Microsoft Azure | intel

White paper

Updated Technical use cases white paper for Azure Stack HCI - December 2020

Many businesses today continue to make investments in their local infrastructures and host selected applications on premises, even as cloud-based alternatives are growing in popularity. Whether for reasons of security, compliance, or simple practicality, local hosting is sometimes seen as the best choice for workloads such as edge applications, remote desktop virtualization, databases and database engines, and storage. For these locally hosted services and applications, customers have especially been turning to hyperconverged infrastructure (HCI) as a means to reduce costs, raise performance, and improve availability.

But as customers are increasingly considering HCI to gain efficiencies for their on-premises workloads, they can still also plan to draw upon cloud services to supplement local infrastructure and make their HCI implementations even better. Embracing HCI, in other words, does not need to exclude cloud integration. Hybrid cloud is a journey, and the challenge for many organizations is to find an HCI solution that not only is well-suited to their specific hosting needs, but that also provides a simple, optional on-ramp to cloud services—all while offering excellent performance at an affordable price.

Microsoft Azure Stack HCI (AzS HCI) is a hyperconverged infrastructure host platform from Microsoft delivered as an Azure service providing the latest security, performance, and feature updates through an Azure Subscription. As a member of the Azure Stack family of technologies, AzS HCI offers an excellent price/performance ratio as an on-premises HCI solution, with a wide variety of deployment options, while also providing optional (and easy) integration with cloud-based services in Microsoft Azure. What's more, you can use the new AzS HCI Technical Use Case program (described below) to find particular AzS HCI solutions from hardware vendors that Microsoft has validated to support the most common customer use cases for HCI.

The Azure Stack HCI solution

The AzS HCI solutions are offered through Microsoft hardware partners, prebuilt, and typically either preconfigured or bundled with simple configuration software. Over 200 AzS HCI solutions are available today from over 25 Microsoft hardware partners (these numbers are often refreshed but are provided to represent scope). These partners offer hardware configurations that Microsoft has validated to ensure optimal performance and reliability for AzS HCI.

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

- **Hyper-V** to run virtual machines (VMs) on all physical hosts
- **Software Defined Networking (SDN) (optional)** for network virtualization
- **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
    - **Azure services** integrated into Windows Admin Center for optional offsite backups, site recovery, cloud-based monitoring, and other benefits
  - **PowerShell** for scripting and automation
  - **Azure Arc** and **Azure integration** present a new level of management options

The complete AzS HCI solution is depicted graphically in Figure 1.

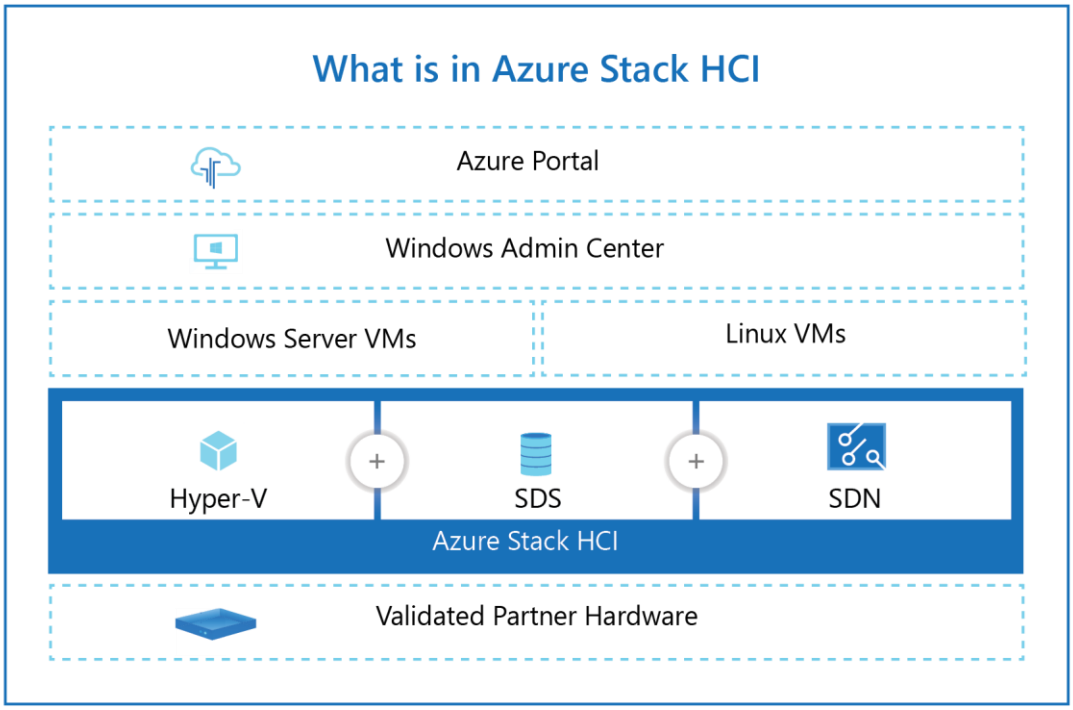


Figure 1. The components of the Azure Stack HCI solution

Introducing the Azure Stack HCI technical use cases

The Microsoft AzS HCI Technical Use Case program enables you to quickly find hardware configurations for AzS HCI that Microsoft has validated for various needs. The most common needs correspond to the five AzS HCI solution capabilities and technical use cases shown in Figure 2. Through the Azure Stack HCI Technical Use Case program, Microsoft validates AzS HCI solution offerings to determine the suitability of their hardware configurations in supporting any of these five scenarios.

