DELL EMC VMAX ALL FLASH STORAGE FOR MICROSOFT HYPER-V DEPLOYMENT

July 2017

Abstract

This white paper examines deployment of the Microsoft Windows Server Hyper-V virtualization solution on Dell EMC VMAX All Flash arrays, with focus on storage efficiency, availability, scalability, and best practices.

H16434R

This document is not intended for audiences in China, Hong Kong, Taiwan, and Macao.

WHITE PAPER

Copyright

Contents

Chapter 1 Executive Summary

This chapter presents the following topics:

Summary

Microsoft Hyper-V provides customers with an ideal platform for key virtualization scenarios, such as production server consolidation, business continuity management, software test and development, and development of a dynamic data center. Scalability and high performance can be achieved by supporting features such as guest multiprocessing support and 64-bit guest and host support. Hyper-V also allows customers to achieve significant space, power, and cooling savings while maintaining availability and performance targets. Dell EMC™ VMAX™ All Flash storage systems can bring additional value to customers by providing the ability to consolidate storage resources, implement advanced high-availability solutions, and provide seamless multisite protection of customer data assets in Windows Server environments.

Many of the advanced features of the Hyper-V environment are either facilitated by, or enhanced with, the implementation of high performance, reliable, and scalable storage by using a VMAX All Flash array.

VMAX All Flash arrays (VMAX 250, VMAX 450, VMAX 850, and newly introduced VMAX 950 arrays) are available with F and FX software package options that offer data protection and availability features that are beneficial for most deployments. The new VMAX architecture uses the latest, most cost-efficient 3D NAND Flash drive technology, using multi-dimensional scale, large write-cache buffering, back-end write aggregation, high IOPS, high bandwidth, and low latency.

Hyper-V deployment on a VMAX All Flash array offers a unique set of benefits, such as:

- Rapid deployment and protection of Hyper-V environments
- Block and file services, multiple connectivity and accessibility options, and multiprotocol data access including integrations with Server Message Block (SMB) 3.0 when using VMAX eNAS file services
- Storage design and deployment best practices for high availability (HA) and performance
- Integration between Microsoft and Dell EMC storage technologies for ease of use of management, backup, and recovery and application awareness

This white paper explains how the features of the VMAX All Flash array provide a high performance and scalable environment to run a Hyper-V based virtualized computing environment with HA for Hyper-V hosts and virtual machines (VMs). This white paper also explains how to use complementary Dell EMC technologies such as Dell EMC AppSync™ software, the VMAX Volume Shadow Copy Services (VSS) provider, and Dell EMC Solutions Enabler for Hyper-V environments to improve availability and efficiency in Hyper-V landscapes.

Audience

This white paper is intended for system administrators, storage administrators, and system architects who are responsible for implementing, managing, and maintaining Hyper-V virtualization solutions on Windows Server 2012 R2 and Windows Server 2016 using VMAX All Flash storage systems. Readers must have some familiarity with SQL

Server and the VMAX family of storage arrays, and want to achieve higher database availability, increased performance, and easier storage management.

We value your feedback

Dell EMC and the authors of this document welcome your feedback on the solution and the solution documentation. Contact EMC.Solution.Feedback@emc.com with your comments.

Authors: Udgith Mankad, Sanjeev Sharma, Penelope Howe-Mailly

Chapter 2 Dell EMC VMAX All Flash Storage Array

This chapter presents the following topics:

Overview

The Dell EMC VMAX family of storage arrays is built on the strategy of simple, intelligent, modular storage. It incorporates a Dynamic Virtual Matrix interface that connects and shares resources across all VMAX engines, enabling the storage array to grow seamlessly from an entry-level configuration into a storage system with massive capacity. It provides the highest levels of performance, scalability, and availability, and also features advanced hardware and software capabilities.

In 2016, Dell EMC announced the new VMAX All Flash VMAX 250F, VMAX 450F, and VMAX 850F arrays, and in May 2017, introduced the VMAX 950F array. Figure 1 shows the VMAX 950F and VMAX 250F arrays.

Figure 1. VMAX 950F and VMAX 250F arrays

VMAX All Flash benefits

VMAX All Flash storage, as shown in Figure 2, provides a combination of ease of use, scalability, high performance, and a robust set of data services that makes it an ideal choice for database deployments.

- **Ease of use**—VMAX All Flash storage provides simplicity for planning, ordering, and management. It uses virtual provisioning to create storage devices in seconds. All VMAX arrays are thin, consuming only the storage capacity that is actually written to, which increases storage efficiency without compromising performance. VMAX arrays are grouped into storage groups and managed as a unit for operations such as device masking to hosts, performance monitoring, local and remote replications, compression, host I/O limits, and so on. In addition, you can manage VMAX arrays by using Dell EMC Unisphere™ for VMAX, the Solutions Enabler command line interface (CLI), or REST APIs.
- **Scale-up and scale-out flexibility**—The VMAX All Flash arrays use V-Bricks to scale out.

  - The VMAX 250 starter V-Brick consists of one VMAX engine and 11 TBu of capacity. The V-Brick scales up to two V-Bricks with Flash Capacity Packs in 11 TBu increments.
  - VMAX 450, VMAX 850, and VMAX 950 starter V-Bricks consist of a VMAX engine and 53 TBu of capacity. These V-Bricks scale up with Flash Capacity Packs in increments of 13 TBu.

  You can order the following VMAX All Flash storage systems:

  - **F package**—An entry package with pre-packaged software bundles
  - **FX package**—A more encompassing package

  The packages also include embedded Unisphere for VMAX management and monitoring.

- **High performance**—VMAX All Flash storage is designed for high performance and low latency. It scales from one engine up to eight engines (V-Bricks). Each engine consists of dual directors. Each director includes two-socket Intel CPUs, front-end and back-end connectivity, a hardware compression module, InfiniBand internal fabric, and a large mirrored and persistent cache. All writes are acknowledged to the host as soon as they are registered with VMAX cache. Writes are subsequently, after multiple updates, written to flash. Reads also benefit from the large VMAX cache. When a read is requested for data that is not already in cache, FlashBoost technology delivers the I/O directly from the back-end (flash) to the front-end (host). Reads are only later staged in the cache for possible future access. VMAX All Flash storage also excels in servicing high bandwidth sequential workloads that leverage pre-fetch algorithms, optimized writes, and fast front-end and back-end interfaces.
- **Copy Data Management**, **Disaster Recovery (DR), and HA**—VMAX All Flash storage offers a strong set of data services. It natively protects all data with T10-DIF from the moment data enters the array until it leaves (including replications). With Dell EMC SnapVX™ and Dell EMC SRDF™, VMAX All Flash storage provides many topologies for consistent local and remote replications. VMAX All Flash storage provides optional D@RE, integrations with Dell EMC Data Domain™ software, such as Dell EMC ProtectPoint™ software, or cloud gateways with Dell EMC CloudArray™ software. Other VMAX data services include Quality of Service (QoS), compression, the “Call-Home” support feature, non-disruptive upgrades (NDU), non-disruptive migrations (NDM), and so on. In virtual environments, VMAX All Flash storage also supports vStorage APIs for Array Integration (VAAI) primitives such as `write-same` and `xcopy`.

  VMAX All Flash storage also provides automatic, scheduled, and application-consistent snapshots for Microsoft SQL Server and other applications for creating point-in-time copies for backup, reporting, or test/dev natively using SnapVX software. AppSync software provides policy-driven, automated, and self-service snapshots for applications with tighter integration between SnapVX and Microsoft VSS and SQL Server Virtual Device Interface (VDI).

With the introduction of the HyperMax OS Q3 2016 microcode release, VMAX All Flash systems are now capable of performing data compression to increase the effective capacity of the array significantly. With this microcode release, HyperMax OS now uses the Adaptive Compression Engine (ACE) to compress data and efficiently optimize system resources to balance overall system performance.

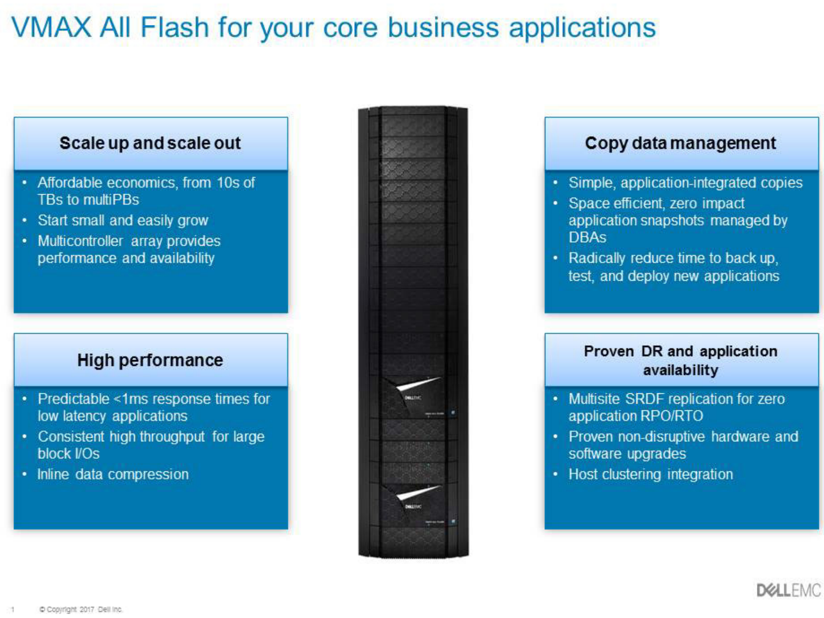

Figure 2. VMAX All Flash benefits for Microsoft environments

Chapter 3 Windows Server and VMAX All Flash design considerations

This chapter presents the following topics:

Overview

VMAX storage provisioning has become simpler than in previous releases. The following sections discuss the principles and considerations for storage connectivity and provisioning for Microsoft Server. When using Windows Server with VMAX All Flash storage, consider the overall system design and configuration to gain the most benefits and avoid artificial bottlenecks.

