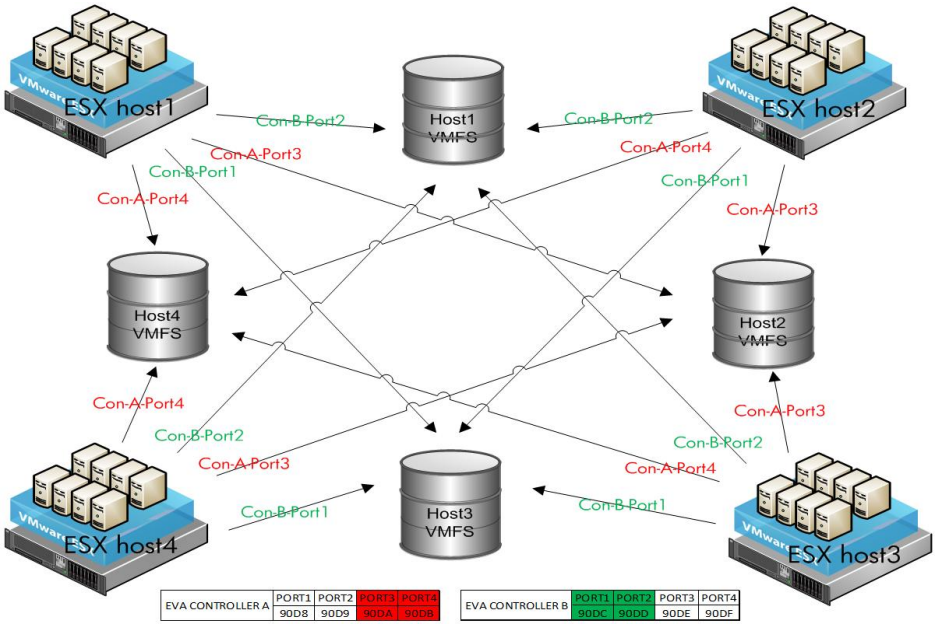
Figure 26 shows a color-coded ESX host-to-storage port mapping. Configuring preferred path mappings in this way provides the following performance benefits:

- Assignment of dedicated storage port resources per clustered VMFS volume
- Segregation of production (ports 3 and 4 on controller A and ports 1 and 2 on controller B) and backup workloads (ports 1 and 2 on controller A and ports 3 and 4 on controller B) on controller storage ports

Prior to ESX 4.0, which introduced Asymmetrical Logical Unit Access (ALUA) support and supported multipath options, controller host port management was essential with all VMware implementations. For storage systems that are not ALUA compliant, it might still be necessary to actively manage controller host ports with ESX 4.0.

The HP test environment is limited to VCB with ESX 3.5U4. The value achieved through efficient controller host port management cannot be overstated in the test environment. The default ESX behavior is to present all volumes to the first storage adapter. If left unmodified, these settings might have both VCB backup and production VM workloads processed through a single controller host port. Administrators can realize significant I/O efficiencies on the backend storage array by segregating VCB backup and VM production path workloads. This is accomplished in conjunction with the volume mapping and zoning recommendations described in Figure 25, but in addition to the preferred path settings on ESX 3.5 and earlier. For an example of preferred path settings in the test environment, see Figure 26.