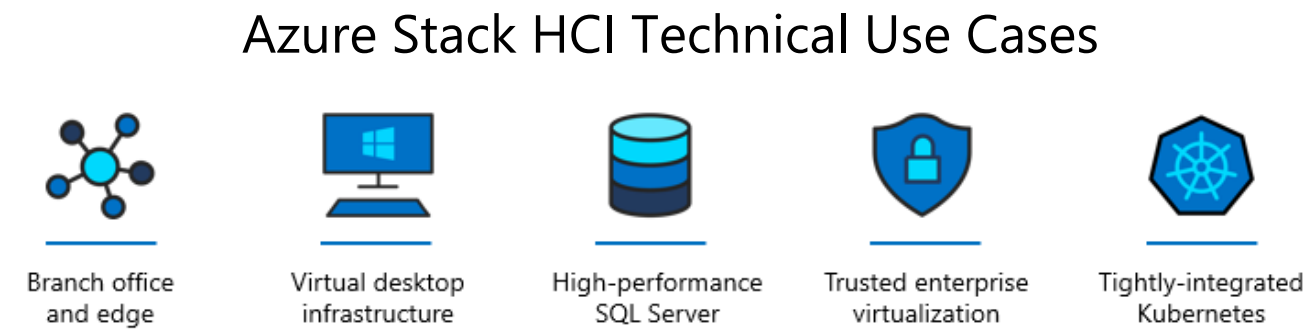


Figure 2. The Microsoft AzS HCI Technical Use Case program approves hardware configurations for the five most typical technical use cases

The following sections describe each of these five technical use cases in more detail.

Technical use case 1: Azure Stack HCI for remote or branch office (ROBO) and edge

The remote or branch office and edge technical use case for AzS HCI meets the typical requirements for retail stores, branch offices, field sites, and other edge sites. Customers deploying AzS 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 for new edge workloads built on containers and Azure IoT Edge. AzS HCI solutions tailored for this use case are therefore designed to offer fault tolerance and resilience in a highly cost-effective way.

Affordable HCI for small deployments

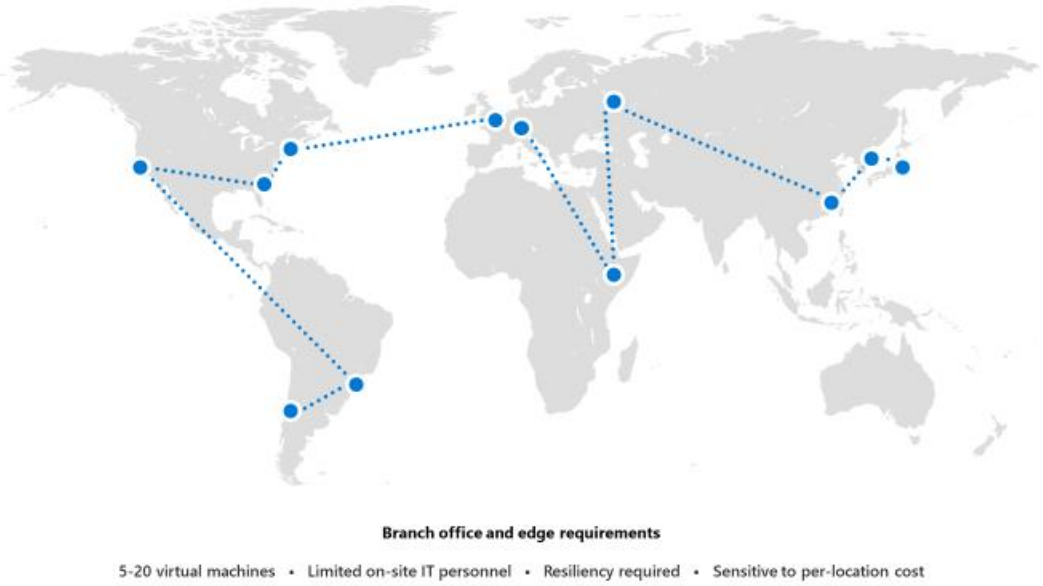
The price for ROBO and edge solutions for AzS HCI can remain low thanks to a minimal hardware footprint that starts at just two nodes. This small footprint also provides advantages such as space, cooling and overall power requirement and is made possible in part by lightweight witness technology. (A witness is a disk or file share that is needed to keep the services online and the data consistent if either of the two nodes fails.) In a typical HCI solution, the witness resides on a remote server on premises, which is an added expense in offices where no other server is available. AzS HCI avoids this requirement for additional hardware, however, by making use of the USB slot in a top-of-rack switch to enable a small USB drive to act as the witness. Alternatively, AzS HCI can take advantage of Azure Cloud Witness to perform the same witness function for the solution.

Aside from lightweight witness technology, another way the two-node configuration of AzS HCI reduces costs is through its support of switchless networking. Switchless networking enables a direct connection between the two nodes via a crossover cable, which eliminates the cost of a switch (10 gigabit Ethernet [GbE] or greater). With AzS HCI, switchless networking is fully supported with no performance degradation, including with remote direct memory access (RDMA) for ultra-low latency and high throughput. Some competing HCI vendors do not support RDMA as of this writing.

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

The minimal hardware requirements and cost-reducing features in two-node deployments for AzS 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.

Additionally, AzS HCI users can now monitor all clusters as needed from a single pane of glass via Azure Arc. Arc simplifies complex and distributed environments across on-premises, edge and multicloud. It allows IT personnel to organize, manage, and govern resources remotely at scale with scripting tools.



High availability and resiliency at the edge

High availability and resiliency are key benefits of an AzS HCI solution, and failover clustering is the main underlying technology that helps ensure high availability for hosted services and applications. Another feature, nested resiliency, adds even more storage resiliency to two-node deployments. With nested resiliency, a two-node AzS HCI cluster can lose one node *and* one disk drive on the other node without failing or losing any data. In other words, nested resiliency offers 4-way mirroring for two-node AzS HCI deployments, without requiring the added expense of a hardware RAID adapter. Many other HCI vendors do not offer this feature.

