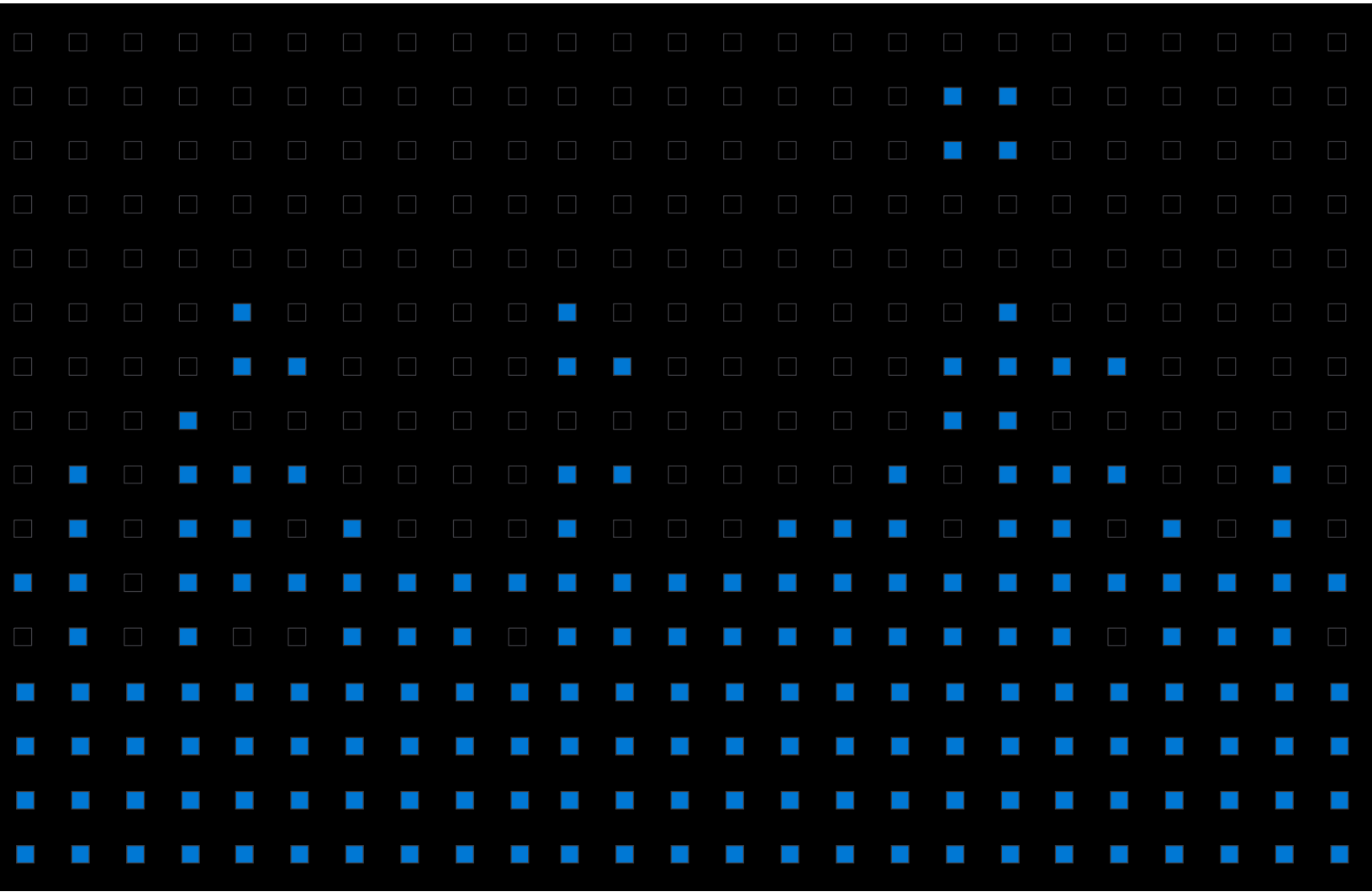
White paper

Microsoft Azure | intel

Azure Stack HCI: Use cases and scenarios



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 file servers. For these locally hosted services and applications, customers have been turning in particular to hyperconverged infrastructure (HCI) as a way 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), a member of the Azure Stack family of technologies, offers an excellent price/performance ratio as an on-premises HCI solution, 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

AzS HCI is offered through Microsoft hardware partners, prebuilt, and typically either preconfigured or bundled with simple configuration software. More than 175 AzS HCI solutions are available today from at least 20 Microsoft hardware partners. These partners offer hardware configurations that Microsoft has validated to ensure optimal performance and reliability for AzS HCI.

