Hewlett Packard Enterprise

HPE Reference Architecture for High Availability/Disaster Recovery solution of virtualized Microsoft SQL Server 2017 on HPE 3PAR storage

HPE 3PAR StoreServ 9450 storage using HPE 3PAR Remote Copy and Peer Persistence software

Reference Architecture

Contents

Executive summary

With online transaction processing, milliseconds of downtime or delayed response time can impact a customer's business by millions of dollars. All-flash HPE 3PAR StoreServ storage (HPE 3PAR) is a cost-effective product that provides high performance and meets resiliency demands of mission critical Microsoft® SQL Server 2017 databases. With HPE 3PAR Peer Persistence software (Peer Persistence), HPE 3PAR storage systems located at metropolitan distances can act as peers to each other, presenting a highly available transparent failover storage system to hosts and servers connected to them. This capability enables customers to configure a high-availability (HA), disaster recovery (DR) solution between two sites or data centers where switchover and switchback remains transparent to the hosts and applications running on those hosts. This secures business continuity without impacting performance and provides load balancing across systems and sites.

This Reference Architecture provides a solution for organizations interested in recovery of virtualized Microsoft SQL Server 2017 on HPE 3PAR storage.

This Reference Architecture demonstrates:

- Key advantages of transparent failover across data centers with federated storage
- Load balancing across systems or sites

Target audience: This Reference Architecture is intended for those who maintain, evaluate, or design backup and storage systems, including IT administrators, Microsoft Windows® administrators, storage administrators, and solution architects planning disaster recovery of Microsoft SQL Server 2017 virtualization deployment with HPE 3PAR storage.

Document purpose: This document highlights recognizable benefits to technical audiences. The objective is to showcase HPE 3PAR StoreServ 9450 storage with HPE 3PAR Peer Persistence and Remote Copy software for disaster-tolerance and high availability in a Microsoft SQL Server 2017 environment, virtualized by Hyper-V.

This white paper describes a project developed by Hewlett Packard Enterprise in November 2017.

Note
Although this document is for HPE 3PAR 9450 storage, it is applicable to the entire HPE 3PAR StoreServ storage portfolio.

Solution overview

Customers deploy HPE 3PAR 9450 storage for its all-flash capability that provides high performance and high availability in SQL Server installations. Virtualizing SQL Server instances is also a key trend in these environments as part of a consolidation strategy to utilize resources more effectively and reduce the amount of physical hardware running these database instances. However, to protect their mission-critical database against local area disasters, a robust replication and failover solution should also be deployed.

This solution provides an overview for HPE Remote Copy and focuses on Microsoft SQL Server failover clusters with HPE 3PAR Peer Persistence. Peer Persistence provides synchronous replication for any HPE 3PAR storage installation, including Windows Server®, Linux®, or VMware® environments.

This solution uses storage replication to maintain synchronized copies of the database files on two HPE 3PAR storage devices. Data is replicated via HPE 3PAR Remote Copy with HPE Peer Persistence software[1] providing additional automated failover functionality for fast recovery times and no data loss (see Figure 1).

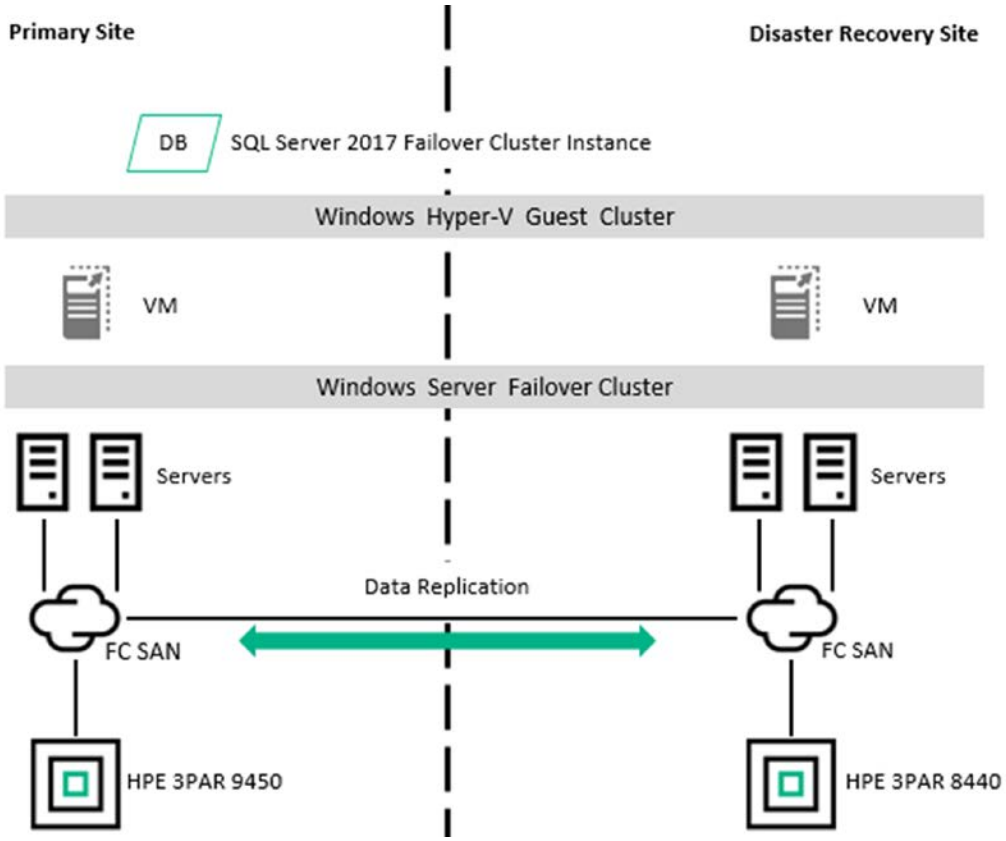

Figure 1. Microsoft SQL Server failover cluster instance with HPE 3PAR Peer Persistence highly available synchronous disaster-tolerant replication solution

Design principles

HPE 3PAR Remote Copy software provides the foundation for HPE 3PAR synchronous and asynchronous solutions by delivering the connectivity path for replication between storage systems. With a rich set of features that can be used to design disaster-tolerant solutions, HPE 3PAR cost effectively addresses the challenges of achieving a highly available, disaster-tolerant solution with HPE 3PAR Remote Copy. Remote Copy is a simple to use, efficient, and flexible replication technology that protects and shares data from any application[2].

Disaster-tolerant solutions

This section discusses some of the basic principles important to SQL Server disaster-tolerant solutions, including:

- Recovery Point Objective (RPO)
- Recovery Time Objective (RTO)
- Synchronous replication
- Asynchronous replication
- Fault domains

[1] *Break the boundaries of storage, HPE 3PAR Peer Persistence software,* http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA4-3533ENW
[2] *Disaster Tolerant Solutions with HPE 3PAR Remote Copy*, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA3-8318ENW

Solution cost tradeoff

Figure 2 displays a disaster-tolerant cost versus data-loss tradeoff. This analysis is unique to every solution. Costs are driven by the cost of the replication links between the storage systems—the greater the distance and link speed, the higher the cost. In any data replication-based, disaster-tolerant solution, the network used to replicate data between the primary site and the disaster recovery site is the key part of the solution.

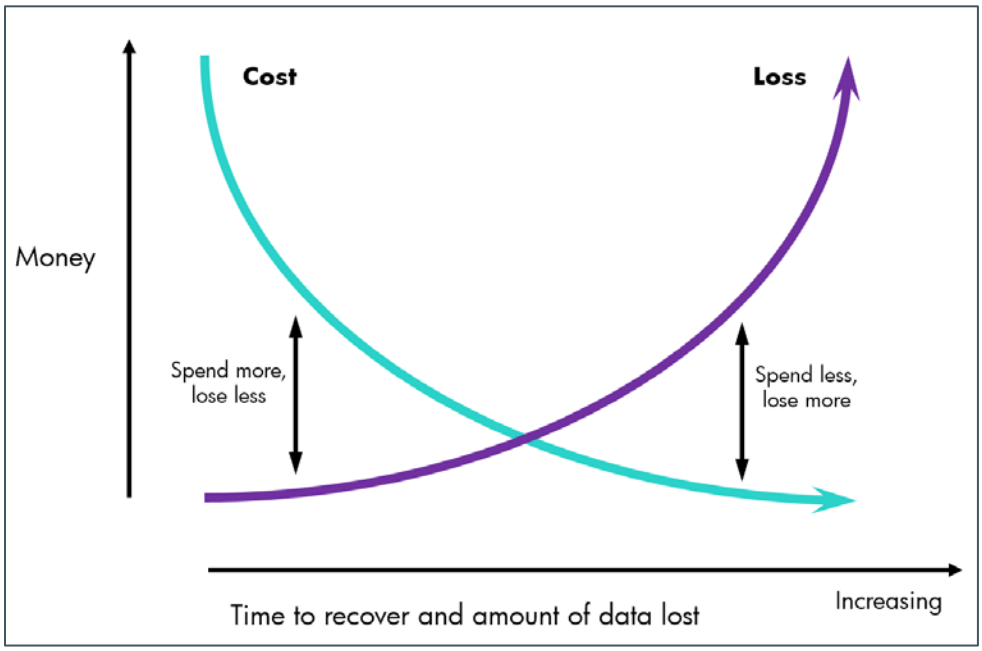

Figure 2. Disaster-tolerant solution cost-versus-loss analysis

The primary metrics on which a disaster-tolerant solution is designed and measured are Recovery Point Objective (RPO) and Recovery Time Objective (RTO). For many enterprise customers, RTO is not a very important metric. However, the maximum amount of data that can be lost in the event of a disaster (RPO) is very important. At the very core of any disaster-tolerant solution is the ability to ensure that data loss does not exceed the defined RPO for the solution following recovery from a disaster.

RPO is generally defined as an amount of time prior to the disaster where all committed transactions can be recovered. It is not defined as a given quantity of data. For example, an RPO of 20 minutes would guarantee that following a disaster that occurs at 12:20 p.m., all transactions that were committed up to and including 12 p.m. would be present in the replicated copy of the database after recovery completes. The RPO could contain transactions committed after 12 p.m., but the specified 20 minute RPO ensures transactions committed up to 12 p.m. are present.

RTO is the maximum amount of downtime that elapses following a disaster before the database should be up and running and processing new transactions. RTO is not a guarantee, but an objective to plan and build the solution. For example, specifying an RTO of 5 minutes has an objective that the database is functional by 12:25 p.m. after a disaster that occurs at 12:20 p.m.

In summary, if replication distances are short, and the link costs are not too high, a synchronous solution providing zero data loss is preferred. However, if replication link costs are too high, consider a managed data-loss solution using asynchronous data copies.

Synchronous Disaster Recovery solution

HPE 3PAR Peer Persistence software[3] is a synchronous remote-copy solution providing a zero RTO and zero RPO solution for short-distance replication.

HPE Peer Persistence is a high-availability (HA) solution between two sites with data in synchronous replication. HPE Peer Persistence can be used within VMware or Windows Server environments to provide HA storage that can extend beyond a single data center. This solution uses HPE Peer Persistence with Windows Server Failover Clusters to provide HA storage to cluster nodes that can extend beyond a single data center. Storage volumes created on one storage system are replicated to the other system using synchronous remote copy to ensure the volumes are in sync at all times.

HPE 3PAR Peer Persistence software enables the transparent replication of data from one HPE 3PAR storage system to another, while maintaining business continuity for virtualized data centers.[4] In the event of a failure in the source storage system, host traffic is redirected to the secondary storage system transparently with minimal to no impact to the hosts.

Peer Persistence helps ensure that key applications are unaffected by:

- Data loss
- Windows Live migration across sites
- Storage system failure
- Site failure

Key Peer Persistence synchronous replication features include:

- RPO (Recovery Point Objective) of 0 seconds
- RTO (Recovery Time Objective) of a few seconds of application interruption
- Automated storage disaster recovery

Asynchronous Disaster Recovery solution

Asynchronous replication provides replication without the additional host latency associated with synchronous mode replication. For asynchronous replication solutions, network bandwidth is efficiently utilized with HPE 3PAR Remote Copy Asynchronous Periodic and Asynchronous Streaming modes.

Efficiencies are built into Remote Copy algorithms. For example, the initial copy creation of target volumes does not require replication of "zero" data across the replication network (regardless of the target volume type, thick or thin). This results in faster initial synchronization and more efficient network utilization.

Periodic Asynchronous mode ensures RPO in minutes (10 minutes or greater) and can tolerate larger fluctuations in I/O traffic to a storage system for a given replication link bandwidth. In Periodic Asynchronous mode, changed data within an HPE 3PAR Remote Copy Volume Group is transferred only once—no matter how many times it might have changed—between synchronization intervals.

Asynchronous Streaming mode is the preferred solution for asynchronous replication with near-zero RPO times. Asynchronous Streaming solutions generally require the replication link speed to be sized so that 70% of the replication link bandwidth is within 95–99% of the maximum data generation rate of the data being replicated. This ensures cache on the primary array is not saturated and very small RPO is delivered.

[3] *Disaster Tolerant Solutions with HPE 3PAR Remote Copy*, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA3-8318ENW

[4] *HPE 3PAR Remote Copy Software User Guide*, page. 158, http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143

Figure 3 shows a graph of replication mode and link speeds versus RPO cost tradeoffs.

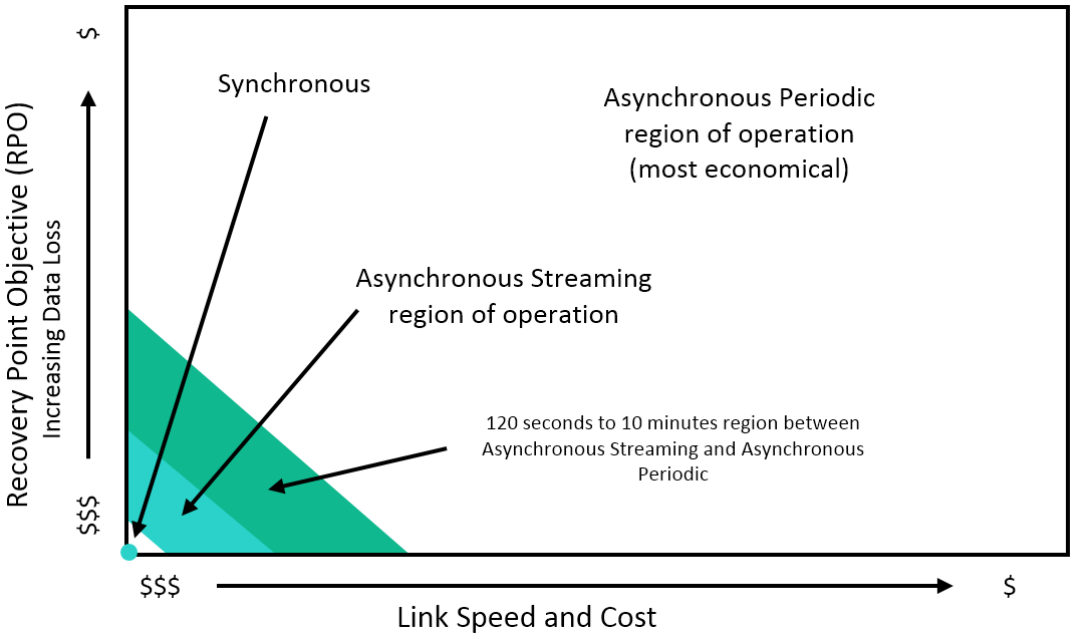

Figure 3. Link-speed cost vs. RPO cost tradeoff

Key Asynchronous Streaming replication features include:

- RPO (Recovery Point Objective) measured in seconds or even less (120 seconds maximum)
- Remote copy can withstand unlimited link failure times
- SSD performance for local storage while providing near synchronous data replication to a disaster recovery site storage

Fault domains

Fault domains are physical zones or areas of fault isolation that add physical or logical separation from the primary storage site[5]. The fault domain (also called failure domain) is a key factor for selecting either an HA or a DR solution. The zones start with the equipment rack or data center and move out to the local metro area and beyond to geographically dispersed areas. Faults include power outages, floods, fires, earthquakes, hurricanes, and other catastrophes.

