Hewlett Packard
Enterprise

HPE StoreVirtual Architecture

Simplify and protect your IT environment with
a scalable, highly available, efficient storage solution

Technical white paper

Contents

Executive summary

HPE StoreVirtual technology, built on the LeftHand operating system, delivers simple, open standards storage platform solutions that require no special IT training to deploy and manage. StoreVirtual is designed to run on most x86-based hardware, with multiple hypervisor options to match business demands for rapid ROI and investment protection. Software-defined architecture gives StoreVirtual the flexibility to be deployed as a virtual machine, a hyper-converged appliance, a dedicated storage array, or a hybrid-cloud building block. The comprehensive enterprise-class feature set includes sophisticated storage management capabilities that enhance the fault-tolerant, shared storage infrastructure. The StoreVirtual architecture lets customers scale-out their IT environment as business needs evolve. This agile and scalable approach unlocks the full benefits of server virtualization.

This white paper describes the key features of StoreVirtual technology and the advantages of a simple, scalable, efficient storage architecture that keeps data highly available and protected. Additional information and recommendations for storage deployments are available in the technical papers and best practice guides listed in the reference section at the end of this white paper.

Five key storage architecture considerations

From a management perspective, single vendor, open platform, converged architectures are optimal, whether a business is seeking build-it-yourself storage, hyper-converged data-center-in-a-box solutions, or a combination of the two. While dedicated hardware devices are still viable (especially for low latency workloads), siloed devices can be more expensive and often require specially trained administrators and specialists.

To reduce complexity in IT environments and keep resources in line, many administrators approach new storage technology with these five questions in mind:

1. **Is it simple?** Simplicity is core to a modern storage environment—simple integration, deployment, management, and upgradeability. Data centers have evolved to host a myriad of technologies, platforms, and devices. Storage solutions must be adaptable, so they can be readily deployed alongside existing technology.
2. **Is it easily scalable?** Traditional "scale up" storage architectures can be difficult to scale. They require customers to determine performance and capacity needs well ahead of time, and manage controller growth to make sure business needs don't outpace the system's capabilities. It's not always feasible to have an expert onsite to stay ahead of possible bottlenecks as data requirements grow.
3. **Does it offer high availability?** Resilient shared storage that provides continuous availability across racks, computer rooms, or campuses is critical to business continuity and revenue streams.
4. **Is it efficient?** One of the driving forces behind virtualization is efficiency: the need to trim operating expenses while increasing ROI and availability. Storage administrators are being asked to deploy more applications with better performance while using less of the data centers' limited space, power, and cooling resources.
5. **Will it effectively protect business data?** Hardware failures due to human error, power outages, or even natural disasters are an ongoing threat to data. Storage systems must be capable of recovering data quickly to protect it for mission-critical business activities.

Storage architecture designed for virtualized environments

Innovative and flexible, StoreVirtual technology facilitates fast-paced virtualized environments that support sophisticated applications running on powerful systems based on multi-core processors. The proliferation of virtual machines (VMs) has driven up the ante, since VMs consolidate workloads onto servers, increasing server utilization levels and placing greater demands on storage performance and throughput.

A dynamic architecture based on software-defined storage (SDS)—where storage functionality is delivered as software, not delivered in a specific chassis—helps reduce cost and complexity in the data center. To address challenges faced by midsize businesses and remote and branch offices, and to address the specific needs around variable, high volume workloads such as VDI, businesses are turning to scalable, resilient SDS designed for virtualized environments.

The shared software-defined storage created by StoreVirtual presents out to the network as an iSCSI storage target. This allows any physical or virtual server in the environment to connect to the internal storage via the Ethernet network, providing administrators with flexible options for growth.

For example, a remote office that currently uses DAS storage may grow to need more than just a single server. If that server is not obsolete, one option would be to purchase an additional server, install StoreVirtual VSA on both, and turn DAS storage into shared storage. A second option is to replace the original server with an HPE Hyper Converged system that provides multiple servers and storage pre-configured in an appliance form-factor. Using this option, the infrastructure can be set up in minutes. In either case, the remote office infrastructure moves from siloed storage to shared storage that can be replicated off-site using the StoreVirtual Remote Copy feature.

Open-standards platform

StoreVirtual technology takes a comprehensive, converged approach that allows customers to construct highly available, fault-tolerant data centers using simple, scalable building blocks. Built on an open-standards platform, StoreVirtual VSA software can run on any modern x86-based hardware in VMware vSphere, Microsoft Hyper-V, and Linux KVM hypervisors. The same resilient StoreVirtual technology is built into HPE Hyper Converged appliances.

As shown in Figure 1, StoreVirtual technology components include targets (storage systems and storage clusters) in a networked infrastructure. Servers and virtual machines act as initiators with access to the shared storage. The LeftHand OS leverages industry-standard iSCSI protocol over Ethernet to provide block-based storage to application servers on the network.

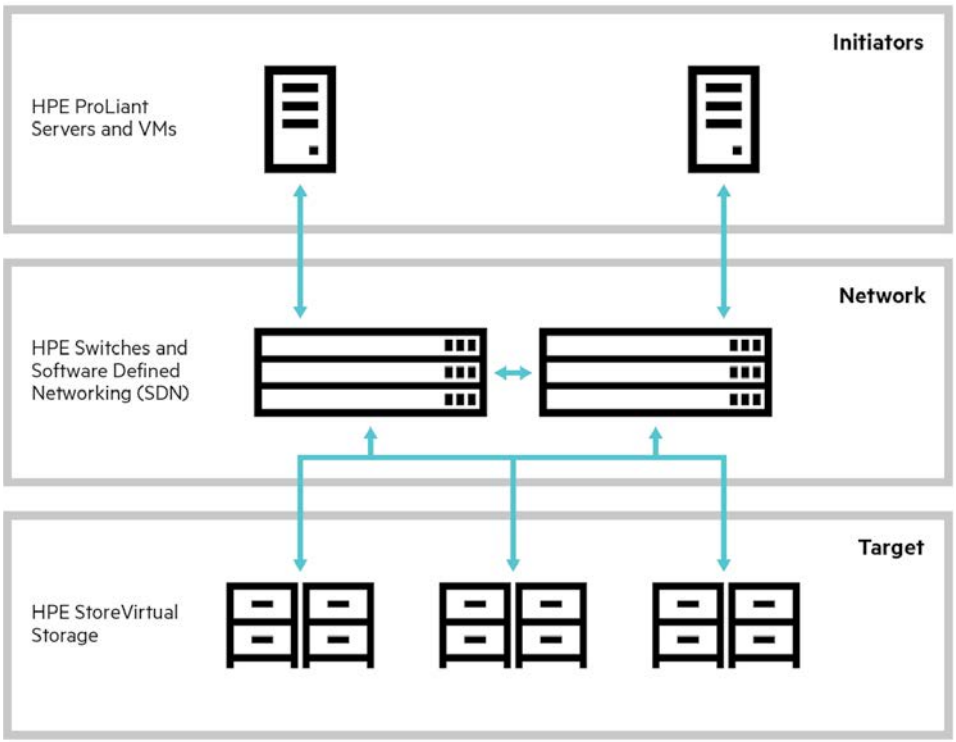

Figure 1. HPE StoreVirtual architecture overview

Nodes, clusters, and management groups

Originally introduced by LeftHand Networks prior to the acquisition by Hewlett-Packard, LeftHand OS is the operating system behind the StoreVirtual architecture. It provides a resilient, highly available foundation that is easy to deploy, integrate with applications, and manage. Leveraging IP networking and iSCSI protocol, the LeftHand OS supports a distributed, scale-out architecture where storage capacity can be expanded incrementally by attaching additional storage nodes to existing clusters on the network.

A storage node is a server that virtualizes its direct-attached storage. In the case of StoreVirtual nodes, each one includes controller functionality—no external controllers needed. Storage nodes require minimal user configuration to bring them online as available systems within the Centralized Management Console (CMC) or as part of the hyper-converged system installation.