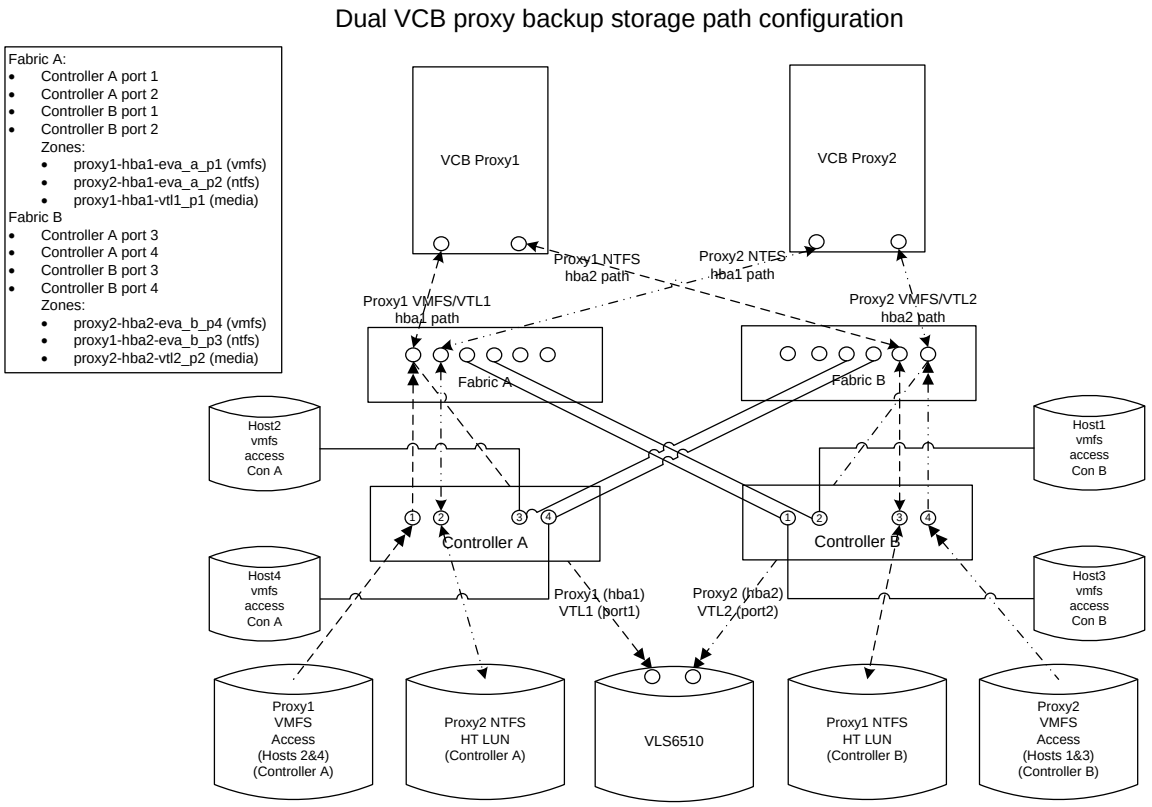
Figure 26. ESX preferred storage paths



Note

While ESX 4.0 with multipath configuration settings (RR) has not been tested, administrators should carefully monitor performance with this recently supported option, especially in mixed workload environments (small block random and large bock sequential). It might continue to be a best practice to zone storage ports based on workload requirements with ESX 4.0. For more information, see *Configuration best practices for HP StorageWorks Enterprise Virtual Array (EVA) family and VMware vSphere 4*.

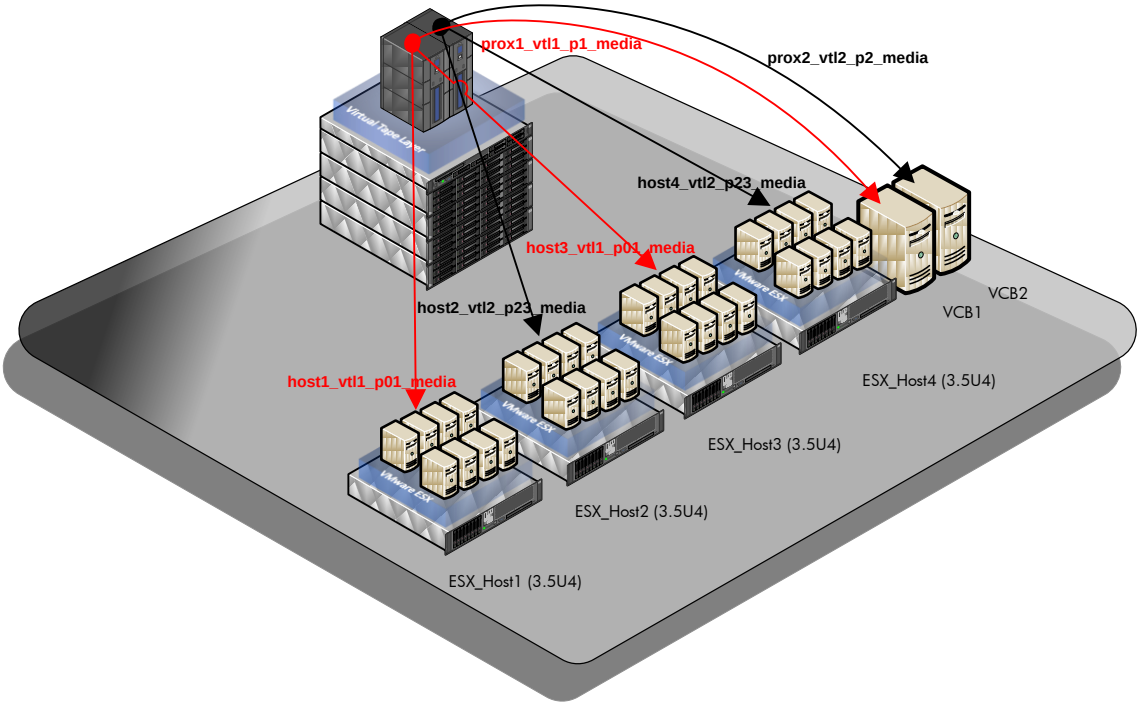
Figure 27. Referenced dual VCB proxy configuration



