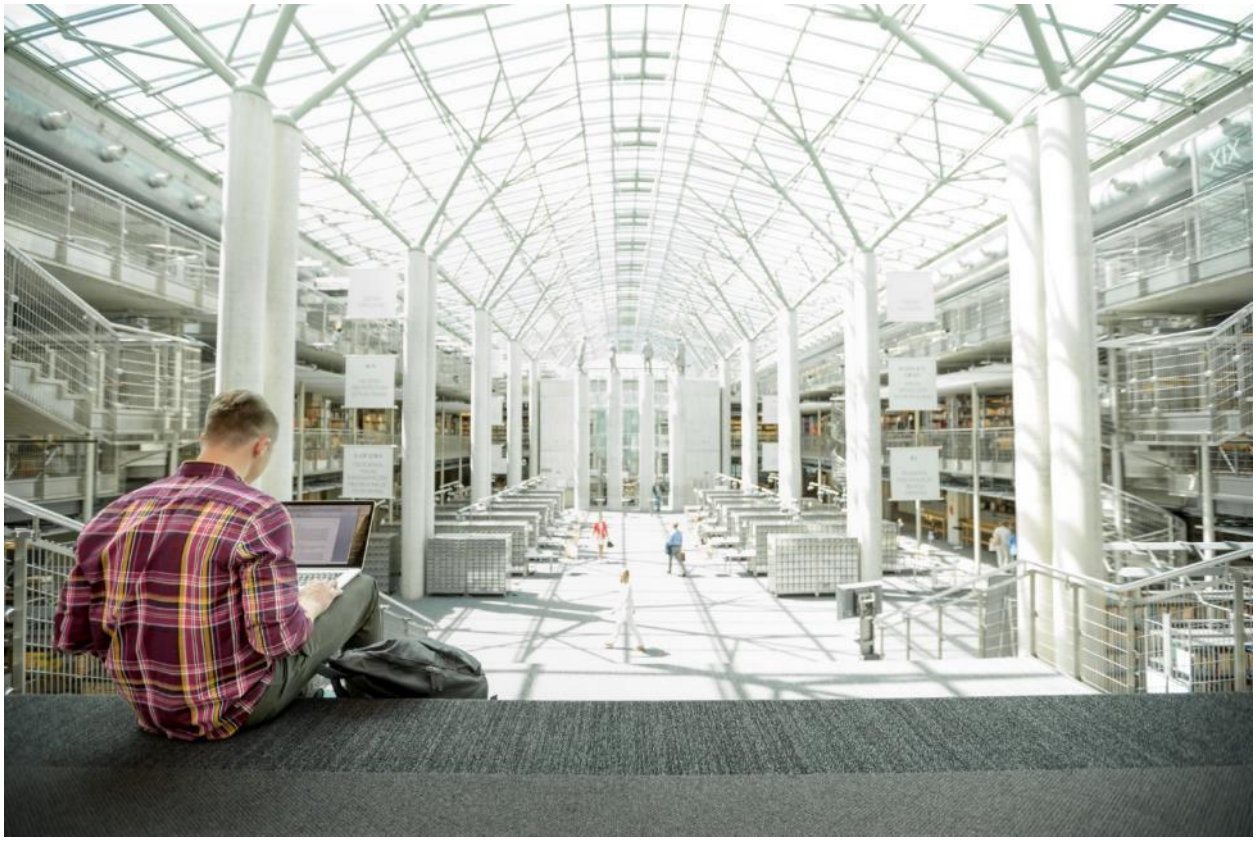


Cisco HyperFlex 3.5 All-Flash Systems for Deploying Microsoft SQL Server 2016 Databases with Hyper-V

Deployment Best Practices and Recommendations for Microsoft SQL Server 2016 Databases on Cisco HyperFlex 3.5.1a and Cisco UCS C240 M5 All-Flash Systems with Windows Server 2016 Hyper-V

Last Updated: December 14, 2018

About the Cisco Validated Design Program

The Cisco Validated Design (CVD) program consists of systems and solutions designed, tested, and documented to facilitate faster, more reliable, and more predictable customer deployments. For more information, see:

http://www.cisco.com/go/designzone.

Table of Contents

Executive Summary

Cisco HyperFlex™ Systems deliver complete hyperconvergence, combining software-defined networking and computing with the next-generation Cisco HyperFlex HX Data Platform. Engineered on the Cisco Unified Computing System™ (Cisco UCS®), Cisco HyperFlex Systems deliver the operational requirements for agility, scalability, and pay-as-you-grow economics of the cloud—with the benefits of on-premises infrastructure. With a hybrid or All-flash-memory storage configurations and a choice of management tools, Cisco HyperFlex Systems deliver a pre-integrated cluster with a unified pool of resources that you can quickly deploy, adapt, scale, and manage to efficiently power your applications and your business.

With the latest All-Flash storage configurations, a low latency, high performing hyperconverged storage platform has become a reality. This makes the storage platform optimal to host the latency sensitive applications like Microsoft SQL Server. This document provides the considerations and deployment guidelines to have a Microsoft SQL server virtual machine setup on an All-Flash Cisco HyperFlex Storage Platform.

Solution Overview

Introduction

Cisco HyperFlex™ Systems unlock the potential of hyperconvergence. The systems are based on an end-to-end software-defined infrastructure, combining software-defined computing in the form of Cisco Unified Computing System servers; software-defined storage with the powerful Cisco HX Data Platform and software-defined networking with the Cisco UCS fabric. Together with a single point of connectivity and hardware management, these technologies deliver a pre-integrated and an adaptable cluster that is ready to provide a unified pool of resources to power applications as your business needs dictate.

Microsoft SQL Server 2016 is the relational database engine release from Microsoft, which has new features and enhancements to the relational and analytical engines. It is built to provide a consistent and reliable database experience to applications delivering high performance. Currently, more and more database deployments are virtualized and hyperconverged storage solutions are gaining popularity in the enterprise space. The Cisco HyperFlex All-Flash system is the latest hyperconverged storage solution providing a high performing and cost-effective storage solution making use of the high speed SSDs locally attached to the Windows Hyper-V hosts. It is crucial to understand the best practices and implementation guidelines that enable customers to run a consistently high performing SQL server database solution on a hyperconverged All-Flash solution.

Audience

This document is intended for system administrators, database specialists and storage architects who are planning, designing and implementing Microsoft SQL Server database solution on Cisco HyperFlex All-Flash storage solution.

Purpose of this Document

This document discusses reference architecture and implementation guidelines for deployment of SQL Server 2016 database instances on Cisco HyperFlex All Flash solution.

What's New in this Release?

The list below provides the new features and enhancements added in HyperFlex 3.5.1a release on Hyper-V.

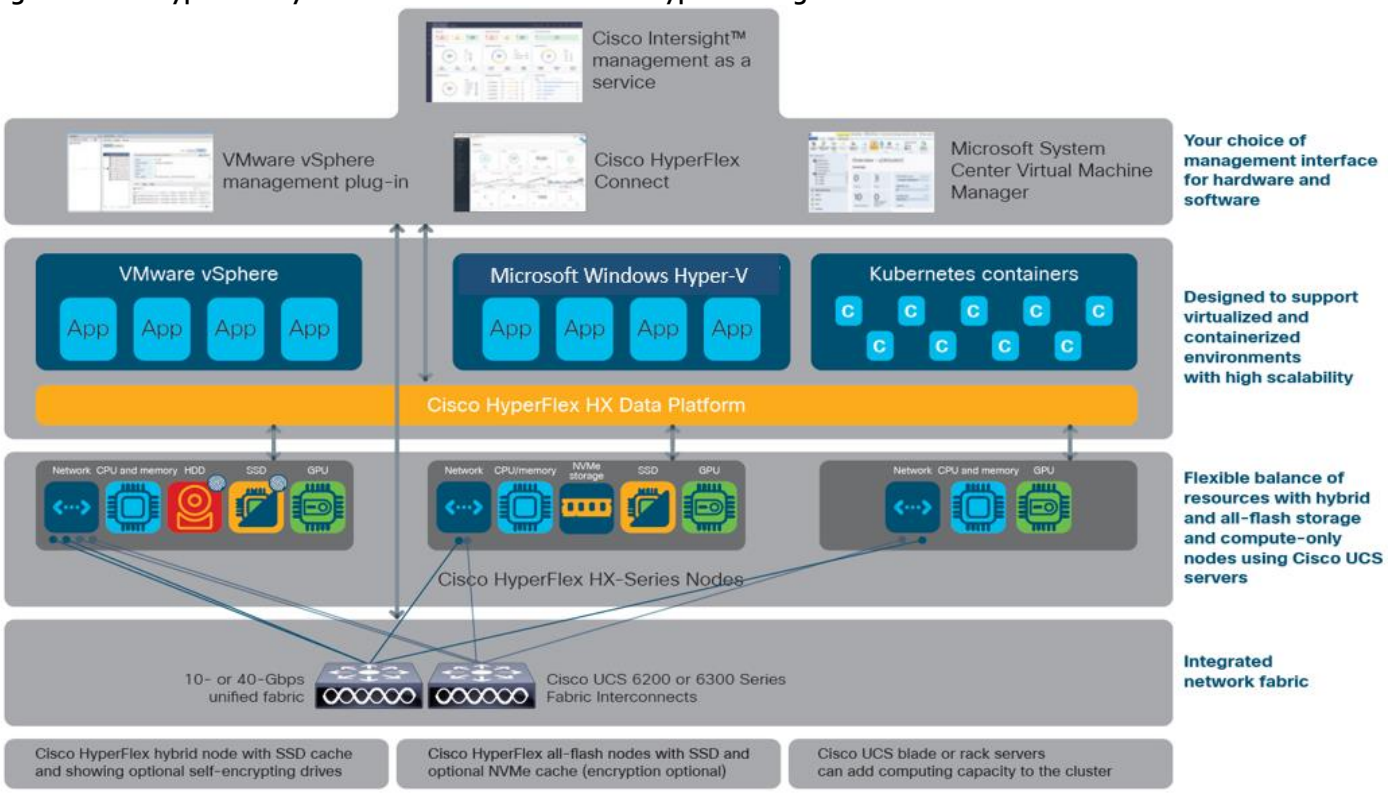
- Cluster Expansion for Hyper-V converged nodes
- Windows Server OS bare metal installation is included as part of HyperFlex cluster creation workflow
- New in-band Cisco IMC access management option added. It is recommended and default option for Hyper-V
- Large form factor (LFF) drives are supported with this release on HyperFlex with Hyper-V

Technology Overview

HyperFlex Data Platform

Cisco HyperFlex Systems are designed with an end-to-end software-defined infrastructure that eliminates the compromises found in first-generation products. Cisco HyperFlex Systems combine software-defined computing in the form of Cisco UCS® servers, software-defined storage with the powerful Cisco HyperFlex HX Data Platform Software, and software-defined networking (SDN) with the Cisco® unified fabric that integrates smoothly with Cisco Application Centric Infrastructure (Cisco ACI™). With All-Flash memory storage configurations, and a choice of management tools, Cisco HyperFlex Systems deliver a pre-integrated cluster that is up and running in an hour or less and that scales resources independently to closely match your application resource needs (Figure 1).

Figure 1 Cisco HyperFlex Systems Offer Next-Generation Hyperconverged Solutions



The Cisco HyperFlex All Flash HX Data Platform includes:

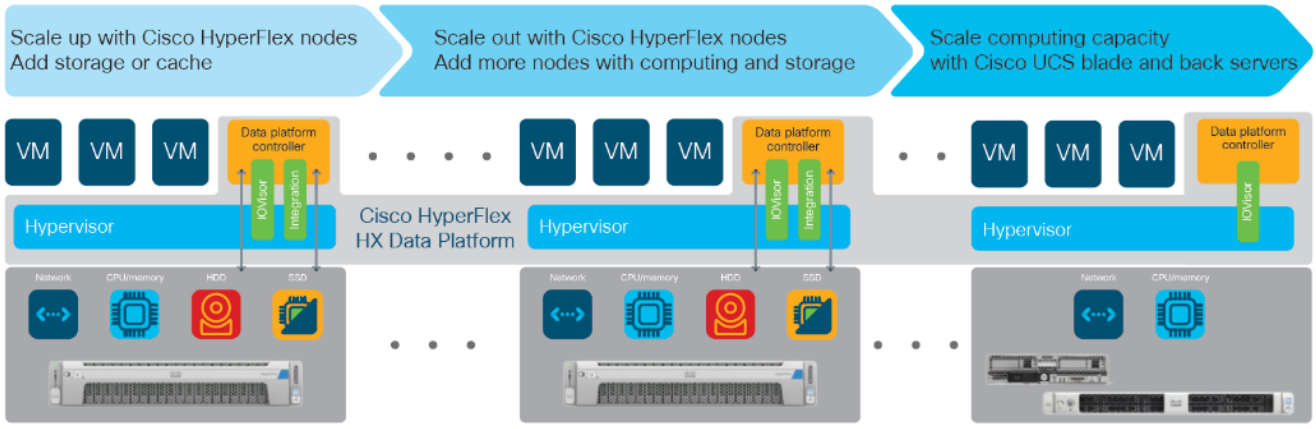
- Enterprise-class data management features that are required for complete lifecycle management and enhanced data protection in distributed storage environments—including replication, always on inline deduplication, always on inline compression, thin provisioning, instantaneous space efficient clones, and snapshots.
- Simplified data management that integrates storage functions into existing management tools, allowing instant provisioning, cloning, and pointer-based snapshots of applications for dramatically simplified daily operations.
- Improved control with advanced automation and orchestration capabilities, robust reporting, and analytics features that deliver improved visibility and insight into IT operation.

- Independent scaling of the computing and capacity tiers, giving you the flexibility to scale out the environment based on evolving business needs for predictable, pay-as-you-grow efficiency. As you add resources, data is automatically rebalanced across the cluster, without disruption, to take advantage of the new resources.
- Continuous data optimization with inline data deduplication and compression that increases resource utilization with more headroom for data scaling.
- Dynamic data placement optimizes performance and resilience by making it possible for all cluster resources to participate in I/O responsiveness. All-Flash nodes use SSD drives for caching layer as well as capacity layer. This approach helps eliminate storage hotspots and makes the performance capabilities of the cluster available to every virtual machine. If a drive fails, reconstruction can proceed quickly as the aggregate bandwidth of the remaining components in the cluster can be used to access data.
- Enterprise data protection with a highly-available, self-healing architecture that supports non-disruptive, rolling upgrades and offers call-home and onsite 24x7 support options
- API-based data platform architecture that provides data virtualization flexibility to support existing and new cloud-native data types

Architecture

In Cisco HyperFlex Systems, the data platform spans three or more Cisco HyperFlex HX-Series nodes to create a highly available cluster. Each node includes a Cisco HyperFlex HX Data Platform controller that implements the scale-out and distributed file system using internal flash-based SSD drives to store data. The controllers communicate with each other over 10 or 40 Gigabit Ethernet to present a single pool of storage that spans the nodes in the cluster (Figure 2). Nodes access data through a data layer using file, block, object, and API plug-ins. As nodes are added, the cluster scales linearly to deliver computing, storage capacity, and I/O performance.

Figure 2 Distributed Cisco HyperFlex System



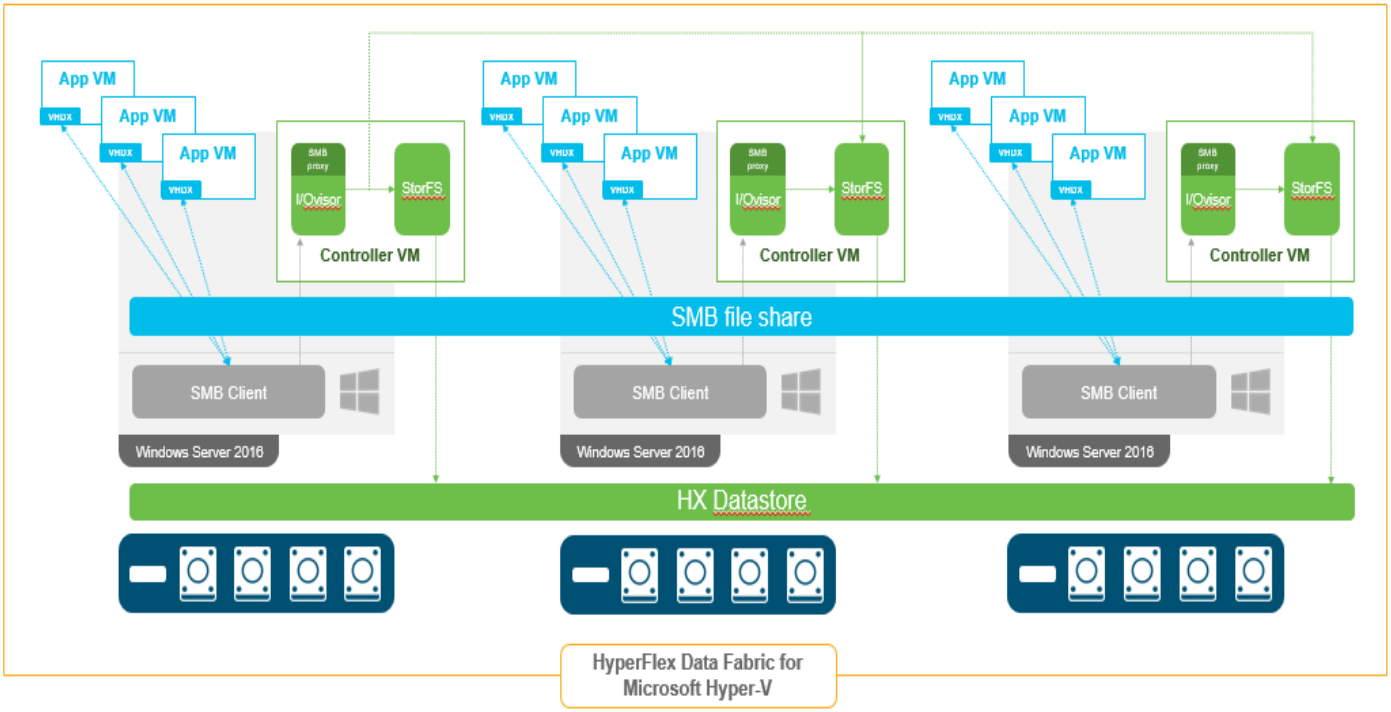
A Cisco HyperFlex Data Platform controller resides on each node and implements a distributed file system. The controller runs in user space within a virtual machine and intercepts and handles all I/O from guest virtual machines. Dedicated CPU cores and memory allow the controller to deliver consistent performance without affecting performance of the other virtual machines in the cluster. The data platform has modules to support the specific hypervisor or container platform in use. The controller accesses all of the node's disk storage through hypervisor bypass mechanisms (Discrete Device Assignment feature introduced in Windows Server 2016) for excellent performance. It uses the node's memory and dedicated SSD drives as part of a distributed caching layer, and it uses the node's other SSD drives, for distributed storage. The data platform controller interfaces with the hypervisor in two ways:

- IOVisor: The data platform controller intercepts all I/O requests and routes them to the nodes responsible for storing or retrieving the blocks. The IOVisor makes the existence of the hyperconvergence layer transparent to the hypervisor.