As stated previously, synchronous copy solutions are preferred for a local fault domain, such as a data center or a metropolitan area, because they can provide zero data loss. In contrast, asynchronous copy solutions are preferred for geographically dispersed solutions, because they eliminate the latency impact of synchronous data copy solutions.

Local fault domain solutions are typically considered to be HA, because they can withstand a data center power failure, but not a local metro area flood. DR solutions typically can withstand metro area disasters and even geographic disasters, such as earthquakes and hurricanes. The tradeoff for these solutions is the RPO, RTO, and the solution impact to performance.

[5] Physical separation includes a different rack, city, or country. Logical separation includes different network subnets or service providers.

HPE Peer Persistence is optimally positioned as an HA solution for local metro and near geographic areas. The tradeoffs that must be considered are the needs for RPO=0 and RTO=0, minimizing replication link latency impact on host I/O, and trying to increase the fault zone size so a local disaster does not shut down all data centers. HPE Asynchronous Streaming is optimally positioned in the same zone areas as HPE Peer Persistence and slightly beyond with the tradeoff of an RPO that is not quite 0, and causing no impact to host I/O. HPE Asynchronous Periodic is optimally positioned for geographically dispersed solutions. The HPE Remote Copy modes can be joined to combine the strengths of each method into a unified replication HA/DR solution with one primary storage system that targets and replicates data to two backup systems[6].

Solution components

Figure 1 illustrates a logical solution layout. Key solution components provided by HPE and Microsoft are:

- HPE ProLiant DL380 Gen9 server
- HPE 3PAR StoreServ storage
- HPE 3PAR Peer Persistence software
- HPE SN6000B Fibre Channel switch
- Microsoft components

HPE ProLiant DL380 Gen9 server

This solution is based on the HPE ProLiant DL380 Gen9 server, which provides exceptional performance and value in a 2U form factor with a myriad of available options for processor, memory, storage, and networking configurations.

The HPE ProLiant DL380 Gen9 server is designed to address the needs of large enterprises to remote office and branch office environments. With Gen9, it delivers high performance and expandability in the Hewlett Packard Enterprise 2-processor (2P) rack portfolio. Reliability, serviceability, and near-continuous availability, backed by a comprehensive warranty, make it ideal for any environment.

Key features:

- The HPE ProLiant DL380 Gen9 server supports industry standard Intel® Xeon® E5-2600 v3 and E5-2600 v4 processors with up to 22 cores, 12 Gb/s SAS and 40 Gb NICs with a broad range compute options, and leverages 2400 MHz DDR4 HPE Smart Memory supporting up to 3.0 TB.
- It has redundant HPE Flexible Slot power supplies, an HPE Flexible Slot Battery Backup module, and support for HPE Power Discovery Services offering.
- The redesigned chassis can accommodate 8 to 24 SFF drives or 4 to 12 LFF drives, also NVMe options, and a choice of embedded 4x1 GbE, HPE FlexibleLOM, or PCIe standup 1 GbE to 40 GbE adapters. In conjunction with the embedded SATA HPE Dynamic Smart Array B140i controller for boot, data and media needs, the redesigned HPE Flexible Smart Array and HPE Smart SAS HBA controllers allow the flexibility to choose the optimal 12 Gb/s controller for your needs.
- The HPE ProLiant DL380 Gen9 server can be managed and monitored by HPE Integrated Lights-Out (iLO), HPE OneView, or HPE Smart Update Manager (SUM), with each tool providing a variety of management functions.

Figure 4 shows the HPE ProLiant DL380 Gen9 server hardware. See the *HPE ProLiant DL380 Gen9 server QuickSpecs* for more details at http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=c04375627

[6] *HPE 3PAR Remote Copy Software User Guide, page 55 Setting up SLD remote copy,* http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143

Figure 4. HPE ProLiant DL380 Gen9 server (no bezel, front facing, 8 SFF drives)

HPE 3PAR StoreServ storage

HPE 3PAR 9450 storage

HPE 3PAR 9000 storage is an enterprise-class all-flash array (AFA) that helps you consolidate primary storage workloads—for file, block, and object—without compromising performance, scalability, data services, or resiliency. This HPE 3PAR model based on the proven HPE 3PAR architecture is purpose built for all-flash consolidation, delivering the performance, simplicity, and agility needed to support your hybrid IT environment. HPE 3PAR 9000 storage is available in a single all-flash model—the 9450—that offers rich Tier-1 data services, quad-node resiliency, fine-grained Quality of Service (QoS), seamless data mobility between systems, high availability through a complete set of persistent technologies, and simple and efficient data protection with a flat backup to HPE StoreOnce Backup appliances.

Figure 5 shows the HPE 3PAR 9000 storage hardware. See the *HPE 3PAR 9000 storage QuickSpecs* for more details at http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=a00005876enw

HPE 3PAR 8440 storage

HPE 3PAR 8000 storage offers enterprise Tier 1 storage at a midrange price. HPE 3PAR 8000 storage delivers the performance advantages of a purpose-built, flash-optimized architecture without compromising resiliency, efficiency, or data mobility. With unmatched versatility, performance, and density, HPE 3PAR 8000 storage provides a range of options that support true convergence of block and file protocols, AFA performance, and the use of spinning media to further optimize costs. HPE 3PAR 8000 storage offers rich, Tier-1 data services, quad-node resiliency, seamless data mobility between systems, high availability through a complete set of persistent technologies, and simple and efficient data protection with a flat backup to HPE StoreOnce Backup appliances. Four models are available: 8200, 8400, 8440, and 8450. You can start small and grow without painful upgrades down the road.

Figure 5 shows the HPE 3PAR 8000 storage hardware. See the *HPE 3PAR 8000 storage QuickSpecs* for more details at http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=c04607918

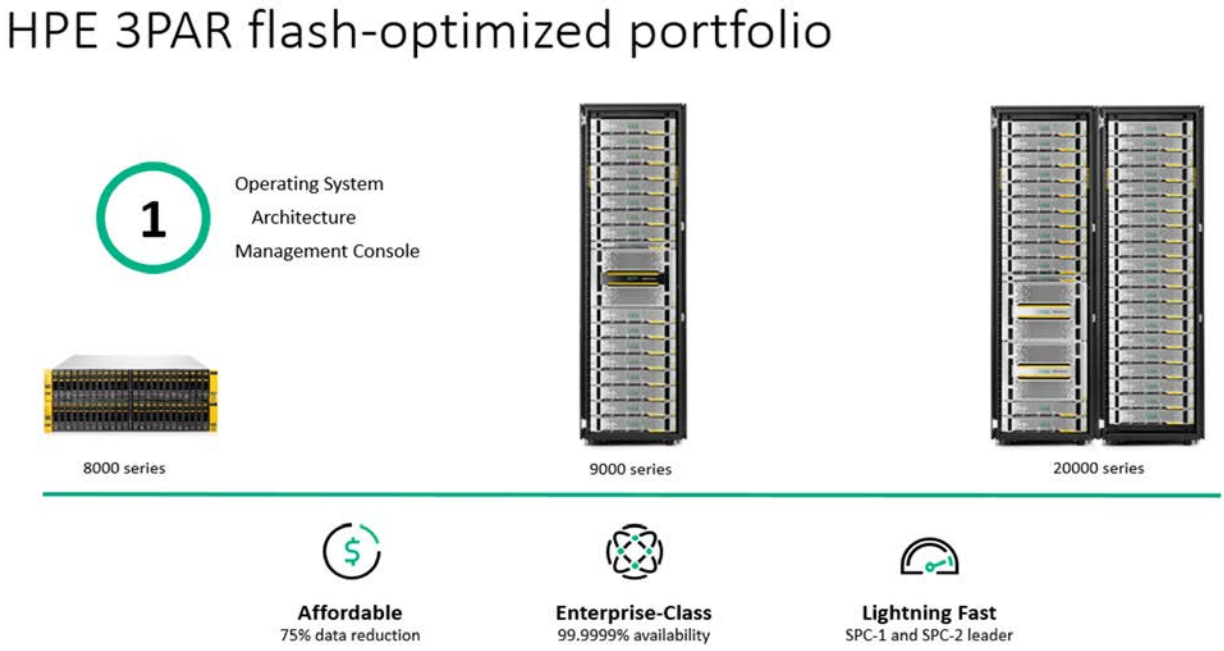

Figure 5. HPE 3PAR StoreServ storage flash-optimized portfolio

HPE 3PAR Peer Persistence software[7]

With HPE 3PAR Peer Persistence (Peer Persistence) software, HPE 3PAR storage systems located at metropolitan distances can act as peers to each other, presenting a highly available transparent failover storage system to hosts and servers connected to them. This capability allows you to configure a high-availability solution between two sites or data centers where failover and failback remains completely transparent to the hosts and applications running on those hosts.

Compared to the traditional failover models, where upon failover the hosts must be restarted, HPE Peer Persistence software allows hosts to remain online serving their business applications—even when their storage switches from their original site to the disaster recovery (DR) site, resulting in a much improved recovery time. HPE Peer Persistence software achieves this key enhancement by taking advantage of the Asymmetric Logical Unit Access (ALUA) capability that marks paths to a SCSI device with different characteristics.

HPE Peer Persistence software can failover the storage either manually or automatically from the primary to the secondary storage system transparently to the hosts, reducing service disruptions. Using ALUA, the paths to the source volumes are presented as active, while the paths to the target volumes are presented as standby. During a Peer Persistence failover operation these replication roles are reversed—the source volume becomes the target and the target volume becomes the source.[8]

HPE Peer Persistence is an HA solution in a disaster-tolerant configuration between two sites with data in synchronous replication. In this Reference Architecture, HPE Peer Persistence is used with Windows Failover Clusters to provide HA storage to cluster nodes that can extend beyond a single data center. Storage volumes created on one storage system are replicated to the other system using synchronous remote copy to ensure volumes remain in sync.

HPE Peer Persistence software enables the transparent migration of application data from one HPE 3PAR storage system to another, maintaining business continuity for virtualized data centers. In the event of a failure in the source storage system, host traffic is redirected to the secondary storage system transparently and without major impact to the hosts.

[7] *Disaster Tolerant Solutions with HPE 3PAR Remote Copy*, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA3-8318ENW
[8] *HPE 3PAR Remote Copy Software User Guide, page 174 Path management policies,* http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143

HPE 3PAR StoreServ Management Console

HPE 3PAR StoreServ Management Console (SSMC) is the HPE 3PAR management and reporting console that offers converged management of both File and Block on HPE 3PAR storage systems.

HPE SSMC is installed on another server that has network access to the HPE 3PAR storage systems. After installing HPE SSMC, configure an administrator and then add the storage systems.

Figure 6 shows the HPE SSMC Dashboard view after the storage systems have been added. The drop-down menu for selecting other views is labeled **3PAR StoreServ** and shown in the upper left of the figure. A help facility is available in the upper right that opens the frame to the right with additional help selections. A good place to start is the **Help → Navigation tutorial** shown in the **Help** area, which provides a quick guide for using the user-friendly management console.

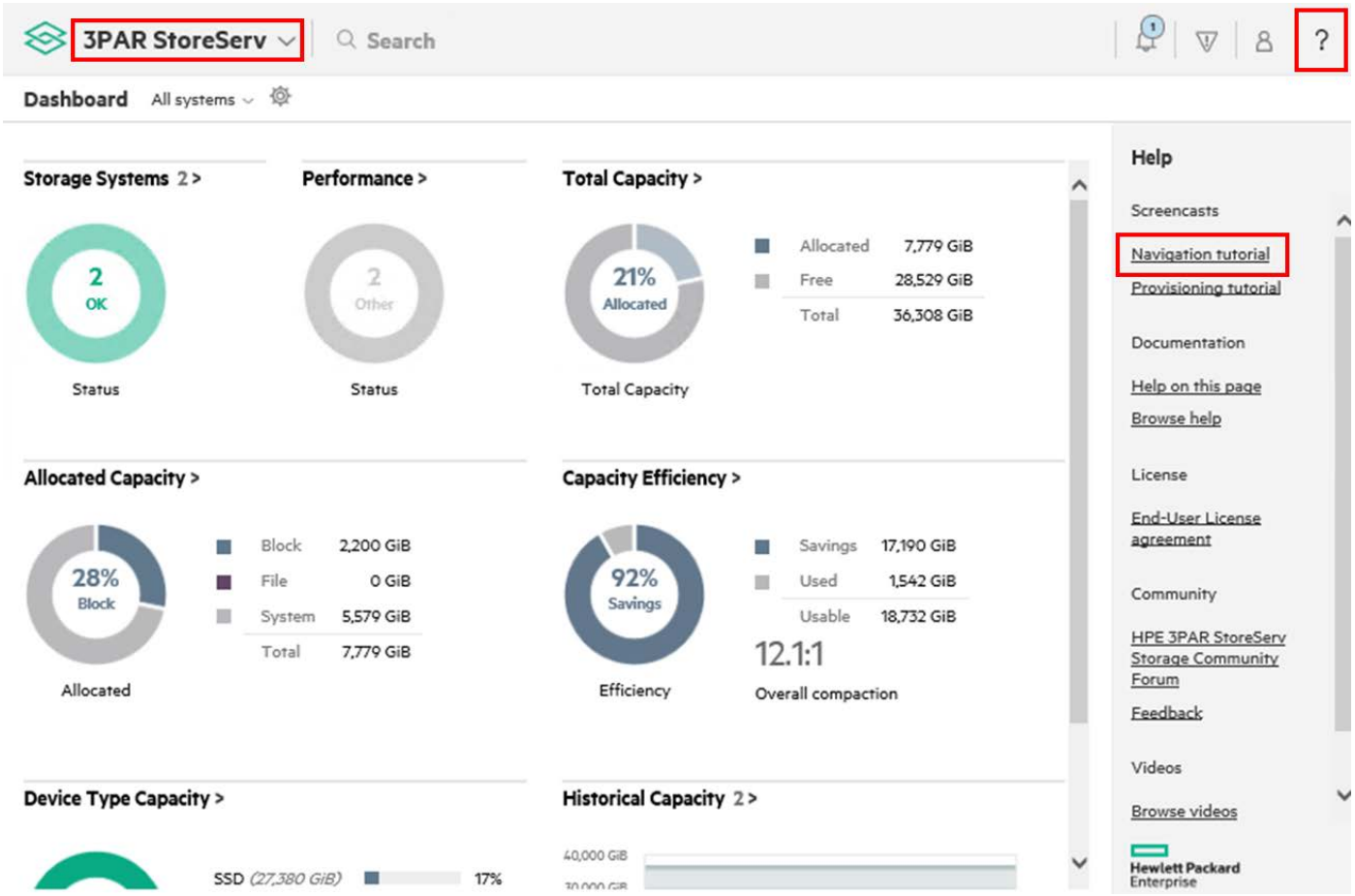

Figure 6. HPE SSMC interface showing Dashboard view, drop-down menu, and help facility

Figure 7 shows an example of the Navigation tutorial help. The Main Menu selector in the upper left allows you to select other views, such as Remote Copy and volume configuration areas. The details panel drop-down selector provides access to additional views and configuration areas.