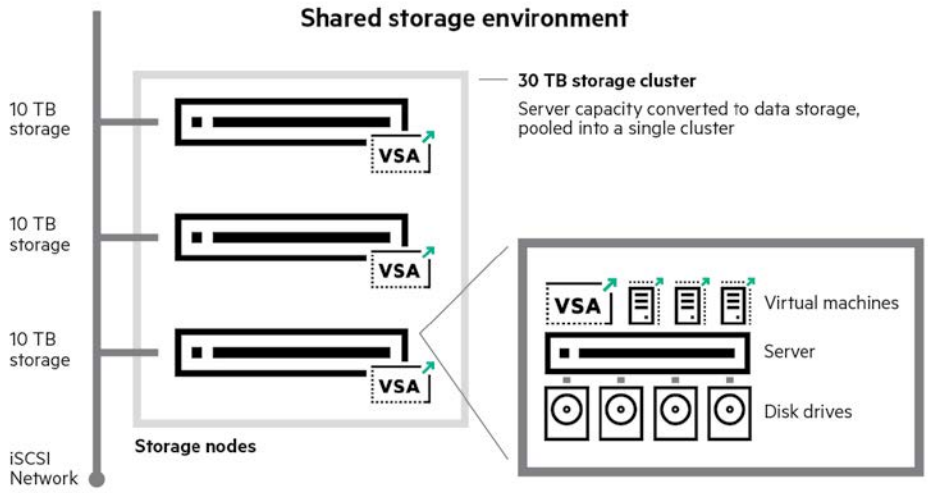

Figure 2. StoreVirtual storage clustering in a virtualized environment

SDS separates hardware and data services from management to optimize costs. IT administrators can aggregate two or more storage nodes into a flexible pool of storage, called a storage cluster (Figure 2). Multiple clusters can be aggregated together into management groups. Volumes, clusters and management groups within the shared storage architecture can be managed centrally through a single console.

The SDS portfolio

The StoreVirtual portfolio offers multiple options for deployment. Users can choose turnkey appliances or mix and match individual products to fit their environment, from hardware agnostic virtual software appliances and compact arrays to hyper-converged systems and data services that are built into a private cloud. The StoreVirtual portfolio includes:

- **HPE Hyper Converged systems:** Pre-configured compute and storage appliances
- **HPE StoreVirtual VSA software:** Flexible, software-defined storage in a virtual storage appliance
- **HPE StoreVirtual arrays:** Purpose-built storage appliances
- **HPE Helion OpenStack software with StoreVirtual VSA:** Built-in data fabric for cloud deployments

When mission-critical support and expert services are factored in, this portfolio delivers on all elements of the software-defined data center—compute, networking, storage, and management—offering a complete SDS strategy and vision based on simplicity, efficiency, and openness that make storage availability a top priority.

The advantages of StoreVirtual technology

One key advantage of StoreVirtual architecture is the all-inclusive feature set. These enterprise-class features are included across the StoreVirtual product portfolio, providing simple, scalable, highly available, and efficient storage solutions:

- Simple deployment and management features
- Flexible hypervisor integration and multi-pathing options
- Pay-as-you-grow scalability
- Peer Motion data mobility
- Network RAID synchronous replication
- Failover management
- Thin provisioning and Space Reclamation
- Adaptive Optimization
- Data replication with Remote Copy, snapshots and SmartClone
- Multi-site stretch clusters, Site Recovery Manager and Split Networks

This agile and scalable approach unlocks the full benefits of server virtualization, particularly for remote and branch offices, private cloud and VDI deployments.

Simple deployment & management

StoreVirtual requires no special IT training to deploy and manage, and the open standards platform and straightforward deployment features make it simple to build a mixed, shared storage infrastructure from the ground up. By integrating StoreVirtual technology into an existing infrastructure, or even installing the VSA software on existing servers, administrators can add new resources seamlessly. They can use the Centralized Management Console (CMC), or the existing hypervisor management tool via HPE OneView plug-ins, to manage the resources through a central console.

Simple deployment features

HPE provides choices for deploying StoreVirtual technology depending on the appliance, size of the environment, and skill set of the data center administrator. The “Zero to VSA” feature automates the installation process for easy deployment of StoreVirtual virtual storage appliances. HPE StoreVirtual VSA Installer for VMware vSphere can virtualize storage and present it as VMFS datastore or RDM to standalone vSphere hosts and vSphere hosts managed by VMware vCenter.

As an alternative, administrators can deploy StoreVirtual VSA in a VMware environment, where the IT generalist or remote administrator will benefit from features contained in HPE OneView for VMware vCenter. HPE OneView can install VSA and provision storage from StoreVirtual. The VMware administrator accesses management capabilities using the vSphere Web Client.

HPE also provides a standalone VSA installer for Microsoft environments. StoreVirtual VSA can be installed through Microsoft System Center Virtual Machine Manager (SCVMM) with the HPE Storage UI add-in that is available from OneView for Microsoft System Center. The UI add-in interface leads the administrator through a sequence of steps to deploy VSA instances to the Hyper-V servers managed by SCVMM. The installation wizard will prompt for host names, network configuration, and the name of the virtual machine. It will also help to define the storage and tiering configuration for the StoreVirtual instance.

HPE Hyper Converged Systems are pre-configured infrastructure building blocks comprised of StoreVirtual technology on HPE ProLiant Server hardware, providing compute and storage in a compact footprint. Using HPE OneView InstantOn, the entire "datacenter-in-a-box" can be deployed in as little as 15 minutes. Once deployed, the solution can be scaled non-disruptively using additional HPE Hyper Converged systems or other StoreVirtual technology appliances.

HPE Helion provides sample configuration files that describe infrastructure components. A simple, straightforward modification to this sample file is required for HPE Helion to build and deploy the StoreVirtual instance in the chosen configuration.

Application consistent point-in-time states

Application integration ensures straightforward deployment and data management for VMware vSphere and Microsoft Windows environments, and mitigates risk by allowing data to be backed up from an application-consistent state. The HPE StoreVirtual Application Aware Snapshot Manager (AASM) manages snapshots for StoreVirtual volumes after they have been quiesced, thereby creating application-consistent, point-in-time states for reliable recovery. Without this integration option, snapshots are crash-consistent only—in-flight cached data may not have been fully written and may not represent an application recovery point.

To create application-managed snapshots in Microsoft Windows and VMware vSphere environments, the StoreVirtual AASM must first be integrated into the platform-specific methods to quiesce volumes (see Figure 3). The instant nature of StoreVirtual snapshots helps to shrink backup windows when backups are coordinated through AASM. These snapshots are similar to regular StoreVirtual snapshots in that they can be created manually, scheduled individually, or scheduled as part of a Remote Copy setup. The resulting snapshots can be used as a recovery point in a disaster recovery strategy, single file/VM restores, or for test and development.

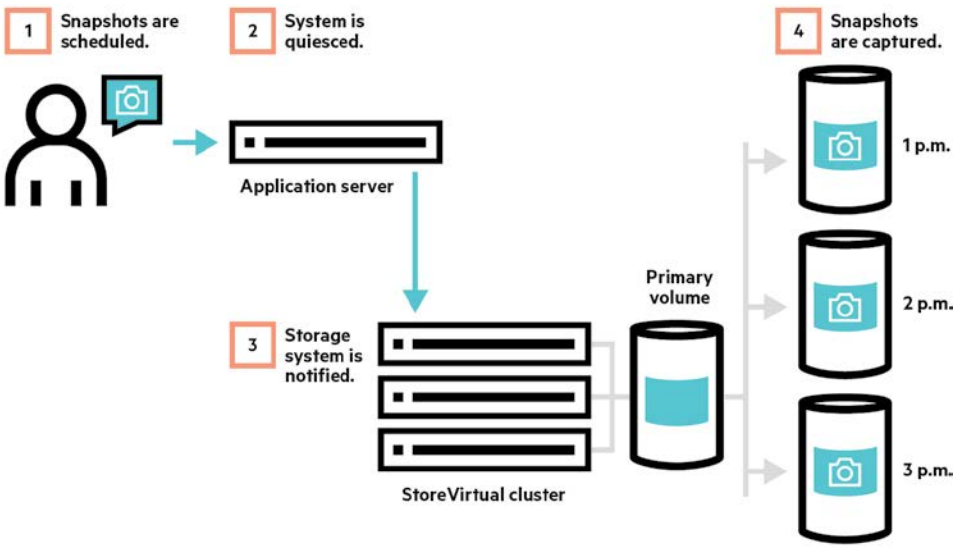

Figure 3. StoreVirtual AASM coordinates snapshots of consistent application data.

Flexible hypervisor integration

StoreVirtual provides tight management integration with VMware vCenter, Microsoft System Center and OpenStack. The flexible StoreVirtual architecture also provides custom integration points. In environments that require high levels of automation for the storage systems and storage provisioning, administrators can automate best practices through software and scripts using command-line interfaces (CLiQs) and APIs. Whether it is mass provisioning of storage to larger installations or custom reporting on capacity utilization, administrators can automate the system. The StoreVirtual CLiQ interface is available for Windows and on every storage node via SSH. Starting with LeftHand OS 11.0, REST APIs can be used to script storage provisioning and snapshots. Refer to the StoreVirtual user guide linked at the end of this document for more detail.

