Microsoft Azure

intel

Five Hybrid Cloud Use Cases for Azure Stack HCI

Table of Contents

Introduction 3

Use case 1: Remote or branch office (ROBO) and edge 6

Use case 2: Virtual Desktop infrastructure (VDI) 11

Use case 3: High-performance Microsoft SQL Server 15

Use case 4: Trusted enterprise virtualization 18

Use case 5: Azure Kubernetes service 23

Conclusion 28

Appendix 1: Hardware guidance: ROBO 29

Appendix 2: Hardware guidance: VDI 31

Appendix 3: Hardware guidance: High-performance Microsoft SQL Server 33

Appendix 4: Hardware guidance: Trusted enterprise virtualization 35

Introduction

Operate your hybrid cloud seamlessly with hyperconverged infrastructure (HCI) integration

Many businesses today continue to invest in local infrastructures and host applications on-premises, even as cloud-based alternatives continue to grow in popularity. Whether for security, compliance, or practicality, some customers prefer local hosting for workloads such as edge applications, remote desktop virtualization, databases, database engines, and storage. For these locally hosted services and applications, customers are turning to HCI to reduce costs, raise performance, and improve availability.

While considering HCI for on-premises workloads, you can leverage cloud services to supplement local infrastructure and make your HCI implementations even better. In other words, embracing HCI does not need to exclude cloud integration. Moving to the hybrid cloud is a journey, and many organizations face the challenge of finding an HCI solution that both meets their specific hosting needs and provides an optional on-ramp to cloud services—all while offering excellent performance at an affordable price.

Azure Stack HCI is a hyperconverged infrastructure host platform from Microsoft delivered as an Azure service that provides the latest security, performance, and feature updates through an Azure subscription. As a member of the Azure Stack family, Azure Stack HCI offers an excellent, affordable on-premises HCI solution, including a wide variety of deployment options and simple integration with cloud-based services in Microsoft Azure. The new Azure Stack HCI Technical Use Case program provides specific hardware configurations for Azure Stack HCI solutions that Microsoft has validated to support the most common HCI customer use cases.

Azure Stack HCI solutions

Offered through Microsoft hardware partners, Azure Stack HCI solutions are prebuilt and typically either preconfigured or bundled with simple configuration software. As of this writing, over 200 Azure Stack HCI solutions are available from over 25 Microsoft hardware partners. These partners offer hardware configurations that Microsoft has validated to ensure optimal performance and reliability for Azure Stack HCI.

Azure Stack HCI combines this validated hardware from Microsoft partners with the following software-defined datacenter components and management tools:

- **Software-defined datacenter components:**
    - **Microsoft Hyper-V** to run virtual machines (VMs) on all physical hosts
    - **Software-Defined Networking (SDN) (optional)** for network virtualization
    - **Microsoft Storage Spaces Direct** for software-defined storage
- **Management tools:**
    - **Windows Admin Center** for deployment and central, comprehensive management of local and remote servers through a graphical interface. Integrated with Azure Services for optional offsite backups, site recovery, cloud-based monitoring, and other benefits
    - **PowerShell** for scripting and automation
    - **Azure Arc** and **Azure Integration** for more management options

The complete Azure Stack HCI solution is depicted in Figure 1.

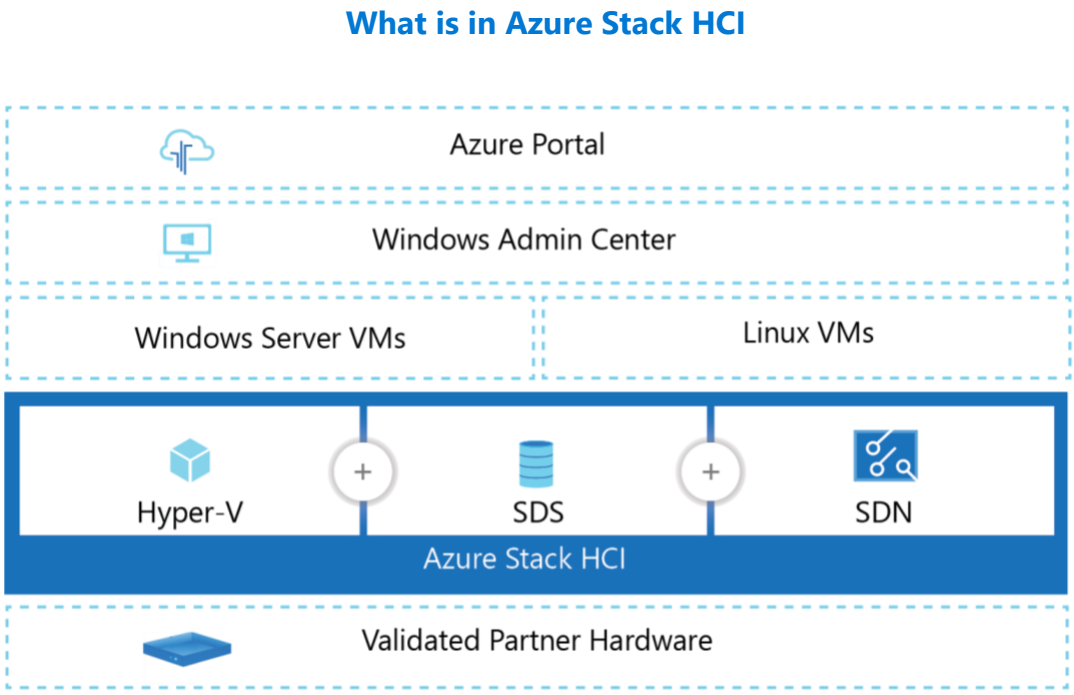


Figure 1. The components of the Azure Stack HCI solution

Introducing the Azure Stack HCI Technical Use Case program

The Microsoft Azure Stack HCI Technical Use Case program enables you to quickly find hardware configurations for Azure Stack HCI that Microsoft has validated for various needs. The most common needs correspond to the five Azure Stack HCI solution capabilities and technical use cases shown in Figure 2.