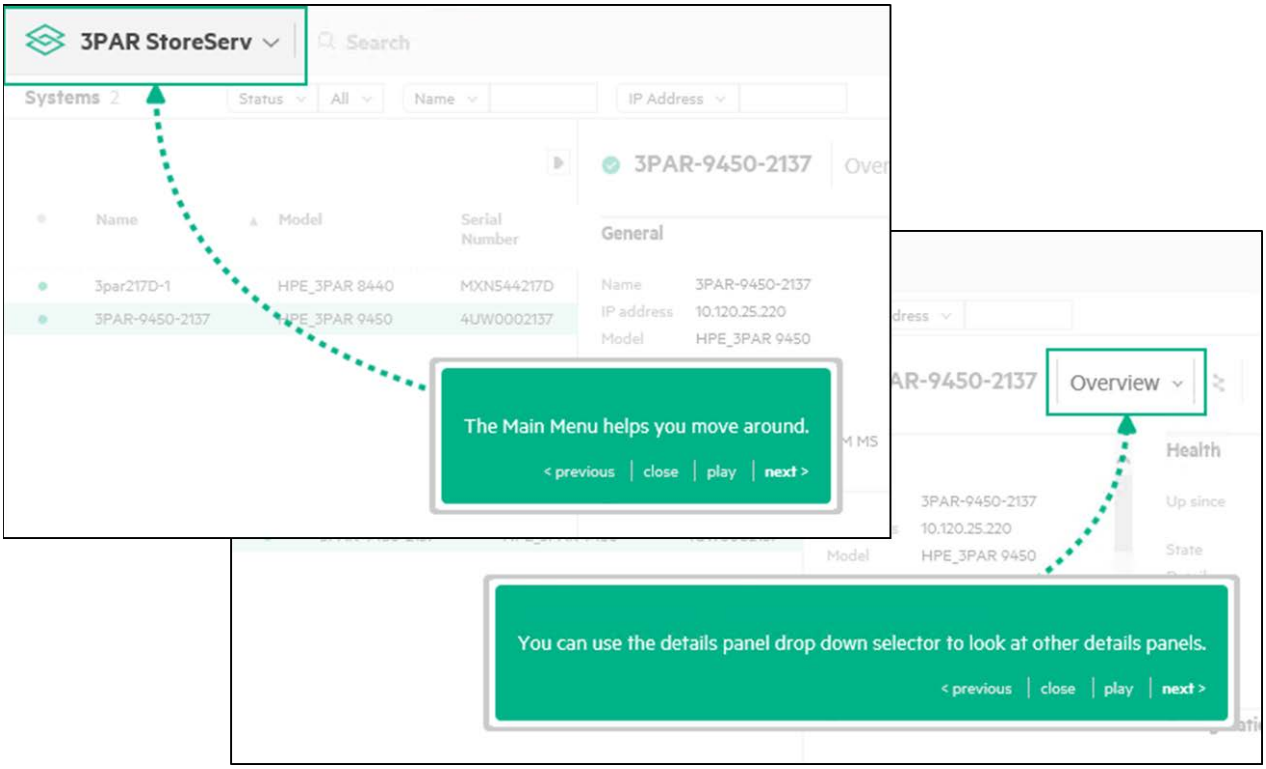

Figure 7. HPE SSMC showing two Navigation tutorial screenshots: the Main Menu location, and the details panel drop-down selector

For this solution configuration, the **Remote Copy Configurations** area is found by selecting the Main Menu **Show all** toggle, as seen in upper right of Figure 8. The **Remote Copy Configurations** allows for easier Remote Copy configuration and monitoring of the HPE Peer Persistence solution.

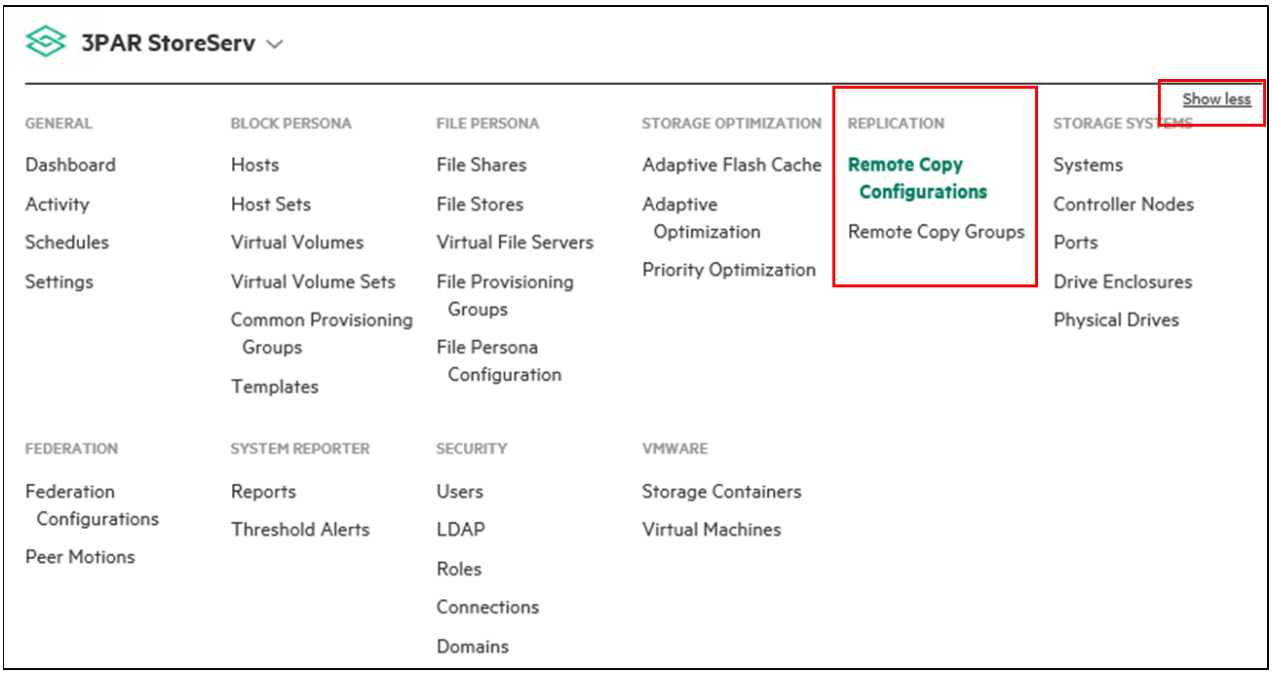

Figure 8. Main Menu after the "Show less" | "Show all" toggle is set to Show all (Replication area showcased)

HPE SN6000B Fibre Channel switch

In this solution, the HPE SN6000B 16 Gb 48-Port Fibre Channel switches are used for SAN connectivity.

The HPE SN6000B Fibre Channel switch meets the demands of hyper-scale, private cloud storage environments by delivering market-leading 16 Gb Fibre Channel technology and capabilities that support highly virtualized environments. Designed to enable maximum flexibility and investment protection, the HPE SN6000B Fibre Channel switch is configurable in 24, 36, or 48 ports and supports 4, 8, 10, or 16 Gbps speeds in an efficiently designed 1U package. It also provides a simplified deployment process and a point-and-click user interface making it both powerful and easy to use. The HPE SN6000B Fibre Channel switch offers low-cost access to industry-leading Storage Area Network (SAN) technology while providing “pay-as-you-grow” scalability to meet the needs of an evolving storage environment.

Figure 9 shows the rear view of the HPE SN6000B Fibre Channel switch. For more details, see the *HPE SN6000B Fibre Channel Switch QuickSpecs* at http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=c04111492.

Figure 9. HPE SN6000B Fibre Channel switch (rear view)

Microsoft components

Microsoft Windows Server 2016 is the operating system used for the solution, providing key business components such as Hyper-V, failover clustering, guest clustering[9], and storage connectivity drivers.

Windows Server 2016

The Hewlett Packard Enterprise and Microsoft partnership combines technologies to help your business reach its full potential and operate with increased agility, extra layers of security, higher resiliency, and performance for your Windows-based servers. Microsoft Windows Server 2016 is the cloud-ready operating system that supports your most demanding workloads while introducing new technologies that make it easy to integrate hybrid cloud services. It delivers powerful new layers of security along with innovations to elevate your business to new levels of efficiency and productivity. Windows Server 2016 paired with HPE servers, storage, and networking solutions provides the technology that helps achieve your business goals.[10]

Shared virtual hard disks (VHD Sets) are a new feature in Windows Server 2016. The shared VHD Set is for use with Windows Server 2016 guest operating system. VHD Sets are used in this solution[11].

Windows Failover Clustering (WSFC)

This solution uses two levels of clustering at the server level. As shown Figure 1, the parent servers are clustered, and then the virtual machines belonging to the parent cluster are members of their own Hyper-V guest cluster.

Clustering provides HA, which means it provides recoverability in the event of a failure. A benefit of Hyper-V guest clustering is that it expands the benefits of Windows virtualization in the event of a failed VM, because the application can restart on another running VM without waiting for the VM to restart. Using clustered VMs helps to minimize the client impact on the application.

This solution does not use Windows Server Network Load Balancing clusters, which provide different functionality.

Microsoft SQL Server 2017[12]

Microsoft SQL Server 2017 enables customers to build mission-critical applications and Big Data solutions using high-performance, in-memory technology across online transaction processing (OLTP), data warehousing, business intelligence, and analytics workloads. Microsoft SQL Server 2017 uses a common set of tools to deploy and manage databases both on-premises and in the cloud, which makes it easier for customers with existing skills to take advantage of the cloud.

SQL Server Always On Failover Cluster Instance

A SQL Server Always On Failover Cluster Instance (FCI) is a single instance of SQL Server installed on the Windows Server Failover Clustering (WSFC) nodes.[13] The SQL Server FCI runs on only one cluster node (server) at a time. A benefit of an FCI is a highly available database; if one cluster node fails, another node is available for the SQL Server instance to start on.

SQL Server Always On FCI can also be used in a stretched cluster with replicated shared storage at a local and remote Disaster Recovery site for a highly available, disaster-tolerant solution. HPE 3PAR Peer Persistence software helps enable a disaster-tolerant synchronous HPE 3PAR storage solution with SQL Server.

A benefit of using HPE 3PAR Peer Persistence software with Windows clustering is that all I/O to storage can be synchronously mirrored to another storage system at the storage level without server interaction or overhead. What this means for SQL Server is that even if there is a site failure, when SQL Server starts again at the Disaster Recovery site, all data acknowledged by the storage solution is fully available for a crash-consistent startup. The RPO is 0, which helps reduce the business impact of catastrophic events and helps minimize the database recovery time.

[9] Microsoft, *Using Guest Clustering for High Availability*, https://technet.microsoft.com/en-us/library/dn440540.aspx
[10] *Microsoft Windows Server 2016*, hpe.com/us/en/product-catalog/detail/pip.microsoft-windows-server-2016.1009480090.html, *HPE OEM Microsoft Windows Server 2016*, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA6-8899ENW
[11] Microsoft, *Create Hyper-V VHD Set files*, https://docs.microsoft.com/en-us/windows-server/virtualization/hyper-v/manage/create-vhdset-file
[12] Microsoft, *SQL Server 2017*, microsoft.com/en-us/sql-server/sql-server-2017
[13] Microsoft, *Always On Failover Cluster Instances (SQL Server)*, https://docs.microsoft.com/en-us/sql/sql-server/failover-clusters/windows/always-on-failover-cluster-instances-sql-server

Best practices and configuration guidance for the solution

Several documents to consult for this solution are:

- *HPE 3PAR Windows Server 2016/2012/2008 Implementation Guide*, http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c04448812
- *HPE 3PAR StoreServ Storage Concepts Guide*, http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c04204225
- *HPE 3PAR StoreServ Storage best practices guide*, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA4-4524ENW
- *HPE 3PAR Remote Copy Software User Guide*, http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143
- *HPE Single Point of Connectivity Knowledge* (requires HPE SPOCK log in), hpe.com/storage/spock

HPE offers a Peer Persistence startup service for Peer Persistence called the HPE 3PAR Peer Persistence Software Installation and Startup Service. (See *HPE 3PAR Peer Persistence Software Installation and Startup Service* at http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA4-2772ENW for details.)

HPE ProLiant servers

The servers were configured to boot from internal storage and were each configured with a 2-port 16 Gb FC adapter for redundancy and throughput capability. All FC ports were attached to the redundant SAN for storage connectivity.

Best practices are:

- The servers should all be as identical as possible, with a few exceptions such as using different network cards.
- Confirm the most recent HPE Service Pack for ProLiant (SPP)[14] is applied before OS installation. Confirm that the server is fully patched with the most current HPE SPP after installing the OS.

FC SAN

Refer to the HPE SAN design guide[15] for supported SAN configurations. Chapter 16 describes the HPE 3PAR StoreServ storage specific rules that apply to this solution.

Follow these two requirements for this HPE Peer Persistence solution (requirements might vary for different replication modes):

- Volumes must be presented and zoned to hosts at the local and remote site, which allows the disk to be used as a source or target device of each cluster node. As an example, the database LUN must be zoned so it is visible to all servers in the cluster.
- Remote Copy port pairs should be zoned to each other for synchronous communications links.

Remote Copy zones

Peer Persistence builds on an HPE 3PAR StoreServ Remote Copy (RC) implementation.

A dedicated Remote Copy (RC1) Ethernet port on each storage node can be used for Remote Copy connectivity, but this could limit the synchronous copy throughput if any of the links are 1 Gbps. Using the 1 Gbps ports is adequate for some Remote Copy solutions, but not this SSD solution, because the HPE 3PAR 8440 1 Gbps ports would limit replicated host I/O to about 1000 Mbps.

Best practices are:

- Use at least one port per node for Remote Copy functionality.
- Match identically numbered ports between storage systems for easier documentation and troubleshooting. For example, match array A node 0 port 0:1:1 to array B node 0 port 0:1:1. The port numbering format is node:slot:port.

[14] *HPE ProLiant Server SPP information on Hewlett Packard Enterprise Information Library*, http://h17007.www1.hpe.com/us/en/enterprise/servers/solutions/info-library/index.aspx?cat=smartupdate&subcat=spp#.WcP9bU2WxoC

[15] Find the *SAN Design Guide* in the Design Guides area in the left website frame. Found at *HPE Storage Single Point of Connectivity Knowledge* (requires HPE SPOCK log in), at https://h20272.www2.hpe.com/spock/

For this solution with differing storage port layout options, the RC port numbers are not able to be matched up between the arrays. For example, the HPE 3PAR 8440 storage has four built-in 16 Gb/sec FC ports per node pair, yet the HPE 3PAR 9450 storage does not have any built in FC ports.

In this solution, the HPE 3PAR 9450 storage uses a single four port FC adapter in each node for both RC and host connect, as shown in Figure 10.