AzS HCI combines this pre-validated hardware from Microsoft partners with the following Windows Server 2019 Datacenter components and management tools:

- **Windows Server 2019 Datacenter roles and features:**
  - **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 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

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

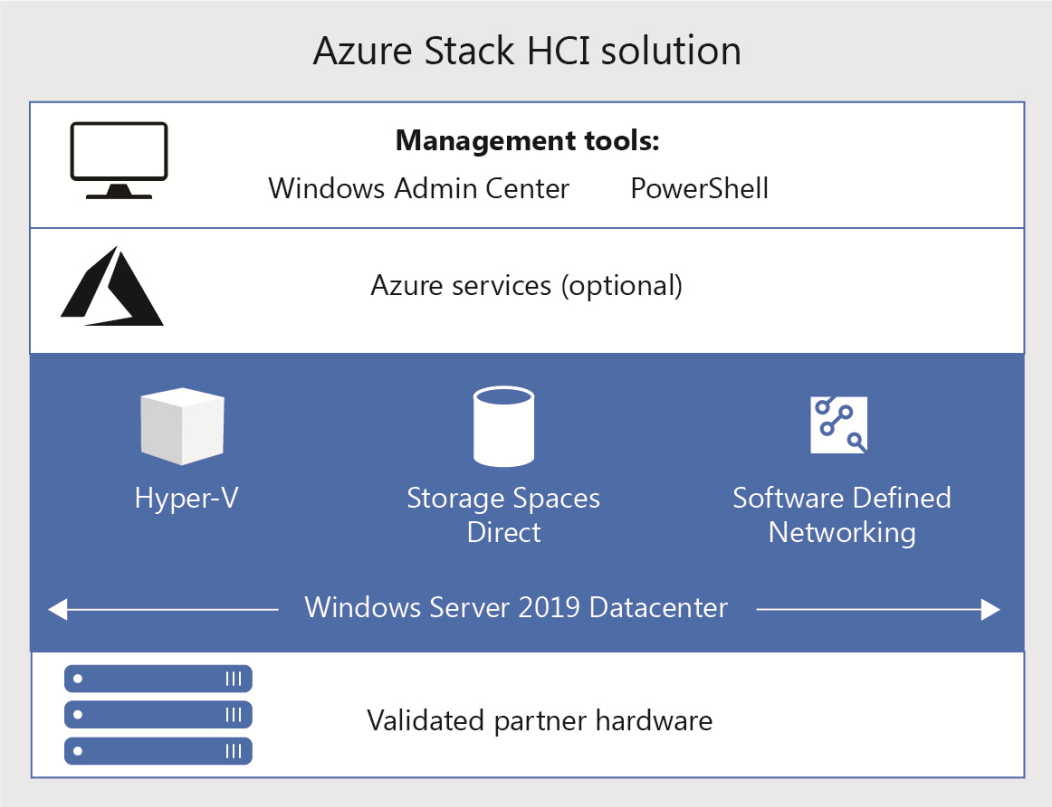


Figure 1. The components of the Azure Stack HCI solution

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 certified 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 certifies AzS HCI solution offerings to validate the suitability of their hardware configurations in supporting any of these five scenarios.