Azure Monitor for the remote office and edge

Windows Admin Center is the main administration console for AzS HCI. Windows Admin Center makes a number of Azure services available to facilitate AzS HCI administration through the Azure hybrid services page, which is found in the navigation menu. One such service that is particularly useful to the remote office and edge use case is Azure Monitor, shown in Figure 4.

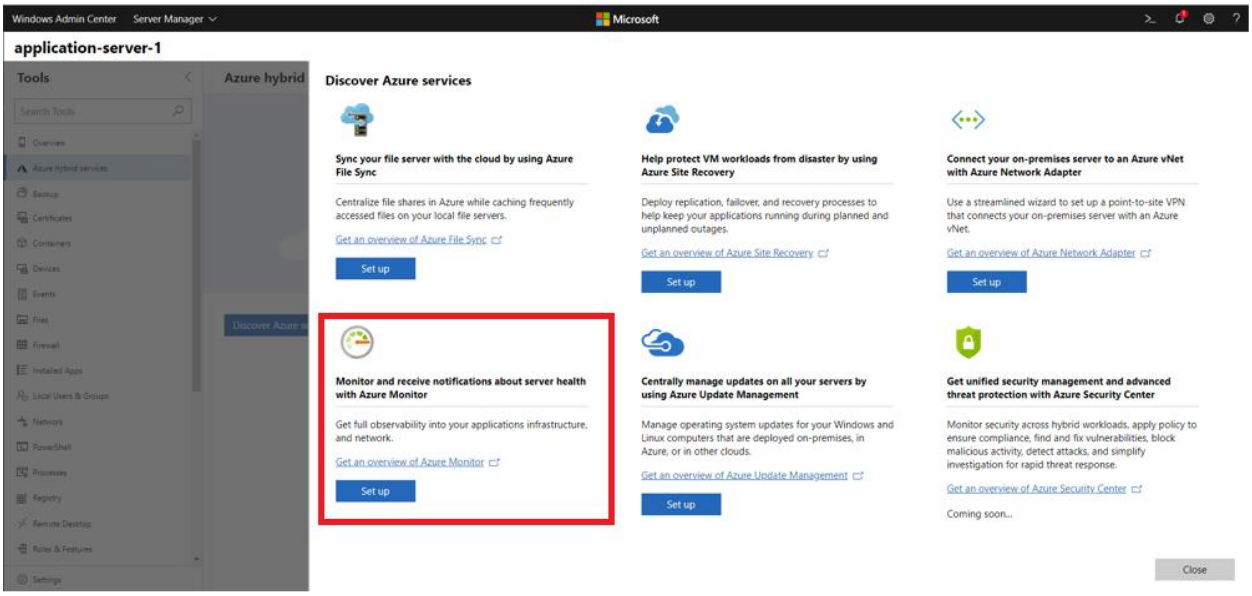


Figure 4. You can access Azure Monitor and other hybrid services through Windows Admin Center

Azure Monitor provides a comprehensive way to collect, analyze, and respond to data in your infrastructure. For example, it can provide comprehensive health monitoring of the nodes and all guest VMs in your AzS HCI cluster. It can also provide insight into 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 the branch office and edge scenario because it enables remote IT staff to monitor AzS HCI and take action as needed. For example, IT staff can configure alerts and dashboards to remotely monitor the health of AzS HCI nodes, guest VMs, and applications.

AzS offers an additional benefit for ROBO scenarios in the ability to monitor all clusters through Azure Arc integration. This allows users 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 for remote office and branch office deployments

The branch office and edge scenario for AzS HCI typically calls for an inexpensive, simple, and reliable solution, although some edge deployments are running analytics on local data, and require a more performant solution. To meet these requirements, AzS 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, fast, reliable, and inexpensive Intel® solid state drives (SSDs) can be used in a single tier, or, in an even lower-cost option, as a cache tier to support hard disk drives (HDDs) in the capacity tier.

Intel® Select Solutions for Microsoft Azure Stack HCI

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 that are designed to support specific workloads in basic ("Base") and advanced ("Plus") configurations to meet the needs of different edge use cases. To review the Intel Select Solutions for Azure Stack HCI, visit the Intel Select Solutions for Microsoft Azure Stack HCI brief.

Resource planning for the branch office and edge technical use case

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

Technical use case 2: Azure Stack HCI for virtual desktop infrastructure (VDI)

For the VDI use case, an AzS 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 AzS 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.

Another advantage is scalability. The new Azure Stack HCI scales performance and number of desktops nearly linearly as more nodes are added, making capacity planning much easier. Competing options do not scale linearly, giving AzS HCI users a key advantage in forecasting and purchasing additional nodes as their business grows.

HCI provides a reliable platform for VDI. An advantage of a VDI workload is that its demands for resources are often proportional to the number of users requiring desktop virtualization, and through HCI you can scale resources easily and predictably to meet those 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 AzS HCI offers particular advantages. First of all, when the client operating system is a version of Windows (as is fairly common), then both the guest and host leverage technologies from Windows, which enable advantages for administration, performance, and technical support. The shared platform between host and guest, in addition, allows native access to other network resources, such as file shares that are stored on the AzS 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, whose interface is shown in Figure 5, greatly eases the administrative overhead associated with updating the many client VMs hosted on AzS HCI. Through its cloud interface, Azure Update Management lets you quickly assess the status of available updates on all client VMs hosted in the AzS HCI cluster and manage the process of installing those updates. Azure Security Center, for its part, is a unified infrastructure-security management system that strengthens the security posture of your datacenters and provides advanced threat protection across all your machines.

