

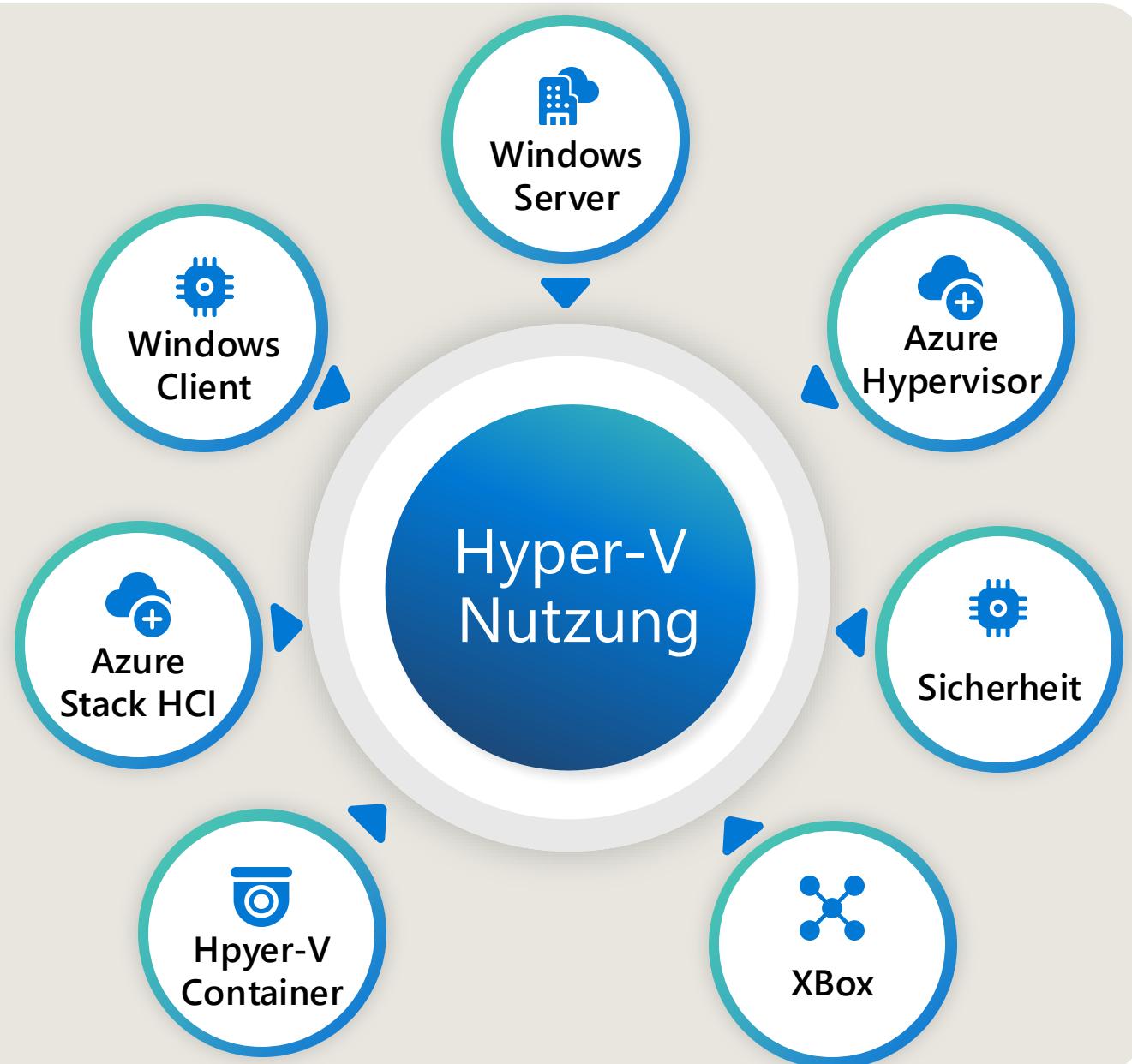


Windows Server 2025 – Virtualizierung on Steroids

**Carsten Rachfaß – Microsoft Azure und Cloud and
Datacenter Management MVP**



Hyper-V Überblick



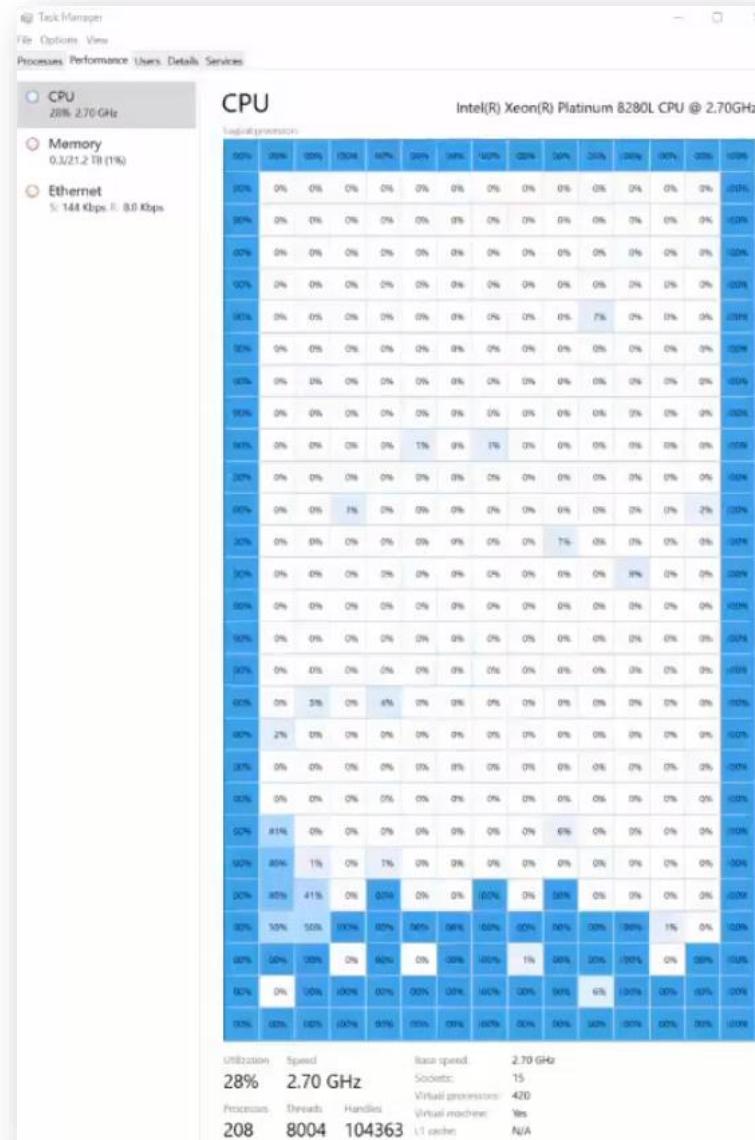
Wo wird Hyper-V genutzt?

Windows Server
Azure Stack HCI
Xbox
Azure Hypervisor
Windows Client
Hyper-V Container
Sicherheit in Betriebssystemen