- Hypervisor agent: A module uses the hypervisor APIs to support advanced storage system operations such as snapshots and cloning. These are accessed through the hypervisor so that the hyperconvergence layer appears just as if it were enterprise-shared storage. The controller accelerates operations by manipulating metadata rather than actual data copying, providing rapid response, and thus rapid deployment of new application environments.

Figure 3 illustrates the storage controller VM architecture with IO path for HyperFlex with Hyper-V.

Figure 3 Storage controller VM architecture for HyperFlex on Hyper-V



Physical Infrastructure

Cisco Unified Computing System

Cisco Unified Computing System is a next-generation data center platform that unites compute, network and storage access. The platform, optimized for virtual environments, is designed using open industry-standard technologies and aims to reduce the total cost of ownership (TCO) and increase the business agility. The system integrates a low-latency; lossless 10 or 40 Gigabit Ethernet unified network fabric with enterprise-class, x86-architecture servers. All resources participate in a unified management domain in an integrated, scalable, multi-chassis platform.

Cisco Unified Computing System consists of the following components:

- Compute - The system is based on an entirely new class of computing system that incorporates rack mount and blade servers based on Intel® Xeon® scalable processors product family.

- Network - The system is integrated onto a low-latency, lossless, 40-Gbps unified network fabric. This network foundation consolidates Local Area Networks (LAN's), Storage Area Networks (SANs), and high-performance computing networks that are separate networks today. The unified fabric lowers costs by reducing the number of network adapters, switches, and cables, and by decreasing the power and cooling requirements.

- Virtualization - The system unleashes the full potential of virtualization by enhancing the scalability, performance, and operational control of virtual environments. Cisco security, policy enforcement, and diagnostic features are now extended into virtualized environments to better support changing business and IT requirements.

- Storage access - The system provides consolidated access to both SAN storage and Network Attached Storage (NAS) over the unified fabric. It is also an ideal system for Software defined Storage (SDS). Combining the benefits of single framework to manage both the compute and Storage servers in a single pane, Quality of Service (QOS) can be implemented if needed to inject IO throttling in the system. In addition, the server administrators can pre-assign storage-access policies to storage resources, for simplified storage connectivity and management leading to increased productivity. In addition to external storage, both rack and blade servers have internal storage that can be accessed through built-in hardware RAID controllers. With storage profile and disk configuration policy configured in Cisco UCS Manager, storage needs for the host OS and application data is fulfilled by user defined RAID groups for high availability and better performance.

- Management - the system uniquely integrates all system components to enable the entire solution to be managed as a single entity by the Cisco UCS Manager. The Cisco UCS Manager has an intuitive graphical user interface (GUI), a command-line interface (CLI), and a powerful scripting library module for Microsoft PowerShell built on a robust application programming interface (API) to manage all system configuration and operations.

Cisco Unified Computing System is designed to deliver:

- A reduced Total Cost of Ownership and increased business agility.

- Increased IT staff productivity through just-in-time provisioning and mobility support.

- A cohesive, integrated system that unifies the technology in the data center. The system is managed, services and tested as a whole.

- Scalability through a design for hundreds of discrete servers and thousands of virtual machines and the capability to scale I/O bandwidth to match the demand.

- Industry standard supported by a partner ecosystem of industry leaders.

Cisco UCS Fabric Interconnect

The Cisco UCS Fabric Interconnect (FI) is a core part of the Cisco Unified Computing System, providing both network connectivity and management capabilities for the system. Depending on the model chosen, the Cisco UCS Fabric Interconnect offers line-rate, low-latency, lossless 10 Gigabit or 40 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE) and Fibre Channel connectivity. Cisco UCS Fabric Interconnects provide the management and communication backbone for the Cisco UCS C-Series, S-Series and HX-Series Rack- Mount Servers , Cisco UCS B-Series Blade Servers and Cisco UCS 5100 Series Blade Server Chassis. All servers and chassis, and therefore all blades, attached to the Cisco UCS Fabric Interconnects become part of a single, highly available management domain. In addition, by supporting unified fabrics, the Cisco UCS Fabric Interconnects provide both the LAN and SAN connectivity for all servers within its domain.

From a networking perspective, the Cisco UCS 6200 Series uses a cut-through architecture, supporting deterministic, low latency, line rate 10 Gigabit Ethernet on all ports, up to 1.92 Tbps switching capacity and 160 Gbps bandwidth per chassis, independent of packet size and enabled services. The product family supports Cisco low - latency, lossless 10 Gigabit Ethernet unified network fabric capabilities, which increase the reliability, efficiency, and scalability of Ethernet networks. The Fabric Interconnect supports multiple traffic classes over the Ethernet fabric from the servers to the uplinks. Significant TCO savings come from an FCoE-optimized server design in which network interface cards (NICs), host bus adapters (HBAs), cables, and switches can be consolidated.

The Cisco UCS 6300 Series offers the same features while supporting even higher performance, low latency, lossless, line rate 40 Gigabit Ethernet, with up to 2.56 Tbps of switching capacity . Backward compatibility and scalability are assured with the ability to configure 40 Gbps quad SFP (QSFP) ports as breakout ports using 4x10GbE breakout cables. Existing

Cisco UCS servers with 10GbE interfaces can be connected in this manner, although Cisco HyperFlex nodes must use a 40GbE VIC adapter in order to connect to a Cisco UCS 6300 Series Fabric Interconnect.

Listed below are the Cisco UCS Fabric Interconnects supported by the HyperFlex System. For detailed information about these FIs, see Cisco UCS Fabric Interconnects and Fabric Extenders.

- Cisco UCS 6248UP
- Cisco UCS 6296UP
- Cisco UCS 6332
- Cisco UCS 6332-16UP

Cisco HyperFlex HX-Series Nodes

A HyperFlex cluster requires a minimum of three HX-Series "converged" nodes (with disk storage). Data is replicated across at least two of these nodes, and a third node is required for continuous operation in the event of a single-node failure. Each node that has disk storage is equipped with at least one high-performance SSD drive for data caching and rapid acknowledgment of write requests. Each node also is equipped with additional disks, up to the platform's physical limit, for long-term storage and capacity.

The list below provides the supported HX-Series All Flash converged nodes. For detailed information about the following nodes, see Cisco HyperFlex Models.

- Cisco HyperFlex HXAF220c-M5SX All-Flash Node
- Cisco HyperFlex HXAF240c-M5SX All-Flash Node

Cisco VIC 1227 and 1387 MLOM Interface Cards

The Cisco UCS Virtual Interface Card (VIC) 1227 is a dual-port Enhanced Small Form-Factor Pluggable (SFP+) 10-Gbps Ethernet and Fibre Channel over Ethernet (FCoE)-capable PCI Express (PCIe) modular LAN-on-motherboard (mLOM) adapter installed in the Cisco UCS HX-Series Rack Servers. The VIC 1227 is used in conjunction with the Cisco UCS 6248UP or 6296UP model Fabric Interconnects.

The Cisco UCS VIC 1387 Card is a dual-port Enhanced Quad Small Form-Factor Pluggable (QSFP+) 40-Gbps Ethernet and Fibre Channel over Ethernet (FCoE)-capable PCI Express (PCIe) modular LAN-on-motherboard (mLOM) adapter installed in the Cisco UCS HX-Series Rack Servers. The VIC 1387 is used in conjunction with the Cisco UCS 6332 or 6332-16UP model Fabric Interconnects.

The mLOM slot can be used to install a Cisco VIC without consuming a PCIe slot, which provides greater I/O expandability. It incorporates next-generation converged network adapter (CNA) technology from Cisco, providing investment protection for future feature releases. The card enables a policy-based, stateless, agile server infrastructure that can present up to 256 PCIe standards-compliant interfaces to the host, each dynamically configured as either a network interface card (NICs) or host bus adapter (HBA). The personality of the interfaces is set programmatically using the service profile associated with the server. The number, type (NIC or HBA), identity (MAC address and World Wide Name [WWN]), failover policy, adapter settings, bandwidth, and quality-of-service (QoS) policies of the PCIe interfaces are all specified using the service profile.

Hardware revision V03 or later of the Cisco VIC 1387 is required for the Cisco HyperFlex HX-series servers.

Cisco HyperFlex Systems Details

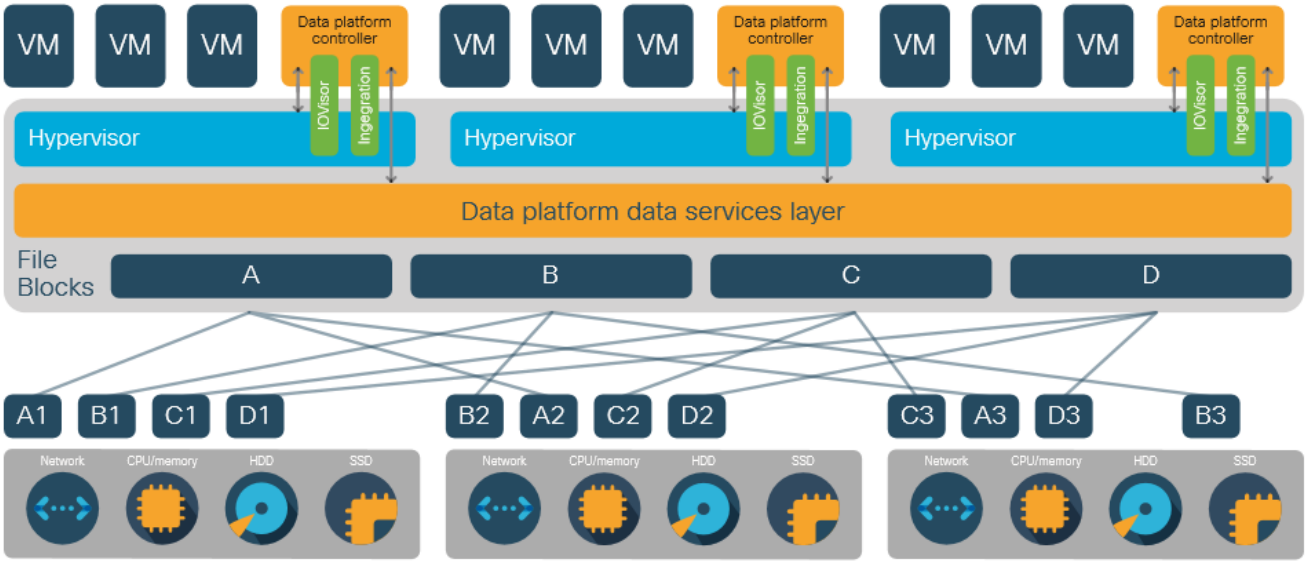
Engineered on the successful Cisco UCS platform, Cisco HyperFlex Systems deliver a hyperconverged solution that truly integrates all components in the data center infrastructure—compute, storage, and networking. The HX Data Platform starts with three or more nodes to form a highly available cluster. Each of these nodes has a software controller called the Cisco HyperFlex Controller. It takes control of the internal flash-based SSDs or a combination of flash-based SSDs and HDDs to store persistent data into a single distributed, multitier, object-based data store. The controllers communicate with each other over low-latency 10 or 40 Gigabit Ethernet fabric, to present a single pool of storage that spans across all the nodes in the cluster so that data availability is not affected if single or multiple components fail.

Data Distribution

The HX Data Platform controller handles all read and write requests for volumes that the hypervisor accesses and thus intermediates all I/O from the virtual machines and containers. Recognizing the importance of data distribution, the HX Data Platform is designed to exploit low network latencies and parallelism, in contrast to other approaches that build on node-local affinity and can easily cause data hotspots.

With data distribution, the data platform stripes data evenly across all nodes, with the number of data replicas determined by the policies you set (Figure 3). This approach helps prevent both network and storage hot spots and makes I/O performance the same regardless of virtual machine location. This feature gives you more flexibility in workload placement and contrasts with other architectures in which a data locality approach does not fully utilize all available networking and I/O resources.

Figure 4 Data Blocks are Replicated Across the Cluster



- **Data write operations:** For write operations, data is written to the local SSD or NVMe cache, and the replicas are written to remote caches in parallel before the write operation is acknowledged. Writes are later synchronously flushed to the capacity layer HDDs (for hybrid nodes) or SSD drives (for all-flash nodes) or NVMe storage (for NVMe nodes).

- **Data read operations:** For read operations in all-flash nodes, local and remote data is read directly from storage in the distributed capacity layer. For read operations in hybrid configurations, data that is local usually is read directly from the cache. This process allows the platform to use all solid-state storage for read operations, reducing bottlenecks and delivering excellent performance. In addition, when migrating a virtual machine to a new location, the data platform does not require data movement because any virtual machine can read its data from any location. Thus, moving virtual machines has no performance impact or cost.

In addition, when migrating a virtual machine to a new location, the data platform does not require data movement because any virtual machine can read its data from any location. Thus, moving virtual machines has no performance impact or cost.

Data Operations

The data platform implements a distributed, log-structured file system that changes how it handles caching and storage capacity depending on the node configuration.

In the All-Flash-memory configuration, the data platform uses a caching layer in SSDs to accelerate write responses, and it implements the capacity layer in SSDs. Read requests are fulfilled directly from data obtained from the SSDs in the capacity layer. A dedicated read cache is not required to accelerate read operations.

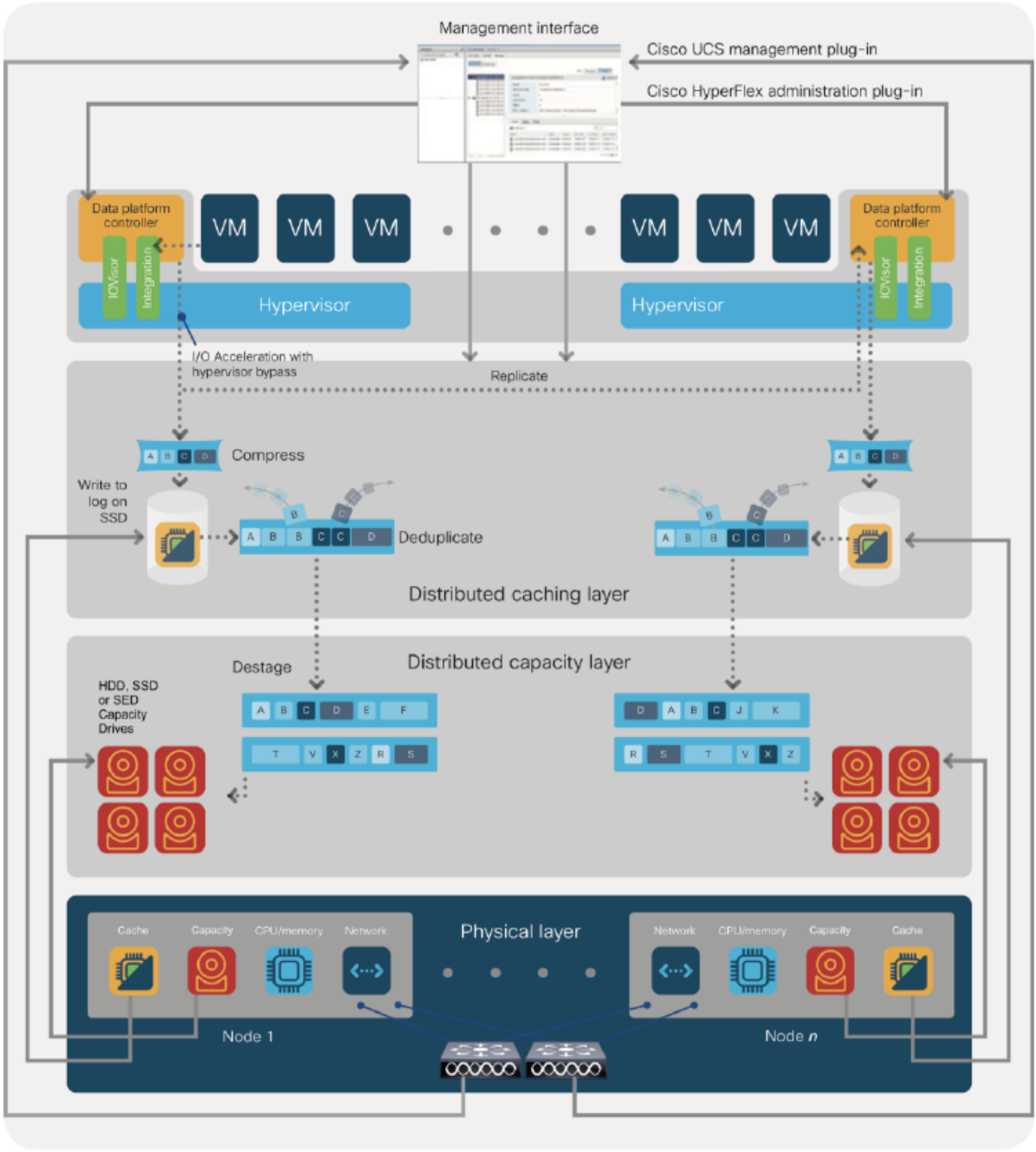
Incoming data is striped across the number of nodes required to satisfy availability requirements—usually two or three nodes. Based on policies you set, incoming write operations are acknowledged as persistent after they are replicated to the SSD drives in other nodes in the cluster. This approach reduces the likelihood of data loss due to SSD or node failures. The write operations are then de-staged to SSDs in the capacity layer in the All-Flash memory configuration for long-term storage.

The log-structured file system writes sequentially to one of two write logs (three in case of RF=3) until it is full. It then switches to the other write log while de-staging data from the first to the capacity tier. When existing data is (logically) overwritten, the log-structured approach simply appends a new block and updates the metadata. This layout benefits SSD configurations in which seek operations are not time consuming. It reduces the write amplification levels of SSDs and the total number of writes the flash media experiences due to incoming writes and random overwrite operations of the data.

When data is de-staged to the capacity tier in each node, the data is deduplicated and compressed. This process occurs after the write operation is acknowledged, so no performance penalty is incurred for these operations. A small deduplication block size helps increase the deduplication rate. Compression further reduces the data footprint. Data is then moved to the capacity tier as write cache segments are released for reuse (Figure 5).

Read operations in hybrid nodes cache data on the SSD drives and in main memory for high performance. In all-flash and NVMe nodes, they read directly from storage. Having the most frequently used data stored in the caching layer helps make Cisco HyperFlex Systems perform well for virtualized applications. When virtual machines modify data, the original block is likely read from the cache, so there is often no need to read and then expand the data on a spinning disk. The data platform decouples the caching tier from the capacity tier and allows independent scaling of I/O performance and storage capacity.