Storage connectivity options for Hyper-V Server

Flexible VMAX architecture provides three connectivity options to provide storage for Windows 2016 Server and Hyper-V VMs. These options include block storage access by Fibre Channel and iSCSI, as well as file-based access using SMB 3.0 and NFS protocols over Ethernet interfaces. For block access, host HBA ports (initiators) and VMAX storage ports (targets) are connected to an FC or Ethernet switch that is based on the connectivity requirements. FC connectivity requires that you create zones on the switch and define which initiator has access to which target. The zones create an I/O path between the host and storage. iSCSI connectivity requires that you set up IP network connectivity between the host and VMAX array. File-based connectivity also requires that you set up the IP network between the host and VMAX-embedded file server eNAS. Figure 3 shows all connectivity options available between a VMAX array and a Windows Server.

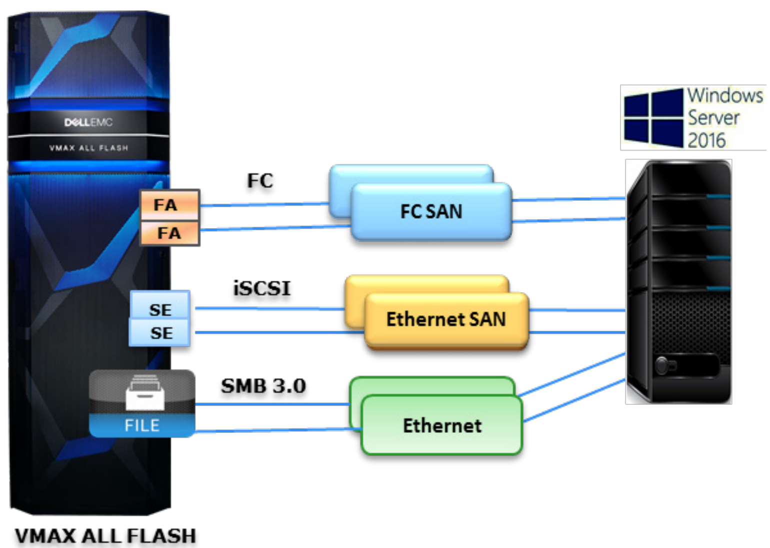

Figure 3. VMAX to Windows Server connectivity options

FC connectivity

Use at least two HBAs for each server to enable better availability and scale. Use multipathing software such as Dell EMC PowerPath® or Microsoft Windows Multipath I/O (MPIO) to load balance and fail over automatically or recover paths. To check that all paths are visible and active, use either of the following PowerPath commands:

```
powermt display paths
multipath -l
```

When zoning host initiators to storage target ports, ensure that each pair is on the same switch. Performance bottlenecks are often created when I/Os travel through ISL (paths between switches) because they are shared and limited.

Consider the ports' speed and count when planning bandwidth requirements. Each 8 Gb FC port can deliver up to approximately 800 MB/s. Therefore, a server with four ports cannot deliver more than approximately 3 GB/s. Also, consider that between the host initiator, storage port, and switch, the lowest speed that is supported by either of these components is negotiated and used for the I/O paths that are serviced by those components.

iSCSI connectivity

Windows Server and Hyper-V Server (2012 and later) include native support for a software iSCSI initiator as well as Multipath IO for resiliency and load balancing of storage I/O over multiple network paths. Storage connectivity using iSCSI is more cost-effective and flexible. VMAX All Flash storage provides iSCSI connectivity over multiple 10 GbE interfaces. Support for VLANs offers network partitioning and traffic isolation in multitenant environments and Challenge-Handshake Authentication Protocol (CHAP) addresses iSCSI security concerns by enabling access to clients that supply valid authentication.

Consider the following guidelines:

- For clustered environments, disable the cluster network communication for any network interfaces that you plan to use for iSCSI.
- Use VLANs dedicated to iSCSI setup. VLANs allow logical grouping of network end points, minimizing network bandwidth, contention for iSCSI traffic and eliminating impact on iSCSI traffic due to noisy neighbors.
- If all network devices in the iSCSI communication paths support jumbo frames, use jumbo frames on Ethernet to improve iSCSI performance.
- To minimize host CPU impact due to network traffic, ensure that Transmission Control Protocol (TCP) offloading is enabled on a host network interface card (NIC), which offloads processing of the TCP stack to the NIC and eases impact on the CPU.
- As with FC connectivity, use of PowerPath software or native multipathing for Windows helps load balance and ease queueing issues for iSCSI traffic through the host NICs.

SMB 3.0 connectivity using eNAS

Windows 2016 server can use SMB 3.0 file shares from the VMAX eNAS-embedded file server to store VMs or their copies. With this capability, Hyper-V can store VM files, which include configuration, virtual hard disk (VHD) files, and snapshots, on SMB file shares. Using the file share for Hyper-V provides increased flexibility because the existing converged network can be used for storage connectivity and you can dynamically migrate VMs or databases in the data center.

The advanced features of VMAX3 eNAS for Microsoft environments include offloaded data transfer (ODX), MPIO, and jumbo frame support, which allow users to make optimal use of resources for best performance. VMAX3 eNAS supports data protection for files using easy-to-schedule periodic snapshots as well as local and remote file system replication. eNAS also supports the Continuous Availability (CA) feature that allows Windows-based clients to access SMB shares persistently without the loss of the session state in case of a failover.

Chapter 4 Storage choice for VMs

This chapter presents the following topics:

Overview

There are two basic methods by which VMs can be provided access to VMAX storage. Connectivity can be either through the Hyper-V Server or directly to the VM. VHDs are created on the Hyper-V server and then made available to the VMs. The virtual Fibre Channel or iSCSI initiator on the VM can be used to connect VMs directly to a VMAX array.

Virtual Machine storage using VHD and VHDX

You can create and manage Virtual Machine storage by using the Hyper-V Manager tool or PowerShell commands. VHD is a legacy storage format. Starting with Windows Server 2012, Microsoft introduced VHDX. Both VHD and VHDX formats are available; however, VHDX has distinct advantages over the legacy VHD format. VHDX has a larger storage capacity compared the older VHD format. It also provides data corruption protection during power failures and optimizes structural alignments of dynamic and differencing disks to prevent performance degradation on new, large-sector physical disks. A 4 KB logical sector virtual disk allows for increased performance when used by applications and workloads that are designed for 4 KB sectors. When you create a VHDX file, you can configure it to be a fixed size or dynamic. A fixed-size VHDX file has all space allocated when the file is created. A dynamic VHDX file grows as data is written to it, which provides space efficiency. There might be slight overhead for future growth of dynamic VHDX files, but dynamic VHDX on VMAX All Flash storage minimizes performance degradation due to wide striping and virtual provisioned storage on the VMAX All Flash back end.

A VHD Set, a new type of virtual disk model for a guest cluster in Windows Server 2016, is a VHD created with a `.vhds` extension. The `.vhds` file is only 260 KB. A file with a `.vhdx` extension is also created to store the actual data and can be either a dynamic or fixed-size file. The main purpose of a VHD Set is to share the virtual disk between multiple VMs. The VHD Set is useful for deploying guest clusters for SQL Server, file servers, and other services that require shared storage.

Using virtual Fibre Channel for VM connectivity

The Hyper-V virtual Fibre Channel connectivity feature can provide VMs with direct access to VMAX arrays. Virtual Fibre Channel-based storage access provides better storage control, direct access to VMAX storage devices on the VM, and security to the VM. With direct connectivity, the Hyper-V server does not manage storage. Although the VM provides storage management, access to storage is over the virtualized network and Fibre Channel adapters that the Hyper-V server manages.

To support Hyper-V virtual Fibre Channel, you must use N_Port ID Virtualization (NPIV)-capable FC host bus adapters (HBAs) and NPIV-capable FC switches. NPIV assigns World Wide Names (WWNs) to the virtual Fibre Channel adapters that are presented to a VM. Zoning and masking can then be performed between VMAX front-end ports and the virtual WWNs that NPIV created for the VMs. No zoning or masking is necessary for the Hyper-V server itself. Figure 4 shows the mapping between the virtual Fibre Channel and the physical HBA on a server. For information about how to configure the virtual Fibre Channel, refer to Figure 4.