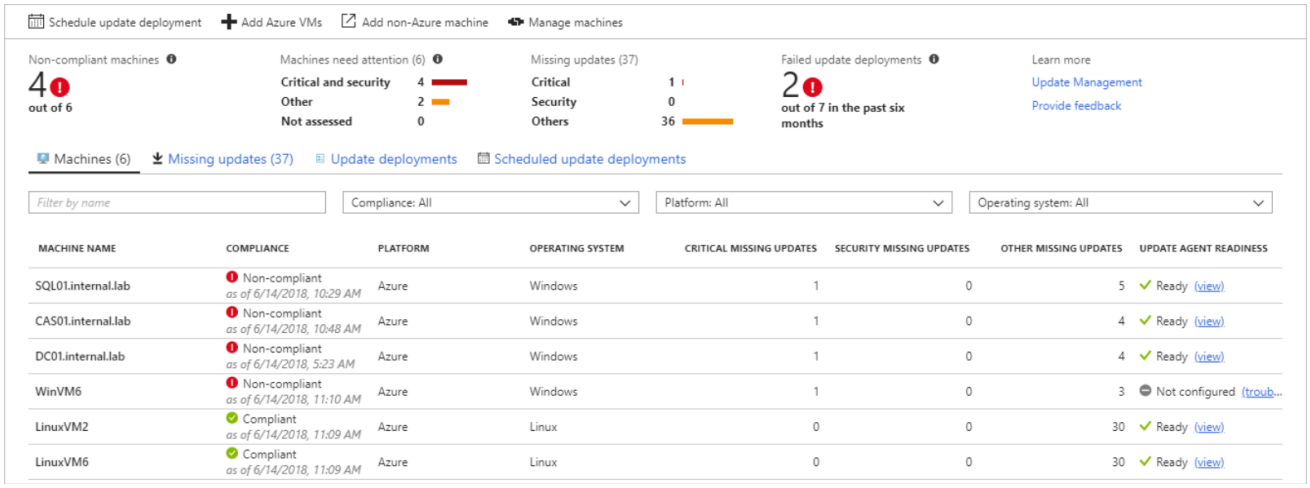


Figure 5. Viewing update assessments in Azure Update Management

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

Intel technologies to support the VDI use case

VDI is first and foremost a compute-intensive workload. To ensure a productive user experience, it is essential to provide adequate virtual CPU resources to support the types of workers and applications being supported. Intel Xeon Scalable processors offer a wide range of options to allow you to tailor the infrastructure to meet users' needs. AzS HCI also supports Intel® Optane™ persistent memory modules, which increase efficiency for VDI deployments. Intel Optane DC persistent memory provides affordable memory and support for data persistence. You can use persistent memory 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-DRAM solutions. (Note that persistent memory requires a 2nd Generation Intel Xeon Scalable processor.)

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.

Technical use case 3: Azure Stack HCI for high-performance Microsoft SQL Server

AzS HCI offers a high-performing, scalable, and manageable platform for implementing Microsoft SQL Server. Excellent performance is often crucial for SQL Server deployments, and, given this requirement, it's important to note that legacy AzS HCI already offered the best-performing HCI solution on the market.[1] We anticipate that the newest version will have similar or better performance, and will publish benchmarks soon. This excellent performance comes at an affordable price for the host software. Windows Server 2019 is then

run as a guest in a VM on top of the native AzS HCI, rather than as the host. AzS HCI also offers flexibility in that it can host VMs based on both Windows and Linux, which in turn gives customers the freedom to choose either the Windows or Linux version of SQL Server — whichever best suits their needs. It is also possible to run SQL Server on a Linux VM if that is more appropriate for the environment. Of course a valid license for SQL Server is needed in both cases.

> "The [Azure Stack HCI] cluster we've been testing has posted tremendous numbers, the fastest we've seen in a mid-market 4-node HCI cluster."[1]
>
> — StorageReview.com

When used to host SQL Server, AzS HCI also offers the benefit of a single vendor for its hypervisor, host operating system, and database server. This advantage allows the underlying code to be optimized for performance, and it also provides 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 organizations to protect the SQL Server back end of an application to help keep your data safe and your apps and workloads online when planned and unplanned outages occur.

AzS HCI also comes with native stretch clustering capability for business continuity and disaster recovery. Stretch clustering provides automatic failover to restore production quickly, and without the need for manual intervention. Storage Replica provides 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

AzS HCI deployments built to support database servers are typically optimized for performance. To support the highest performance, you can use high core count and 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 the typical workload demands and resource requirements for the high-performance SQL Server use case, see Appendix 3.

Technical use case 4: Azure Stack HCI for trusted enterprise virtualization

This technical use case for Azure Stack HCI is broad and can overlap all other use cases. Despite the name, it can also apply to small, medium and mid-market companies. It is an enterprise-class use case, but by no means enterprise only. It pertains to general scenarios in which customers are looking to serve any applications and a mix of workloads hosted in VMs with both high security and high availability. AzS 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 (called 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. With these increased protections offered by VBS, even if malware gains access to the operating system kernel, the possible exploits can be greatly limited and contained 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 AzS 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.