Figure 5 Data Write Operation Flow Through the Cisco HyperFlex HX Data Platform



Data Optimization

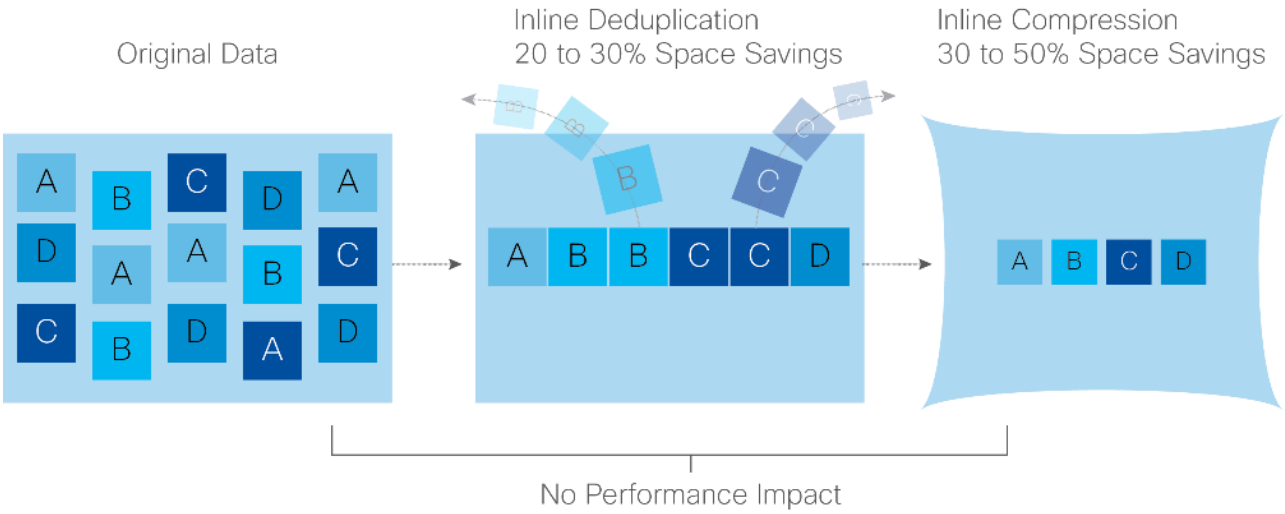
The Cisco HyperFlex HX Data Platform provides finely detailed inline deduplication and variable block inline compression that is always on for objects in the cache (SSD and memory) and capacity (SSD or HDD) layers. Unlike other solutions,

which require you to turn off these features to maintain performance, the deduplication and compression capabilities in the Cisco data platform are designed to sustain and enhance performance and significantly reduce physical storage capacity requirements.

Data Deduplication

Data deduplication is used on all storage in the cluster, including memory and SSD drives. Based on a patent-pending Top-K Majority algorithm, the platform uses conclusions from empirical research that show that most data, when sliced into small data blocks, has significant deduplication potential based on a minority of the data blocks. By fingerprinting and indexing just these frequently used blocks, high rates of deduplication can be achieved with only a small amount of memory, which is a high-value resource in cluster nodes (Figure 6).

Figure 6 Cisco HyperFlex HX Data Platform Optimizes Data Storage with No Performance Impact



Inline Compression

The Cisco HyperFlex HX Data Platform uses high-performance inline compression on data sets to save storage capacity. Although other products offer compression capabilities, many negatively affect performance. In contrast, the Cisco data platform uses CPU-offload instructions to reduce the performance impact of compression operations. In addition, the log-structured distributed-objects layer has no effect on modifications (write operations) to previously compressed data. Instead, incoming modifications are compressed and written to a new location, and the existing (old) data is marked for deletion, unless the data needs to be retained in a snapshot.

The data that is being modified does not need to be read prior to the write operation. This feature avoids typical read-modify-write penalties and significantly improves write performance.

Log-Structured Distributed Objects

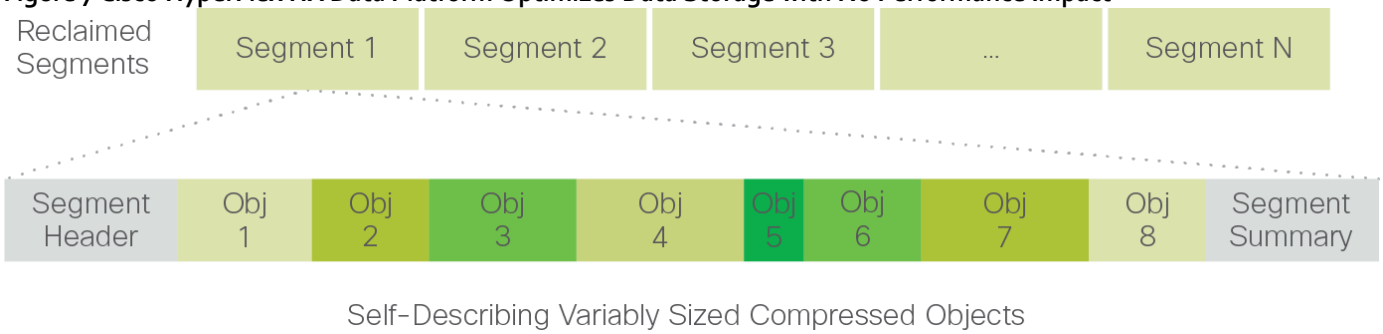
In the Cisco HyperFlex HX Data Platform, the log-structured distributed-object store layer groups and compresses data that filters through the deduplication engine into self-addressable objects. These objects are written to disk in a log-structured, sequential manner. All incoming I/O—including random I/O—is written sequentially to both the caching (SSD and memory) and persistent (SSD or HDD) tiers. The objects are distributed across all nodes in the cluster to make uniform use of storage capacity.

By using a sequential layout, the platform helps increase flash-memory endurance. Because read-modify-write operations are not used, there is little or no performance impact of compression, snapshot operations, and cloning on overall performance.

Data blocks are compressed into objects and sequentially laid out in fixed-size segments, which in turn are sequentially laid out in a log-structured manner (Figure 7). Each compressed object in the log-structured segment is uniquely addressable

using a key, with each key fingerprinted and stored with a checksum to provide high levels of data integrity. In addition, the chronological writing of objects helps the platform quickly recover from media or node failures by rewriting only the data that came into the system after it was truncated due to a failure.

Figure 7 Cisco HyperFlex HX Data Platform Optimizes Data Storage with No Performance Impact



Encryption

Securely encrypted storage optionally encrypts both the caching and persistent layers of the data platform. Integrated with enterprise key management software, or with passphrase-protected keys, encrypting data at rest helps you comply with HIPAA, PCI-DSS, FISMA, and SOX regulations. The platform itself is hardened to Federal Information Processing Standard (FIPS) 140-1 and the encrypted drives with key management comply with the FIPS 140-2 standard.

Data Services

The Cisco HyperFlex HX Data Platform provides a scalable implementation of space-efficient data services, including thin provisioning, space reclamation, pointer-based snapshots, and clones—without affecting performance.

Thin Provisioning

The platform makes efficient use of storage by eliminating the need to forecast, purchase, and install disk capacity that may remain unused for a long time. Virtual data containers can present any amount of logical space to applications, whereas the amount of physical storage space that is needed is determined by the data that is written. You can expand storage on existing nodes and expand your cluster by adding more storage-intensive nodes as your business requirements dictate, eliminating the need to purchase large amounts of storage before you need it.

Fast, Space-Efficient Clones

In the Cisco HyperFlex HX Data Platform, clones are writable snapshots that can be used to rapidly provision items such as virtual desktops and applications for test and development environments. These fast, space-efficient clones rapidly replicate storage volumes so that virtual machines can be replicated through just metadata operations, with actual data copying performed only for write operations. With this approach, hundreds of clones can be created and deleted in minutes. Compared to full-copy methods, this approach can save a significant amount of time, increase IT agility, and improve IT productivity.

Clones are deduplicated when they are created. When clones start diverging from one another, data that is common between them is shared, with only unique data occupying new storage space. The deduplication engine eliminates data duplicates in the diverged clones to reduce the clone's storage footprint.

Data Replication and Availability

In the Cisco HyperFlex HX Data Platform, the log-structured distributed-object layer replicates incoming data, improving data availability. Based on policies that you set, data that is written to the write cache is synchronously replicated to one or two other SSD drives located in different nodes before the write operation is acknowledged to the application. This approach allows incoming writes to be acknowledged quickly while protecting data from SSD or node failures. If an SSD or node fails, the replica is quickly re-created on other SSD drives or nodes using the available copies of the data.

The log-structured distributed-object layer also replicates data that is moved from the write cache to the capacity layer. This replicated data is likewise protected from SSD or node failures. With two replicas, or three data copies, the cluster can survive uncorrelated failures of two SSD drives or two nodes without the risk of data loss. Uncorrelated failures are failures that occur on different physical nodes. Failures that occur on the same node affect the same copy of data and are treated as a single failure. For example, if one disk in a node fails and subsequently another disk on the same node fails, these correlated failures count as one failure in the system. In this case, the cluster could withstand another uncorrelated failure on a different node. See the Cisco HyperFlex HX Data Platform system administrator's guide for a complete list of fault-tolerant configurations and settings.

If a problem occurs in the Cisco HyperFlex HX controller software, data requests from the applications residing in that node are automatically routed to other controllers in the cluster. This same capability can be used to upgrade or perform maintenance on the controller software on a rolling basis without affecting the availability of the cluster or data. This self-healing capability is one of the reasons that the Cisco HyperFlex HX Data Platform is well suited for production applications.

Data Rebalancing

A distributed file system requires a robust data rebalancing capability. In the Cisco HyperFlex HX Data Platform, no overhead is associated with metadata access, and rebalancing is extremely efficient. Rebalancing is a non-disruptive online process that occurs in both the caching and persistent layers, and data is moved at a fine level of specificity to improve the use of storage capacity. The platform automatically rebalances existing data when nodes and drives are added or removed or when they fail. When a new node is added to the cluster, its capacity and performance is made available to new and existing data. The rebalancing engine distributes existing data to the new node and helps ensure that all nodes in the cluster are used uniformly from capacity and performance perspectives. If a node fails or is removed from the cluster, the rebalancing engine rebuilds and distributes copies of the data from the failed or removed node to available nodes in the clusters.

Online Upgrades

Cisco HyperFlex HX-Series systems and the HX Data Platform support online upgrades so that you can expand and update your environment without business disruption. You can easily expand your physical resources; add processing capacity; and download and install BIOS, driver, hypervisor, firmware, and Cisco UCS Manager updates, enhancements, and bug fixes.

Why to use HyperFlex All-Flash systems for Database Deployments

SQL server database systems act as the backend to many critical and performance hungry applications. It is very important to ensure that it delivers consistent performance with predictable latency throughout. Below are some of the major advantages of Cisco HyperFlex All-Flash hyperconverged systems that makes it ideally suited for SQL Server database implementations.

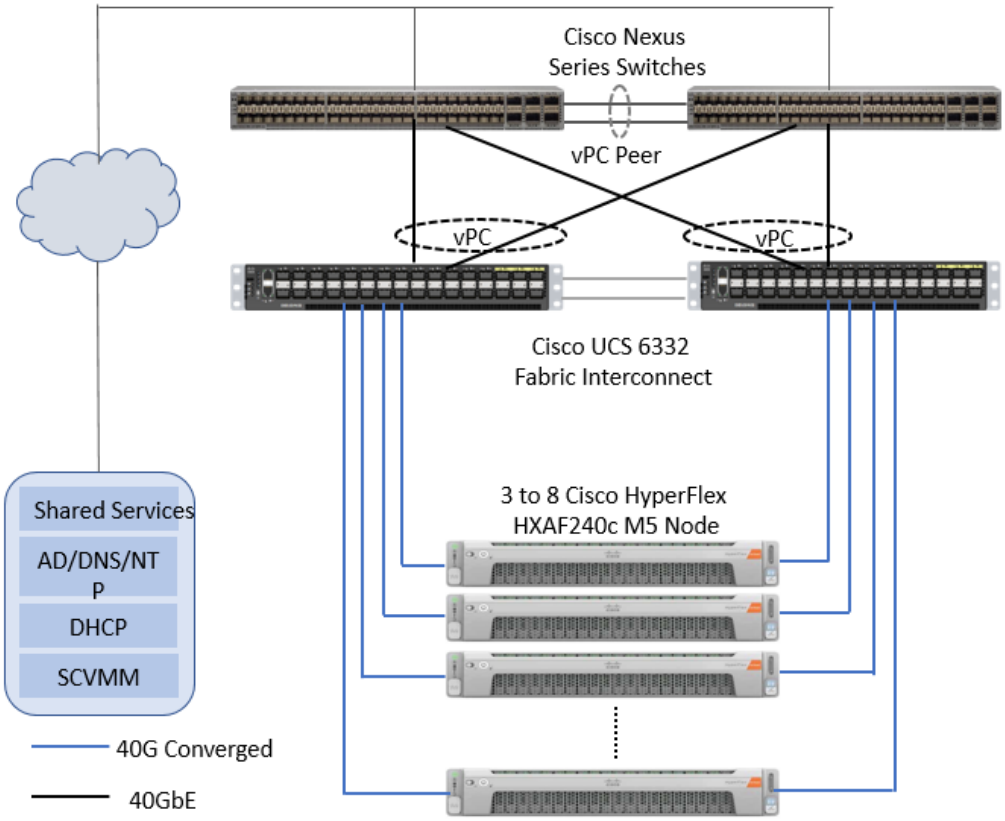
- **Low latency with consistent performance**: Cisco HyperFlex All-Flash nodes provides excellent platform for critical database deployment by offering low latency, consistent performance and exceeds most of the database service level agreements.
- **Data protection** (fast clones and snapshots, replication factor, VM replication): The HyperFlex systems are engineered with robust data protection techniques that enable quick backup and recovery of the applications in case of any failures.
- **Storage optimization**: All the data that comes in the HyperFlex systems are by default optimized using inline deduplication and data compression techniques. Additionally, the HX Data Platform's log-structured file system ensures data blocks are written to flash devices in a sequential manner thereby increasing flash-memory endurance. HX System makes efficient use of flash storage by using Thin Provisioning storage optimization technique.
- **Performance and Capacity Online Scalability**: The flexible and independent scalability of the capacity and compute tiers of HyperFlex systems provide immense opportunities to adapt to the growing performance demands without any application disruption.

- **No Performance Hotspots**: The distributed architecture of HyperFlex Data Platform makes sure that every VM is able to leverage the storage IOPS and capacity of the entire cluster, irrespective of the physical node in which it is residing. This is especially important for SQL Server VMs as they frequently need higher performance to handle bursts of application or user activity.
- **Non-disruptive System maintenance**: Cisco HyperFlex Systems enables distributed computing and storage environment that enables the administrators to perform system maintenance tasks without disruption.

Solution Design

This section details the architectural components of a HyperFlex with Hyper-V to host Microsoft SQL Server databases in virtual environment. Figure 8 depicts a sample Cisco HyperFlex hyperconverged reference architecture comprising HX-Series rack mount servers.

Figure 8 Cisco HyperFlex Reference Architecture using HXAF240c Nodes



Cisco HyperFlex with Hyper-V is composed of a pair of Cisco UCS Fabric Interconnects along with up to eight HX-Series rack-mount server per cluster. Up to eight separate HX clusters can be installed under a single pair of Fabric Interconnects. The two Fabric Interconnects both connect to every HX-Series rack-mount server and both connect to every Cisco UCS 5108 blade chassis, and Cisco UCS rack-mount server. Upstream network connections, also referred as "north bound" network, are made from the Fabric Interconnects to the customer datacenter network at the time of installation. For more details about physical connectivity of HX-Series services, compute-only servers, Fabric Interconnect to the northbound network, please refer VSI CVD here.

Infrastructure services such as Active Directory, DNS, DHCP and SCVMM are typically installed outside the HyperFlex cluster. Customers can leverage these existing services deploying and managing the HyperFlex Hyper-V cluster.

The HyperFlex storage solution has several data protection techniques, as explained in detail in the Technology overview section, one of which is data replication that needs to be configured on HyperFlex cluster creation. Based on the specific performance and data protection requirements, customer can choose either a replication factor of two (RF2) or three (RF3). For the solution validation (described in the "Solution Testing and Validation" later in this document), we had configured the test HyperFlex cluster to be of replication factor 3 (RF3).

As described in section Technology Overview, Cisco HyperFlex distributed file system software runs inside a controller VM that is installed on each cluster node. These controller VMs pool and manage all the storage devices and exposes the

underlying storage as SMB share to the Hyper-V nodes. The Hyper-V exposes these SMB shares as datastores to the guest virtual machines to store their data.

Logical Network design

In the Cisco HyperFlex All Flash system, Cisco VIC 1387 is used to provide the required logical network interfaces on each host in the cluster. The communication pathways in Cisco HyperFlex system can be categorized in to four different traffic zones as described below.

- **Management Zone:** This zone comprises the connections needed to manage the physical hardware, the hypervisor hosts, and the storage platform controller virtual machines (SCVM). These interfaces and IP addresses need to be available to all staff who will administer the HX system, throughout the LAN/WAN. This zone must provide access to Domain Name System (DNS) and Network Time Protocol (NTP) services, and allow Secure Shell (SSH) communication. In this zone are multiple physical and virtual components:
  - Fabric Interconnect management ports.
  - Cisco UCS external management interfaces used by the servers, which answer via the FI management ports.
  - Hyper-V host management interfaces.
  - Storage Controller VM management interfaces.
  - A roaming HX cluster management interface.
  - Storage Controller VM Management interfaces.
- **VM Zone:** This zone comprises the connections needed to service network IO to the guest VMs that will run inside the HyperFlex hyperconverged system. This zone typically contains multiple VLANs that are trunked to the Cisco UCS Fabric Interconnects via the network uplinks, and tagged with 802.1Q VLAN IDs. These interfaces and IP addresses need to be available to all staff and other computer endpoints which need to communicate with the guest VMs in the HX system, throughout the LAN/WAN.
- **Storage Zone:** This zone comprises the connections used by the Cisco HX Data Platform software, Hyper-V hosts, and the storage controller VMs to service the HX Distributed Data Filesystem. These interfaces and IP addresses need to be able to communicate with each other at all times for proper operation. During normal operation, this traffic all occurs within the Cisco UCS domain, however there are hardware failure scenarios where this traffic would need to traverse the network northbound of the Cisco UCS domain. For that reason, the VLAN used for HX storage traffic must be able to traverse the network uplinks from the Cisco UCS domain, reaching FI A from FI B, and vice-versa. This zone is primarily jumbo frame traffic therefore, jumbo frames must be enabled on the Cisco UCS uplinks. In this zone are multiple components:
  - A teamed interface is used for storage traffic on each Hyper-V host in the HX cluster.
  - Storage Controller VM storage interfaces.
  - A roaming HX cluster storage interface.
- **Live Migration Zone:** This zone comprises the connections used by the Hyper-V hosts to enable live migration of the guest VMs from host to host. During normal operation, this traffic all occurs within the Cisco UCS domain, however there are hardware failure scenarios where this traffic would need to traverse the network northbound of the Cisco UCS domain. For that reason, the VLAN used for HX live migration traffic must be able to traverse the network uplinks from the Cisco UCS domain, reaching FI A from FI B, and vice-versa.