Hyper-V VM Gigantonomie

Größter Kunde hat mehr als 4 Millionen Hyper-V Hosts

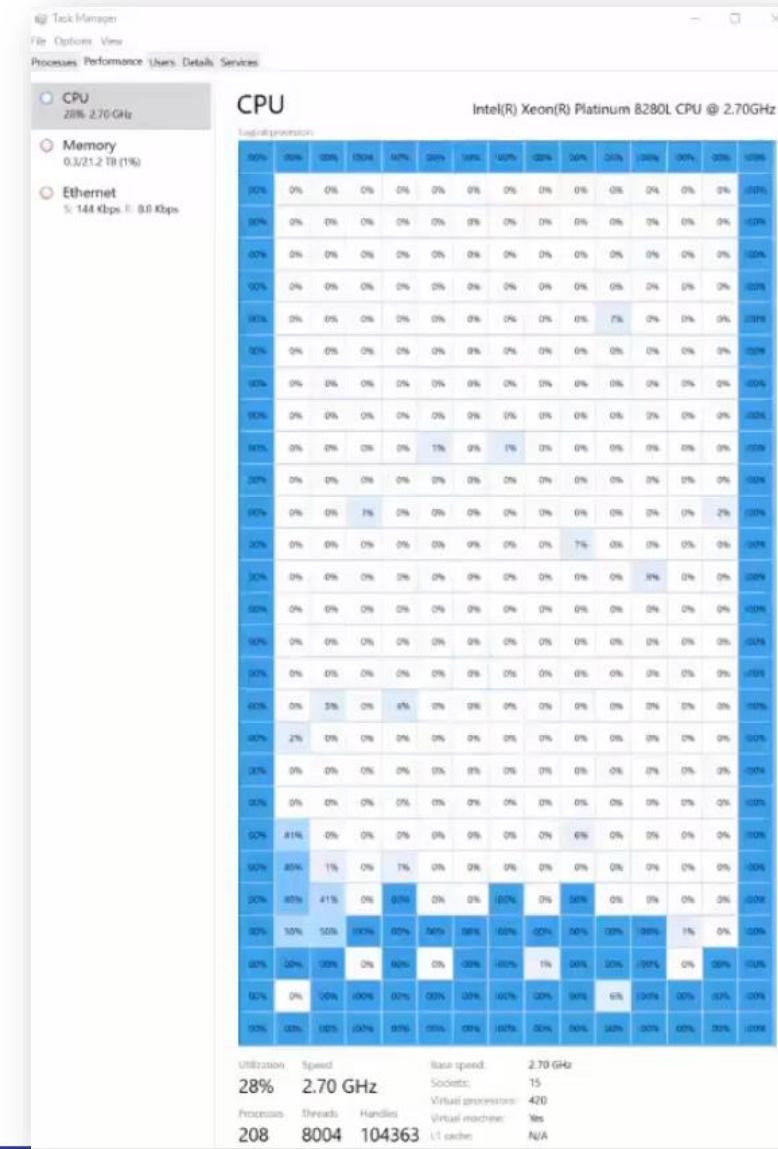
Microsoft Azure



Hyper-V VM Gigantonomie

Unterstützte Server Hardware

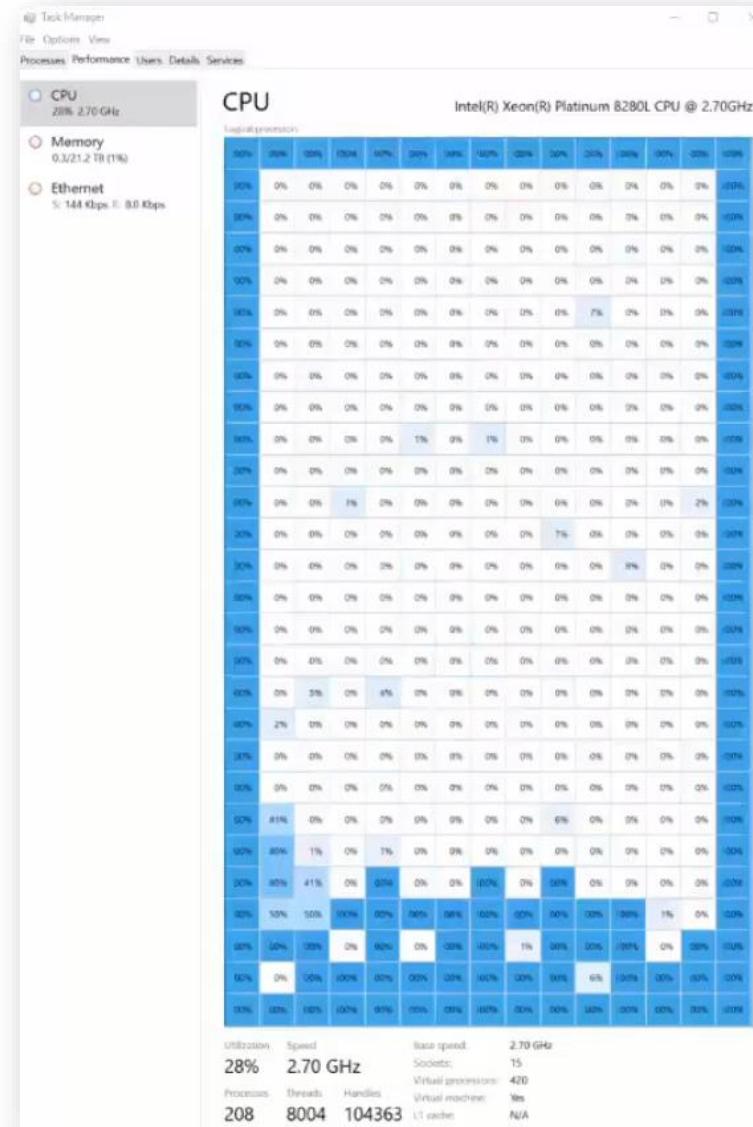
4 PB RAM
2048 CPU-Threads



Hyper-V VM Gigantonomie

Unterstütze VM-Size

240 TB RAM
2048 virtuelle CPUs



Was kann Hyper-V

Einige Features

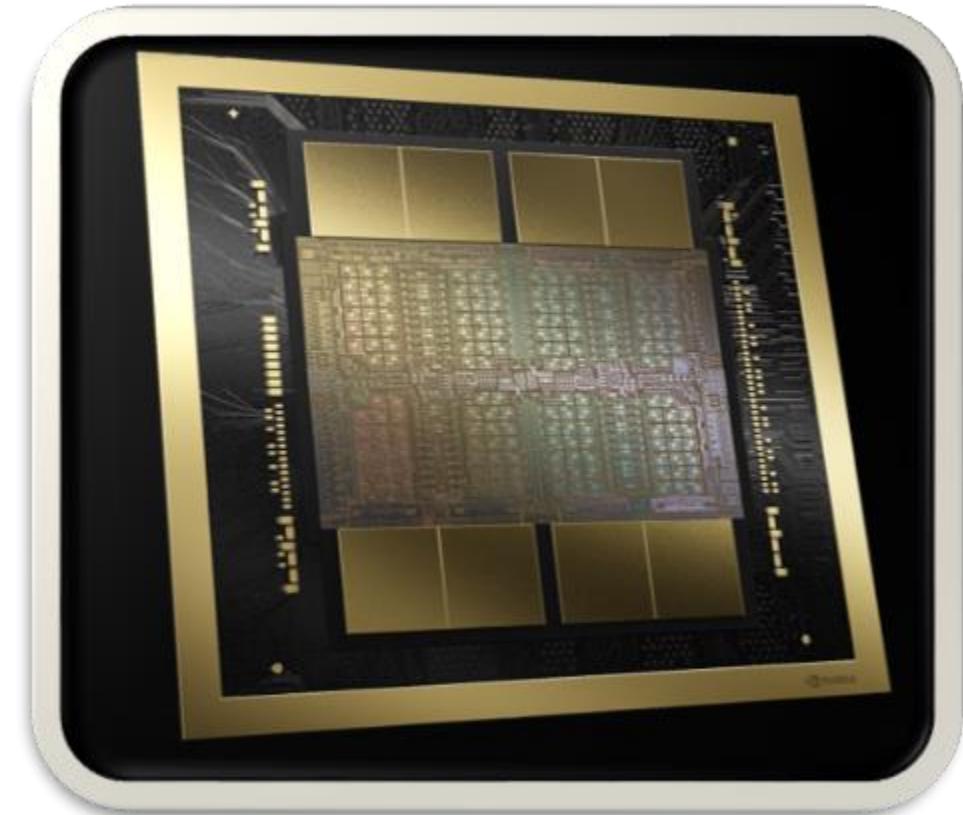
- Generation 1 VM (BIOS)
- Generation 2 VM (UEFI)
- Dynamic Memory
- Integration Components incl. In the OS since Windows 7/Server 2008
- VM with 240 vCPUs and 12 TB Memory (Generation 2)
- Live Migration
- Shared Nothing Live Migrartion
- Live Storage Migration
- Hyper-V Replica
- Storage Replica
- High availabe virtual machines
- Snapshots, checkpoints, and differencing disks
- Container and Hyper-V Container
- Shielded VMs
- Hyper-Converged Infrastructure
- SMB3 and SMB Direct
- SR-IOV and DDA-Support
- SDN

Hyper-V und GPUs



Warum GPUs in VMs

- GPU-Anforderungen in der Virtualisierung unterscheiden sich stark von denen von vor 10 Jahren
- Microsoft hat seit der Abschaffung von RemoteFX vGPU in 2020 keine zeitgemäße GPU-Unterstützung in Hyper-V (Azure Local und Windows Server)
- GPU DDA und GPU-P adressiert diesen Mangel und bietet VMs die Möglichkeit genau ihre unterschiedlichen GPU-Anforderungen zu erfüllen



Quelle: <https://nvidianews.nvidia.com/multimedia/data-center-cloud/enterprise-hpc#gallery-6>

GPU-Unterstützung in VMs

Viele Anwendung benötigen GPUs

- grafikintensive Anwendungen
- Multisession Hosts
- KI-Anwendungen

Supportete Technologien

- Discrete Device Assignment (DDA)
- GPU Partitioning (GPU-P)

Attaching GPUs on Azure Local

You can attach your GPUs in one of two ways for Azure Local:

- [Discrete Device Assignment \(DDA\)](#) - allows you to dedicate a physical GPU to your workload. In a DDA deployment, virtualized workloads run on the native driver and typically have full access to the GPU's functionality. DDA offers the highest level of app compatibility and potential performance.
- [GPU Partitioning \(GPU-P\)](#) - allows you to share a GPU with multiple workloads by splitting the GPU into dedicated fractional partitions.

Consider the following functionality and support differences between the two options of using your GPUs:

Description	Discrete Device Assignment	GPU Partitioning
GPU resource model	Entire device	Equally partitioned device
VM density	Low (one GPU to one VM)	High (one GPU to many VMs)
App compatibility	All GPU capabilities provided by vendor (DX 12, OpenGL, CUDA)	All GPU capabilities provided by vendor (DX 12, OpenGL, CUDA)
GPU VRAM	Up to VRAM supported by the GPU	Up to VRAM supported by the GPU per partition
GPU driver in guest	GPU vendor driver (NVIDIA)	GPU vendor driver (NVIDIA)

Quelle: <https://learn.microsoft.com/en-us/azure/azure-local/manage/gpu-preparation?view=azloc-24112#attaching-gpus-on-azure-local>