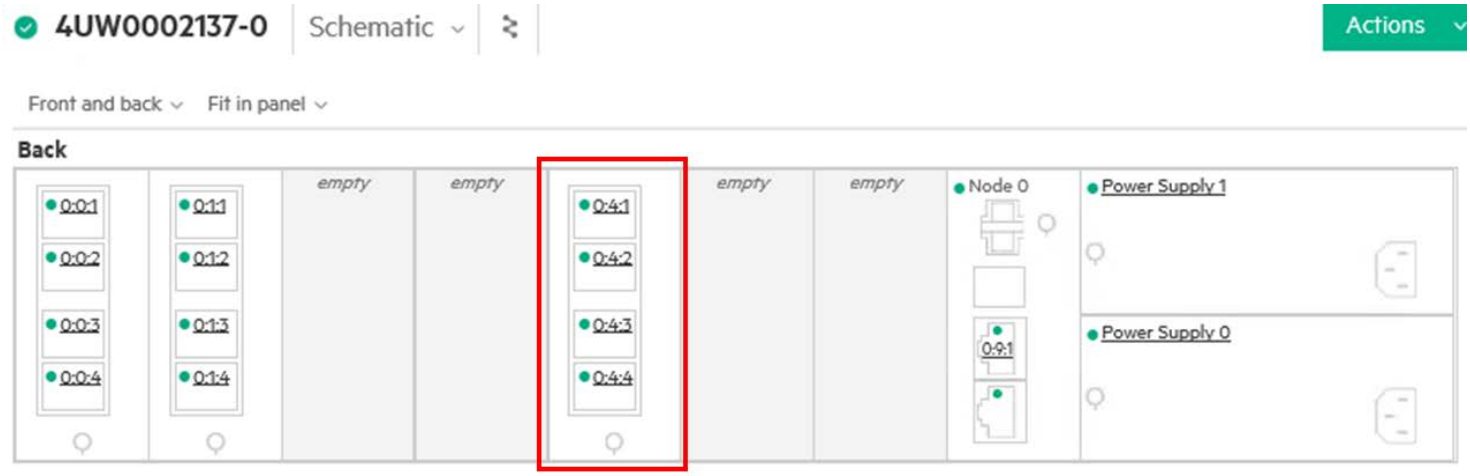

Figure 10. HPE 3PAR 9450 storage node 0 (typical) SSMC schematic view showing FC adapter ports used for Remote Copy and host connectivity

For the HPE 3PAR 8440 storage, the four (two per node) built-in FC-1 and FC-2 ports are used for Remote Copy. Optional FC adapters are used for host connectivity, as shown in Figure 11.

Figure 11. HPE 3PAR 8440 storage node 0 (typical) SSMC schematic view showing FC adapter ports used for host connectivity and built-in ports used for Remote Copy

The RC FC zone information is shown in Table 1. The RC transport layer is configured later as discussed in the HPE 3PAR RC transport configuration section.

Table 1. Remote Copy FC SAN zones for HPE 3PAR Peer Persistence software

Zone name	Fabric A members	Fabric B members
ZoneRCFC_0	9450 port 0:4:1 8440 port 0:0:1	Not applicable
ZoneRCFC_1	Not applicable	9450 port 1:4:2 8440 port 1:0:2
ZoneRCFC_2	9450 port 2:4:1 8440 port 1:0:1	Not applicable
ZoneRCFC_3	Not applicable	9450 port 3:4:2 8440 port 0:0:2

This RCFC zoning layout follows RC node-pair rules and fabric-connection best practices by:

- Connecting odd ports to one fabric and even ports to the other fabric
- Using a single zone per a connected pair of ports
- Alternating port connections between fabrics A and B

Server connectivity zones

FC SAN zones were used for all storage connects. Zoning by host bus adapter (HBA) port is an HPE-recommended zoning method for HPE 3PAR storage (a single initiator port to multiple target ports per zone using World Wide Port Name [WWPN]). For example, a zone consists of a single server FC HBA port and multiple storage target ports. Table 2 lists the zoning for one cluster server node and is typical for the other three servers. Only the zone name changes. Refer to Figure 10 and Figure 11 for an array schematic view of typical port numbers.

Table 2. Cluster node FC SAN zones for one cluster node, typical for each node

Zone name	Server alias	Fabric A members	Fabric B members
Zone_Clus63_20s18_1	Node20s18_HBA1p1	node20s18 HBA 1p1 9450 port 0:4:3 9450 port 1:4:3 8440 port 0:2::3 8440 port 1:2::3	Not Applicable
Zone_Clus63_20s18_2	Node20s18_HBA1p2	Not Applicable	node20s18 HBA 1p2 9450 port 2:4:4 9450 port 3:4:4 8440 port 0:2::4 8440 port 1:2::4

Important

For zoning, all hosts must have access to all storage systems, because all HPE Peer Persistence volumes must be accessible from both storage systems for high availability and disaster recovery.

HPE 3PAR Remote Copy with Peer Persistence synchronous replication

Major items to configure for the Remote Copy Peer Persistence configuration are:

- Controller node connections
- HPE 3PAR Peer Persistence software

- Fibre Channel switch zoning for RC
- HPE 3PAR RC port configuration
- HPE 3PAR RC transport configuration
- HPE Peer Persistence Quorum Witness
- HPE 3PAR RC Group
- HPE Peer Persistence manual switchover and automatic failover

The HPE 3PAR configuration steps can be configured via the storage system CLI interface or via the HPE SSMC. Consult the *HPE 3PAR Remote Copy Software User Guide* at http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143 for details and for CLI instructions. HPE Peer Persistence with transparent failover information is discussed.

Controller node connections

HPE 3PAR storage is connected to the network for management access and for Peer Persistence Quorum Witness (QW) usage. Network connectivity is via the MGMT port on each storage node. For redundancy, at least two separate network connections should be established for each storage system. A single IP address is configured for each HPE 3PAR storage system.

Console access for each storage node is via the RS-232 **Mfg** port.

The RC1 Ethernet port on the storage nodes was not used in this solution as discussed in the FC SAN section of this paper.

For the HPE 3PAR 8440, when using Remote Copy over Fibre Channel, additional optional FC HBAs are recommended, because the built-in ports will not offer sufficient connectivity[16]. For the HPE 3PAR 9450, there are no built in FC ports, so an optional FC adapter is added to each node for server and RC usage.

HPE 3PAR Peer Persistence software

HPE Peer Persistence software and HPE 3PAR OS version 3.3.1 is a license update. For multiple systems, add a single All-inclusive Multi-system software option for Remote Copy, Peer Motion, Peer Persistence, and CLX[17].

HPE 3PAR Peer Persistence software using Hyper-V Quorum Witness and HPE 3PAR 9000 storage requires the following conditions:

- HPE 3PAR StoreServ storage OS version 3.3.1 or newer
- Synchronous mode Remote Copy 1-to-1 configuration
- Maximum round trip network latency for Remote Copy synchronous mode replication link is 10 milliseconds[18]
- Quorum Witness virtual machine installed at a third site
- The Quorum Witness reachable from both storage systems
- Static IP addresses for storage and the Quorum Witness
- Quorum Witness maximum round trip network latency (RTT) is 250 milliseconds[19].
- Each Peer Persistence volume pair presented to the servers with the same volume Worldwide Name (WWN). All volumes presented to all servers.
- Storage volumes with snapshot space, such as commonly provisioned virtual volumes (CPVVs) or thinly provisioned virtual volumes (TPVVs)
- Data replication links to maintain the synchronous copy of data between the storage systems

[16] *HPE 3PAR StoreServ 8000 storage QuickSpecs*, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=c04607918. Go to Customer Self Installation → Step 2 - Choose Host Adapter, see Note under Adapter Configuration table

[17] *HPE 3PAR StoreServ 9000 storage QuickSpecs*, *Step 8 Choose Software*, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=a00005876enw

[18] *HPE 3PAR StoreServ 3PAR OS 3.3.1 Support Matrix* found on HPE SPOCK at https://h20272.www2.hpe.com/spock/ under Other Hardware → 3PAR → Other HPE 3PAR Support Matrices → HPE 3PAR Support Matrices

[19] *HPE 3PAR Remote Copy Software User Guide*, *3PAR Quorum Witness requirements*, http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143

Fibre Channel switch zoning for RC

Ensure zone members are not in any other zones and the zones consist of only the two RC port members. Refer to the **FC SAN** Remote Copy zones section of this document for additional information.

HPE 3PAR RC port configuration

HPE 3PAR ports used for RC over FC must be set to type **RCFC**. This is easily accomplished via the HPE SSMC. Set the connection mode to RCFC; do not set the connection mode to Peer. The configuration steps are (Figure 12):

1. Select **Ports** from the Main menu selector.
2. Select the port to configure.
3. Select **Settings** from the details drop-down selector.
4. Hover to the right of the **FC Settings** area to make the **Edit** selector visible. Specify the **Connection mode** to be **RCFC**.
5. After the configuration process completes, the **Port Type** should display as **RC**, as shown in Figure 12.

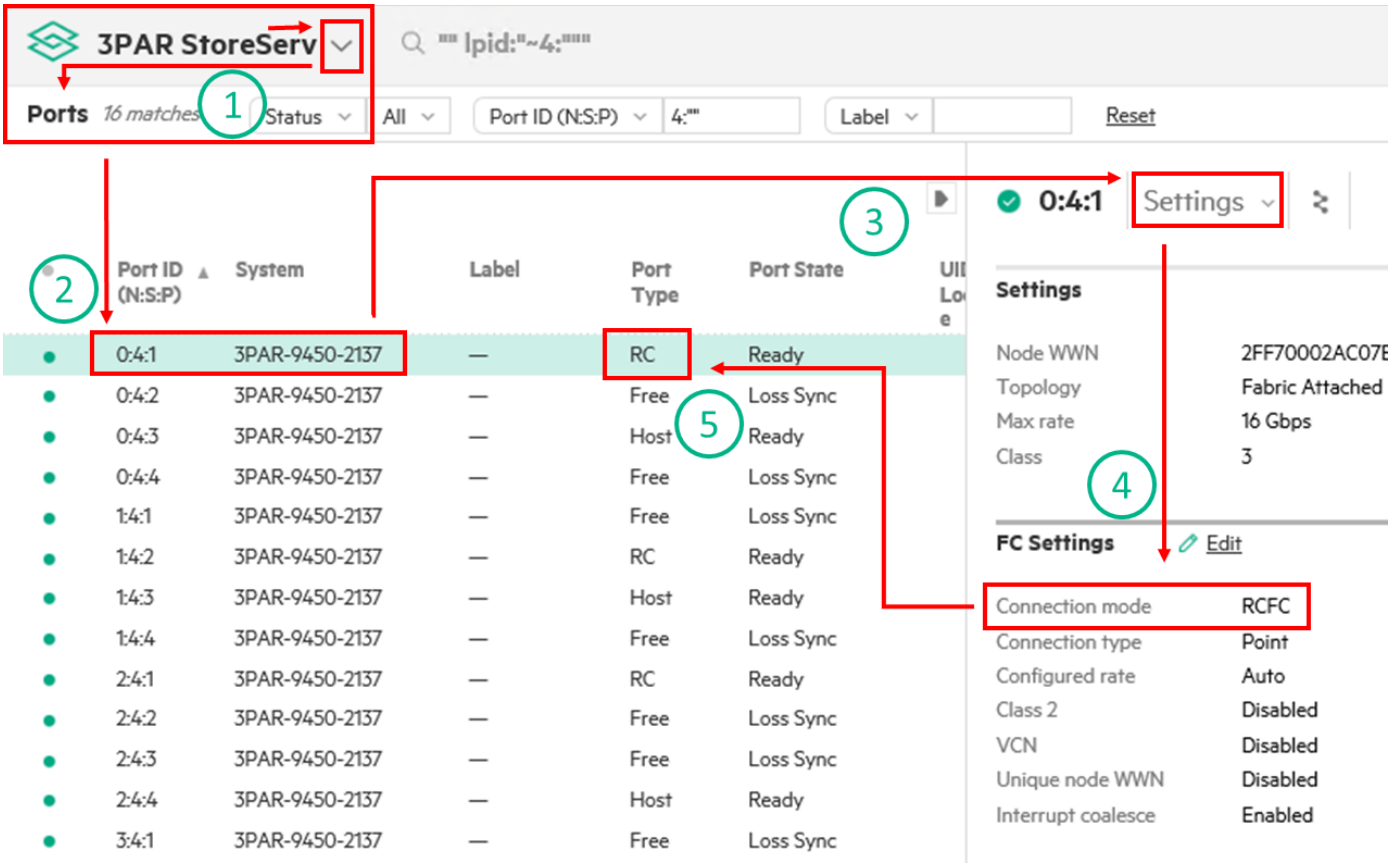

Figure 12. HPE SSMC RC port connection mode set to RCFC in **Port** → **Settings** → **FC Settings** area

HPE 3PAR RC transport configuration

After RC ports are zoned and have their port type set for RC (earlier steps), create the RC transport connections. Transport connections are also referred to as RC links.

From HPE SSMC, the RC transport links are configured in the Remote Copy Configurations area as shown in Figure 13. Use the **Create configuration** selector to create the transport links. Use the details panel drop-down selector to view link status and select other information areas to display.

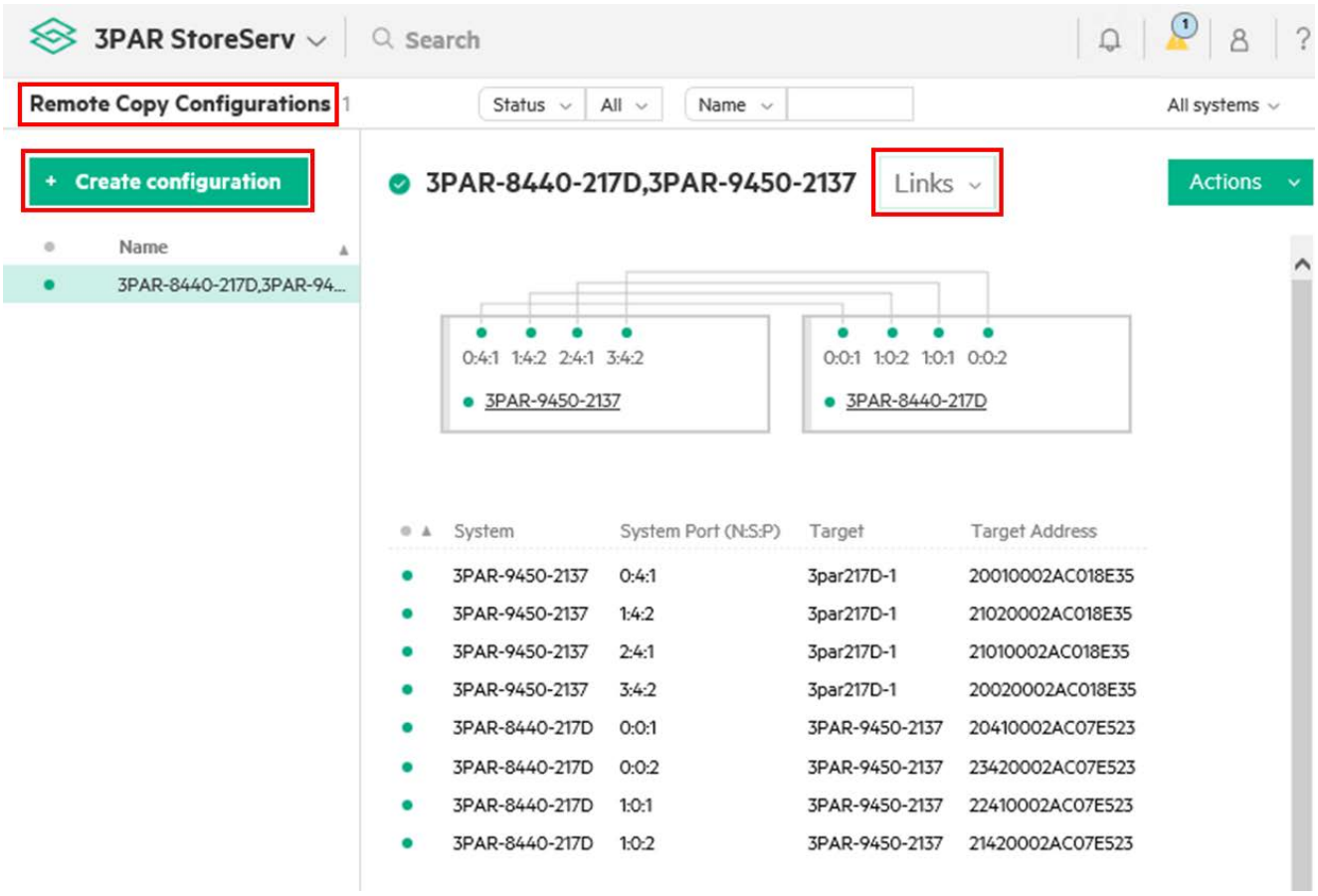

Figure 13. HPE SSMC **Remote Copy Configurations** area used to create the transport links and review remote copy information.

