Technical Report

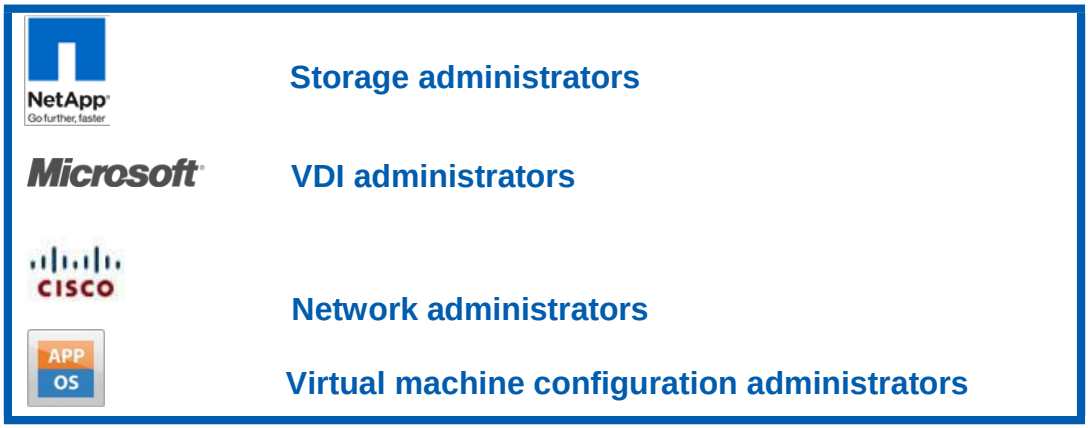
Best Practices for Microsoft Hyper-V with Citrix XenDesktop VDI on NetApp Storage

Rob Briggs, Pavel Lobanov, NetApp

February 2012 | TR-4042

TABLE OF CONTENTS

1 INTRODUCTION



1.1 EXECUTIVE SUMMARY

NetApp® technology enables companies to extend their virtual infrastructures to include the benefits of advanced storage virtualization. Our unified storage platforms offer industry-leading technologies in the areas of storage efficiencies, instantaneous virtual machine (VM) and data store cloning for virtual servers and virtual desktops, and virtual data center backup and business continuance solutions.

This technical report reviews the best practices for implementing a virtual desktop infrastructure (VDI) with Microsoft® Hyper-V™ and Citrix XenDesktop with NetApp unified storage arrays.

NetApp has been providing advanced storage features to Microsoft virtualization solutions since 2004. During this time, NetApp has developed operational guidelines for storage arrays running Data ONTAP® and Hyper-V servers. These techniques, which are described in this report, have been documented and are referred to as “best practices.”

1.2 IMPLEMENTING BEST PRACTICES

Unless stated otherwise, the recommendations and best practices presented in this document should be considered as deployment requirements. NetApp, Microsoft, and Citrix will still provide support even if you do not implement all of these best practices. However, disregarding any of these practices commonly results in the need to implement them at a later date, in a much larger environment, and often with application downtime. For these reasons, NetApp recommends implementing all of the best practices defined in this document as a part of your initial deployment or migration.

All recommendations in this document apply specifically to deploying Microsoft Hyper-V and Citrix XenDesktop on NetApp. Therefore this document supersedes all recommendations and best practices described in other NetApp documents.

NetApp and our partners offer professional services to architect and deploy the designs described in this document. These services can provide optimal virtual storage architecture for your virtual data center.

1.3 AUDIENCE

This best practice document is part of the NetApp Technical Library and is intended for use by individuals who are responsible for architecting, designing, managing, and supporting Microsoft virtual infrastructures. Readers should, at a minimum, be familiar with concepts pertaining to Microsoft Hyper-V, Citrix XenDesktop, and NetApp Data ONTAP 8.

The administrative roles required to implement the technology and/or configurations are presented at the beginning of each section.

1.4 DOCUMENT ROADMAP

This technical report is the main document in a set of NetApp documents covering virtualization on NetApp with Microsoft and Citrix products. This document specifically discusses Hyper-V and XenDesktop.

2 NETAPP STORAGE BEST PRACTICES

This section applies to:

Storage administrators

VDI administrators

2.1 AN INTRODUCTION TO STORAGE IN A VIRTUAL INFRASTRUCTURE

In a Hyper-V environment, the availability and performance of the shared storage infrastructure are more critical than those of the individual servers running the virtualized server environment. It is therefore vital to factor in the required level of availability and performance when selecting and designing the storage solution for the virtualized server environment. NetApp offers a comprehensive set of software and hardware solutions to address the most stringent requirements for availability and performance of large, scalable Hyper-V environments.

2.2 THE VALUE OF MULTIPROTOCOL STORAGE ARRAYS

The virtualization of a data center results in physical systems being virtualized as part of a cost-saving effort to reduce both capital expenditures and operational expenditures through infrastructure consolidation and increased operational efficiencies. These efforts result in multiple VMs sharing physical resources, including shared storage pools known as *datastores.* Virtualizing demanding, business-critical applications such as e-mail or database servers results in gains in operational efficiencies. Database Servers might share server resources but are typically configured with exclusive access to the storage it requires.

Both Microsoft and NetApp offer technologies that natively support multiple storage protocols. These technologies allow customers to deploy best-in-class virtual data centers that leverage the strengths inherent in using these technologies together. This report goes beyond comparing storage area network (SAN) with network-attached storage (NAS) to consider the operational value based on the type of storage network interconnect available to a virtual data center.

Whether your storage network is Fiber Channel (FC) or Ethernet (NFS, iSCSI, and FCoE), these technologies combine with NetApp storage to scale the largest consolidation efforts and to virtualize the most demanding applications without sacrifice or the need to deploy separate hardware to meet the needs of either environment. This is virtualization, and it's valuable in a storage array platform.

2.3 THE 80/20 RULE

When designing the storage architecture for a virtual data center, you can apply the 80/20 rule: 80% of all systems virtualized are for consolidation efforts. The remaining 20% of systems are classified as business-critical applications. Although these applications can be virtualized successfully, they tend to be deployed on shared storage pools but in isolated datasets.

THE CHARACTERISTICS OF CONSOLIDATION DATASETS

Consolidation datasets have the following characteristics:

- The VMs do not require application-specific backup and restore agents.
- The dataset is the largest in terms of the number of VMs and potentially the total amount of storage addressed.

- Individually, each VM might not address a large dataset or have demanding IOP requirements; however, the size of the collective whole might be considerable.
- These datasets are ideally served by large, shared, policy-driven storage pools (or datastores).

THE CHARACTERISTICS OF ISOLATED DATASETS (FOR BUSINESS-CRITICAL APPLICATIONS)

Isolated datasets have the following characteristics:

- The VMs require application-specific backup and restore agents.
- Each individual VM might address a large amount of storage and/or have high I/O requirements.
- Storage design and planning apply in the same way as with physical servers.
- These datasets are ideally served by individual, high-performing, nonshared datastores.

Unless your data center is globally unique, the evolution of your data center from physical to virtual will follow the 80/20 rule.

2.4 NETAPP STORAGE BEST PRACTICES

NetApp's scalable, unified storage and data management solution for VDI offers the following benefits:

- **Storage efficiency.** Significant cost savings with multiple levels of storage efficiency for all the virtual machine data components
- **Performance.** Enhanced user experience with transparent read and write I/O optimization that strongly complements NetApp's storage efficiency capabilities
- **Data protection.** Enhanced protection of both the virtual desktop OS data and the user data, with very low overhead for both cost and operations

SINGLE SCALABLE UNIFIED ARCHITECTURE

The NetApp Unified Storage Architecture gives customers an agile and scalable storage platform. NetApp's innovative storage solutions offer customers new alternatives and expanded possibilities over traditional storage vendors. All NetApp storage systems use the Data ONTAP operating system to provide SAN (FCoE, Fibre Channel, and iSCSI), NAS (CIFS and NFS), primary storage, and secondary storage in a single unified platform so that all virtual desktop data components can be hosted on the same storage array. A single process for activities such as installation, provisioning, mirroring, backup, and upgrading is used throughout the entire product line, from the entry level to enterprise-class controllers. Having a single set of software and processes brings great simplicity to even the most complex enterprise data management challenges.

Unifying storage and data management software and processes reduces the complexity of data ownership, enables companies to adapt to their changing business needs without interruption, and results in a dramatic reduction in total cost of ownership.

For large, scalable VDI environments, the NetApp solution offers the following benefits:

- At least 50% savings in storage, power, and cooling requirements
- Agile and operationally efficient storage solutions
- Best-in-class data protection and business continuance solutions to address any level of data availability demands

STORAGE EFFICIENCY

One critical barrier to VDI adoption is the increased cost of using shared storage to obtain a highly available enterprise-quality infrastructure. Virtual desktop deployment creates a high level of data redundancy, especially for the virtual machine OS data. Using traditional storage, this means that you need storage equal to the sum of the storage required by each VM. For example, if each VM is 20GB in

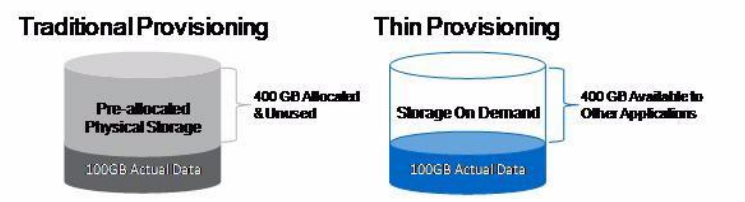
size and there are supposed to be 1,000 virtual machines in the solution, the solution would require at least 20TB of occupied hard drive space usable on the shared storage.

Thin provisioning, data deduplication, volume and LUN FlexClone® thin-cloning technology, and sub-LUN cloning are the critical components of the NetApp solution, offering multiple levels of storage efficiency across the virtual desktop OS data, installed applications, and user data. This helps customers to save 50% to 90% on the cost associated with shared storage (based on existing customer deployments and NetApp solutions lab validation). NetApp is the only storage vendor that offers block-level data deduplication for live virtual machines, without any negative trade-offs.

THIN PROVISIONING

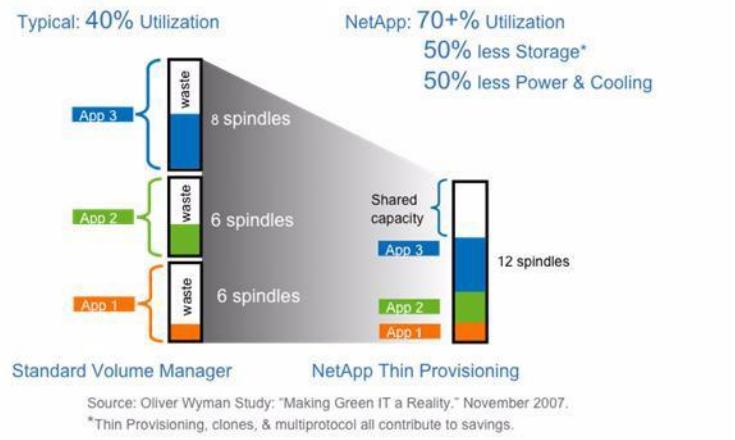
Thin provisioning is a way of logically presenting more storage to hosts than is physically available. With thin provisioning, the storage administrator can use a pool of physical disks (known as an *aggregate*) and create logical volumes for different applications to use, while not preallocating space to those volumes. The space gets allocated only when the host needs it. The unused aggregate space is available for the existing thinly provisioned volumes to expand or for use in creating new volumes. For details about thin provisioning, refer to NetApp TR 3563: NetApp Thin Provisioning.

Figure 1) Traditional and thin provisioning.



NetApp recommends using thinly provisioned LUNs where possible in the Hyper-V environment for maximum storage efficiency. When using thin provisioning, it is important to monitor capacity use. In addition, administrators should configure storage management policies on the volumes that contain the thin-provisioned LUNs. The use of these policies aids in providing the thin-provisioned LUNs with storage capacity as they require it. The policies include automatic sizing of a volume, automatic snapshot deletion, and LUN fractional reserve.

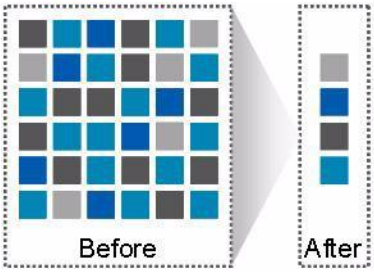
Figure 2) Increased disk use with NetApp thin provisioning.



NETAPP DEDUPLICATION

NetApp deduplication saves space on primary storage by removing redundant copies of blocks within a volume hosting hundreds of virtual desktops. This process is transparent to the application and user, and it can be enabled and disabled on the fly. In a VDI environment, deduplication offers significant space savings, given that each virtual machine is an identical copy of the OS, applications, and patches. The savings are also achieved for the user data hosted on CIFS home directories. For more information about NetApp deduplication, refer to NetApp TR-3505: NetApp Deduplication for FAS, Deployment and Implementation Guide.

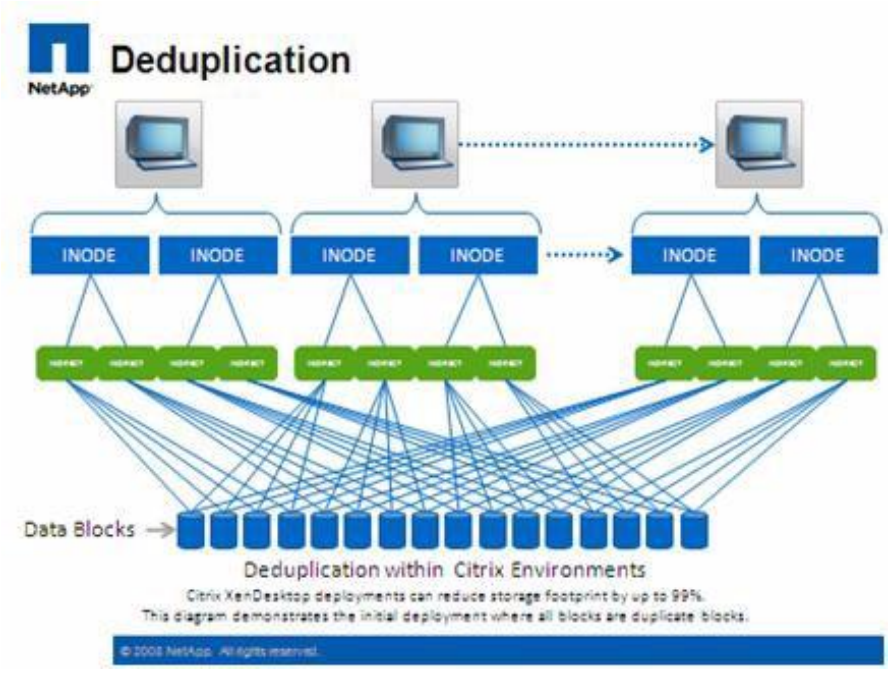
Figure 3) NetApp deduplication.



Using NetApp deduplication and file FlexClone can reduce the overall storage footprint of virtual desktops, and it can also improve performance by using transparent storage cache sharing. Data that is deduplicated (or nonduplicated, in the case of file FlexClone data) on disk exists in the storage array cache only once per volume. All subsequent reads from any of the virtual machine disks of a block that is already in cache are read from cache and not from disk, improving performance tenfold.

Any nondeduplicated data that is not in cache must be read from disk. Data that is deduplicated but that does not have as many block references as heavily deduplicated data appears in cache only once; however, based on the frequency of access, this data might be evicted earlier than data that has many references or is heavily used.

Figure 4) NetApp deduplication and FlexClone.



For more information about deduplication, refer to NetApp TR-3505: NetApp Deduplication for FAS, Deployment and Implementation Guide.

NETAPP FLEXCLONE

NetApp FlexClone technology is hardware-assisted rapid creation of space-efficient, writable, point-in-time images of individual files, LUNs, or flexible volumes. The use of FlexClone technology in VDI deployments offers the flexibility to provision and redeploy thousands of virtual machines rapidly.

FlexClone adds a new level of agility and efficiency to storage operations. FlexClone volumes take only seconds to create and are nondisruptive to the parent FlexVol® volume or virtual machine. FlexClone copies share the same physical data space as the source and occupy negligible space (metadata) on the storage system. FlexClone file-level and volume-level clones use space very efficiently, leveraging the Data ONTAP architecture to store only data that changes between the source and the clone. In addition to all of these benefits, file-level and volume-level FlexClone volumes have the same high performance as other FlexVol volumes or files hosted on the volumes. Also, FlexClone technology offers significant benefits with disaster recovery (DR) testing. DR testing with FlexClone is risk free and can be done during operational hours. For more information about FlexClone technology, refer to NetApp TR-3347: FlexClone Volumes: A Thorough Introduction.

PERFORMANCE

Virtual desktops can be both read and write intensive at different times during the lifecycle of the desktop, depending on the user activity and the desktop maintenance cycle. Most large-scale deployments experience performance-intensive activities, referred to as *storm activities,* such as:

- Boot storms
- Login storms
- Virus scan and/or definition update storms

With physical desktops, this was not a problem because each machine had its own disks and I/O was contained in a single desktop. With VDI using a shared storage infrastructure, significant performance issues can arise during these critical operations. This essentially means that the solution would require a large number of additional spindles to meet performance requirements, resulting in increased overall solution cost.