By leveraging Cisco UCS vNIC templates, LAN connectivity policies and vNIC placement policies in service profile, eight vNICs are carved out from Cisco VIC 1387 on each HX-Series server for network traffic zones mentioned above. Every HX-

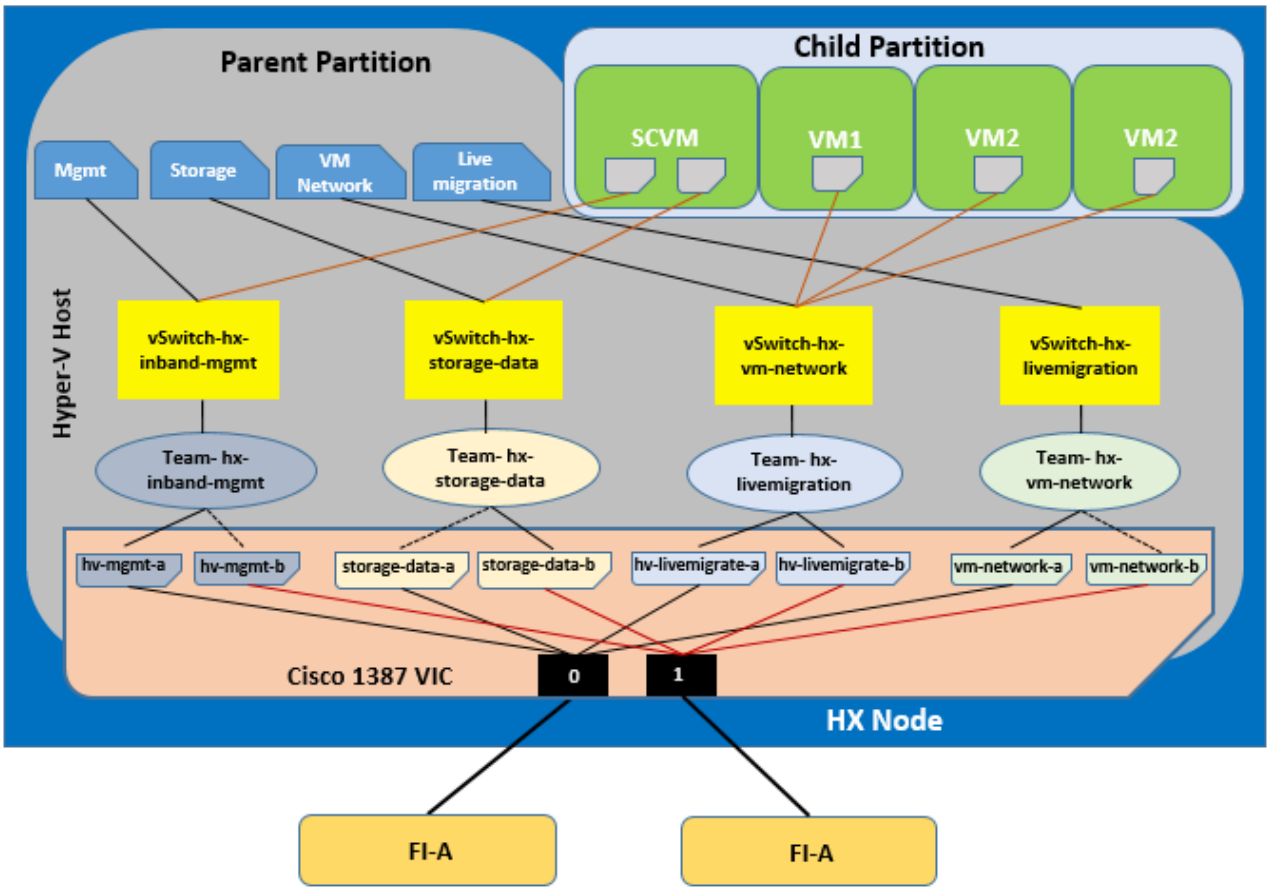
Series server will detect the network interfaces in the same order and they will always be connected to the same VLANs via the same network fabrics. Table 1 lists the vNICs and other configuration details being used in the solution.

Table 1 HX vNICs Template Details

vNIC Template Name:	**hv-mgmt-a**	**hv-mgmt-b**	**storage-data-a**	**storage-data-b**	**hv-livemigrate-a**	**hv-livemigrate-b**	**vm-network-a**	**vm-network-b**
Setting	**Value**	**Value**	**Value**	**Value**	**Value**	**Value**	**Value**	**Value**
Fabric ID	A	B	A	B	A	B	A	B
Fabric Failover	Disabled	Disabled	Disabled	Disabled	Disabled	Disabled	Disabled	Disabled
Target	Adapter	Adapter	Adapter	Adapter	Adapter	Adapter	Adapter	Adapter
Type	Updating Template	Updating Template	Updating Template	Updating Template	Updating Template	Updating Template	Updating Template	Updating Template
MTU	1500	1500	9000	9000	9000	9000	1500	1500
MAC Pool	hv-mgmt-a	hv-mgmt-b	storage-data-a	storage-data-b	hv-livemigrate-a	hv-livemigrate-b	vm-network-a	vm-network-b
QoS Policy	silver	silver	platinum	platinum	bronze	bronze	gold	gold
Network Control Policy	HyperFlex-infra	HyperFlex-infra	HyperFlex-infra	HyperFlex-infra	HyperFlex-infra	HyperFlex-infra	HyperFlex-vm	HyperFlex-vm
VLANs	<<hx-inband-mgmt>>	<<hx-inband-mgmt>>	<<hx-storage-data>>	<<hx-storage-data>>	<<hx-livemigrate>>	<<hx-livemigrate>>	<<vm-network>>	<<vm-network>>
Native VLAN	No	No	No	No	No	No	No	No

Figure 9 illustrates logical network design of a HX-Series server of HyperFlex cluster.

Figure 9 HX-Series Server Logical Network Diagram



The Cisco HyperFlex system has a pre-defined virtual network design at the Hyper-V hypervisor level. As shown in the figure above, four virtual switches are configured for four traffic zones. Each virtual switch is configured to use a teamed interface with two member adapters connected to both the Fabric Interconnects. The network adapters in the team for Storage, Management and Live Migration networks are configured in active and standby fashion. However, the network adapters in the team for VM Network are configured in active and active fashion. This makes sure that the data path for guest VMs traffic has the aggregated bandwidth for the specific traffic type.

Enabling jumbo frames on the Storage traffic zone benefits the following SQL Server database use case scenarios:

- Heavy write SQL Server guest VMs caused by the activities such as database restoring, rebuilding indexes and importing data, etc.
- Heavy read SQL Server guest VMs caused by the typical maintenance activities such as backup database, export data, report queries and rebuilding indexes, etc.

Enabling jumbo frames on Live Migration traffic zone help the system quickly failover the SQL VMs to other hosts there by reducing the overall database downtime.

Creating a separate logical network (using two dedicated vNICs) for guest VMs is beneficial with the following advantages:

- Isolating guest traffic from other traffic such as management, backup, etc.
- A dedicated MAC pool can be assigned to each vNIC that would simplify troubleshooting the connectivity issues.

- As shown in Figure 9, the VM Network switch is configured to use a teamed interface with two network adapters as members in active and active fashion to provide two active data paths that will result in aggregated bandwidth.

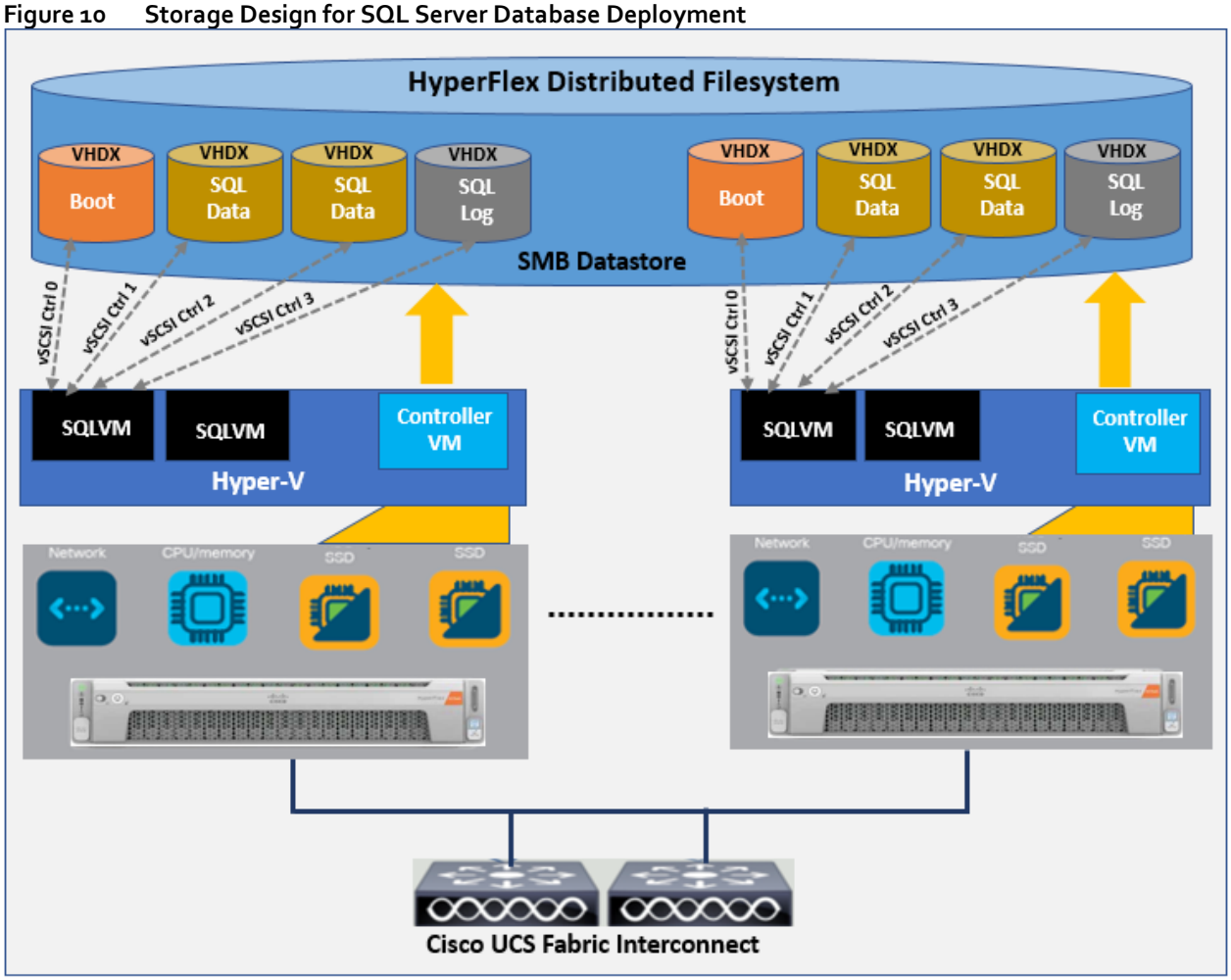
For more information about network configuration for the HyperFlex HX-Server node, using Cisco UCS network policies, templates and service profiles, refer to the HyperFlex Network Design guidelines section of the Cisco HyperFlex 3.0 for Virtual Server Infrastructure with Microsoft Hyper-V CVD.

The following sections provide more details on configuration and deployment best practices for deploying SQL server databases on HyperFlex All Flash nodes.

Storage Configuration for SQL Guest VMs

Figure 10 illustrates the storage configuration recommendations for virtual machines running SQL server databases on HyperFlex All Flash nodes. Separate SCSI controllers are configured to host OS, SQL server data and log volumes. For large scale and high performing SQL deployments, it is recommended to spread the SQL data files across two or more different SCSI controllers for better performance as shown in the following figure. Additional performance guidelines are explained in the "Deployment Planning" section.

Figure 10 Storage Design for SQL Server Database Deployment



Deployment Planning

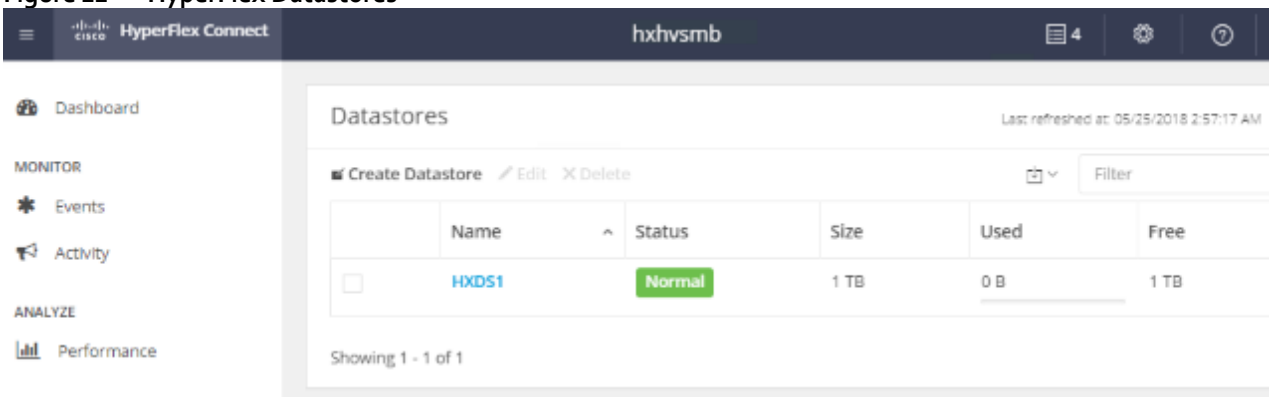
It is crucial to follow and implement configuration best practices and recommendations in order to achieve best performance from any underlying system. This section discusses the major design and configuration best practices that should be followed when deploying SQL Server databases on All Flash HyperFlex with Hyper-V systems.

Datastore Recommendation

The recommendations described in this section can be followed while deploying SQL Server virtual machines on HyperFlex with Hyper-V All-Flash Systems.

All the virtual machine's virtual disks comprising guest Operating System, swap file, SQL data and TempDB files and database log files can be placed on a single datastore exposed as SMB share to the Hyper-V nodes. Start deploying multiple SQL guest virtual machines using single datastore although HyperFlex supports creation of multiple data stores. Single datastore approach simplifies the management tasks and take advantage of HyperFlex's inline deduplication and compression.

Figure 11 HyperFlex Datastores

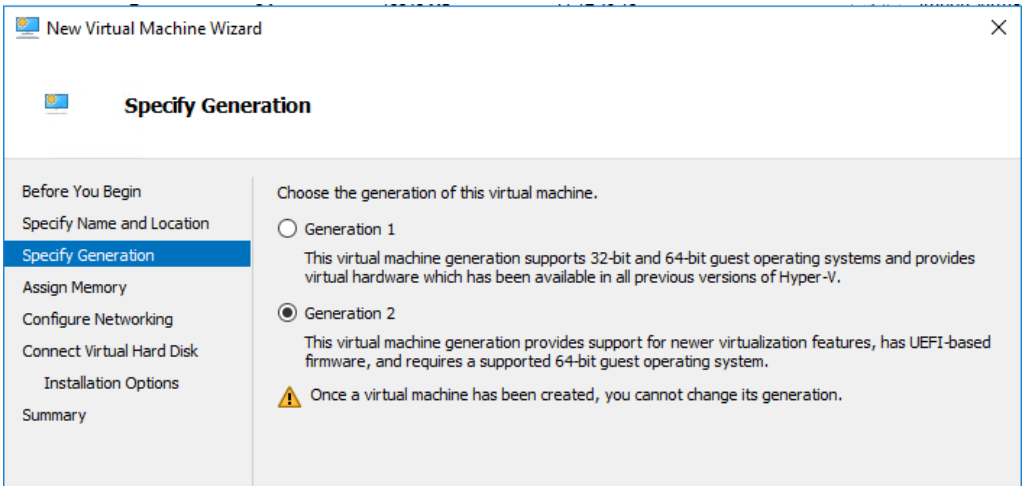


SQL Virtual Machine Configuration Recommendation

While creating a VM for deploying a SQL Server instance on a HyperFlex Hyper-V All-Flash system, the recommendations described in the following sections should be followed for performance and better administration.