HPE Peer Persistence Quorum Witness

The Quorum Witness virtual machine is installed at a third site and used by the arrays to provide arbitration for Peer Persistence functionality.

HPE Peer Persistence Quorum Witness (QW) software facilitates automatic transparent failover within a Peer Persistence environment. It is run on a separate server (ESX or Hyper-V environment) and is not part of the host cluster. It is configured and connected to both arrays via a non-remote-copy Ethernet link.

Review current Peer Persistence support via the SPOCK[20] website in **Software** (found the left frame of the window) → **Array SW: 3PAR** → **HPE 3PAR Operating System Software: Array Software** → *HPE 3PAR Peer Persistence Host OS Support Matrix* (https://h20272.www2.hpe.com/spock/utility/document.aspx?docurl=Shared%20Documents/sw/array_3par/3par_peer_persistence.2017_03_30.pdf)

Quorum Witness deployment information is found in the *HPE 3PAR Remote Copy Software User Guide*[21] in the *3PAR Quorum Witness deployment* section. If importing the Hyper-V image, considerations are:

- If the import operation has trouble and reports a file format issue, try copying the files from the mounted ISO to a different location on an NTFS volume, such as the C drive.
- Make sure the virtual machine network settings are correct after importing the VM, including any VLAN ID settings that might have been imported.
- Confirm the `ifcfg-eth0` file has the correct BROADCAST, NETMASK, and NETWORK info, and consult the *HPE 3PAR Remote Copy Software User Guide*, *Configuring a static IP address* section.

After the QW is configured and running, add the QW to the RC configuration via the HPE SSMC. Add the QW IP address via the Configure quorum witness action (**SSMC → Main Menu → Remote Copy Configuration → Actions → Configure quorum witness**).

HPE 3PAR RC Group

Table 3 shows important RC group Peer Persistence settings for correct and automated functionality.

When creating Peer Persistence RC Groups, ensure the **Path Management** policy is set. This option ensures the source and target volumes report the correct multipath ALUA information as discussed in the HPE 3PAR Peer Persistence software section.

The Peer Persistence RC Group **Auto Failover** policy is set to enable the automatic failover of an RC group when used in conjunction with the Quorum Witness functionality. If the auto failover policy is set, in the event of a disaster at the primary data center, Peer Persistence will be allowed to automatically failover the volumes at the disaster recovery site. A more thorough discussion is provided in the HPE 3PAR Remote Copy software user guide.[22]

The auto recover policy is not set by default, requiring manual intervention to restart the synchronization process after the remote-copy links go down and back up. Enabling auto recover specifies that if the remote copy is stopped as a result of the remote-copy links going down, the remote-copy group is restarted automatically after the links come back up. The best practice is to enable auto recover. [23]

All database volumes and files (including temp) and the volumes that contain the virtual machines are included in the same RC Group for consistency of replicated I/O to the DR site. In an asynchronous DR solution, a disaster at the primary data center will have data loss; only the database data and transaction log volumes would be included in a consistency group. For this synchronous DR solution, a disaster at the primary data center will have no data loss. All acknowledged data will be at both data centers when the data replication links are up and the volume pair sync/status is **Synced**.

To help understand why all the database and virtual machine volumes would be included in the same remote copy consistency group, remember that any I/O to any of those volumes must be 100% consistent in the event that a switchover occurs. In this situation, both arrays must have exactly the same data in the same write order for both the database volumes as well as the virtual machine volume running the database.

[20] *HPE Storage Single Point of Connectivity Knowledge* (SPOCK), https://h20272.www2.hpe.com/spock/
[21] *HPE 3PAR Remote Copy Software User Guide*, http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143
[22] HPE 3PAR Remote Copy Software User Guide, http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143, page 155, 3PAR Peer Persistence with transparent failover
[23] HPE 3PAR Remote Copy Software User Guide, http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143, page 185, Restarting synchronous volume groups after link recovery

Table 3. RC group configuration information for Peer Persistence

Item	Best Practice	Comment
User and Copy CPG	Same	A copy CPG is needed for RC functionality. Use the same CPG for both the virtual volume (VV) user CPG and the copy CPG.
Path management	Set	Set this policy from the CLI with the `setrcopygroup` command or from the SSMC for the RC group used with Peer Persistence.
Auto failover	Set	Set this policy from the CLI with the `setrcopygroup` command for the RC group used with Peer Persistence for automatic failover.
Auto recover[24]	Enabled	Specifies that if the remote copy is stopped as a result of the remote-copy links going down, the group is restarted automatically after the links come back up.

Figure 14 shows the HPE SSMC RC group interface for important Peer Persistence settings for correct and automated functionality. The path management and auto failover policies are not set by default and must be manually configured.

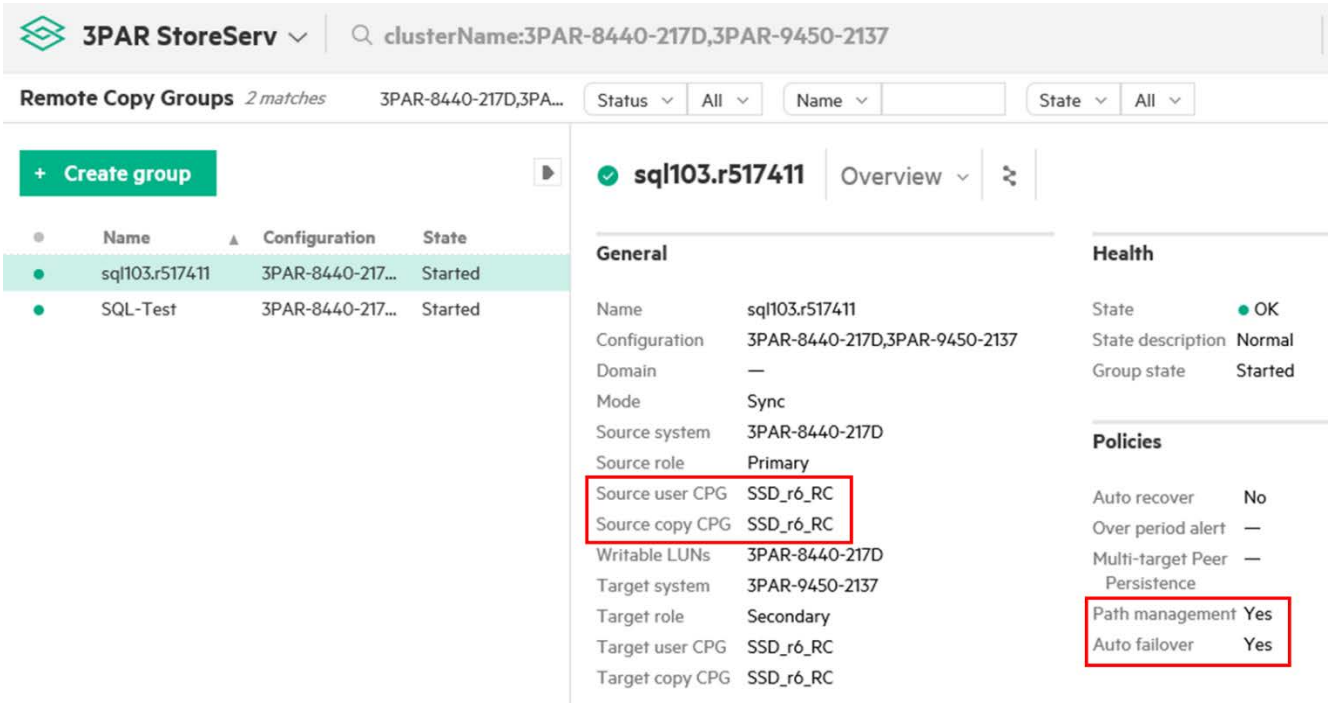

Figure 14. HPE SSMC Remote Copy Group display showing the user and copy CPG settings as well as the important path management and auto failover policies

HPE Peer Persistence manual switchover and automatic failover

HPE Peer Persistence provides for a manual switchover or an automatic failover of the Remote Copy group. A manual switchover can be used for testing and maintenance, while automatic failover can be enabled for disaster recovery operations.

The manual transparent switchover is initiated in one of two ways:

- The `setrcopygroup switchover` command from the CLI interface
- The `switchover` command from the HPE SSMC

[24] Auto recover is set via the CLI setrcopygroup, auto_recover policy, or via the HPE SSMC → Remote Copy Groups → Edit →Advanced options → Auto recover → Enabled

The `switchover` command is only supported for synchronous remote copy groups that are started and synced. The command must be issued on the primary array and upon completion of the command, the remote copy group will be restarted in the reverse direction. Use `showtask -d <task#>` and `showrcopy` to confirm results.

The automatic transparent failover, when enabled, is initiated when the primary data center experiences a catastrophic event. The automatic transparent failover is triggered by the Quorum Witness infrastructure and related software, which begins state transition and activation of the secondary host paths for all applicable secondary volume groups. For an automatic failover operation, replication will stop, and remain that way until the failed system is recovered.

Multipath I/O (MPIO)

Multipath functionality must be configured correctly for regular HPE 3PAR storage usage and then for HPE Peer Persistence.

MPIO HPE 3PAR device

Install and configure the Microsoft Multipath I/O (MPIO) feature on the failover cluster server nodes. Until the MPIO feature is installed and configured, each HPE 3PAR presented volume will appear multiple times, depending on the configured paths. Configure the MPIO properties for the HPE 3PAR storage by using the **3PARdataVV** device hardware ID (see Figure 15). A reboot will be required.

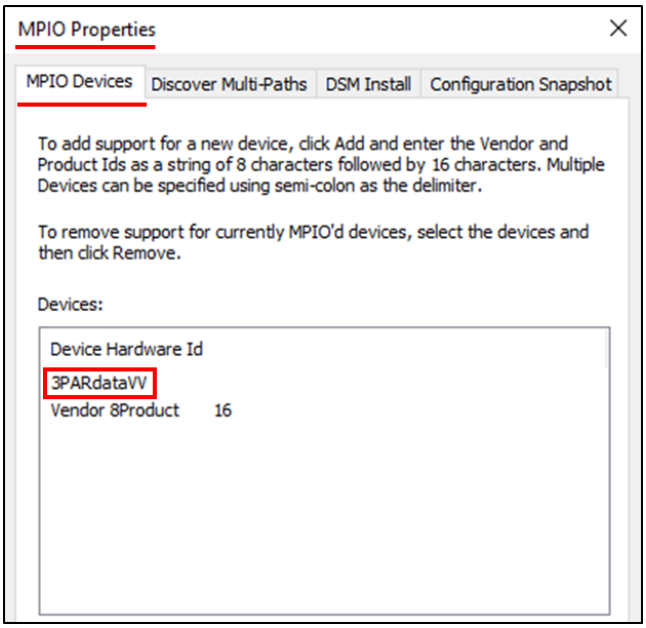

Figure 15. 3PARdataVV device string added to MPIO Devices

MPIO Path Verify Enabled

Path Verify Enabled must be set in the MPIO settings on Windows Server 2016/2012/2008 hosts configured for an HPE 3PAR Peer Persistence implementation.

After adding the **3PARdataVV** device ID and rebooting, set the HPE 3PAR Device Specific Module (DSM) **Path Verify Enabled** option for volumes used with Remote Copy and Peer Persistence. This should be configured on all servers using Remote Copy volumes.

See *HPE 3PAR Peer Persistence* on page 11 of the *HPE 3PAR Windows Server 2016/2012/2008 Implementation Guide* at http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c04448812

Capacity and sizing

SSD usage

CPGs composed of solid-state drives (SSDs) should follow these best practices:

- Use RAID 6 for increased resiliency.
- Use RAID 1 for increased performance if that is the goal. The tradeoff is lower available capacity.
- If using RAID 6, use a set size of 6+2 rather than the default of 4+2.

Using RAID 6 (6+2) provides the same percentage capacity utilization as RAID 5 with a "set size" of 3+1 while providing increased resiliency. RAID 5 is not recommended due to lower resiliency over RAID 6 and lower performance over RAID 1.

As discussed in the Design principles section earlier, there are various tradeoffs to consider in a DR solution. For synchronous replication DR solutions there is a slight delay for the I/O acknowledgement to come back from the remote storage system; in this case, RAID 1 might not be as beneficial over a RAID 6 CPG.

Storage provisioning

All volumes are thinly provisioned. All volumes except the backup volume are configured to use SSD storage in a RAID 6 common provisioning group (CPG). The backup volume is configured to use HDD storage in a RAID 6 CPG. See the *HPE 3PAR Best Practice Guide*[25] for additional information.

Table 4 lists the volumes provisioned for the solution.

Table 4. HPE 3PAR volumes used for SQL Server solution with Peer Persistence

Disk	HPE 3PAR Volume	RC Group	Disk Size (GiB)	Formatted Size (GB)	Comment
DB	win.clus63.db104	SQL103	2,048	2,047	For database
Transaction Log	win.clus63.tlog104	SQL103	2,096	2,095	For database transaction logs
Tempdb.0	win.clus63.tempdb104.0	SQL103	1,472	1,471	For tempdb
Tempdb.1	win.clus63.tempdb104.1	SQL103	1,472	1,471	For tempdb
Bkp[26]	win.clus63.bkp104	N/A	4,096	4095	On 3PAR 8440 only (HDD storage only)
DTC	win.clus63.msdtc105	SQL103	2	2	For Microsoft DTC
vm	win.clus63.vm	SQL103	2,048	2,047	For virtual machines

File system

For optimal performance with HPE 3PAR storage volumes, format the disks presented to the Windows disks in allocation unit size multiples of 16 KB[27]. Use the `diskpart filesystems` command to confirm the formatted volume allocation unit size.

The Windows Server 2016 NTFS file system with an allocation unit size of 64 KB is used for all storage volumes presented to the cluster.