To solve this problem, the NetApp solution contains transparent storage cache sharing (TSCS). TSCS is a core component of Data ONTAP and is extended with NetApp Flash Cache. These solution components save customers money by:

- Requiring far fewer disks and less cache
- Serving read data from cache, freeing up disk I/O to perform writes
- Providing better throughput and system use
- Providing faster response times and a better overall end-user experience

TRANSPARENT STORAGE CACHE SHARING

Transparent storage cache sharing (TSCS) allows customers to benefit from NetApp's storage efficiency and at the same time to significantly increase I/O performance. TSCS is natively built into the Data ONTAP operating system and works by using block-sharing technologies such as NetApp primary storage deduplication and file and volume FlexClone to reduce the amount of cache required and to eliminate

duplicate disk reads. Only one instance of any duplicate block is read into cache, thus requiring less cache than traditional storage solutions. Because VDI implementations can see initial space savings as high as 99% (validated in the NetApp solutions lab) by using NetApp space-efficient cloning technologies, this translates into higher cache deduplication and high cache hit rates. TSCS is especially effective in addressing the simultaneous system boot, or boot storm, of hundreds to thousands of virtual desktop systems, which can overload a traditional storage system.

Here are the main benefits of transparent storage cache sharing:

- **Increased performance.** With transparent storage cache sharing, in combination with FlexClone and deduplication, latencies decrease by a factor of 10 versus serving data from the fastest-spinning disks available, giving sub-millisecond data access. Decreasing the latency results in higher throughput and lower disk utilization, which directly translate into fewer disks reads.
- **Lowering TCO.** Requiring fewer disks and getting better performance means that customers can increase the number of virtual machines on a given storage platform, resulting in a lower total cost of ownership.
- **Green benefits.** Power and cooling costs are reduced because the energy needed to run and cool the Flash Cache module is significantly less than even a single shelf of Fibre Channel disks. A standard disk shelf of 300GB 15K RPM disks can consume as much as 340 watts/hr and generate heat up to 1394BTU/hr. In contrast, the Flash Cache module consumes a mere 18W/hr and generates 90BTU/hr. By not deploying a single shelf, the power savings alone can be as much as 3000kW/hr/year per shelf. In addition to the environmental benefits of heating and cooling, you can save 3U of rack space per shelf. For a real-world deployment, a NetApp solution (with Flash Cache as a primary component) can typically replace several such storage shelves, so the savings could be considerably higher

NETAPP FLASH CACHE

The NetApp Performance Acceleration Module (PAM) and Flash Cache (formerly PAM II) are hardware devices that extend the native Data ONTAP TSCS capabilities. Flash Cache increases the amount of available cache, which helps reduce virtual desktop storm activities. For more information about NetApp Flash Cache technology, refer to http://www.netapp.com/us/products/storage-systems/flash-cache/flash-cache-tech-specs.html.

NETAPP WRITE OPTIMIZATION

Virtual desktop I/O patterns are often random in nature. Random writes are the most expensive operation for almost all RAID types because each write operation requires more than one disk operation. The ratio of VDI client operation to disk operation also depends on the RAID type for the back-end storage array. In a RAID 5 configuration on a traditional storage array, each client write operation requires up to four disk operations. Large write cache might help, but traditional storage arrays still require at least two disk operations. (Some coalescing of requests happens if you have a big enough write cache. Also, there is a chance that one of the reads might come from the read cache.) In a RAID 10 configuration, each client write operation requires two disk operations. The cost of RAID 10 is very high compared to RAID 5. However, RAID 5 offers less resiliency (protection against single disk failure). Imagine dual disk failure in the middle of the day, making hundreds to thousands of users unproductive.

With NetApp, write operations are optimized for NetApp RAID-DP® by the Data ONTAP core operating system and WAFL® (Write Anywhere File System). NetApp arrays coalesce multiple client write operations and send them to disk as a single IOP. Therefore the ratio of client operations to disk operations is always less than 1, as compared to traditional storage arrays with RAID 5 or RAID 10, which require at least two disk operations per client operation. Also, RAID-DP provides the desired resiliency (protection against dual disk failure) and performance, comparable to RAID 10 but at the cost of RAID 5.

FLEXIBLE VOLUMES AND AGGREGATES

Flexible volumes (also known as FlexVol volumes) and aggregates provide pools of storage. This storage virtualization allows the performance and capacity to be shared by all desktops in the volume or

aggregate. Much like the way that Citrix virtualizes computing resources, NetApp virtualizes the storage resources.

DATA PROTECTION

The availability of thousands of virtual desktops depends on the availability of the shared storage on which the virtual desktops are hosted. Therefore using the proper RAID technology is critical, as is being able to protect the virtual desktop images and/or user data. RAID-DP, the Citrix StorageLink virtual machine backup and recovery function, NetApp SnapMirror® replication technology, and NetApp Snapshot™ copies are critical components of the NetApp solution that address storage availability.

RAID-DP

With any Citrix XenDesktop deployment, data protection is critical, because any RAID failure could result in hundreds to thousands of end users being disconnected from their desktops, resulting in lost productivity. RAID DP provides performance that is comparable to that of RAID 10 but requires fewer disks to achieve equivalent protection. RAID DP protects against double disk failure as compared to RAID 5, which can protect against only one disk failure per RAID group. For more information about RAID DP, refer to NetApp TR-3298: RAID-DP: NetApp Implementation of RAID Double Parity for Data Protection.

2.5 STORAGE SIZING BEST PRACTICES

To estimate the storage required for deploying a VDI solution on NetApp, perform the following tasks:

- Gather essential solution requirement information
- Estimate performance-based and capacity-based storage
- Get recommendations on storage system physical and logical configuration

GATHER ESSENTIAL SOLUTION REQUIREMENT INFORMATION

The first step in the storage sizing process is to gather the solution requirements. This is essential to size the storage system correctly in terms of the model and the number of required NetApp storage controllers, type and quantity of disk spindles, software features, and general configuration recommendations.

The main storage sizing elements are:

- Total number of virtual machines for which the system has to be designed (for example, 2,000 VMs),
- The types and percentage of different types of desktops being deployed. For example, if Citrix XenDesktop is used, different desktop delivery models might require special storage considerations.
- Size per virtual machine (for example, 20GB C: drive, 2GB data disk).
- Virtual machine OS (for example, Windows® XP, Windows 7, and so on).
- Worker workload profile (type of applications on the virtual machine, IOPS requirement, read-write ratio, if known).
- Number of years for which the storage growth has to be considered.
- Disaster recovery and business continuance requirements.
- Size of NAS (CIFS) home directories.
- NetApp strongly recommends storing user data on NAS (CIFS) home drives. By using NAS home drives, companies can more efficiently manage and protect their user data and eliminate the need to back up the virtual desktops.
- For most Citrix XenDesktop deployments, companies may also plan to implement roaming profiles

and/or folder redirection. For information about implementing these technologies, refer to the following documentation:

- Microsoft Configuring Roaming User Profiles
- NetApp TR-3367: NetApp Systems in a Microsoft Windows Environment
- Microsoft Configuring Folder Redirection

When implementing Citrix XenDesktop, decide on the following:

- Types of desktops that will be deployed for different user profiles.
- Data protection requirements for different data components (OS disk, user data disk, CIFS home directories) for each desktop type being implemented.
- For Citrix Provisioning Server pooled desktops, the write back cache size needs to be calculated based on how often the user reboots the desktop and what applications the user uses. NetApp recommends using NFS for write back cache for space efficiency and easy management.
- NetApp thin provisioning and deduplication and NetApp Snapshot can be used to achieve the desired storage efficiency and data protection for the user data disk.

ESTIMATE PERFORMANCE-BASED AND CAPACITY-BASED STORAGE

There are two important considerations for sizing storage for Citrix XenDesktop. The storage system must be able to meet both the performance and capacity requirements of the project, and it must be scalable to accommodate future growth.

The steps for calculating these storage requirements are:

- Determine the storage sizing building block
- Perform detailed performance estimation
- Perform detailed capacity estimation

GET RECOMMENDATIONS ON STORAGE SYSTEM PHYSICAL AND LOGICAL CONFIGURATION

After determining the total capacity and performance requirements, contact your NetApp technical resource to determine the appropriate storage system configuration. Give the total capacity and performance requirements to the NetApp SE to obtain the appropriate storage system configuration. If required, NetApp can help you in every phase of the process described in this section. NetApp has detailed sizing tools that are specific to VDI to help architect deployments of any scale. The tools are designed to factor in all the NetApp storage efficiency and performance acceleration components discussed earlier.

This step also involves planning the logical architecture (the total number of templates and the associated FlexClone volumes that should be provisioned per aggregate). NetApp recommends provisioning fewer large aggregates rather than provisioning more smaller aggregates. The advantage of larger aggregates is that the I/O has more disks to write across, thereby increasing the performance of all volumes contained in the aggregate.

Based on the estimated volume size from the capacity estimation described earlier in this section, determine the number of templates and associated FlexClone volumes that can be hosted in the largest possible aggregate. It is also a good idea to leave some room to grow the aggregates to handle unexpected growth. Also, disable scheduled aggregate Snapshot copies and set the aggregate snap reserve to zero. Make sure that the data disk in the aggregate satisfies the performance requirements for the proposed number of virtual machines for volumes to be hosted in the aggregate.

2.6 STORAGE ARCHITECTURE BEST PRACTICES

In a VDI environment, the availability and performance of the storage infrastructure are critical because thousands of users will be affected by storage outages or performance issues. The storage architecture must provide the necessary level of availability and performance for business-critical applications. NetApp has all the software and hardware solutions to address the availability and performance for large, scalable VDI environments. For a complete Citrix XenDesktop deployment guide, refer to NetApp TR-3795: XenDesktop on Hyper-V with NetApp.

This section is a high-level overview of the components and features to consider when deploying a Citrix XenDesktop infrastructure on NetApp. For detailed information about storage resiliency, refer to the following documents:

- NetApp TR-3437: Storage Best Practices and Resiliency Guide
- NetApp TR-3450: Active-Active Controller Overview and Best Practices Guidelines

BUILDING A RESILIENT STORAGE ARCHITECTURE

- **Active-active NetApp controllers.** If it is not designed correctly, the controller in a storage system can be a single point of failure. Active-active controllers deliver enterprise-class availability by providing controller redundancy and simple automatic transparent failover in the event of a controller failure . Transparent recovery from component failure is critical because all desktops rely on the shared storage. For more information, go to www.netapp.com/us/products/platform-os/active-active.html.

- **Multipath high availability (HA).** Multipath HA storage configuration further enhances the resiliency and performance of active-active controller configurations. Multipath HA-configured storage enhances storage resiliency by reducing unnecessary takeover by a partner node due to a storage fault, improving overall system availability and promoting higher performance consistency. Multipath HA provides added protection against various storage faults, including HBA or port failure, controller-to-shelf cable failure, shelf module failure, dual inter-shelf cable failure, and secondary path failure. Multipath HA enhances consistent performance in active-active configurations by providing larger aggregate storage loop bandwidth. For more information, go to http://media.netapp.com/documents/tr-3437.pdf.

- **RAID data protection.** Data protection against disk drive failure using RAID is a standard feature of most shared storage devices, but with the capacity and subsequent rebuild times of current hard drives where exposure to another drive failure can be catastrophic, protection against double disk failure is now essential. NetApp RAID-DP is an advanced RAID technology that is the default RAID level on all FAS systems. RAID-DP performance is comparable to that of RAID 10, with much higher resiliency. It offers protection against double disk failure as compared to RAID 5, which can protect against only one disk failure. NetApp strongly recommends using RAID-DP on all RAID groups that store Citrix XenDesktop data. For more information about RAID-DP, refer to NetApp white paper 3298 at http://www.netapp.com/us/library/white-papers/wp_3298.html.

- **Service Processor (SP) for remote system management.** The SP enables administrators to access the storage system remotely to diagnose, shut down, power cycle, or reboot the system, regardless of the state of the storage controller. It is powered by a standby voltage, which is available as long as the system has input power to at least one of its power supplies. It is connected to the system through the serial console. You can log in to the SP by using a Secure Shell client application from an administration host and then use the SP CLI to monitor and troubleshoot the system remotely. In addition, you can use the SP to access the system console and run Data ONTAP commands remotely, and you can open both an SP CLI session and a separate system console session simultaneously.

  The SP monitors the system temperatures, voltages, currents, and fan speeds. When the SP detects that an environmental sensor has reached an abnormal condition, it logs the abnormal readings, notifies Data ONTAP of the issue, and proactively sends alerts and “down system” notifications through an

AutoSupport™ message if necessary. The SP also logs system events such as boot progress, field replaceable unit (FRU) changes, Data ONTAP generated events, and SP command history. Hardware-assisted takeover is available on systems that support the SP and that have the SP configured.

For more information about the Service Processor, refer to the NetApp SP FAQ at https://kb.netapp.com/support/index?page=content&id=3012997.

- **Networking infrastructure design.** A network infrastructure (FCoE, Fibre Channel, or IP) should not have a single point of failure. A highly available solution includes having two or more Fibre Channel, FCoE, or IP network switches; two or more CNAs, HBAs, or NICs per host; and two or more target ports or NICs per storage controller. In addition, if using Fibre Channel, two or more redundant FC or Ethernet paths provided by independent FC fabrics or network/FCoE switches are required to have a truly redundant architecture.

TOP RESILIENCY PRACTICES

- Use RAID-DP, the NetApp high-performance implementation of RAID 6, for better data protection.
- Use multipath HA with active-active storage configurations to improve overall system availability and to promote higher performance consistency.
- Use the default RAID group size (16) when creating aggregates.

  Note: The RAID group size may be different, based on disk type. Don't create incomplete RAID groups, because they can ruin the performance of the aggregate.
- Allow Data ONTAP to select disks automatically when creating aggregates or volumes.
- Use the latest Data ONTAP general availability release, available on the NetApp Support (formerly NOW®) site.
- Use the latest storage controller, shelf, and disk firmware, available on the NetApp Support site.
- Disk drive differences are Fibre Channel, SAS, SATA disk drive types, disk size, and rotational speed (RPM).
- Maintain two hot spares for each type of disk drive in the storage system to take advantage of the Maintenance Center.
- Except for FAS2000 series systems, do not put user data into the root volume, due to lack of disk spindles.
- Replicate data with SnapMirror or SnapVault® for disaster recovery protection.
- Replicate to remote locations to increase data protection levels.
- Use an active-active storage controller configuration to eliminate single points of failure.
- Deploy SyncMirror® and RAID-DP for the highest level of storage resiliency.

For more information, refer to NetApp TR-3437: Storage Best Practices and Resiliency Guide.

BUILDING A HIGH-PERFORMANCE STORAGE ARCHITECTURE

A VDI workload can be very I/O intensive, especially during the simultaneous boot up, login, and virus scan in the virtual desktops. Depending on how many servers and guests are attached to the storage, a boot storm can create a significant impact on performance if the storage is not sized properly. A boot storm can affect both the speed at which the desktops are available to the customer and the overall customer experience. A virus scan storm is similar to a boot storm in I/O, but it might last longer and can significantly affect customer experience.

It is important to make sure that the storage is architected to eliminate or decrease the effect of these events.

- **Aggregate sizing.** An aggregate is NetApp's virtualization layer, which abstracts physical disks from logical datasets (referred to as *flexible volumes*). Aggregates are the means by which the total IOPS available to all of the physical disks are pooled as a resource. This design is well suited to meet the needs of an unpredictable and mixed workload. NetApp recommends that whenever possible a small aggregate should be used as the root aggregate. The root aggregate stores the files required for running and providing GUI management tools for the storage system. The remaining storage should be placed into a small number of large aggregates. The overall disk I/O from virtualization environments is traditionally random in nature, so this storage design gives optimal performance because a large number of physical spindles are available to service I/O requests. On smaller storage systems, it might not be practical to have more than a single aggregate, due to the restricted number of disk drives on the system. In this case, it is acceptable to have only a single aggregate.
- **Disk configuration summary.** When sizing the disk solution, consider the number of desktops being served by the storage controller or disk system and the number of IOPS per desktop. Then it is possible to calculate the number and size of the disks needed to serve the given workload. Remember to keep the aggregates large, spindle count high, and rotational speed fast. When one factor needs to be adjusted, Flash Cache can help eliminate potential bottlenecks to the disk.
- **Flexible volumes.** Flexible volumes contain either LUNs or virtual disk files that are accessed by Citrix XenDesktop servers. NetApp recommends a one-to-one alignment of Citrix XenDesktop datastores to flexible volumes. This design offers an easy means to understand the Citrix XenDesktop data layout when viewing the storage configuration from the storage system. This mapping model also makes it easy to implement Snapshot backups and SnapMirror replication policies at the datastore level, because NetApp implements these storage-side features at the flexible volume level.
- **Flash Cache.** Flash Cache enables transparent storage cache sharing and improves read performance, which in turn increases throughput and decreases latency. It provides greater system scalability by removing IOPS limitations due to individual disk I/O bottlenecks, and it lowers cost by providing the equivalent performance with fewer disks. Using Flash Cache in a dense (deduplicated) volume allows all the shared blocks to be accessed directly from the intelligent, faster Flash Cache versus disk. Flash Cache offers great benefits in Citrix XenDesktop environments, especially during a boot storm, login storm, or virus storm, because only one copy of deduplicated data needs to be read from the disk (per volume). Each subsequent access of a shared block is read from Flash Cache and not from disk, increasing performance and decreasing latency and overall disk utilization.

2.7 ADDITIONAL NETAPP SOFTWARE

STORAGE PROVISIONING AND MANAGEMENT USING NETAPP ONCOMMAND PLUG-IN 3.0

The NetApp OnCommand™ plug-in 3.0 for Microsoft is the successor to ApplianceWatch™ PRO 2.1.1. It is an enterprise-class storage monitoring application that provides integration with Microsoft System Center Operations Manager (SCOM) and System Center Virtual Machine Manager (SCVMM). It enables administrators to monitor, manage, and report on NetApp storage.