Integration with VMware vSphere

StoreVirtual is certified to integrate with VMware vSphere software, a virtualization platform and software cloud infrastructure that virtualizes server resources to enable critical business applications. Virtualization resources such as VMware vMotion, VMware High Availability, and VMware Fault-Tolerance require the use of highly available shared storage that certified Storage Virtual Appliances, such as StoreVirtual VSA, provide.

In vSphere environments, the StoreVirtual AASM is installed on the vCenter server to orchestrate the creation of application-integrated snapshots. It communicates with VMware Virtual Center Server to create VMware snapshots of virtual machines whose datastores reside on StoreVirtual volumes. Application-managed snapshots use vSphere snapshots to create VMware snapshots which can also include VM memory and quiesced guest file systems before the StoreVirtual snapshot is created.

StoreVirtual supports vSphere storage APIs for Array Integration (VAAI), which improves efficiency and performance in larger vSphere clusters and virtual machine provisioning and cloning operation by offloading certain operations to the storage layer. Through the OneView for vCenter plugin, StoreVirtual also exposes vSphere APIs for Storage Awareness (VASA) 1.0 capabilities, which bring array information into DRS and profile-driven features.

StoreVirtual can also work in conjunction with vCenter Site Recovery Manager (SRM) for integrated storage failover and failback. The StoreVirtual Site Recovery Adapter (SRA) integrates with VMware Site Recovery Manager to respond quickly and accurately to disasters that are geographic in scope. In a recovery operation, SRM brings applications up on remote servers in a choreographed process to ensure proper sequence.

HPE provides recommendations for configuring StoreVirtual SDS in a VMware environment in StoreVirtual best practice guides, hyperlinked at the end of this white paper.

Integration with Microsoft Hyper-V and Windows

StoreVirtual is certified for interoperability with Microsoft Windows, including applications such as Microsoft Exchange, Microsoft SQL Server, Microsoft SharePoint, and Microsoft Hyper-V. All StoreVirtual features can be used by applications on standalone Windows systems or by Windows Failover Clusters.

Application-consistent snapshots in Windows environments are orchestrated in much the same way as they are in VMware environments. The StoreVirtual AASM is installed on application servers to orchestrate the creation of application-integrated snapshots, encompassing all nodes in a Windows Failover Cluster. The Windows Volume Shadow-copy Services (VSS) framework coordinates the creation of snapshots for backups, creating consistent, point-in-time copies of volumes. Snapshot sets (consistency groups) ensure that multiple volumes are captured at the same time so that application data spread across volumes is backed up consistently. For example, StoreVirtual AASM would create a snapshot set that includes both the Microsoft SQL Server database and log volume. VSS integration also supports Microsoft Windows Cluster Shared Volumes (CSV) and can automatically quiesce CSV volumes when taking snapshots for improved application integration.

Provisioning storage for new VMs in Microsoft Hyper-V environments is easy with Microsoft System Center Virtual Machine Manager (SCVMM). Microsoft SCVMM 2012 can integrate with storage pools on StoreVirtual via SMI-S. This integration allows for provisioning of new volumes to Hyper-V hosts and rapid provisioning of VMs from the SCVMM library using Microsoft SANcopy without switching to the StoreVirtual CMC and without an additional agent on the SCVMM server. Users simply add StoreVirtual to the array management in SCVMM.

Windows Active Directory framework
The flexibility of centralized management is further enhanced by integration with the Windows Active Directory (AD) framework. Within AD, administrators can:

- Manage user authentication to HPE StoreVirtual
- Create different storage array admin groups with various permission levels
- Log into the StoreVirtual CMC using AD credentials
- Add and remove users within AD without accessing the CMC
- Track and audit AD users from the CMC or command-line tool (CLIQ)

Windows AD integration helps businesses organize and manage user directories across applications and devices, mapping AD user groups to user roles and consolidating credentials.

High performance multi-pathing for Windows & VMware
Every StoreVirtual solution is enabled for highly available iSCSI data path operations by simply including at least two storage nodes in a cluster. There are two connection methods for volumes on the array. By default, all volumes are enabled for iSCSI load-balancing connection for the application servers. Alternatively, for even higher levels of connection availability, organizations can implement multi-pathing I/O (MPIO), with specific optimizations for Microsoft Windows® and VMware vSphere environments.

By default, LeftHand OS load balances all iSCSI sessions (via iSCSI login redirect) from client iSCSI initiators across all the storage nodes in the cluster, and then processes all subsequent data I/O requests from that client through that same node. The storage node then redirects each client request to the storage node owning the desired data block, and then redirects the response back to the client. In the event that a storage node hosting the iSCSI session for the volume goes offline, the iSCSI session is automatically redirected to one of the remaining storage nodes in the cluster. This failover of the iSCSI session occurs quickly enough that the session on the application server does not reset. This avoids having an impact on the application server's connectivity to the volume.

The LeftHand OS Multi-Pathing Extension Module (MEM) for VMware and the device-specific module (DSM) for the Microsoft Windows MPIO iSCSI plug-in contribute to availability in similar ways. Both features establish independent iSCSI sessions from the application server to each storage node in the cluster for the volume, thus establishing multiple I/O paths for the volume that can be used concurrently. In the event that a node goes offline, the surviving connections continue to support I/O for the volume. When the failed node is brought back online, the MEM (in VMware environments) or the DSM for MPIO (in Windows environments) automatically reestablish sessions to the restored node.

Multi-pathing Extension Module (MEM) for VMware
The StoreVirtual multi-pathing solution for VMware provides advanced capability for StoreVirtual cluster deployments in a VMware environment. Once the StoreVirtual MEM is installed on the ESXi server, all supported StoreVirtual iSCSI volume connections will be claimed by the StoreVirtual MEM. You do not need to configure the connections manually.

MEM provides awareness of the data layout in the StoreVirtual clusters to the ESXi server. By establishing iSCSI sessions to every storage node in the cluster, it optimizes the data paths to the data stored on the node. It was designed to reduce latency and increase throughput in environments with storage nodes with SSDs and more complex multi-site StoreVirtual clusters.

MEM is a useful addition to vSphere's native multipathing, which utilizes multiple network adapters for increased throughput and resiliency.

DSM for MPIO

Similar to MEM, the StoreVirtual DSM for the Microsoft Windows MPIO delivers stronger performance and reduced latency by allowing independent logical connections between application servers and storage nodes. The overall degree of performance increases linearly with scaling of the cluster size (Figure 4).

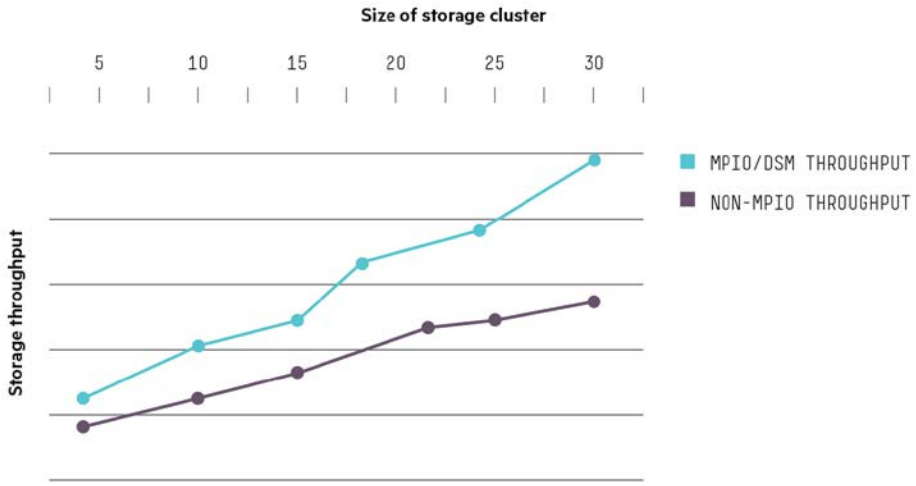

Figure 4. DSM for MPIO allows performance to scale more linearly with capacity, improving sequential I/O performance over standard load balancing.

The DSM for MPIO integrates with the Microsoft Windows iSCSI MPIO framework for increased performance and availability. The DSM establishes an independent iSCSI session between the application server and each storage node in a storage cluster, helping to increase parallelism for performance and availability.