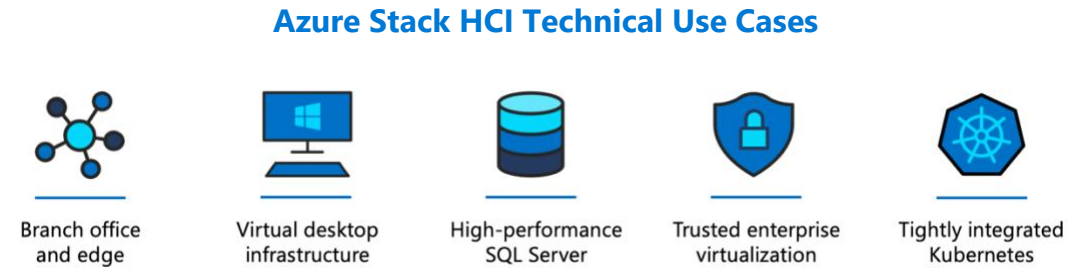


Figure 2. The Microsoft Azure Stack HCI Technical Use Case program approves hardware configurations for the five most typical technical use cases

Use case 1:

Remote or branch office (ROBO) and edge

Ensure high availability and resilient, low-cost storage for applications everywhere

The ROBO and edge technical use case for Azure Stack HCI meets the typical requirements for retail stores, branch offices, field sites, and other edge sites. Customers deploying Azure Stack HCI in this scenario are often seeking high availability and resilient storage for their applications at an affordable price, both for familiar business-critical applications and new edge workloads built on containers and Azure IoT Edge. Azure Stack HCI solutions tailored for this use case are therefore designed to offer fault tolerance and resilience in a highly cost-effective way.

Additionally, Azure Stack HCI users can now monitor all clusters from a single pane through Azure Arc. As shown in Figure 3, Azure Arc helps simplify complex and distributed environments across on-premises, edge, and multicloud, allowing you to organize, manage, and govern resources remotely at scale with scripting tools.

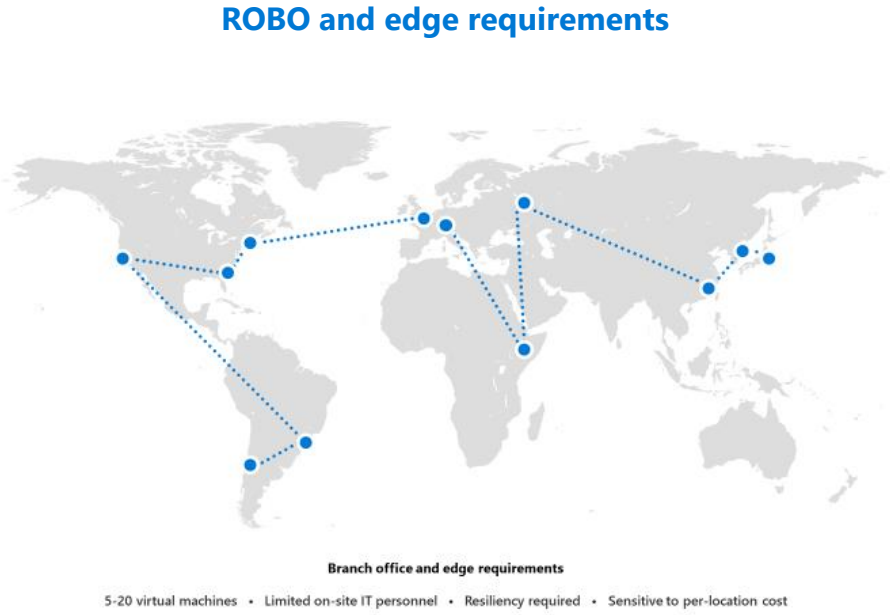


Figure 3. ROBO and edge requirements

Affordable HCI for small deployments

The price for ROBO and edge solutions for Azure Stack HCI remains low because of a minimal hardware footprint that starts at just two nodes. The small footprint also provides advantages such as space, cooling, and overall power requirement thanks to a disk or file share, known as a "witness," that keeps services running in the event an outage causes one node to fail. In a typical HCI solution, the witness resides on a remote server on-premises and is an added expense in offices where no other server is available. Azure Stack HCI eradicates the need for additional hardware by using a USB slot in a top-of-rack switch to enable a small USB drive to act as the witness. Alternatively, Azure Stack HCI can take advantage of Azure Cloud Witness to perform the same witness function for the solution.

Aside from lightweight witness technology, the two-node configuration of Azure Stack HCI reduces costs by supporting switchless networking. This type of networking enables a direct connection between the two nodes via a crossover cable, which eliminates the cost of a switch that's 10 gigabit Ethernet (GbE) or greater. With Azure Stack HCI, switchless networking is fully supported with no performance degradation, even with remote direct memory access (RDMA) for ultra-low latency and high throughput.

The new Azure Stack HCI business and billing model helps deployment, particularly in smaller environments where most operations run 12 or fewer VMs on a site. Azure Stack HCI allows you to buy smaller, 8-core clusters, enabling a significant cost savings versus traditional solutions. For instance, the Azure Stack HCI catalog features servers that utilize 4-core CPUs, and running two such servers would cost only $80 per month.

The minimal hardware requirements and cost-reducing features in two-node deployments for Azure Stack HCI make HCI affordable even for small businesses. And for companies with many sites, such as retail chains with many stores, this can add up to tremendous savings.

High availability and resiliency at the edge

Failover clustering helps ensure high availability for hosted services and applications, while nested resiliency adds even more storage resiliency to two-node deployments. With nested resiliency, a two-node Azure Stack HCI cluster can lose one node *and* one disk drive on the other node without failing or losing any data. Simply, nested resiliency offers four-way mirroring for two-node Azure Stack HCI deployments without requiring the added expense of a hardware redundant array of inexpensive disks (RAID) adapter.

Azure Monitor for ROBO and edge

Windows Admin Center is the main administration console for Azure Stack HCI and makes a number of Azure services available to facilitate Azure Stack HCI administration through the Azure hybrid services page, found in the navigation menu. Azure Monitor, shown in Figure 4, is one such service that's particularly beneficial to ROBO and edge use.