The key features of this application are:

- LUN and VHD misalignment detection alerts, reports, and tips for performance and resource optimization (PRO).
- Enables agentless monitoring and discovery functionality on Hyper-V hosts.
- Support for cluster shared volumes (CSVs) for highly available VMs.
- Limited support for NetApp MultiStore®.
- Discovers, monitors, and reports on vFiler® units similarly to physical controllers.
- Displays the relationship between vFiler units and their physical controllers.

- Presents volume-level and LUN-level statistics in the form of views and reports.
- Presents storage efficiency statistics for volumes, aggregates, and controllers in the form of views and reports.
- Rapidly provisions and clones VMs by using Windows PowerShell™ scripts or Microsoft Opalis. This is done by building workflows using objects from the Opalis Integration pack.
- Performs disaster recovery across geographically dispersed sites and domains.

The OnCommand plug-in 3.0 for Microsoft discovers hardware and storage layouts of NetApp storage systems and provides alerts, health views, and various performance views. Customers can also dynamically manage their virtualized environments along with Performance and Resource Optimization (PRO). PRO is a feature of System Center Virtual Machine Manager 2008 that enables dynamic management of a virtualized infrastructure. PRO provides an open and extensible framework for the creation of management packs for virtualized applications or associated hardware.

A good understanding of Windows administration, SCOM, SCVMM, and Opalis, as well as an understanding of NetApp storage concepts, is necessary to use the NetApp OnCommand plug-in 3.0 for Microsoft. The recommendations in this document are guidelines to assist you with the configuration of the OnCommand plug-in 3.0 for Microsoft. NetApp recommends that you refer to the following documents before using this technical report:

- OnCommand Plug-In 3.0 for Microsoft Installation and Administration Guide
- OnCommand Plug-in 3.0 for Microsoft release notes
- Rapid Provisioning and Cloning Command Reference Guide

DATA ONTAP DSM 3.5 FOR WINDOWS MPIO

The Data ONTAP DSM 3.5 for Windows MPIO is a device-specific module (DSM) that enables NetApp storage systems to integrate with Microsoft Multipath I/O (MPIO) to provide applications with high-availability connections to NetApp storage. The Data ONTAP DSM is used with NetApp storage system configurations in which Windows hosts have more than one physical path to the NetApp system.
You can have multiple optimized paths and multiple non-optimized paths. If all of the optimized paths fail, the DSM automatically switches to the non-optimized paths, maintaining the host's access to its storage.

WINDOWS HOST UTILITIES 6.0

Windows Host Utilities 6.0 contains the software components you need to configure a Microsoft Windows Server® 2003, Windows Server 2008, or Windows Server 2008 R2 system to access virtual disks (LUNs) on a NetApp storage system by using the Fibre Channel, FCoE, or iSCSI protocol.

Windows Host Utilities 6.0 also supports Windows guest operating systems running on Windows Hyper-V virtual machines. These guest operating systems include Windows Server 2008, Windows Server 2003, Windows XP, Windows Vista®, and SUSE Linux® Enterprise Server.

Windows Host Utilities 6.0 supports Fibre Channel, FCoE, iSCSI, and mixed FC and iSCSI connections. It also supports iSCSI host bus adapters (HBAs).

Note: If you want to use multipathing to create a highly available connection to the storage system, you must also install a supported version of multipathing software such as the Data ONTAP DSM for Windows MPIO.

For the currently supported multipathing software versions and related requirements, see the NetApp Interoperability Matrix.

3 HYPER-V STORAGE NETWORK DESIGN

This section applies to:

Storage administrators

Microsoft

VDI administrators

The goal of any storage network is to provide uninterrupted service to all nodes that connect to it. This section focuses primarily on how to establish a highly available Ethernet storage network. There are two reasons for focusing on Ethernet. First, FC storage networks provide a single service, Fibre Channel. These single-purpose networks are simpler to design and deploy in a highly available configuration. Second, the current industry trend is solely focused on multipurposed Ethernet networks (converged networks) that provide storage, voice, and user access.

Regardless of protocol, a production storage network must address the following goals:

- Be redundant across switches in a multiswitch environment
- Use as many available physical paths as possible
- Be scalable across multiple physical interfaces or ports

3.1 SAN AND NAS STORAGE NETWORKING BASICS

A network infrastructure (FC or IP) should have no single point of failure. A highly available solution includes two or more FC or IP network switches, two or more HBAs or NICs per Hyper-V server, and two or more target ports or NICs per storage controller. In addition, if using Fibre Channel, two fabrics are required for a truly redundant architecture. For more information about designing and deploying an FC or iSCSI solution, refer to the NetApp Fibre Channel and iSCSI Configuration Guide for your version of Data ONTAP on the NetApp Support site.

3.2 FIBRE CHANNEL STORAGE NETWORKING BASICS

FC storage networks make up a large percentage of mature Hyper-V storage infrastructures. This section covers best practices for deploying Hyper-V on FC with NetApp storage arrays.

CONNECTIVITY BASICS

Hyper-V hosts and NetApp storage arrays connect to a SAN fabric by using host bus adapters (HBAs). Connectivity to FCoE fabrics is enabled through converged network adapters (CNAs). Each HBA or CNA can run as either an initiator (iSCSI) or a target (NetApp). If you are using an HBA on the host, use the FC initiator. If you are using CNA on the host, use the FC or iSCSI initiator. If you are using an HBA on NetApp, use the FC target If you are using CNA on NetApp, use the FC or iSCSI target.

Each adapter has a global unique address (a World Wide Name, WWN). Each WWN must be known to configure LUN access on a NetApp storage array.

Note: Both NetApp and Microsoft highly recommend that as a best practice each Hyper-V host should have at least two adapter ports.

FABRIC ZONING RECOMMENDATION

Many devices and nodes can be attached to a SAN. Implementing zones is a method that is used to secure access and to optimize I/O access to these devices. SAN zoning is a method of arranging FC devices into logical groups over the physical configuration of the fabric or FC network.

Zoning is available in hardware (hard zoning) or in software (soft zoning). An option that is available with both implementations is called *port zoning.* With this option, physical ports define security zones. A host's access to a LUN is determined by what physical port it connects to. With port zoning, zone information must be updated every time a user changes switch ports.

Another form of zoning is WWN zoning, where the fabric leverages its name servers to either allow or block access to particular World Wide Port Names (WWPNs) in the fabric. A major advantage of WWPN zoning is the ability to recable the fabric without having to modify the zone members.

When provisioning storage for access by using FCP or iSCSI, the storage must be masked so that the appropriate Hyper-V parent and child partitions can connect to them. With a NetApp storage system, LUN masking is handled by the creation of initiator groups (also known as igroups).

If NetApp SnapDrive® for Windows is installed in the Hyper-V server or guest OS, then it can be used to configure igroups though the Manage IGroup Wizard under Disks > Actions. However, if you are not using NetApp SnapDrive, then you must follow a manual process, in which you first obtain the IQN or WWPN for the configured storage connectivity.

Note: NetApp recommends creating an igroup for each Hyper-V server, each Windows failover cluster (a combination of multiple Hyper-V servers), or each guest OS (for the option of direct LUN access by guest OS using the Microsoft iSCSI Software Initiator).

If a Hyper-V server or cluster uses both Fibre Channel and iSCSI protocols, separate igroups must be created for Fibre Channel and iSCSI.

NetApp also recommends including the name of the Hyper-V server, Windows failover cluster, or guest OS and the protocol type (for example, DC1_FCP or DC1_iSCSI) in the naming convention for the igroup. This naming convention and method simplify the management of igroups by reducing the total number created. It also means that all Hyper-V servers in the Windows failover cluster see each LUN at the same ID. Each initiator group includes all of the FCP Worldwide Port Names (WWPNs) or iSCSI qualified names (IQNs) of the Hyper-V servers in the cluster.

NetApp recommends using NetApp SnapDrive to manage igroups for the configured storage.

3.3 IP STORAGE NETWORKING

NETAPP VIRTUAL INTERFACES

A virtual network interface (VIF) is a mechanism that supports the aggregation of network interfaces into one logical interface unit. Once created, a VIF is indistinguishable from a physical network interface. VIFs are used to provide fault tolerance of the network connection and in some cases higher throughput to the storage device.

Multimode VIFs are compliant with IEEE 802.3ad. In a multimode VIF, all of the physical connections in the VIF are simultaneously active and can carry traffic. This mode requires that all of the interfaces are connected to a switch that supports trunking or aggregation over multiple port connections. The switch must be configured to understand that all the port connections share a common MAC address and are part of a single logical interface. Figure 5 is an example of a multimode VIF; interfaces e0, e1, e2, and e3 are part of the MultiTrunk1 multimode VIF. All four interfaces in the MultiTrunk1 multimode VIF are active.

In a single-mode VIF, only one of the physical connections is active at a time. If the storage controller detects a fault in the active connection, then a standby connection is activated. No configuration is

necessary on the switch to use a single-mode VIF, and the physical interfaces that make up the VIF do not have to connect to the same switch. IP load balancing is not supported on single-mode VIFs. Figure 6 is an example of a single-mode VIF; in the figure, e0 and e1 are part of the SingleTrunk1 single-mode VIF. If the active interface (e0) fails, the standby e1 interface takes over and maintains the connection to the switch.

Figure 5) Multimode VIF.

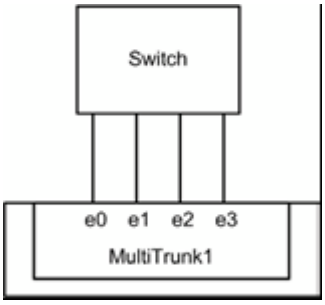
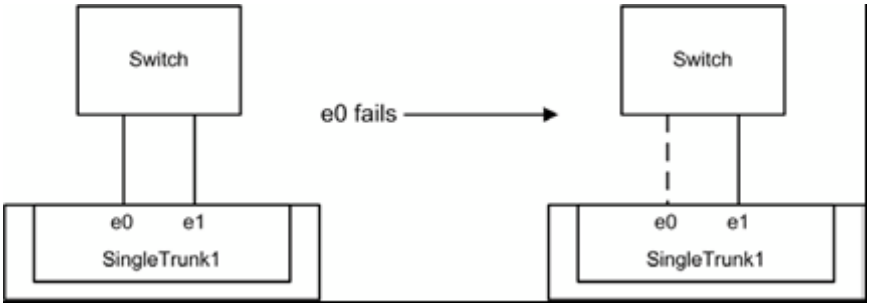
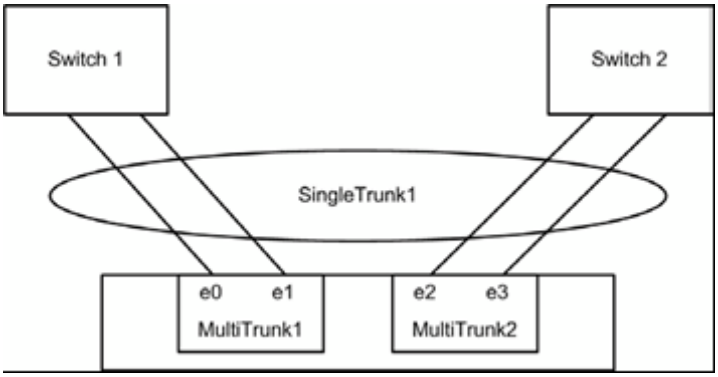


Figure 6) Single-mode VIF.



It is also possible to create second-level single-mode or multimode VIFs, as shown in Figure 7. By using second-level VIFs, it is possible to take advantage of both the link aggregation features of a multimode VIF and the failover capability of a single-mode VIF. In this configuration, two multimode VIFs are created, each one to a different switch. A single-mode VIF is then created composed of the two multimode VIFs. In normal operation, traffic flows over only one of the multimode VIFs, but in the event of an interface or switch failure, the storage controller moves the network traffic to the other multimode VIF. For detailed information about the different types of VIF, refer to the Data ONTAP Network and File Access Management Guide for your version of Data ONTAP.

Figure 7) Second-level VIF.



Note: The storage controllers should have two or more target ports to be sure that redundant paths are available between the NetApp storage system and the Hyper-V servers.

The use of LACP (dynamic multimode VIF) is also supported, beginning with Data ONTAP 7.2.1. However, dynamic multimode VIFs have some special requirements:

- They must be connected to a switch that supports LACP.
- They must be configured as first-level VIFs.
- They should be configured to use the IP-based load-balancing method.

ISCSI TRAFFIC SECURITY

NetApp storage controllers also allow the restriction of the iSCSI protocol to specific interfaces or VLAN tags. These simple configuration settings have an enormous effect on the security and availability of IP-based host disks.

Note: For Hyper-V environments that will be deployed using iSCSI, NetApp strongly recommends implementing separate physical networks for handling the general IP communication between servers and virtual machines and for handling all storage-related IP communication. At a minimum, the traffic should be virtually segregated by using 802.1Q VLAN tagging.

MASKING AND ZONING

When storage is provisioned for access by using FCP or iSCSI, the storage must be masked so that the appropriate Hyper-V parent and child partitions can connect to it. With a NetApp storage system, LUN masking is handled by the creation of igroups.

If NetApp SnapDrive for Windows is installed in the Hyper-V server or a guest OS, then it can be used to configure igroups through the Manage IGroup wizard under Disks > Actions. However, if you are not using NetApp SnapDrive, then you must follow a manual process, in which you first obtain the iSCSI qualified names (IQNs) or Worldwide Port Names (WWPNs) for the configured storage connectivity.

Note: NetApp recommends creating an igroup for each Hyper-V server, each Windows failover cluster (a combination of multiple Hyper-V servers), or each guest OS (for the option of direct LUN access by the guest OS using the Microsoft iSCSI Software Initiator).

If a Hyper-V server or cluster uses both Fibre Channel and iSCSI protocols, separate igroups must be created for Fibre Channel and iSCSI.

NetApp also recommends including the name of the Hyper-V server, Windows failover cluster, or guest OS and the protocol type (for example, DC1_FCP or DC1_iSCSI) in the naming convention for the igroup. This naming convention and method simplify the management of igroups by reducing the total number created. It also means that all Hyper-V servers in the Windows failover cluster see each LUN at the same ID. Each initiator group includes all of the FCP WWPNs or IQNs of the Hyper-V servers in the cluster.

NetApp recommends using NetApp SnapDrive to manage igroups for the configured storage.

Obtaining the IQN (iSCSI)

Obtain the IQN from the Hyper-V server or guest OS by opening a command prompt and running the following command: **`iscsicli`**. For more information about obtaining the IQN for all configured iSCSI initiators, see “Configure Windows Server 2008 Initiator Groups on NetApp Storage” in TR-3701: NetApp and Microsoft Virtualization: Solution and Implementation Guide.

Obtaining the WWPN (FC)

Obtain the WWPN for the FC ports by using the NetApp Windows Host Utility (discussed in section 2.7,

"Additional NetApp Software") or the HBA vendor host utility (for example, Qlogic SAN Surfer) on the Hyper-V server. For more information about obtaining the WWPN of all Fibre Channel HBAs installed, see "Fibre Channel Zoning Configuration" in TR-3701: NetApp and Microsoft Virtualization: Solution and Implementation Guide.

3.4 HYPER-V SERVER NETWORKING CONSIDERATIONS

This section discusses the following aspects of Hyper-V server networking:

- Physical network adapters
- Virtual networks
- Network feature support
- Network naming standard
- Network adapter binding order and metric values

PHYSICAL NETWORK ADAPTERS

Most Hyper-V servers have four or more physical network adapters installed to handle Hyper-V management, virtual machine connectivity, IP storage connectivity, Windows failover cluster (WFC) or WFC heartbeat communication, live migration communication, and cluster shared volume (CSV) communication. Smaller environments require a minimum of 2 or 3 network adapters, and larger environments require at least 4 or 5 network adapters. Why are multiple physical network adapters necessary?

- **Hyper-V management.** As a best practice, Microsoft has always recommended that the Hyper-V parent partition, also known as the management operating system (MOS), should have a dedicated physical network adapter for management of the Hyper-V server. Communication is necessary to manage the Hyper-V server and any VMs it hosts remotely, from another system or from SCVMM; therefore you should consider using network interface card (NIC) teaming for redundancy. For more information, see Configuring Virtual Networks on Microsoft TechNet.

- **Virtual machines.** VMs can communicate over external, internal, and private virtual networks that are implemented through the Hyper-V parent partition. Each external virtual switch must map to an individual physical network adapter or logical network adapter from the result of NIC teaming, which is discussed later in this section. To provide redundancy in a production environment, you can assign an external virtual switch to a network team or use multiple external virtual switches; for both configurations, a minimum of two physical network adapters is needed to provide redundancy. For more information, see Configuring Virtual Networks on Microsoft TechNet.

- **IP storage.** As a best practice, Microsoft recommends that IP storage communication should be separate from virtual machine and cluster communications. NetApp supports this practice. Therefore a minimum of one physical network adapter is required to support iSCSI communication from the Hyper-V parent partition. If you want to use multipathing or multipath input/output (MPIO) from the Hyper-V parent partition, a minimum of two physical network adapters is required. If you are enabling Windows failover clustering for Hyper-V, maintaining separation of IP storage traffic becomes a requirement for configuration before validating a failover cluster. For more information, see Hyper-V: Using Hyper-V and Failover Clustering and Hyper-V: Using Live Migration with Cluster Shared Volumes in Windows Server 2008 R2 on Microsoft TechNet.