Aware of layout algorithms for the storage cluster, the DSM for MPIO can calculate the location of any block in any virtual volume. The iSCSI driver can contact the storage node that owns the block directly, without the redirection used by the standard iSCSI load-balancing approach. Figure 5 illustrates a redirected login sequence, the SCSI Mode Sense command that loads the cluster-specific information into the driver, and the separate I/O path that the driver establishes to each server in the cluster.

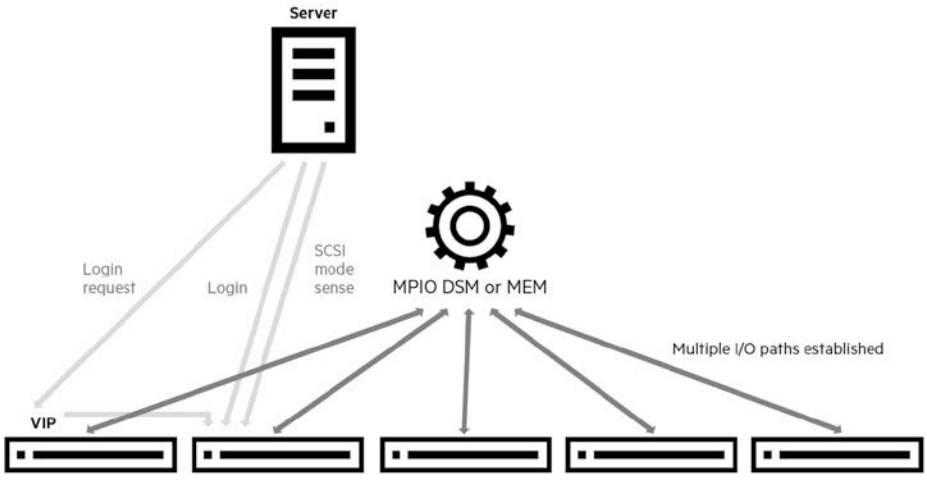

Figure 5. The StoreVirtual MEM and the DSM for the Microsoft MPIO driver establish parallel I/O paths–one to each storage node in the cluster.

Integration with OpenStack and container drivers

Businesses can meet unpredictable workloads and cloud storage demands with StoreVirtual's integration into OpenStack's block driver, Cinder. A 50TB StoreVirtual VSA license is included with Helion OpenStack.

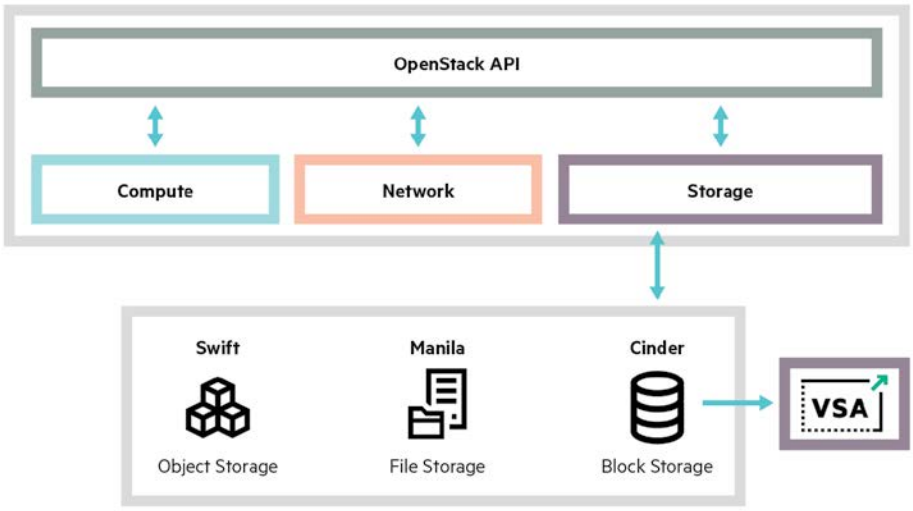

Figure 6. StoreVirtual integration into OpenStack Cinder driver.

StoreVirtual's iSCSI integration into OpenStack's Cinder driver for block storage provides complete array functionality on top of Linux KVM/ESXi environments without the need for external array hardware (see Figure 6). StoreVirtual features simplify provisioning and deploying cloud storage with HPE Helion OpenStack, while delivering the added benefits of:

- **Heterogeneous environments**—supports mixed cloud and hypervisor environments on commodity hardware
- **Non-disruptive scalability**—provides client provisioning when and where needed
- **Centralized management**—converged management of compute storage and virtual machines through OpenStack's Horizon management interface
- **High availability**—supports business continuity with 99.999 percent high availability

The CMC manages all of the StoreVirtual instances attached to an HPE Helion OpenStack cloud from overall cluster health to volume protection levels, alarms and software upgrades. Administrators and users are able to provision storage for use within their virtual compute environment through the OpenStack Horizon interface.

To meet the rapid deployment of lighter-weight virtualization within the cloud environment, HPE Storage supports portable and stateful containers for storage through ClusterHQ Flocker's OpenStack Cinder driver for StoreVirtual. Administrators can deploy Docker in an OpenStack environment to take greater advantage of the flexibility and simplicity of StoreVirtual VSA features.

Intuitive infrastructure management

Intuitive interfaces make the everyday operation of StoreVirtual solutions simple enough for IT generalists to manage without specialized training. StoreVirtual VSA and arrays can be managed through the CMC, with built-in wizards for complex operations. Hyper-converged and HPE Helion cloud solutions abstract storage infrastructure management for day-to-day operations, and provide clear, easy-to-read dashboards via OneView and Horizon interfaces, respectively.

To make things even simpler, StoreVirtual is well integrated with VMware, Microsoft, and OpenStack, allowing administrators who are already comfortable with vSphere, vCenter, and Horizon to continue to use those management interfaces, especially for provisioning storage for new virtualized workloads or applications.

Centralized Management Console

Included with StoreVirtual VSA and StoreVirtual arrays, the CMC provides a single point of management for all storage features. No matter how many StoreVirtual solutions an organization installs, local or remote, they can all be managed through a single, intuitive, graphical user interface (GUI). It is as easy to manage one cluster as it is to manage many, regardless of the size. Provisioning a volume, changing its attributes, copying volumes, or taking snapshots are easy-to-execute tasks. For complex operations such as geographic failover and failback, CMC wizards take the risk out of performing critical functions in times of stress. The CMC provides access to all the information needed to manage storage, including detailed performance and capacity statistics that can be leveraged for capacity planning.

Performance Monitor

StoreVirtual architecture provides the performance metrics needed to monitor, manage, and troubleshoot storage performance. The Performance Monitor, found within the CMC, reduces the abstraction that can make performance management in virtualized environments quite complex.

Storage metrics are available based on application servers, VMs, logical volumes, snapshots, storage clusters, and storage nodes, providing valuable insight to system operation without searching through irrelevant statistics. The GUI is simple and elegant, presenting each metric together with a detailed explanation of what it means for storage performance. For automation purposes and data extraction into other management tools and suites, all the performance information for the SAN is available via Simple Network Management Protocol (SNMP), management information bases (MIBs), and CLI calls.

Best Practice Analyzer

The Best Practice Analyzer in the CMC continually monitors the storage to make sure that LeftHand OS best practices, such as cluster size, disk RAID consistency, and Network Interface Card (NIC) bonding configurations, are adhered to. The data is continually updated and displayed on a dashboard in the CMC, informing users proactively about changes that need to be made.

Scalable architecture

The StoreVirtual platform is engineered to be easily scalable. The basic components of a data center make up HPE SDS solutions and hyper-converged appliances: virtualized servers, StoreVirtual shared storage, integrated management software, plus applications. This consolidated, converged approach makes it easier for administrators to deploy a virtualized infrastructure, roll out projects, and keep up with business requirements.

Linear scalability

Scalability means the ability to add more resources to a system, resulting in a directly proportionate increase in the system's performance. StoreVirtual solutions are built with a scale-out architecture, which means that both capacity and performance scale linearly, without disrupting data access. The LeftHand OS manages each storage node in parallel. A scale-out architecture offers the flexibility to grow storage on demand, maximizing return on investment.

Each storage node contributes its own disk drives, RAID controller, cache, memory, CPU, and networking resources to the cluster. Every volume is striped across each node in the cluster, leveraging all of their critical resources. The OS manages each storage node in parallel, so the cluster's performance increases in lockstep with its storage capacity (Figure 7). This means that when users add nodes to a cluster, they get both the performance and capacity they need without incurring downtime.