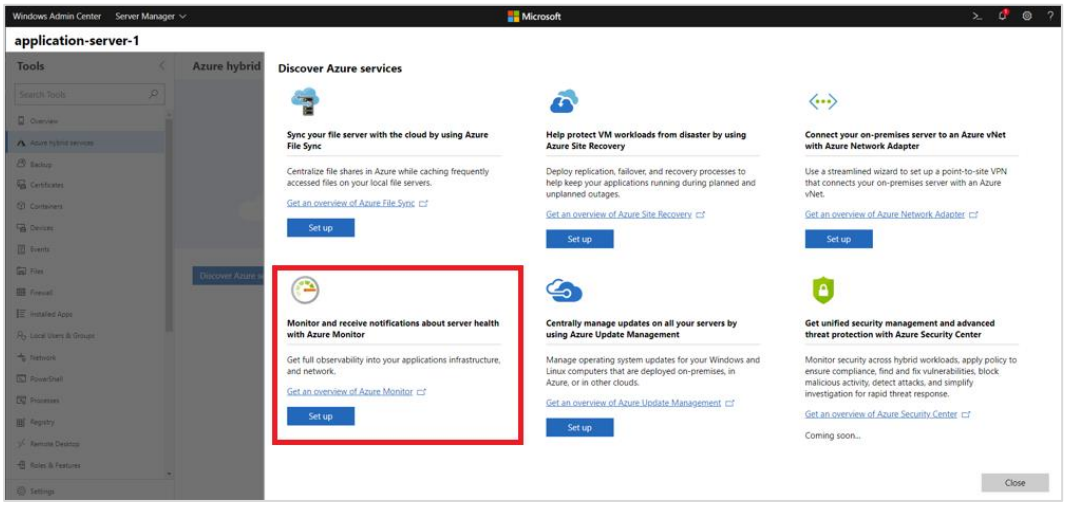


Figure 4. You can access Azure Monitor and other hybrid services through the Windows Admin Center

Azure Monitor provides a comprehensive way to collect, analyze, and respond to data in your infrastructure. It also offers insight on how applications are performing, allowing for proactive identification of issues affecting those applications and the resources they depend on.

Azure Monitor is especially useful for ROBO and edge scenarios because it enables remote IT staff to monitor Azure Stack HCI and take action as needed. For example, IT staff can configure alerts and dashboards to remotely monitor the health of Azure Stack HCI nodes, guest VMs, and applications. Azure Stack HCI offers an additional benefit for ROBO scenarios that allow you to monitor and control all clusters from one view.

For more information about Azure Monitor, visit "Monitor servers and configure alerts with Azure Monitor from Windows Admin Center."

Intel® technologies and Select Solutions for ROBO deployments

ROBO and edge scenarios for Azure Stack HCI typically call for an inexpensive, simple, and reliable solution, but some edge deployments run analytics on local data and require a solution that delivers better performance. To meet these requirements, Azure Stack HCI solutions can support single-socket servers that run on Intel® Xeon® Scalable processors with as few as four cores and a frequency as low as 1.4 GHz with a broad range of SKUs available to scale to meet requirements. For storage, Intel® solid state drives (SSDs) are a fast, reliable, and inexpensive solution that can be used in a single tier, or, as an even lower-cost option, a cache tier to support hard disk drives.

Intel has defined Intel Select Solutions for Azure Stack HCI, including reference designs for an edge scenario. Intel Select Solutions are predefined and verified combinations of Intel® compute, memory, storage, and network products designed to support specific workloads in basic ("Base") and advanced ("Plus") configurations in different edge use cases. For more information about Intel Select Solutions for Azure Stack HCI, read the "Intel Select Solutions for Microsoft Azure Stack HCI" brief.

Resource planning for the ROBO and edge technical use case

For more information about typical workload demands and resource requirements for the ROBO and edge use case, see Appendix 1.

Use case 2:

Virtual desktop infrastructure (VDI)

Easily and predictably scale resources to meet your desktop virtualization needs

For the virtual desktop infrastructure (VDI) use case, an Azure Stack HCI cluster is used to implement remote desktop virtualization on a large scale. VDI delivers user desktops through a virtual desktop broker, such as Microsoft Remote Desktop Services, Citrix Virtual Apps and Desktops (formerly Citrix XenDesktop), or VMware Horizon. These virtual desktops connect back to VMs and central storage on the Azure Stack HCI cluster. A significant advantage of VDI is security; it offers businesses a secure way to deliver client desktops on a wide range of devices without allowing users to store any data locally or upload any data from those local devices.

VDI also offers significant scalability. The new Azure Stack HCI scales performance and number of desktops nearly linearly as more nodes are added, making capacity planning, forecasting, and purchasing additional nodes much easier as your business grows. This is a real advantage over competing options, which do not scale linearly.

HCI provides a reliable platform for VDI. A VDI workload's demand for resources is often proportional to the number of users requiring desktop virtualization, and HCI enables you to scale resources easily and predictably to meet these needs. In a similar way, HCI hardware solution providers can accurately predict which hardware configuration is suitable for your VDI use case based on the number of users you need to support. The maximum number of nodes for a single site is 16.

Running VDI workloads on Azure Stack HCI offers additional advantages. If you're running a Windows operating system (which is fairly common), both the guest and host can leverage shared technologies from Microsoft, enabling administration, performance, and technical support advantages. The shared platform between host and guest, in addition, allows native access to other network resources, such as file shares stored on the Azure Stack HCI cluster.

Azure services support the VDI use case

Windows Admin Center provides access to two cloud-based services that are especially useful for VDI: Azure Update Management and Azure Security Center. Azure Update Management, shown in Figure 5, greatly eases the administrative overhead associated with updating the many client VMs hosted on Azure Stack HCI. Through its cloud interface, Azure Update Management lets you quickly assess the status of available updates on all client VMs hosted in the Azure Stack HCI cluster and manage the installation of those updates. Azure Security Center is a unified infrastructure-security management system that strengthens the security posture of your datacenters and provides advanced threat protection across all your machines.

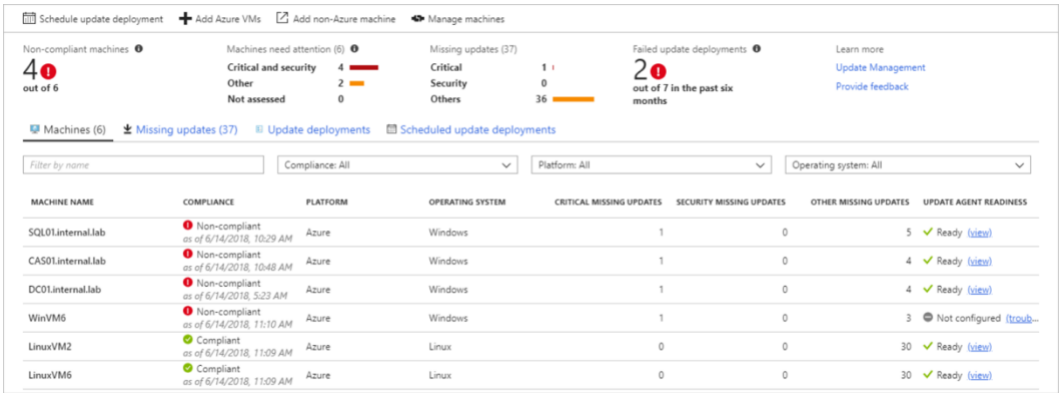


Figure 5. Viewing update assessments in Azure Update Management

For information about the Azure Security Center, see "Azure Security Center." For more information on Azure Update Management, see "Update Management solution in Azure."