- **Windows failover cluster private network.** If you create a Windows failover cluster for Hyper-V, it requires a cluster private network and therefore might require a dedicated physical network adapter. In previous versions of Windows Server®, this was used primarily for the cluster heartbeat communications, but with R2 it is also used for cluster shared volumes or live migration communications (see "Live migration" and "Cluster shared volumes" in this section). For more information, see Hyper-V: Using Hyper-V and Failover Clustering and Hyper-V: Using Live Migration with Cluster Shared Volumes in Windows Server 2008 R2 on Microsoft TechNet.

- **Live migration.** This is a new feature for Windows Server 2008 R2; it does not apply to versions of Hyper-V before R2. When live migration of virtual machines is taking place, the communications for facilitating this migration traverse the network. Microsoft recommends configuring a dedicated physical network adapter for only live migration traffic in the Failover Cluster Manager MMC or using Windows PowerShell. For more information, see Hyper-V Live Migration Overview and Architecture in the Microsoft Download Center and Hyper-V: Using Live Migration with Cluster Shared Volumes in Windows Server 2008 R2 on Microsoft TechNet.

- **Cluster shared volumes.** CSVs are also a new feature for R2, so this section does not apply to versions of Hyper-V before R2. When CSVs are enabled in a Windows failover cluster for Hyper-V, communication between Hyper-V cluster nodes that are owners and non-owners of a particular CSV include health checks and dynamic I/O redirection. Microsoft recommends a dedicated physical network adapter to make sure that the necessary bandwidth is available to support these operations and to minimize the possibility of a failover caused by the inability to support CSV communication between nodes. For more information, see Hyper-V: Using Live Migration with Cluster Shared Volumes in Windows Server 2008 R2 on Microsoft TechNet.

NETWORK ADAPTER TEAMING

In the past, there was much confusion about the support of NIC teaming with Microsoft Hyper-V. Microsoft explicitly states that it does not support use of NIC teaming with Microsoft Hyper-V. This means that Microsoft does not have a proprietary driver for all types of NICs installed in the Hyper-V server that supports NIC teaming. Only the manufacturers of the NICs, such as Intel and Broadcom, have these drivers, and their driver software supports the use of NIC teaming.

Therefore you can use NIC teaming in Windows Server 2008 R2 as long as the manufacturer of the NIC supports it. Microsoft does not contest enabling the Hyper-V role and then assigning a logical network interface that represents the teamed NICs to a virtual switch. The benefits of this procedure are obvious: Not only does the virtual switch have access to increased bandwidth, it also has access to increased availability thanks to the redundancy of teaming multiple NICs together. However, some of the network features are not available on the logical NIC that is created as a result of teaming one or more physical NICs; for more information, see “Network Feature Support,” later in this section.

Microsoft does not support the use of NIC teaming for networks that are used for iSCSI communications. Therefore you cannot use a logical network interface for iSCSI communications when using the Microsoft iSCSI Software Initiator to present LUNs to the Hyper-V parent partition. In addition, when a logical NIC interface is assigned to a virtual switch, the VMs that are connected to that specific virtual switch must not have the Microsoft iSCSI Software Initiator enabled for that virtual NIC.

Note: For functional areas with more than one connection, such as the multiple network adapters used for VM communications, the connections should be spread across different network adapters, especially if multiple port network adapters are installed. This allows those functional areas to maintain connectivity to the network when configured properly so that connectivity is not lost in the event of a port or adapter failure in the Hyper-V server.

Because NIC teaming is not supported for iSCSI communications, NetApp recommends configuring multiple paths for storage connectivity to provide redundancy and additional bandwidth in some configurations that use multipathing/MPIO.

NETWORK FEATURE SUPPORT

With the release of Windows Server 2008 R2, support for several new networking features has been added, including jumbo frame support for GbE networks and TCP chimney support for 10GbE networks. These network technologies allow Hyper-V R2 to take advantage of network offload technologies, so instead of the

Hyper-V CPUs processing network packets, these packets can now be handled by the offload technologies to help improve performance by reducing CPU use.

Windows Server 2008 R2 Hyper-V can fall victim to a few different performance bottlenecks in the network architecture. These bottlenecks can be classified into two categories: receive path and transmit path.

Large Send Offload and Checksum Offload

Large send offload (LSO) and checksum offload (CSO) are supported by the virtual networks in Hyper-V. In addition, if the physical network adapters support it as well, the virtual traffic is offloaded to the physical network as necessary. Most network adapters support LSO and CSO, but check with your NIC manufacturer to be sure. If it is supported, it is often enabled by default.

Note: Where LSO and CSO are supported, NetApp strongly recommends making sure that they are enabled (if they are not already enabled by default).

Jumbo Frames

Jumbo frames support was initially added with the introduction of Windows Server 2008, but additional improvements have been made to supportability with R2, and the NIC vendors have added a wider variety of NICs that support it. With Windows 2008 R2, enhancements converge to support up to six times the payload per packet, making a huge difference in overall throughput and reducing CPU utilization for large file transfers. In addition, jumbo frames are supported not only on the physical network adapters (as with Windows 2008 before R2), but also on the virtual networking, including switches and adapters. NetApp recommends the use of jumbo frames with NICs that are configured to handle iSCSI communications, but they should be enabled only if there is end-to-end support across all hardware.

Use `ping -l 8000 -f -n 1 <target-ip>` to help determine if there is end-to-end support. Using an 8,000-byte datagram does not make it fragment at layer, and use of the `-f` option prevents false positives.

TCP Chimney

TCP chimney offload supports making connection offload capabilities available to the TCP/IP stack in the child partition. The virtual NIC in the child partition advertises the connection offload capabilities, and then the VM switch in the parent partition offloads the child partition TCP connections to the physical NIC. Applications with long-lived connections with large data transfers and applications with preposted buffers benefit the most from TCP chimney support.

The major benefits of TCP chimney offload support in a Microsoft Hyper-V virtual environment are:

- Significant CPU utilization reduction on end-to-end workloads
- 10GbE connections can be fully used
- Avoids excessive chatter between child and parent partitions
  - Avoids the overhead of parent/child context switching
  - The connection state is fully maintained by the physical NIC
- Live migration support; connections are uploaded to the host stack during the live migration process

Note: Where TCP chimney is supported by the NIC manufacturer, NetApp recommends making sure that it is enabled. (It is disabled by default.)

Virtual Machine Queue

The virtual machine queue (VMQ) helps significantly by classifying and grouping the received packets, parsing them in the hardware, and thereby improving performance. It also applies VLAN filtering in the hardware, dropping all packets with invalid VLAN IDs at the NIC and using a switch on the NIC to do route lookup on transmits, and it avoids a copy of NIC receive buffers to VM address space. All of the VMQ processing happens concurrently in multiple queues that are serviced by different processors.

NICs that support VMQ have embedded switches that allow receive queues to be paired with transmit queues, where each queue pair is a switch port. The switch requires no MAC address learning, and the switch inspects all transmit packets for destination MAC address and VLAN ID. If the packets pass the filter set on the receive queue, they DMA to that queue; otherwise they are sent on the wire. This is a huge advantage in VM-to-VM communication because it avoids route lookup in the software, avoids packet copies, and takes advantage of offload support in the hardware. Even with VM-to-physical communication, route lookup is still avoided, providing some performance benefit.

Overall, VMQ improves network throughput by distributing network traffic for multiple VMs across multiple processors, while reducing processor use by offloading packet classification to the hardware and avoiding both network data copy and route lookup on transmit paths. VMQ is compatible with most other task offloads, and therefore it can coexist with large send offload and jumbo frames; however, where TCP chimney is supported by the NIC, VMQ takes precedence. VMQ is secure and supports live migration in R2. By far the best performance gains are seen with VMQ enabled on 10GbE network interfaces.

VMQ is supported only with specific adapters in Windows Server 2008 R2, and therefore VMQ is disabled by default. Before enabling it, check with Microsoft and your NIC vendor to make sure that your configuration supports VMQ.

Note: Where VMQ is supported by the NIC manufacturer, NetApp recommends considering enabling it in your environment, especially if you have deployed Hyper-V with 10GbE.

NETWORK NAMING STANDARD

Network Naming Considerations

Because there are so many physical and virtual network adapters in a specific Hyper-V server, managing them is difficult. Most administrators establish an easy-to-understand naming standard for all network adapters, both physical and virtual. When deciding on a naming configuration, consider the following points:

- The name should identify whether the adapter is a physical or a virtual network adapter.
- The naming convention should standardize the number of characters allowed, regardless of whether it is a physical or a virtual network adapter.
- For a physical network adapter, identify the physical network adapter hardware.
  - If the physical network adapter is located on the motherboard of the server, consider abbreviating as follows: <M for motherboard><Double-Digit Port #>.
  - If the physical network adapter is an add-on to the server, consider abbreviating as follows: <A for Add-on><PCIe/x Slot #><Single-Digit Port #>.
- For a physical network adapter connected to a physical network:
  - Use an abbreviated descriptor such as LAN for local area network, SAN for IP storage network, and so on.

- Create an abbreviation for the network subnet, using letters for class and three digits for the subnet identifier.
- If using VLAN tagging, use the VLAN ID of the network.

- For a physical network adapter connected to a virtual network or switch:
  - Use an abbreviated descriptor such as `D` for dedicated, `E` for external, `I` for internal, and `P` for private.
  - Use a two-digit code to differentiate the virtual network types, because you may have multiple virtual networks of the same type.

- If a virtual or physical network adapter is connected to an external or dedicated virtual network, identify the virtual network type.

  - Use an abbreviated descriptor such as `D` for dedicated, `E` for external, `I` for internal, and `P` for private.
  - Use a two-digit code to differentiate the virtual network types, because you may have multiple virtual network of the same type.

- If a virtual or physical network adapter is connected to an external or dedicated virtual network, describe the network to which it is connected:

  - Use an abbreviated descriptor such as LAN for local area network, SAN for IP storage network, and so on.

  - If using VLAN tagging, use the VLAN ID of the network.

  - Create an abbreviation for the network subnet. First, use a single alpha character to identify the subnet class. Second, use two, three, or five digits to identify the subnet, with three digits for the class octet and/or two digits for the subnet mask.

NETWORK ADAPTER BINDING ORDER AND METRIC VALUES

Network Adapter Binding Order

Because most Hyper-V servers have a multitude of network adapters, both physical and virtual, the network adapter binding order may be configured incorrectly by Windows. The administrator should verify that the network adapter binding order is correct for each Hyper-V server. This is especially important for Hyper-V servers that are configured as part of a Windows failover cluster. Modifying the network adapter binding order can be accomplished through Network Connections, under Advanced > Advanced Settings, in the Connections field. Network Connections can be found under Control Panel > Network and Internet in Windows Server 2008.

For a server core environment, the `nvspbind` tool can be used to modify network bindings from the command line. For details, refer to http://archive.msdn.microsoft.com/nvspbind.

Note: After all the physical network adapters have been installed, all the Hyper-V virtual networks have been created, and all the networks have been named according to any standard, you must modify the network adapter binding order appropriately.

The first adapter in the binding order should be the adapter that is used to manage the Hyper-V parent partition. The adapters for iSCSI, live migration, and CSV communications should follow, with the private network used for cluster heartbeat and all adapters associated with virtual networks done last.

4 MICROSOFT VIRTUALIZATION BEST PRACTICES

This section applies to:

Storage administrators

VDI administrators

4.1 HYPER-V

SERVER CONFIGURATION

This section describes best practices for selecting hardware for virtualization servers and installing and setting up Windows Server 2008 R2 for the Hyper-V server role.

Hardware Selection

The hardware considerations for Hyper-V servers generally resemble those for nonvirtualized servers, but Hyper-V servers can exhibit increased CPU usage, consume more memory, and need larger I/O bandwidth because of server consolidation.

- **Processors.** Hyper-V in Windows Server 2008 R2 presents the logical processors as one or more virtual processors to each active virtual machine. You can achieve additional run-time efficiency by using processors that support Second Level Address Translation (SLAT) technologies such as EPT and NPT.

  Hyper-V in Windows Server 2008 R2 adds support for deep CPU idle states, timer coalescing, core parking, and guest idle state. These features allow better energy efficiency over previous versions of Hyper-V.
- **Cache.** Hyper-V can benefit from larger processor caches, especially for loads that have a large working set in memory and in VM configurations in which the ratio of virtual processors to logical processors is high.
- **Memory.** The physical server requires sufficient memory for the root and child partitions. Hyper-V first allocates the memory for child partitions, which should be sized based on the needs of the expected load for each VM. Having additional memory available allows the root to efficiently perform I/Os on behalf of the VMs and operations such as VM snapshots.
- **Networking.** If the expected loads are network intensive, the virtualization server can benefit from having multiple network adapters or multiport network adapters. Each network adapter is assigned to its own virtual switch, allowing each virtual switch to service a subset of virtual machines. When hosting multiple VMs, using multiple network adapters allows distribution of the network traffic among the adapters for better overall performance.

  To reduce the CPU usage of network I/Os from VMs, Hyper-V can use hardware offloads such as large send offload (LSOv1), TCPv4 checksum offload, chimney, and VMQ.
- **Storage.** The storage hardware should have sufficient I/O bandwidth and capacity to meet current and future needs of the VMs that the physical server hosts. Consider these requirements when you select storage controllers and disks and choose the RAID configuration. Placing VMs with highly disk-intensive workloads on different physical disks is likely to improve overall performance. For example, if four VMs share a single disk and actively use it, each VM can yield only 25% of the bandwidth of that disk.

Server Core Installation Option

Windows Server 2008 and Windows Server 2008 R2 feature the Server Core installation option. Server Core offers a minimal environment for hosting a select set of server roles, including Hyper-V. It features a smaller disk, memory profile, and attack surface. Therefore NetApp highly recommends that Hyper-V virtualization servers use the Server Core installation option. Using Server Core in the root partition leaves additional memory for the VMs to use (approximately 80MB for commit charge on 64-bit Windows).

Server Core offers a console window only when the user is logged on, but Hyper-V exposes management features through WMI so that administrators can manage it.

Dedicated Server Role

The root partition should be dedicated to the virtualization server role. Additional server roles can adversely affect the performance of the virtualization server, especially if they consume significant CPU, memory, or I/O bandwidth. Minimizing the server roles in the root partition has additional benefits, such as reducing the attack surface and the frequency of updates.

System administrators should consider carefully what software is installed in the root partition, because some software can adversely affect the overall performance of the virtualization server.

Guest Operating Systems

Hyper-V supports and has been tuned for a number of different guest operating systems. The number of virtual processors that are supported per guest depends on the guest operating system.

CPU Statistics

Hyper-V publishes performance counters to help characterize the behavior of the virtualization server and break out the resource usage. The standard set of tools for viewing performance counters in Windows includes Performance Monitor (`Perfmon.exe`) and `Logman.exe`, which can display and log the Hyper-V performance counters. The names of the relevant counter objects are prefixed with “Hyper-V.”

You should always measure the CPU usage of the physical system by using the Hyper-V Hypervisor Logical Processor performance counters. The CPU utilization counters that Task Manager and Performance Monitor report in the root and child partitions do not accurately capture the CPU usage. Use the following performance counters to monitor performance:

\Hyper-V Hypervisor Logical Processor (*) \% Total Run Time

The counter represents the total non-idle time of the logical processors.

\Hyper-V Hypervisor Logical Processor (*) \% Guest Run Time

The counter represents the time spent executing cycles in a guest or in the host.

\Hyper-V Hypervisor Logical Processor (*) \% Hypervisor Run Time

The counter represents the time spent executing in the hypervisor.

\Hyper-V Hypervisor Root Virtual Processor (*) \ *

The counters measure the CPU usage of the root partition.

\Hyper-V Hypervisor Virtual Processor (*) \ *

The counters measure the CPU usage of guest partitions.

PROCESSOR PERFORMANCE

The hypervisor virtualizes the physical processors by time-slicing between the virtual processors. To perform the required emulation, certain instructions and operations require the hypervisor and virtualization stack to run. Moving a workload into a VM increases the CPU usage; this guide describes best practices for minimizing that overhead.

VM Integration Services

The VM Integration Services include enlightened drivers for the synthetic I/O devices, which significantly reduce CPU overhead for I/O compared to emulated devices. You should install the latest version of the VM Integration Services in every supported guest. The services decrease the CPU usage of the guests, from idle guests to heavily used guests, and improve the I/O throughput. This is the first step in tuning a Hyper-V server for performance. For the list of supported guest operating systems, see the documentation that is provided with the Hyper-V installation.

Enlightened Guests

The operating system kernels in Windows Vista SP1, Windows 7, Windows Server 2008, and Windows Server 2008 R2 feature enlightenments that optimize their operation for VMs. For best performance, NetApp recommends that you use Windows Server 2008 R2 or Windows Server 2008 as a guest operating system. The enlightenments present in Windows Server 2008 R2 and Windows Server 2008 decrease the CPU overhead of Windows that runs in a VM. The integration services provide additional enlightenments for I/O. Depending on the server load, it can be appropriate to host a server application in a Windows Server guest for better performance.

Virtual Processors

Hyper-V in Windows Server 2008 R2 supports a maximum of four virtual processors (VPs) per VM. Virtual machines that have loads that are not CPU intensive should be configured to use one virtual processor. This is because of the additional overhead that is associated with multiple virtual processors, such as additional synchronization costs in the guest operating system. More CPU-intensive loads should be placed in 2-VP to 4-VP VMs if the VM requires more than one CPU of processing under peak load. The documentation that is provided with the Hyper-V installation lists the supported guest operating systems and the number of virtual processors supported for each operating system.

Windows Server 2008 R2 features enlightenments to the core operating system that improve scalability in multiprocessor VMs. Workloads can benefit from the scalability improvements in Windows Server 2008 R2 if they run in 2-VP to 4-VP VMs.