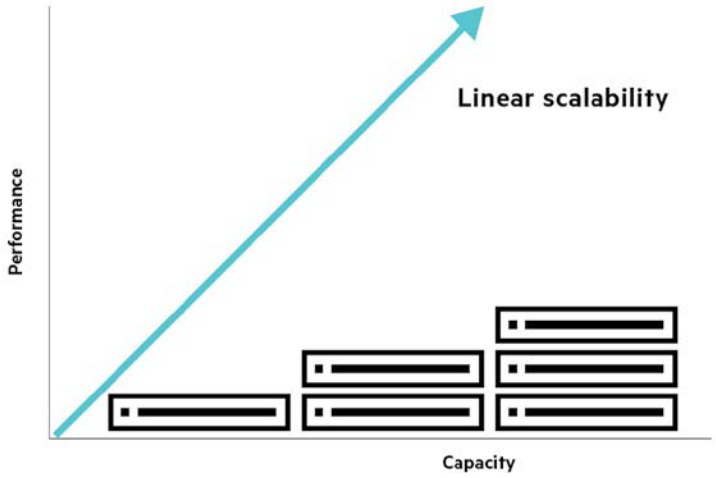

Figure 7. Capacity and performance grow in parallel as new nodes are added to the cluster.

This contrasts to the traditional methods used by most other SANs and some network attached storage (NAS) appliances that consist of purpose-built controllers and disk trays. These systems scale by adding disk trays, increasing performance only to the point at which the controller or the Fibre Channel interconnects becomes a bottleneck. At that point, organizations must either upgrade to higher-power controllers or add new storage systems. Either choice can result in significant amounts of application downtime.

An often-forgotten aspect of enterprise-class scalability is the need for a system to scale without increasing in complexity, especially in terms of management. The StoreVirtual CMC manages all storage clusters in the environment no matter the size or location of the storage clusters. The storage cluster serves as a layer of abstraction so the administrator does not need to manage each storage node individually, but rather as a single entity. Once the storage node is added to a cluster, all other operations are performed against the storage cluster as a whole. Adding a new storage node to a cluster is as simple as discovering the storage node in the CMC and assigning it to a cluster. The management console and the cluster itself coordinate the integration of the storage node with no additional intervention from administrators. This allows the array (or storage pool) to increase in size without increasing in management complexity.

Pay-as-you-grow modularity

Linear scalability gives "pay-as-you-grow" departments and businesses the agility to respond quickly to changes in storage requirements, and the cluster architecture makes it much simpler to size a StoreVirtual storage infrastructure. Combined with Network RAID, which spreads data across the cluster for data protection, the clustering architecture allows the capacity of the array to grow over time without disrupting application availability. Instead of estimating growth ahead of time and buying a higher capacity storage solution today, administrators can add storage nodes to the cluster as the business need arises.

This "pay-as-you-grow" architecture stands in sharp contrast to traditional architectures, where estimates are made for future requirements, and storage is purchased up front. StoreVirtual technology lets organizations purchase storage in affordable increments rather than having to justify the purchase of future storage today.

Peer Motion

LeftHand OS Peer Motion data mobility technology enables simplified movement of volumes across all deployed StoreVirtual nodes regardless of platform or drive type. Peer Motion can move data between clusters in a management group and simplifies upgrading to new-generation technology without scheduled downtime or taking data offline. It offers data and workload movement between storage clusters to address performance and capacity requirements. Peer Motion enables seamless StoreVirtual storage technology refreshes, eliminating downtime and service interruption during migration and upgrade activities.

Additional benefits of Peer Motion include:

- Promotes federated asset management by moving data from retiring storage systems to new storage systems non-disruptively
- Allows flexible migration of data volumes between multi-site clusters to align with the growing demand for cloud-based storage
- Enables clients to transparently move application workloads between storage tiers in virtualized and cloud computing environments
- Increases return on investment by combining HPE thin provisioning and volume migration technology

Highly available, continuous storage

StoreVirtual helps achieve five-9's (99.999%) availability with storage clustering and built-in Network RAID that stripes and protects multiple copies of data across a cluster of nodes, eliminating any single point of failure. Storage clustering turns a set of storage nodes into a storage pool from which volumes are allocated to application servers. Network RAID provides high availability by synchronously protecting data across a cluster. By storing multiple copies or parity-protected copies of a volume's data across the storage cluster, volumes can be configured for very high reliability levels where multiple failures in the cluster can be tolerated while still having access to the data. If any storage node is offline for maintenance, a network cable is accidentally pulled, or another error takes place, a copy of the volume's data is available from another storage node in the cluster. Applications have continuous data availability in the event of a disk, controller, storage node, power, network, or site failure.

For high availability storage and transparent failover, data centers require two nodes for data redundancy, plus either a failover manager (FOM) on a third device. If a third device is not available on site, the LeftHand OS can provide high availability with as few as two nodes through a feature called 2-Node Quorum.

Network RAID synchronous replication

Network RAID offers data protection in case of hardware failures. Storing one or two copies of data on a single disk puts that data at risk. To mitigate that risk, Network RAID spreads data across the storage nodes in a cluster, and the nodes work in parallel to satisfy application server data requests.

Specifically, Network RAID stripes and replicates data blocks across the cluster as dictated by the cluster size and individual volume's availability requirements, storing data on multiple hard disks to eliminate any single point of failure (Figure 8).

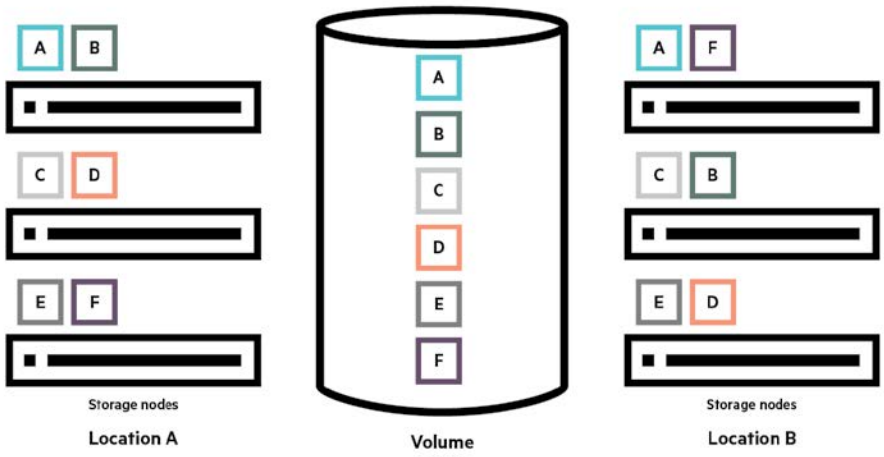

Figure 8. Network RAID 10 configuration. Each of the data blocks in the volume is distributed across two nodes.

As application servers write data to a volume, the data is transparently distributed over the storage nodes in the cluster based on the volume's Network RAID level. At its most basic level, recommended for test and development environments, Network RAID can simply stripe data across storage nodes. For production workloads, Network RAID can stripe and mirror two, three, or four copies before the write acknowledgment is sent back to the application server. This process delivers consistent results, and ensures that all copies of the data are identical across the storage cluster.

Network RAID 10 is most commonly implemented for high availability for production level volumes. This level of Network RAID is recommended for most workloads, including server/desktop virtualization and databases workloads. Using a minimum of two nodes, Network RAID 10 synchronously replicates copies of each of the volume's data blocks across the storage cluster. The volume is guaranteed to have two copies of every block available, and because those copies reside on separate nodes, it allows for the system to sustain any single storage node failure. Network RAID 10 also provides the synchronous replication between two locations in a stretched, multi-site cluster. For data that requires higher levels of availability, administrators can implement Network RAID 10+1 (3 copies) or 10+2 (4 copies).

Administrators who seek to drive cost effective high-availability for volumes that contain mostly read workloads, can implement parity-based Network RAID levels: Network RAID 5 & Network RAID 6. These parity-based Network RAID levels stripe the data and a parity block across the nodes in the cluster and therefore provide better capacity utilization than Network RAID 10. Network RAID 5 can sustain any single node failure, and Network RAID 6 provides dual parity for extra protection. In the event of a node failure, the parity stripe acts as a reference and the missing block can be recalculated. All of this is managed internally in the cluster and is not visible to applications. Parity-based Network RAID levels cannot be used in a stretched, multi-site cluster.

StoreVirtual architecture allows a single cluster to host volumes with different Network RAID levels, matching each volume's availability and/or performance level to the needs of the application. Network RAID is an attribute of each volume, so the level can be changed as needed without having to move a volume between RAID groups or taking it offline, which is the case with many traditional storage architectures. The result is greater storage capacity efficiency and flexibility, since only the amount of storage required to support the desired availability level is used for individual volumes.