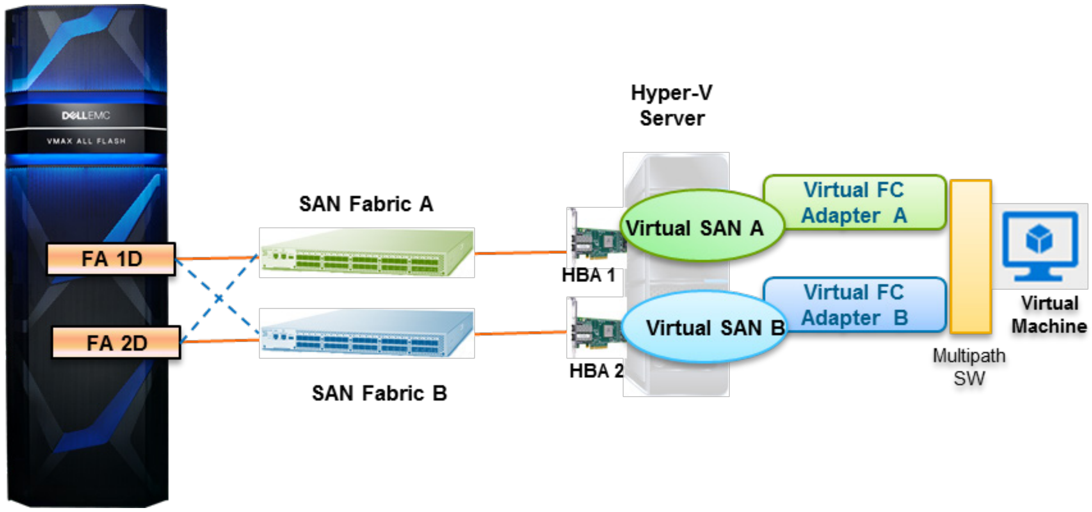

Figure 4. Mapping between the virtual Fibre Channel and the server HBA

The VM identifies VMAX volumes that are presented on the virtual Fibre Channel as directly accessible devices. The concept is similar to how physical servers (or Hyper-V servers) see VMAX volumes by using their HBAs. However, virtual Fibre Channel is an abstraction on the physical FC HBA and the single FC HBA port is logically shared by multiple virtual Fibre Channel ports. A VM can use both virtual Fibre Channel and traditional types of volumes simultaneously. With mixed storage provisioning, the volumes on the virtual Fibre Channel are fully under the VM's control and the Hyper-V server manages the VHDX volumes presented through non-virtual Fibre Channel. Figure 5 shows the PowerShell `Get-Disk` command that lists disks connected over the virtual Fibre Channel and VHDX.

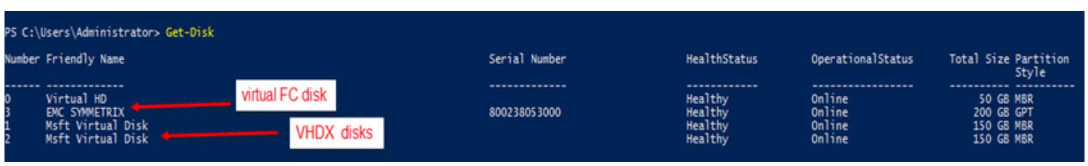

Figure 5. List of VFC and VHDX disks on a VM

During testing, we observed that for an online transaction processing (OLTP) workload, VHDX fixed virtual disk performance was 26 percent lower and dynamic virtual disk performance was 28 percent lower compared to virtual Fibre Channel storage. However, virtual Fibre Channel storage consumed 35 percent more CPU cycles on VMs because the VMs perform all I/O processing. Refer to Figure 6 for the performance comparison.

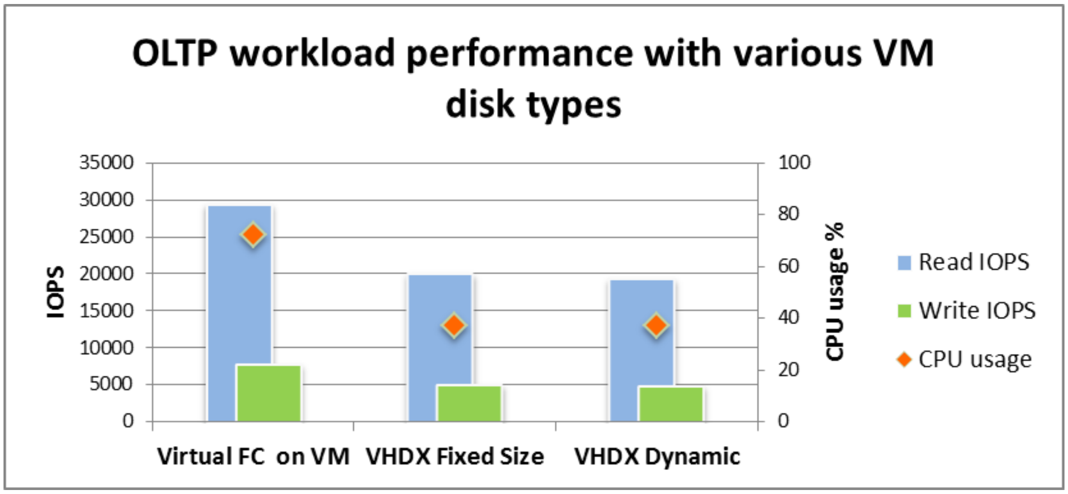

Figure 6. Performance comparison between virtual Fibre Channel and VHDX

iSCSI connectivity for VMs

You can provide VMs direct access to a VMAX array over iSCSI by configuring the iSCSI target on the VMAX array and the Microsoft iSCSI initiator software on the VM. Virtual network interfaces on the VM and iSCSI interfaces on the VMAX array provide network connectivity for iSCSI. The VM operating system is responsible for disk and volume management. Devices on the VMAX array are masked to the VM initiator iSCSI Qualified Name (IQN) and are only visible to the VM that provides isolation and security. However, because the iSCSI protocol stack runs directly on the VM, it uses additional CPU and memory allocated to the VM.

For HA, we recommend that you use at least two virtual network adapters and multipathing software such as PowerPath software or Microsoft Multipath I/O (MPIO) to load balance and automatically fail over or recover paths. We also recommend using jumbo frames for I/O intensive applications by setting the Maximum Transmission Unit (MTU) to 9,000. The MTU must be the same for the storage array, network infrastructure (switch or fabric interconnect), Hyper-V server network cards servicing iSCSI traffic, and on the virtual NICs for the virtual machine. For clustered environments, disable the cluster network communication for any network interfaces that you plan to use for iSCSI.

SMB 3.0 file shares for VMs

SMB file shares from the file server can be presented directly to a VM for storage use or as a target to support VHDs used by VMs. The eNAS file server, which is embedded in the VMAX array, supports SMB 3.0. Figure 7 shows the PowerShell command that creates a VHD over the SMB share.

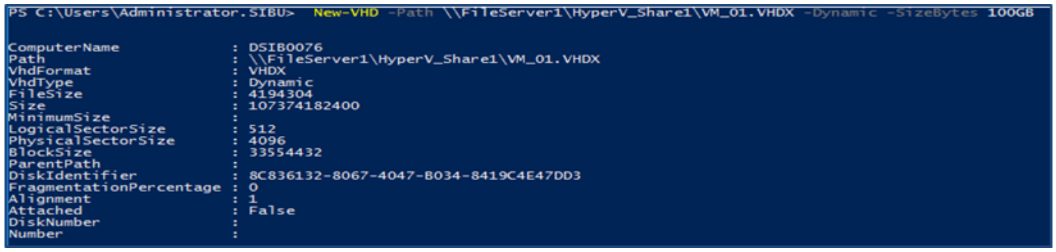

Figure 7. Creating a VHD over an SMB share

Existing VHDs or VHDXs on a file server can also be mapped to a VM by specifying the Universal Naming Convention (UNC) path for the location of the VHD or VHDX. Figure 8 shows how to map an existing VHDX from a file server to a VM.

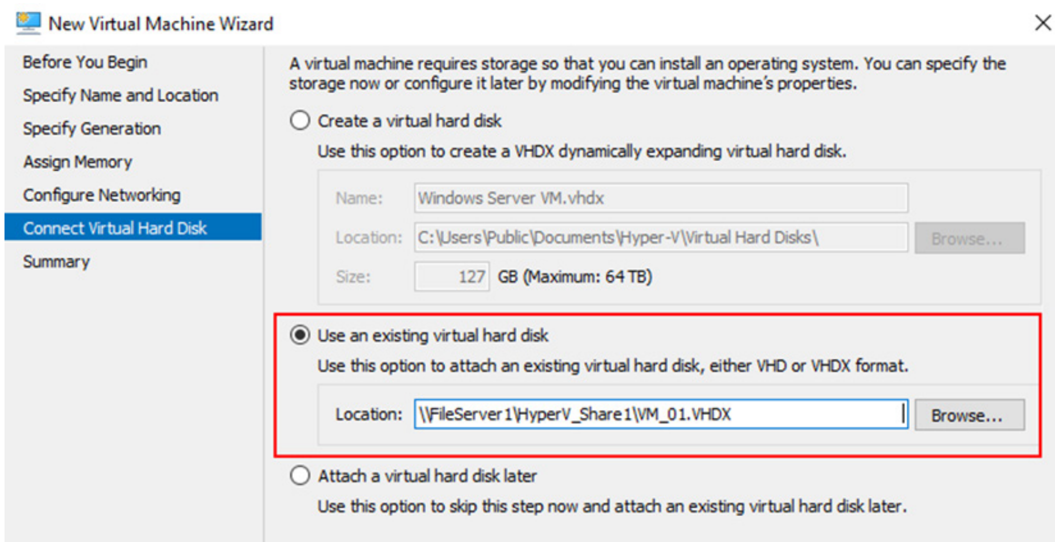

Figure 8. Mapping an existing VHDX on a file server to a VM

Chapter 5 Availability and Mobility for VMs

This chapter presents the following topics:

Overview

Microsoft Windows Server Failover Clustering (WSFC) is a feature of the Windows Server platform that improves the HA of applications and services. You can administer WSFC through the Failover Cluster Manager snap-in. Windows Failover clusters provide HA and scalability to many server workloads. These workloads include server applications such as Microsoft Exchange Server, Hyper-V, Microsoft SQL Server, and file servers. In a failover cluster, if one or more of the clustered servers (nodes) fails, other nodes begin to provide service. For a failover cluster to work, each server's hardware specifications must be the same and the servers must share storage.