On close examination, Figure 27 shows many of the recommendations discussed throughout this white paper. For ESX datacenters with sizable amounts of data to be backed up with VCB image (because of the disk export operations), the referenced configuration can serve as a roadmap for administrators to achieve the best possible solution performance.

Figure 28 shows how a VLS device can be carved into two separate VTL nodes and shared across ESX COS and VCB proxy installations.

Figure 28. Two VTL node ESX/VCB presentation



In the configuration shown in Figure 28, two VTL nodes (with independent 4 Gb/s host port interconnects) running in parallel with eight total backup streams fill the available disk bandwidth (1.4 TB/h) of a single 48 disk VLS node. With the addition of more VLS nodes, performance can scale out as datacenter backup requirements dictate.

Implementing a proof-of-concept

As a best practice for all deployments, HP recommends that you implement a proof-of-concept by using a test environment that matches the planned production environment as closely as possible. In this way, you can obtain appropriate performance and scalability characterizations. For help with a proof-of-concept, contact an HP Services representative (http://www.hp.com/hps/contacts/index.html) or your HP partner.

Summary

This white paper outlines many planning, implementation, and management considerations when implementing Data Protector with VMware ESX. Strong emphasis is placed on effectively implementing a VMware solution that matches datacenter data types with the wide selection of Data Protector backup options. Administrators have the following backup choices with Data Protector.

VCB image

Seamlessly integrated with vCenter server, administrators can select, schedule, backup, and restore all VMs within the datacenter without any scripting or additional plug-in requirements. Data Protector VCB image provides the following solution benefits:

- All backup and restore management through a single user interface
- Seamless integration with all datacenter VM operating systems
- Seamless VM restores with both COS and COS-less (with V6.11 only) ESX installations

VCB file

Best Practice
VCB file is the preferred file backup solution with VMware ESX and Data Protector.

Data Protector integration with VCB file provides backup administrators with a scalable and efficient file-level backup solution and provides the following high-level solution benefits:

- Direct and efficient write-to-media operations
- Low storage resource usage
- All backup and restore management through a single user interface
- Seamless integration with all datacenter VM file stores (Windows only)
- Seamless file restores with Data Protector disk agent installation
- Scalable performance with file snapshot concurrency

Data Protector snapshot

Best Practice
Data Protector snapshot is the preferred image solution with VMware ESX.

Data Protector snapshot integration with VMware ESX is a powerful alternative to VCB image, providing administrators with a scalable and efficient image backup solution when implemented on the COS. Administrators can expect the following high-level solution benefits:

- Direct and efficient write-to-media operations
- Low storage resource usage
- All backup and restore management through a single user interface
- Full, incremental, differential, and mixed-mode snapshots available
- Seamless VM restores with both COS and COS-less (with V6.11 only) ESX installations
- SAN-attached VTL option available on the ESX host
- Highly scalable solution for large datacenter installations

Best Practice
Data Protector ZDB/IR is the preferred application solution with VMware ESX.