Failover Manager and 2-Node Quorum

The Failover Manager (FOM) is designed to provide automated and transparent failover capability. For fault tolerance in a single-site configuration, the FOM runs as a virtual appliance in either a VMware vSphere, Microsoft Hyper-V Server, or Linux KVM environment, and must be installed on storage that is not provided by the StoreVirtual installation it is protecting.

The FOM participates in the management group as a manager; however, it performs quorum operations only, it does not perform data movement operations. It is especially useful in a multi-site stretch cluster to manage quorum for the multi-site configuration without requiring additional storage systems to act as managers in the sites.

For each management group, the StoreVirtual Management Group Wizard will set up at least three management devices at each site. FOM manages latency and bandwidth across these devices, continually checking for data availability by comparing one online node against another. If a node should fail, FOM will discover a discrepancy between the two online nodes and the one offline node – at which point it will notify the administrator. This process requires at least three devices, with at least two devices active and aware at any given time to check the system for reasonableness: if one node fails, and a second node remains online, the FOM will rely on the third node to maintain quorum, acting as a "witness" to attest that the second node is a reliable source for data.

In smaller environments with only two storage nodes and no third device available to provide quorum, administrators have two options:

A. Supply a third node onsite with StoreVirtual VSA installed and use FOM to maintain quorum

B. Set up 2-Node Quorum on a shared disk, using LeftHand OS 12.5 or later versions

StoreVirtual 2-Node Quorum is a mechanism developed to ensure high availability and transparent failover between 2-node management groups in any number of satellite sites, such as remote offices or retail stores. A cost-effective, low-bandwidth alternative to the FOM, the feature does not require a virtual machine in that site, relying instead on a centralized Quorum Witness in the form of an NFSv3 file share as the tie-breaker between two storage nodes (Figure 9).

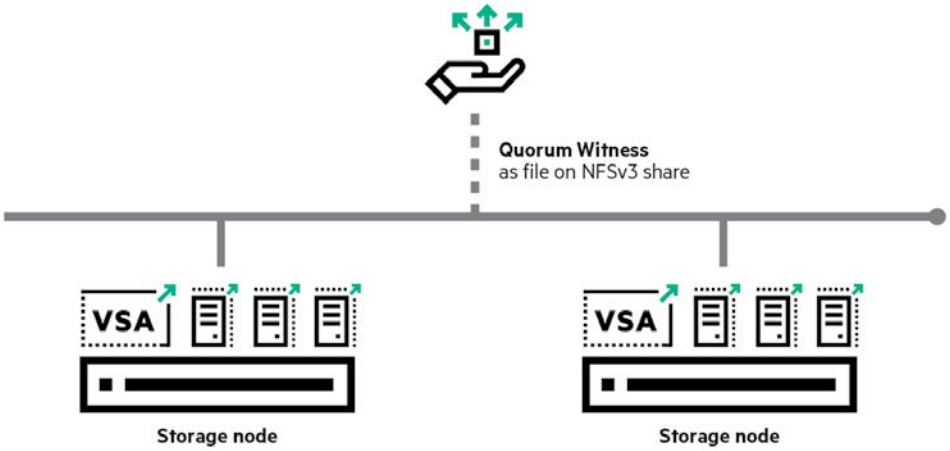

Figure 9. StoreVirtual 2-Node Quorum for high availability with 2-node cluster

Quorum Witness uses a shared disk to determine which of the two nodes should be considered a reliable resource in the event of a failure. The shared disk is an NFS share that both nodes in the management group can access.

Efficient use of resources

The need to trim operating expenses while increasing availability is a driving force behind virtualization, storage tiering, and performance monitoring technology. Organizations need to deploy more applications with better performance while using less of the data centers' limited space, power, and cooling resources. Legacy storage architectures are inefficient, forcing IT organizations to purchase space for future needs at the time their systems are deployed—a costly proposition with significant risk in ROI.

Improved ROI begins with storage that integrates efficiency into its architecture. LeftHand OS allows storage to be purchased only as it is needed, and thin provisioning keeps unused capacity available and reduces the need to hold large amounts of storage in reserve. Capacity-saving snapshots and Space Reclamation make this software-defined architecture more efficient in its use of space, and improve the performance of StoreVirtual features, from sub-volume auto-tiering to intelligent multi-pathing.

Unconditional Thin Provisioning

Thin Provisioning, integrated throughout the StoreVirtual architecture, allows storage for volumes, snapshots, remote copies, and volume clones to use only the storage they need. This helps to increase storage utilization and reduce total cost of ownership. The thin provisioning engine allocates storage from the storage cluster as data is written to the volume (snapshot, remote copy, or clone) to be perpetually optimal for storage capacity utilization. All snapshots, remote copies, and clones are thin provisioned, eliminating the need for the administrator to manage the storage reservations for those items.

Thin Provisioning allows volumes to be sized large from the beginning with no impact on cost because no storage is actually allocated when the volume is created. This approach, also known as "over-provisioning" allows the administrator to present more capacity to the application servers than is actually in the storage cluster. As storage volumes have blocks allocated to them and the storage cluster begins to fill to capacity, additional storage nodes can be added to the storage cluster seamlessly. The storage pool capacity grows underneath those volumes with no impact to application availability.

Space Reclamation

The Space Reclamation feature allows supported operating systems to manage thin volumes more efficiently by reclaiming space from blocks that have been deleted. Space Reclamation recovers space on thinly and fully provisioned volumes used by Windows Server 2012 and vSphere 5.5 or later. If this feature is not enabled, the space doesn't automatically get returned to the usable capacity pool once data is written and then deleted on a block. Space Reclamation allows the cluster to reclaim unused capacity from volumes where data has been deleted to maximize efficiency (Figure 10).

Figure 10. Space Reclamation

Adaptive Optimization

Data-intensive environments can be made much more efficient and cost-effective by reserving space on flash disks for files that need to be most accessible, and keeping files that are infrequently accessed on a slower storage tier. This method contrasts with keeping all files equally accessible, which creates inefficiency unless the storage infrastructure is comprised entirely of high-performance solid-state drives (SSDs).

A simple and smart sub-volume auto-tiering feature, StoreVirtual Adaptive Optimization allows both SSDs and HDDs to be used within a single StoreVirtual instance, creating a higher level of overall performance. This built-in capability moves more frequently accessed blocks to the Tier-0 flash storage tier and keeps the infrequently accessed blocks on Tier-1 with lower performance and potentially lower-cost hard disk drive (HDD) storage.

Adaptive Optimization provides intelligent, automatic, and transparent performance optimization by monitoring data access patterns at a granular level based on a page size of 256 KB. The system maintains a data "heat map" to track access frequency. While many storage systems only update the data blocks at set times throughout the day, the LeftHand OS monitors data and moves pages continually and in real time for the most efficient use of SSD capacity.

Figure 11 depicts the dynamic movement between the flash tier and the HDD tier. The data block that has been accessed 146 times on the left will automatically move down to Tier-1, while the data block on the right, accessed 149 times, will move to Tier-0 due to the increased frequency of transactions.

Figure 11. Adaptive Optimization - data balanced dynamically between tiers

By enforcing extremely efficient use of valuable SSD capacity, Adaptive Optimization keeps the flash tier full and applies change thresholds to prevent data thrashing. For additional efficiency, only a single copy of data is maintained on Tier-0, regardless of which Network RAID protection level is implemented. As a result, the most valuable capacity real estate is reserved for application data.

Automatic upgrades

The Automatic Online Upgrades feature proactively identifies the latest system updates and enhancements, allowing IT administrators to download upgraded software components whenever HPE makes them available. Full support for HTTP and SOCKS proxies provide greater flexibility for IT teams. (High-security sites that restrict Internet access can upgrade by simply taking the CMC outside of the dark site to download all upgrade components.)

Flexible data protection options

While high availability is a key reliability indicator, data recovery and protecting against data loss are also crucial. LeftHand OS provides multiple layers of protection against data loss. Multi-fault protection avoids data loss due to the failure of any single component in a storage node, and it protects against the failure of multiple components across storage nodes.

Reliability begins with the use of enterprise-class server technology, such as HPE ProLiant. StoreVirtual storage solutions are built with redundancy as a best practice, which contributes to data protection. Depending on the system and the organization's requirements, disk RAID levels 5, 6, and 10 can be configured on each node and then combined with Network RAID across the storage nodes to protect against multiple concurrent component failures.