Windows failover clustering can be implemented for Hyper-V VMs in the same way as for the Windows cluster environment for other applications such as SQL Server or Exchange Server. VMs become another form of application that failover clustering can manage and protect

Using SRDF/Metro for HA

A WSFC cluster is a group of independent servers that work together to increase the availability of applications and services. Hyper-V Server takes advantage of WSFC services and capabilities to support HA of VMs. Failover clusters also provide Cluster Shared Volume (CSV) functionality that provides a consistent, distributed namespace that clustered roles can use to access shared storage from all nodes. With the Failover Clustering feature, users experience a minimum of disruptions in service. WSFC that uses SRDF/Metro™ enables read/write access to VMAX storage devices to all nodes of the cluster and ensures that the devices are accessible to all the nodes before a failover. SRDF/Metro configuration includes the following steps:

1. Set up a WFSC cluster using multi-path enabled nodes.
2. Provision the VMAX devices for failover cluster needs on both R1 and R2 sites and set up SRDF/Metro using the storage groups and full synchronization of both storage groups.
3. Using multipathing solutions such as Windows MPIO or PowerPath, ensure that VMAX devices are accessible to servers from the local and remote VMAX arrays. This accessibility ensures storage visibility to all failover cluster nodes from both SRDF/Metro-managed storage groups. Multipathing solutions must use round-robin configurations to ensure that all available paths are used optimally.
4. Configure a physical VMAX witness or Virtual Witness (vWitness) to detect and protect from cluster split-brain situations.
5. Scan provisioned VMAX devices by using Windows disk manager, and then bring the disks online and format them.
6. Add the disks as the CSV cluster resource.
7. Use CSV to provision storage for the VMs.

Figure 9 shows the SDRF/Metro setup for data protection.

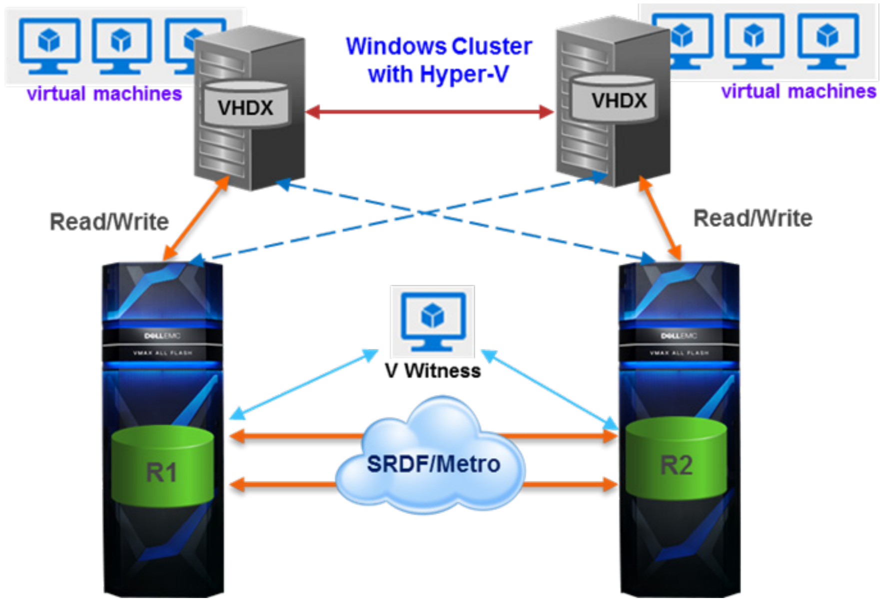

Figure 9. SRDF/Metro Setup for data protection

When one side of an SRDF device pair goes to the Not Ready (NR) state, SRDF/Metro responds by choosing one side of the SRDF pair to remain accessible to the hosts to avoid a split-brain condition. SRDF/Metro chooses by using either of the following methods:

- **Bias mode**—In bias mode, if the SRDF device pair becomes NR on the SRDF link, the bias side always prevails and SRDF/Metro ensures that the bias side remains accessible to the hosts. If still operational, SRDF/Metro then makes the non-bias side inaccessible to the hosts. Inevitably, if the bias side experiences an error, manual intervention might be required to restore state of SRDF site.
- **Quorum mode**— A third-party process (witness) monitors failures in the configuration and takes appropriate action to ensure paired-device availability to hosts wherever possible. This mode provides a higher level of availability compared to the bias-based implementation. When quorum mode is configured, it supersedes the bias functionality. SRDF/Metro also provides asynchronous third-party site support to provide DR over geographical distances to a tertiary site by using SRDF/Asynchronous mode (SRDF/A) from R1 and R2. If one of the Metro sites goes down, customers can continue to replicate from a surviving Metro site to a tertiary site.

Using Snap VX for VM mobility

In a non-clustered environment, you can migrate VMs easily from one Windows Hyper-V Server to another Windows Hyper-V Server that is connected to the same VMAX array. Using Hyper-V live migration, you do not have to perform a physical copy of the data because both Hyper-V servers have access to the same storage devices. Using SnapVX technology, a replica of the VM's VHDs can be created quickly and then presented to the target Windows Server to which the VM must be moved. This approach is faster than the traditional server-to-server copy method, which can be time consuming depending on network speed and the size of the VM. SnapVX can also be used to deploy multiple copies of a VM rapidly. For more information, refer to Appendix A.

Chapter 6 Microsoft integration with EMC storage technologies

This chapter presents the following topics:

SnapVX software

SnapVX provides local storage replication for increased availability and faster data recovery. By using the Dell EMC TimeFinder™/SnapVX feature, up to 256 space-saving copies of live data can be produced within the VMAX array. These snapshots are tracked as versions of the original volume. They do not have to be paired with another device and cause minimal overhead because write operations are made to new thin tracks. A single operation can create or restore (with consistency) snapshots of a storage group. A device can have up to 256 snapshots and up to 1024 linked targets. These copies are immediately available for read/write access by host systems.

AppSync software

AppSync software is an application that enables Integrated Copy Data Management (iCDM) with VMAX storage systems. AppSync software simplifies and automates the process of generating and consuming copies of production data. By abstracting the underlying storage and replication technologies, AppSync software empowers application owners to satisfy copy demand for operational recovery and data repurposing on their own. Storage administrators must only be concerned with initial setup and policy management, resulting in an agile, frictionless environment. AppSync software orchestrates all the activities that are required from copy creation and validation through mounting at the target host and launching or recovering the application. Supported workflows also include refresh, expire, and restore production. AppSync software supports SnapVX Snap (no copy) and SnapVX Clone (copy) for VMAX All Flash storage as the supported copy technologies. For information about how to use AppSync software for data protections in the Windows server environment, refer to Appendix A.

Solutions Enabler and Unisphere for VMAX

The Solutions Enabler CLI (SYMCLI) can be used to manage the VMAX storage array from Hyper-V or a VM that is running on a Hyper-V server. Using the Powershell and SYMCLI command lines, all storage-related operations can be performed from a single server manually or by using a script. Solutions Enabler, which includes the SYMCLI command line interface, is supported on Windows 2012 and 2016 Servers.

Unisphere for VMAX is an advanced Graphical User Interface (GUI) that provides a common Dell EMC user experience across storage platforms. Unisphere for VMAX enables customers to provision, manage, and monitor VMAX environments easily. Unisphere for VMAX can be installed on any Windows operating system running on physical hardware or as a VM.

VSS Hardware Provider

The Dell EMC Volume Shadow Copy Service (VSS) hardware provider can be used to create consistent point-in-time shadow copies of Microsoft applications by using Microsoft’s VSS framework. VSS is a set of COM API interfaces that implement a framework to allow volume backups to be performed while applications on a system continue to write to the volumes. VSS requestors such as DiskShadow and VShadow along with the Dell EMC VSS provider can use VMAX's native snapshot technology SnapVX to create storage-efficient point-in-time copies of storage volumes and restore them as needed. For information about how to use the VSS provider with SnapVX for data protection in a HyperV environment, refer to Appendix A.

Chapter 7 Best practices for optimal performance

This chapter presents the following topics:

Overview

The following best practices are recommended to receive optimal performance from VMAX All Flash storage systems in the Windows Hyper-V environment.