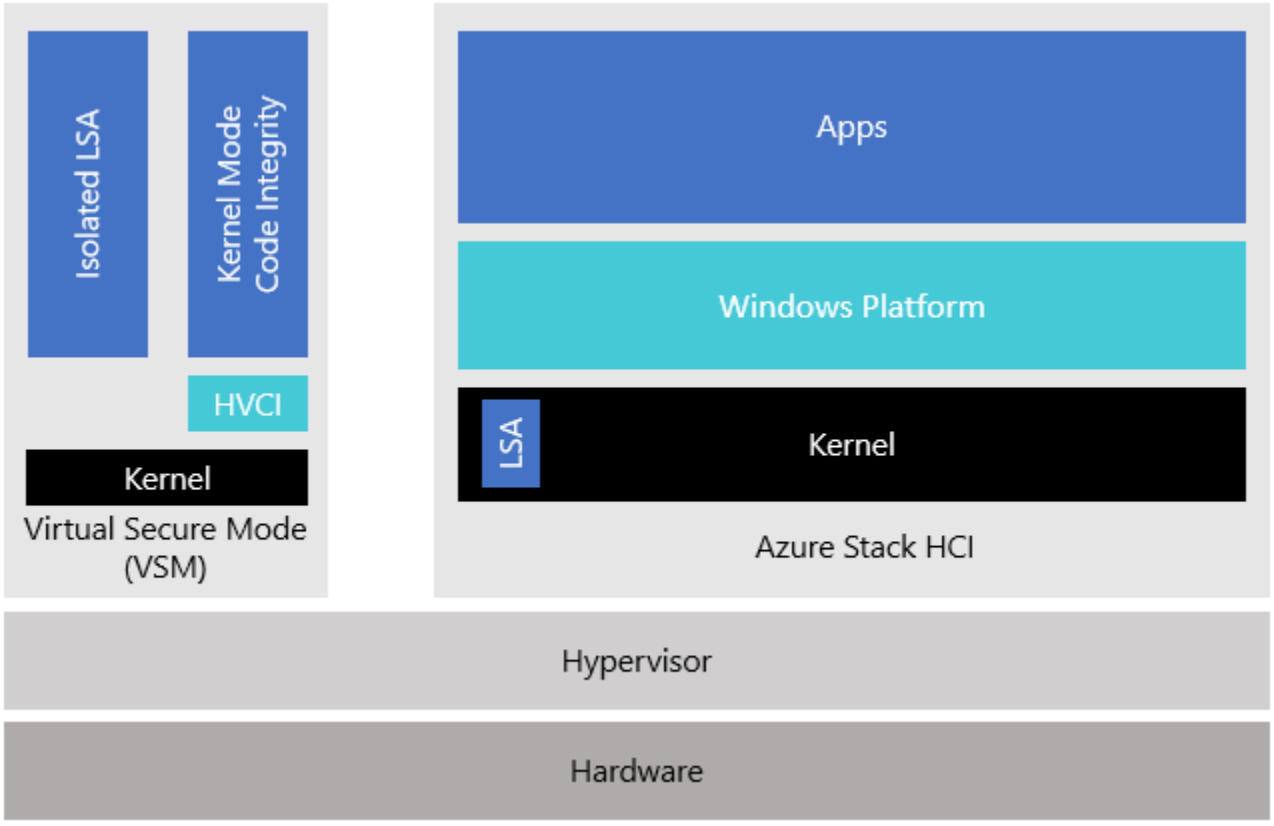


Figure 6. Virtualization-based security (VBS) creates a secure memory enclave that isolates sensitive code and processes from the host operating system

Two features enabled by VBS include Hypervisor-Enforced Code Integrity (HVCI) and Credential Guard. HVCI uses VBS to significantly strengthen code-integrity policy enforcement. (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 makes use of 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, for its part, isolates users' 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 AzS HCI hardware solutions.

For protected workloads on VMs to fulfill their security promise, it is imperative that the host machine is equipped, as are all validated AzS HCI solutions, with a TPM chip, and that the host has enabled both VBS and HVCI. Figure 7 illustrates the advantages that AzS 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). Hosts that are not protected (bottom right) are not able to pass along the same level of security to VMs, even if the guest is protected with a virtual TPM (vTPM).