Intel® technologies to support the VDI use case

VDI is a compute-intensive workload. To ensure a productive user experience, it's essential to provide adequate virtual CPU resources to support workers and applications. Intel Xeon Scalable processors offer a wide range of options to allow you to tailor your infrastructure to meet your users' needs. Azure Stack HCI also supports Intel® Optane™ persistent memory modules, which increase efficiency for VDI deployments and provides affordable memory and support for data persistence. You can use it in memory mode to help increase the number of user VMs that can be hosted per node, at a lower cost per VM than is possible with all-dynamic random-access memory solutions.*

For more information on Intel Optane persistent memory, visit the Intel Optane persistent memory website.

Resource planning for the VDI technical use case

For information about the typical workload demands and resource requirements for the VDI use case, see Appendix 2.

*Note that persistent memory requires a 2nd Generation Intel Xeon Scalable processor.

Use case 3:

High-performance Microsoft SQL Server

Run Microsoft SQL Server on a trusted platform optimized for performance

Azure Stack HCI offers a high-performing, scalable, and manageable platform for implementing Microsoft SQL Server. Excellent performance is crucial for SQL Server deployments. Given this requirement, it's important to note that Azure Stack HCI is already the best-performing HCI solution on the market[1] at an affordable price point. Azure Stack HCI also offers flexibility; it can host Windows VMs and Linux VMs, giving you the freedom to run the Windows or Linux version of SQL Server depending on your needs.

> **"The [Azure Stack HCI] cluster we've been testing has posted tremendous numbers, the fastest we've seen in a mid-market four-node HCI cluster."[1]**
>
> — StorageReview.com

When used to host SQL Server, Azure Stack HCI enables a single vendor to manage the hypervisor, host operating system, and database server. This capability allows the underlying code to be optimized for performance, and it also allows a single party to be responsible for resolving issues when they appear.

Azure Site Recovery offers a business continuity and disaster recovery (BCDR) strategy

Azure Site Recovery offers a business continuity and disaster recovery (BCDR) strategy. Setting up disaster recovery for SQL Server allows you to protect the SQL Server's back end of an application to help protect data and keep apps and workloads online when outages occur, whether planned or unplanned.

Azure Stack HCI also comes with native stretch clustering capability for BCDR. Stretch clustering provides automatic failover to restore production quickly without the need for manual intervention. And Storage Replica enables the replication of volumes across sites for disaster recovery with all servers staying in sync. Storage Replica supports both synchronous and asynchronous replication.

For more information on Azure Site Recovery, see "Azure Site Recovery."

Intel technologies to support the high-performance SQL Server use case

Azure Stack HCI deployments built to support database servers are typically optimized for performance. To support the highest performance, you can use high-core-count and high-frequency Intel Xeon Scalable processors, coupled with high-bandwidth Intel SSDs with NVM Express (NVMe) in a single storage tier, including the Intel® P4610 SSD. Larger clusters would benefit from 25GbE Intel® Ethernet adapters with support for RDMA, such as the Intel® Ethernet 800 series.

Resource planning for the high-performance SQL Server technical use case

For information about typical workload demands and resource requirements for the high-performance SQL Server use case, see Appendix 3.

Use case 4:

Trusted enterprise virtualization

Create a highly secure infrastructure for apps and workloads, no matter the size of your organization

This technical use case for Azure Stack HCI is broad and can overlap all other use cases. While this is an enterprise-class use case, it's by no means exclusive. It can also apply to small, medium, and mid-market companies, as it pertains to general scenarios in which customers are looking to serve applications and workloads hosted in VMs with both high security and high availability. Azure Stack HCI meets customer requirements for trusted enterprise virtualization by providing a highly secure infrastructure for workloads through virtualization-based security (VBS).

VBS uses the Hyper-V hypervisor to create and isolate a secure region of memory, virtual secure mode (VSM), from the normal operating system. When VBS is enabled, security-sensitive operations can occur in this secure memory enclave, independent of the host operating system. The host operating system sees the specific processes and their associated memory in VSM as belonging to a separate operating system. Even if malware gains access to the operating system kernel, these increased protections can greatly limit and contain the possible exploits because the hypervisor can prevent the malware from executing code or accessing platform secrets.

To ensure the highest level of security for workloads, all hardware solutions certified for Azure Stack HCI are ensured to meet the requirements needed for VBS. The relationship of the hypervisor, VSM, and the host operating system are illustrated in Figure 6.

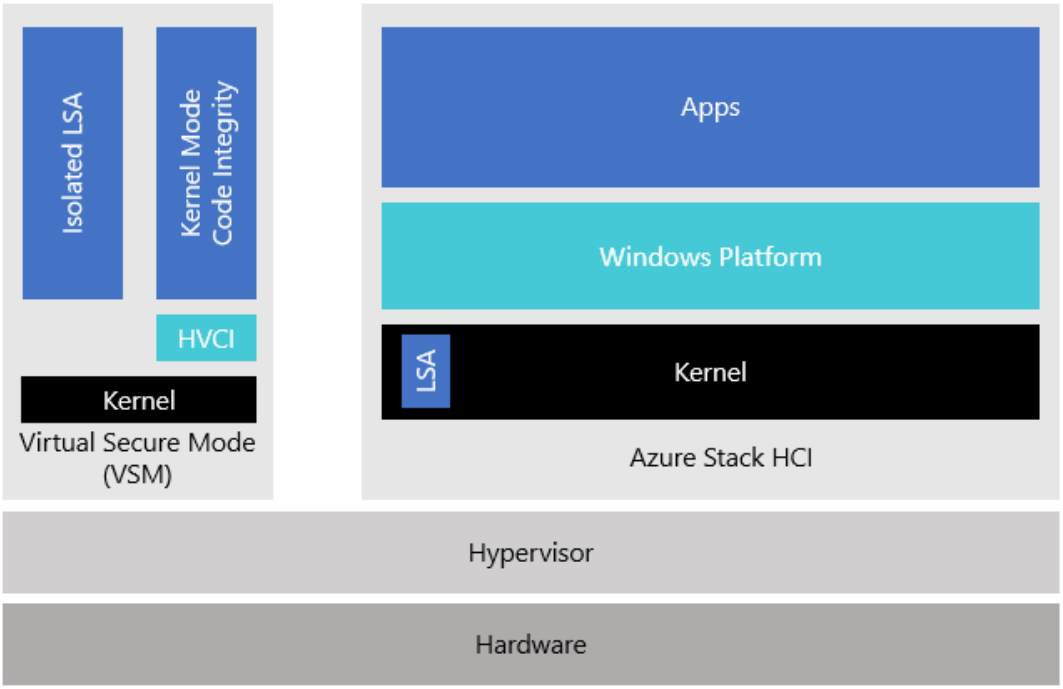


Figure 6. Virtualization-based security (VBS) creates a secure memory enclave that isolates sensitive code and processes from the host operating system