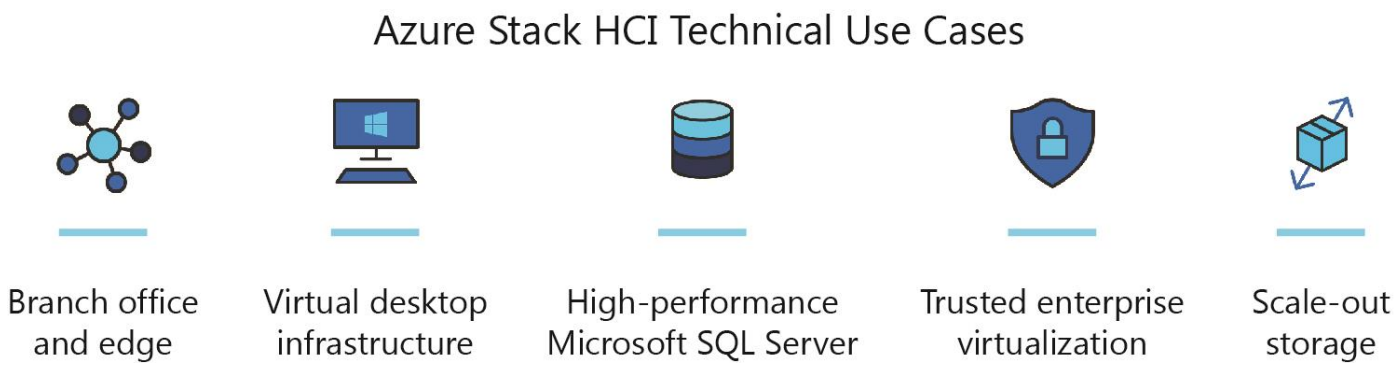


Figure 2. The Microsoft AzS HCI Technical Use Case program approves hardware configurations for the five most typical technical use cases

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

Technical use case 1: Azure Stack HCI for branch office and edge

The 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 branch office and edge solutions for AzS HCI can remain low thanks to a minimal hardware footprint that starts at just two nodes. This small footprint 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 high-speed (10 gigabit Ethernet [GbE] or greater) switch. 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. Competing HCI vendors do not support this implementation as of this writing.

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, they can add up to tremendous savings.

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 the equivalent of RAID 5 + 1 for two-node AzS HCI deployments, without requiring the added expense of a hardware RAID. No other HCI solution offers this feature.

Using AzS HCI with Azure IoT Edge to deploy containerized applications

For offices with minimal IT staff, Azure IoT Edge can be used to ease the deployment of containerized applications to an AzS HCI cluster with the help of an administrator working anywhere in the world. Azure IoT Edge is an engine that can be installed on a VM in AzS HCI and that enables containers for the cluster. Azure IoT Edge also has Internet of Things (IoT) gateway functionality included, which enables the device on which it is installed to be managed remotely from the cloud via Azure IoT Hub.

Through Azure IoT Hub, administrators can see the edge device and determine which software packages (such as containerized applications) the device should download. The Azure IoT Edge runtime installed on the device later contacts Azure IoT Hub to download and execute those approved containers. This process is illustrated in Figure 3.

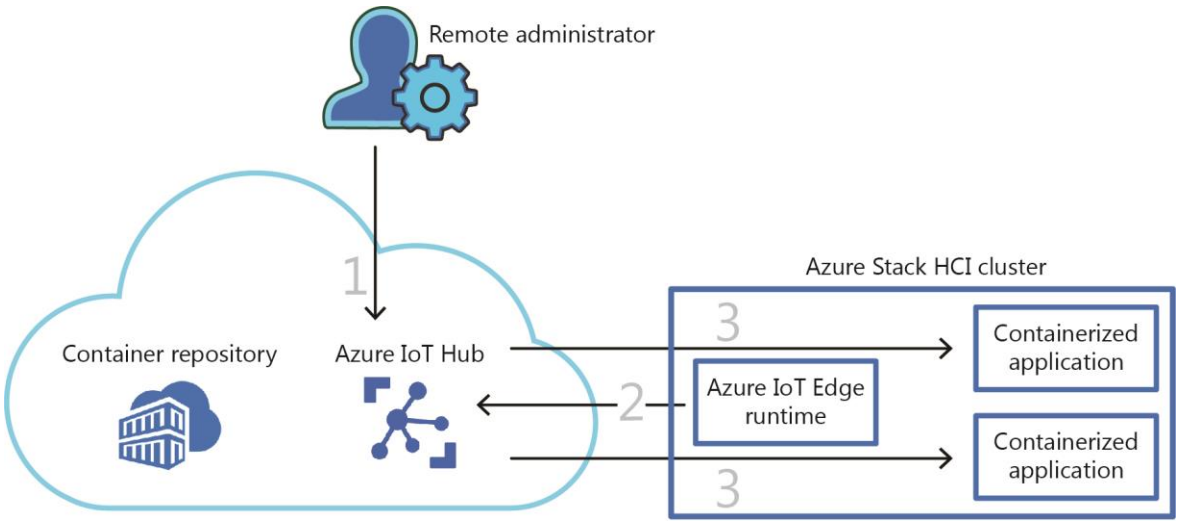


Figure 3. Azure IoT Edge enables remote administrators to push approved applications to an Azure Stack HCI cluster

Application deployment through Azure IoT Edge and Azure IoT Hub offers a particular advantage to companies with many remote offices or retails stores: it allows a remote administrator to create groups of devices and manage them together. In this way, the remote administrator can deploy (or remove) the same applications to many branch offices at once, which can dramatically reduce administrative overhead.