Virtual disk VHD sets are used for all virtual machine disks including the DTC disk and all database disks other than the backup disk. During the VHD creation, a fixed size (not dynamic) is used for all volumes for optimal performance of the virtual disk. Thin provisioning of the volume at the HPE 3PAR storage side is used for storage optimization, which is more efficient than at the OS level.

The VHD sizes are as follows:

- db = 2000 GB
- tlog = 2000 GB

[25] *HPE 3PAR StoreServ Storage Best Practices Guide*, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA4-4524ENW
[26] The backup volume is not included in the RC Group since consistency is not needed there. The backup volume stands on its own.
[27] *HPE 3PAR StoreServ Storage Best Practice Guide*, page 14, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA4-4524ENW

- tempdb = 1400 GB
- DTC = 1 GB
- Virtual machines = 128 GB

Workload description

The workload used is an OLTP-type financial brokerage account workload utility driving a 1 TB fully populated database. Workload characteristics are not altered; only user and global intensity settings are set.

Normally, client applications will retry requests for any anomalous application responses. However, like many test tools, this workload generator does not handle retries well; so, the workload was restarted after failover tests as needed and documented in the test cases.

Performance observations

For optimal performance, link latency effects must be considered. Effects will vary depending on the distance between data center sites and the speed of the replication links.

Optimal performance is achieved when the SQL Server storage disks are in the same data center as the physical server nodes and virtual machines hosting the active SQL Server cluster role. The reason is that for synchronous solutions, I/O sent to storage from the database must be replicated to the remote storage system prior to the acknowledgement back to the server and considered complete. After the acknowledgement is received by the server, the I/O is considered complete from the database perspective.

Figure 16 shows the best-case scenario with the database I/O and storage in the same data center and the only link latency incurred is for storage replicating the data to the disaster recovery site. The sequence of host data write, replication, and acknowledgement is:

1. Host data is written to the local (primary) storage system.
2. Host data is replicated to and acknowledged by the disaster recovery storage system.
3. The local storage system acknowledges the host data write as complete.

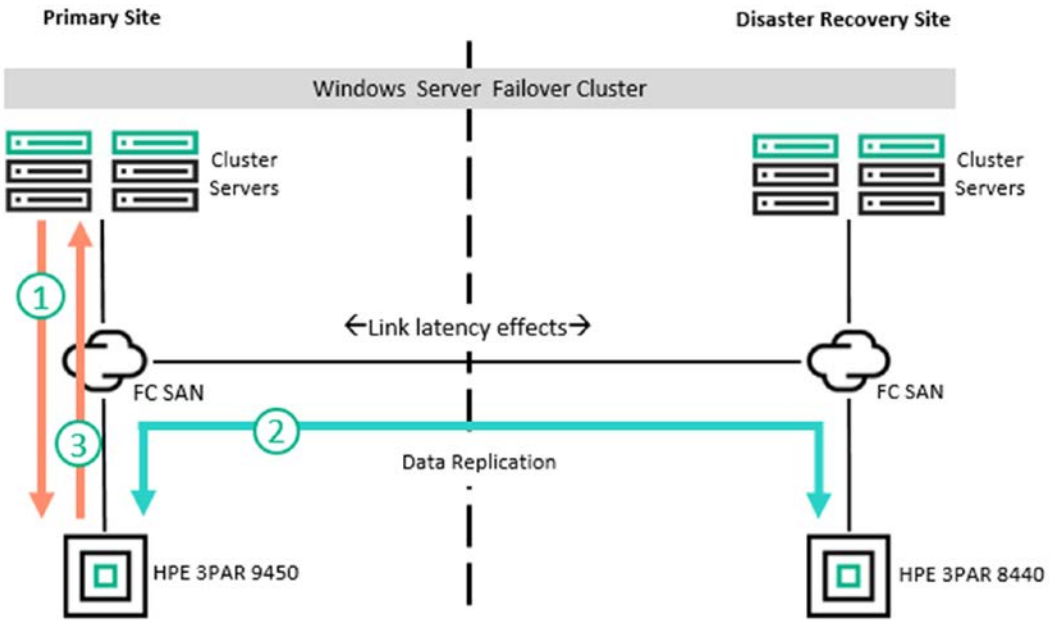

Figure 16. SQL Server database I/O to local storage system, replicated to remote storage system, prior to acknowledgment of database I/O

For long-term operations, avoid having both the server to storage I/O and the storage-replicated I/O traversing the remote copy link. Not only does this impact performance, but it uses up expensive network bandwidth.

Figure 17 illustrates the situation to avoid for long-term operations. The sequence of host data write, replication, and acknowledgement is:

1. Host data is written to the remote (disaster recovery) storage system.
2. Host data is replicated to and acknowledged by the local (primary) storage system.
3. The remote storage system acknowledges the host data write as complete.

In this case, if the link latency is 1 ms one way, then there will be at least a minimum of 4 ms in total link latency, along with the extra overhead of the I/O traversing the link multiple times.

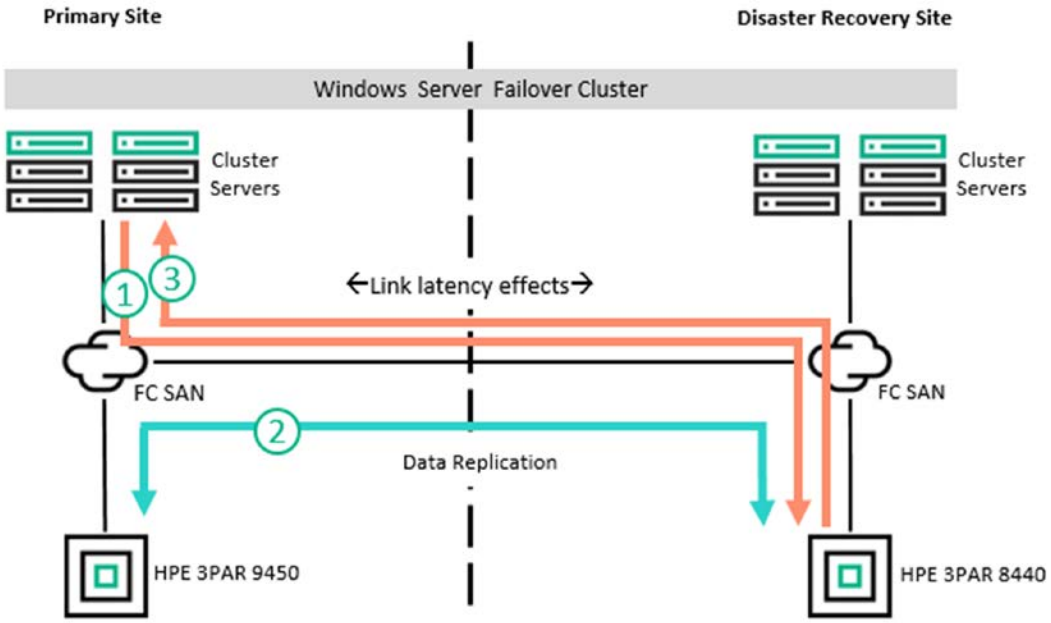

Figure 17. SQL Server database I/O to remote storage system, replicated to local storage system, prior to acknowledgment of database I/O

Use cases

No database or cluster timeouts were changed for these use cases. The following use cases were performed:

Use case #1: Manual switch of source and target copy direction

This use case confirms manual switchover functionality of the Peer Persistence solution, including storage usage from both sites.

Use case #2: Move the running SQL Server instance to another cluster node

This use case confirms SQL Server database storage access from other nodes.

This might also be performed during maintenance operations.

Use case #3: Live migration of running VM with SQL Server instance

This use case confirms Hyper-V functionality of the parent cluster and storage access from other nodes.

This might also be performed during maintenance operations.

Use case #4: Catastrophic cluster server node failure

This use case confirms the solution operates as expected during a catastrophic node failure.

This use case tests a catastrophic event.

Use case #5: Catastrophic site failure

This use case confirms the solution operates as expected during a catastrophic site failure.

The result was essentially the same as use case #4, but with the SQL Server instance restarting at the disaster recovery data center.

This use case tests a catastrophic event.

Use case #6: Catastrophic storage system failure and HPE Peer Persistence automated recovery

This use case confirms the solution operates as expected during a catastrophic storage system failure.

This use case tests a catastrophic event. The use case demonstrates the HPE Peer Persistence ability to recover and maintain application uptime.

Use case #1: Manual switch of source and target copy direction

This use case shows a manual switchover of the Remote Copy group four times over several minutes. Each time the switchover is performed, the array servicing the database I/O switches from the role of primary storage system to secondary, and the previous secondary storage system becomes the new primary storage system. The new primary storage system takes over host I/O communication.

At the time of switchover, the synchronous copy direction is also switched. There is a slight delay expected as the server and array change pathing information to the arrays. The switchover is performed without any interaction from the applications.

This is an interesting use case because it is similar to a later WSFC Live Migration, except in this case, it is the storage system performing a switchover while the virtual machine remains stationary. The manual switchover operation might be performed during scheduled maintenance operations while under full administrator control.

Figure 18 shows two manual switchover events while the database is active and storage is servicing I/O. SQL Server Management Studio (SSMS) has the Activity Monitor set to a refresh rate of 5 seconds and displays two switchover events.

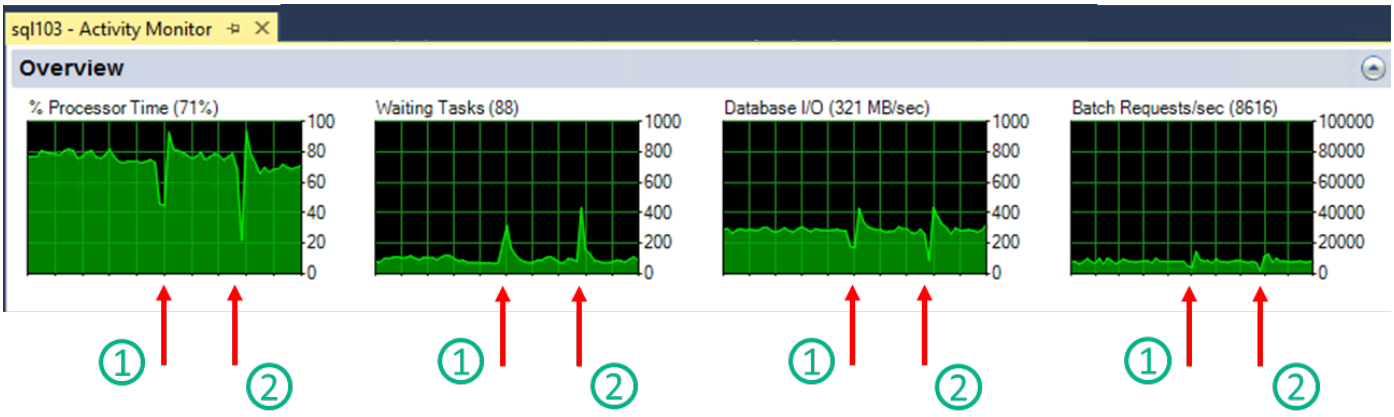

Figure 18. SSMS Activity Monitor output during two switchover events (see HPE 3PAR SSMC Figure 19). The first switchover = 1; the second switchover = 2.

Figure 19 shows the four switchover events from HPE SSMC System Report graphs. The display from two selected graphs are stacked to show how the two storage systems work together. HPE Peer Persistence software hosts connectivity and synchronous data replication for this disaster recovery solution as seen in these manual switchover operations.

Both the SSMS and HPE SSMC graphs are set to use a refresh interval of 5 seconds.

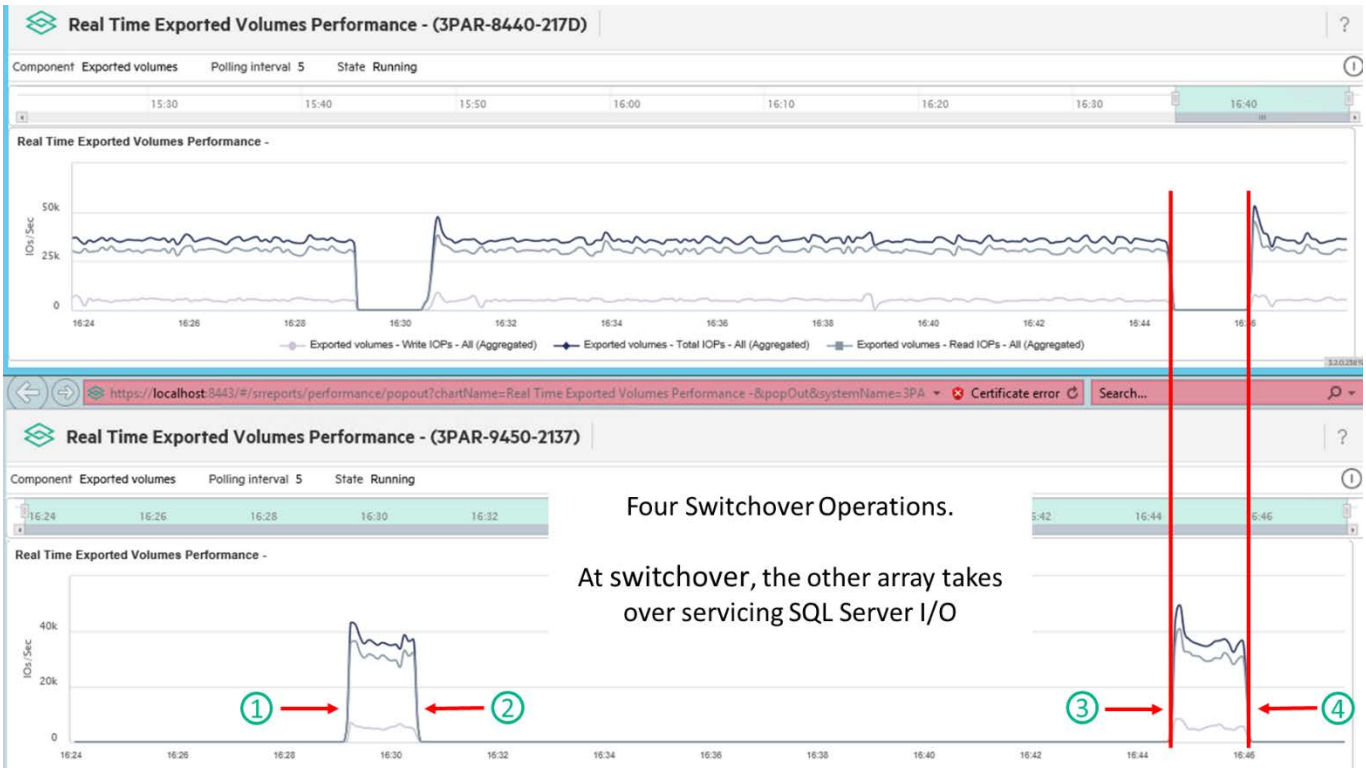

Figure 19. HPE SSMC System Reporter graph of both arrays showing four switchover operations while servicing SQL Server database I/O

Use case #2: Move the running SQL Server instance to another cluster node

Figure 20 shows a performance graph of the I/O workload, as seen from the HPE SSMC. The time period includes a move of the SQL Server instance to the other virtual machine node. Moving the SQL Server cluster role requires the application to stop and then restart on another node. Such an operation might occur during a database or cluster node maintenance operation.