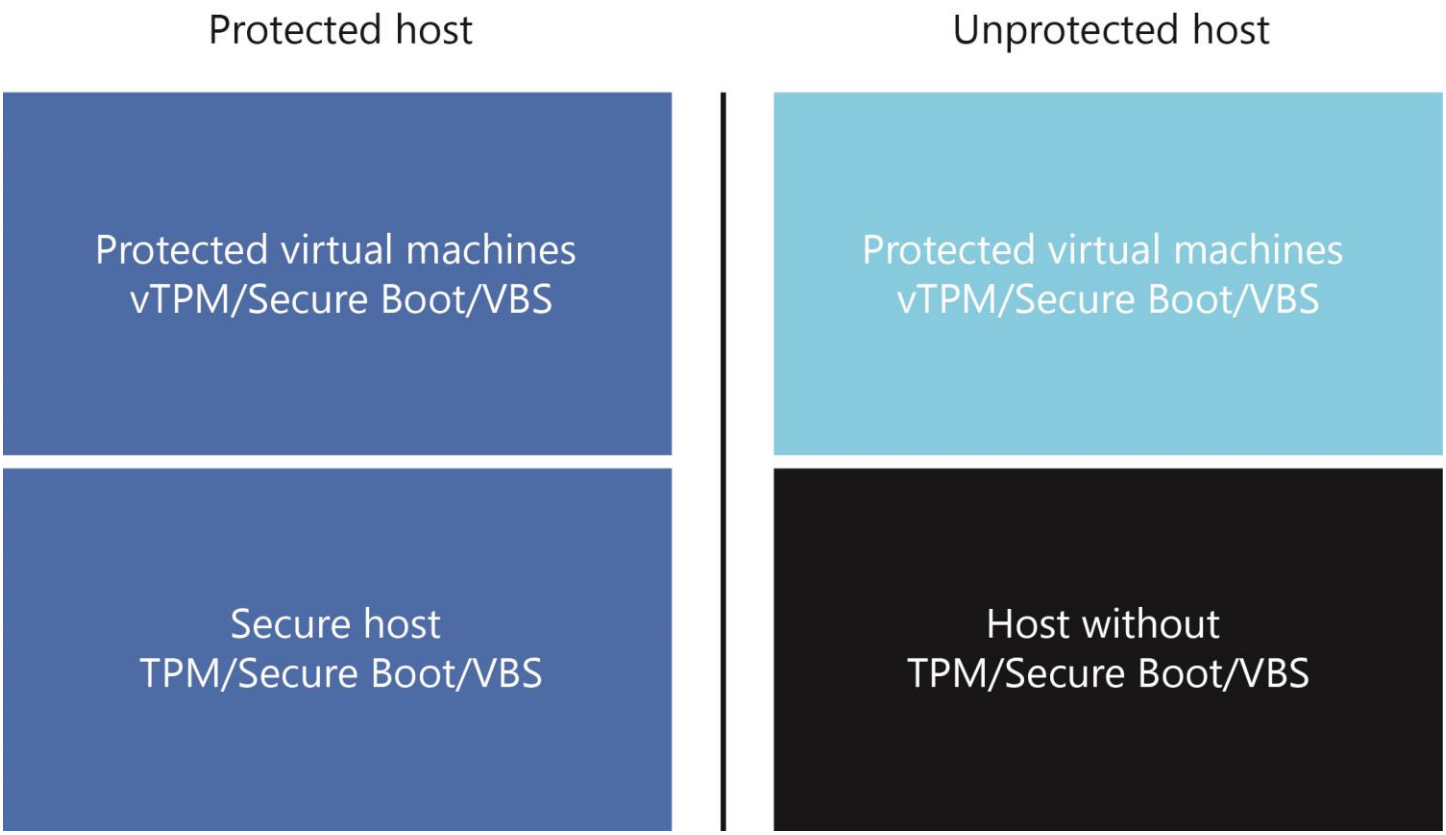
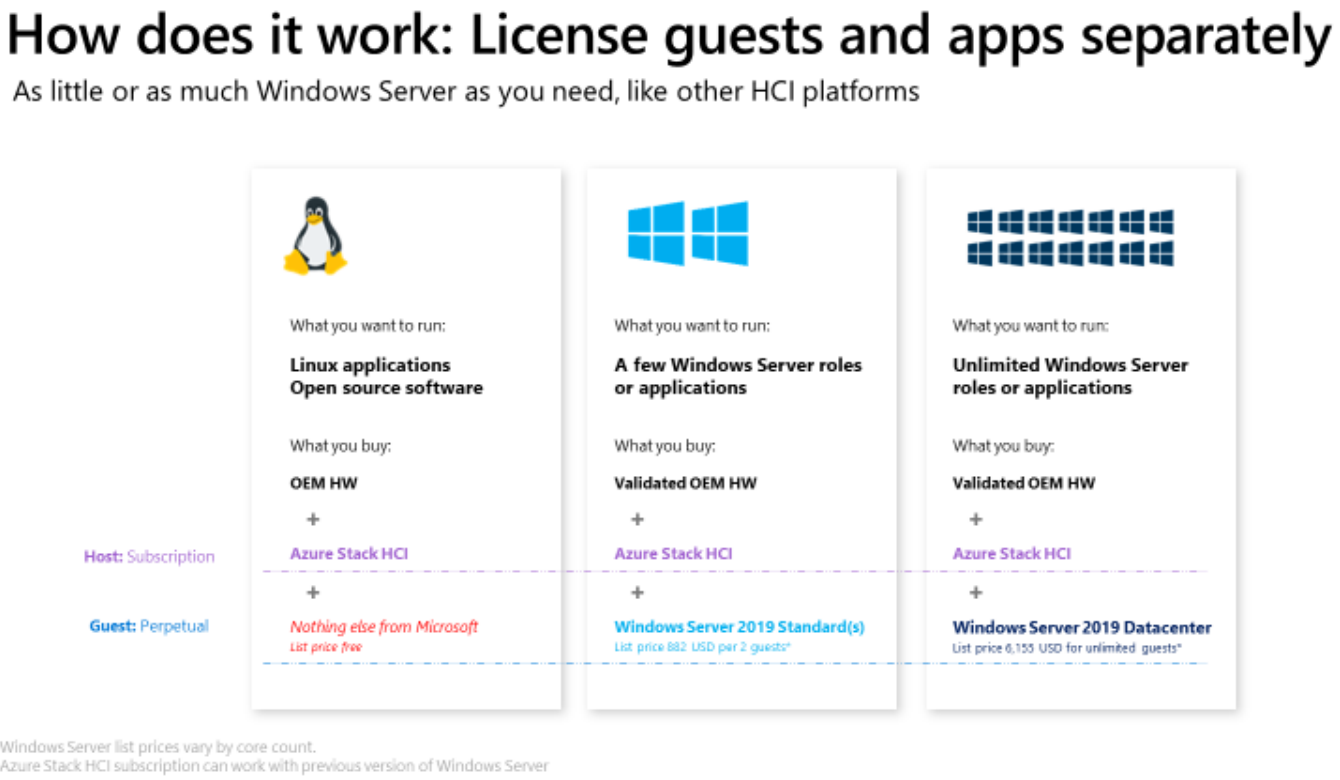


Figure 7. All validated AzS HCI solutions are able to provide a high level of protection to VMs through a physical TPM

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

What sort of licensing is needed for this use case will vary based on intended workload. **The following licensing section applies to all use cases not just Trusted Enterprise Virtualization.** For all Azure Stack HCI operations, an Azure Stack HCI subscription is required, since Azure Stack HCI functions as the host in all cases. For Linux applications or open source software, nothing else needs to be purchased from Microsoft. To run a few Windows Server roles or applications, only a standard Windows Server 2019 license is needed. For unlimited Windows Server roles or applications, a Windows Server 2019 Datacenter license is required. These options allow AzS HCI customers to select the licensing option that's best for their specific operations and budget.