Data Protector ZDB/IR provides administrators with a highly effective VM application-consistent backup and restore solution and provides the following high-level solution benefits:

- Integrated with the EVA storage array's snapclone functionality, providing near instantaneous database restores with IR
- Integrated with the EVA storage array's snapshot functionality, providing administrators with an excellent choice for offloaded Disk-to-Disk-to-Tape (D2D2T), data mining, and test replicas
- ESX host, VM, and application resources offloaded from the backup process
- All backup and restore management through a single user interface

In addition to these options, this white paper covers a variety of components associated with virtualized backup environments. Extensive testing with HP Data Protector clearly demonstrates backup flexibility, scalability, backup performance, and ease of use; and should be considered a market-leading backup software solution when paired with VMware ESX.

Appendix A

The storage system disk IOPS rate is critical beyond VCB backup operations and should be sized first and foremost to adequately service production VM loads. Generally speaking, if production VM IOPS is being serviced in a timely manner by the backend storage system, backup loads also perform adequately provided both loads are not in contention for the same IOPS resources. For this reason, administrators who are planning to eliminate backup windows must take into account IOPS sizing of both production and VCB backup loads, ensuring ample storage overhead is available in the datacenter.

It is critical that administrators approximate the maximum number of VMs per clustered volume with VMware ESX. The ratio of VMs to clustered volume is driven by storage and ESX host constraints. Storage sizing for any number of VMs is a very straightforward exercise (number of VMs x each VM disk resource requirements = total storage) and is not the primary consideration in this context. Instead, administrators must know the number of the following:

- VMs that can coexist per clustered volume shared across many ESX hosts
- VMs that can coexist per ESX host volume

To determine both limits, administrators must approximate or determine the values described in the following sections.

Average active SCSI command rate per VM (a)

This value identifies the workload demand per VM in the environment. Administrators with live or test environments can capture average active SCSI command rates with the ESX host command-line utilities ESXTop (local console) or RESXTop (remote interface). These utilities allow for monitoring, capturing, and measuring VM active SCSI command rates. For information about these utilities and their use, see the *vSphere Resource Management Guide*.

Maximum outstanding storage array queue depth (n)

The EVA in the test environment supports a maximum outstanding queue depth of 1,536 (6,144 per controller, 12,288 per array) per storage port before returning "queue full" to an initiator. However, administrators are cautioned against using these maximum values for the following reasons:

- Adequate servicing of queue depths is too dependent on the VM workload (application IOPS and latency requirements) and the quality of disk resources (disk protocol/speed and RAID) for such a generalized value.
- It is always a best practice to size the storage port queue depths for a single controller, yet balance those values across both controllers (in the event of controller failure).
- Ongoing controller maintenance (leveling and rebuilding) consumes disk and controller resources, reducing the queue depth servicing rates.
- Disk I/O consumes queue depth resources for both data transfer (block data) and SCSI commands (instruction sets) and must be factored accordingly.

The variables involved with VM workload and array disk resources are complex and unique to each environment. For this reason, a one-size-fits-all queue depth value cannot be universally applied with VMware ESX and the EVA storage array. Generally speaking, it is always a best practice to first ensure that disk resources are adequately servicing the VM I/O requests through comprehensive preproduction testing before deployment. In addition to determining the precise datacenter IOPS measurements and ensuring that the application latency requirements are met, administrators should always factor in potential controller failure events and never size the queue depths to the maximum values of both units. In the event of controller failure, such sizing practices might potentially render

production VMs with latency-sensitive applications in a degraded state. Nevertheless, administrators can closely approximate these values when initially sizing storage resources.

Note
Appendix B provides four IOPS and maximum storage port queue depth use cases with the tested EVA storage array.

ESX host queue depth (d)

The ESX server defaults to a storage adapter queue depth of 32. Additionally, VMware recommends that this setting be no higher than 64 or 128 if an ESX host has exclusive volume access. Administrators must keep in mind that apart from the ESX host storage adapter setting is the ESX host HBA adapter queue depth. Typically, most vendors default to 32, matching the ESX host storage adapter setting. However, it is still a best practice to verify and align both values on the ESX host for the best performance.