GPU DDA

1

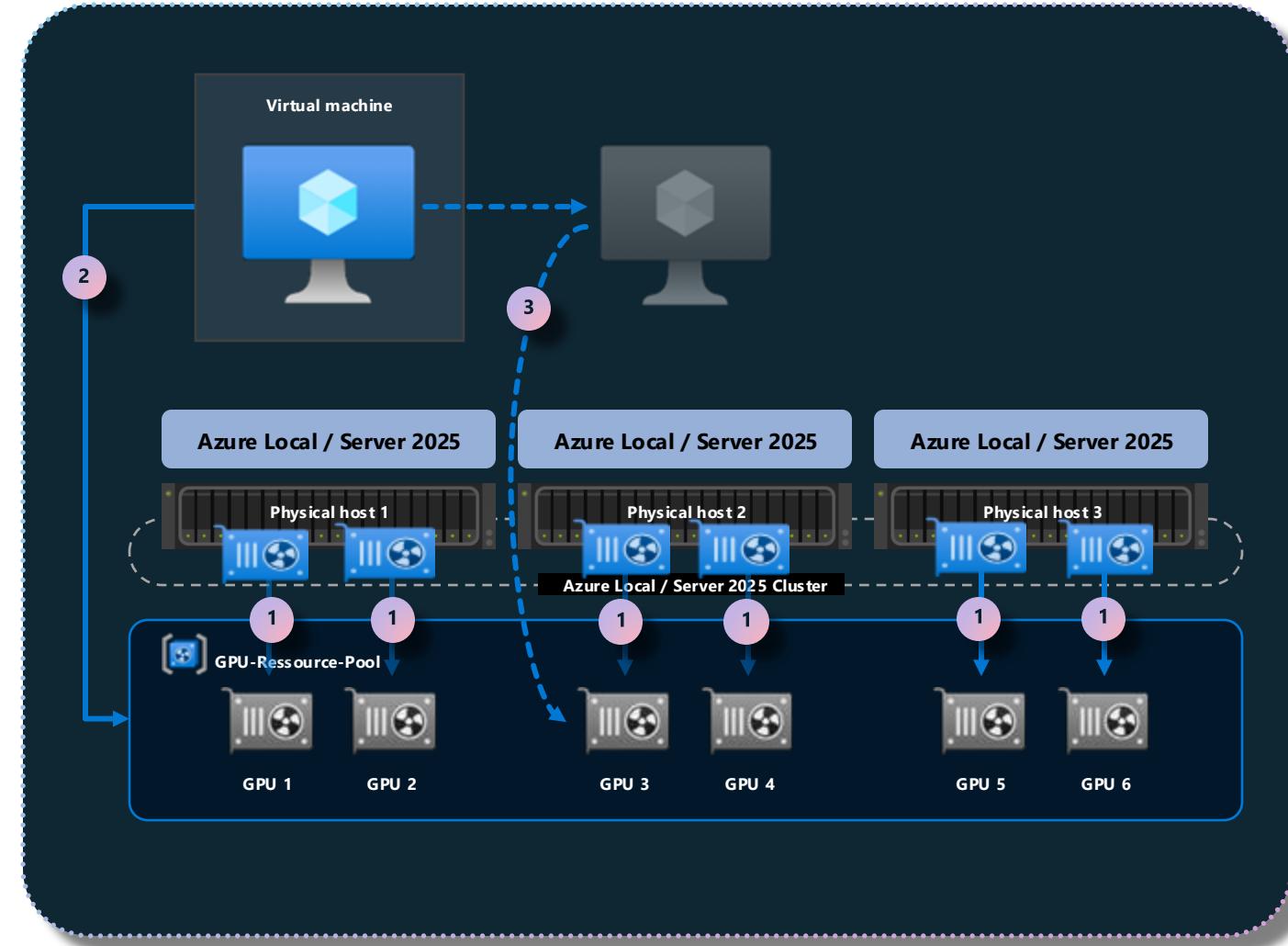
Auf jedem Knoten wird ein PCIe Ressource Pool erstellt und GPUs hinzugefügt

2

VMs werden dem GPU-Pool zugewiesen anstatt einer individuellen GPU

3

Bei einem Failover oder Move startet die VM auf einem anderen Host neu, sofern eine GPU verfügbar ist.



GPU DDA

Vorteile

- Leistung einer dedizierten GPU in der VM

Nachteile

- Nvidia Lizenzmodell
- Keine Live Migration
- Zuweisung ganzer GPU
- Wenige unterstützte GPUs
 - Nvidia A2
 - Nvidia A16

Supported GPU models

To see the full list of supported solutions and GPUs available, see [Azure Local Solutions](#) and select GPU support in the left menu for options.

NVIDIA supports their workloads separately with their virtual GPU software. For more information, see [Microsoft Azure Local - Supported NVIDIA GPUs and Validated Server Platforms](#).

For AKS workloads, see [GPUs for AKS for Arc](#).

The following GPU models are supported using both DDA and GPU-P for Arc VM workloads:

- NVIDIA A2
- NVIDIA A16

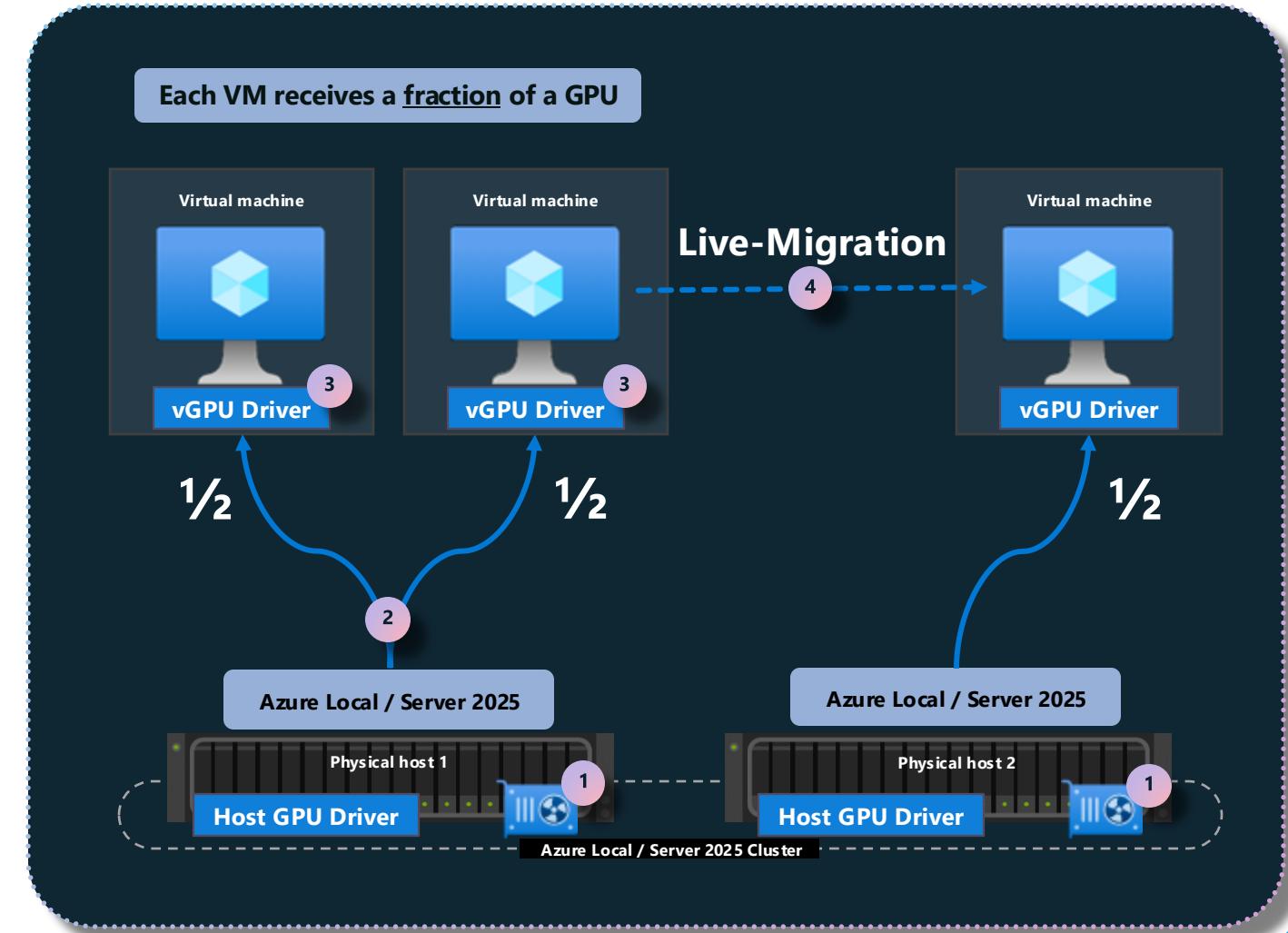
These additional GPU models are supported using GPU-P (only) for Arc VM workloads:

- NVIDIA A10
- NVIDIA A40
- NVIDIA L4
- NVIDIA L40
- NVIDIA L40S

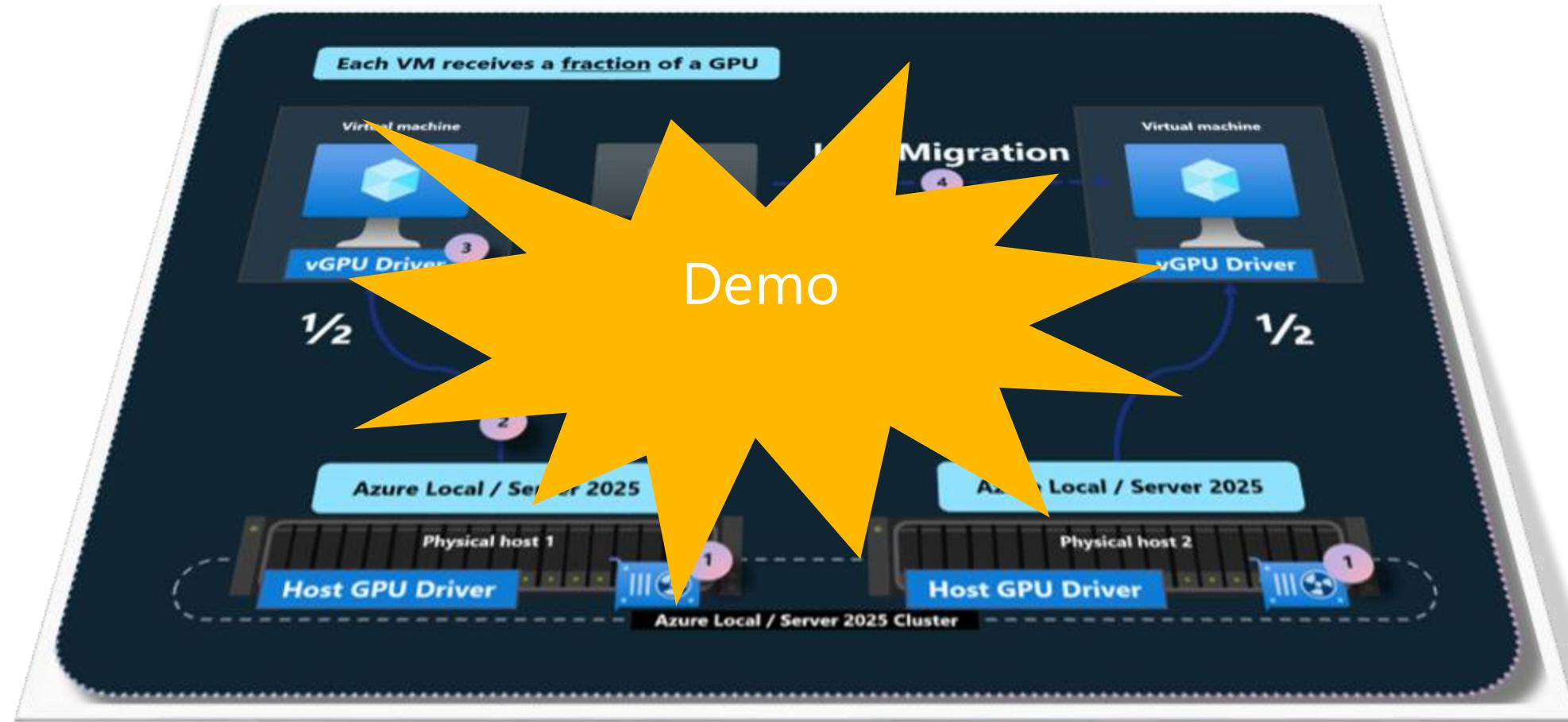
Quelle: <https://learn.microsoft.com/en-us/azure/azure-local/manage/gpu-preparation?view=azloc-24112#supported-gpu-models>