VBS enables two features: hypervisor-enforced code integrity (HVCI) and Credential Guard. HVCI uses VBS to significantly strengthen code-integrity policy enforcement. For reference, "code integrity" is a threat-protection feature that checks the drivers and system files on your device for signs of corruption or malicious software. HVCI uses VBS to run the code-integrity service inside VSM, providing stronger protections against kernel viruses and malware before they're loaded in the host operating system. Credential Guard isolates user sign-in information in VSM to protect this sensitive data in case the system is compromised. Both of these features also take advantage of an installed Trusted Platform Module (TPM) chip for root of trust, which is part of all Azure Stack HCI hardware solutions.

For protected workloads on VMs to fulfill their security promise, it's imperative that the host machine is equipped with a TPM chip, and that the host has enabled both VBS and HVCI. Figure 7 illustrates the advantages Azure Stack HCI solutions have over solutions that do not use a TPM chip for root of trust. On the bottom left, a secure host is protected through a TPM chip, secure boot, and VBS. Such a host is able to provide a secure foundation for protected VMs that run on top of it (top left). Unprotected hosts (bottom right) are unable to pass along the same level of security to VMs, even if the guest is protected with a virtual TPM (vTPM).

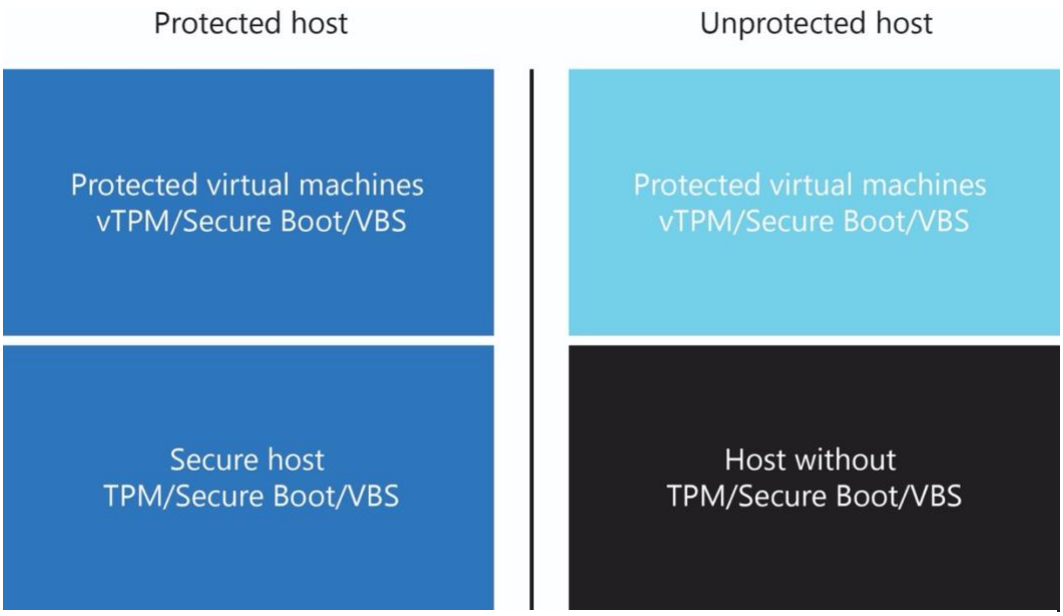


Figure 7. All validated Azure Stack HCI solutions provide a high level of protection to VMs through a physical TPM

For more information about VBS, see "Virtualization-based Security (VBS)."

Azure Security Center enhances the trusted enterprise virtualization use case

To enhance Azure Stack HCI deployments requiring high security, you can utilize Azure Security Center features integrated in Windows Admin Center. Connecting your Azure Stack HCI cluster to Azure Security Center gives you a comprehensive view of the security status of your environment and its vulnerability to attacks. It also lets you monitor access control according to security policies configured in Azure. Note that bringing on-premises machines within the scope of Azure Security Center monitoring requires an agent to install them, which is facilitated through Windows Admin Center.

For more information on Azure Security Center, see "Azure Security Center."

Resource planning for the trusted enterprise virtualization technical use case

For information about typical workload demands and resource requirements for the trusted enterprise virtualization use case, see Appendix 4.

Use case 5:

Azure Kubernetes service

Run containerized workloads on-prem with simplicity and confidence

This technical use case details Azure Stack HCI as an ideal solution for deploying Kubernetes-based workloads. Azure Kubernetes Service on Azure Stack HCI (AKS-HCI) is a Kubernetes-based orchestrator that automates running containerized applications on clusters that use Azure Stack HCI. Orchestrators such as the open-source Kubernetes automate much of the work involved with deploying and managing multiple containers. However, Kubernetes can be complex to set up and maintain. AKS-HCI helps simplify setting up Kubernetes on-premises, making it quicker to start hosting Linux and Windows containers. By bringing the power of AKS from Azure to your on-premises locations, you benefit from the advanced security, management, and deployment experiences for running your workloads, all while reducing the complexity and learning curve for your users.

AKS-HCI is ideal for running containerized workloads within on-premises environments to meet network or regulatory goals, whether you're operating from an edge remote site or the datacenter. With an existing Azure Stack HCI cluster in place, and after validating several prerequisites, Windows Admin Center guides you through configuring an AKS management cluster. This cluster is responsible for provisioning and managing target clusters where your workloads run, and includes an API server and a load balancer. The wizard streamlines deployment of all key components, but PowerShell is also an option for this step.

