IMPLEMENTATION GUIDE

Data Center
Software-Defined Storage

Implementing High-Performance Software-Defined Storage

Optimize the performance of Microsoft Azure Stack HCI with 2nd Generation Intel® Xeon® Scalable processors and Intel® Optane™ technology

Table of Contents

Introduction

To remain competitive in today's fast-moving and digitally demanding marketplace, businesses need to explore ways to increase data center efficiency. One excellent way to do so is to deploy hyperconverged infrastructure (HCI), which can simplify data center complexity, improve scalability, increase reliability and manageability, and optimize resource utilization—all of which helps reduce data center costs and increases business agility. Microsoft Azure Stack HCI makes it possible to run virtualized applications on premises as well as connect to Azure for cloud services.

Microsoft Azure Stack HCI combines highly virtualized compute, storage, and networking on industry-standard servers and components and is optimized for 2nd Generation Intel® Xeon® Scalable processors. This workload-optimized platform is designed to provide enhanced performance and advanced capabilities. Adding Intel® Optane™ technology to Microsoft Azure Stack HCI can help organizations increase data throughput, reduce latency, affordably increase memory capacity, and quickly extract value from large datasets for timely, actionable insights, all while potentially consolidating workloads on a smaller data center footprint.

Intel has created multiple reference architectures (see "Summary of Reference Architectures") to accelerate infrastructure decisions and solve storage efficiency and memory capacity issues. Organizations can look to these modern solutions to help meet today's storage and memory requirements across a wide variety of use cases.

Solution Overview

Azure Stack HCI is offered through Microsoft hardware partners. It is typically preconfigured or bundled with simple configuration software, and can be designed to support a wide array of use cases. Its minimal hardware requirements help enable a small-footprint, two-node deployment if needed. However, it can also scale up to 16 nodes (or more through a Windows Server 2019 feature called cluster sets). Figure 1 shows the basic system architecture; system requirements are discussed in the following sub-sections. You can contact your preferred hardware vendor about Azure Stack HCI solutions.

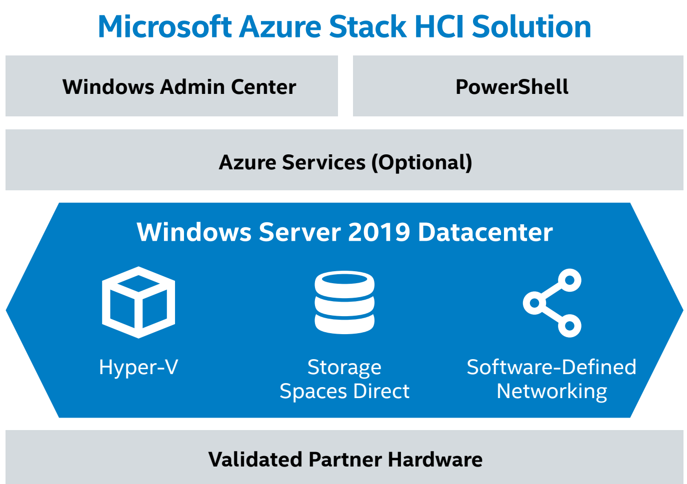


Figure 1. Basic architecture for Microsoft Azure HCI solution.

Hardware Requirements

Microsoft Azure Stack HCI supports Intel® Optane™ technology in Intel Optane persistent memory modules and in Intel Optane solid state drives (SSDs). You can use persistent memory in Memory Mode to maximize the number of virtual machines (VMs) hosted per node. You can also use Intel Optane SSDs as cache drives to deliver increased throughput for Storage Spaces Direct. Table 1 shows the solution's minimum hardware requirements and maximum supported specifications.