GPU-Partitioning (GPU-P)

- 1 GPU-P fähige GPU in jedem Node GPU-P Host-Treiber installieren
- 2 GPU wird auf mehrere Teile (Partitionen) aufgeteilt (mittels WAC oder PowerShell)
- 3 VM erhält den festgelegten Teil der GPU-Ressource NVIDIA vGPU-Treiber installieren
- 4 Live-Migration der VMs zwischen den Nodes möglich



GPU-P Demo



GPU-P

Vorteile

- Stücke einer GPU in VM verwendbar
- GPU Live Migration Support
- Mehr unterstützte GPUs

Nachteile

- Nvidia Lizenzmodel

Supported GPU models

To see the full list of supported solutions and GPUs available, see [Azure Local Solutions](#) and select GPU support in the left menu for options.

NVIDIA supports their workloads separately with their virtual GPU software. For more information, see [Microsoft Azure Local - Supported NVIDIA GPUs and Validated Server Platforms](#).

For AKS workloads, see [GPUs for AKS for Arc](#).

The following GPU models are supported using both DDA and GPU-P for Arc VM workloads:

- NVIDIA A2
- NVIDIA A16

These additional GPU models are supported using GPU-P (only) for Arc VM workloads:

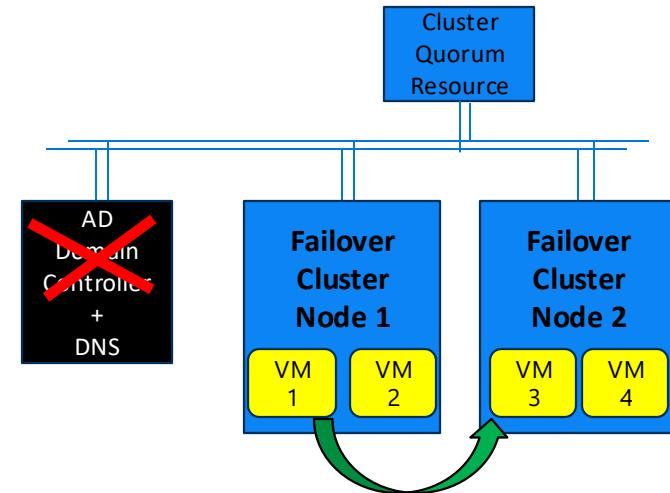
- NVIDIA A10
- NVIDIA A40
- NVIDIA L4
- NVIDIA L40
- NVIDIA L40S

Quelle: <https://learn.microsoft.com/en-us/azure/azure-local/manage/gpu-preparation?view=azloc-24112#supported-gpu-models>

Hyper-V Workgroup Cluster

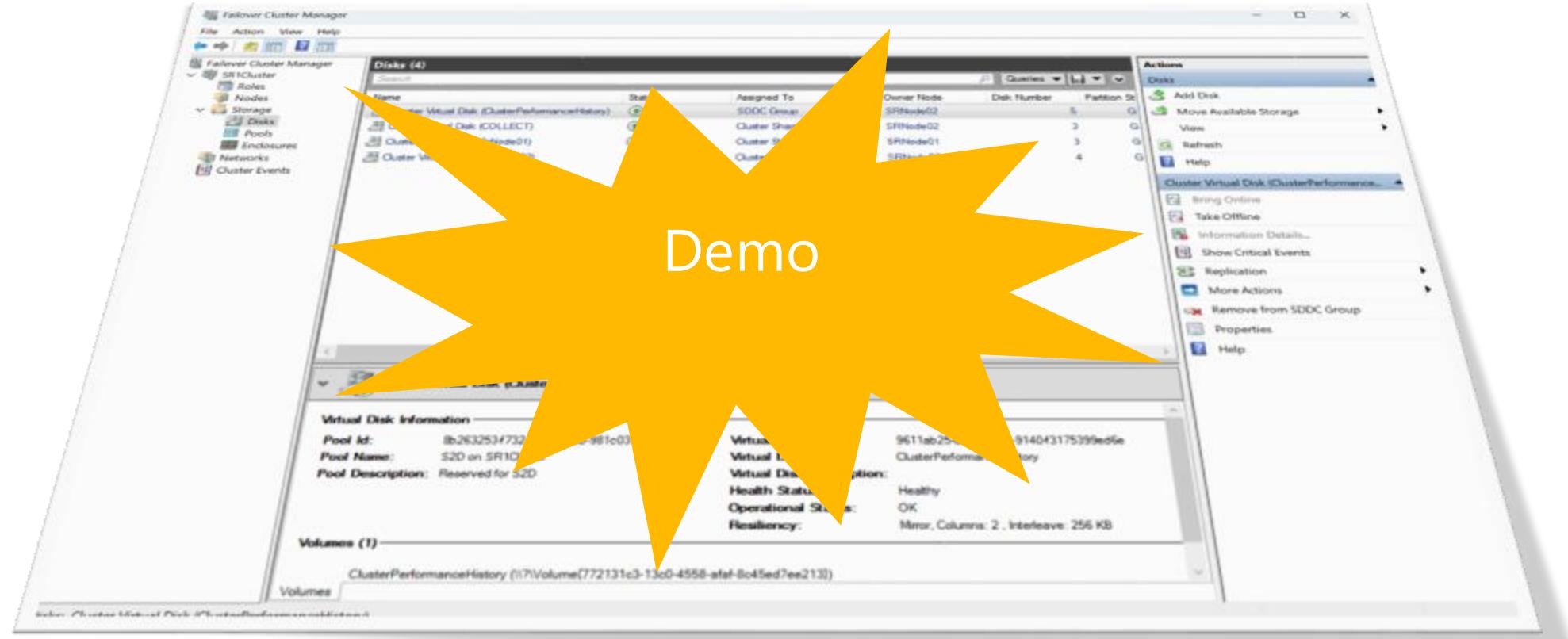
Workgroup Cluster VM Livemigration

- Workgroup Clusters eingeführt mit Windows Server 2016
- Usecase SQL Server Availability Groups ohne AD um SQL Server Database Mirroring zu ersetzen
- Kunden möchten auch Hyper-V Failover Cluster ohne Active Directory einsetzen
- Man kann auch Hyper-V Workgroup Cluster einsetzen aber die Live Migration funktioniert nicht, zumindest bis jetzt



Empfohlen für **Hyper-V VMs**, und **SQL Server Availability Groups** – nicht empfohlen für File Servers (Auth Probleme) oder SQL Server FCI

Workgroup Cluster



Beispiel: Erzeugen eines Workgroup Cluster

```
#Install WorkGroup Hyper-V Cluster
# stolen from Window Server Summit Demo https://www.youtube.com/watch?v=XJNht06_wYQ&t=276s by Rob Hindman
$clusterName = 'SR1Cluster'
$clusterIP = '192.168.209.41'
$Nodes = 'SrNode01', 'SRNode02'
$SecureStringPassword = Read-Host -Prompt "Enter your password" -AsSecureString
$Password =
[System.Runtime.InteropServices.Marshal]::PtrToStringAuto([System.Runtime.InteropServices.Marshal]::SecureStringToBSTR($SecureStringPassword))

#region install cluster
foreach ($Node in $Nodes) {
    Invoke-Command -ComputerName $Node -ArgumentList $Nodes[0], $Nodes[1] -ScriptBlock {
        param ($Node1, $Node2)
        New-ItemProperty -Path HKLM:\SOFTWARE\Microsoft\Windows\CurrentVersion\Policies\System -Name
LocalAccountTokenFilterPolicy -Value 1
        Set-Item WSMan:\localhost\Client\TrustedHosts -value "$Node1,$Node2" -Force
    }
}

New-Cluster -Name $clusterName -Node $Nodes -StaticAddress $clusterIP -NoStorage -AdministrativeAccessPoint Dns
#endregion

#region enable S2D and create Volumes
Enable-Clusters2D -Confirm:$false -verbose

New-Volume -FriendlyName COLLECT -StoragePoolFriendlyName S2D* -Filesystem CSVFS_ReFS -Size 200GB -ProvisioningType Thin
foreach ($Node in $Nodes) {
    New-Volume -FriendlyName $Node -StoragePoolFriendlyName S2D* -Filesystem CSVFS_ReFS -Size 400GB -ProvisioningType Thin
}
#endregion

...
```