VM Generation

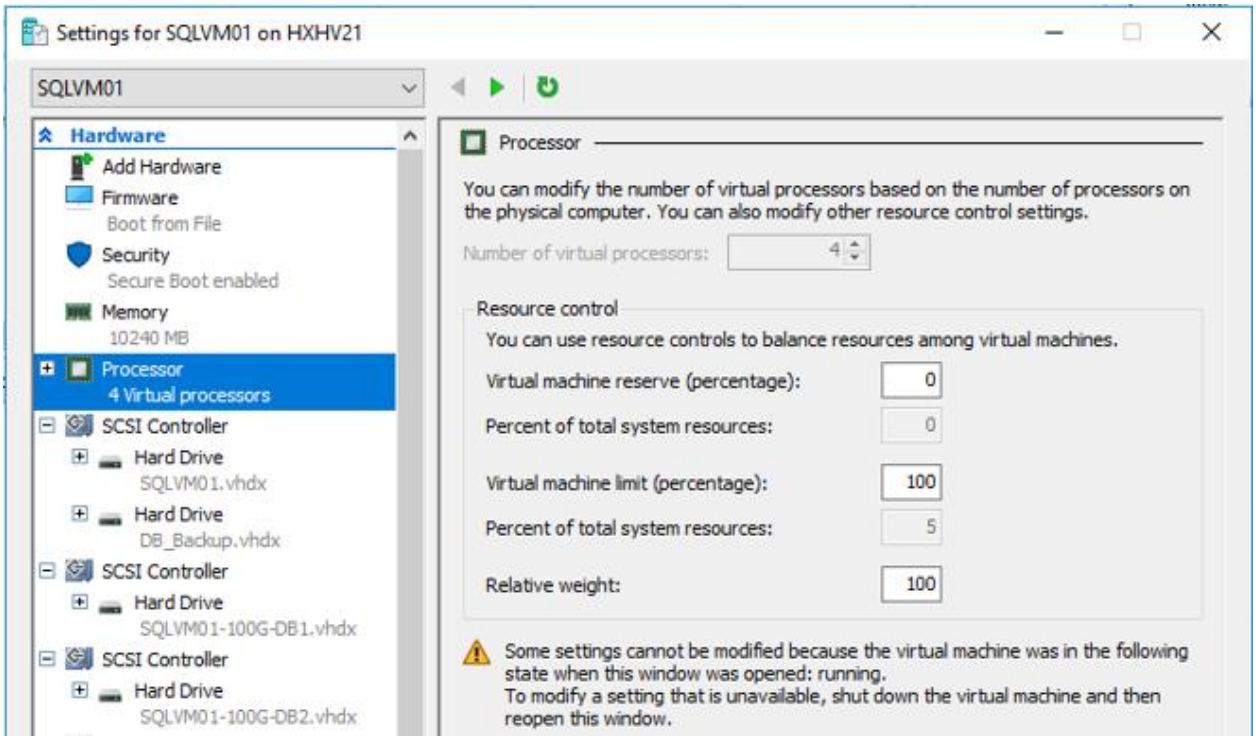
Create generation 2 VMs for running SQL Server as shown below:



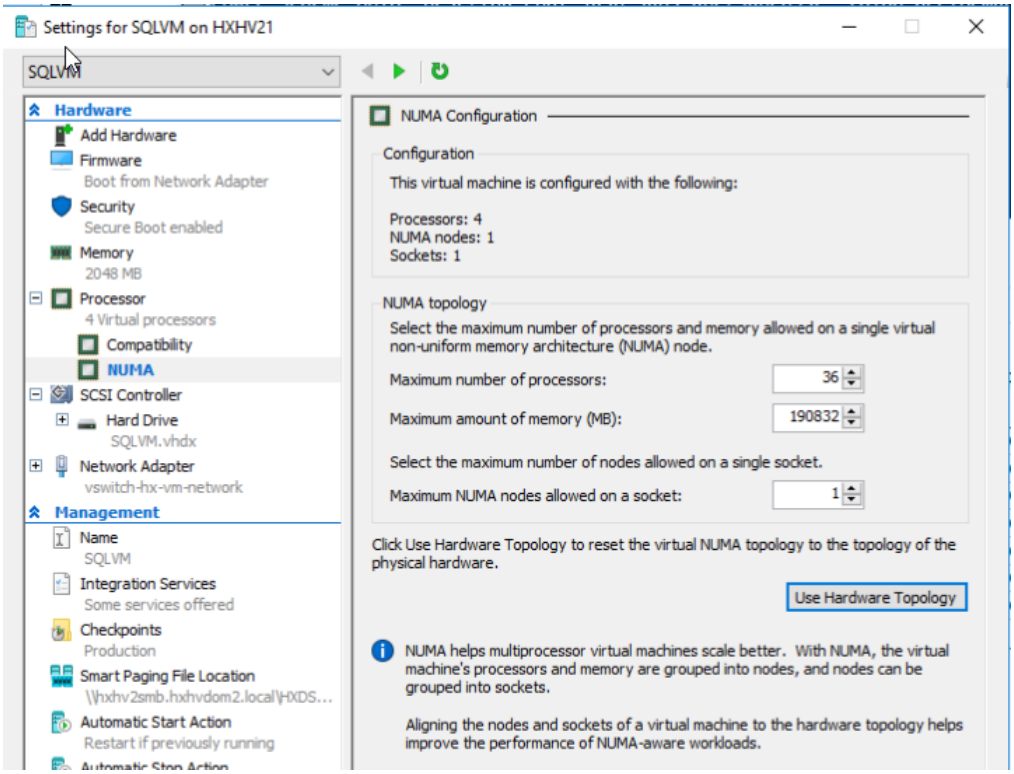
vCPU and NUMA Settings

Virtual machines with multiple virtual processors have additional overhead related to synchronization costs in guest operating systems. Multiple virtual processors should only be configured in cases where the virtual machine requires more processing power under peak loads. Set weights and reserves on the virtual processors based on the intensity of the loads the VMs bear. In this way, you can make sure that a large amount of the CPU cycle is available for virtual machines/virtual processors having high-intensity loads when there is CPU resource contention.

The weights and reserves for vCPU are left at defaults for this solution's testing and validation.



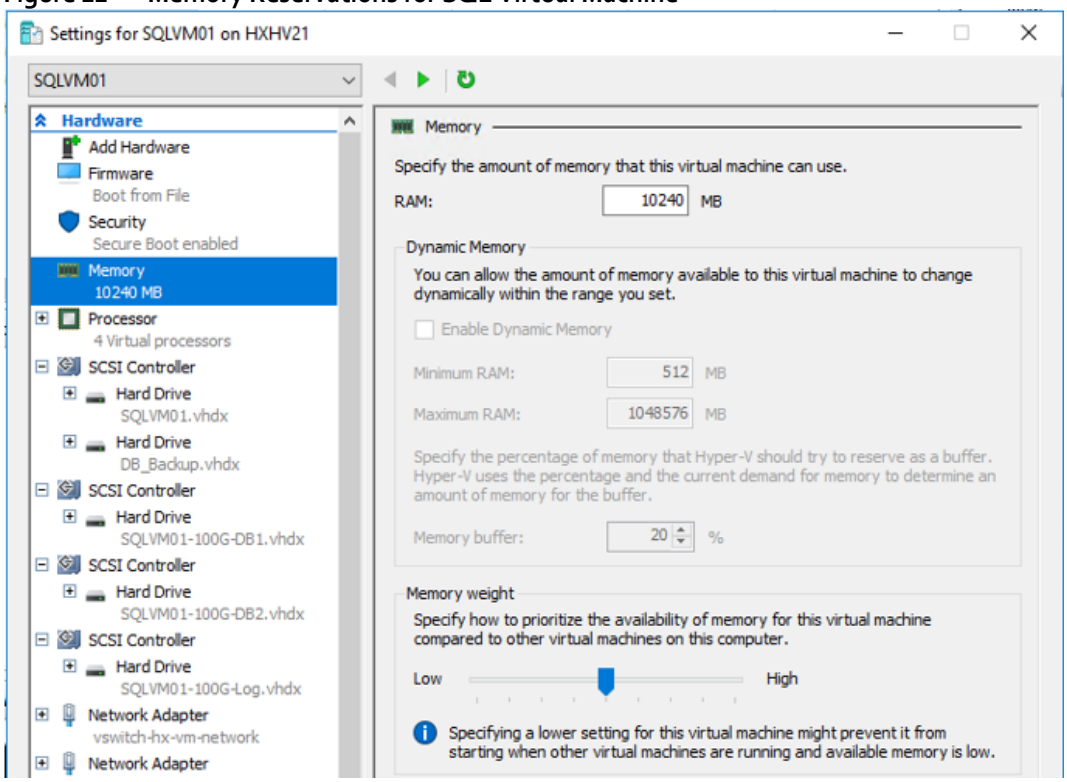
By default, the virtual NUMA topology maps to the topology of the underlying physical hardware. Defaults settings are recommended for better performance of NUMA-aware application like SQL Server. However, there are few scenarios (for example, licensing constraints, ensuring vNUMA is aligned with Physical NUMA) wherein you may want to change the vNUMA settings. Typically, it is not recommended to change this setting, unless the changes made to these settings are thoroughly tested on the given environment.



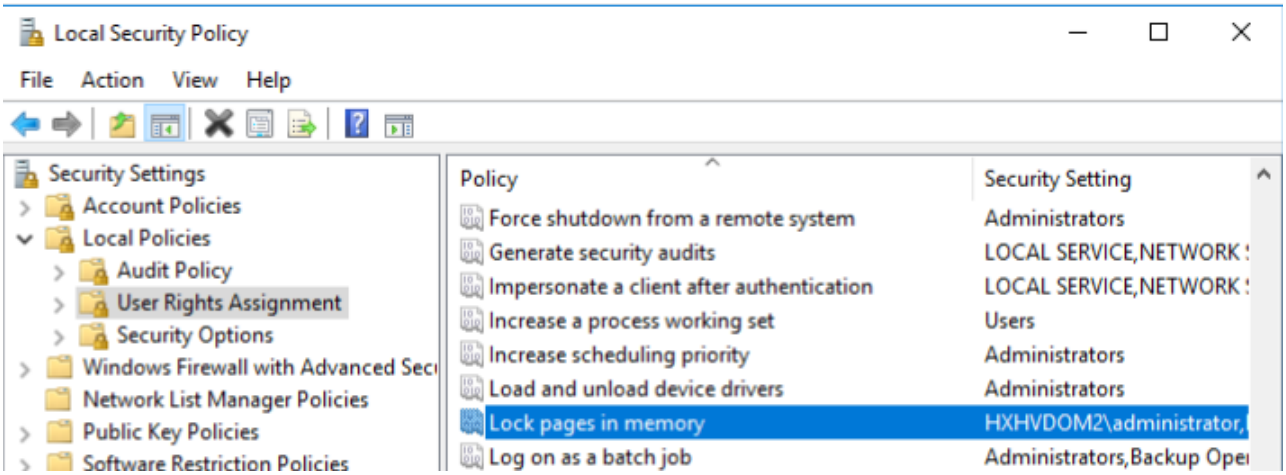
Memory Settings

SQL Server database transactions are usually CPU and memory intensive. In a heavy OLTP database system, it is recommended to assign static memory to the SQL Virtual Machines. This makes sure that the assigned memory to the SQL VM is committed and will eliminate the possibility of ballooning and paging the memory out by the hypervisor.

Figure 12 Memory Reservations for SQL Virtual Machine

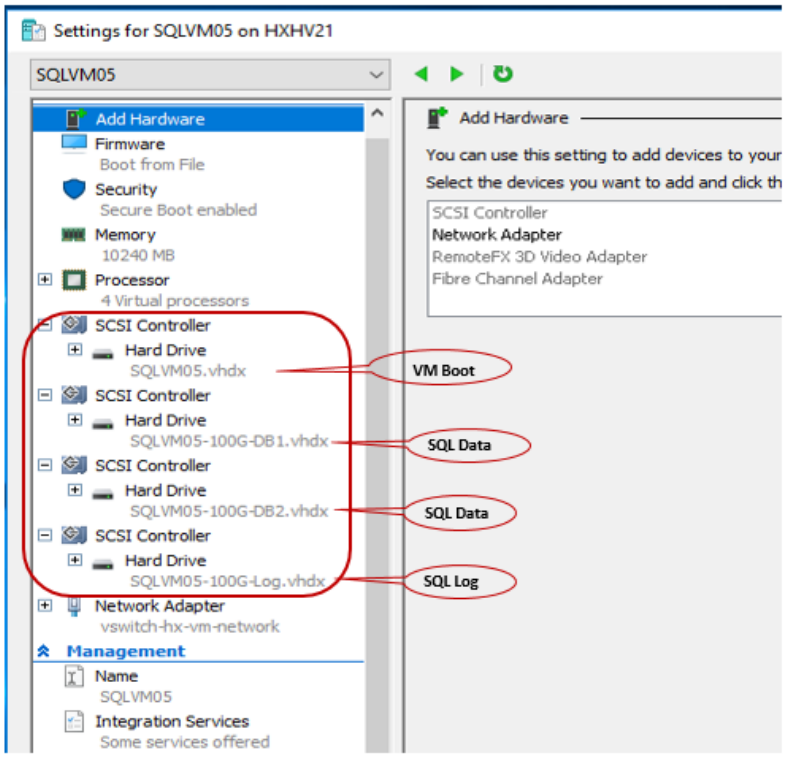


To provide better stability to a virtual machine workload, grant Lock Pages in Memory user rights to the SQL Server service account. This helps when Hyper-V Dynamic Memory is trying to reduce the virtual machine's memory. In such cases, it prevents Windows from paging out a large amount of buffer pool memory from the process, thereby providing a positive performance impact.



SCSI Controller Recommendations

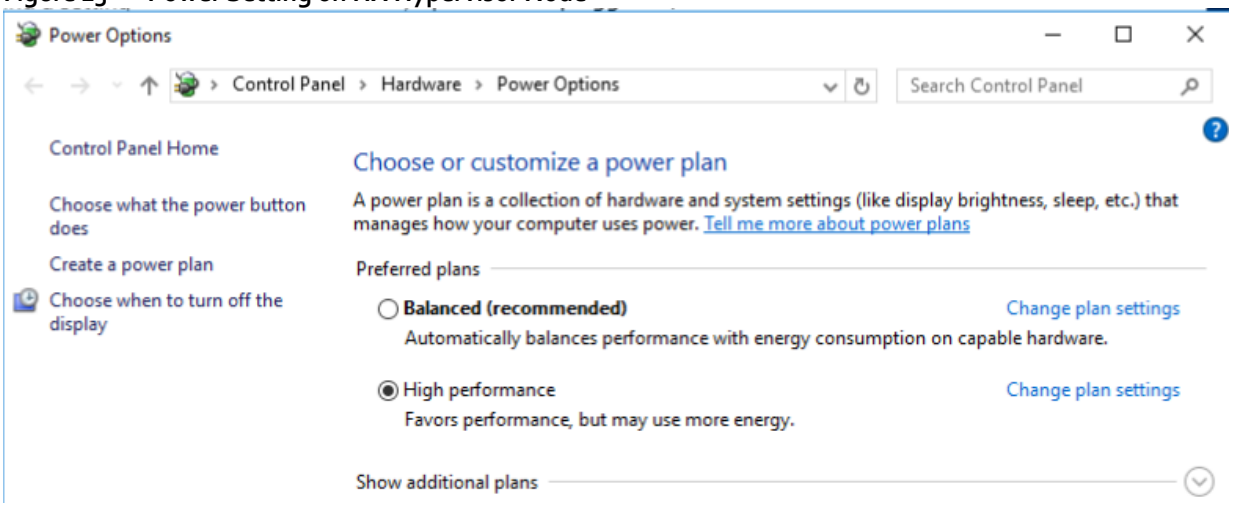
For large-scale and high IO databases, it is always recommended to use multiple virtual disks and have those virtual disks distributed across multiple SCSI controller adapters than assigning all of them to a single SCSI controller. This makes sure that the guest VM will access multiple virtual SCSI controllers (four SCSI controllers maximum per guest VM) and hence enabling greater concurrency, utilizing the multiple queues available for the SCSI controllers.



Guest Power Scheme Settings

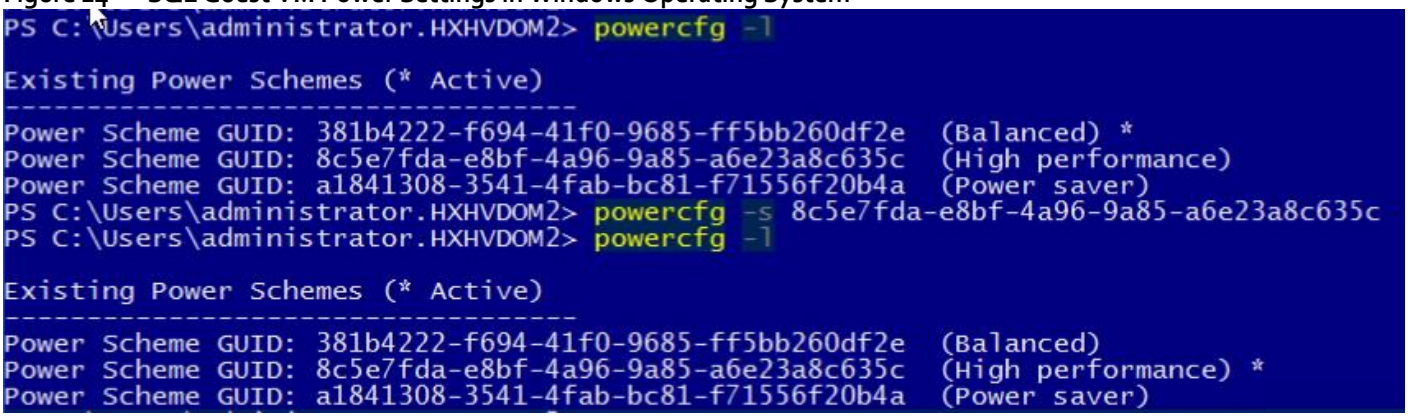
HX Servers are optimally configured with appropriate BIOS policy settings at the host level and does not require any changes. Similarly, at Hyper-V host power management options are also set to "High performance" at the time of HX installation by installer.

Figure 13 Power Setting on HX Hypervisor Node



Inside SQL Server guest VM, it is recommended to set the power management option to "High Performance" for optimal database performance as shown in Figure 14.

Figure 14 SQL Guest VM Power Settings in Windows Operating System



For more information about SQL server specific configuration recommendations in virtualized environments, see http://download.microsoft.com/download/6/1/d/61dde9b6-ab46-48ca-8380-d7714c9cb1ab/best_practices_for_virtualizing_and_managing_sql_server_2012.pdf

Achieving Database High Availability

Cisco HyperFlex storage systems incorporates efficient storage level availability techniques such as data mirroring (Replication Factor 2/3), native snapshot etc. to make sure continuous data access to the guest VMs hosted on the cluster. More details of the HX Data Platform Cluster Tolerated Failures are detailed in the Cisco HyperFlex Data Platform Management Guide, Release 3.0.

This section discusses the high availability techniques that will be helpful in enhancing the availability of virtualized SQL Server databases (apart from the storage level availability that comes with HyperFlex solutions).

The availability of the individual SQL server database instance and virtual machines can be enhanced using the technology listed below.

- Failover Cluster Manager – VM as clustered role for HA
- SQL Server Always On: To achieve database level high availability

Single VM / SQL Instance Level High Availability using Failover Cluster Manager Roles

Cisco HyperFlex solution leverages Hyper-V clustering to provide high availability to the hosted virtual machines. Since the exposed SMB share is accessible on all of the hosts in the cluster, they act as the shared storage environment to help migrate the VMs between the hosts. This configuration helps migrate the VMs seamlessly in case of planned as well as unplanned outage.