Software Requirements

The Microsoft Azure Stack HCI solution is intended to be preinstalled and (partially or completely) preconfigured by Microsoft partners on validated, industry-standard hardware. It is built from the following components:

- Management tools:
  - Windows Admin Center for central, comprehensive management of local and remote servers through a graphical interface
  - Azure services integrated into Windows Admin Center for offsite backups, site recovery, cloud-based monitoring, and other benefits (optional)
  - PowerShell for scripting and automation
- Windows Server 2019 Datacenter roles and features:
  - Hyper-V to run VMs on all physical hosts
  - Software-Defined networking (SDN) for network virtualization (optional)
  - Storage Spaces Direct for storage virtualization

Table 1. Microsoft Azure Stack HCI Hardware Requirements

MINIMUM HARDWARE REQUIREMENTS	
Number of physical servers per cluster	2
Intel® CPU (per node)	• Intel® Xeon® processor[a] • 1.4 GHz 64-bit processor • Compatible with x64 instruction set
RAM (per node)	32 GB
MAXIMUM SUPPORTED HARDWARE SPECIFICATIONS	
Number of physical servers per cluster	16 (clusters can also be combined using cluster sets to create an HCI platform of hundreds of nodes)
Maximum number of VMs per host	1,024
Maximum number of disks per VM (SCSI)	256
Maximum storage per cluster	4 PB
Maximum number of logical processors per host	512
MAXIMUM RAM	
Per host	24 TB
Per VM	• 12 TB for 2 VMs • 1 TB for 1 VM
MAXIMUM NUMBER OF VIRTUAL PROCESSORS	
Per host	2,048
Per VM	• 240 vCPUs for Gen2 VM • 64 vCPUs for Gen1 VM

[a] 2nd Generation Intel® Xeon® Scalable processor is required to support Intel® Optane™ persistent memory.

Summary of Reference Architectures

Depending on workload needs, we recommend using one of the following three reference architectures, as detailed in Tables 2–4 and visually represented in Appendix A. Note that for the storage network it is highly recommended to use a network interface card (NIC) that supports remote direct memory access (RDMA) over iWARP.

Option 1. Increase caching speed. Use Intel Optane SSDs as cache, plus SATA-based Intel SSDs for the capacity tier, to speed caching and increase VM density—leading to server consolidation.

Option 2. Increase available memory. Workloads that need more memory can benefit from Intel Optane persistent memory in Memory Mode, in addition to the Intel Optane SSDs in the cache tier.

Option 3. Increase memory *and* cache bandwidth. Workloads that not only need additional memory, but also need extremely low latency can combine Intel Optane persistent memory in Memory Mode plus App Direct-Dual Mode, where the cache layer uses a two-tier architecture that allows for increased bandwidth and frees up drive bays for more capacity. The App Direct Mode persistent memory replaces Intel Optane SSDs in the cache tier. This configuration is ideal for high-bandwidth scenarios.

Table 2. Bill of Materials for Option 1 Reference Architecture Focused on Increasing Caching Speed

	DESCRIPTION	QUANTITY PER NODE
CPU	Intel® Xeon® Gold 6244 processor (8 cores)	2
Boot Drive	Intel® SSD D3-S4510 480 GB	1
Memory	16x 16 GB DDR4	256 GB
Cache Tier	Intel® Optane™ SSD DC P4800X 375 GB	2
Capacity Tier	Intel SSD D3-S4510 3.84 TB	2+
Storage Network (Data)	Intel® Ethernet Network Connection X722 10 GbE Dual Port with RDMA capabilities	1
Management Network	1 GbE Network	1

Table 3. Bill of Materials for Option 2 Reference Architecture Focused on Increasing Available Memory

	DESCRIPTION	QUANTITY PER NODE
CPU	Intel® Xeon® Gold 6242 processor (16 cores)	2
Boot Drive	Intel SSD D3-S4510 480 GB	1
Memory	• Up to 8x 128 GB Intel® Optane™ persistent memory in Memory Mode (addressable memory) • 12x 16 GB DDR4 (memory cache)	Up to 1 TB 192 GB
Cache Tier	Intel Optane SSD DC P4800X 375 GB	2
Capacity Tier	Intel SSD D3-S4510 3.84 TB *or* Intel® SSD D5-P4320 7.84 TB	2+
Storage Network (Data)	Intel® Ethernet Network Connection X722 10 GbE Quad Port with RDMA capabilities	1
Management Network	1 GbE Network	1

Table 4. Bill of Materials for Option 3 Reference Architecture Focused on Increasing Caching Bandwidth ***and*** Available Memory