Beispiel: Erzeugen eines Workgroup Cluster

```
...
#region Configure Cluster Networks with Network ATC
Add-NetIntent -Name computer_management -Compute -Management -AdapterName IntelX540NIC1, IntelX540NIC2
Add-NetIntent -Name storage -Storage -AdapterName MLNXNIC1, MLNXNIC2

sleep 600

#Configure Hyper-V by modifieng the GlobalClusterOverrides
$ClusterOverrides = New-NetIntentGlobalClusterOverrides
$ClusterOverrides.MaximumVirtualMachineMigrations = 5
$ClusterOverrides.MaximumSMBMigrationBandwidthInGbps = 50
Set-NetIntent -GlobalClusterOverrides $ClusterOverrides
#endregion

#region deploy VMFleet
New-Item -Path C:\ClusterStorage\COLLECT\Images -ItemType Directory
robocopy C:\System\Images\ C:\ClusterStorage\COLLECT\Images\ VMFleetTemplate.vhdx
Install-Module VMFleet -Force
Install-Fleet
New-Fleet -BaseVHD C:\ClusterStorage\COLLECT\Images\VMFleetTemplate.vhdx -VMs 10 -Groups pk -AdminPass $Password -Admin
Administrator -ConnectPass $Password -ConnectUser Administrator

#Disable IPv6 on VMFleet Internal Adapter (to hide them in ClusterNetworks)
foreach ($Node in $Nodes) {
    Invoke-Command -ComputerName $Node -ScriptBlock {
        Get-NetAdapter | where Name -like "*FleetInternal*" | Set-NetAdapterBinding -ComponentID ms_tcpip6 -Enabled
>false
    }
}
#endregion
```

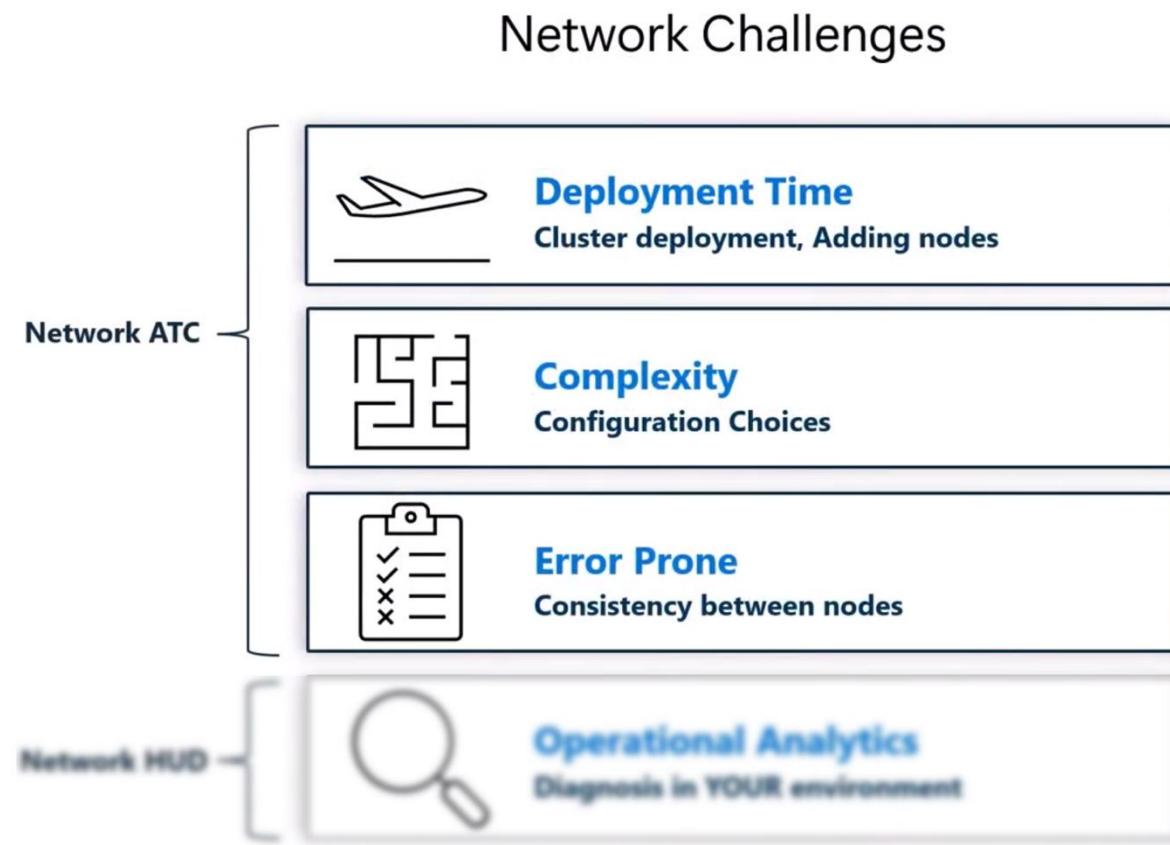
Network ATC and Network HUD



Network ATC

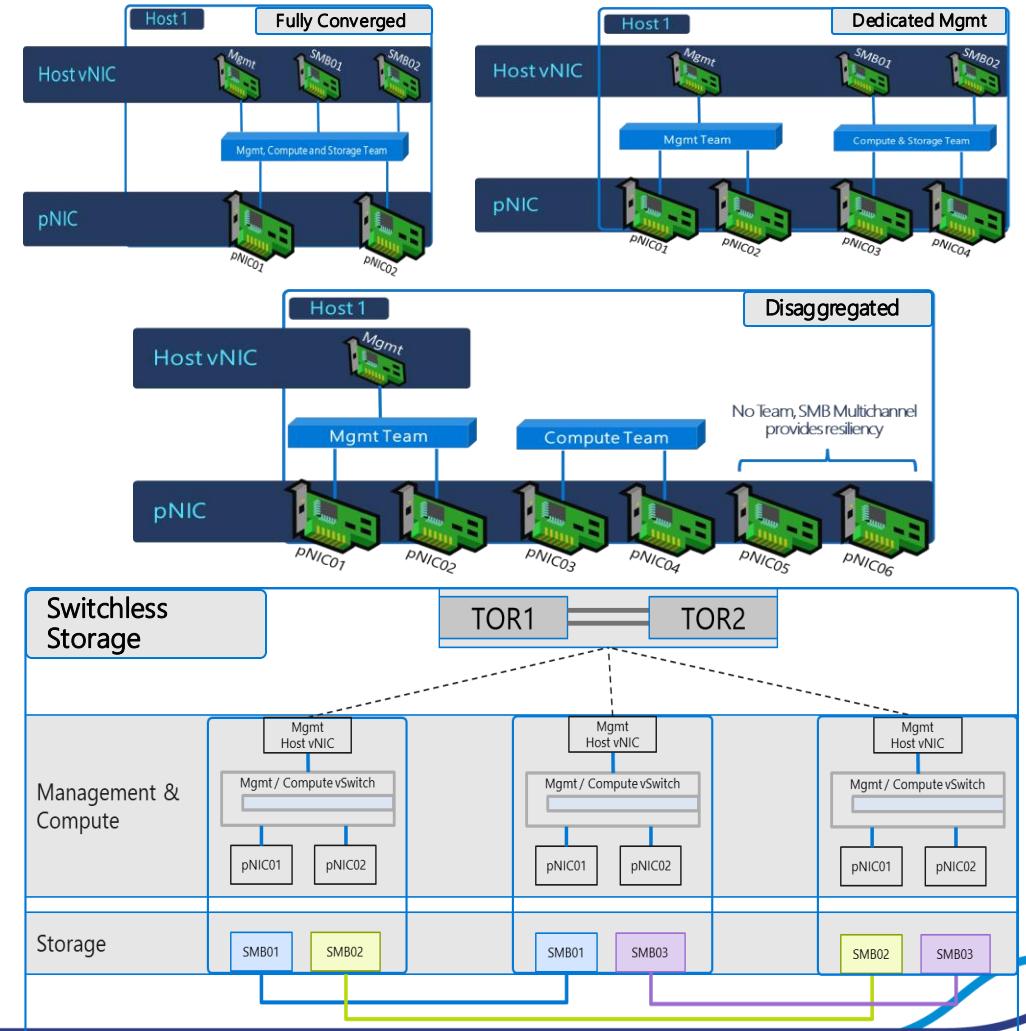
Network ATC adressiert
Herausforderungen im Netzwerk

- Deployment Zeit
- Komplexität
- Wiederholbarkeit
- Veränderung (Drift)
- Error Anfälligkeit