Azure Kubernetes Service on Azure Stack HCI

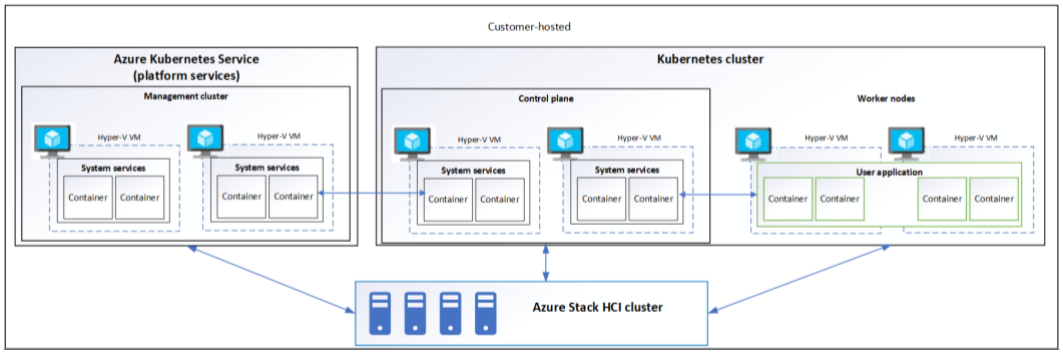


Figure 8. Azure Kubernetes Service on Azure Stack HCI Architecture

With the management cluster in place, you can now deploy your workload clusters, which will ultimately run your workloads. Windows Admin Center streamlines this experience, collecting information related to cluster and host selection, node sizing, number of Windows and Linux nodes, storage and network info, and optionally, Azure Arc connectivity. With these selections in place, upon clicking create, you'll have a new Kubernetes workload cluster ready for your applications in just a few minutes. For those who have embraced AKS in Azure, you'll notice the steps in Windows Admin Center mirror those native to AKS in Azure, but if you prefer defining your infrastructure programmatically, PowerShell can also be used to deploy these workload clusters quickly and easily.

For workload clusters created with Azure Arc integration enabled, you can navigate to the Azure Portal to visualize and manage your new Kubernetes workload clusters, side by side with your native AKS clusters running in Azure.

To deploy your workloads, you can harness the power of Azure Arc and its GitOps approach to centrally deploying applications and configurations to Kubernetes environments. Alternatively, you can use your existing tools and repositories to deploy your containerized applications. From there, you can integrate with additional Azure services, such as Azure Monitor, to gain insights into your Kubernetes clusters and applications. This solution also forms a powerful platform for running additional Microsoft applications and services, such as Azure Arc enabled data services.

With AKS-HCI, organizations benefit from an end-to-end, Microsoft-supported solution, from virtualization layer to containerized workloads, with native integration with Azure services. Through its streamlined deployment and scaled management, the combination of Azure Stack HCI and AKS-HCI provides improved time to value and a highly scalable, reliable, and secure solution to run traditional and containerized Windows and Linux-based workloads.

Intel software features that enhance Azure Stack HCI solutions

Beyond the Intel hardware components used to support Azure Stack HCI, Intel also provides many software components to support performance, reliability, and security in Azure Stack HCI solutions:

Intel® Deep Learning Boost (Intel® DL Boost), available on 2nd Generation Intel® Xeon® Scalable processors, takes embedded artificial-intelligence (AI) performance to the next level. Intel Xeon Scalable processors are built specifically for the flexibility to run complex AI workloads on the same hardware as your existing workloads. With Intel DL Boost, some enterprises have experienced a 57 percent performance improvement.[2]

Intel® Advanced Vector Extensions 512 (Intel® AVX-512) is a set of CPU instructions that impacts compute, storage, and network functions. The number 512 refers to the width, in bits, of the register file, which sets the parameters for how much data a set of instructions can operate on at a time. Intel AVX-512 enables twice the number of floating-point operations per second (FLOPS) per clock cycle compared to its predecessor, Intel AVX2.[3]

Intel® Run Sure Technology delivers advanced reliability, availability, and serviceability, adding more resilience to Intel Xeon Scalable processors and helping to ensure the highest levels of uptime for your mission-critical workloads.[4]

Intel® Trusted Execution Technology (Intel® TXT) with one-touch activation is a powerful component of enterprise data protection.4 Intel TXT creates a hardware root of trust and a measured launch environment, which helps ensure that your server is running "known-good" configurations of your critical software components (firmware, BIOS, operating system, and hypervisors).

Licensing requirements

Note that the following licensing section applies to all use cases. The licensing needed for the trusted enterprise virtualization use case will vary based on intended workload. A subscription is required for all Azure Stack HCI operations, because Azure Stack HCI functions as the host in all cases. Linux applications and open source software do not require additional purchases. Running a few Windows Server roles or applications requires only a standard Windows Server 2019 license. Running unlimited Windows Server roles or applications requires a Windows Server 2019 Datacenter license. These options allow you to select the best licensing option for your specific operations and budgets.

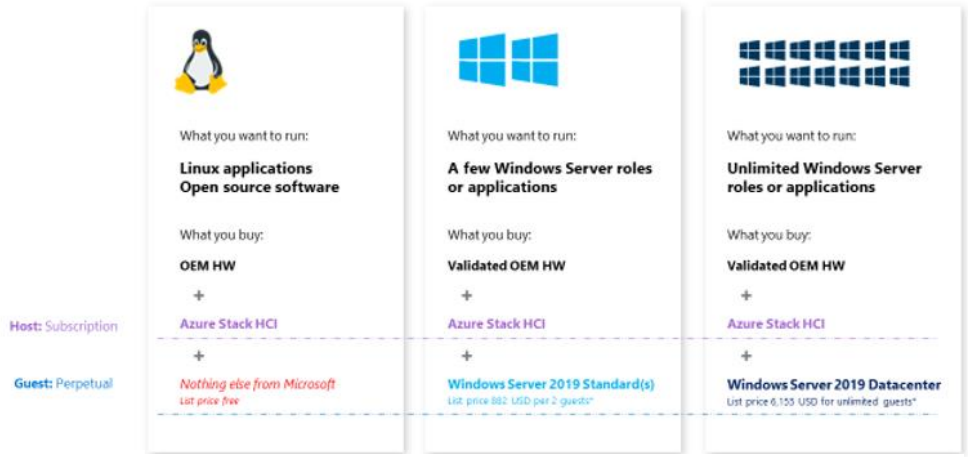


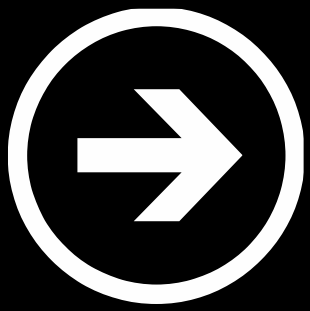
Figure 9. Azure Stack HCI host & guest licensing

Conclusion

Get the best-performing HCI solution to suit your needs, amplified by hybrid services

Azure Stack HCI offers industry-leading performance for HCI for $10 per month, per core activated. Azure Stack HCI is also the only HCI solution that includes Azure hybrid services to ease administration, monitoring, adoption, and secure configuration.