Type, number, and access paths of devices for virtual environments

- **Use "Quick Format" NTFS volumes for thin space saving**—VMAX devices are already thin-provisioned and offering the highest level of storage efficiency. Use of Quick Format allows disk space allocation on an as-needed basis, rather than pre-allocating space that might not be used immediately.
- **Avoid dynamic VHDX for heavy workloads**—A fixed-size VHDX performs better than a dynamic VHDX by roughly five percent to 10 percent in most scenarios. Compared to VHDX, storage access using virtual Fibre Channel provides better performance.
- **Monitor the accumulated I/O load for CSVs**—CSVs are a shared resource. Therefore, I/O can come from any number of parent nodes in a cluster in parallel.
- **Capacity allocation**—VMAX All Flash storage uses thin devices exclusively. On creation, unless you specifically request that devices allocate their capacity fully in the Storage Resource Pool (SRP), the devices only consume as much capacity as the application writes to them. For example, by default, a 1 TB device that is not written to does not consume any storage capacity. This approach enables capacity saving as storage is only consumed based on demand and not during device creation. However, if applications require a guarantee for their allocated capacity such as database logs, these devices can be created fully and allocated in the SRP.
- **Provide enough devices for concurrency (multiple I/O queues), but not so many that they increase management overhead—**Each path to a device creates a SCSI representation on the host. Each representation provides a host I/O queue for that path. Each queue can service a tunable, but limited number (often 32) of I/Os simultaneously. Also, when multiple host devices are used, the storage array can use more parallelism internally for operations such as data movement and local or remote replications. Provide enough devices for concurrency (multiple I/O queues), but not so many that they increase management overhead.

Storage connectivity considerations

When planning storage connectivity for performance and availability, we recommend that you to connect storage ports across different engines and directors than use all the ports on a single director. If a component fails, the storage can continue to service host I/Os.

Dynamic core allocation is a new VMAX feature. Each VMAX director provides services such as front-end connectivity, back-end connectivity, or data management. Each service has its own set of cores on each director that are pooled together to provide CPU resources that can be allocated as necessary. For example, even if host I/Os arrive by a single front-end port on the director, the front-end pool with all its CPU cores is available to service that port. Because I/Os arriving at other directors have their own core pools, we recommend connecting each host to ports on different directors before using additional ports on the same director. This method ensures favorable performance and availability.

Physical host connectivity considerations

For physical host connectivity, we recommend considering the number and speed of the HBA ports (initiators) and the number and size of host devices.

HBA ports

Each HBA port (initiator) creates a path for I/Os between the host and the SAN switch, which then continues to the VMAX array. If a host uses a single HBA port, it has a single I/O path that serves all I/Os. This design is not advisable because a single path does not provide HA and risks a potential bottleneck during high I/O activity for the lack of additional ports for load balancing.

A preferable design is to provide each database server with at least two HBA ports on two separate HBAs. The additional ports provide more connectivity and allow multipathing software, such as PowerPath or Microsoft Multipath I/O (MPIO), to load balance and fail over across HBA paths.

Each path that is configured between the host HBA ports and the VMAX front end by using SAN switch zoning creates a SCSI device representation on the host. The host sees a single VMAX device that is accessible through multiple paths. PowerPath or native Microsoft MPIO multipathing software leverages these paths to allow load balancing and HA of host I/O by using the multiple paths created between host HBA ports and the VMAX FA ports.

While modern operating systems can manage hundreds of devices, it is not advisable or necessary. It also burdens the user with complex tracking and storage provisioning management overhead. We recommend that you have enough HBA ports to support workload concurrency, availability, and throughput. Also use 1:1 relationships for storage front-end ports, and do not have each HBA port zoned and masked to all VMAX front-end ports. This method provides enough connectivity, availability, and concurrency, yet reduces unnecessary complexity of the host registering lots of SCSI devices.

Number and size of host devices

VMAX All Flash storage introduces the ability to create host devices with a capacity from a few megabytes to multiple terabytes. Because VMAX All Flash storage provides native striping across the data pools, you might be tempted to create only a few very large host devices.

For example, a 1 TB Microsoft SQL Server database can reside on a 1 x 1 TB host device or on 10 x 100 GB host devices. While either option satisfies the capacity requirement, we recommend that you use a reasonable number of host devices and a reasonable size. In this example, if the database capacity rises above 1 TB, it is likely that the database administrator might add another device of the same capacity, even if they do not need 2

TB in total. Therefore, large host devices create very large building blocks when additional storage is needed.

Each host device also creates its own host I/O queue for the host operating system. Each queue can service a tunable, but limited, number of I/Os that can be transmitted simultaneously. If, for example, the host has four HBA ports, and a single 1 TB LUN with multipathing software, the host has only four paths available to queue I/Os. A high level of database activity generates more I/Os than the queues can service, resulting in artificially elongated latencies. In this example, two or more host devices are advisable to alleviate such an artificial bottleneck. Host software, such as PowerPath or Windows Perfmon, can monitor host I/O queues to ensure that the number of devices and paths is adequate for the workload.

Another benefit of using multiple host devices is that the storage array can use more parallelism internally when operations such as local and remote replications take place. By performing more copy operations simultaneously, the overall operation takes less time.

While there is no specific number for the size and number of host devices, we recommend that you find a reasonably low number that offers enough concurrency, provides an adequate building block for capacity when additional storage is needed, and does not become too large to manage.

Dell EMC Virtual Provisioning™ and thin devices

All VMAX host devices use Virtual Provisioning (also known as thin provisioning). Devices are actually a set of pointers to capacity allocated at 128 KB extent granularity in the storage data pools. However, to the host, they appear and respond like regular LUNs. Using pointers enables better capacity efficiency for local replication using SnapVX software by sharing extents when data does not change between snapshots.

Virtual provisioning offers the following choices:

- To allocate the host device capacity fully. A fully allocated device consumes all its capacity in the data pool on creation. Therefore, there is no risk that future writes might fail if the SRP has no capacity left.
- To perform allocation on demand. Allocation on demand enables over-provisioning. Although the storage devices are created and look to the host as being available with full capacity, actual capacity is only allocated in the data pools when host writes occur. This is a common cost saving practice.

Use allocation on demand when:

- The capacity growth rate for an application is unknown.
- You do not want to commit large amounts of storage ahead of time, as it might never be used.
- You do not want to disrupt host operations at a later time by adding more devices.

If you use allocation on demand, capacity is physically assigned only when it is needed to meet application requirements.

Host I/O Limits and multi-tenancy

The Host I/O Limits Quality of Service (QOS) feature was introduced in the previous generation of VMAX arrays but it continues to provide the option to place specific IOPs or bandwidth limits on any storage group. Assigning a specific Host I/O Limit for IOPS, for example, to a storage group with low performance requirements, can ensure that a spike in I/O demands does not saturate or overload the storage, affecting performance of critical applications. As the test cases for multitenant application workloads show later in this paper, although VMAX arrays can maintain high performance at low latency, using Host I/O Limits might limit the impact of noisy neighbors on mission critical applications.

Resilient File System (ReFS)

Microsoft's ReFS was introduced in Windows Server 2012, but its visibility has been limited because installations still use NTFS by default. However, ReFS is recommended for Hyper-V hosts in Server 2016. It is faster for certain operations that Hyper-V uses, including creating a fixed-size VHDX and performing a file merge, which is used by Hyper-V checkpoints and Hyper-V Replica.

Using Thin Provisioning

VMAX All Flash storage is a fully thin-provisioned storage system that provides the highest level of storage efficiency as the SQL server storage groups only consume storage capacity for the data that is actually written by the host and grow as needed with host writes. So the storage efficiency has already realized when SQL Server storage groups are deployed on VMAX devices. All VMAX data services continue to offer highly efficient data copy and capacity utilization even when using VMAX Snap VX for periodic snapshots or SRDF for remote replication.

Using compression

The VMAX Adaptive Compression Engine (ACE) uses a versatile architecture to achieve a high storage compression ratio (CR) while taking into account data access patterns and activity level. In general, the CR depends on the actual data. In our testing, for Microsoft VM images, we achieved a CR of 1.6:1 for the operating system. The CR for the data might be higher depending on the data type. ACE might reduce the effective CR based on the amount of active data in the storage group.

Chapter 8 Conclusion

This chapter presents the following topic:

Conclusion

VMAX storage arrays provide all forms of storage connectivity that is required by Windows Server 2012 R2 and Windows Server 2016 Hyper-V deployments. You can deploy any form of Hyper-V Server managed by, or directed to VM connectivity, and can even combine multiple forms of connectivity to satisfy application-level requirements.

Each form of storage connectivity provides different management or operational features. For example, storage that is provisioned directly to a VM, using virtual Fibre Channel, iSCSI, SMB 3.0, or pass-through disks, restricts that storage volume only to VMs. Conversely, storage that is allocated in VHD devices that are created on volumes within the Hyper-V server allows a single LUN to be collocated among any number of VMs. The various VHD devices are also shared with, and on the parent-managed volume. When you use a common volume to collocate VHD devices, you can also affect some HA or mobility solutions, because a change to the single LUN affects all VMs that are located on the LUN. This action can affect configurations using failover clustering.