	DESCRIPTION	QUANTITY PER NODE
CPU	Intel® Xeon® Gold 6252 processor (24 cores)	2
Boot Drive	Intel SSD D3-S4510 480 GB	1
Memory/Cache Tier[a]	• Up to 12x 128 GB Intel® Optane™ persistent memory – 6x used in Memory Mode for available system memory – 6x used in App Direct-Dual Mode to replace cache tier for increased caching bandwidth • 12x 16 GB DDR4 (memory cache)	 768 GB 768 GB 192 GB
Capacity Tier	Intel SSD D5-P4320 7.84 TB	4+
Storage Network (Data)	25 GbE Dual Port Network Adapter with RDMA capabilities	1
Management Network	1 GbE Network	1

[a] Ratio of Memory Mode and App Direct-Dual Mode is configurable.

Installation and Configuration

Installing Windows Admin Center

Windows Admin Center is installed independently of Windows Server 2019, and its MSI file is available for download here. For remote administration, the tool can be installed on any machine (except a domain controller) running Windows 10 or Windows Server 2012 or later and using a modern browser such as Microsoft Edge or Google Chrome. Running the MSI file sets up a web server (Internet Information Services [IIS] is not required) on the local machine, which communicates over port 6516 by default. It also installs a component called the Windows Admin Center gateway. The gateway enables management of remote servers through remote PowerShell and Windows Management Instrumentation (WMI) over Windows Remote Management (WinRM). If you are not using domain admin credentials to manage the remote servers, you should let Windows Admin Center manage your TrustedHosts lists. Otherwise, you must configure Trusted Hosts manually.

For more information, visit the Windows Admin Center.

Installing Windows Server 2019 Datacenter

Windows Server 2019 is the recommended operating system. For details about installation, review the article "Install, upgrade, or migrate to Windows Server."

Configuring Intel Optane Persistent Memory in Memory Mode

The steps for setting up Memory Mode vary depending on server model. Refer to your OEM of choice for specific steps. Here is an overview of the process:

1. Follow your hardware vendor's BIOS setup and memory placement steps for Intel Optane persistent memory to enable Memory Mode.
2. Reboot.
3. Make sure the latest version of ipmctl is installed (see https://github.com/intel/ipmctl).
4. Check the installed memory capacity using Task Manager → System Information.

Configuring Intel Optane Persistent Memory in App Direct-Dual Mode

App Direct-Dual Mode combines Memory Mode and App Direct Mode, thereby extending volatile system memory as well as providing persistent memory. To enable App Direct-Dual Mode, use the above steps for the Intel Optane persistent memory modules you want to use in Memory Mode, then set up the remaining modules as persistent memory.

Best Practices

Build a Balanced Design

It is a best practice to increase all subsystems by the same factor, thereby avoiding creating bottlenecks by lack of one or more subsystems. With the increase in options to design HCI solutions, it is imperative that all resources support one another in a balanced approach. The key is not to under-provision nor over-provision any subsystem—achieving a balanced design approach for HCI solutions. For example, the storage subsystem historically has been considered a key limiting factor on scaling any solution, because hard disk drives were inherently slow and limited top system performance. With the advent of Non-Volatile Memory Express (NVMe), storage is no longer a limiting factor; a single enterprise-grade NVMe-based SSD can generally provide the IOPS needed for many applications. As a result, CPU cores and memory are now potential bottlenecks. And in HCI solutions, the network can significantly affect system performance as well.

The balanced approach takes into account all four major subsystems—CPU cores, memory, storage, and network—because every subsystem is linked to all the others. Memory is not isolated from the network, storage is not isolated from CPU cores, and so on. Peak performance and density depend on how you balance each subsystem to make sure none of them are causing a bottleneck.

It is important to note that CPU cores generate the load to storage, which in turn is dependent on east/west traffic for data replication and remote reads in HCI. So, the network for storage needs to be as fast as possible. That means it should support RDMA and have enough bandwidth to sustain traffic at 10, 25, 50, or 100 Gbps. The available memory dictates the number of VMs that can be hosted in a node; it also directly relates to the CPU NUMA design.

In conclusion, every subsystem needs to grow when others grow, otherwise bottlenecks are highly likely.