Beyond the management simplicity enabled by Azure IoT Edge, another advantage provided by this service is that it helps deploy useful IoT applications. Many IoT applications can be particularly beneficial for remote offices, such as retail shops (for example, for power monitoring and control).

For more information about Azure IoT Edge, visit "Azure IoT Edge documentation."

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.

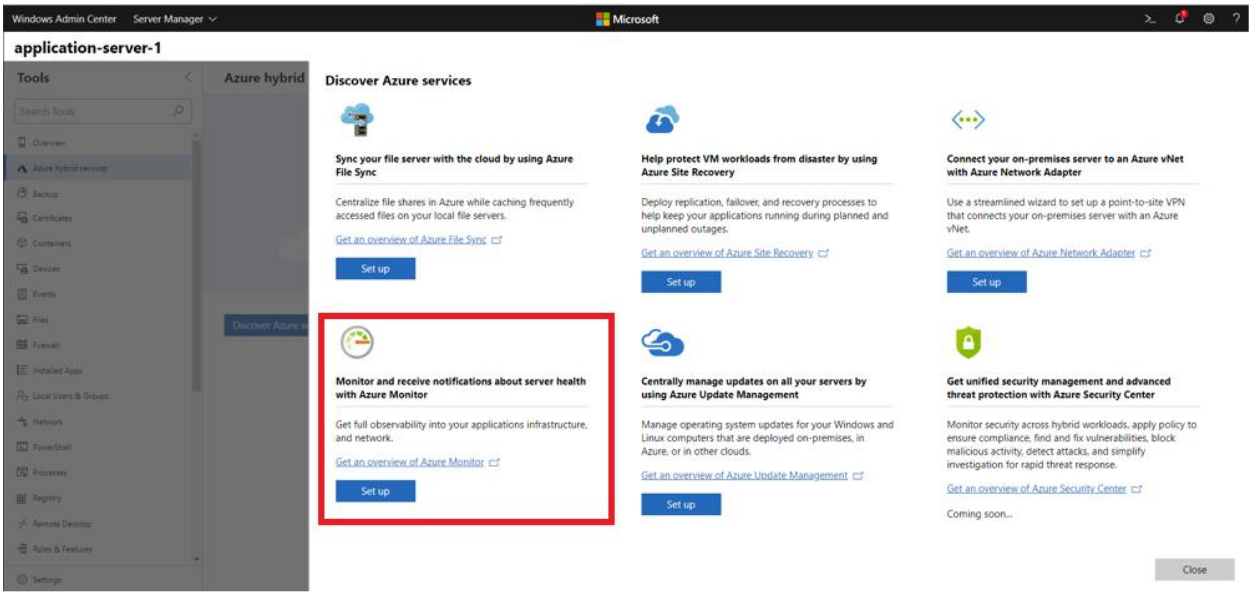


Figure 4. You can access Azure Monitor and other hybrid services through Windows Admin Center

Azure Monitor provides a comprehensive way to collect, analyze, and respond to data in an 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.

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. To meet these requirements, AzS HCI solutions can support single-socket servers that run on Intel® Xeon® processors with as few as eight cores and a frequency as low as 1.4 GHz. 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 combinations of Intel compute, memory, storage, and network products that are designed to support specific workloads in basic ("Base") and advanced ("Plus") configurations. 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 technical 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.

HCI provides a reliable platform for VDI. An advantage of a VDI workload is that its demands for resources are typically 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.

Running VDI workloads on AzS HCI offers particular advantages. First of all, when the client operating system is a version of Windows (as is typical), then both the guest and host are running versions of Windows, which enables 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.