Note
The advanced ESX host setting `Disk.SchedNumReqOutstanding` can be used to throttle queue depths when many VMs are sharing a clustered volume resource. Typically, this advanced setting is set to the same value as the storage adapter and host HBA setting.

After you have determined the preceding values, you can use the following calculations to determine the one-to-many and one-to-one volume-to-esx-host VM threshold levels:

$$VMs\ per\ ESX\ clustered\ volume\ (m) = \frac{storage\ array\ queue\ depth\ (n)}{average\ active\ VM\ SCSI\ command\ rate\ (a)}$$

$$VMs\ per\ ESX\ host\ volume\ (m) = \frac{ESX\ host\ queue\ depth\ (d)}{average\ active\ VM\ SCSI\ command\ rate\ (a)}$$

It is important to note the following with these calculations:

- For the maximum VMs per ESX clustered volume, both the preferred storage port (single port queue depth only) and the multipath solutions can be sized. If a multipath solution is available (ESX 4.x), then the calculated values are a multiple of the participating controller host port maximums in the multipath configuration. For the EVA, ALUA compliance allows for sizing on the active controller (volume-owning resource) host ports only.
- For the maximum VMs per ESX host on a volume resource, the calculations assume a single ESX host storage adapter. However, an additional ESX host storage adapter can be configured on a separate volume resource, thereby increasing the number of hosted VMs per ESX server.

Appendix B

Note
The tables in this appendix reduce the per port maximums by approximately 25% to account for disk array normalization, disk rebuilding, and SCSI command set overhead.

Figure 29 provides the IOPS calculations and recommended controller port maximums based on specific prerequisites. Per port maximums are based on disk latency requirements of less than 20 ms. Be extremely careful when using these recommendations outside of these specifications and without preproduction testing.

Figure 29. 168 FC 15K disk group

Scenario 1: Small block random workload (8k transfer)			
Disk group spindle count	168		
Disk protocol	FC 15k		
Protection level	Vraid5		
Leveling active	no		
Latency requirements	<20ms		
typical workloads			
IOPS (0% reads)	9114	per disk IOPS	54
IOPS (60% reads)	19457	per disk IOPS	116
IOPS (100% reads)	47040	per disk IOPS	280
per port queue recommendations			
port maximum	46	< 25% (overhead)	34
port maximum	97	< 25% (overhead)	73
port maximum	235	< 25% (overhead)	176

Scenario 2: Small block random workload (8k transfer)			
Disk group spindle count	168		
Disk protocol	FC 15k		
Protection level	Vraid1		
Leveling active	no		
Latency requirements	<20ms		
typical workloads			
IOPS (0% reads)	18228	per disk IOPS	109
IOPS (60% reads)	30576	per disk IOPS	182
IOPS (100% reads)	47040	per disk IOPS	280
per port queue recommendations			
port maximum	91	< 25% (overhead)	68
port maximum	153	< 25% (overhead)	115
port maximum	235	< 25% (overhead)	176

Scenario 3: Large block sequential workload (128k transfer)			
Disk group spindle count	168		
Disk protocol	FC 15k		
Protection level	Vraid5		
Leveling active	no		
Latency requirements	<20ms		
typical workloads			
IOPS (0% reads)	5838	per disk IOPS	35
IOPS (60% reads)	12218	per disk IOPS	73
IOPS (100% reads)	29232	per disk IOPS	174
per port queue recommendations			
port maximum	29	< 25% (overhead)	22
port maximum	61	< 25% (overhead)	46
port maximum	146	< 25% (overhead)	110

Scenario 4: Large block sequential workload (128k transfer)			
Disk group spindle count	168		
Disk protocol	FC 15k		
Protection level	Vraid1		
Leveling active	no		
Latency requirements	<20ms		
typical workloads			
IOPS (0% reads)	11676	per disk IOPS	70
IOPS (60% reads)	19200	per disk IOPS	114
IOPS (100% reads)	29232	per disk IOPS	174
per port queue recommendations			
port maximum	58	< 25% (overhead)	44
port maximum	96	< 25% (overhead)	72
port maximum	146	< 25% (overhead)	110

Figure 30 provides the IOPS calculations and recommended controller port maximums based on the tested storage resources in both Vraid5 and Vraid1 scenarios. Figure 30 also shows an estimate of the number of VMs supported per EVA controller port on a clustered VMFS volume. An average active VM SCSI command rate of four and a disk latency of less than 20 ms is used for illustration purposes. For more information, see Appendix A. Be extremely careful when using these recommendations outside of these specifications and without preproduction testing.