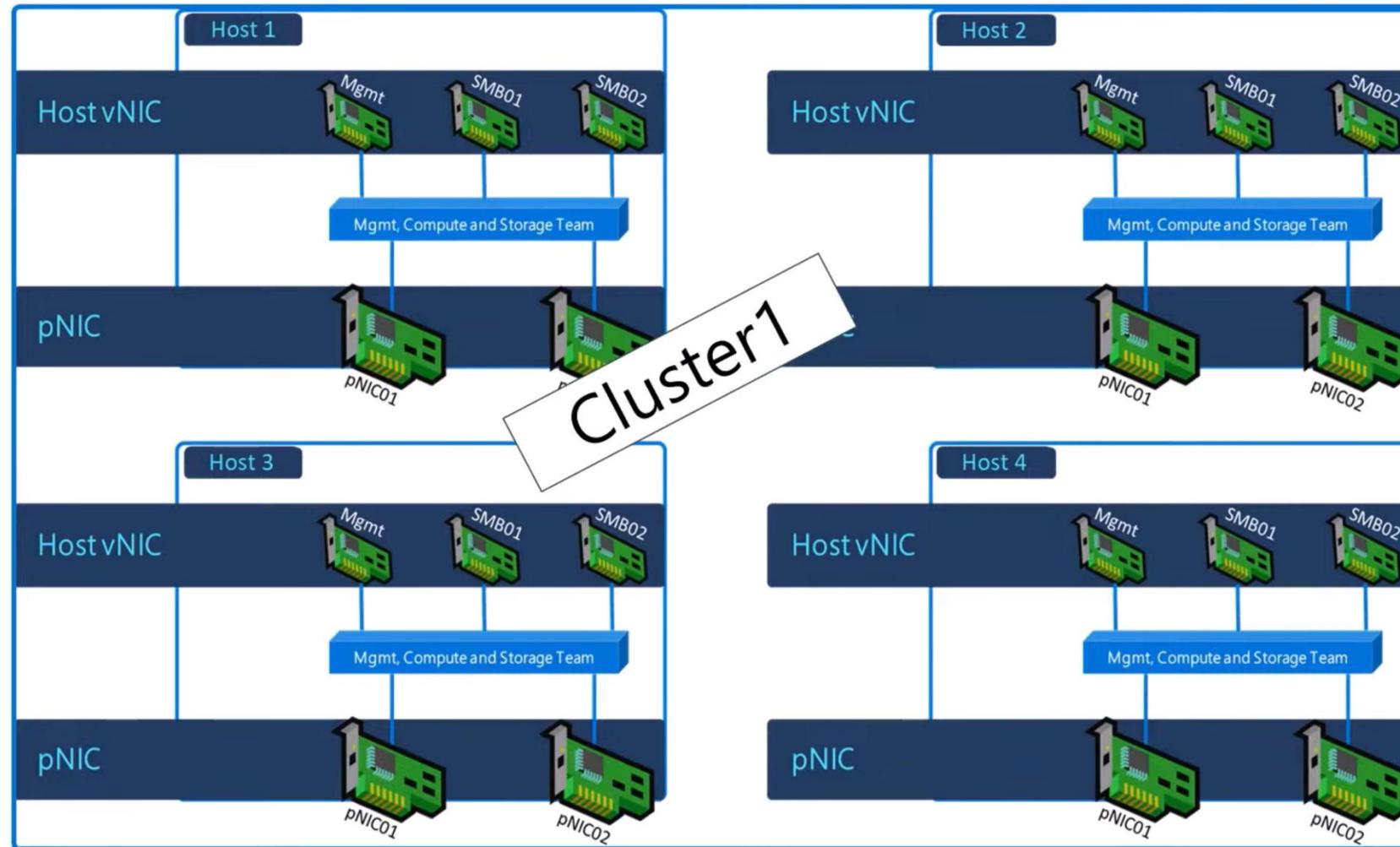


Netzwerk Herausforderungen

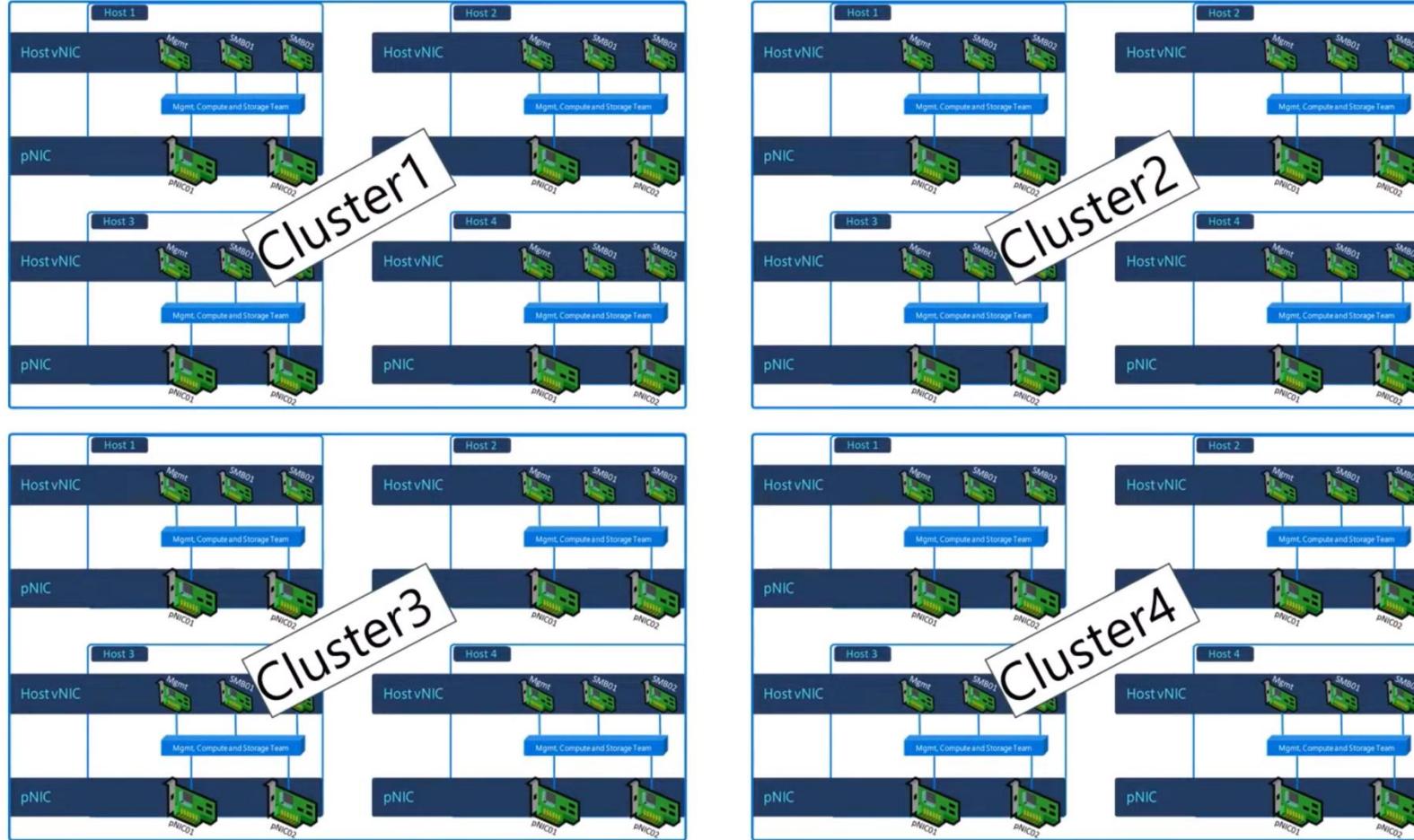
- **Adapter Settings**
 - Advanced Properties
 - RSS/VMQ/VMMQ
- **DCB QoS**
- **vSwitch Configuration**
 - Teaming, Load Balancing Algorithm
 - IOV Enablement
 - IEEE Priority Tagging
 - Default Queue Configuration
- **Miscellaneous**
 - Management and Storage VLANs
 - MTU Configuration, DNS Registration, etc.
 - Host vNIC Team mapping
- **WAC Deployment (Day-0)**
- **Storage Auto-IP**
- **Local and Network Symmetry**
- **Cluster Settings**
 - Network Naming
 - Live Migration Transport
 - Max VM Migrations, LM Bandwidth Limits
- **Proxy Settings**
- **Stretch Clustering**
- **Improved Diagnostics**
- **WS22 New Adapter Properties**
- **Full 2-Node Switchless deployment**
- **WAC Deployment (Day-100)**



Netzwerk Herausforderungen



Netzwerk Herausforderungen



Network ATC

- Übersetzt meine Absicht (Intent) für die Netzkonfiguration und führt diese durch
- Setzt die letzten Microsoft validierten und supporteten Best Practice um
- Sichert Konfigurationskonsistenz im ganzen Cluster
- Eliminiert Konfigurationsdrift
- Reduziert Hostnetzwerk Deployment Zeit, Komplexität und Fehleranfälligkeit
- Reduziert Supportcalls für euch und Microsoft ;-)
- Muss in Azure Local, Option in Windows Server 2025

Network Intents

Management

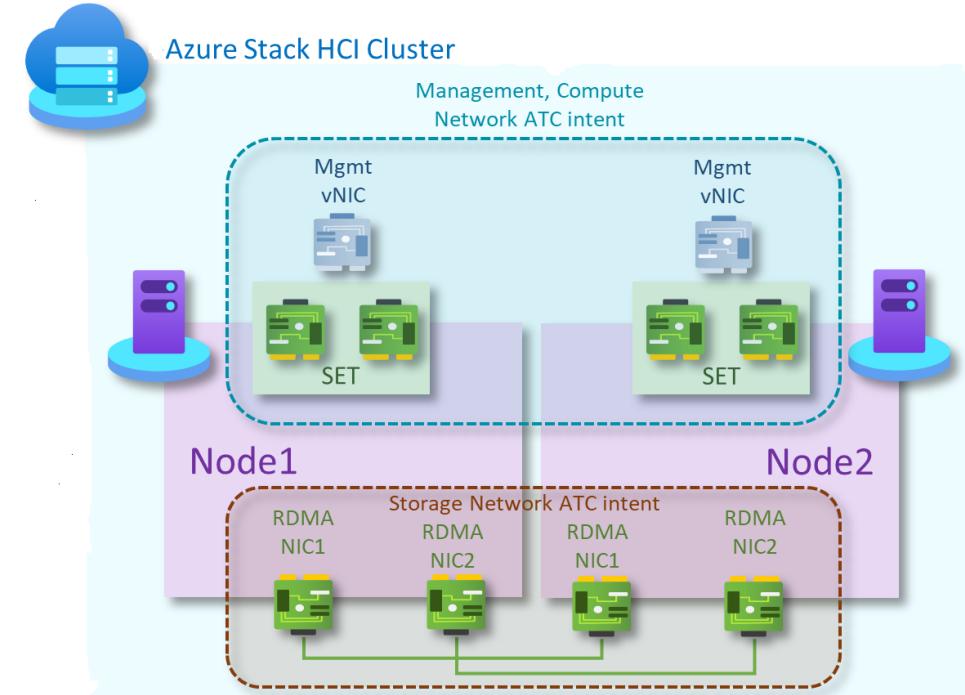
- Über diesen Intent läuft alle externe Knotenkommunikation
- AD, DNS, DHCP, Update, Management, Monitoring, Internet, Backup, Heartbeat, ...