For more information, see https://docs.microsoft.com/en-us/powershell/module/failoverclusters/add-clustervirtualmachinerole?view=win10-ps

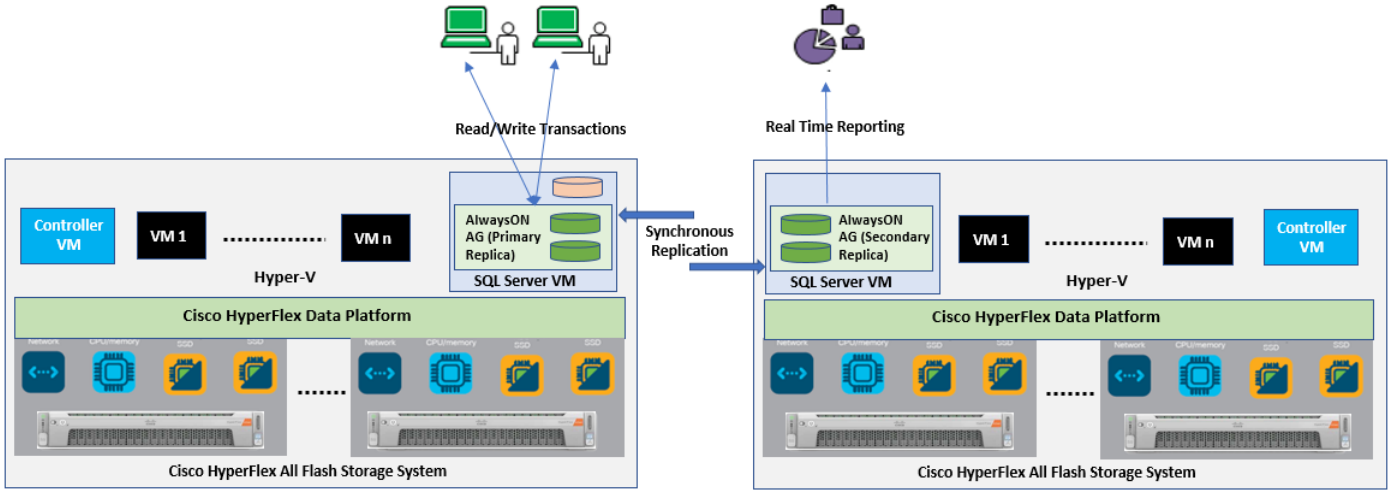
Database Level High Availability using SQL Always On Availability Group Feature

The database level availability of a single or multiple set of databases can still be ensured using SQL Server Always On feature that comes as part of the SQL server enterprise edition. Introduced in SQL Server 2012, Always On Availability Groups maximizes the availability of a set of user databases for an enterprise. An availability group supports a failover environment for a discrete set of user databases, known as availability databases, that failover together. An availability group supports a set of read-write primary databases and one to eight sets of corresponding secondary databases. Optionally, secondary databases can be made available for read-only access and/or some backup operations. For more information about this feature, see: https://msdn.microsoft.com/en-us/library/hh510230.aspx.

SQL Server Always On Availability Groups take advantage of Windows Server Failover Clustering (WSFC) as a platform technology. WSFC uses a quorum-based approach to monitor the overall cluster health and maximize node-level fault tolerance. The Always On Availability Groups will be configured as WSFC cluster resources and the availability of the same will depend on the underlying WSFC quorum modes and voting configuration explained in https://docs.microsoft.com/en-us/sql/sql-server/failover-clusters/windows/wsfc-quorum-modes-and-voting-configuration-sql-server.

Using Always On Availability Groups with synchronous replication that supported automatic failover capabilities, enterprises will be able to achieve the seamless database availability across the database replicas configured. The following figure depicts the scenario where an Always On availability group is configured between the SQL Server instances running on two separate HyperFlex Storage systems. To make sure that the involved databases can be protected with guaranteed high performance and no data loss in the event of failure, proper planning need to be done to maintain a low latency replication network link between the clusters.

Figure 15 Synchronous Always On Configuration Across HyperFlex All-Flash Systems



Although there are no specific rules about the infrastructure used for hosting a secondary replica, listed below are some of the guidelines that should be followed to have the primary replica on an All-Flash High Performing cluster:

- In case of synchronous replication (no data loss)
  - The replicas need to be hosted on similar hardware configurations to ensure that the database performance is not compromised in waiting for the acknowledgment from the replicas.
  - A high-speed low latency network connection between the replicas needs to be ensured.
- In case of asynchronous replication (minimal data loss)
  - The performance of the primary replica does not depend on the secondary replica, so it can be hosted on low cost hardware solutions as well.
  - The amount to data loss will depend on the network characteristics and the performance of the replicas.

The link below is for the Microsoft article that describes the considerations for deploying Always On availability groups, including prerequisites, restrictions, and recommendations for host computers, Windows Server failover clusters (WSFC), server instances, and availability groups:

https://docs.microsoft.com/en-us/sql/database-engine/availability-groups/windows/prereqs-restrictions-recommendations-always-on-availability?view=sql-server-2017

Deployment of Microsoft SQL Server

Cisco HyperFlex 3.5.1a Installation and Deployment on Hyper-V

This CVD focuses on the Microsoft SQL Server virtual machine deployment and assumes the availability of an already running healthy All-Flash HyperFlex 3.5.1a cluster on Hyper-V.

The step-by-step process to deploy and configure the HyperFlex system on Hyper-V is not in the scope of this document. For more information about deploying the Cisco HyperFlex 3.5.1a on Hyper-V All-Flash System, please refer to the installation guide:

https://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HyperFlex_HX_DataPlatformSoftware/HX_Hyper-V_Installation_Guide/3_5/b_Cisco_HyperFlex_Systems_Installation_Guide_for_Microsoft_HyperV_3_5.html

Deployment Procedure

This section provides step-by-step deployment procedure of setting up a test Microsoft SQL server 2016 on Windows Server 2016 virtual machine on a Cisco HyperFlex All-Flash system. Cisco recommends following the guidelines mentioned here: http://download.microsoft.com/download/6/1/d/61dde9b6-ab46-48ca-8380-d7714c9cb1ab/best_practices_for_virtualizing_and_managing_sql_server_2012.pdf to have an optimally performing SQL Server database configuration.

Before proceeding with creating guest VM and installing SQL Server on the guest, it is important to gather the following information. It is assumed that information such as IP addresses, Server names, DNS / NTP / VLAN details of HyperFlex Systems are available before proceeding with SQL VM deployment on HX All-Flash System. An example of the database checklist is shown in Table 2

Table 2 Virtual Interface Order with in HX-Series Server

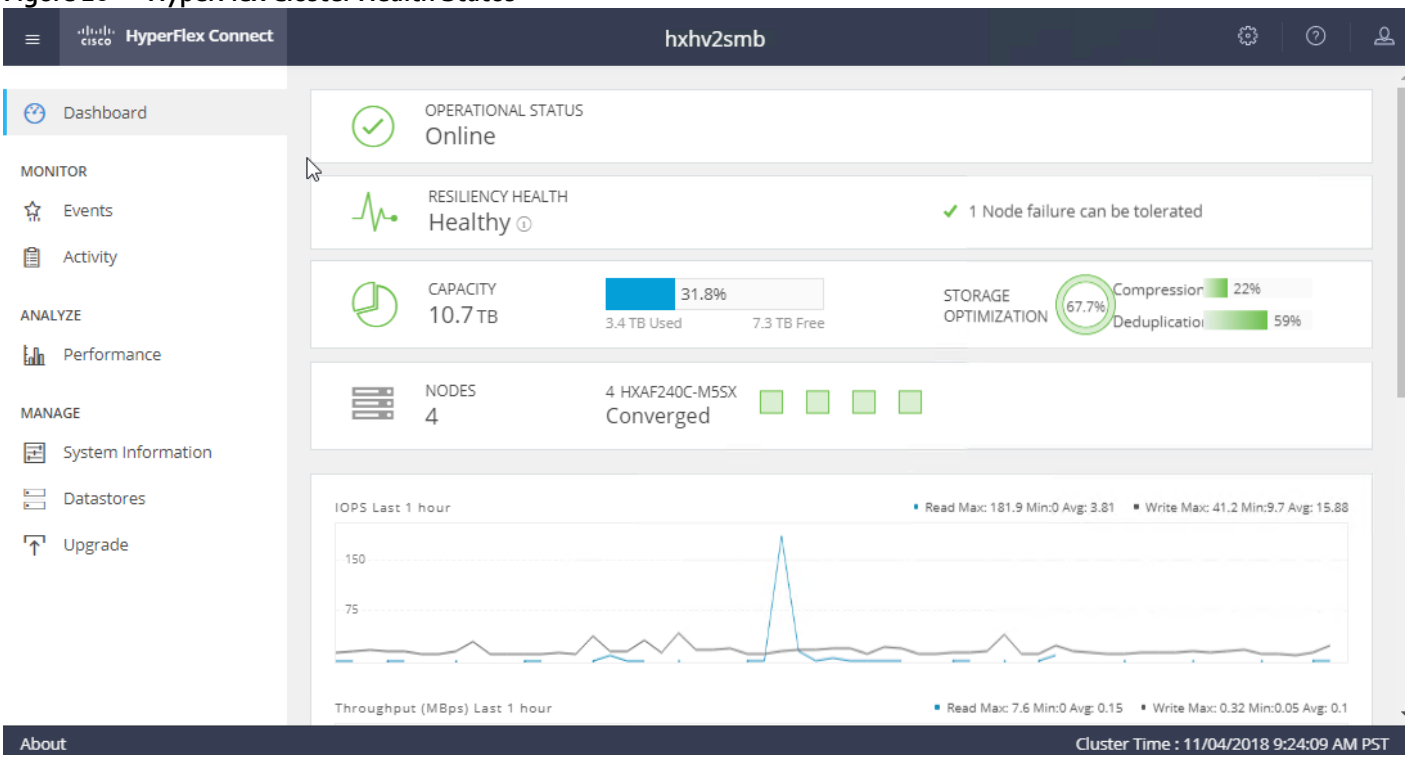
Component	Details
Cisco UCSM user name /password	admin / <<password>>
HyperFlex cluster credentials	admin / <<password>>
Hxadmin service account in AD	Domain name\hxadmin / <<password>>
Datastores names and their sizes to be used for SQL VM deployments	HXDS1 – 5 TB
Windows and SQL Server ISO location	\HXDS1\ISOs\
VM Configuration: vCPUs, memory, vmdk files and sizes	vCPUs : 4 Memory: 10GB OS : 75GB DATA volumes: SQL-DATA1 – 350GB, SQL-DATA2: 350GB Log volume: SQL-Log : 150GB

Component	Details
	All of these files to be stored in HXDS1 datastore.
Windows and SQL Server License Keys	<<Client provided>>
Drive letters for OS, Swap, SQL data and Log files	OS: C:\ SQL-Data1: D:\ SQL-Data2: E:\ SQL-Log: F:\

1. Verify HyperFlex Cluster System is healthy and configured correctly. This can be done in the following two ways.

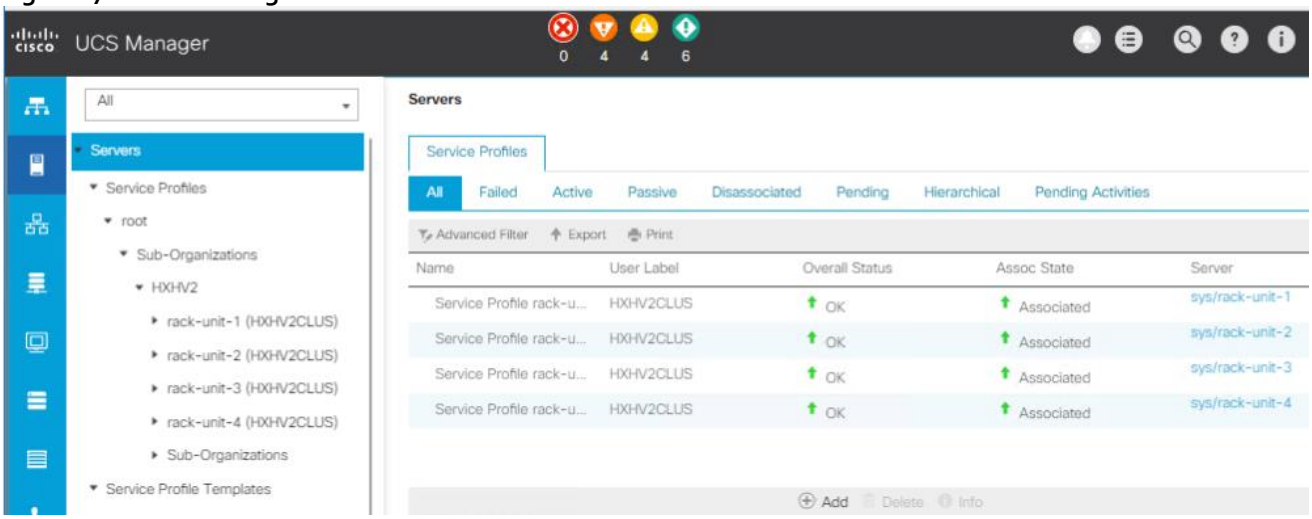
2. By logging in to HX Connect dashboard using the HyperFlex Cluster IP address as shown below.

Figure 16 HyperFlex Cluster Health Status



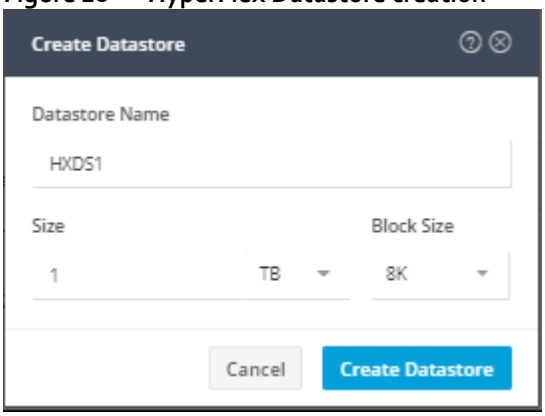
3. Make sure that Windows Hyper-V Host service profiles in the Cisco UCS Manager are all healthy without any errors. Following is the service profile status summary screenshot from Cisco UCS Manager GUI.

Figure 17 UCS Manager Service Profile



4. Create datastores for deploying SQL guest VMs and make sure the datastores are mounted on all the HX cluster nodes. The procedure for adding datastores to the HyperFlex system is provided in the HX Administration Guide. The following figure shows the creation of a sample datastore. 8K block size is chosen for datastore creation, as it is appropriate for SQL Server database.

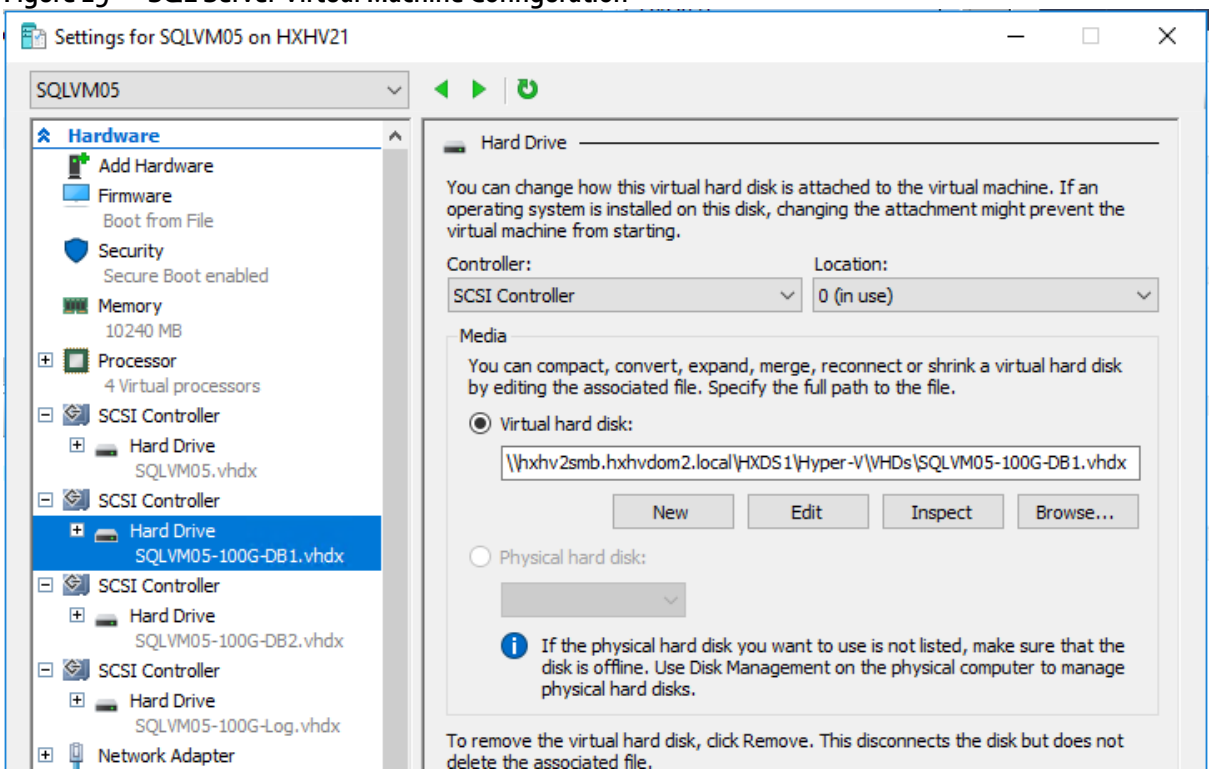
Figure 18 HyperFlex Datastore creation



5. Install the Windows Server 2016 virtual machine using the instructions mentioned in Microsoft article here. As described before in the Deployment Procedure section of this guide, make sure that the OS, data and log files are segregated and balanced by configuring separate virtual SCSI controllers. In Hyper-V Manager, select and right-click the VM and click Settings to change the VM configuration as shown below.

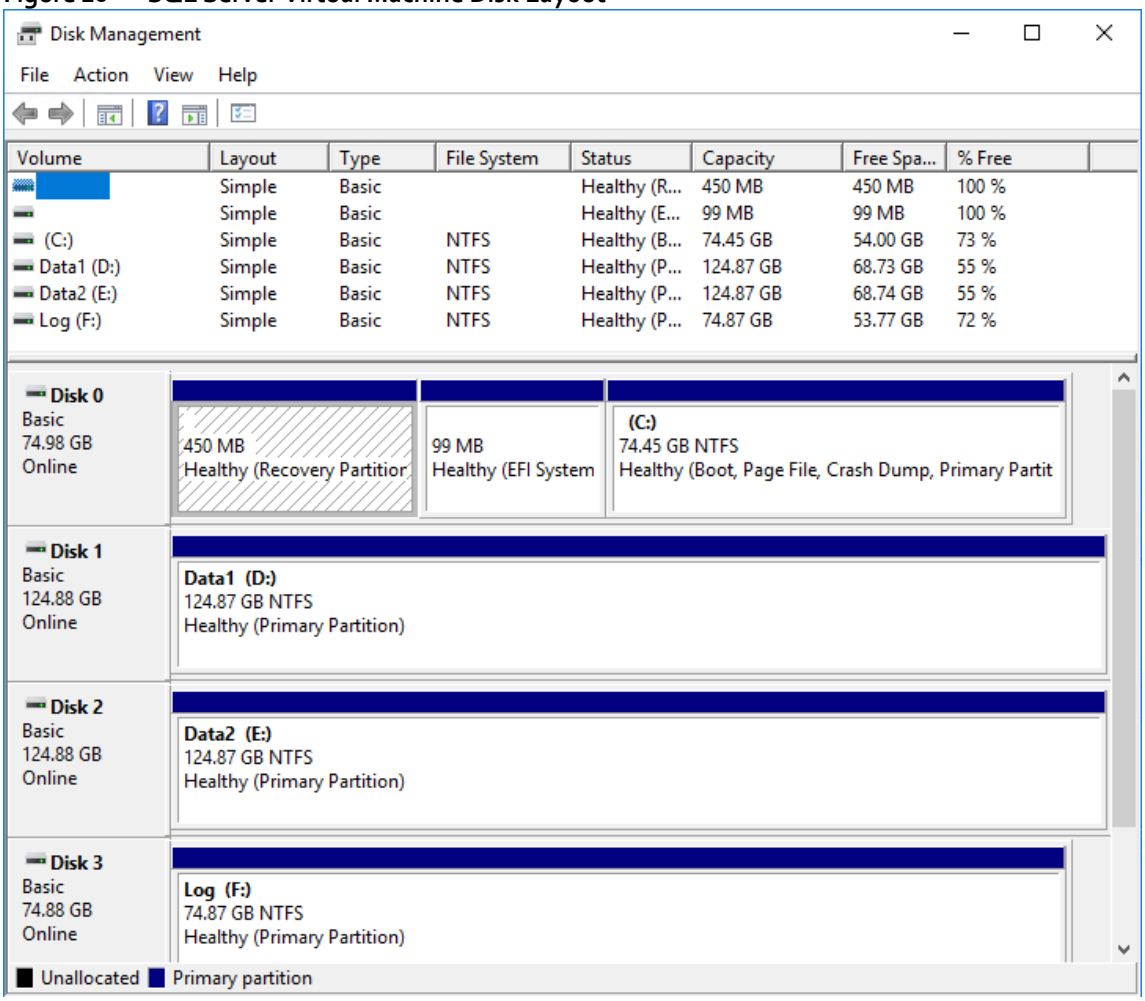
Fixed-size VHDX uses the full amount of space specified during VHD creation and it can deliver better throughput than dynamic VHDX.