You can use the Technical Use Case program to easily identify Azure Stack HCI solutions that are built for your requirements. Whether you need an HCI solution to run edge workloads, VDI, SQL Server, highly secure workloads, or a Kubernetes cluster, you can find an Azure Stack HCI solution that meets your requirements on-premises while also offering optional hybrid services through Azure.



To find Azure Stack HCI solutions from hardware vendors that have been approved for the most common customer needs for HCI, view the Azure Stack HCI catalog.

Try Azure Stack HCI for free

Appendix 1:

Azure Stack HCI for ROBO and edge—hardware guidance

The ROBO and edge technical use case is typified by a small, two-node footprint, with low CPU- and storage-capacity requirements. However, a fairly large amount of RAM per node (for example, 384 GB) is recommended to optimize performance. The recommended storage design is single-tier, with all NVMe solid state drives (SSDs). For more information, see Table 1.

Hardware guidance for ROBO and edge

Workload assumptions

- Tenants: Small number of mixed tenants
- Containers: Smaller resource footprint than virtual machines (VMs)
- VMs: Small to medium in size for traditional business-application stacks
- Performance per tenant: Resources fall somewhere between the virtual desktop infrastructure (VDI) and SQL Server use cases

Design guidance

- Keep physical footprint small and optimize cost for desired level of performance (this typically will not resemble a full datacenter deployment; for example: switchless, two-node).
- Balance performance and the cost of the hardware bill of materials.

Hardware considerations (per node)

CPU requirements

Core counts:	Frequency:	> 1 TB memory/CPU?
Low–medium	Low–medium	No

- Start with a general-purpose CPU with sufficient headroom for Storage Spaces Direct
- No need for large amounts of memory per socket

Storage requirements

Capacity:	Performance/tiered layout:
Low	One-tier, all NVMe

Account for nested resiliency when computing raw needs

<table>
<tr><td rowspan="5"></td><td colspan="4">(Needs vary based on resiliency model employed)</td></tr>
<tr><td colspan="4"><h2>Memory requirements</h2></td></tr>
<tr><td colspan="2">Capacity:</td><td colspan="2">Performance:</td></tr>
<tr><td colspan="2">High</td><td colspan="2">Medium</td></tr>
<tr><td colspan="2">256 GB</td><td colspan="2">Possible candidate for Intel® Optane™ persistent-memory (PMem) for designs that need more memory</td></tr>
<tr><td colspan="5"><h2>Network (east/west) requirements</h2></td></tr>
<tr><td colspan="5">Bandwidth: Medium
Driven by storage configuration
Switchless option</td></tr>
<tr><td rowspan="2">Example node configuration</td><td>CPU</td><td>Storage</td><td>Memory</td><td>Network</td></tr>
<tr><td>2 x Intel® Xeon® Gold 5218 processors</td><td>Boot: 2 x Intel® SSD D3-S4510 (M.2)

Capacity: 4 x Intel SSD DC P4610</td><td>16 x 16 GB DDR4</td><td>2 x 10 gigabits per second (Gbps) or 2 x 25 Gbps

Intel® Ethernet Network Adapter X722 or Intel Ethernet Network Adapter 800 series</td></tr>
</table>

Table 1. Hardware guidance for ROBO and edge technical use case

Appendix 2:

Azure Stack HCI for VDI—hardware guidance

The hardware requirements for VDI tend to scale in a fairly predictable manner as the number of users grows. In a typical implementation, CPU and storage capacity needs are average, but memory requirements are high. For more information, see Table 2.

Hardware guidance for VDI
Workload assumptions User types are mixed, but this example provides an overview. Access further guidance from Microsoft. • Light: 1 vCPU, 4 GB RAM, 60 GB disk, 20 average input/output operations per second (IOPS) • Heavy: 2 vCPU, 8 GB RAM, 80 GB disk, 30 average IOPS • Power: 4 vCPU, 16 GB RAM, 100 GB disk, 50 average IOPS
Design guidance • Target 40 users per node (10 light, 10 heavy, and 20 power users) with minimal overcommit. • Enable deduplication and compression. • Maintain service during boot and sign-in storms. • Leave extra resources during times of maintenance/node failures.

<table>
<tr><td rowspan="8">Hardware considerations (per node)</td><td colspan="4">CPU requirements</td></tr>
<tr><td>Core counts:</td><td>Frequency:</td><td colspan="2">> 1 TB memory/CPU?</td></tr>
<tr><td>Medium</td><td>Medium–high</td><td colspan="2">No</td></tr>
<tr><td colspan="4">• 110 vCPU (minimum for workloads; not accounting for Azure Stack HCI infrastructure roles)
• Performance per core a consideration
• No need for large memory per socket</td></tr>
<tr><td colspan="4">Storage requirements</td></tr>
<tr><td colspan="2">Capacity:</td><td colspan="2">Performance/tiered layout:</td></tr>
<tr><td colspan="2">Medium</td><td colspan="2">Two-tier, all-flash</td></tr>
<tr><td colspan="4">• 15,000 IOPS sufficient for boot/sign-in storms
• 3.4 TB usable capacity (10.2 TB raw)
• ~1 TB of cache</td></tr>
<tr><td colspan="4">Memory requirements</td></tr>
<tr><td colspan="2">Capacity:</td><td colspan="2">Performance:</td></tr>
<tr><td colspan="2">High</td><td colspan="2">Medium</td></tr>
<tr><td colspan="2">Example: 768 GB</td><td colspan="2">Candidate for PMEM</td></tr>
<tr><td colspan="4">Network (east/west) requirements</td></tr>
<tr><td colspan="4">Bandwidth: Medium (driven by storage configuration)</td></tr>
<tr><td rowspan="2">Example node configuration</td><td>CPU</td><td>Storage</td><td>Memory</td><td>Network</td></tr>
<tr><td>2 x Intel® Xeon® Gold 6240 processors</td><td>Boot: 1 x Intel® SSD D3-S4510 (M.2)

Cache: 4 x Intel® Optane™ SSD P4800X (U.2, 375 GB)

Capacity: 4 x Intel® SSD D3-S4510 (2.5”, 3.84 TB)</td><td>12 x 16 GB DDR4

6 x 128 GB Intel® Optane™ persistent memory</td><td>2–4 x 25 Gbps Intel® 800 Series</td></tr>
</table>

Table 2. Hardware guidance for the VDI use case

Appendix 3:

Azure Stack HCI for high-performance SQL Server—hardware guidance

For the SQL Server use case, CPU, storage, and memory requirements all remain high in order to optimize performance. To optimize storage performance in particular, a single tier with only NVMe SSDs is recommended. For more information, see Table 3.