However, the implementation of CSV Windows Server failover clustering addresses the need for HA of consolidated VHD deployments. When you collocate VHD devices on a single storage LUN, consider how to address the cumulative workload. In cases where an application such as Microsoft SQL Server or Microsoft Exchange Server is deployed in a VM, use sizing best practices to ensure that the underlying storage can support the anticipated workload. When collocated VHD devices are placed on a common storage volume (LUN), provision the device to ensure that it can satisfy the cumulative workload of all applications and operating systems that are located on the VHDs. When storage has been under-provisioned from a performance perspective, all collocated applications and VMs can be adversely affected.

VMAX All Flash storage is well suited to run mission-critical and high-performance SQL Server workloads while leveraging VMAX features, which include data protection, data services, resiliency, and availability.

The key value points for running Hyper-V Server on VMAX All Flash storage include:

- **Scale**—VMAX All Flash storage provides a flexible and scalable storage design that can start as low as 11 terabytes per V-Brick for VMAX 250 arrays and can go as high has 4 petabytes capacity for VMAX 850 arrays.
- **Write cache**—VMAX cache eliminates flash-related write latency delays, which can affect log writes during commits and write rates during batch or database file initialization.
- **Connectivity**—VMAX All Flash storage provides a flexible choice of connectivity and protocols, including FC, iSCSI, and SMB 3.0 (with eNAS).
- **Data integrity**—All data in VMAX All Flash storage is protected with T10-DIF, including local and remote replications.
- **Remote replications**—Remote replications provide robust HA, business continuity (BC), and DR support for many topologies and modes, including synchronous and asynchronous replications, Cascaded, Extended Distance Protection, STAR, and active/active with SRDF Metro.

- **Local replications**—The 256 snapshots for any host device with up to 1,024 linked targets enable better HA and an easier way of creating database replicas, backups, and so on.
- **Multitenancy**—By using the Host I/O Limits feature, ensure that applications that are not performance-critical can be limited by how many resources they get, using IOPS or bandwidth semantics. This method allows better consolidation and sharing of the VMAX All Flash storage system across many applications with varied business priorities.
- **Self-service automated copy data management using AppSync software**—AppSync software support for VMAX All Flash storage offers self-service, automated service plan-driven copy data management for Microsoft Windows Server. This supports periodic backup, instant restore and recovery, and repurposed copies for test, dev, and reporting, with tighter integration with Windows Server Virtual Device Interface (VDI).
- **Adaptive Compression Engine (ACE)**—With the Q3 2016 HyperMax OS release, VMAX All Flash storage now uses the ACE to compress data and efficiently optimize system resources to balance overall system performance.

Chapter 9 References

This chapter presents the following topic:

Dell EMC documentation

The following documentation on Dell EMC.com or Online Support provides additional and relevant information. Access to these documents depends on your login credentials. If you do not have access to a document, contact your Dell EMC representative.

- *EMC VMAX All Flash Product Guide*
- *VMAX All Flash Storage Family Sheet Detailed Overview*
- *Dell EMC VMAX All Flash Storage for Mission-Critical SQL Server Databases*
- *Microsoft SQL Server High Availability using VMAX All Flash SRDF/Metro*
- *VMAX eNAS Deployment for Microsoft Windows and SQL Server Environments*
- *EMC Symmetrix VMAX Virtual Provisioning Space Reclamation and Application Considerations*
- *Deployment Best Practices for Microsoft SQL Server with VMAX3 SLO Management*
- AppSync Support Matrix
- *EMC AppSync User and Administration Guide*
- *AppSync with VMAX All Flash White Paper*
- *EMC Unisphere for VMAX REST API Concepts and Programmer's Guide*

Appendix A Configuring a Deployment

This appendix presents the following topics:

Configuring virtual Fibre Channel for Hyper-V VMs

For the initial configuration of virtual Fibre Channel:

1. Create a new virtual Fibre Channel SAN within the Hyper-V Virtual SAN Manager, as shown in Figure 10.

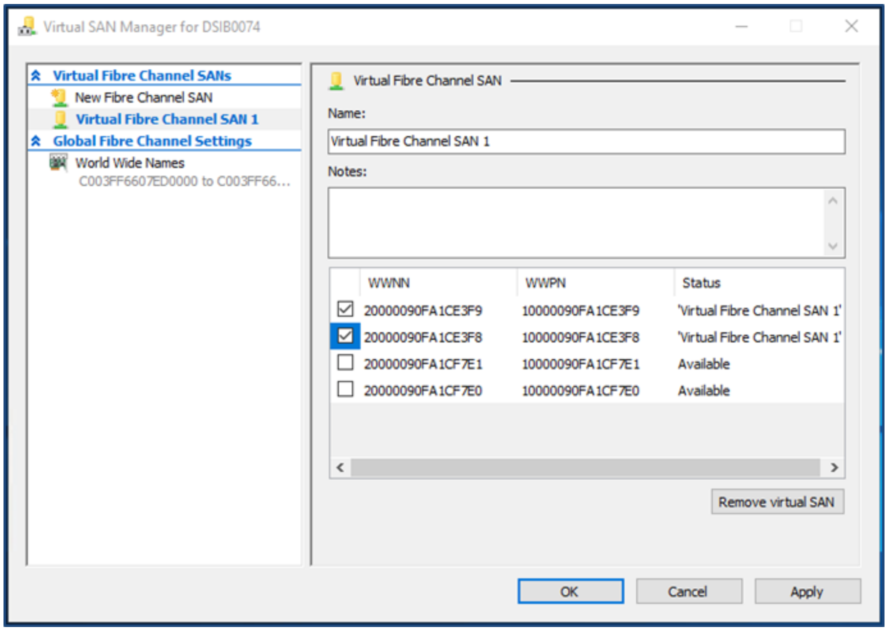

Figure 10. Creating a virtual Fibre Channel SAN

The Hyper-V SAN is a logical construct where physical HBA ports are assigned to multiple VMs by using virtual Fibre Channel. You can place one or multiple HBAs in a virtual Fibre Channel SAN for port isolation or for deterministic fabric redundancy. Use the same virtual SAN configuration and naming conventions on all Hyper-V servers in a clustered environment. This method ensures that each node can take ownership and host a highly available guest with virtual Fibre Channel

2. Add a Fibre Channel adapter to the VM and attach it to the virtual SAN. If required, you can use a pre-created WWPN for the virtual Fibre Channel adapter, as shown in Figure 11, or modify them. We recommend that you add a minimum of two virtual adapters to a VM, each assigned to a different physical FC adapter. Physical adapters must be on a separate fabric to ensure that there is no single point of failure.

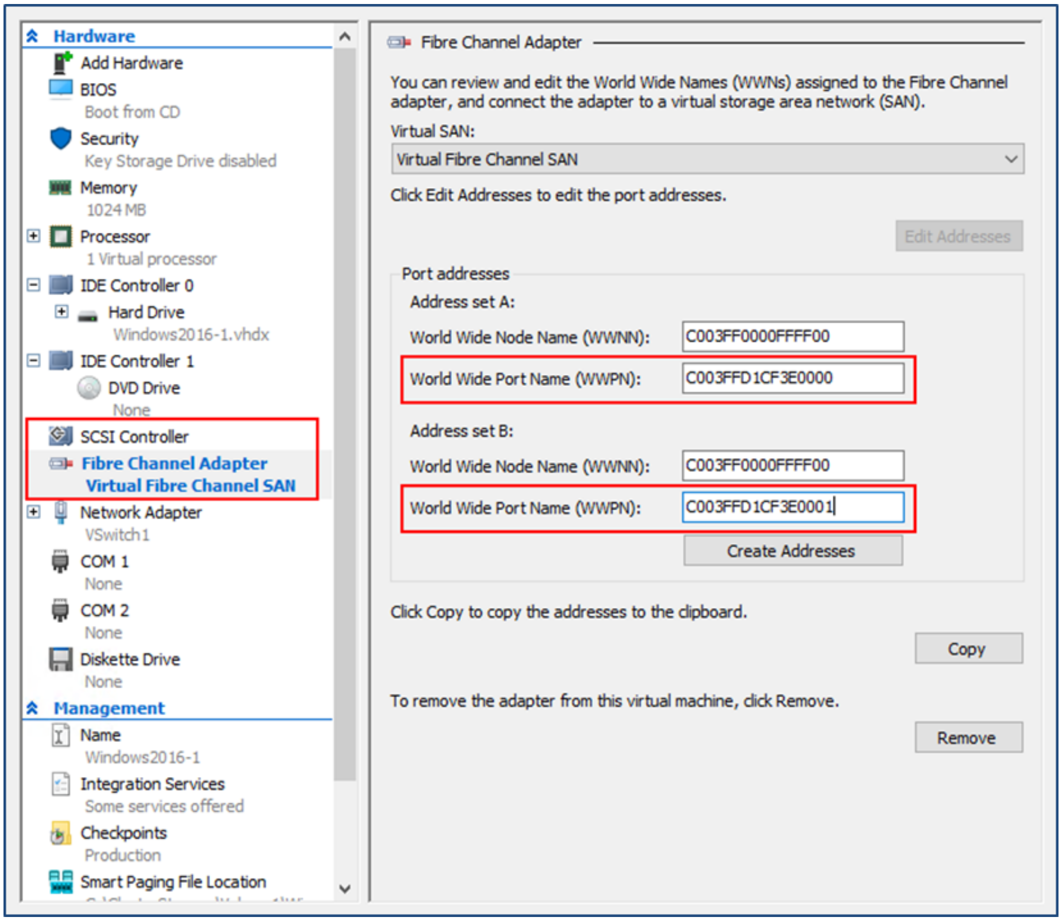

Figure 11. Adding a virtual Fibre Channel Adapter to a VM

3. Use the `Get-InitiatorPort` command to list the WWNs of the virtual Fibre Channel, as shown in Figure 12. Use the port address to configure zoning on the switch and for creating a masking view on the VMAX array.

```
PS C:\Users\Administrator.SIBU> Get-InitiatorPort

InstanceName                                                                     NodeAddress      PortAddress      ConnectionType
------------                                                                     -----------      -----------      --------------
VMBUS\{2f9bcc4a-0069-4af3-b76b-6fd0be528cda}\{a8fb2395-bed6-40cd-83f6-618024571386}_0 c003ff0000ffff00 c003ff6607ed0002 Fibre Channel
VMBUS\{2f9bcc4a-0069-4af3-b76b-6fd0be528cda}\{b44be2af-3f11-4c10-bb21-6ef318391e50}_0 c003ff0000ffff00 c003ff6607ed0004 Fibre Channel
```