Background Activity

Minimizing the background activity in idle VMs releases CPU cycles that can be used elsewhere by other VMs or saved to reduce energy consumption. Windows guests typically use less than 1% percent of one CPU when they are idle. Here are several best practices for minimizing the background CPU usage of a VM:

- Install the latest version of the VM Integration Services.
- Remove the emulated network adapter through the VM settings dialog box (use the Microsoft synthetic adapter).
- Remove unused devices such as the CD-ROM and COM port, or disconnect their media.
- Keep the Windows guest at the logon screen when it is not being used.
- Use Windows Server 2008 or Windows Server 2008 R2 for the guest operating system.
- Disable the screen saver.
- Disable, throttle, or stagger periodic activity such as backup and defragmentation.
- Review the scheduled tasks and services that are enabled by default.
- Improve server applications such as timers to reduce periodic activity.
- Use the Balanced power plan instead of the High Performance power plan.

Here are additional best practices for configuring a *client version* of Windows in a VM to reduce the overall CPU usage:

- Disable background services such as SuperFetch and Windows Search.

- Disable scheduled tasks such as Scheduled Defrag.
- Disable AeroGlass and other user interface effects (through the System application in the Control Panel).

Weights and Reserves

Hyper-V supports setting the weight of a virtual processor to grant it a larger or smaller share of CPU cycles than average and specifying the reserve of a virtual processor to make sure that it gets a minimal percentage of CPU cycles. The CPU that a virtual processor consumes can also be limited by specifying usage limits. System administrators can use these features to prioritize specific VMs, but NetApp recommends the default values unless you have a compelling reason to alter them.

Weights and reserves prioritize or deprioritize specific VMs if CPU resources are overcommitted. This makes sure that those VMs receive a larger or smaller share of the CPU. Highly intensive loads can benefit from adding more virtual processors instead, especially when they are close to saturating an entire physical CPU.

Tuning NUMA Node Preference

On Non-Uniform Memory Access (NUMA) hardware, each VM has a default NUMA node preference. Hyper-V uses this NUMA node preference when assigning physical memory to the VM and when scheduling the VM's virtual processors. A VM performs optimally when its virtual processors and memory are on the same NUMA node.

By default, the system assigns the VM to its preferred NUMA node every time the VM is run. An imbalance of NUMA node assignments can occur depending on the memory requirements of each VM and the order in which each VM is started. This can lead to a disproportionate number of VMs being assigned to a single NUMA node.

Use Perfmon to check the NUMA node preference setting for each running VM by examining the \Hyper-V VM Vid Partition (*)\ NumaNodeIndex counter.

You can change NUMA node preference assignments by using the Hyper-V WMI API. To set the NUMA node preference for a VM, set the `NumaNodeList` property of the `Msvm_VirtualSystemSettingData` class.

MEMORY PERFORMANCE

The hypervisor virtualizes the guest physical memory to isolate VMs from each other and provide a contiguous, zero-based memory space for each guest operating system. In general, memory virtualization can increase the CPU cost of accessing memory. On non-SLAT-based hardware, frequent modification of the virtual address space in the guest operating system can significantly increase the cost.

Enlightened Guests

Windows Server 2008 R2 and Windows Server 2008 include kernel enlightenments and optimizations to the memory manager to reduce the CPU overhead from Hyper-V memory virtualization. Workloads that have a large working set in memory can benefit from using Windows Server 2008 R2 or Windows Server 2008 as a guest. These enlightenments reduce the CPU cost of context switching between processes and accessing memory. Additionally, they improve the multiprocessor scalability of Windows Server guests.

Correct Memory Sizing for Child Partitions

You should size VM memory as you typically do for server applications on a physical machine. You must size it to reasonably handle the expected load at ordinary and peak times, because insufficient memory can significantly increase response times and CPU or I/O usage.

You can enable Dynamic Memory to allow Windows to size VM memory dynamically. The recommended initial memory size for Windows Server 2008 R2 guests is at least 512MB. With Dynamic Memory, if applications in the VM experience launching problems, you can increase the pagefile size for the VM. To increase the VM pagefile size, navigate to Control Panel > System > Advanced System Settings > Advanced. From this tab, navigate to Performance Settings > Advanced > Virtual memory. For the Custom Size selection, configure the Initial Size to the amount of memory assigned to the VM by Hyper-V Dynamic Memory when VM reaches its steady state, and set the Maximum Size to three times the Initial Size.

When running Windows in the child partition, you can use the performance counters described in Table 1 in a child partition to identify whether the child partition is experiencing memory pressure and is likely to perform better with a higher VM memory size.

Performance counter	Suggested threshold value
Memory – Standby Cache Reserve Bytes	The sum of Standby Cache Reserve Bytes and Free and Zero Page List Bytes should be 200MB or more on systems with 1GB, and 300MB or more on systems with 2GB or more of visible RAM.
Memory – Free and Zero Page List Bytes	Sum of Standby Cache Reserve Bytes and Free and Zero Page List Bytes should be 200MB or more on systems with 1GB, and 300MB or more on systems with 2GB or more of visible RAM.
Memory – Pages Input/Sec	Average over a 1-hour period is less than 10.

Correct Memory Sizing for Root Partition

The root partition must have sufficient memory to provide services such as I/O virtualization, snapshot, and management to support the child partitions. The root partition should have at least 512MB available. When Dynamic Memory is enabled, the root reserve is calculated automatically based on root physical memory and NUMA architecture. This logic applies for supported scenarios with no applications running in the root.

A good standard for the memory overhead of each VM is 32MB for the first 1GB of virtual RAM plus another 8MB for each additional GB of virtual RAM. This should be factored in the calculations of how many VMs to host on a physical server. The memory overhead varies depending on the actual load and amount of memory that is assigned to each VM.

STORAGE I/O PERFORMANCE

Hyper-V supports synthetic and emulated storage devices in VMs, but the synthetic devices can generally offer significantly better throughput and response times and reduced CPU overhead. The exception is if a filter driver can be loaded and reroutes I/Os to the synthetic storage device. Virtual hard disks (VHDs) can be backed by three types of VHD files or raw disks. This section describes the different options and considerations for tuning storage I/O performance.

Synthetic SCSI Controller

The synthetic storage controller offers significantly better performance on storage I/Os with less CPU overhead than the emulated IDE device. The VM Integration Services include the enlightened driver for this storage device and are required for the guest operating system to detect it. The operating system disk must be mounted on the IDE device for the operating system to boot correctly, but the VM integration services load a filter driver that reroutes IDE device I/Os to the synthetic storage device. NetApp strongly recommends that you mount the data drives directly to the synthetic SCSI controller, because that configuration has reduced CPU overhead. You should also mount log files and the operating system paging file directly to the synthetic SCSI controller if their expected I/O rate is high.

For highly intensive storage I/O workloads that span multiple data drives, each VHD should be attached to a separate synthetic SCSI controller for better overall performance. In addition, each VHD should be stored on a separate physical disk.

Virtual Hard Disk Types

There are three types of VHD files. NetApp recommends that production servers use fixed-sized VHD files for better performance and also to make sure that the virtualization server has sufficient disk space to expand the VHD file at run time. Here are the performance characteristics and trade-offs between the three VHD types:

- **Dynamically expanding VHD.** Space for the VHD is allocated on demand. The blocks in the disk start as zeroed blocks, but they are not backed by any actual space in the file. Reads from such blocks return a block of zeros. When a block is first written to, the virtualization stack must allocate space in the VHD file for the block and then update the metadata. This increases the number of necessary disk I/Os for the

write and increases CPU usage. Reads and writes to existing blocks incur both disk access and CPU overhead when looking up the mapping of the blocks in the metadata.

- **Fixed-size VHD.** Space for the VHD is first allocated when the VHD file is created. This type of VHD is less apt to fragment, which reduces the I/O throughput when a single I/O is split into multiple I/Os. It has the lowest CPU overhead of the three VHD types because reads and writes do not need to look up the mapping of the block.

  Note: NetApp highly recommends thinly provisioned fixed VHDs.

- **Differencing VHD.** The VHD points to a parent VHD file. Any writes to blocks never written to before result in space being allocated in the VHD file, as with a dynamically expanding VHD. Reads are serviced from the VHD file if the block has been written to. Otherwise, they are serviced from the parent VHD file. In both cases, the metadata is read to determine the mapping of the block. Reads and writes to this VHD can consume more CPU and result in more I/Os than a fixed-sized VHD.

Snapshots of a VM create a differencing VHD to store the writes to the disks since the snapshot was taken. Having only a few snapshots can elevate the CPU usage of storage I/Os, but it might not noticeably affect performance except in highly I/O-intensive server workloads.

However, having a large chain of snapshots can noticeably affect performance because reading from the VHD can require checking for the requested blocks in many differencing VHDs. Keeping snapshot chains short is important for maintaining good disk I/O performance.

Pass-through Disks

The VHD in a VM can be mapped directly to a physical disk or logical unit number (LUN) instead of to a VHD file. The benefit is that this configuration bypasses the file system (NTFS) in the root partition, which reduces the CPU usage of storage I/O. The risk is that physical disks or LUNs can be more difficult to move between machines than VHD files.

Large data drives can be prime candidates for passthrough disks, especially if they are I/O intensive. VMs that can be migrated between virtualization servers (such as quick migration) must also use drives that reside on a LUN of a shared storage device.

Disabling File Last-Access Time Check

Windows Server 2003 and earlier Windows operating systems update the last-accessed time of a file when applications open, read, or write to the file. This increases the number of disk I/Os, which further increases the CPU overhead of virtualization. If applications do not use the last-accessed time on a server, system administrators should consider setting this registry key to disable these updates.

```
NTFSDisableLastAccessUpdate
HKLM\System\CurrentControlSet\Control\FileSystem\ (REG_DWORD)
```

By default, Windows Server 2008 R2 disables the last-access time updates.

Physical Disk Topology

VHDs that I/O-intensive VMs use generally should not be placed on the same physical disks because this can cause the disks to become a bottleneck. If possible, they should also not be placed on the same physical disks that the root partition uses.

I/O Balancer Controls

The virtualization stack balances storage I/O streams from different VMs so that each VM has similar I/O response times when the system's I/O bandwidth is saturated. The following registry keys can be used to adjust the balancing algorithm, but the virtualization stack tries to fully use the I/O device's throughput while providing reasonable balance. The first path should be used for storage scenarios, and the second path should be used for networking scenarios:

```
HKLM\System\CurrentControlSet\Services\StorVsp\<Key> = (REG_DWORD)
HKLM\System\CurrentControlSet\Services\VmSwitch\<Key> = (REG_DWORD)
```

Both storage and networking have three registry keys at the preceding `StorVsp` and `VmSwitch` paths, respectively. Each value is a DWORD and operates as follows. NetApp does not recommend this advanced tuning option unless you have a specific reason to use it.

Note: These registry keys may be removed in future releases.

- **IOBalance_Enabled.** The balancer is enabled when set to a nonzero value and disabled when set to 0. The default is enabled for storage and disabled for networking. Enabling the balancing for networking can add significant CPU overhead in some scenarios.
- **IOBalance_KeepHwBusyLatencyTarget_Microseconds.** This parameter controls how much work, represented by a latency value, the balancer allows to be issued to the hardware before throttling to provide better balance. The default is 83 ms for storage and 2 ms for networking. Lowering this value can improve balance but reduces some throughput. Lowering it too much significantly affects overall throughput. Storage systems with high throughput and high latencies can show added overall throughput with a higher value for this parameter.
- **IOBalance_AllowedPercentOverheadDueToFlowSwitching.** This parameter controls how much work the balancer issues from a VM before switching to another VM. This setting is primarily for storage where finely interleaving I/Os from different VMs can increase the number of disk seeks. The default is 8% percent for both storage and networking.

NETWORK I/O PERFORMANCE

Hyper-V supports synthetic and emulated network adapters in the VMs, but the synthetic devices offer significantly better performance and reduced CPU overhead. Each of these adapters is connected to a virtual network switch, which can be connected to a physical network adapter if external network connectivity is needed.

Synthetic Network Adapter

Hyper-V features a synthetic network adapter that is designed specifically for VMs to achieve significantly reduced CPU overhead on network I/O as compared to the emulated network adapter that mimics existing hardware. The synthetic network adapter communicates between the child and root partitions over VMBus by using shared memory for more efficient data transfer.

The emulated network adapter should be removed through the VM Settings dialog box and replaced with a synthetic network adapter. The guest requires the VM integration services to be installed.

Perfmon counters representing the network statistics for the installed synthetic network adapters are available under the counter set `\Hyper-V Virtual Network Adapter (*) \ *`.

Install Multiple Synthetic Network Adapters on Multiprocessor VMs

Virtual machines with more than one virtual processor may benefit from having more than one synthetic network adaptor installed in the VM. Workloads that are network intensive, such as a Web server, can make use of greater parallelism in the virtual network stack if a second synthetic NIC is installed in a VM.

Offload Hardware

As with the native scenario, offload capabilities in the physical network adapter reduce the CPU usage of network I/Os in VM scenarios. Hyper-V currently uses LSOv1 and TCPv4 checksum offload. The offload capabilities must be enabled in the driver for the physical network adapter in the root partition.

Drivers for certain network adapters disable LSOv1 but enable LSOv2 by default. System administrators must explicitly enable LSOv1 by using the driver Properties dialog box in Device Manager.

Network Switch Topology

Hyper-V supports creating multiple virtual network switches, each of which can be attached to a physical network adapter if necessary. Each network adapter in a VM can be connected to a virtual network switch. If the physical server has multiple network adapters, VMs with network-intensive loads can benefit from being connected to different virtual switches to better use the physical network adapters.

Perfmon counters representing the network statistics for the installed synthetic switches are available under the counter set `\Hyper-V Virtual Switch (*) \ *`.

Interrupt Affinity

System administrators can use the IntPolicy tool to bind device interrupts to specific processors.

VLAN Performance

The Hyper-V synthetic network adapter supports VLAN tagging. It provides significantly better network performance if the physical network adapter supports NDIS_ENCAPSULATION_IEEE_802_3_P_AND_Q_IN_OOB encapsulation for both large send and checksum offload. Without this support, Hyper-V cannot use hardware offload for packets that require VLAN tagging, and network performance can be decreased.

VMQ

Windows Server 2008 R2 introduces support for VMQ-enabled network adapters. These adapters can maintain a separate hardware queue for each VM, up to the limit supported by each network adapter.

Because limited hardware queues are available, you can use the Hyper-V WMI API to make sure that the VMs that are using the network bandwidth are assigned a hardware queue.

VM Chimney

Windows Server 2008 R2 introduces support for VM Chimney. Network connections with long lifetimes benefit the most, due to the increase in CPU required for to establish a connection when VM Chimney is enabled.

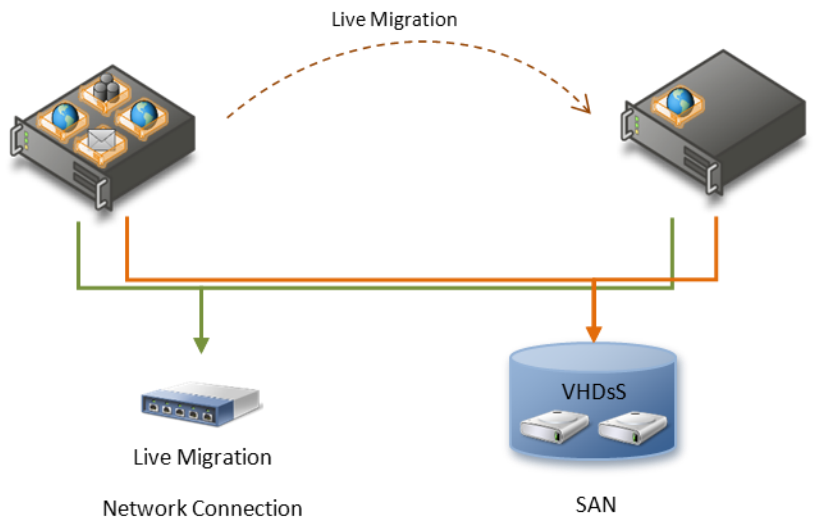
Live Migration

Live migration enables you to transparently move running virtual machines from one node of the failover cluster to another node in the same cluster without a dropped network connection or perceived downtime. In addition, failover clustering requires shared storage for the cluster nodes.

The process of moving a running virtual machine can be broken down into two major phases. The first phase is the copying of the memory contents on the VM from the current host to the new host. The second phase is the transfer of the VM state from the current host to the new host. The duration of both phases is greatly determined by the speed at which data can be transferred from the current host to the new host.

Providing a dedicated network for live migration traffic helps to minimize the time required to complete a live migration, and it ensures consistent migration times.

Figure 8. Example Hyper-V live migration configuration.



In addition, increasing the number of receive and send buffers on each network adapter involved in the migration can improve migration performance.

4.2 MICROSOFT SYSTEM CENTER VIRTUAL MACHINE MANAGER

For IT professionals who are responsible for managing virtual infrastructures, System Center Virtual Machine Manager 2008 R2 (VMM 2008 R2) — a member of the System Center family of management products —is a straightforward and cost-effective solution for unified management of physical and virtual machines. VMM 2008 R2 provides Performance and Resource Optimization (PRO) for dynamic and responsive management of virtual infrastructures, consolidation of underused physical servers, and rapid provisioning of new virtual machines by leveraging the expertise and investments in Microsoft Windows Server technology.