Azure Security Center enhances the trusted enterprise virtualization use case

To enhance AzS HCI deployments requiring high security, customers can take advantages of the Azure Security Center features integrated in Windows Admin Center. Connecting your AzS 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 the installation of an agent, a procedure that 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

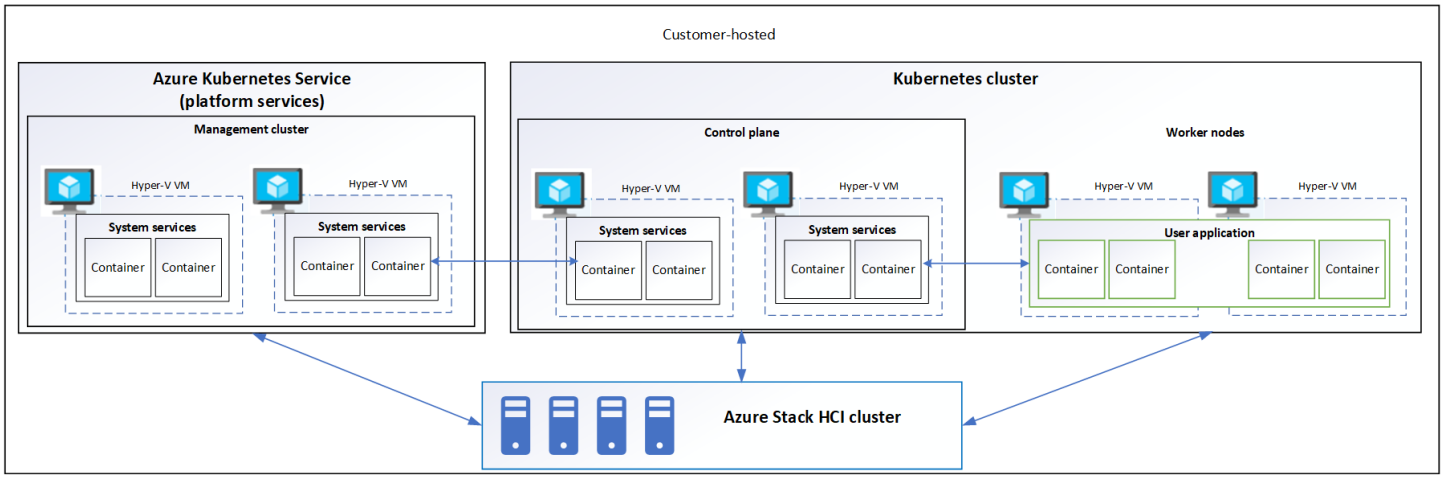
For information about the typical workload demands and resource requirements for the trusted enterprise virtualization use case, see Appendix 4.

Technical use case 5: Azure Kubernetes Service on Azure Stack HCI (AKS-HCI)

This technical use case details AzS 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 get started 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 customers who want to run containerized workloads within their on-premises environments, whether that be at an edge remote site, or in the datacenter, to meet network or regulatory goals. With an existing Azure Stack HCI cluster in place, and after validating several prerequisites, Windows Admin Center guides administrators through first configuring an AKS management cluster. This 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 the key components, but if you prefer, you could use PowerShell for this step too.

Azure Kubernetes Service on Azure Stack HCI



With the management cluster in place, you can now deploy your workload clusters, which will ultimate run your workloads. Again, the Windows Admin Center streamlines the experience, asking for a few pieces of information relating to cluster and host selection, node sizing, number of Windows and Linux nodes, some storage and network info, and optionally, Azure Arc connectivity. With those 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 users who have embraced AKS in Azure, you'll notice the steps in the Windows Admin Center mirror those native to AKS in

Azure, but for those of you who prefer defining your infrastructure programmatically, PowerShell can also be used to deploy these workload clusters, quickly and easily.

For those workload clusters that were 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 it's GitOps approach to centrally deploying applications and configurations to kubernetes environments, or if you prefer, use your existing, familiar 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 through to containerized workloads, with native integration with Azure services. Through it's streamlined deployment, and scaled management, the combination of Azure Stack HCI, and AKS-HCI, provides an incredibly time to value, and a highly scaled, reliable and secure solution to run your 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 (AzS HCI), Intel also provides many software components, such as the following, to support performance, reliability, and security in AzS HCI solutions:

- Intel® Deep Learning Boost (Intel® DL Boost), available on 2nd Generation Intel Xeon Scalable processors, takes embedded artificial-intelligence (Al) performance to the next level. Intel Xeon Scalable processors are built specifically for the flexibility to run complex Al workloads on the same hardware as your existing workloads. With Intel DL Boost, some enterprises have experienced 57 percent performance improvements.[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 upon 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 (RAS), 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 (MLE), which helps ensure that your server is running "known-good" configurations of your critical software components (firmware, BIOS, operating system, and hypervisors).

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

AzS HCI offers industry-leading performance for HCI at an affordable price of an Azure subscription with a cost of $10 per month, per core activated. AzS 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 AzS HCI solutions that are built for your technical use case. Whether you need an HCI solution to run edge workloads, VDI, SQL Server, highly secure workloads, or a Kubernetes cluster, you can find an AzS HCI solution that meets your requirements on-premises while also offering optional hybrid services through Azure.

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

Appendix 1: Azure Stack HCI for branch office and edge—hardware guidance

The branch office 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, 384GB) is recommended to optimize performance. The recommended storage design is single-tier, with all NVMe SSDs. For more information, see Table 1.