For added data protection, space-efficient reservationless snapshots can create point-in-time copies of data for backup purposes, and Remote Copy provides space- and bandwidth-efficient asynchronous replication. Snapshots help to recover from data corruption, logical file system problems, and accidental or malicious destruction of data. Remote Copy enables centralized backup and disaster recovery on a per-volume basis and leverages application integrated snapshots for faster recovery.

Snapshots

Snapshots provide instant, point-in-time volume copies that are readable, writeable, and mountable for use by applications and backup software. LeftHand OS is uniquely designed to take snapshots that are thinly provisioned volume copies requiring no space reservation, allowing more efficient use of storage.

Snapshots can be used to restore entire volumes through a simple rollback process, or to restore individual files, since, unlike many other array products, snapshots can be mounted and individual files retrieved from them. They can help to prevent a high rate of data loss due to human error by allowing multiple, point-in-time copies of a volume with the ability of instantly rolling back to a known point in time. Because snapshots are thinly provisioned, administrators can make numerous snapshots of a volume over time and only the changes are committed.

Snapshots can be created manually, scheduled, scripted, or created through programmable application programming interfaces (APIs), synchronized with application state to create a stable, consistent backup or source for tape backup or remote copy. StoreVirtual AASM provides automated quiescing and integrates into Windows Server VSS and VMware vCenter, making the data in the snapshot consistent with the application's view of the data. StoreVirtual Recovery Manager for Windows leverages snapshots to quickly identify files or folders for recovery to any location.

SmartClone

StoreVirtual's SmartClone feature helps businesses respond quickly to changing conditions; create new VMs for development, test, or deployment purposes; and make virtual desktop environments efficient, secure, and cost-effective. SmartClone clones are space-efficient copies of existing volumes or snapshots. Using the snapshot mechanism to clone volumes for use by virtual or physical servers, SmartClone instantly turns any volume or snapshot into one or many full, permanent, read-write volumes. They appear as multiple volumes that share a common snapshot, called a clone point. It is as easy to create a new volume copy as it is to create the new VM to use it, making SmartClone an ideal companion to VM cloning features.

Volume clones use redirect-on-write semantics to avoid copying or duplicating data, making SmartClone an instant, space-efficient way to increase storage utilization and improve storage ROI (Figure 12).

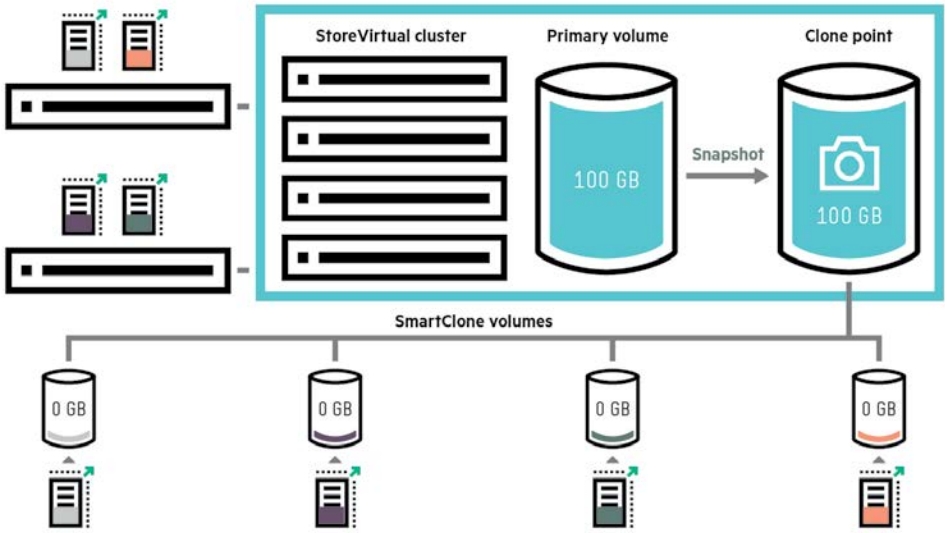

Figure 12. SmartClone volumes

SmartClone can be used to duplicate configurations or environments quickly and without consuming disk space for duplicated data. These fully featured, writable volumes are dependent on the clone point, minimizing space used on the array. For example, a server's operating system takes up considerable storage but does not change frequently, so a volume could be created with the specific OS configuration. Then multiple SmartClone volumes could be created from that master image without using additional storage capacity. A single instance of the configuration is all that is required. The volumes consume additional space on the array only as additional data is written to the individual SmartClone volumes.

Remote Copy

The Remote Copy feature replicates snapshots between StoreVirtual-based systems at primary and/or remote locations. Based on thinly provisioned snapshots, Remote Copy provides asynchronous replication for disaster recovery, supporting one-to-one, one-to-many, many-to-one, and round robin configurations. This allows it to meet today's disaster recovery (DR) needs. Precise, consistent, point-in-time copies of data are stored at a remote location, and they can be used to continue business operations at an alternate location in the event of a site failure.

In traditional storage architectures a 100-percent reserve used to be required for a local snapshot, a 100-percent reserve was required for the remote copy, and the remote copy itself had to be copied before it could be mounted and used at the remote site.

StoreVirtual requires only two copies of volume data for remote copies, and the primary volume is always thinly provisioned. The integration of thin provisioning with Remote Copy results in highly efficient use of storage for disaster recovery (Figure 13). Both the snapshot that initiates the remote copy and the remote copy itself are created without reserve.

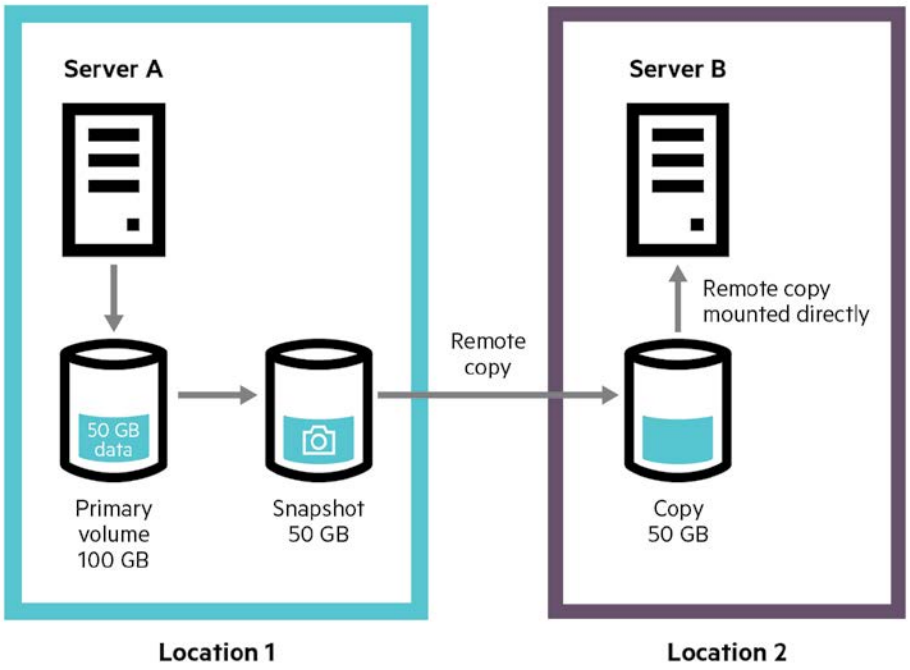

Figure 13. In this example, StoreVirtual requires 200 GB for Remote Copy operations: 100 GB for the primary volume (which contains 50 GB of data), plus 50 GB for the thinly provisioned snapshot, and 50 GB for the remote copy.

Remote copies require the same number of blocks to be allocated on the local and the remote site: the number of blocks that are actually in use. For example, if a volume is 40 percent utilized on the primary site, the remote copy of the volume in the remote site requires only 40 percent capacity utilization.

Asynchronous replication is simply a series of scheduled remote copies that support the same application integration as standard snapshots. The cluster understands the relationship between any remote copy to the sync point on the primary copy, so only the blocks that have changed since the last remote copy are asynchronously replicated. This simplifies failover for DR, because there is no issue with incomplete or inconsistent writes. Failback is efficient because only the incremental, changed data blocks must be copied back to the primary location once service has been restored.

StoreVirtual remote copies are space-efficient, and custom bandwidth management helps to share wide-area network (WAN) resources and maintain network quality-of-service levels. Bandwidth used for copying data between sites is managed with preset limits, so the WAN resources can be shared efficiently with other network services. If network conditions deteriorate for any reason, Remote Copy automatically adapts its packet flow without requiring complex network quality-of-service management.