Figure 12. Retrieving host initiator ports using the PowerShell command

Using AppSync software for Hyper-V server copy data management

Microsoft Windows Servers running Hyper-V must be fully operational at all times and the data must be continuously protected depending on the required recovery point objective (RPO) and recovery point objective (RTO). This emphasizes the need for fast and efficient data protection and replications, and the ability to meet these requirements without overhead or operations disruption. Examples include the ability to create local and remote replicas in seconds without disruption to host CPU or I/O activity for purposes such as patch testing, running reports, creating development sandbox environments, publishing data to analytic systems, offloading backups from Production, DR strategy, and so on.

Traditional solutions rely on host-based replications. The disadvantages are that additional host I/O and CPU cycles are consumed by the need to create such replicas and the complexity of monitoring and maintaining them, especially across multiple servers, and the elongated time and complexity associated with their recovery.

AppSync software provides simple, automated, and self-service copy management automation for Hyper-V Server that runs on VMAX All Flash arrays. The following steps describe how to use AppSync software for Windows Hyper-V server with SnapVX on VMAX All Flash arrays:

1. Configure the AppSync server.
2. Run VMAX All Flash storage discovery and provisioning for Hyper-V storage.
3. Run host agent registration and discovery of Microsoft Windows Server.
4. Select or configure service plans and subscribe the file system or volume to the plan.
5. Schedule the copies and use them for repurposing or restore.

Refer to the EMC AppSync User and Administration Guide for more details about Dell EMC AppSync software, refer to the *EMC AppSync User and Administration Guide*.

Configuring the AppSync Server

The AppSync server can be installed on a supported Windows environment. Refer to the Dell EMC E-lab Navigator for supported environments. After the AppSync server is configured, it can be accessed from any web browser.

Discovering and provisioning storage

The AppSync server can discover a VMAX All Flash array by registering a pre-configured SMI-S provider. As part of the discovery process, the AppSync software registers devices not associated with any existing masking views as available replication storage. You can run the discovery process periodically or on demand to add new devices to be used for copy management. Figure 13 shows the storage discovery and provisioning for copy management. The AppSync software can discover and manage multiple storage arrays.

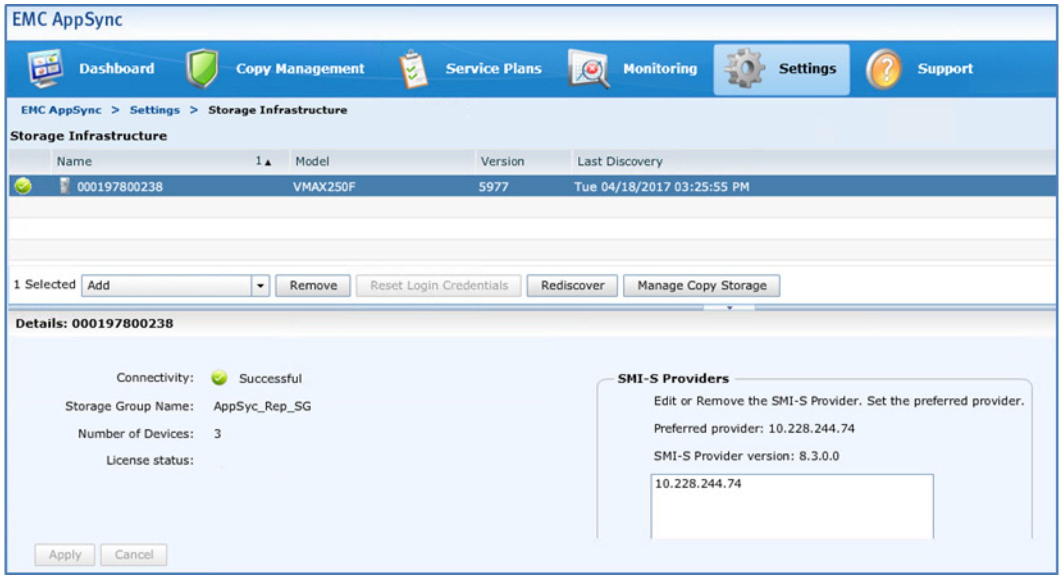

Figure 13. AppSync storage discovery

Registering host agents

The AppSync server enables push install of Microsoft Windows host agents using the Windows administrator account credentials, as shown in Figure 14. After the host has been successfully registered, the AppSync software shows all volumes that are eligible for copy management.

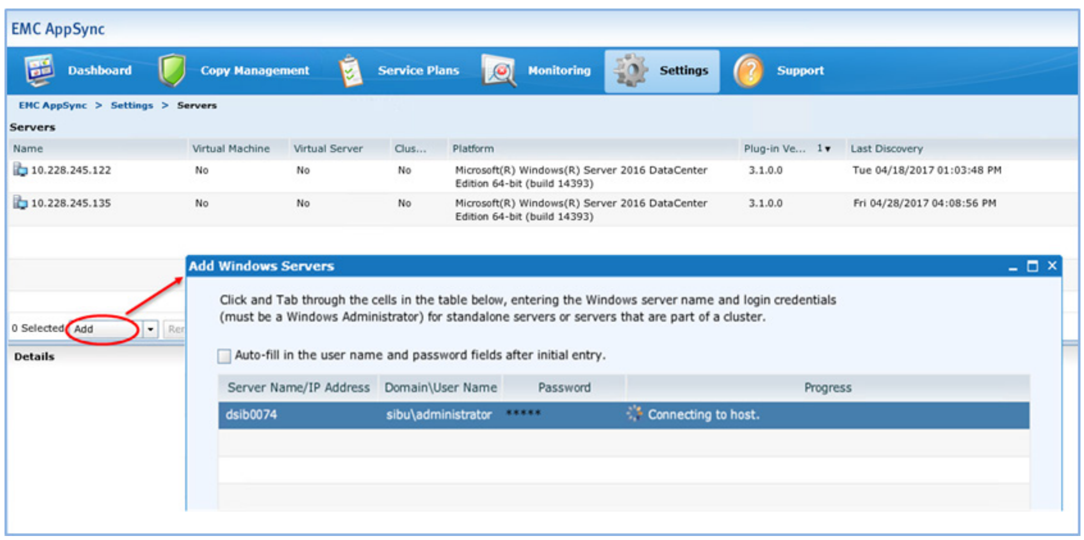

Figure 14. Registering the hose agents

Selecting or configuring service plans

The AppSync software comes with pre-configured service plans for copy management:

- The Bronze service plan provides local replication.
- The Silver service plan provides remote replication.
- The Gold service plan provides both local and remote replication capabilities with default settings.

You can also customize service plans by using existing service plans as a template to meet the needs for the copy data management for Windows Server with Hyper-V. Customized service plans can specify the list of tasks that are associated with the plan, the recurrence pattern, and the schedule. When the service plan is configured, the server volumes can subscribe to the plans for copy management. Figure 15 shows how to create a customized plan for an hourly backup.