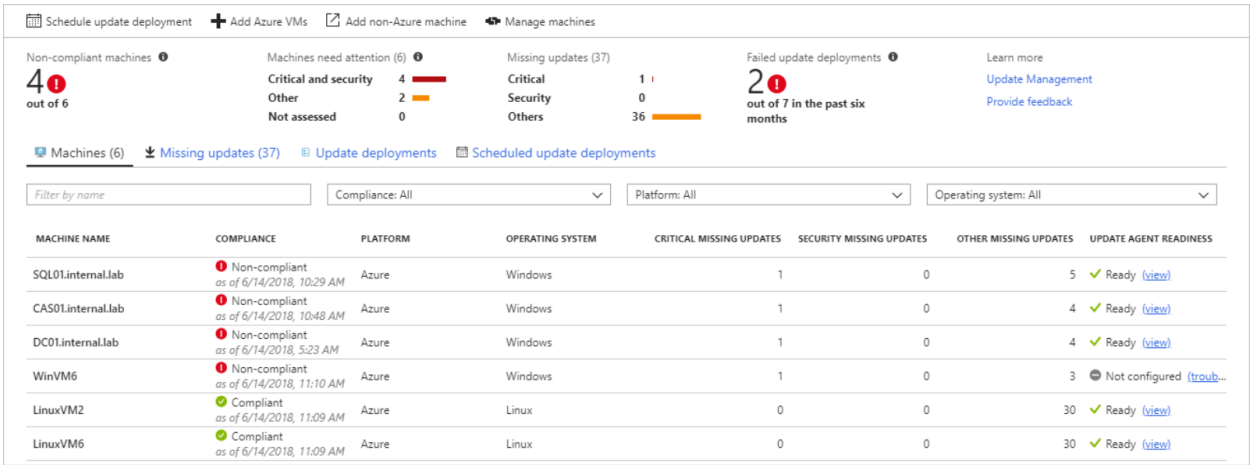


Figure 5. Viewing update assessments in Azure Update Management

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

AzS HCI supports Intel® Optane™ DC persistent memory modules, which are useful 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 client 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 DC persistent memory, visit the Intel Optane DC 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 implementation for Microsoft SQL Server. Although the native SQL Server feature already provides high availability for databases (AlwaysOn Availability Groups), running SQL Server in an AzS HCI cluster allows you to run SQL Server and its associated applications with the added resiliency of virtualization.

Excellent performance is often crucial for SQL Server deployments, and, given this requirement, it's important to note that AzS HCI offers the best-performing HCI solution on the market.[1] What's more, this excellent performance comes at an affordable price, which is simply the cost of a Windows Server 2019 Datacenter license.

> "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. 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.

Azure Backup enhances the high-performance SQL Server use case

You can use Azure Backup to provide comprehensive management of your database backups. Through Azure Backup, you can back up machines on premises to a local server—by relying on either Microsoft System Center Data Protection Manager (DPM) or Microsoft Azure Backup Server (MABS). You can then back up that server to an Azure Recovery Services vault. (Alternatively, you can back up on-premises Windows machines directly to Azure.)

Another advantage of Azure Backup is that it allows you to back up the machine state of machines running on premises, along with the data, allowing for application-consistent backups.

For more information on Azure Backup, see "What is the Azure Backup service?"

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-bandwidth Intel SSDs with NVM Express (NVMe) in a single storage tier.

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 AzS HCI is broad and can overlap with all other use cases. It pertains to general scenarios in which customers are looking to serve any applications 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.