VDI MANAGEMENT

VDI management has different usage patterns compared to server virtualization management. For example, VDI VMs are powered on in the morning when the users need to log on to their desktops and they are powered off in the evening after users log off (all this is controllable via policy). So it's common for large number of VMs to be powered on in a short period of time, which causes a spike in load on the VMM to handle a large number of parallel jobs. These spikes can cause the system to become overloaded; future versions of SCVMM will be improved to handle such scenarios. This section describes best practices for configuring SCVMM 2008 R2 for managing VDI environments. The size of the environment described is around 1,000 desktop VMs; if your environment is larger, you will probably need to use multiple instances of SCVMM.

SYSTEM REQUIREMENTS

This subsection looks at the key system requirements for SCVMM and SQL Server®. For SCVMM system requirements, See "System Requirements" in the System Center Virtual Machine Manager (VMM) 2008R2 documentation.

For managing a VDI environment of 1,000 VMs:

- **SCVMM server.** Quad-core, 2GHz (x64) or larger with 8GB of memory. When running SCVMM server in a VM, 4 Vproc with 8GB of memory, use fixed size VHD.

- **SQL Server.** With SQL Server 2008 R2 Enterprise, NetApp highly recommends using Fibre Channel disks for better I/O throughput and configuring the database and log files to be on different disks. When running SQL Server in a VM, make sure that data and log files for the SCVMM DB are passthrough disks to Fibre Channel disks.

REFRESHERS

SCVMM uses refreshers, which are basically periodic polling to get the latest configuration of hosts, VMs, network, storage, and so on so that the VMM database reflects the "truth" in the data center. Refreshers are necessary because configuration can be changed out of band to SCVMM; that is, users can make changes to the environment by going directly to the host or VM. However, in a controlled environment, the amount of out-of-band changes can be minimized, reducing the frequency of refreshers. Because refreshers take up some amount of system resources, reducing their frequency frees up SCVMM to handle the spike in loads that occurs in VDI scenarios.

Below is a list of refresher intervals and recommended values.

Regkey	Default	Min	Max	Recommended value	Registry value (in seconds)
VMUpdateInterval - Periodic VM refresher	30 min	0	24 hour	120 min	7200
HostUpdateInterval - Host and User Role Refresher	30 min	0	24 hour	120 min	7200
VMPropertiesUpdateInterval – VM light refresher (subset of properties)	2 min	0	24 hour	30 min	1800
VHDMountTimeoutSeconds – Used when multiple VMs are being created in parallel from the same base disk, which causes disk conflicts	10 min	10 min	1 hour	1 hour	

Note:

- Must be created under the key `HKLM\SOFTWARE\Microsoft\Microsoft System Center Virtual Machine Manager Server\Settings`.
- All values must be specified in seconds.
- All values are of type `REG_DWORD`.
- The VMM service must be restarted after changing the reg keys.

GARBAGE COLLECT OLDER JOBS

SCVMM retains jobs in the database for a period of time for auditing purposes. In VDI scenarios, a large number of jobs (start/stop VMs) can be collected in the database, which can result in performance issues, especially when applications try to get job objects when querying for job completion status.

Table 3) <<Add caption.>>

Regkey	Default	Recommended Value
TaskGC	90 (days)	7

Regkey: HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Microsoft System Center Virtual Machine Manager Server\Settings\Sql\TaskGC

WCF TIMEOUT

SCVMM uses Windows Communication Foundation (WCF) for communication between the client PowerShell layer and the server. This channel is used for delivering requests from client and sending events from server to client. When there are a large number of parallel requests, the channel can get overloaded, causing delays that can result in timeouts. NetApp recommends increasing the timeout value to handle high loads.

Regkey	Default	Recommended Value
IndigoSendTimeout	120 (seconds)	300

Regkey: HKLM\Software\Microsoft\Microsoft System Center Virtual Machine Manager Server\Settings\IndigoSendTimeout

VHD MOUNT TIMEOUT

In VDI, it's fairly common to create VMs by using differencing disks that use a common base disk because they share a common golden image. When multiple VMs are created in parallel from the same base disk, SCVMM needs to mount the same base disk for making checks, so a small window exists when failures can occur due to disk conflicts that cause SCVMM to retry the operation. When a large number of VMs are created in parallel, NetApp recommends increasing the timeout interval to reduce the chance of failure.

Regkey	Default	Min	Max	Recommended value	Registry value (in seconds)
VHDMountTimeoutSeconds	10 min	10 min	1 hour		

SERVER-OPTIMIZED GARBAGE COLLECTOR

Enable the server-optimized garbage collector (GC) on the VMM server instead of the default workstation GC. This can significantly reduce the CPU utilization on the VMM server and improve performance for parallel VMM operations.

To enable the server-optimized garbage collector on the VMM server, create a file named `vmmservice.exe.config` and place it in the `%SYSTEMDRIVE%\Program Files\Microsoft System Center Virtual Machine Manager 2008 R2\Bin` directory on the VMM server. The file should contain the following:

```
<configuration>

<runtime>

<gcServer enabled="true"/>

</runtime>

</configuration>
```

4.3 FAILOVER CLUSTERING

The Failover Clustering feature enables you to create and manage failover clusters. A failover cluster is a group of independent computers that work together to increase the availability of applications and services. The clustered servers (called nodes) are connected by physical cables and by software. If one of the cluster nodes fails, another node begins to provide service (a process known as failover). Users experience a minimum of disruptions in service.

This section describes how to use these two technologies together to make a virtual machine highly available. To do this, you create a simple two-node cluster and a virtual machine, and then verify the setup by failing over the virtual machine from one node to the other.

REQUIREMENTS FOR USING HYPER-V AND FAILOVER CLUSTERING

To use the Hyper-V role on a failover cluster with two nodes, you need the hardware, software, accounts, and network infrastructure described in the following subsections.

A new Failover Clustering feature called Cluster Shared Volumes was introduced in Windows Server 2008 R2. With CSV, the configuration of clustered virtual machines is much simpler than before. For information about requirements for using Hyper-V with Cluster Shared Volumes, see Hyper-V: Using Live Migration with Cluster Shared Volumes in Windows Server 2008 R2.

HARDWARE REQUIREMENTS FOR HYPER-V

Hyper-V requires an x64-based processor, hardware-assisted virtualization, and hardware-enforced Data Execution Prevention. Specifically, you must enable the Intel® XD bit (execute disable bit) or AMD NX bit (no execute bit).You can identify systems that support the x64 architecture and Hyper-V by searching the Windows Server catalog for Hyper-V as an additional qualification. The Windows Server catalog is available at http://go.microsoft.com/fwlink/?LinkId=111228.

HARDWARE REQUIREMENTS FOR A TWO-NODE FAILOVER CLUSTER

You will need the following hardware for a two-node failover cluster:

- **Servers.** NetApp recommends using a set of matching computers that contain either the same or similar features.

  Note: Microsoft supports a failover cluster solution only if all the hardware features are marked as "Certified for Windows Server 2008 R2." In addition, the complete configuration (servers, network, and storage) must pass all tests in the Validate a Configuration Wizard, which is included in the Failover Cluster Manager snap-in.

  For information about hardware compatibility for Windows Server 2008 R2, see http://go.microsoft.com/fwlink/?LinkId=139145.

- **Network adapters and cable (for network communication)**. The network hardware, like other features in the failover cluster solution, must be marked as “Certified for Windows Server 2008 R2.” If you use iSCSI, your network adapters should be dedicated to either network communication or iSCSI, not both.

  In the network infrastructure that connects your cluster nodes, avoid having single points of failure. There are multiple ways of accomplishing this. You can connect your cluster nodes by multiple, distinct networks. Alternatively, you can connect your cluster nodes with one network that is constructed with teamed network adapters, redundant switches, redundant routers, or similar hardware that removes single points of failure.

  Note: If you connect cluster nodes with a single network, the network passes the redundancy requirement in the Validate a Configuration Wizard. However, the report from the wizard will include a warning that the network should not have single points of failure.

  For more information about the network configuration required for a failover cluster, see Network infrastructure and domain account requirements for a two-node failover cluster, later in this section.

- **Device controllers and appropriate adapters for the storage**

  - **For SAS or Fibre Channel.** If you are using Serial Attached SCSI or Fibre Channel, in all clustered servers, the mass-storage device controllers that are dedicated to the cluster storage should be identical. They should also use the same firmware version.

    Note: With Windows Server 2008 R2, you cannot use parallel SCSI to connect the storage to the clustered servers. This is also true for Windows Server 2008.

  - **For iSCSI.** If you are using iSCSI, each clustered server must have one or more network adapters or host bus adapters that are dedicated to the cluster storage. The network that you use for iSCSI cannot be used for network communication. In all clustered servers, the network adapters that you use to connect to the iSCSI storage target should be identical, and NetApp recommends using Gigabit Ethernet or a faster network adapter

    Note: You cannot use teamed network adapters, because they are not supported with iSCSI.

  For more information about iSCSI, see the iSCSI FAQ on the Microsoft Web site (http://go.microsoft.com/fwlink/?LinkId=61375).

- **Storage.** You must use shared storage that is compatible with Windows Server 2008 R2.

  Cluster Shared Volumes (CSV) is a feature of failover clusters that is specifically designed to enhance the availability and manageability of virtual machines. Cluster Shared Volumes are volumes in a failover cluster that multiple nodes can read from and write to at the same time. This feature enables multiple nodes to concurrently access a single shared volume. The CSV feature is supported for use only with Hyper-V and other technologies specified by Microsoft.

  On a failover cluster that uses Cluster Shared Volumes, multiple clustered virtual machines that are distributed across multiple cluster nodes can all access their virtual hard disk (VHD) files at the same time, even if the VHD files are on a single disk (LUN) in the storage. This means that the clustered virtual machines can fail over independently of one another, even if they use only a single LUN. When CSV is not enabled, a single disk (LUN) can be accessed by only a single node at a time. This means that clustered virtual machines can fail over independently only if each virtual machine has its own LUN, which makes the management of LUNs and clustered virtual machines more difficult.

  For a two-node failover cluster, the storage should contain at least two separate volumes (LUNs), configured at the hardware level. Do not expose the clustered volumes to servers that are not in the cluster. One volume functions as the witness disk (described later in this section). One volume contains the files that are being shared between the cluster nodes. This volume serves as the shared storage on which you will create the virtual machine and the virtual hard disk. To complete the steps as described in this document, you need to expose only one volume.

  Storage requirements include:

  - To use the native disk support included in failover clustering, use basic disks, not dynamic disks.

  - NetApp recommends formatting the partitions with NTFS (for the witness disk, the partition must be NTFS). If you have a witness disk or use CSVs, the partition for each of those must be NTFS.

    For Cluster Shared Volumes, there are no special requirements other than the requirement for NTFS.

  - For the partition style of the disk, you can use either master boot record (MBR) or GUID partition table (GPT).

A witness disk is a disk in the cluster storage that is designated to hold a copy of the cluster configuration database. A failover cluster has a witness disk only if this is specified as part of the quorum configuration. For this two-node cluster, the quorum configuration is Node and Disk Majority, the default for a cluster with an even number of nodes. Node and Disk Majority means that the nodes and the witness disk each contain copies of the cluster configuration, and the cluster has a quorum as long as a majority (two out of three) of these copies are available.

DEPLOYING STORAGE AREA NETWORKS WITH FAILOVER CLUSTERS

When deploying a storage area network (SAN) with a failover cluster, follow these guidelines:

- **Confirm compatibility of the storage.** Confirm with manufacturers and vendors that the storage, including drivers, firmware, and software used for the storage, are compatible with failover clusters in Windows Server 2008 R2.

  Note: Storage that was compatible with server clusters in Windows Server 2003 might not be compatible with failover clusters in Windows Server 2008 R2. Contact your vendor to make sure that your storage is compatible with failover clusters in Windows Server 2008 R2.

- **Failover clusters** include the following new requirements for storage:

  - Improvements in failover clusters (as compared to server clusters in Windows Server 2003) require that the storage respond correctly to specific SCSI commands. To confirm that your storage is compatible, run the Validate a Configuration Wizard. In addition, you can contact the storage vendor.

  - The miniport driver used for the storage must work with the Microsoft Storport storage driver.

- **Isolate storage devices, one cluster per device.** Servers from different clusters must not be able to access the same storage devices. In most cases, a LUN that is used for one set of cluster servers should be isolated from all other servers through LUN masking or zoning.

- **Consider using multipath I/O software.** In a highly available storage fabric, you can deploy failover clusters with multiple host bus adapters by using multipath I/O software. This provides the highest level of redundancy and availability. For Windows Server 2008 R2, your multipath solution must be based on Microsoft Multipath I/O (MPIO). Your hardware vendor usually supplies an MPIO device-specific module (DSM) for your hardware, although Windows Server 2008 R2 includes one or more DSMs as part of the operating system.

  Note: Host bus adapters and multipath I/O software can be very version sensitive. If you are implementing a multipath solution for your cluster, you should work closely with your hardware vendor to choose the correct adapters, firmware, and software for Windows Server 2008 R2.

SOFTWARE REQUIREMENTS FOR USING HYPER-V AND FAILOVER CLUSTERING

Here are the software requirements for using Hyper-V and the Failover Clustering feature:

- All the servers in a failover cluster must run either the x64-based version or the Intel Itanium® architecture-based version of Windows Server 2008 R2 (nodes in a single failover cluster cannot run different versions).

- All the servers should have the same software updates (patches) and service packs.

- Windows Server 2008 R2 Enterprise or Windows Server 2008 R2 Datacenter must be used for the physical computers. These servers must run the same version of Windows Server 2008 R2, including the

same type of installation. That is, both servers must be either a full installation or a Server Core installation.

- If you do not want to install Windows Server 2008 R2 Enterprise or Windows Server 2008 R2 Datacenter on the test virtual machine, you will need the installation media for the operating system. The instructions in this guide assume that you will install Windows Server 2008 R2 on the virtual machine.

NETWORK INFRASTRUCTURE AND DOMAIN ACCOUNT REQUIREMENTS FOR A TWO-NODE FAILOVER CLUSTER

You will need the following network infrastructure for a two-node failover cluster and an administrative account with the following domain permissions:

- **Network settings and IP addresses.** When you use identical network adapters for a network, you must also use identical communication settings on those adapters (for example, Speed, Duplex Mode, Flow Control, and Media Type must be identical). Also, compare the settings between the network adapter and the switch it connects to and make sure that no settings are in conflict.

  If you have private networks that are not routed to the rest of your network infrastructure, make sure that each of these private networks uses a unique subnet. This is necessary even if you give each network adapter a unique IP address. For example, if you have a cluster node in a central office that uses one physical network, and another node in a branch office that uses a separate physical network, do not specify 10.0.0.0/24 for both networks, even if you give each adapter a unique IP address.

  For more information about the network adapters, see Hardware requirements for a two-node failover cluster, earlier in this section.

- **DNS.** The servers in the cluster must be using Domain Name System (DNS) for name resolution. The DNS dynamic update protocol can be used.

- **Domain role.** All servers in the cluster must be in the same Active Directory® domain. As a best practice, all clustered servers should have the same domain role (either member server or domain controller). The recommended role is member server.

- **Domain controller.** NetApp recommends that your clustered servers should be member servers. If they are, you need an additional server that acts as the domain controller in the domain that contains your failover cluster.

- **Clients.** As needed, you can connect one or more networked clients to the failover cluster that you create and then observe the effect on a client when you move or fail over the highly available virtual machine from one cluster node to the other.

- **Account for administering the cluster.** When you first create a cluster or add servers to it, you must be logged on to the domain with an account that has administrator rights and permissions on all servers in that cluster. The account does not need to be a Domain Admins account; it can be a Domain Users account that is in the Administrators group on each clustered server. In addition, if the account is not a Domain Admins account, the account (or the group that the account is a member of) must be given the Create Computer Objects and Read All Properties permissions in the domain.

  Note: There is a change in the way the Cluster Service runs in Windows Server 2008 R2, as compared to Windows Server 2003. In Windows Server 2008 R2, there is no Cluster Service account. Instead, the Cluster Service automatically runs in a special context that provides the specific permissions and privileges that are necessary for the service (similar to the local system context, but with reduced privileges).

LIMITATIONS FOR USING HYPER-V AND FAILOVER CLUSTERING

Here are the specific limitations for using Hyper-V and the Failover Clustering feature:

- A maximum of 16 nodes is allowed in the failover cluster.
- You can have a maximum of 1,000 virtual machines per cluster for server computer virtualization, with a maximum of 384 on any one node. When Hyper-V is used in conjunction with virtual desktop infrastructure (VDI) for client computer virtualization, you can have a maximum of 1,000 VDI (Windows XP, Windows Vista, Windows 7) virtual machines per cluster, with a maximum of 384 on any one node.
- The number of virtual machines allowed for each node does not change regardless of the size of the cluster.

5 BEST PRACTICES FOR CITRIX XENDESKTOP WITH PROVISIONING SERVER

This section applies to:

Storage administrators

VDI administrators

It is a NetApp best practice to use thin provisioned fixed size VHDs throughout an enterprise virtual environment. Please refer to TR-3483: Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment for more information on how to create thin provisioned VHDs. Engineers have discovered that there are substantial performance hits when using Dynamic VHDs in a SAN environment due to block level misalignment. Please refer to TR-3747: Best Practices for File System Alignment in Virtual Environments for more information regarding this issue.

In light of this, it is a NetApp preference that Citrix Provisioning Server be used when deploying Citrix XenDesktop and provisioning virtual desktops. Due to its use of Dynamic VHDs, Citrix Machine Creation Services (MCS), as a means of provisioning virtual desktops, is not recommended at this time.