For more information on configuring App Direct-Dual Mode, refer to the following resources:

- Quick Start Guide: Provision Intel Optane persistent memory
- Understand and deploy persistent memory
- Cmdlets for configuring persistent memory devices for Hyper-V VMs

Determine How Many Nodes You Need

There is no one-size-fits-all recommendation for node count, since it depends on customer requirements and server form factor. However, node count does influence storage efficiency, replication, and fault tolerance. Table 5 provides some node count recommendations for Storage Spaces Direct. For greater storage efficiency you can choose to use dual parity instead of three-way mirror, which increases the storage efficiency (available usable capacity) with a higher node count.

Table 5. Node Count Recommendations for Fault Tolerance in Storage Spaces Direct

RESILIENCY	MINIMUM REQUIRED FAULT DOMAINS
Two-way mirror	2
Three-way mirror	3
Dual parity	4
Mixed	4

Other Recommendations

- **Storage sizing.** To determine optimal storage sizing, start with the overall usable storage capacity for all-flash HCI, then select the storage type. We recommend using a combination of Intel Optane SSDs and SATA SSDs.
- **Storage network.** To increase storage performance, we recommend using a NIC with RDMA/iWARP capabilities.

Summary

By deploying Azure Stack HCI on Intel® architecture, businesses can keep up with today's demanding big data workloads and increase data center efficiency. The reference architectures recommended in this guide support running virtualized applications on premises; Azure Stack HCI also helps enable organizations to connect to Azure for cloud services. Intel Optane technology—including Intel Optane SSDs and Intel Optane persistent memory—can help improve data throughput, achieve low latency, and increase memory capacity for a broad range of use cases (see Appendix A).

Find the solution that is right for your organization. Contact your Intel representative or learn more about **Intel Optane technology**.

Learn More

You may also find the following resources useful:

- Microsoft Azure Stack HCI
- Intel Select Solutions for Azure Stack HCI
- Intel Optane Technology
- Intel Optane SSD Series
- Intel Optane persistent memory
- Intel Optane persistent memory Brief
- 2nd Generation Intel Xeon Scalable Processors

Appendix A. Three Ways to Improve Microsoft Azure Stack HCI with Intel® Optane™ Technology

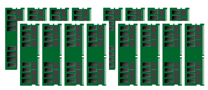
Refer to Tables 2-4 above for the bill of materials used in each option.

OPTION 1
Increase Caching Speed

*Upgrade the cache tier with **Intel Optane SSDs** to increase caching speed*

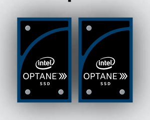
256 GB System Memory

16x 16 GB DDR4



+

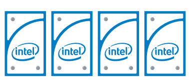
Upgrade Cache Tier to 2x Intel Optane SSDs



+

Capacity Tier

Intel SATA SSDs

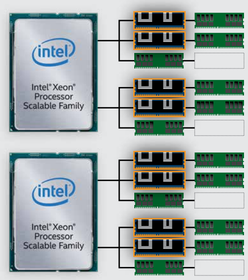


OPTION 2
Increase Available Memory

*Replace some DDR4 with **Intel Optane persistent memory** and enable **Memory Mode** to increase memory*

Upgrade to 8x 128 GB Intel Optane Persistent Memory

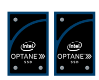
Intel Optane persistent memory (1 TB) used in 'Memory Mode' for available system memory + 12x 16 GB DDR4 for memory caching



+

Cache Tier

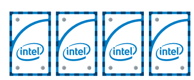
Intel Optane DC SSDs



+

Capacity Tier

Intel SATA or QLC SSDs

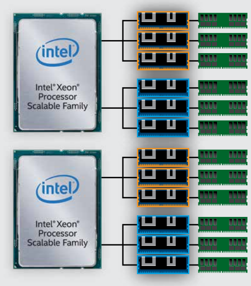


OPTION 3
Increase Memory *and* Cache Bandwidth

*Variably configure **Intel Optane persistent memory** with **Memory Mode** for increasing memory, and **App Direct-Dual Mode** for increasing cache*

Upgrade to 12x 128 GB Intel Optane Persistent Memory

- 6x used in 'Memory Mode' (768 GB) for available system memory
- 6x used in 'App Direct-Dual Mode' (768 GB) to replace cache tier for increased caching bandwidth
- 12x 16 GB DDR4 for memory caching



+

Capacity Tier

Intel QLC SSDs