Figure 19 SQL Server Virtual Machine Configuration



6. Initialize, format and label the volumes for Windows OS files, SQL Server data and log files. Use 64K as allocation unit size when formatting the volumes. The following screenshot (disk management utility of Windows OS) shows a sample logical volume layout of our test virtual machine.

Figure 20 SQL Server Virtual Machine Disk Layout



7. When the Windows Guest Operating System is installed in the virtual machine, it is highly recommended to have the latest VM Integration Services enabled and running as shown in the below figure. For more information about managing Hyper-V integration services click here.

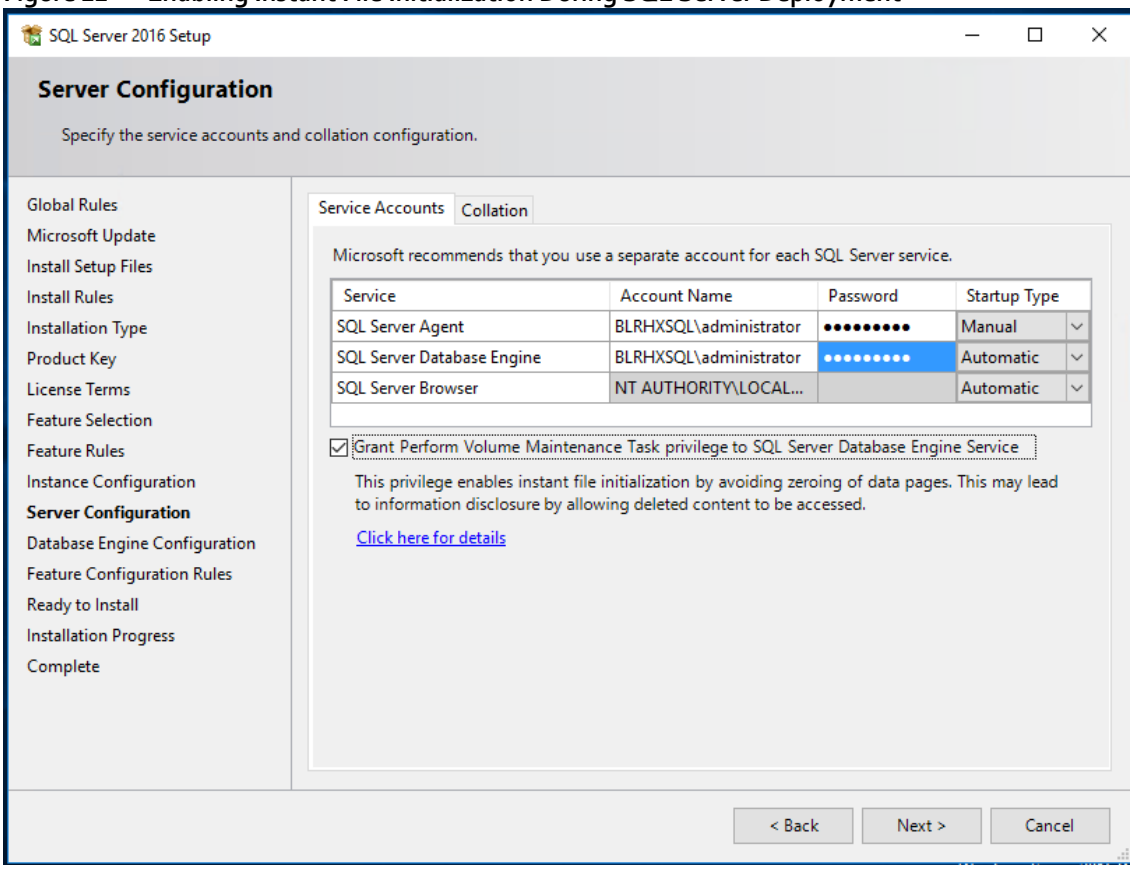
```
PS C:\Users\Administrator.HXHVDOM2> Get-VMIntegrationService -VMName SQLVM12

VMName  Name                    Enabled PrimaryStatusDescription SecondaryStatusDescription
------  ----                    ------- ------------------------ --------------------------
SQLVM12 Guest Service Interface False   OK
SQLVM12 Heartbeat               True    OK                       OK
SQLVM12 Key-Value Pair Exchange True    OK
SQLVM12 Shutdown                True    OK
SQLVM12 Time Synchronization    True    OK
SQLVM12 VSS                     True    OK
```

8. Install Microsoft SQL server 2016 on the Windows Server 2016 virtual machine. Follow the Microsoft documentation to install the database engine on the guest VM.

9. Download and mount the required edition of Microsoft SQL server 2016 SP1 ISO to virtual machine from the Hyper-V Manager. The choice of Standard or Enterprise edition of SQL server 2016 can be selected based on the application requirements

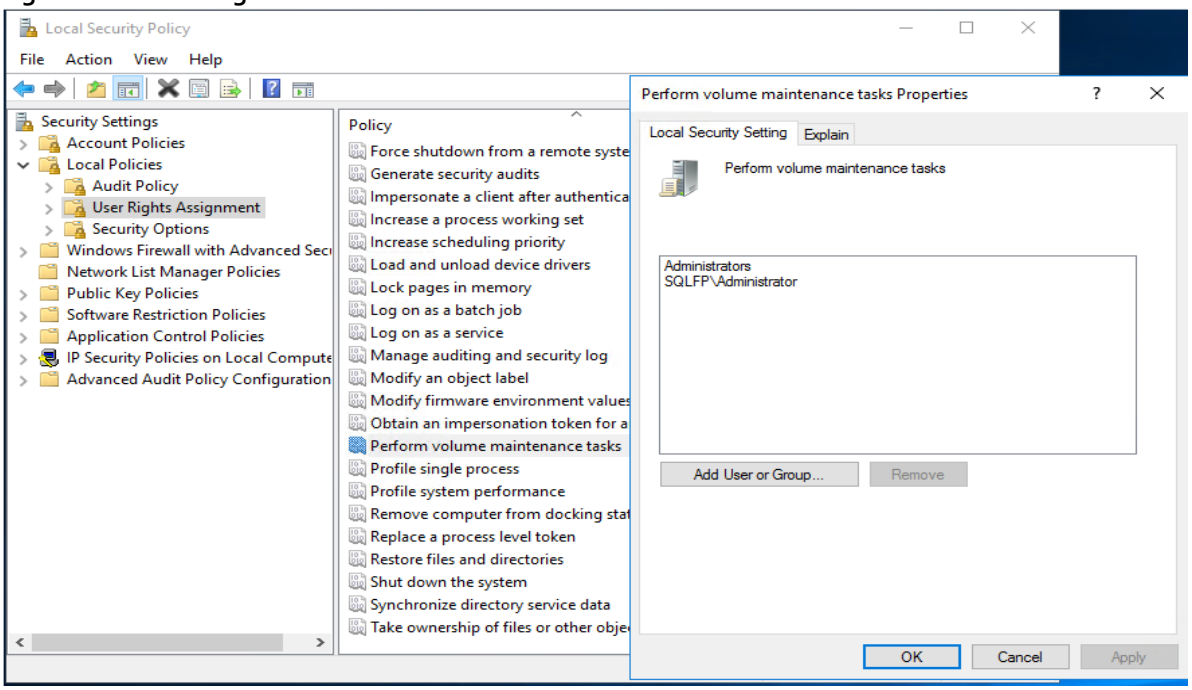
10. On the **Server Configuration** window of SQL Server installation, make sure that instant file initialization is enabled by enabling check box as shown in Figure 21. This enables the SQL server data files are instantly initialized avowing zero-ing operations.

Figure 21 Enabling Instant File Initialization During SQL Server Deployment



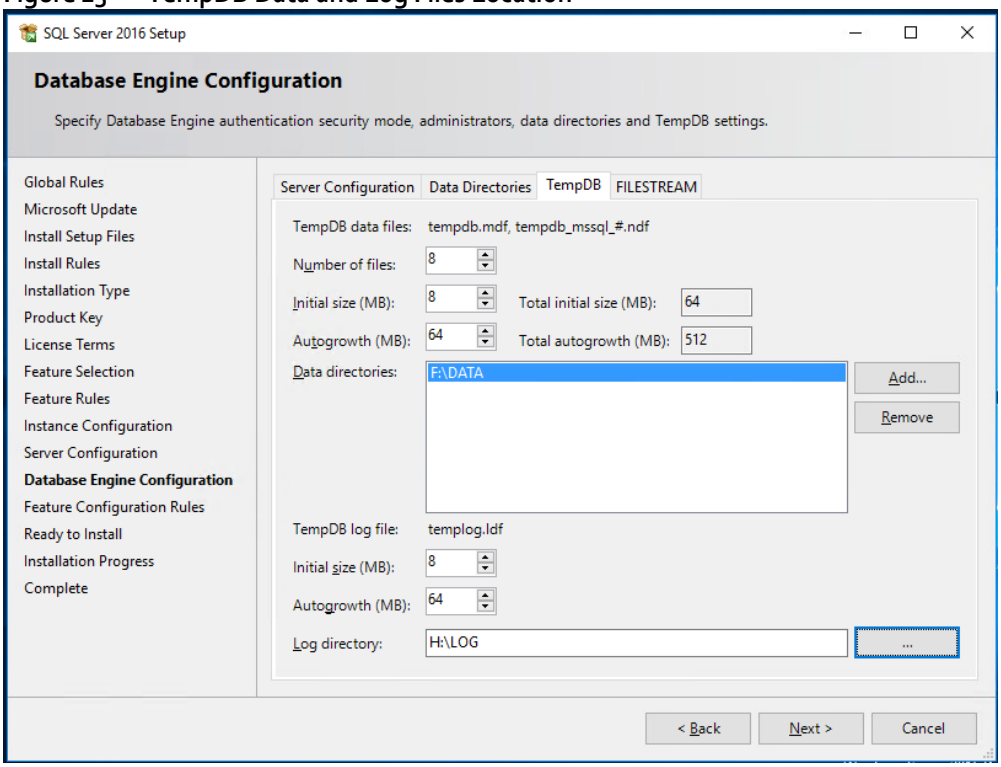
11. If the domain account which is used as SQL Server service account is not member of local administrator group, then add SQL Server service account to the "Perform volume maintenance tasks" policy using **Local Security Policy** editor as shown below:

Figure 22 Granting Volume Maintenance Task Permissions to the SQL Server Service Account



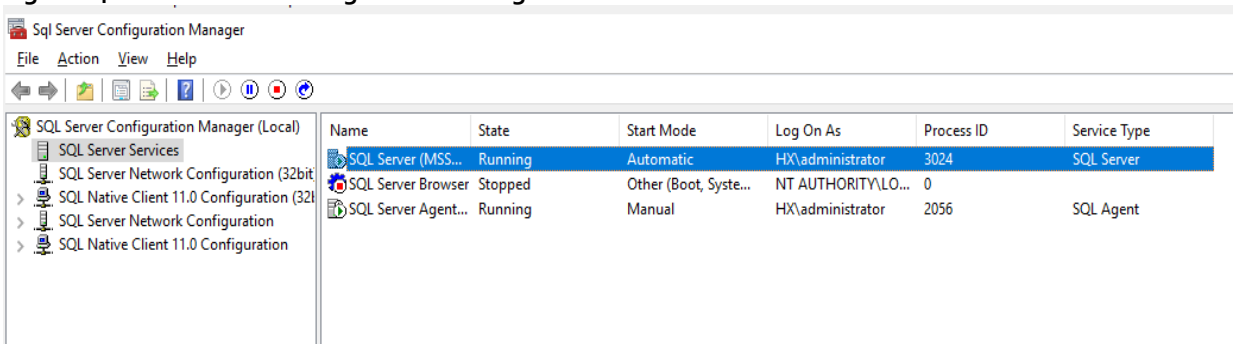
12. In the **Database Engine Configuration** window under the TempDB tab, make sure the number of TempDB data files are equal to 8 when the vCPUs or logical processors of the SQL VM is less than or equal to 8. If the number of logical processors are more than 8, start with 8 data files and try to add data files in the multiple of 4 when the contention is noticed on the TempDB resources. The following diagram shows that there are 8 TempDB files chosen for a SQL virtual machine which has 8 vCPUs. Also, as general best practice, keep the TempDB data and log files are two different volumes.

Figure 23 TempDB Data and Log Files Location



13. When the SQL Server is successfully installed, use SQL Server Configuration manager to verify that the SQL server service is up and running as shown below.

Figure 24 SQL Server Configuration Manager



14. Create a user database using SQL Server Management studio or Transact-SQL so that the database logical file layout is in line with the desired volume layout. Detailed instructions are provided here.

Solution Resiliency Testing and Validation

This section discusses some of the solution tests conducted to validate the robustness of the solution. These tests were conducted on a HyperFlex cluster built with four HXAF240c M4 All-Flash nodes. Table 3 lists the component details of the test setup. Other failure scenarios (failures of disk, network, etc.) are out of the scope of this document. The test configuration used for validation is described below.

Table 3 Hardware and Software Component Details used in HyperFlex All-Flash Testing and Validation

Component	Details
Cisco HyperFlex HX data platform	Cisco HyperFlex HX Data Platform software version 3.5(1a) Replication Factor : 3 Inline data dedupe/ compression : Enabled(default)
Fabric Interconnects	2x Cisco UCS 3rd Gen UCS 6332-16UP UCS Manager Firmware: 4.0 (1b)
Servers	4 x Cisco UCS HX-series C240c M5
Processors per node	2x Intel® Xeon® Gold 6140 CPUs @2.30GHz, 18 Cores each
Memory Per Node	384GB (12 x 32GB) at 2666 MHz
Cache Drives Per Node	1x 400GB 2.5 inch Ent. Performance 12G SAS SSD (10X Endurance)
Capacity Drives Per Node	10x 960GB 2.5 inch Enterprise Value 6G SATA SSD
OS/Hypervisor	Windows Server 2016 Hyper-V (v1607 Build # 14393.1884)
Network switches (optional)	2x Cisco Nexus 9396PX (9000 series)
Guest Operating System	Windows 2016 Standard Edition
Database	SQL Server 2016 SP1
Database Workload	Online Transaction Processing With 70:30 Read Write Mix

The major tests conducted on the setup are as follows and will be described in detail in this section.

- Node failure Test
- Fabric Interconnect Failure Test
- Database Maintenance Tests

Notice that in all the tests (above), the HammerDB testing tool is used to generate required stress on the guest SQL VM. A separate client machine located outside the HyperFlex cluster is used to run the testing tool and generate the database workload.

Node Failure Test

The intention of this failure test is to analyze how the HyperFlex system behaves when failure is introduced into the cluster on an active node (running multiple guest VMs). The expectation is that the Cisco HyperFlex system should be able to detect the failure, initiate the VM migration from a failed node and retain the pre-failure state with an acceptable limit of performance degradation.

In our testing, node failure is introduced when the cluster is stressed with eight SQL VMs utilizing 60-70 percent of cluster storage capacity and 70 percent CPU utilization. When one node was powered off (unplug both power cables), all the SQL guest VMs running on the failed node successfully failed over to other node. No database consistency errors were reported in SQL Server logs of the migrated VMs. After the VMs migrated to the other nodes, database workload is manually restarted. Around 5 percent of impact is observed on the overall performance because of failed node in the cluster. Later when the failed node was powered up, it rejoined the cluster automatically and started syncing up with the cluster. The cluster returned to the pre-failure performance within 5-10 minutes after the failed node was brought online (including the cluster sync-up time).

Fabric Interconnect Failure Test

The intention of this failure test is to analyze the performance impact in case of a fabric interconnect switch failure. Since Cisco UCS Fabric Interconnects (FIs) are always deployed in pairs and operating as a single cluster, failure of a Fabric Interconnect should not impact systems connected to them.

In our tests, we introduced the Fabric Interconnect failure when the cluster was stressed with SQL VMs utilizing 60-70 percent of cluster storage capacity and 70 percent CPU utilization. When one of the FIs was powered off (unplugged both power supplies), no impact was observed on the VMs running the workload. All the VMs remained fully functional and no VM migrations were observed in the cluster.

Database Maintenance Tests

In any hyperconverged systems, it is important to assess the impact caused by database maintenance tasks to the regular SQL workloads running on the same cluster. The maintenance activities typically have sequential IO pattern as opposed to random IO pattern of regular OLTP workloads. Usually in typical hyperconverged shared storage environments, caution must be exercised while running DB maintenance activities during the business hours as they may impact the regular operational workloads.

Following maintenance activities were carried out on a SQL guest VM deployed on an All-Flash cluster to assess the impact caused by these maintenance activates on the ongoing workload in the cluster. The cluster setup used is the same as detailed in Table 3 .