5.1 CITRIX XENDESKTOP AND PROVISIONING SERVER OVERVIEW

Citrix XenDesktop offers a next-generation, user-centric desktop virtualization solution that provides a complete system for desktop delivery. For information technology (IT) organizations, XenDesktop greatly simplifies the desktop lifecycle management process and drives down the cost of desktop ownership by separating the delivery of the desktop operating system from applications and user settings. Citrix XenDesktop, in concert with the virtual desktop provisioning capabilities of Provisioning Server for Desktops, is a powerful solution that simplifies the delivery of desktops to end users.

When designing a VDI solution, it is important to take into account how XenDesktop with Provisioning Server will interact with an organization's existing infrastructure and services. These components include directory services, network and security architecture, server hardware types, storage infrastructure, virtualization strategies, and desktop operating system types.

CITRIX XENDESKTOP

Citrix XenDesktop provides a complete virtual desktop delivery solution by integrating several components with advanced configuration tools that simplify the creation and real-time management of the virtual desktop infrastructure.

The core components of XenDesktop discussed in this section are:

- **Desktop Delivery Controller (DDC).** Installed as an infrastructure component on servers, the controller authenticates users, manages the assembly of users' virtual desktop environments, and brokers connections between users and their virtual desktops.

- **Virtual Desktop Agent (VDA).** Installed on each virtual desktop, the agent communicates with the DDC and enables a direct Independent Computing Architecture (ICA) connection between the virtual desktop and the users' endpoint device.

- **Desktop Receiver.** Installed on users' endpoint devices, the Desktop Receiver enables direct ICA connections from endpoint devices to virtual desktops. The XenApp plug-in can be used in place of the Desktop Receiver, but the new features in the ICA toolbar would not be available.

CITRIX PROVISIONING SERVER

The Provisioning Server infrastructure is based on software-streaming technology. Using Provisioning Server, administrators prepare a device (Master Target Device) to be imaged by installing an operating system and any required software on that device. A virtual disk (vDisk) image is then created from the Master Target Device's hard drive and saved to the network (on Provisioning Server or on a back-end storage device). Once the vDisk is available from the network, a target device no longer needs its local hard drive to operate, because it boots directly from the network. The Provisioning Server streams the contents of the vDisk to the target device on demand, in real time. The target device behaves as if it is running from its local drive. Unlike thin-client technology, processing takes place on the target device.

The following components of Citrix Provisioning Server are discussed in this section:

- **Provisioning Server.** A Provisioning Server is any server that has the Stream Service installed. It is used to stream software from vDisks to target devices as needed. In some implementations, vDisks reside directly on the Provisioning Server. In larger implementations, Provisioning Servers obtain the vDisk from a network storage device. Provisioning Servers also retrieve and provide configuration information to and from the Provisioning Server database. Provisioning Server architecture includes options to ensure high availability and load balancing of connections between target devices and their vDisks.

- **vDisk images.** vDisks are disk image files on a Provisioning Server or on a shared storage device. vDisks are configured to be in either Private Image mode, where changes made by the user are kept on the image, or Standard Image mode (read-only), where changes made by the end user are discarded upon shutdown.

- **Target devices**. A device, such as a desktop computer or server, that boots and gets its operating system and software from a vDisk on the network, is considered a target device.

- **Write cache files**. When using a standard image mode, the vDisk is set to read-only mode. Each target device then builds a cache that stores any writes that the operating system needs to perform, such as application streaming, routine application processing tasks, or system paging. Several scenarios and options for storing the write cache files are discussed in section 5.2.

- **Network storage.** vDisks are generally stored on network storage devices (SAN, NAS, Windows File Servers, and so on). Leveraging network storage devices allows redundancy and high availability of the vDisk images.

5.2 CITRIX XENDESKTOP AND PROVISIONING SERVER DEPLOYMENT

This section contains the best practices for a XenDesktop and Provisioning Server deployment. These best practices are divided into the following areas:

- Networking

- Storage

- XenDesktop Desktop Delivery Controller

- Provisioning Server for Desktops

- Virtual desktop images and target devices
- Scalability

Note: This section contains recommendations that require using the Registry Editor. Using the Registry Editor incorrectly can cause serious problems that may require reinstalling the operating system. NetApp cannot guarantee that problems resulting from the incorrect use of the Registry Editor can be solved. Use the Registry Editor at your own risk. Always back up the registry before editing any values.

NETWORKING

Domain Name System

A key infrastructure component used when registering target devices with the DDC is the existing Domain Name System (DNS) environment. When the target VM begins the registration process, the DDC performs a DNS query and tries to communicate with the target device by using the Fully Qualified Domain Name (FQDN) of the registering machine. The target devices, which typically obtain their IP addresses via DHCP, need to be able to dynamically update their Host (A) records in DNS. The DDCs should also be able to update their own DNS records if additional DDCs are brought online, changed, moved, or added to the environment.

Although standard file-based storage such as a hosts file provides proper DNS resolution in proof of concept (POC) environments, NetApp and Citrix recommend implementing Active Directory integrated DNS in enterprise deployments. If the DNS zone is integrated with Active Directory, organizations have the benefit of *secure* dynamic updates, as well as the ability to use Access Control List (ACL) editing features to control which machines can update the DNS system. This enables an organization to specify the computers that are allowed to update their DNS records.

Dynamic updates are a key feature of DNS, allowing domain computers to register their name and IP address with the DNS server automatically when they come online or change IP addresses through the DHCP server. This form of update eliminates the need for manual entry of names and IP addresses into the DNS database. The security aspect comes into play when an automatic update from a client to the DNS database could possibly open an attack vector for a malicious entry. Secure dynamic updates verify that the computer that is requesting the update to the DNS server also has an entry in the Active Directory database. This means that only computers that have joined the Active Directory domain can dynamically update the DNS database.

Note: When using Windows based standard zone storage, the DNS Server service is configured to disallow dynamic updates on its zones by default; therefore dynamic updates must be explicitly allowed for the zone. It is also important to note that reverse lookup zones are used when the virtual desktops are added to a desktop group in the Access Management Console. The IP address of the virtual desktop is mapped to the FQDN using the reverse lookup DNS zone.

Provisioning Server

To provide optimal throughput, NetApp and Citrix recommend using multiple network interface cards (NICs) in Provisioning Server machines. One NIC should be configured for Preboot Execution Environment (PXE) communication on the network. A teamed pair should be configured for streaming the vDisks via the PVS Stream Service. A second teamed pair may be required for network access to enterprise storage systems or file shares. When possible, streaming vDisk data should be isolated from normal production network traffic such as Internet browsing, printing, file sharing, and so on by using dedicated networks or VLANs.

STORAGE

Storage requirements for Provisioning Server implementations greatly depend on the number of vDisk images to be created and maintained, the type of vDisks to be used (Standard Image Mode or Private Image Mode), the operating system installed in the image, what components or applications are installed on each image, plans for potential future application installations, and storage location of the write cache file for each provisioned workstation.

Types of VDisks

When using Standard Image Mode vDisks, many target devices boot from the same vDisk. However, space is required to store only one copy of each vDisk. A second copy should be created and used for updates to the vDisk. When using Private Mode vDisks, each target device must have its own copy of the vDisk, requiring additional storage space for each target device. When using a Private Mode disk as the base image for a Standard Image deployment, do not add the Vdisk to Active Directory if it will be used as the Standard Image. When a previously installed Private Mode image boots the vDisk in Standard Image Mode, the Domain Controller already contains the unique machine information (SID) in Active Directory and an error message occurs. Instead, let PVS manage the Active Directory machine accounts.

VDisk Size Estimates

The size of the vDisk depends largely on the operating system and the number of applications to be installed on the vDisk. The vDisk grows larger as more applications are installed in it. NetApp and Citrix recommend creating vDisks larger than initially necessary to leave room for additional application installations or patches if necessary. Based on past experience, NetApp and Citrix recommend that organizations use the general sizing estimates in Table 6 when determining storage requirements. The estimates were made based on an operating system with only a few applications installed, such as Microsoft Office. These estimates may not accurately reflect the required sizes of vDisks in every environment; each organization should determine the space requirements for their vDisk images individually.

To estimate the size of the vDisk required, follow these guidelines:

- **Target device space.** Examine the properties of the disk to identify the amount of space currently in use by the master target device. A typical installation of Windows 7 with no local applications added could consume between 6 and 10GB of space on a disk drive.

- **Application sizing.** As applications are installed, the amount of space in use increases. If the plan is to add applications to the vDisk image after capture, then the vDisk should be sized to allow for the additional space requirements. NetApp and Citrix recommend adding an additional 50% of space to the initial footprint of the image, allowing room for future application installations.

To minimize the storage space required, NetApp and Citrix recommend the following:

- **Minimize the number of applications on each vDisk.** Minimizing the number of installed applications helps reduce the footprint of each vDisk. In addition to reducing the storage capacity, keeping the vDisk as generic as possible allows an organization to leverage the same vDisk for many workloads. Delivering personalized application sets to the desktop via published applications hosted on XenApp servers or application streaming via Citrix Streaming Server further minimizes the applications on each vDisk.

- **Minimize the number of vDisks.** Each vDisk requires physical disk space equal to the size of the vDisk. Consequently, reducing the number of vDisks reduces the storage space requirements. NetApp and Citrix recommend keeping at least one backup copy of each vDisk to be used to make updates when required. Organizations should consider allocating twice the amount of disk space required for each vDisk that it plans to maintain.

Write Cache File

Each target machine contains a volatile write cache file that is deleted upon each reboot cycle. The size of the cache file for each VM depends on several factors, including the types of applications used, user workloads, and reboot frequency. A general estimate of the file cache size for a provisioned workstation running only text-based 6 applications such as Microsoft Word and Outlook that is rebooted daily is about 300 to 500MB. If workstations are rebooted less often, or if graphic-intensive applications (such as Microsoft PowerPoint®, Visual Studio®, and CAD/CAM applications) are used, cache file sizes can grow much larger. Because the application workload of each environment can vary, NetApp and Citrix recommend that each organization perform a detailed analysis to determine the expected cache file size required for their environment.

Cache File Location

There are several options for storing the cache file for provisioned desktops. Common locations include target-based physical local storage, client disk (virtual machine vDisks), and enterprise storage-based (SAN/NAS) disk environments. Each of these options has benefits and limitations; therefore it is important to evaluate the specific requirements of the organization before determining the cache file location. Here are the benefits and limitations of each option:

- **Physical local storage.** If physical desktops or blade servers are being used, Citrix recommends leveraging the target device's RAM or hard disk to store the cache files. When locating the cache file on a local physical storage device, consider the following:

    - Because accessing RAM is considerably faster than reading and writing data to a hard disk, placing the cache file in RAM yields better performance when compared to a hard disk. Placing the cache file in RAM often provides the best performance for a single-application environment.

    - When RAM is used for the cache file, the cache file is limited in size by the amount of physical RAM available in the target machine. If the cache file is expected to grow larger than the amount of available physical RAM in the device, the device's hard drive should be used instead. If the RAM-based cache file becomes filled, undesired result and errors may occur because there is not enough space to write the cached data.

    - There is generally more space available for the cache file on the hard drive on a target machine than in memory. If the target devices are diskless and do not contain hard drives or enough memory for the required size of the cache file, then the cache files can be stored on a shared enterprise storage device and proxied via the Provisioning Server for Desktops.

- **Client disks (virtual machine vDisks).** When leveraging a virtual machine infrastructure to host the desktops, an organization can also leverage enterprise storage associated with each VM or the VM's allocated RAM for the write cache location. For example, each VM may be created with an additional 1GB disk to store the cache file on the vDisk of each target device. When using virtual machine vDisk-based cache locations, consider the following:

    - Storing the cache file as part of the VM requires additional RAM or disk space equal to that of the cache file to be allocated, in addition to the size required for each unique VM. If the target cache file is hosted on the PVS server, the cache file is not highly available. To enable a cache file that is highly available, use the Provisioning Server (PVS) Proxy Mode model, described next.

    - The PVS Proxy Mode model generally gives good performance when the PVS server contains a Fibre Channel or dedicated HBA card that is connected to a SAN. If Fibre Channel cards are not available, gigabit NICs can be used to connect to CIFS-based network shares. Citrix recommends that the Provisioning Server contain multiple NICs—one teamed pair dedicated for streaming, one teamed pair dedicated for network traffic to the enterprise storage network (iSCSI, CIFS, NAS, and so on), and one configured for PXE traffic.

    - This option offers added resiliency in the event of a failure because only a single VM is affected if the local disk associated with the VM runs out of space.

- For this option, each VM must be able to see the additional disk that is associated with it.

- Using a client disk for the cache file lowers the overall network traffic when compared to the enterprise-server-based option. In addition, a local file is required for full system dump tracing.

- **Enterprise-server-based (PVS Proxy Mode).** The cache file can also be stored on a shared enterprise storage solution (SAN/NAS) accessed via the Provisioning Server. In this case, the Provisioning Server acts as a proxy between the virtual machine and the storage solution. To create HA architectures, at least two Provisioning Servers are required.

  - This solution allows the configuration of PVS in HA mode. If an HA mode failure occurs in the PVS infrastructure, any remaining servers can begin servicing the failed requests without interruption to the target device. If the vDisk file is also placed on the shared storage location, XenMotion features can also be enabled.

  - Proxying the cache traffic through the PVS in this manner affects the network I/O and reduces the scalability of each Provisioning Server. This affects the number of active clients that can be supported by a single PVS architecture.

  - Proper sizing of the storage location is critical in this scenario. If the shared storage location fills up and no disk space remains for cache, all virtual machines may experience performance issues.

  - The write cache must be located on shared storage to benefit from XenServer's HA feature (XenMotion between XenServers).

    For enterprise deployments using HA configurations, Citrix does not recommend storing the cache file locally on the PVS. If the cache file is stored locally on the PVS, each PVS becomes a single point of failure, and HA configurations are not available.

Each organization must determine its individual service-level agreement (SLA) to determine whether an HA configuration is required for its unique desktop use cases. When a small SLA window is required, both XenMotion and PVS in HA mode can be used for a complete fault-tolerant solution. This would require the vDisk and cache file to be located on shared storage at the expense of lowering the overall scalability of the PVS. If a large SLA window is available, then Citrix recommends using locally stored cache files when possible for optimal performance.

Determining Expected Cache File Size

Citrix recommends using a pilot or proof of concept (POC) environment to determine the expected size of the cache files. To determine the file size in the pilot or POC environment, configure the write cache to be located on the Provisioning Server and have several different types of end users (graphical application users, text-based task workers, and so on) work on their provisioned desktops. After several full days of heavy use, the administrators can look at the size of the cache files on the Provisioning Server to get a rough estimate of how large the cache files can grow in the production environment. In the production environment, gracefully rebooting the desktops every day helps reduce the size of the cache file because the files are purged on each reboot cycle.

Reboot Provisioned Workstations Frequently

The write cache file for provisioned workstations can grow quite large, and it continues to grow until the workstation is rebooted. Depending on the applications used and the frequency of reboots, the write cache can grow large enough to affect an organization’s storage solution. The cache file is cleared out upon workstation reboot. Subsequently, the more frequently the workstation is rebooted, the less impact the cache files will have. Citrix recommends rebooting workstations daily if possible. If daily reboots are not possible, Citrix recommends rebooting workstations at least once a week to reduce the storage impact of cache files.

Note: A graceful shutdown or restart is required to clear the cache file. Powering down the machine without a graceful shutdown does not clear out the cache file. XenDesktop allows configuration of logoff behavior, so this process can be automated as well.

XENDESKTOP DESKTOP DELIVERY CONTROLLER

Uninstall Web Interface and IIS

XenDesktop automatically installs IIS and Web Interface as part of the Desktop Delivery Controller (DDC) installation. Most organizations have existing Web server infrastructure that can be leveraged for Web Interface for XenDesktop. Not installing these components helps increase the performance and scalability of the DDCs in a production environment. Organizations that have an existing IIS infrastructure can take one of two approaches:

- During the installation, use the `Setup.exe -nosites` parameter, which will not install Web Interface.
- If the installation has already occurred, consider uninstalling this component or disabling the associated services.

Separate the Farm Master and Controller

By default, in a XenDesktop farm the initial Desktop Delivery Controller (DDC) installed is the farm master, which has specific duties as the data collector, performing desktop resolution operations during desktop launches and managing the hosting infrastructure. This single server has the role of Data Collector and Controller. When there are multiple servers in a farm, it is often desirable to separate the functions of the Controller and the Data Collector by delegating these functions to different servers. These duties can be better performed by other DDC machines in the farm, leaving the farm master machine to concentrate on its own role requirements. To separate the roles requires two kinds of actions: make sure that a particular machine is chosen to be the farm master; and make sure that unnecessary duties are not assigned to that machine.

Each DDC in the farm can potentially become the farm master in an election process. This process can be influenced by settings on the various server machines, and one or more machines can be configured to be the preferred farm master, while other machines are configured to take on the farm master role only if the preferred machines are not available. This preference indication is achieved by setting registry entries on the various server machines. Each machine can be configured to have one of three settings:

- **Master.** Servers with this setting are preferentially chosen as the farm master.
- **Backup.** Servers with this setting are preferentially chosen as the farm master when the master server is unavailable.
- **Member.** Servers with this setting are normally not the farm master but can assume the farm master role when none of the master or backup servers is available.

The desktop launch request is the sole function of the DDC master; it handles the logic of launching connections. When configuring Web Interface to point to a DDC member, Web Interface forwards the launch request to the DDC master. If the DDC master server becomes unavailable, an election occurs and another member server assumes the DDC master role. Once the DDC has established the connection to the virtual desktop, the ICA establishes a direct connection between the virtual desktop and the endpoint (client machine). The DDC is no longer in the line of traffic, so the demand on it is diminished and it can scale sufficiently.