Compute

- Netzwerkzugriff der Virtuellen Maschinen

Storage

- Storage Kommunikation, Heartbeat Clusterkommunikation, Live Migration



Network ATC zusätzliche Features

Cluster Overrides	Duplizierung	Dokumentation
<ul style="list-style-type: none">• Anzahl parallele Livemigrationen• Anzahl Storagemigration• Livemigration Bandbreite• Livemigration Netzwerk Auswahl	<ul style="list-style-type: none">• Exakte Kopie der Einstellungen• Beinhaltet alle Intents und Overrides	<ul style="list-style-type: none">• FAQ-Seite• Allgemeine Fehler Meldungen• PS Module Dokumentation

Network ATC zusätzliche Features

Day-100 Management UI In WAC

- Intents ansehen und verwalten
- Overrides managen und ausrollen
- Intent Status und Health ansehen

Brownfield Migrationen

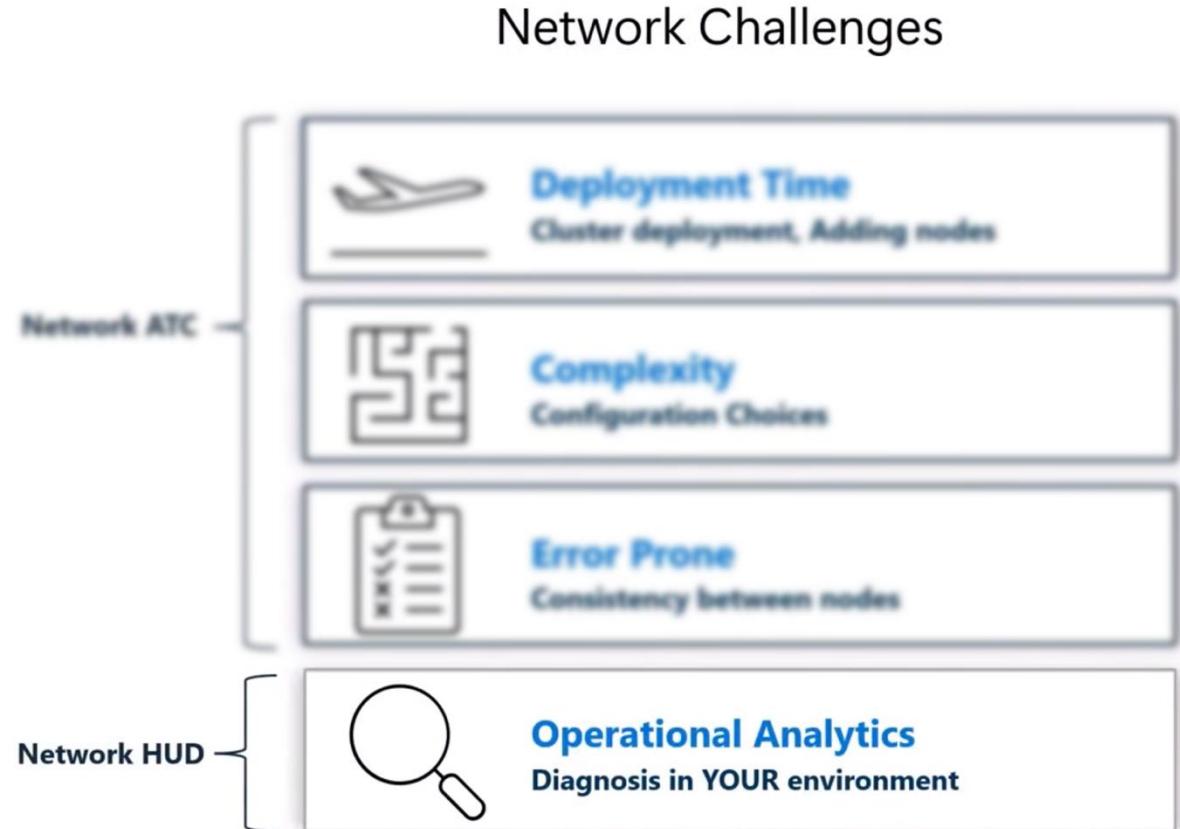
- Migration von Clustern mit klassischem Netzwerk in Network ATC Cluster
- Keine Downtime
- Sofortiges Management, Drift Korrektur mit voreingestellter Konfiguration

Detailliertes, Granulares Event Logging

- Detailliertes step-by-step Eventlogging
- Schnellere Diagnose
- Schnellere Problembehebung
- Einfache Benutzung

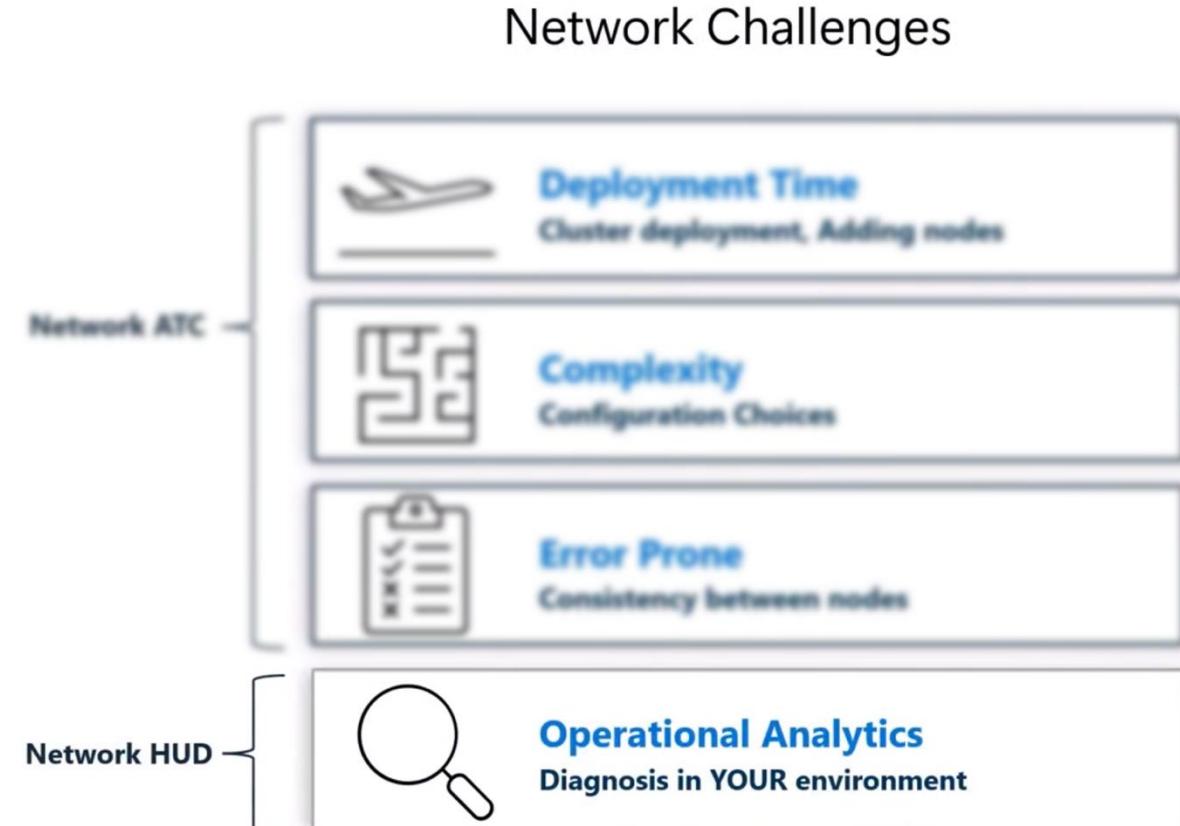
Network HUD

- Release mit Windows Server 2025
- ARC-Extension
- Integriert mit
 - Windows Admin Center
 - Azure Portal
 - Network ATC
- Verwendet Switch Informationen

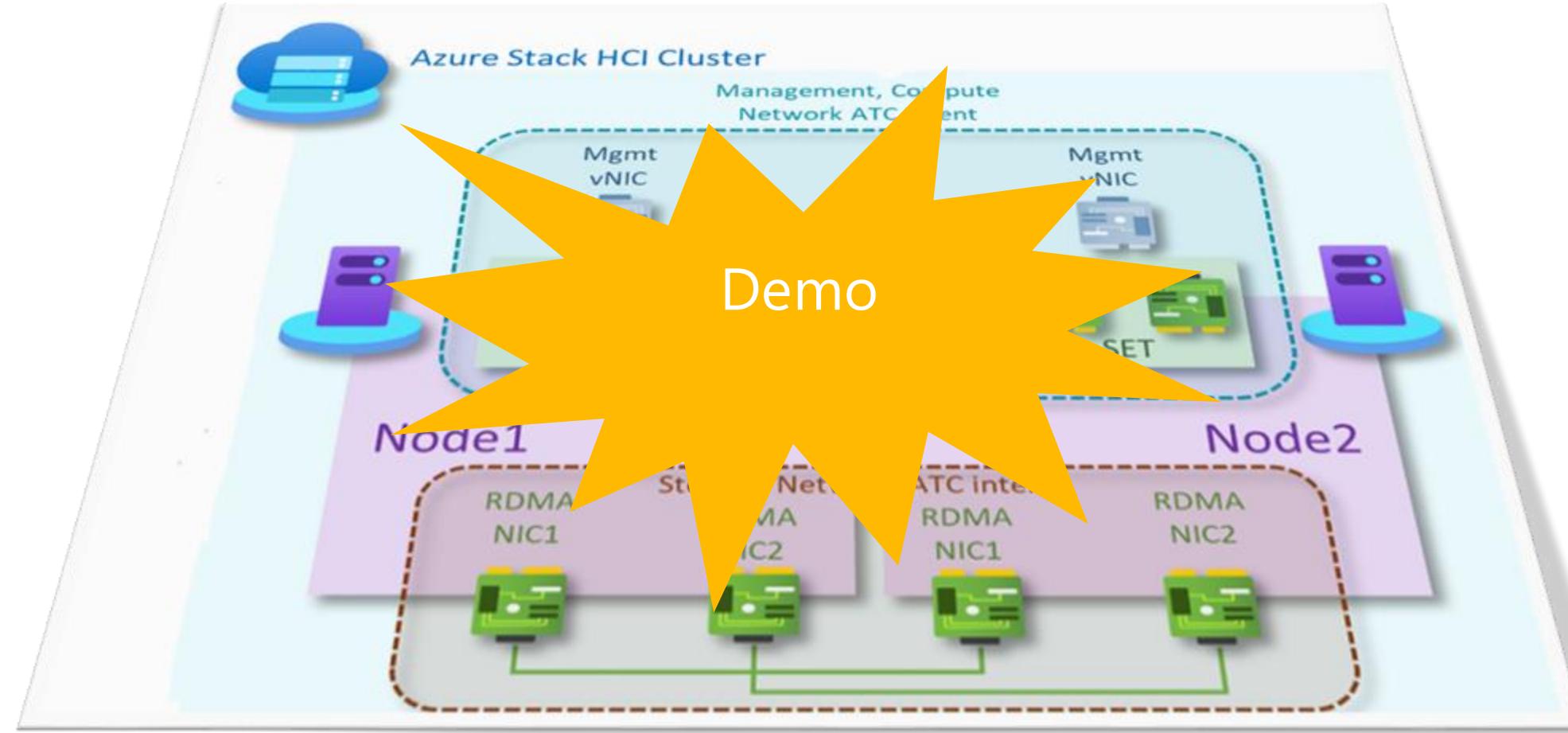


Network HUD

- Erkennt folgende Probleme
- NIC-Treiber und NIC Health
 - Inbox Treiber
 - Warnungen wenn Treiber 2 Jahre alt ist
 - Fehler wenn Treiber 3 Jahre alt ist
 - Unstabile Adapter
 - PCIe Bandbreiten Oversubscription
- Network ATC
 - Fehlende Intent Typen
 - Physische Netzwerk Konfiguration (kommt bald)
 - Fehlende VLANs auf pNICs (kommt bald)