The workload was momentarily stopped for this disruptive event, and then started shortly after the move completed. The operation completed as expected with HPE Peer Persistence providing continuous data availability for SQL Server.

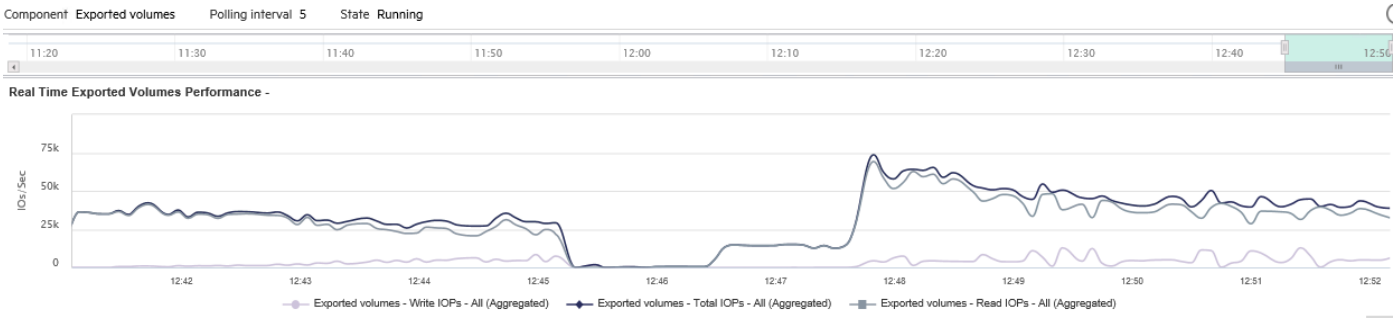

Figure 20. HPE SSMC Performance Report showing I/O from the database prior to and following a move of the SQL Server instance to another cluster node

Use case #3: Live migration of running VM with SQL Server instance

Figure 21 and Figure 22 capture a live migration of a virtual machine while SQL Server is processing client requests. The actual dip in the graphs only occurs during the final phase of the migration when the I/O is momentarily quiesced and virtual machine memory handling is transferred to another node. The HPE Peer Persistence storage solution easily handles all database I/O no matter which server node it comes from.

Both the SSMS and HPE SSMC graphs are set to use a refresh interval of 5 seconds.

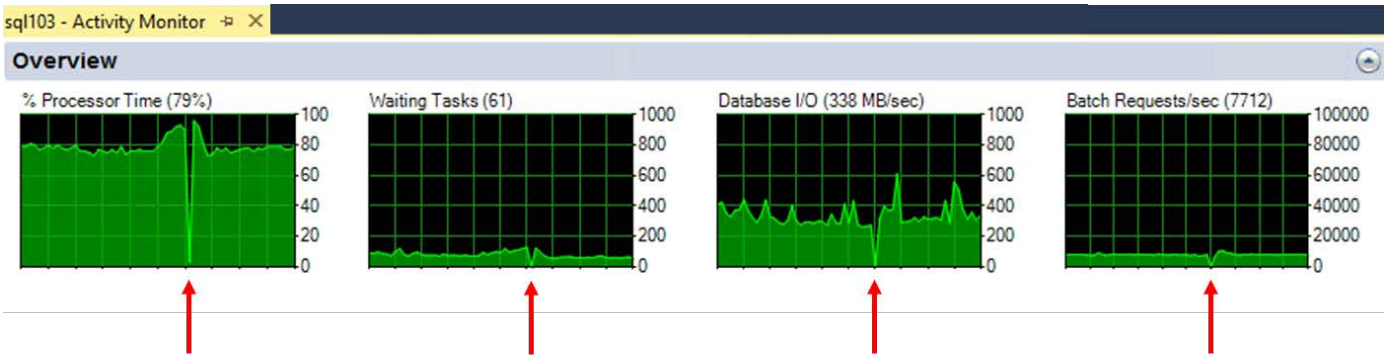

Figure 21. SSMS Activity Monitor output showing live migration event while SQL Server is processing client batch requests

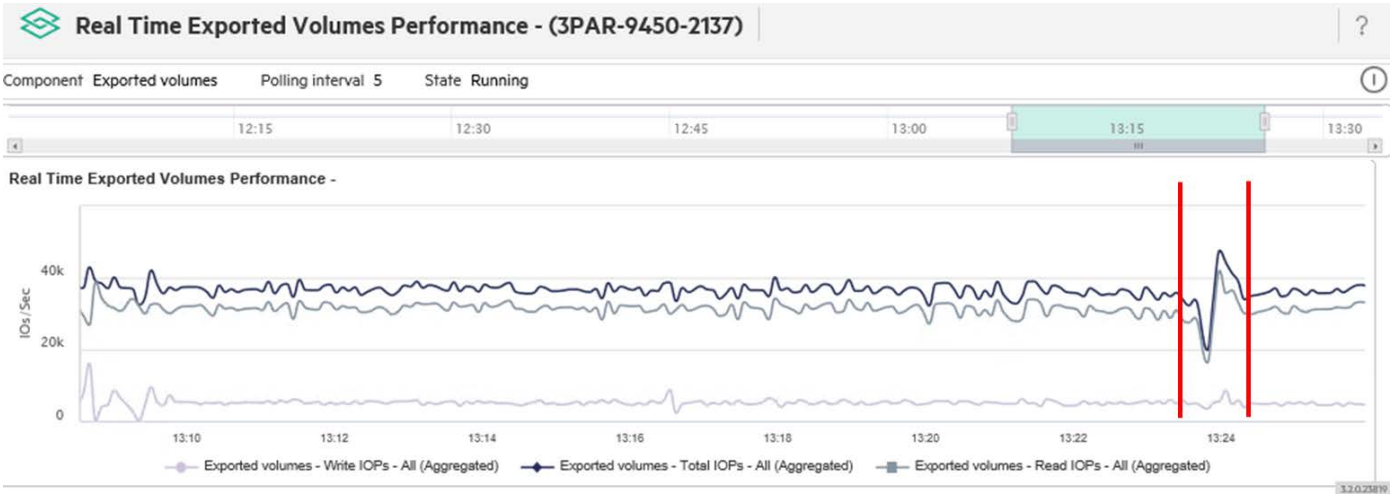

Figure 22. HPE SSMC System Reporter graph showing I/O effect of virtual machine live migration with SQL Server database I/O

Use case #4: Catastrophic cluster server node failure

This use case tested the failure of the parent node that owned the virtual machine running the active SQL Server instance. The catastrophic server node failure was initiated via the HPE Integrated Lights-Out (iLO) interface. Within the iLO Server Power settings area a **Force Power Cycle** event was performed to kill the server without any Windows Server OS notification (see Figure 23).

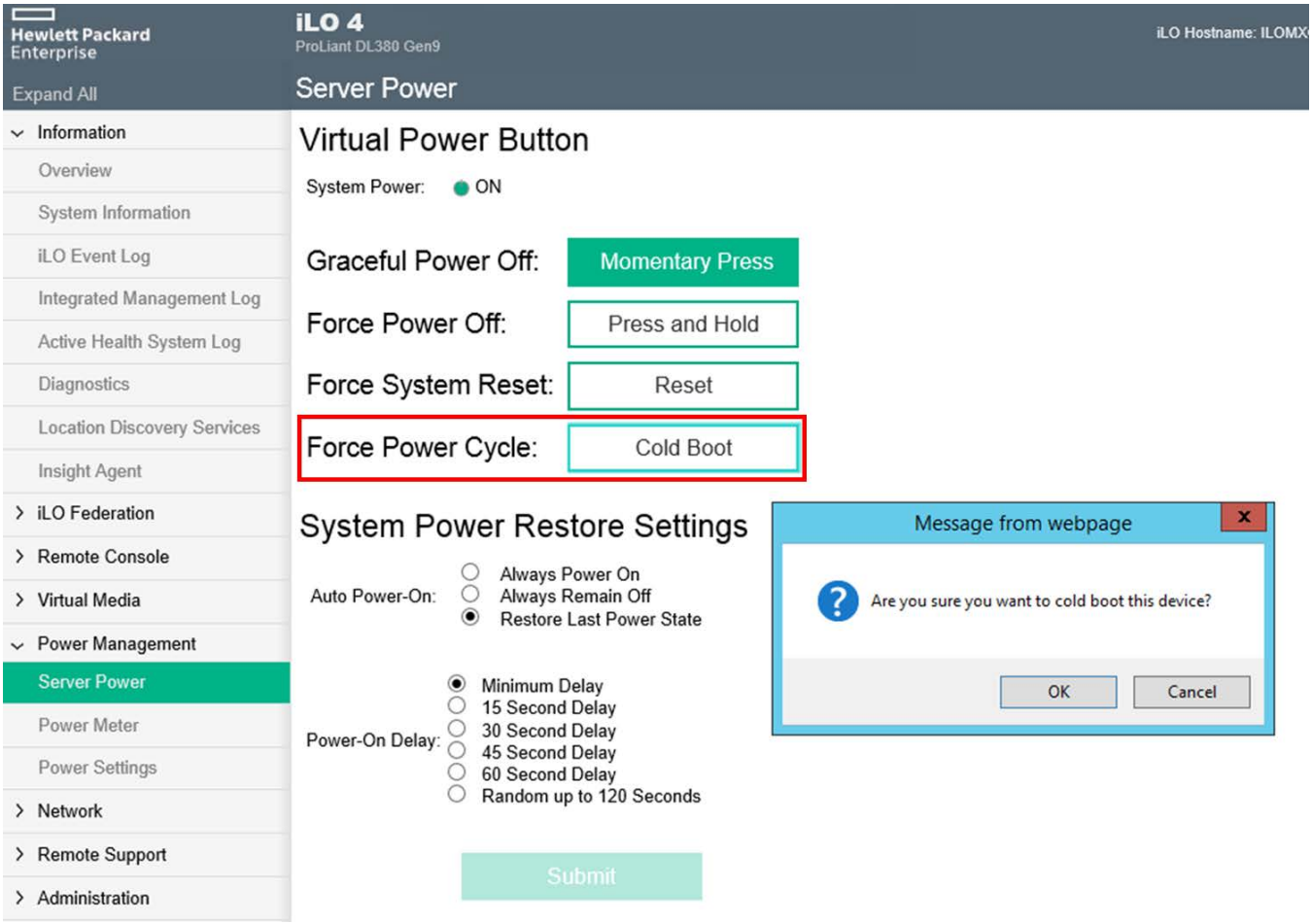

Figure 23. HPE ProLiant iLO interface showing the Force Power Off server OS crash event initiation

The VM running on the node failed as expected when the cluster node crashed. The VM eventually restarted on another available cluster node in the same data center.

The SQL Server instance running on the failed VM in the Hyper-V guest cluster also failed, as expected. An advantage of using SQL Server failover clusters is that the SQL Server instance automatically restarts on another guest cluster VM node without waiting for the original VM to start up first.

The client load was stopped when the server crashed and restarted after the database restarted. Figure 24 captures the I/O activity during the complete event. There was no impact to the HPE Peer Persistence protected storage.

Whether a real or simulated crash event occurs, all data written to the storage system and acknowledged to the server remains intact. This meets a durability requirement of the Atomicity, Consistency, Isolation, Durability (ACID) test. The SQL Server database remained intact and SQL Server performed a crash-consistent recovery after the database restarted.

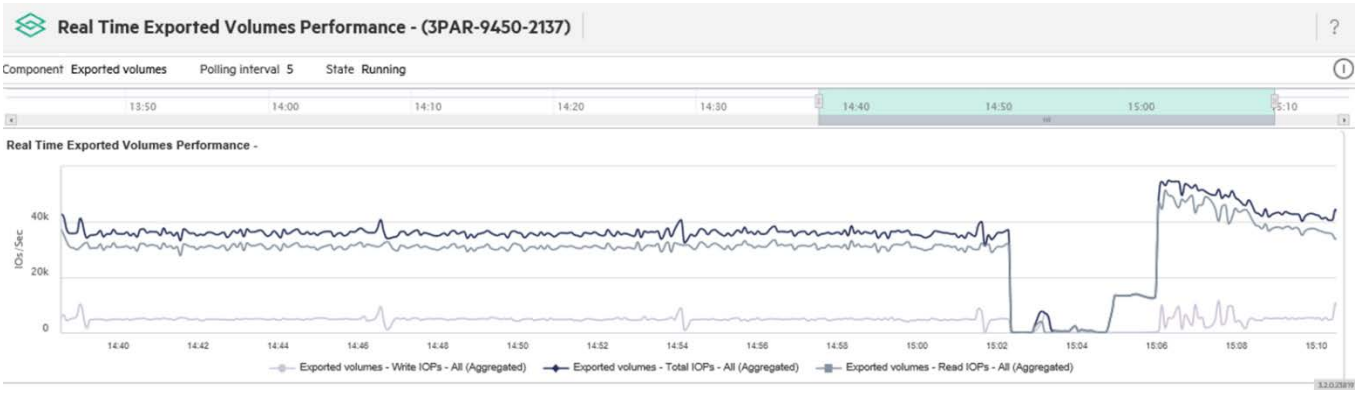

Figure 24. HPE SSMC System Reporter showing client database activity stopping and then being restarted after catastrophic event

Use case #5: Catastrophic site failure

This use case was performed by resetting the ProLiant servers (as in use case #4) and removing power from the HPE 3PAR 9450 storage system. The result is essentially the same as use case #4, but with the SQL Server instance restarting at the disaster recovery data center. The OLTP load generation was restarted after the database came online and the result was the same as if only the server had failed.

Whether a real or simulated crash event occurs, all data written to the storage system and acknowledged to the server remains intact. This meets a durability requirement of the ACID test. The SQL Server database remained intact and SQL Server performed a crash-consistent recovery after the database restarted.

HPE Peer Persistence functionality ensures there is no data lost, ensuring there is an RPO of 0. While this might not be important for some, many SQL Server database solutions running critical applications require this kind of zero-data-loss synchronous solution.

Use case #6: Catastrophic storage system failure and HPE Peer Persistence automated recovery

This use case was performed by removing power from the HPE 3PAR 9450 storage system. The result was similar to use case #1 with interesting and expected recovery timing differences.

As noted in Appendix B: HPE Peer Persistence failure recovery information, HPE Peer Persistence monitors the links and reports storage system failure at a predetermined time. After the specified time, the storage solution initiates a recovery and starts servicing SQL Server database I/O directly from the disaster recovery storage system. Figure 25 shows the interesting and expected behavior of database I/O pausing during the failure detection period. The default HPE Peer Persistence failure detection and the start of recovery time is 10 seconds. The Activity Monitor refresh time of 5 seconds shows the failure detection and recovery time.

Both the SSMS and HPE SSMC graphs are set to use a refresh interval of 5 seconds.