When using multiple DDCs, Citrix recommends dedicating a machine as the master controller and configuring all other servers as member servers. Web Interface servers should be configured to point to

member servers to minimize the CPU load placed on the controller server during periods of heavy logon. To configure a server with these settings, edit the following registry keys and restart the server:

```
HKLM\Software\Citrix\IMA\RUNTIME\UseRegistrySetting

DWORD=UseRegistrySetting

Value=1

HKLM\Software\Citrix\IMA\RUNTIME\MasterRanking

DWORD=Value

Value= 1 indicates 'Master'

       2 indicates 'Backup'

       3 indicates 'Member'
```

For more information, see Citrix Knowledge Base Article CTX117477.

Throttle Commands to Virtual Machine Hosting Infrastructure

When a large number of power on/off commands are sent to the virtual machine hosting infrastructure (Microsoft SCVMM), it could become overwhelmed and unresponsive for a short period of time while all of the requests are queued and processed. By default, the communication between the pool management service on the DDC and the hosting infrastructure is throttled to 10% of the total pool. For example, if there are 500 VMs in the desktop group, only 50 VM power operation requests are sent at a time. When pools grow even larger (more than 1,000 desktops), this could result in approximately 100+ power on/off requests being sent concurrently to the hosting infrastructure.

The hosting infrastructure can become overloaded during the following scenarios:

- When the idle pool count significantly increases for a concurrent peak time.
- When the time between logout and login events is short. For example, suppose that a company employs three shifts a day. A group of users logs in in the morning and logs out when their shift is finished. The next shift of users begins to log in at the same time that the previous group logs out, resulting in a very small window (5 to 15 minutes) during which large numbers of users log out and log in.

It may be necessary to throttle the number of power requests sent to the virtual machine infrastructure at a single time. To modify the number of concurrent requests, edit the following configuration on each DDC. This example throttles the number to 20 concurrent commands.

- Open `C:\Program Files\Citrix\VmManagement\CdsPoolMgr.exe.config`.
- Add the line in bold type:

```
<?xml version="1.0" encoding="utf-8" ?>

<configuration>

<appSettings>

<add key="LogToCdf" value ="1"/>
<add key="LogFileName" value ="C:\cdslogs\VMManager.log"/>
```

```
<add key="LogDebug" value="1"/>

<add key=”MaximumTransitionRate” value=”20”/>

</appSettings>

</configuration>
```

- Save the file and restart the DDC. The DDC or the Pool Management Service must be restarted for the new value in the `.NET` configuration file to be read by the DDC.

The value `”MaximumTransitionRate” value=”20”` in this example should be considered only as a point of reference when configuring the concurrent command values. This value varies based on the unique hardware platform and use case of each environment. Citrix recommends that each organization properly test the configuration before determining the optimal balance between the number of concurrent commands that can be serviced and the performance and responsiveness of the hosting infrastructure when sending power commands from the DDC.

PROVISIONING SERVER FOR DESKTOPS

Disable Checksum Offloading on Network Adapter

Checksum Offload parameters are not compatible with the Provisioning Server network stack and may cause slow performance when enabled on the physical network adapter.

Symptoms of slow performance can include:

- Excessive number of retries
- Slow ICA performance in virtual machines
- Freezing or locking when using Windows XP Service Pack 2 virtual machines
- VDisk retries during normal operation when a Windows XP Service Pack 2 virtual machine is already online and in waiting mode
- Slow hosted application launch when using Presentation Server 4.5 and publishing applications in XenDesktop

Citrix recommends disabling Checksum Offload on the network adapter of both the Provisioning Server and the target devices. On many NICs, checksum offloading can be disabled by opening the NIC Properties page and selecting the Advanced Configuration tab.

Some NICs do not offer this setting in the Properties page (that is, target devices running on virtual machines). To change the Checksum Offload parameter value, create (if it does not already exist) and edit the following registry key:

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters

WORD = DisableTaskOffload

Value = 1
```

Note: This registry value may not exist by default on some systems and may need to be created. For more information, see Microsoft Knowledge Base Article 904946.

Disable TCP Large Send Offload

The TCP Large Send Offload option allows the TCP layer to build a TCP message up to 64KB long and send it in one call via IP and the Ethernet device driver. The adapter then resegments the message into multiple TCP frames for wire transmission. The TCP packets sent on the wire are either 1500 byte frames for a maximum transmission unit (MTU) of 1500 or up to 9000 byte frames for an MTU of 9000 (jumbo frames). Resegmenting and queuing packets to send in large frames can cause latency and timeouts to the Provisioning Server and therefore this option should be disabled on all Provisioning Servers and target devices.

To disable the Large Send Offload option, open the NIC Properties page and select the Advanced Configuration tab.

Some NICs do not offer this setting in the Properties page (that is, target devices running on virtual machines). To disable the Large Send Offload parameter, create and edit the following registry key:

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\BNNS\Parameters\

DWORD = EnableOffload

Value "0"
```

For more information, see Citrix Knowledge Base Article CTX117374.

Disable Auto Negotiation

Auto Negotiation can cause long booting times and PXE timeouts, especially when booting multiple target devices. Citrix recommends hard-coding all Provisioning Server ports (server and client) on the NIC and on the switch port to disable the Auto Negotiation feature and configure the connection speed.

Disable Spanning Tree or Enable PortFast

With Spanning Tree Protocol (STP) and Rapid Spanning Tree Protocol, the ports are placed in a "blocked" state while the switch transmits Bridged Protocol Data Units (BPDU) and listens to make sure that the BPDUs are not in a loopback configuration. The amount of time it takes for this process to converge depends on the size of the switched network; it might allow the Preboot Execution Environment (PXE) to time out, causing the VM to enter a wait state or reboot until the condition is cleared and the PXE process can resume. To resolve this issue, disable STP on edge ports that are connected to clients or enable PortFast or Fast Link, depending on the managed switch brand. Refer to Table 7.

Maximize the System Cache

All traffic that occurs between the vDisk and the target device passes through the Provisioning Server regardless of where the vDisk resides. Using Windows Server 2003 file caching features can improve vDisk deployment efficiency. The operating system caches the file reads and write data operations at the block level. When a single target device is booted from a shared vDisk, subsequent clients do not require disk read I/O to perform similar operations. The OS caching mechanism caches only the blocks that were accessed, not the entire vDisk file. The file read data stay in the cache until it is flushed to make space for new data. To increase the speed at which the vDisk is streamed, the Provisioning Server should be optimized for file caching so that the file cache is contained in server RAM rather than increasing disk I/O by reading the files from the server's hard disk. To maximize the system cache on the Provisioning Server, follow these steps:

1. Right-click My Computer and select the Advanced tab.
2. Click the Settings button in the Performance section of the Advanced tab.
3. Select the Advanced tab in the Performance Options window.
4. Make sure that System Cache is selected in the Memory Usage section.

Alternatively, the system cache value can be configured through the registry by using the following key:

```
KEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\Session
Manager\MemoryManagement

DWORD=LargeSystemCache

Value=1
```

For more information, see Microsoft Knowledge Base Article 837331.

Team NICs for Increased Throughput

Network I/O on the Provisioning Server can be a limiting factor in the scalability of the server. Teaming two NICs for throughput provides the server with a maximum of 2Gb of network I/O, increasing the network performance and helping to alleviate this potential bottleneck. In addition, teaming the NICs eliminates a single point of failure if only one NIC is enabled.

Isolate Streaming Traffic

When possible, vDisk streaming traffic should be isolated from normal production network traffic such as Internet browsing, printing, file sharing, and so on. In this scenario it is best to use multiple NICs, one for PXE and teamed NICs for streaming the vDisks to target machines.

Make a Copy of Each vDisk

In order to modify images created with Provisioning Server, NetApp recommends saving a copy of each vDisk and placing it in Private Image mode. By keeping a copy of the vDisk in this mode, the administrator can modify the vDisk to include any required updates without affecting the production image. When all required modifications have been made, the administrator can use the Provisioning Server Console to configure the target devices to boot from the newly created vDisk image. NetApp also recommends backing up each vDisk for disaster recovery and business continuity purposes.

5.3 SCALABILITY

When considering XenDesktop scalability, it is important to examine the scalability of the constituent parts of XenDesktop rather than just viewing XenDesktop scalability in general. This section describes scalability for the Desktop Delivery Controller, for Provisioning Server, and for the virtual machine infrastructure. It also discusses the factors that affect scalability for the Desktop Delivery Controller and Provisioning Server.

DESKTOP DELIVERY CONTROLLER

The heaviest toll on the Desktop Delivery Controller occurs during peak connection times, such as when users simultaneously log on in the morning or during shift changes. Most of the logon load on the server is caused by IMA, the Desktop Delivery Controller, and the pool management services. Upon connection completion, the ICA protocol connects the endpoint and the virtual desktop – the DDC is engaged only to receive heartbeat notifications from the virtual desktops. These desktop notifications are far less taxing than the activity at times of peak logon.

When considering large deployments, it is important to note that there are several services that run only on the DDC Zone Master and that all connection requests from end users go through the Zone Master. As a result, the master cannot be scaled out by adding servers. It can only be scaled up by upgrading to a more robust server (more processing power). However, a XenDesktop farm can be scaled out for *failover* by adding new DDCs, which handle the XDA keep-alives and registration with the virtual desktops. The Zone Master can become a bottleneck when desktop groups grow large. Therefore Citrix recommends limiting the number of desktops per group to 3,000.

The most common factor that limits scalability of a single DDC is processing power. In order to maximize the number of simultaneous connections a single DDC can handle, NetApp recommends that organizations dedicate servers with the maximum processing power (dual socket or higher with quad core processors) to serve as the Zone Masters in a XenDesktop environment.

PROVISIONING SERVER

The number of target devices that can be supported per Provisioning Server depends on the size of the vDisk, the storage solution for the vDisk placement, the location of the write cache file, and the end users' workflow. The most common bottlenecks that affect scalability of the Provisioning Server are network I/O of the Provisioning Server, disk I/O of the vDisk storage location, and cache file location.

The Provisioning Server Streaming Service manages traffic (that is, a proxy). If the vDisk is located on a share, then the Provisioning Server retrieves the image and distributes it to the target devices (virtual desktops). Therefore, as the number of target devices increases per Provisioning Server, the Provisioning Server's NIC is heavily used and becomes saturated at peak intervals. Teaming multiple NICs can help increase the number of target devices that can be supported before the NIC becomes a bottleneck.

Disk I/O of the storage solution can become a bottleneck as well, because the Provisioning Server (or servers) reads from the disk to obtain the vDisk and cache files. The faster the network storage solution (SAN, NAS, Windows DFS, and so on) can output the data from the hard disk, the better the performance of the Provisioning Server will be, thereby increasing the number of devices that a single Provisioning Server can support without reducing performance. Additionally, scalability and throughput can be affected by the location of the vDisks and cache files in the back-end storage system. When files are stored on the same storage system, NetApp and Citrix recommend placing the vDisks and write cache files on separate dedicated LUNs to decrease and distribute disk spindle loads and increase vDisk and cache file access speeds.

As discussed in section x.x, locating the write cache on the Provisioning Server reduces the scalability of each Provisioning Server by increasing its processor, network I/O, and disk I/O requirements.

NetApp and Citrix recommend that each organization perform scalability testing specific to its environment based on the existing infrastructure and use cases to determine the number of target devices that can be supported by a single Provisioning Server. Additional Provisioning Servers can be added to the architecture to distribute the load, as well as to provide redundancy and high availability. NetApp and Citrix recommend implementing additional Provisioning Servers as part of an organization's HA and disaster recovery plans.

6 PROVISIONING

This section applies to:

Storage administrators

VDI administrators

Virtual machine configuration administrators

6.1 VIRTUAL MACHINE PROVISIONING

Virtual infrastructure solutions, such as Microsoft Hyper-V, empower IT organizations to rapidly deploy virtual machines in all phases: development, test, and production. The tasks involved to deploy virtual machines usually generate many physical copies of virtual machine images, which demand more storage resources to maintain the many virtual machine instances and management resources to execute the many manual steps required to deploy these virtual machines individually.

Integration of Microsoft virtual environments with NetApp storage technology can solve these challenges by helping organizations to reduce the effort spent in deploying individual virtual machines and to reduce the amount of storage required to support the deployment of individual virtual machines, which helps to reduce the costs associated with space, power, and cooling. Use of NetApp Snapshot and FlexClone technology, along with NetApp deduplication, can support the rapid deployment of tens, hundreds, or thousands of virtual machines in minutes while minimizing the total storage supporting such a deployment by 50% or more when compared to a baseline of traditional storage.

Note: NetApp best practice for provisioning virtual desktops is to use automation such as PowerShell scripts. To find examples of PowerShell scripts, see TR-XXXX.

7 SUMMARY

Microsoft Hyper-V with Citrix XenDesktop on NetApp offers customers several methods of providing storage to VMs. All of these storage methods give customers flexibility in their infrastructure design, which in turn offers cost savings, increased storage use, and enhanced data recovery.

This technical report is not intended to be a definitive implementation or solutions guide. Expertise may be required to solve user-specific deployment issues. Contact your NetApp representative to make an appointment to speak with a NetApp Hyper-V solutions expert.

Comments about this technical report are welcome. Feel free to contact the authors by sending an e-mail to xdl-vgibutmevmtr@netapp.com. Refer to TR-4042 in the subject line of your e-mail.

8 DOCUMENT REFERENCES

NETAPP REFERENCES

Data ONTAP Network and File Access Management Guide
http://now.netapp.com/NOW/knowledge/docs/ontap/

High-Availability System Configuration
http://www.netapp.com/us/solutions/infrastructure/data-protection/high-availability.html

NetApp Deduplication for FAS: Deployment and Implementation Guide
http://media.netapp.com/documents/tr-3505.pdf

NetApp Support Site
http://now.netapp.com/

OnCommand Plug-In 3.0 for Microsoft Installation and Administration Guide
http://now.netapp.com/NOW/knowledge/docs/oncommand_plugin_msft/ocplugin30/pdfs/install.pdf

OnCommand Plug-In 3.0 for Microsoft Release Notes
http://now.netapp.com/NOW/knowledge/docs/oncommand_plugin_msft/ocplugin30/pdfs/rnote.pdf

Rapid Provisioning and Cloning Command Reference Guide
http://now.netapp.com/NOW/knowledge/docs/oncommand_plugin_msft/ocplugin30/pdfs/rnote.pdf

TR-3437: Storage Subsystem Resiliency Guide
http://media.netapp.com/documents/tr-3437.pdf

TR-3450: Active-Active Controller Overview and Best Practices Guidelines
http://media.netapp.com/documents/tr-3450.pdf

TR-3701: NetApp and Microsoft Virtualization: Solution and Implementation Guide

http://media.netapp.com/documents/tr-3701.pdf

TR-3702: NetApp Storage Best Practices for Microsoft Virtualization and NetApp SnapManager for Hyper-V
http://media.netapp.com/documents/tr-3702.pdf

MICROSOFT REFERENCES

Configuring Disks and Storage
http://technet.microsoft.com/en-us/library/ee344823(WS.10).aspx

Configuring Virtual Networks
http://technet.microsoft.com/en-us/library/cc816585(WS.10).aspx

Frequently Asked Questions: Virtual Hard Disks in Windows 7 and Windows Server 2008 R2
http://technet.microsoft.com/en-us/library/dd440865(WS.10).aspx#fixed)

How to use the Sysprep tool to Automate Successful Deployment of Windows XP
http://support.microsoft.com/kb/302577

Hyper-V: Using Hyper-V and Failover Clustering
http://technet.microsoft.com/en-us/library/cc732181(WS.10).aspx

Hyper-V: Using Live Migration with Cluster Shared Volumes in Windows Server 2008 R2
http://technet.microsoft.com/en-us/library/dd446679(WS.10).aspx

Hyper-V Live Migration Overview and Architecture
http://www.microsoft.com/download/en/details.aspx?displaylang=en&id=12601

Hyper-V Virtual Machine Snapshots: FAQ
http://technet.microsoft.com/en-us/library/dd560637(WS.10).aspx

NVSPBIND
http://archive.msdn.microsoft.com/nvspbind

Operations Manager 2007 R2 Supported Configurations
http://technet.microsoft.com/en-us/library/bb309428.aspx

Planning for Disks and Storage
http://technet.microsoft.com/en-us/library/dd183729(WS.10).aspx

SCVMM BEST PRACTICES

http://blogs.technet.com/b/vishwa/archive/2011/02/01/tuning-scvmm-for-vdi-deployments.aspx

The Windows Trace Session Manager Service Does Not Start and Event ID 7000 Occurs

http://support.microsoft.com/kb/839803.

How to Suppress the Windows Tour Prompt in Windows XP

http://support.microsoft.com/kb/311489.

How to Enable and Disable System Restore

http://support.microsoft.com/kb/264887.

SDelete v1.6

http://technet.microsoft.com/en-us/sysinternals/bb897443

CITRIX REFERENCES

Best Practices for Citrix XenDesktop with Provisioning Server
http://support.citrix.com/article/ctx119849

Separating the Roles of Farm Master and Controller in the XenDesktop Farm
Citrix Knowledge Base Article CTX117477

Best Practices for Configuring Provisioning Server on a Network
Citrix Knowledge Base Article CTX117374.

9 VERSION HISTORY

Version	Revision Date	Revision Comments
1.0	Feb 2012	Original document

10 ACKNOWLEDGMENTS

10.1 ABOUT THE AUTHORS AND CONTRIBUTORS

The authors of this technical report are members of the NetApp Microsoft Business Group based in Bellevue, Washington.