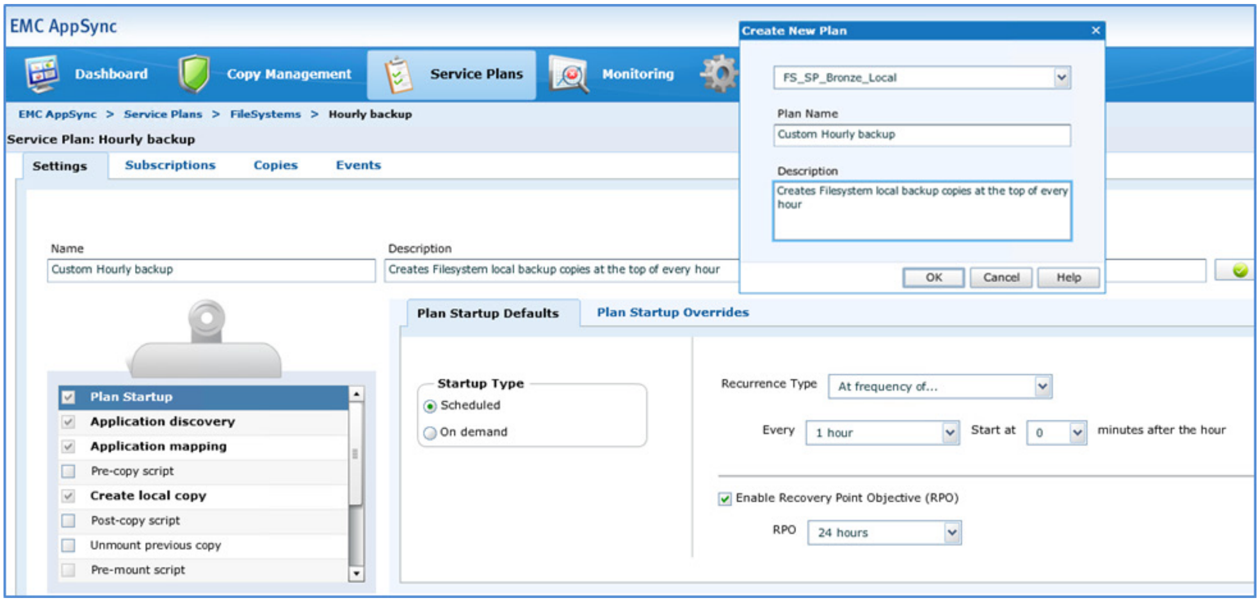

Figure 15. Service plan configuration and customization

Scheduling database copies and using them for repurposing

You can run database copies on demand or schedule them for automated copy management. When the copies are created, they can be used as a mount point for test/dev/reporting or used to restore source database. Figure 16 shows how to schedule VHDX copies

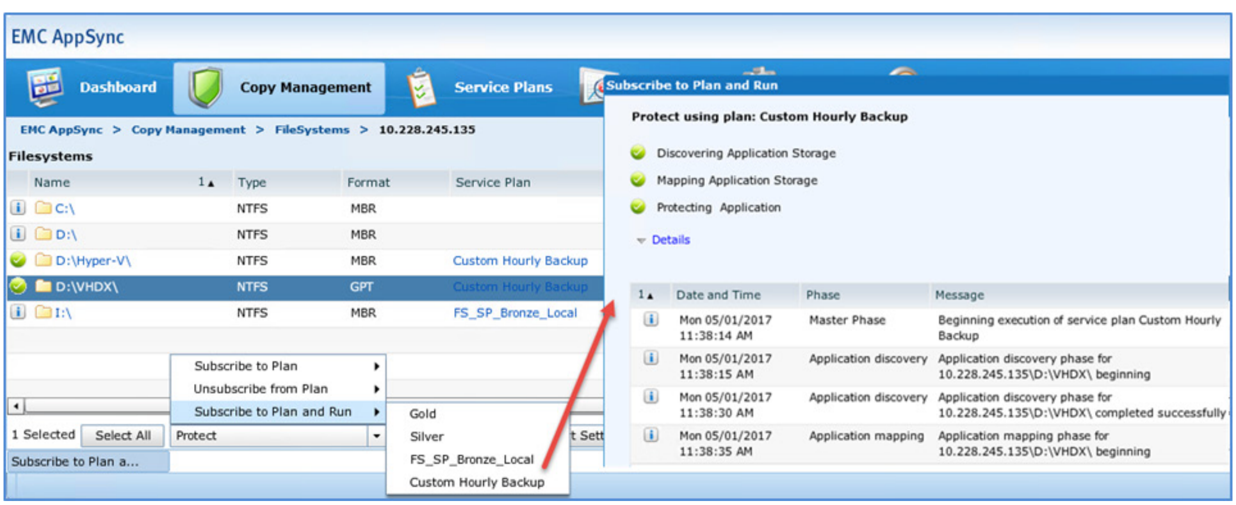

Figure 16. Scheduling VHDX copies

Managing snapshots

To manage snapshots:

1. Before using the VSS hardware provider to take snapshots, use SYMCLI commands to create a snapshot of the volumes being used by Windows Server and link it to a target volume. The target device must be the same size as the source and cannot be part of any storage group. The following example shows that a snapshot is taken of VMAX device 26 and then linked to target device 56.

```
symsnapvx  -sid 238 establish -devs 26 -name HyperV_SNAP_1
symsnapvx  -sid 238 link -devs 26 -lndevs 56 -snapshot_name
SNAP_Test –copy
```

2. To create snapshots of the source volume as needed, use the DiskShadow utility on the Windows server. The latest copy snapshots are automatically linked to the target so that the target volume is refreshed automatically. Use a unique name for the metadata file. This file is required for restore or expose operations of the snapshot later.

```
C:\Users\Administrator>diskshadow
DISKSHADOW> set context persistent
DISKSHADOW> set option transportable plex
DISKSHADOW> set metadata c:\ diskshadowdata\snap_10am.cab
DISKSHADOW> begin backup
DISKSHADOW> add volume I: alias Data_VSS_1
DISKSHADOW> create
DISKSHADOW> end backup
DISKSHADOW> exit
```

3. Expose any snapshot as a drive or mount point for backup or other purposes by using the metadata file. This is a read-only copy or the snapshot.

```
DISKSHADOW> load metadata C:\diskshadowdata\snap.cab
DISKSHADOW> import
DISKSHADOW> expose  % Data_VSS_1% Y:
-> %VSS_SHADOW_1% = {be64fc2e-78df-496f-b820-b3e62ac7dd1b}
```

   In this example, the shadow copy is successfully exposed as `Y:\`.

4. If read/write access is required for an exposed snapshot, use the `break` command to make the copy read/write. The `break` command is relevant only for hardware shadow copies after import. The copy can then be accessed using a drive letter (if assigned), volume name, or other means after executing the break.

```
DISKSHADOW> break readwrite NOREVERTID  %VSS_SHADOW_SET%
-> %VSS_SHADOW_SET% = {66fa135b-66fb-4576-9dc8-be2a350b3581}
The shadow copy is now an independent volume
```

5. To restore snapshots, use the metadata of the snapshot. In the DiskShadow utility,

   a. Load the metadata.

   b. Begin the restore operation.

c. When PreRestore completes, use the SYMCLI command to restore the snapshot followed by `end restore` command. In the following example the SYMCLI `restore` command is in the `restore.cmd` file and executed directly from the DiskShadow utility.

```
DISKSHADOW> load metadata C:\VSSData\Snap.cab
Alias RestoreTest for value {ef923419-c22c-4590-97f8-4ef54ba48617}
set as an environment variable.
Alias VSS_SHADOW_SET for value {3bfb611c-1904-4a29-a0a1-
89c1828c138d} set as an environment variable.
DISKSHADOW> begin restore
Number of writers listed: 10
The PreRestore event was successfully sent to the writers.
The restore operation has started. Please issue END RESTORE when
it is complete.
DISKSHADOW> exec c:\VSSData\restore.cmd
C:\Users\Administrator>symsnapvx -sid 238 restore -dev 2F -
snapshot_name VSS_142743_04282017 -nop
Restore operation execution is in progress for the device
range(s). Please wait...
    Polling for
Restore..............................................Started.
    Polling for
Restore..............................................Done.
Restore operation successfully executed for the device range(s)
DISKSHADOW> end restore
```

You can put all DiskShadow commands in a text file and issue the `diskshadow -s` *`filename`* command to run operations. The following is an example of the commands file:

```
#DiskShadow script file

set context persistent set option transportable plex
set metadata c:\diskshadowdata\snap.cab
set verbose on

begin backup
add volume I:
add volume d: alias DataVolumeShadow
create

expose %DataVolumeShadow% q:

# execute some command or script after creating snapshot
#exec c:\diskshadowdata\backupscript.cmd

end backup
#End of script
```

Using SnapVX for rapid deployment of VMs

TimeFinder SnapVX provides storage-consistent snapshots that use redirect-on-write for added performance, and pointer-based management for deduplication-like data reduction for efficiency. In the Hyper-V environment, snapshots of a VMAX storage group that contains VMs can be used to deploy large number of VMs rapidly and efficiently.

Use the following procedure to deploy VMs using SnapVX:

1. Create a SnapVX snapshot of the VMAX storage containing the VM data disks by using the following SYMCLI command on Unisphere for VMAX:

```
$ symsnapvx -sid  999  -nop -sg hyperv_src  establish -name vm_snap_1
```

2. Create a target storage group that has the identical number of volumes and volume sizes as the source storage group.
3. Link the snapshot to the target storage group. SnapVX allows space-efficient, no-copy point-in-time snapshot linked to the target device that consumes additional storage capacity on future updates because these snaps leverage track sharing with their source device for all non-modified data.

```
 $ symsnapvx -sid  999 –sg hyperv_src  -snapshot_name vm_snap_1
link  -lnsg snaptarget
```

4. To target the Hyper-V server and rescan the server to discover new disks from the VMAX array, mask the linked storage group.
5. Make the newly discovered disks online and assign them a drive letter or a mount path.
6. Create a VM by using the virtual hard disks from the new volumes that are discovered in step 5.
7. To create multiple VMs, repeat step 2 through step 6. Link the same snapshot to multiple target storage groups.