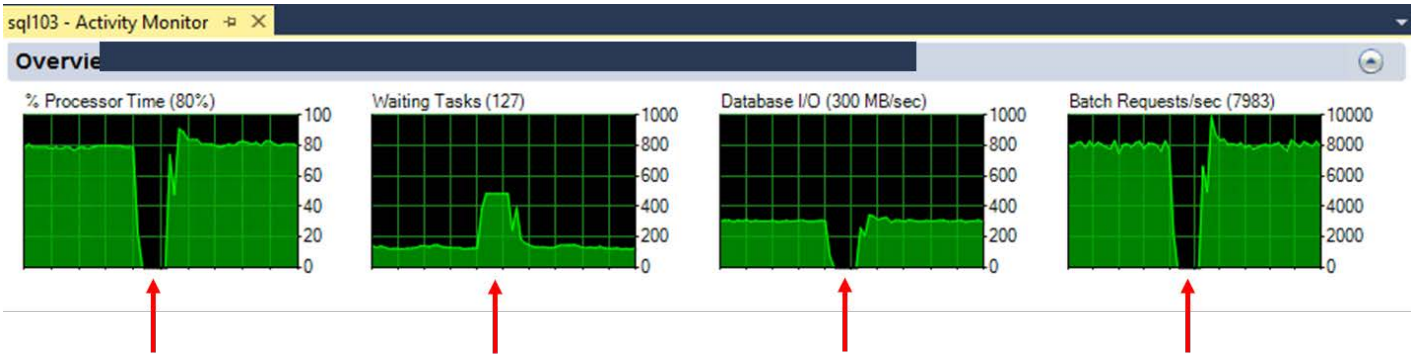

Figure 25. Failure of primary storage system and automated HPE Peer Persistence recovery

The behavior seen is the result of Windows Server Multipath and the HPE Peer Persistence solution detecting the failure and switching over to the disaster recovery site well before other cluster timeouts took effect.

Whether a real or simulated crash event occurs, all data written to the storage system and acknowledged to the server remains intact. This meets a durability requirement of the ACID test. The SQL Server database remained intact and SQL Server performed a crash-consistent recovery after the database restarted.

HPE Peer Persistence functionality ensures there is no data lost, delivering an RPO of 0.

The solution failed over in this disaster recovery use case without any user intervention.

Key findings

Key findings include:

- HPE Peer Persistence provides a SQL Server storage-centric synchronous replication solution that plays well with WSFC.
- HPE Peer Persistence manual switchover functionality provides seamless storage live-migration-type storage failover and failback.
- HPE Peer Persistence synchronous replication provides a zero-data-loss (RPO=0) replication solution.
- HPE Peer Persistence removes replication overhead from servers and works across multiple solution environments.
- HPE Peer Persistence automated failure handling provides a durable storage replication solution to help maintain SQL Server database integrity.
- Switching storage source and target roles was almost indiscernible from the application perspective with no downtime.
- Virtual machine live migrations were helped by the HPE Peer Persistence ability to switchover with them to another site.
- Virtual machine guest clustering provides fast failover capability for the SQL Server resource.
- HPE Peer Persistence monitors and responds quickly to catastrophic storage system failures to maintain 100% uptime.
- HPE Peer Persistence monitors and responds quickly to catastrophic site failures to minimize downtime and maintain 100% data availability.

Summary

All-flash HPE 3PAR StoreServ is cost-effective storage that provides high performance and meets resiliency demands of mission critical Microsoft SQL Server 2017 databases.

HPE 3PAR Remote Copy software provides the foundation for HPE 3PAR synchronous and asynchronous solutions by delivering the connectivity path for replication between storage systems. HPE Remote Copy software is a simple-to-use, efficient, and flexible replication technology that protects and shares data from any application[28].

With HPE 3PAR Peer Persistence software, HPE 3PAR storage systems located at metropolitan distances can act as peers to each other, presenting a highly available transparent failover storage system to hosts and servers connected to them. This capability enables customers to configure a high-availability, disaster recovery solution between two sites or data centers where switchover and switchback remains transparent to the hosts and applications running on those hosts. And in the event of a catastrophic failure, HPE Peer Persistence automated failure recovery secures business continuity without any data loss. Additionally, with active databases at both sites, the HPE 3PAR solution provides load balancing across systems and sites. Unlike traditional failover models, HPE Peer Persistence software utilizes automatic transparent failovers that allow hosts to remain online serving their business applications, even when they switch from their original site to the disaster recovery site, resulting in a much improved recovery time. Your data center can achieve high availability in a multisite federated environment with HPE Peer Persistence.

With HPE you can break the boundaries of storage, enabling your hosts and data to move freely across data centers without impacting your business applications. You can enable your HPE 3PAR storage systems located at metropolitan distances to act as peers to each other, presenting a nearly continuous storage system to hosts and servers connected to them.

HPE Remote Copy and HPE Peer Persistence software features secure business continuity and provide an extremely available, always-on storage solution that can withstand failure of an entire array without an impact to Microsoft SQL applications.

Implementing a proof-of-concept

As a matter of best practice for all deployments, HPE recommends implementing a proof-of-concept using a test environment that closely matches the planned production environment. In this way, appropriate performance and scalability characterizations can be obtained. For help with a proof-of-concept, contact an HPE Services representative (hpe.com/us/en/services/consulting.html) or your HPE partner.

Appendix A: Bill of materials

The following QuickSpecs will help with ordering information:

HPE 3PAR StoreServ 9000 storage QuickSpecs, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=a00005876enw

HPE 3PAR StoreServ 8000 storage QuickSpecs, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=c04607918

HPE ProLiant DL300 servers QuickSpecs, https://www.hpe.com/h20195/v2/Library.aspx?doctype=41&doccompany=HPE&footer=41&filter_doctype=no&filter_doclang=no&country=&filter_country=no&cc=us&lc=en&status=A&filter_status=rw#doctype-41&doccompany-hpe&subcate_oid-15351_241475&status-a&sortorder-csdisplayorder&teasers-off&isRetired-false&isRHParentNode-false&titleCheck-false

HPE 3PAR Software Products QuickSpecs, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=c04199812

The following BOMs contain electronic license to use (E-LTU) parts. Electronic software license delivery is now available in most countries. HPE recommends purchasing electronic products over physical products (when available) for faster delivery and for the convenience of not tracking and managing confidential paper licenses. For more information, please contact your reseller or an HPE representative.

[28] *Disaster Tolerant Solutions with HPE 3PAR Remote Copy*, http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA3-8318ENW

Note

Part numbers are at time of testing and subject to change. The bill of materials does not include complete support options or other rack and power requirements. If you have questions regarding ordering, please consult with your HPE Reseller or HPE Sales Representative for more details (hpe.com/us/en/services/consulting.html).

Table 5. Bill of materials

Qty	Part number	Description
		HPE ProLiant DL380 Gen9 server
4	719064-B21	HPE ProLiant DL380 Gen9 8SFF Configure-to-order Server
4	781913-L21	HPE DL380 Gen9 Intel Xeon E5-2698v3 (2.3GHz/16-core/40MB/135W) FIO Processor Kit (initial processor)
4	781913-B21	HPE DL380 Gen9 Intel Xeon E5-2698v3 (2.3GHz/16-core/40MB/135W) FIO Processor Kit (second processor)
32	726722-B21	HPE 32GB 4Rx4 DDR4-2133P-L (256 GB RAM per server)
4	749974-B21	HPE Smart Array P440ar/2GB FBWC 12Gb 2-ports Int FIO SAS Controller
8	759210-B21	HPE 450GB 12G SAS 15K rpm SFF (2.5-inch) SC Enterprise 3yr Warranty Hard Drive (2 per server)
4	Q0L14A	HPE SN1200E 16Gb 2P FC HBA
4	727060-B21	HPE FlexFabric 10Gb 2-port 556FLR-SFP+ Adapter (one per server)
		HPE 3PAR StoreServ 9450 storage
1	Q0E92A	HPE 3PAR 9450 2-node Storage Base with All-inclusive Single-system Software
1	Q7F41A	HPE 3PAR StoreServ 9000 Controller Nodes (additional node pair)
4	Q0E97A	HPE 3PAR 9000 4-port 16Gb Fibre Channel Host Bus Adapter (one per node)
8	Q0E95A	HPE 3PAR 9000 24-disk 2U SFF (2.5in) SAS Drive Enclosure
16	Q0F40A	HPE 3PAR 9000 400GB SAS SFF (2.5in) SSD with All-inclusive Single-system Software
		HPE 3PAR StoreServ 8440 storage
1	H6Y97B	HPE 3PAR 8440 2-node Storage Base with All-inclusive Single-system Software (factory integrated in HPE rack)
2	H6Z00A	HPE 3PAR StoreServ 8000 4-port 16Gb Fibre Channel Adapter
12	K2P89B	HPE 3PAR 8000 1.92TB+SW SFF SSD
8	K2P93B	HPE 3PAR 8000 1.2TB+SW 10K SFF HDD
		Miscellaneous hardware
4	C8R08A	HPE SN6000B 16Gb 48-port Bundled Fibre Channel Switch[29]
4	T5522AAE (Physical License - T5522A)	HPE B-series Mid Range Switch Power Pack+ E-LTU (for HPE SN6000 switches)
1	JG846A	HPE 5900AF 48XG 4QSFP FB Bundle[30]

[29] HPE StoreFabric SN6000B 16Gb and QuickSpecs
[30] HPE FlexFabric 5900 Switch Series QuickSpecs

Qty	Part number	Description
		HPE software
2	N/A	HPE StoreServ OS 3.3.1 EGA release (included with All-inclusive Single-system Software)
2	N/A	HPE StoreServ Management Console 3.2.0 (included with All-inclusive Single-system Software)
1	L7F04A	HPE 3PAR 9450 All-inclusive Multi-System Software LTU[31]
1	L7E71A	HPE 3PAR 8440 All-inclusive Multi-system Software LTU[32]
2	N/A	HPE 3PAR Remote Copy capabilities (included with All-inclusive Multi-System Software)
2	N/A	HPE 3PAR Peer Persistence Software (included with All-inclusive Multi-System Software)
1	N/A	HPE Peer Persistence Quorum Witness (QW) v2.1.000 Hyper-V virtual machine (included with All-inclusive Single-system Software); both ESX and Hyper-V versions are available
4		SPP (Service Pack for ProLiant) 2017.04.0[33]. All HPE ProLiant servers.
		Microsoft software
8		Microsoft Windows Server 2016 Datacenter Build 14393. Cluster nodes, guest cluster nodes, supporting servers.
4	Add Role	Microsoft Windows Hyper-V (add role to physical cluster server nodes)
4	Add Feature	Microsoft Windows Multipath IO (add feature to physical cluster server nodes)
4	Add Feature	Microsoft Windows Failover Clustering (add feature to physical and guest cluster server nodes)
2		Microsoft SQL Server 2017 (installed on guest cluster nodes)
1		Microsoft SQL Server Management Studio

[31] HPE 3PAR Software Products QuickSpecs
[32] HPE 3PAR Software Products QuickSpecs
[33] HPE SPP used on all ProLiant servers to ensure application of current BIOS, FW, and Drivers

Appendix B: HPE Peer Persistence failure recovery information

Important HPE Peer Persistence information is included here as found in the *HPE 3PAR Remote Copy Software User Guide version OS 3.3.1 MU1* (RC SW UG) at http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=c03618143.

Link and communications failure scenarios with figures are discussed in the RC SW UG starting at page 167.

Timeout values were not altered from the defaults in this paper, but this information is discussed in RC SW UG starting at page 175.

Failure timeout behavior is discussed in the RC SW UG starting at pages 195 and 204. At 5 seconds, the system detects the links have failed. At 10 seconds, the system declares the RC links are down and stops the RC groups. The remaining storage system will take over server I/O after that time.

Nondisruptive failover (switchover) operations are discussed in the RC SW UG on page 196.

Storage system recovery after an automated failover

Storage system recovery is discussed in the RC SW UG starting at page 200.

Note

HPE Peer Persistence is a synchronous solution—not an asynchronous solution, which is discussed in the same area in the user guide.

In the disaster recovery situation, the storage systems are not in a normal state and replication will have stopped, so the `setrcopygroup failover` operation for systems in a normal state (discussed in *Recovering from disaster for 1-to-1 configurations*) is not used (RC SW UG starting at page 205).

There are two basic recovery methods (Method 2 is preferred):

- Method 1 as described in the RC SW UG from page 205 to 213 using key commands:
  - `Setrcopygroup recover`
  - `Setrcopygroup restore`
- Method 2 is preferred and the best practice, and described in the RC SW UG page 294:
  - `Setrcopygroup recover`
  - `Setrcopygroup reverse`
  - `Setrcopygroup switchover`

Method 2 is preferred and is the best practice. One reason method 2 is preferred is that in large HPE Peer Persistence solutions approaching the maximum limits, the `setrcopygroup restore` command might timeout at the limits. If the restore operation times out during recovery, then additional recovery steps might need to be performed.

Example from this solution (3PAR-9450-2137 is the failed original source storage system):

1. `setrcopygroup recover -t 3PAR-9450-2137`
2. `setrcopygroup reverse -natural -t 3PAR-9450-2137`
3. `setrcopygroup switchover -t 3PAR-9450-2137`

Use the `showrcopy` command during the recovery process to confirm the correct array state prior to issuing any of the recovery commands. The switchover is only performed after synchronization is 100% complete for all volumes in the group.

Glossary

Table 6 contains common terms used in this document.

Table 6. Glossary of terms

Term	Description
Failover cluster	A grouping of hardware and software components that provide high availability through redundancy, along with scalability for workloads. See also WSFC below.
Guest clustering	A failover cluster whose server nodes are composed of virtual machines and possibly includes cluster-aware software running on those virtual machines, such as SQL Server 2017.
Peer Persistence	HPE 3PAR Peer Persistence software enables failover or failback between two sites or data centers to be transparent to hosts, without any disruption to applications running on them. HPE Peer Persistence employs a quorum witness at a third site.
Remote Copy	HPE 3PAR Remote Copy is a unique replication technology that allows you to protect and share data from any application more simply, efficiently, and affordably. Remote Copy dramatically reduces the cost of remote data replication and DR.
Recovery Point Objective (RPO)	An amount of time prior to the disaster at which all committed transactions can be recovered.
Recovery Time Objective (RTO)	The maximum amount of downtime that elapses following a disaster before the database should be up and running and processing new transactions.
SCSI LUN	SCSI logical unit number (LUN), which is used with FC devices along with their Worldwide Name (WWN) for identification.
Virtual volume (VV), and VLUN	Either thinly or thickly provisioned storage that is made available for server usage by exporting it, which creates a volume-LUN (VLUN) pairing from the HPE 3PAR perspective and a disk LUN from the server perspective.
TPVV	A thinly provisioned VV (TPVV) is a volume that allocates system storage as needed, while a fully provisioned VV is one with fully allocated or reserved space on the storage system.
Fibre Channel (FC) fabric; FC SAN	One or more FC switches connected together creates an FC fabric. Two switches not connected would be two FC fabrics. An FC SAN is composed of one or more FC fabrics.
SQL Server failover cluster	SQL Server installed as a role within a WSFC
WSFC	Windows Server Failover Clustering. See also Failover clustering above.

Resources and additional links

HPE Reference Architectures
HPE Reference Architecture for Microsoft SQL Server 2014 on HPE 3PAR Storage with Disaster Recovery
http://h20195.www2.hpe.com/V2/GetDocument.aspx?docname=4AA6-3954ENW

HPE Information Library
hpe.com/info/ra

HPE Servers
hpe.com/servers

HPE Storage
hpe.com/storage

HPE Information Library, HPE 3PAR storage documentation
http://h17007.www1.hpe.com/us/en/storage/info-library/index.aspx

HPE Networking
hpe.com/networking

HPE Technology Consulting Services
hpe.com/us/en/services/consulting.html

To help us improve our documents, please provide feedback at hpe.com/contact/feedback.

Learn more at HPE Data Storage

hpe.com/storage

Sign up for updates

a00030078enw, April 2018, Rev. 2