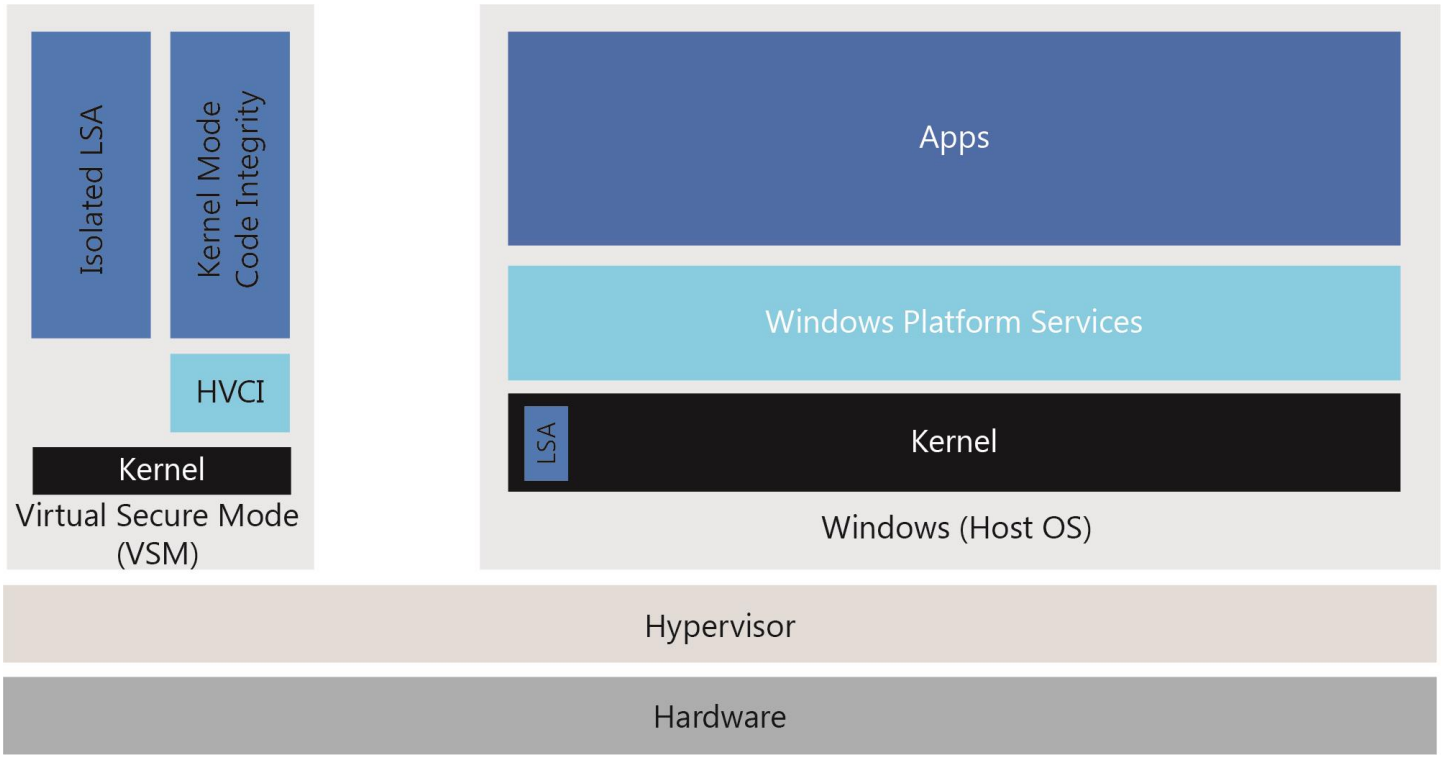


Figure 6. Virtualization-based security (VBS) creates a secure memory enclave that isolates sensitive code and processes from the host operating system

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 available on 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).