<table>
<tr><th colspan="4">Hardware guidance for branch office and edge</th></tr>
<tr><td colspan="4">Workload assumptions
• Tenants: Small number of mixed tenants
• Containers: Smaller resource footprint than VMs
• VMs: Small to medium in size for traditional business-application stacks
• Performance per tenant: Resources fall somewhere between the VDI and SQL Server use cases</td></tr>
<tr><td colspan="4">Design guidance
• Keep physical footprint small and optimize cost for desired level of performance (this typically will not resemble a full-blown datacenter deployment; for example: switchless, two-node)
• Balance of performance and the cost of the hardware bill of materials (BoM)</td></tr>
<tr><td rowspan="11">Hardware considerations (per node)</td><td colspan="3">CPU requirements</td></tr>
<tr><td>Core counts:</td><td>Frequency:</td><td>> 1 TB memory/CPU?</td></tr>
<tr><td>Low–medium</td><td>Low–medium</td><td>No</td></tr>
<tr><td colspan="3">• Start with a general-purpose CPU with sufficient headroom for Storage Spaces Direct
• No need for large amounts of memory per socket</td></tr>
<tr><td colspan="3">Storage requirements</td></tr>
<tr><td colspan="2">Capacity:</td><td>Performance/tiered layout:</td></tr>
<tr><td colspan="2">Low</td><td>One-tier, all NVMe</td></tr>
<tr><td colspan="3">Account for nested resiliency when computing raw needs
(Needs vary based on resiliency model employed)</td></tr>
<tr><td colspan="3">Memory requirements</td></tr>
<tr><td colspan="2">Capacity:</td><td>Performance:</td></tr>
<tr><td colspan="2">High</td><td>Medium</td></tr>
<tr><td colspan="2">256 GB</td><td>Possible candidate for persistent-memory (PMEM) for designs that need more memory</td></tr>
</table>

	Network (east/west) requirements			
	Bandwidth: Medium • Driven by storage configuration • Switchless option			
	CPU	Storage	Memory	Network
Example node configuration	2 x Intel Xeon Gold 5218 processors	Boot: 2 x Intel SSD D3-S4510 (M.2) Capacity: 4 x Intel SSD DC P4610	16 x 16 GB DDR4	2 x 10 gigabits per second (Gbps) or 2 x 25 Gbps Intel Ethernet Network Adapter X722 or Intel Ethernet Network Adapter 800 series

Table 1. Hardware guidance for the branch office 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 – here is an overview example. Microsoft also provides further guidance which you can find HERE • 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="10">Hardware considerations (per node)</td><td colspan="4">CPU requirements</td></tr>
<tr><td colspan="2">Core counts:</td><td>Frequency:</td><td>> 1 TB memory/CPU?</td></tr>
<tr><td colspan="2">Medium</td><td>Medium–high</td><td>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.

<table>
<tr><th colspan="4">Hardware guidance for high-performance Microsoft SQL Server</th></tr>
<tr><td colspan="4">Workload assumptions
Model loosely on Azure DSv2-series approach:
4 vCPU, 32 GB RAM, 1 TB disk, 50,000 IOPS, 500 MB/s throughput</td></tr>
<tr><td colspan="4">Design guidance
• Consistent, high performance is the number-one design goal
• Avoid oversubscription scenarios to maintain quality of service (QoS)
• Tune Storage Spaces Direct for storage performance versus storage efficiency (for example, you need 3 nodes for 3 way mirroring)
• Host up to 12 SQL Server instances per node</td></tr>
<tr><td rowspan="11">Hardware considerations (per node)</td><td colspan="3">CPU requirements</td></tr>
<tr><td>Core counts:</td><td>Frequency:</td><td>> 1 TB memory/CPU?</td></tr>
<tr><td>High</td><td>High</td><td>No</td></tr>
<tr><td colspan="3">• 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</td></tr>
<tr><td colspan="3">Storage requirements</td></tr>
<tr><td colspan="2">Capacity:</td><td>Performance/tiered layout:</td></tr>
<tr><td colspan="2">High</td><td>One tier, all NVMe</td></tr>
<tr><td colspan="3">• Support >= 600,000 IOPS per node for user workloads
• 12 TB usable capacity (36 TB raw). A minimum of 3 nodes is needed to have performance storage tier (3-way mirror).
• Medium-endurance SSDs required in a single-tier design</td></tr>
<tr><td colspan="3">Memory requirements</td></tr>
<tr><td colspan="2">Capacity:</td><td>Performance:</td></tr>
<tr><td colspan="2">Medium–high</td><td>High</td></tr>
<tr><td colspan="2">Example: 768 GB</td><td>Example: CPU with 2,933 MHz DDR4 support
Candidate for Optane Persistent Memory</td></tr>
</table>

	Network (east/west) requirements
	Bandwidth: Medium (driven by storage configuration)

	CPU	Storage	Memory	Network
Example node configuration	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 that is most suitable for your workload.

<table>
<tr><th colspan="4">Azure Stack HCI for trusted enterprise virtualization</th></tr>
<tr><td colspan="4">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 branch office/edge and SQL Server</td></tr>
<tr><td colspan="4">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.</td></tr>
<tr><td rowspan="11">Hardware considerations (per node)</td><td colspan="3">CPU requirements</td></tr>
<tr><td>Core counts:</td><td>Frequency:</td><td>> 1 TB memory/CPU?</td></tr>
<tr><td>Medium</td><td>Medium</td><td>No</td></tr>
<tr><td colspan="3">• 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</td></tr>
<tr><td colspan="3">Storage requirements</td></tr>
<tr><td colspan="2">Capacity:</td><td>Performance/tiered layout:</td></tr>
<tr><td colspan="2">Medium</td><td>Two-tier, all-flash</td></tr>
<tr><td colspan="3">• Use branch office/edge and SQL Server use cases as bookends/guardrails
• Cache should remain between 5–10 percent of raw capacity</td></tr>
<tr><td colspan="3">Memory requirements</td></tr>
<tr><td colspan="2">Capacity:</td><td>Performance:</td></tr>
<tr><td colspan="2">Medium</td><td>Medium</td></tr>
<tr><td colspan="2">Example: 384GB or 768GB</td><td>Candidate for PMEM</td></tr>
<tr><td colspan="3">Network (east/west) requirements</td></tr>
<tr><td colspan="3">Bandwidth: Medium (driven by storage configuration)</td></tr>
</table>

	CPU	Storage	Memory	Network
Example node configuration	2 x Intel Xeon Gold 6230 processors	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	12 x 16 GB DDR4 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 4. Hardware guidance for the trusted enterprise virtualization use case

[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.