Hardware guidance for high-performance Microsoft SQL Server

Workload assumptions

Modeled loosely on Azure DSv2-series approach:

- 4 vCPU, 32 GB RAM, 1 TB disk, 50,000 IOPS, 500 MB/s throughput

Design guidance

- Consistent high performance is the number-one design goal.
- Avoid oversubscription scenarios to maintain quality of service.
- Tune Storage Spaces Direct for storage performance versus storage efficiency (for example, you need three nodes for three-way mirroring).
- Host up to 12 SQL Server instances per node.

Hardware considerations (per node)

CPU requirements

Core counts:	Frequency:	> 1 TB memory/CPU?
High	High	No

- Sufficient cores for both SQL Server and driving storage use
- Performance per core a factor with SQL Server licensing
- No need for large memory per socket

Storage requirements

Capacity:	Performance/tiered layout:
High	One tier, all NVMe

- Support >= 600,000 input/output operations per second, per node, for user workloads
- 12 TB usable capacity (36 TB raw). A minimum of three nodes are needed to have performance storage tier (three-way mirror).

	• Medium-endurance solid state drives (SSDs) required in a single-tier design
	Memory requirements
	Capacity: / **Performance:**
	Medium–high / **High**
	Example: 768 GB / Example: CPU with 2,933 MHz DDR4 support Candidate for Optane persistent memory
	Network (east/west) requirements
	Bandwidth: Medium (driven by storage configuration)

Example node configuration	CPU	Storage	Memory	Network
	2 x Intel® Xeon® Platinum 8268 processors	Boot: 1 x Intel® SSD D3-S4510 (M.2) Capacity: 12 x Intel® SSD DC P4610 (2.5", 3.2 TB)	24 x 32 GB DDR4 Or 4-8 x 128 GB Intel® Optane™ persistent memory	2–4 x 25 Gbps to 2 x 100 Gbps Intel® Ethernet Network Adapter E810

Table 3. Hardware guidance for the high-performance SQL Server use case

Appendix 4:

Azure Stack HCI for trusted enterprise virtualization—hardware guidance

The trusted enterprise virtualization use case is the most variable in terms of its hardware needs. One requirement, however, is a TPM 2.0 module. You can use the general guidance provided in Table 4, combined with the guidance for the other use cases, to help you determine the hardware configuration most suitable for your workload.

Azure Stack HCI for trusted enterprise virtualization

Workload assumptions

Mix of tenants and workloads that are security-sensitive:

- VMs: Small to large (various sizes to support traditional business application stacks)
- Workloads: Can run the gamut compared to homogeneous VMs
- Performance per tenant: Resources range between ROBO/edge and SQL Server

Design guidance

Establish desired profiles of tenant types, resource needs, and mix. Then model on a per-node basis: either design hardware components to meet the desired number of tenants, or start with hardware components and arrive at the number of supported tenants.

Hardware considerations (per node)

CPU requirements

Core counts:	Frequency:	> 1 TB memory/CPU?
Medium	Medium	No

- Meet requirement for TPM 2.0 module (needed for all use cases).
- Start with a general-purpose CPU with sufficient headroom for Storage Spaces Direct.
- No need for large amounts of memory per socket.

Storage requirements

Capacity:	Performance/tiered layout:
Medium	Two-tier, all-flash

- Use ROBO/edge and SQL Server use cases as bookends/guardrails
- Ensure cache remains between 5-10 percent of raw capacity

<table>
<tr><td rowspan="5"></td><td colspan="4">Memory requirements</td></tr>
<tr><td colspan="2">Capacity:</td><td colspan="2">Performance:</td></tr>
<tr><td colspan="2">Medium</td><td colspan="2">Medium</td></tr>
<tr><td colspan="2">Example: 384GB or 768GB</td><td colspan="2">Candidate for PMEM</td></tr>
<tr><td colspan="4">Network (east/west) requirements</td></tr>
<tr><td></td><td colspan="4">Bandwidth: Medium (driven by storage configuration)</td></tr>
<tr><td rowspan="2">Example node configuration</td><td>CPU</td><td>Storage</td><td>Memory</td><td>Network</td></tr>
<tr><td>2 x Intel® Xeon Gold 6230 processors</td><td>Boot: 1 x Intel® SSD D3-S4510 (M.2)

Cache: 2 x or more Intel® Optane™ SSD P4800X

Capacity: 4 x or more Intel® SSD D3-S4510</td><td>12 x 16 GB DDR4

4-8 x 128 GB Intel® Optane™ Persistent Memory</td><td>2–4 x 25 Gbps to 2 x 100 Gbps

Intel Ethernet Network Adapter E810</td></tr>
</table>

Table 4. Hardware guidance for the trusted enterprise virtualization use case

Microsoft Azure

---

1 StorageReview. "Microsoft Azure Stack HCI Review (DataON HCI-224 with Intel Optane NVMe)." September 2019. www.storagereview.com/microsoft_azure_stack_hci_review_dataon_hci224_with_intel_optane_nvme.

2 Intel. "MLPerf Results Validate CPUs for Deep Learning Training." December 2018. www.intel.ai/mlperf-results-validate-cpus-for-dl-training/#gs.4f76xm.

3 To learn more about Intel AVX-512, visit: Intel. "Intel Advanced Vector Extensions 512 (Intel AVX-512)." www.intel.com/content/www/us/en/architecture-and-technology/avx-512-animation.html.

4 No computer system can provide absolute reliability, availability, or serviceability. Requires an Intel Xeon processor with Intel Run Sure Technology. Built-in reliability features available on select Intel processors may require additional software, hardware, services, and/or an internet connection. Results may vary depending upon configuration. Consult your system manufacturer for more details.