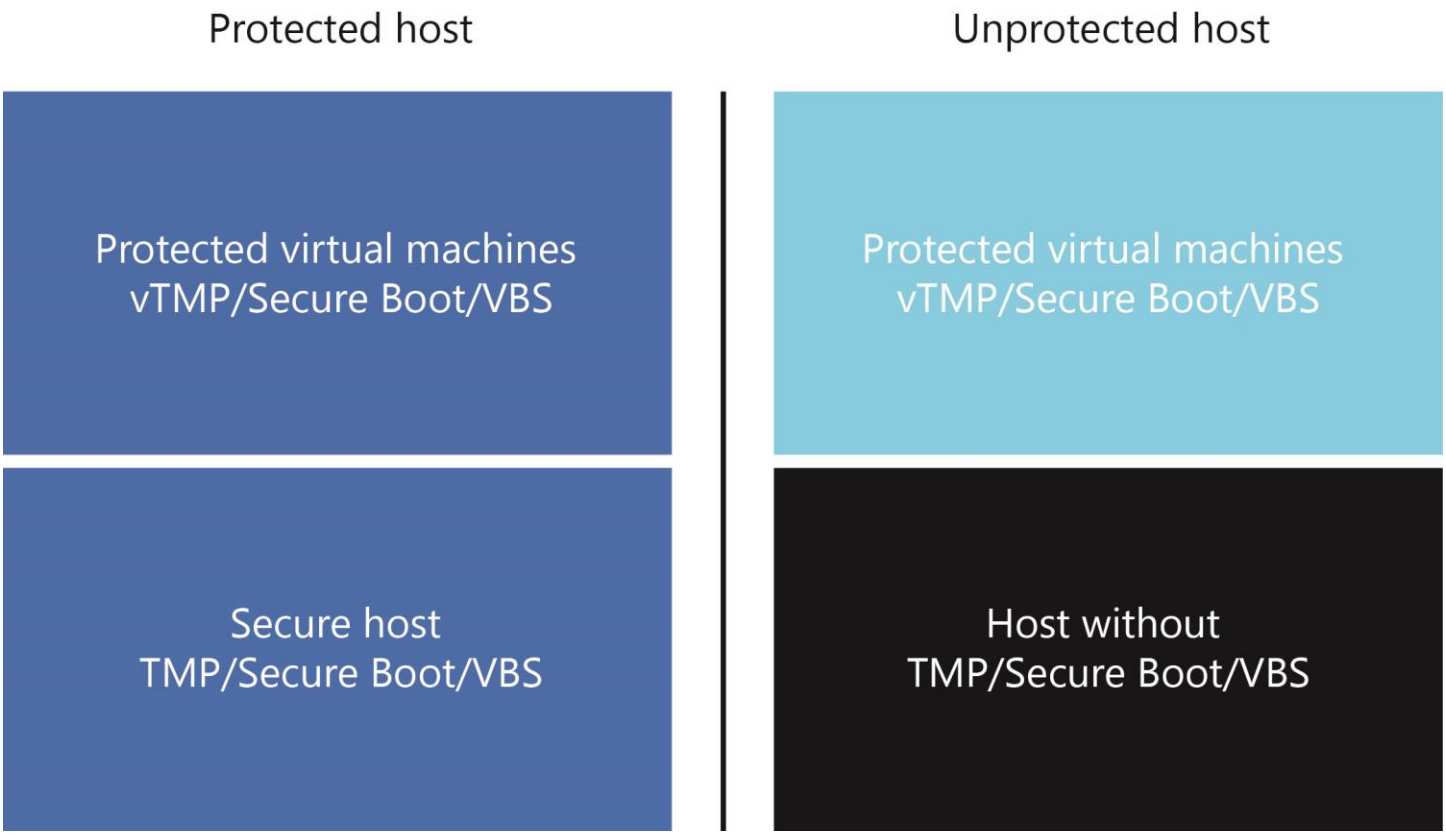


Figure 7. All validated AzS HCI solutions are able to provide a high level of protection to VMs through a physical TPM

The Department of Defense requires both VBS and HVCI

The United States Department of Defense (DoD) considers both VBS and HVCI to be crucial in high-security environments. In fact, the Windows Server 2019 Security Technical Implementation Guide (STIG), which the DoD publishes as a tool to improve the security of its information systems, lists both VBS and HVCI as general security requirements.

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

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 Stack HCI for scale-out storage

This technical use case describes customer requirements for file serving with high scalability, performance, and availability. AzS HCI meets these requirements by offering industry-leading storage performance on validated hardware that can be optimized for density, speed, or performance-to-cost ratio. In addition, AzS HCI provides these benefits in an affordable file server with a small hardware commitment.

The scale-out storage technical use case draws upon the features of Storage Spaces Direct, one of the key building blocks of an AzS HCI solution, which provides shared-nothing storage among nodes. Through storage virtualization, Storage Spaces Direct creates a pool of highly available, highly scalable storage from locally attached drives at a cost that is much lower than traditional storage area network (SAN) or network-attached storage (NAS) arrays.

Many features in Storage Spaces Direct enhance performance and availability for file servers. For example, persistent read/write cache drives are used to improve performance, illustrated in Figure 8, which shows a three-node AzS HCI cluster in which all servers are contributing local storage. For each server, the two fastest drives (NVMe drives) are dedicated to read/write caching.

Figure 8. Storage Spaces Direct uses locally attached storage, and the fastest (NVMe) drives are used for read/write caching

Beyond this caching feature, the mirror-accelerated parity feature in Storage Spaces Direct offers inexpensive, space-efficient storage without sacrificing performance. Nested resiliency, for its part, adds extra fault tolerance for storage in two-node AzS HCI clusters. And deduplication (available with NTFS only) improves storage density.

Azure File Sync supports the scale-out storage use case

Available through Windows Admin Center, Azure File Sync is an Azure service that can significantly ease administration for the scale-out storage use case for AzS HCI. Azure File Sync enables you to centralize your organization's file shares in Azure, while keeping the flexibility, performance, and compatibility of an on-premises file server. For more information about Azure File Sync, visit "Sync your file server with the cloud by using Azure File Sync."