Figure 30. 144 FC 10K disk group

Scenario 1: Small block random workload (8k transfer)				Scenario 2: Small block random workload (8k transfer)			
Disk group spindle count	144			Disk group spindle count	144		
Disk protocol	FC 10k			Disk protocol	FC 10k		
Protection level	Vraid5			Protection level	Vraid1		
Leveling active	no			Leveling active	no		
Latency requirements	<20ms			Latency requirements	<20ms		
typical workloads				typical workloads			
IOPS (0% reads)	5468	per disk IOPS	38	IOPS (0% reads)	10937	per disk IOPS	76
IOPS (60% reads)	11674	per disk IOPS	81	IOPS (60% reads)	18346	per disk IOPS	127
IOPS (100% reads)	28224	per disk IOPS	196	IOPS (100% reads)	28224	per disk IOPS	196
per port queue recommendations				per port queue recommendations			
port maximum	27	< 25% (overhead)	21	port maximum	55	< 25% (overhead)	41
port maximum	58	< 25% (overhead)	44	port maximum	92	< 25% (overhead)	69
port maximum	141	< 25% (overhead)	106	port maximum	141	< 25% (overhead)	106
Scenario 3: Large block sequential workload (128k transfer)				Scenario 4: Large block sequential workload (128k transfer)			
Disk group spindle count	144			Disk group spindle count	144		
Disk protocol	FC 10k			Disk protocol	FC 10k		
Protection level	Vraid5			Protection level	Vraid1		
Leveling active	no			Leveling active	no		
Latency requirements	<20ms			Latency requirements	<20ms		
typical workloads				typical workloads			
IOPS (0% reads)	3503	per disk IOPS	24	IOPS (0% reads) (88% for calculation)	7006	per disk IOPS	49
IOPS (60% reads)	7331	per disk IOPS	51	IOPS (60% reads)	11520	per disk IOPS	80
IOPS (100% reads)	17539	per disk IOPS	122	IOPS (100% reads) (12% for calculation)	17539	per disk IOPS	122
per port queue recommendations				per port queue recommendations			
port maximum	18	< 25% (overhead)	13	port maximum	35	< 25% (overhead)	26
port maximum	37	< 25% (overhead)	27	port maximum	58	< 25% (overhead)	43
port maximum	88	< 25% (overhead)	66	port maximum	88	< 25% (overhead)	66

- calculations based on <20ms application latency requirements
- supported disk IOPS (large block sequential) = (disk IOPS*read workload)+(disk IOPS*write workload) For example: 57=(122*12%)+(49*88%)
- controller port queue max = (IOPS/(latency))*overhead. For example: 175=(11674/(1000ms/20ms))*75%

Note

As a general rule of thumb, subtract approximately 30% from 15K FC spindles when sizing 10K drives.

$$VMs\ per\ controller\ port\ on\ a\ ESX\ clustered\ volume\ (m) = \frac{storage\ array\ (per\ port)\ queue\ depth\ (n)}{average\ active\ VM\ SCSI\ command\ rate\ (a)}$$

$$44 = \frac{175}{4}$$

For more information

HP StorageWorks, http://www.hp.com/go/storageworks

HP Data Protector, http://www.hp.com/go/dataprotector

HP Virtualization with VMware, http://www.hp.com/go/vmware

VMware Storage Solutions from HP, http://www.hp.com/go/storage/vmware

Customer Focus Testing Solutions from HP,
http://h71028.www7.hp.com/enterprise/us/en/solutions/storage-customer-focused-testing.html

HP ActiveAnswers for VMware
http://h71019.www7.hp.com/ActiveAnswers/cache/71086-0-0-0-121.html

To help us improve our documents, please provide feedback at
http://h20219.www2.hp.com/ActiveAnswers/us/en/solutions/technical_tools_feedback.html.

Technology for better business outcomes

4AA0-5675ENW, February 2010