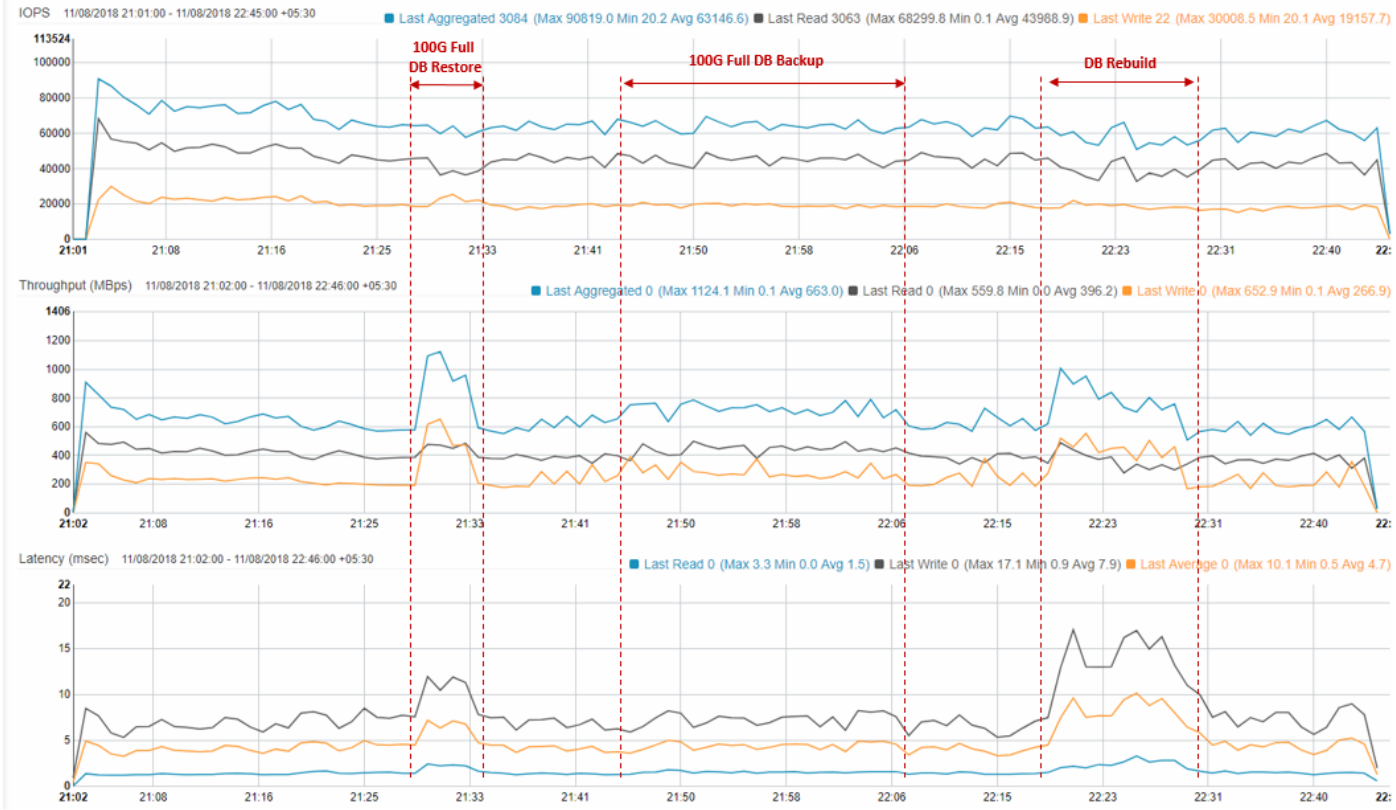
- Full database restore from an external backup source (backup source residing outside of HX cluster)
- Full database consistency check and complete backup to an external backup source (outside of HX cluster)
- Rebuilding indexes of two large tables (of size around 70GB)
- Exporting SQL data to flat files (located within HX cluster)
- Importing data from flat files into SQL (files located within HX Cluster)

Notice that these maintenance activities are run on a separate SQL guest VM in parallel to the regular SQL VMs which were used to exert stress on the cluster up to its 40-50 percent storage capacity utilization. With no maintenance activity, the OLTP workload was generating around an average of 65K IOPs, 600 MBps of throughput and average latency of 4.4 msec. As expected, full database restore caused a drop of around 7 percent drop in IO performance which is understandable given the full database restore activity (100 percent sequential writes activity) done on a cluster which is already exercised to 70

percent resource usage by normal OLTP workload. Other activities (in the above list) had IOPS impact ranging anywhere from 3 to 10 percent with marginal increase in latencies.

The amount of impact caused by the maintenance activities would typically depend on the replication factor and percentage of cluster resource utilization in addition to factors such as back up settings etc. On a properly sized system with appropriate resource headroom, the impact would be much lower. Figure 25 (HX performance dashboard GUI) shows the cluster behavior when the maintenance activities such as restore, backup and rebuild are performed on a VM and when an operational workload is running on the other VMs in the same cluster.

Figure 25 Performance Impact Analysis of Typical Database Maintenance Activities on Ongoing Database Workload



Database Performance Testing

This section contains examples of different ways in which Microsoft SQL server workloads can take advantage of the HyperFlex Data Platform architecture and its performance scaling attributes.

Single Large VM Performance

HyperFlex Data Platform uses a distributed architecture and one of the main advantages of this approach is that the cluster resources form a single, seamless pool of storage capacity and performance resources. This allows any individual VM to take advantage of the overall cluster resources and is not limited to the resources on the local node that hosts the VM. This is a unique capability and significant architectural differentiator that Cisco HyperFlex Data Platform provides.

There are a couple of deployment scenarios that are common in data centers which benefit from Cisco HyperFlex Data Platform:

- VM Hotspot – rarely do all the VMs in any shared virtual infrastructure show uniform utilization of resources. The capacity growth for the individual VMs usually are different and also their performance requirements are different at different points in time. With the HyperFlex distributed architecture those hotspots are easily absorbed by the infrastructure without having capacity or performance hotspots in the infrastructure.

- Large VMs – Since the cluster presents a common pool of resources, it makes it possible to deploy large applications and VMs with performance and capacity requirements that exceed the capability of any single node in the cluster.

This section demonstrates the above mentioned attribute of the architecture through a performance test with a large SQL VM. The cluster setup used is the same as detailed in Table 3 . The details of the VM configuration and the workload are provided in Table 4 . OLTP workload with 70:30 read write ratio was exerted on the guest VM. The workload stressed the VM up to 70 percent of CPU utilization which resulted in 35-40 percent of Hyper-V host CPU utilization.

Table 4 VM and Workload Details used for Single Large VM Test

Configuration	Details
VM	8 vCPUs, 12GB Memory(8G assigned to SQL) Two Data volumes and one Log volume(each with a dedicated SCSI controller)
Workload	Tool Kit: HammerDB Users: 41 Data Warehouses: 8000 DB Size= 800GB RW Ratio: ~70:30

Figure 26 shows the performance seen by the single VM running a large SQL workload for roughly 8 hours. There are few noteworthy points here:

- Large VM with a very large working set size can get sustained high IOPS that is leveraging resources (capacity and performance) from all 4 nodes in the cluster. Note that it is possible to scale to higher IOPS with an even larger VM.

- Dedupe and Compression is on by default and that is the case during this test as well.

Figure 26 Single Large Working SQL Server Database Workload on HyperFlex All-Flash Cluster

This test demonstrates the ability of HyperFlex to leverage the resources from all nodes in the cluster to satisfy the performance (and capacity) needs of any given VM.

Performance Scaling with Multiple VMs

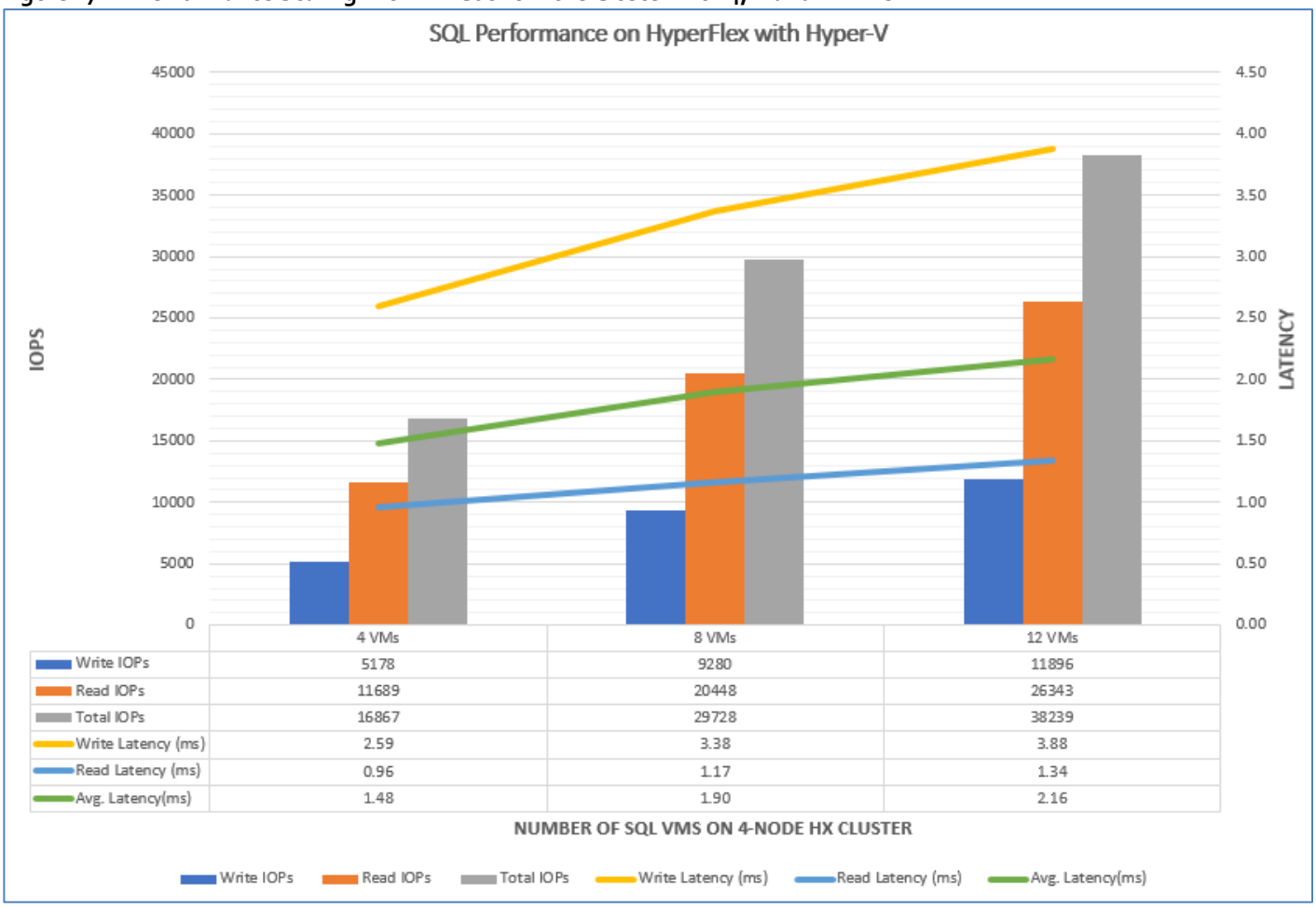
The seamless pool of capacity and performance presented by the HyperFlex cluster can also be accessed by multiple smaller VMs. In this section we are using multiple VMs that are each running a SQL instance with HammerDB. The cluster setup used is the same as detailed in Table 3 . The details of the VM configuration and the workload are listed in Table 5 . OLTP workload with 70:30 read write ratio was exerted on each guest VM. The workload stressed each VM up to 40 percent of guest CPU utilization.

Table 5 VM and Workload Details used for Single Large VM Test

Configuration	Details
VM	4 vCPUs, 10GB Memory (8G is assigned to SQL) Two Data LUNs & one Log LUN VM count scales in units of 4 (1 VM per node).
Workload	Tool Kit: HammerDB Users: 5 Data Warehouses: 1000 DB Size= ~100GB RW Ratio: ~70:30

Figure 27 shows the performance scaling seen by scaling the number of VMs in the cluster from 4 (1 VM per node), 8 (2 VMs per node) and finally 12 (3 VMs per node). The performance data shown in the below graphs is captured using Windows Perfmon tool. Note that these are sustained level of performance. Also deduplication and compression is on by default and that is the case during this test as well. If one or more VMs need additional IOPS / throughout, they will be able to get the increase in storage performance provided the VM itself is not CPU or memory bottlenecked and there is additional performance headroom available in the cluster.

Figure 27 Performance Scaling with VM Count in the Cluster with 4, 8 and 12 VMs



This test demonstrates HyperFlex Data Platform's ability to scale the cluster performance with a large number of VMs.

Common Database Maintenance Scenarios

This section discusses common database maintenance activities and provides a few guidelines for planning database maintenance activities on the SQL VMs deployed on the All-Flash HyperFlex system.

The most common database maintenance activities include export, import, index rebuild, backup, restore and running database consistency checks on regular intervals. The IO pattern of these activities usually differ from business operational workloads hosted on the other VMs in the same cluster. The maintenance activities would typically generate sequential IO when compared to the business transactions which typically generate random IO (in case of transactional workloads). When sequential IO pattern is introduced to the system alongside with random IO pattern, there is a possibility of impact on IO sensitive database applications. Hence caution must be exercised while sizing the environment or controlling the impact by running DB maintenance activities during the business hours in production environments. The following list provide some of the guidelines to run the management activities in order to avoid the impact on business operations.

- As a general best practice all the management activities such as export, import, backup, restore and DB consistency checks must be scheduled to run off business hours when no critical business transactions are running on the underlying HyperFlex system in order to avoid impact on the ongoing business operations. Another way of limiting the impact is to size the system with appropriate headroom.
- In case of any urgency to run the management activities in the business hours, administrators should know the IO limits of hyperconverged systems and plan to run accordingly.
- For clusters running at peak load or near saturation, when exporting a large volume of data from SQL database hosted on any hyperconverged system to any flat files, it should be ensured that the destination files are located outside of the HyperFlex cluster. This will avoid the impact on the other guest VMs running on the same cluster. For small data exports, the destination files can be on the same cluster.
- Most of the import data operations will be followed by recreation of index and statistics in order to update the database metadata pages. Usually Index recreation would cause lot of sequential read and writes hence it is recommended to schedule import data in off business hours.
- Database restore, backup, rebuilding indexes and running database consistency checks typically generate huge sequential IO. Therefore, these activities must be scheduled to run in the out of business hours.

In case of complete guest or database backups, it is not recommended to keep the backups in the same cluster as it would not protect against the scenario where the entire cluster is lost, for example, during a geographic failure or large scale power outage, etc. Data protection of the virtualized applications that are deployed on the hyperconverged systems are becoming one of the major challenges to customers; so there is a need for a flexible, efficient and scalable data protection platform.

Cisco HyperFlex has integration with several backup solutions, for example, Cisco HyperFlex™ System's solution together with Veeam Availability Suite gives customers a flexible, agile, and scalable infrastructure that is protected and easy to deploy.

Workloads are rarely static in their performance needs. They tend to either grow or shrink over time. One of the key advantages of the Cisco HyperFlex architecture is the seamless scalability of the cluster. In scenarios where the existing workload needs to grow; Cisco HyperFlex can handle the scenario by growing the existing cluster's compute, storage capacity or storage performance capabilities depending on the resource requirement. This gives administrators enough flexibility to right size their environment based on today's needs without worrying about future growth.

Troubleshooting Performance

For information about how to use Performance Monitor to identify problems in host and virtual machines that do not perform as expected, see: https://docs.microsoft.com/en-us/windows-server/administration/performance-tuning/role/hyper-v-server/detecting-virtualized-environment-bottlenecks

For information about the best practices for virtualizing and managing SQL server, see: http://download.microsoft.com/download/6/1/d/61dde9b6-ab46-48ca-8380-d7714c9cb1ab/best_practices_for_virtualizing_and_managing_sql_server_2012.pdf

Some of the commonly seen performance problems on virtualized/hyperconverged systems are described in the following sections.

High SQL Guest CPU Utilization

When high CPU utilization with lower disk latencies on SQL guest VMs is observed and CPU utilization on Hyper-V hosts appears to be normal, then it might be the case that virtual machine is experiencing a CPU contention. In that case, the solution may be to add more vCPUs to the virtual machine as the workload is demanding more CPU resources.

When high CPU utilization is observed on both guest and Hosts, then one of the options to be looked at is upgrading to a higher performing processor.

High Disk latency on SQL Guest

The following guidelines can be used to troubleshoot when higher disk latencies are observed on SQL guest VMs.

Use Windows Performance Monitor to identify latencies and follow the options mentioned in section "Deployment Planning."

In case of higher HX storage capacity utilization nearing expected thresholds (above 60 percent usage), SQL VMs might also experience IO latencies at both guest and host level. In such case, it is recommended to scale up the cluster by adding new HX node to the cluster.

Summary

The Cisco HyperFlex HX Data Platform revolutionizes data storage for hyperconverged infrastructure deployments that support new IT consumption models. The platform's architecture and software-defined storage approach gives you a purpose-built, high-performance distributed file system with a wide array of enterprise-class data management services. With innovations that redefine distributed storage technology, the data platform provides you the optimal hyperconverged infrastructure to deliver adaptive IT infrastructure. Cisco HyperFlex systems lower both operating expenses (OpEx) and capital expenditures (CapEx) by allowing you to scale as you grow. They also simplify the convergence of compute, storage, and network resources.

All-Flash configurations enhance the unique architectural capabilities of Cisco Hyperflex systems to cater to the high performance low latency enterprise application requirements. This makes it possible to utilize the entire cluster resources efficiently by the hosted virtual machines regardless the host. This enables the virtualized SQL Server implementations as an excellent candidate for the high-performing Cisco HyperFlex All-Flash systems.

Lab solution resiliency tests detailed in this document show the robustness of the solution to host IO sensitive applications like Microsoft SQL Server. The system performance tuning guidelines described in this document addresses the platform specific tunings that will be beneficial for attaining the optimal performance for a SQL Server virtual machine on Cisco HyperFlex All-Flash System. SQL Server performance observed on Cisco HyperFlex systems focused in this document proves Cisco HyperFlex All-Flash system to be an ideal platform to host high performing low latency applications like Microsoft SQL Server database.

About the Authors

Sanjeev Naldurgkar, Cisco Systems, Inc.

Sanjeev Naldurgkar is a Technical Marketing Engineer with Cisco UCS Data Center Solutions Group. He has been with Cisco for six years focusing on the delivery of customer-driven solutions on Microsoft Hyper-V and VMware vSphere. Sanjeev has over 16 years of experience in the IT Infrastructure, Server virtualization, and Cloud Computing. He holds a Bachelor Degree in Electronics and Communications Engineering, and leading industry certifications from Microsoft and VMware.

Acknowledgements

- Gopu Narasimha Reddy, Cisco Systems, Inc.
- Rajesh Sundaram, Cisco Systems, Inc.
- Babu Mahadevan V, Cisco Systems, Inc.
- Vadi Bhatt, Cisco Systems, Inc.