Intel technologies to support scale-out storage

You can select Intel storage drives that support the goals of your scale-out storage server, whether those goals are to optimize performance or the performance-to-cost ratio. To optimize performance, you can choose high-bandwidth Intel SSDs with NVMe for all storage drives. To gain excellent performance for a low cost, you can use Intel Optane DC SSDs as cache drives, with Intel QLC 3D NAND SSDs or Serial ATA (SATA)-based SSDs or HDDs as capacity drives.

Beyond storage drives, Intel® Ethernet Network Adapters support RDMA, a required technology for Storage Spaces Direct in AzS HCI. Internet Wide-Area RDMA Protocol (iWARP) is one of the network protocol options used to implement RDMA, and the following Intel Ethernet Network Adapters support iWARP:

- Intel Ethernet Network Adapter X722-DA2
- Intel Ethernet Network Adapter X722-DA4

In addition, the new Intel® Ethernet 800 Series Ethernet controllers, such as the Intel® Ethernet Controller E810, support both iWARP and the second RDMA protocol option, RDMA over Converged Ethernet version 2 (RoCE v2). For more information about Intel Ethernet 800 Series controllers and other Intel innovations for Ethernet, visit www.intel.com/ethernet.

Resource planning for the scale-out storage technical use case

For information about the typical workload demands and resource requirements for the scale-out storage use case, see Appendix 5.

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 (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 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, which is the simple cost of a Windows Server 2019 Datacenter license. AzS HCI is also the only HCI solution that includes optional cloud-based enhancements to ease administration.

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 simple file server, 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, 256 GB) 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			
Example node configuration	CPU	Storage	Memory	Network
	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

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: • 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="11">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 minimum vCPU/node without Storage Spaces Direct and failover needs
• 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 DC 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 DC persistent memory</td><td>2–4 x 25 Gbps</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
• Host up to 12 SQL Server instances per node</td></tr>
<tr><td rowspan="13">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)
• 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</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 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	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
• 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: 512 GB</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-S4150 (M.2) Cache: 2 x or more Intel Optane SSD DC P4800X Capacity: 4 x or more Intel SSD D3-S4510	12 x 16 GB DDR4 4 x 128 GB Intel Optane DC 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

Appendix 5: Azure Stack HCI for scale-out storage—hardware guidance

For the scale-out storage technical use case, the AzS HCI cluster is used not to host an application but to provide information workers with highly available storage on file shares. With this scenario in mind, CPU requirements per node are low, and storage capacity requirements per node are high. See Table 5 for more information.

<table>
<tr><th colspan="4">Hardware guidance for scale-out storage</th></tr>
<tr><td colspan="4">Workload assumptions
• The storage is needed to host file shares that are accessed by information workers
• The file shares include a high-availability or performance requirement that merits AzS HCI</td></tr>
<tr><td colspan="4">Design guidance
• Hardware BoM should reflect a balance of cost and performance</td></tr>
<tr><td rowspan="13">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</td><td>Low–medium</td><td>No</td></tr>
<tr><td colspan="3">• <= 16 cores (Windows Server 2019 license minimum)
• CPU cycles only needed for cluster operations
• 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>Two-tier, NVMe (cache) + HDD (capacity)</td></tr>
<tr><td colspan="3">• Medium endurance or higher in cache tier
• HDD in capacity tier for greatest density
• 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">Low</td><td>Medium</td></tr>
<tr><td colspan="2">Example: 192 GB</td><td>Example: 6 x 32 GB DDR4
or 12 x 16 GB DDR4</td></tr>
<tr><td colspan="3">Network (east/west) requirements</td></tr>
<tr><td colspan="3">Bandwidth: Low (driven by storage configuration)</td></tr>
</table>

	CPU	Storage	Memory	Network
Example node configuration	2 x Intel Xeon Silver 4215 processors	Boot: 1 x Intel SSD D3-S4150 (M.2) Cache: Intel SSD DC P4610 or Intel Optane SSD DC P4800X Capacity: Intel SSD D5-P4326 (or HDDs)	12 x 16 GB DDR4	2 x 10 Gbps to 2 x 25 Gbps Intel Ethernet Network Adapter X722

Table 5. Hardware guidance for the scale-out storage 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.