Remote Copy further conserves bandwidth by check-pointing in-flight remote copies at every 10th percentile of the journey. If a network interruption occurs during a remote copy operation, it can resume from the checkpoint rather than having to restart at the beginning. This strategy also helps to reduce the disruption in the event of an interruption in network connectivity. Within a cluster, local network bandwidth can be prioritized for peak storage response times for applications by setting a lower priority for activities such as data rebuilding or restriping.

Files and folders on a Windows Server can be retrieved easily from snapshots using HPE StoreVirtual Recovery Manager. As is the case with regular snapshots, Remote Copy should use application-consistent snapshots via the AASM feature. Smart search capabilities speed recovery and can be used to access up to five of the most recent volume snapshots automatically. Once identified, the file or folders can be recovered to any location—the original location, a file share, or a local file system.

Split Networks

StoreVirtual's Split Networks feature separates high-priority iSCSI storage traffic from lower-priority administration traffic generated by the CMC. This effectively enables IT administrators to meet stringent networking policies for security purposes by preventing management traffic from being exposed to the host (Figure 14).

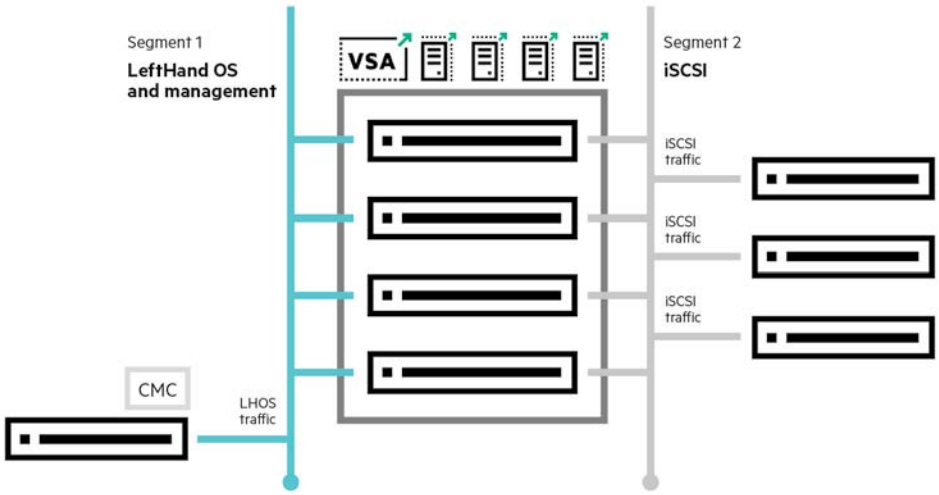

Figure 14. Split Networks securely separates traffic onto separate networks

Multi-Site Stretch Cluster

The StoreVirtual synchronous replication that protects a cluster from a node failure also provides real-time site protection for data centers with the Multi-Site Stretch Cluster feature. Multi-Site Stretch Cluster allows storage clusters to be stretched across physical or logical sites to provide high availability across geographic locations. The technology can also be used to protect against the failure of a logical site—a rack, data center, building, or data centers that are separated by up to 100 km.

The Multi-Site feature assigns storage nodes in the cluster to different sites (racks, rooms, buildings, and cities). Based on this assignment, the software makes configuration and operational decisions to always protect data from a complete site failure—automatically. Using the CMC, an administrator can map the data center configuration to the storage device to completely automate all site-dependent data availability and fault-tolerance decisions within the software. The CMC enables fault tolerance programmatically, and does not allow the administrator to create a non-fault-tolerant configuration.

Enabled through Network RAID, Multi-Site Stretch Cluster provides the ability to stripe multiple copies of data across storage nodes in a cluster. As Figure 15 illustrates, blocks are replicated so that one storage node can go offline at each location without affecting availability of any volume at either location. This is particularly useful in virtualized environments, where the VMs running on Servers A and B could failover in Location 2 following such a failure. Because the volumes are exactly the same whether accessed from Location 1 or 2, no reconfiguring of logical unit numbers (LUNs) needs to take place in order for restarted servers or VMs to continue operations at the alternate site.

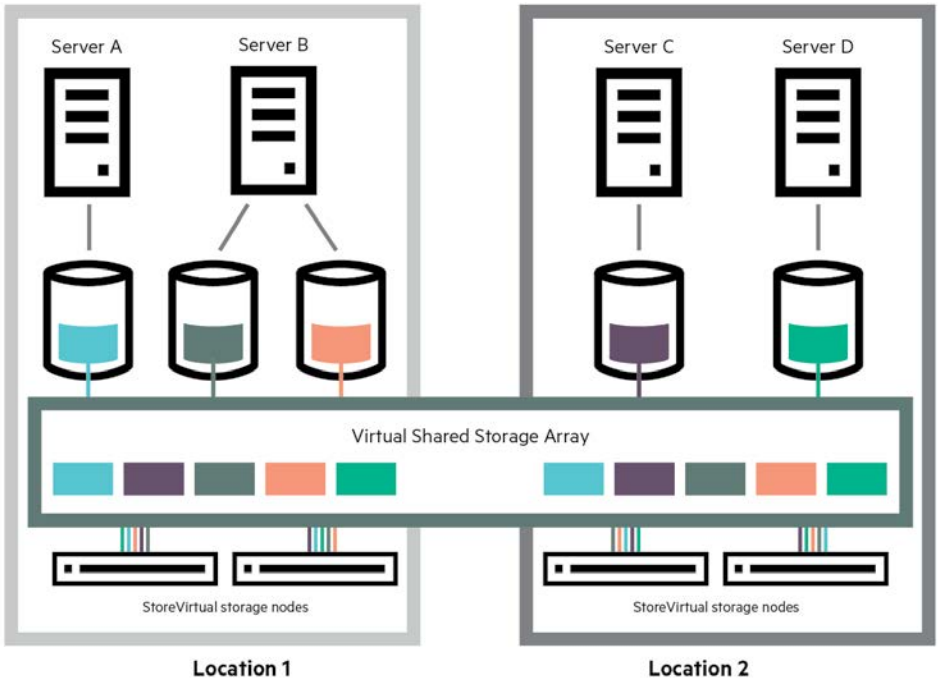

Figure 15. Multi-Site Stretch Cluster configurations support continuous availability in the event of a site failure. When the downed site returns to an online state, the data is rebuilt automatically.

When the failed site comes back online, its storage nodes automatically obtain any changed data blocks so failback is automatic and transparent to the application servers. With traditional storage architectures, this failback process can be time-consuming and error-prone, and it can require application downtime.

While a flat network is recommended, the Multi-Site feature supports other more advanced networking architectures in a multi-data-center environment. Storage clusters can span multiple IP subnets, so the storage pool can accommodate more complex multi-site network topologies. For each subnet and site pair, the storage cluster is accessed using a dedicated virtual IP address (VIP) for iSCSI high availably in each site. Application servers in Location 1 access volumes through the VIP for Location 1, while application servers in Location 2 access volumes through the VIP in Location 2.

Conclusion

Today's more powerful multi-core servers, combined with virtualization technology, place unprecedented demands on storage—demands not being met by traditional storage systems that are based on a controller-and-disk-tray model. Businesses need a storage architecture designed for virtualization.

StoreVirtual meets this need by delivering investment protection through flexible and scalable software-defined storage solutions that are adaptable to a wide variety of environments and business needs. Features including storage clustering, Network RAID, Thin Provisioning, Snapshots, and Remote Copy form the basis for the many advantages of StoreVirtual. With simple deployment and management though the StoreVirtual CMC and its integrated Performance Monitor, administrators do not require specialized training.

StoreVirtual appliances and solutions are engineered to perform optimally in a virtualized environment, based on an architecture that is simple, scalable, highly available, efficient, and protected.

Additional resources

HPE StoreVirtual Storage User Guide
http://h20628.www2.hp.com/km-ext/kmcsdirect/emr_na-c04776097-1.pdf

HPE LeftHand Storage with VMware vSphere: Design considerations and best practices
http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA3-6918ENW&cc=us&lc=en

HPE StoreVirtual Storage: Network design considerations and best practices
http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA2-5615ENW&cc=us&lc=en

HPE StoreVirtual VSA design and configuration guide for Microsoft and VMware environments
http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA4-8440ENW&cc=us&lc=en

Sign up for updates

★ Rate this document

4AA3-0365ENW, May 2016 rev. 2

Hewlett Packard Enterprise