Network ATC Demo



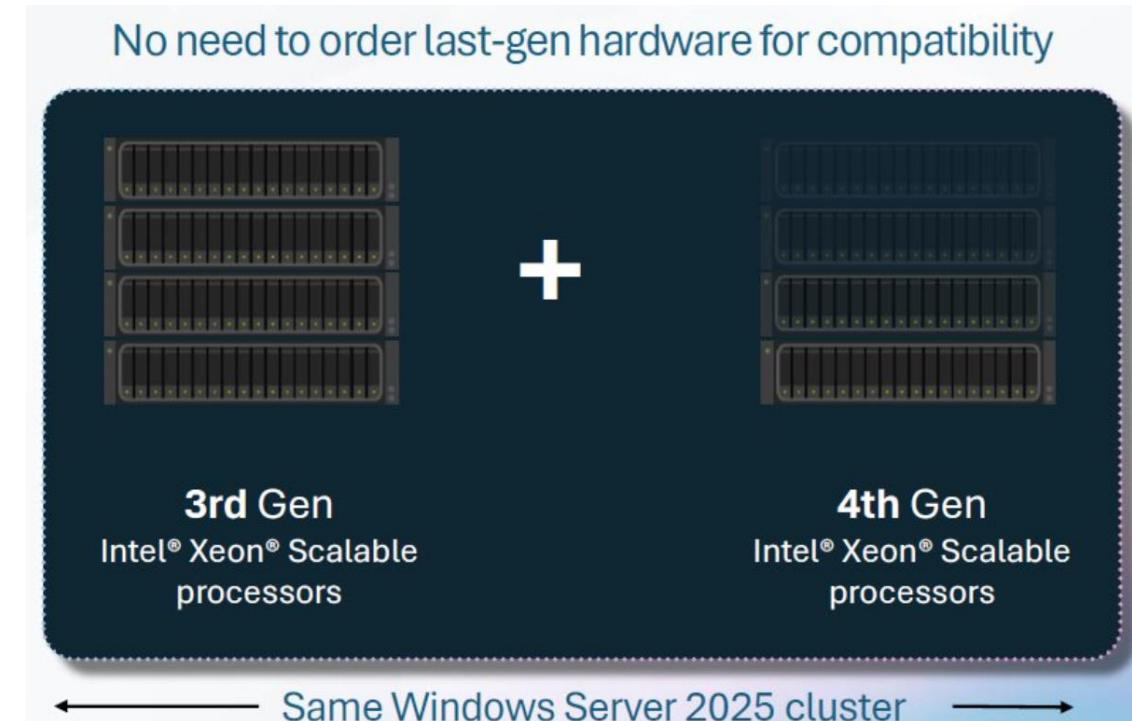
Dynamic Compatibility



Dynamic Processor Compatibility

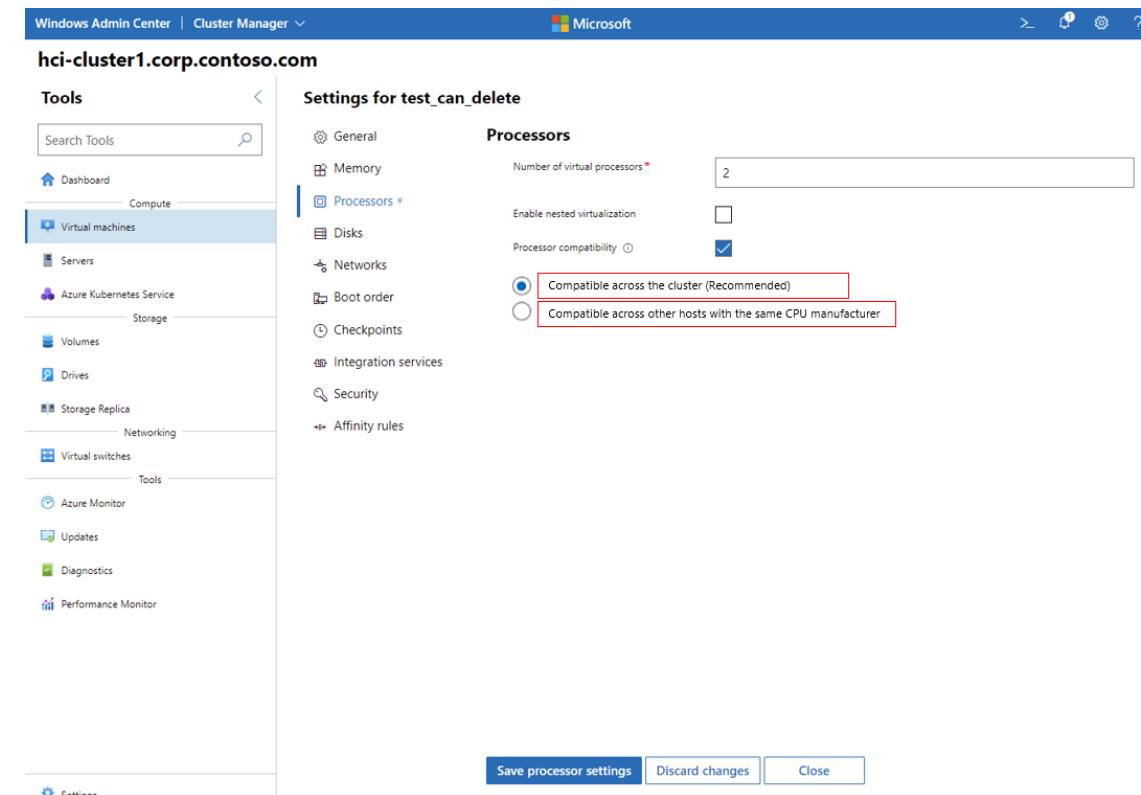
Bisherige Lösung: Prozessor Kompatibilität

- Es wird die CPU auf ein relative altes CPU Feature Set reduziert
- Folge: sowohl VMs auf der neueren Hardware wie auf der älteren Hardware verlieren Prozessor Eigenschaften



Dynamic Processor Compatibility

- Dynamic Processor Compatibility
 - Cluster misst Prozessoreigenschaften der Knoten aus
 - Stellt Prozessorkompatibilität mit maximalem Feature Set ein
- Folge: VMs steht maximale Prozessoreigenschaften der Hardware zur Verfügung



NVMe



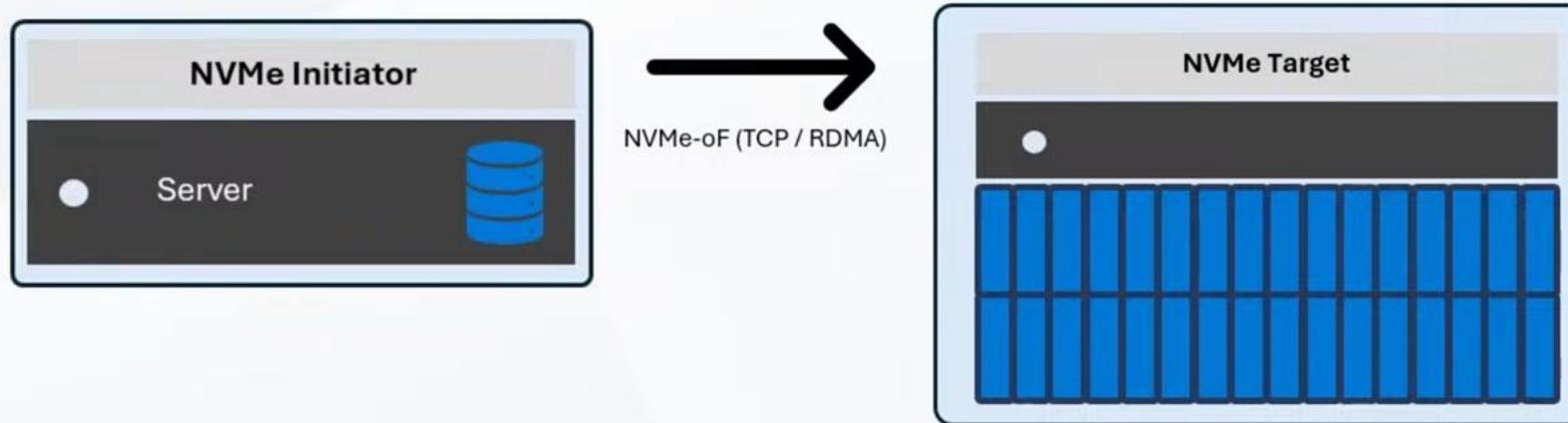
Coming soon to Windows Server 2025

Full stack, Native NVMe Support

- WS2025 delivers up to 70% more IOPS on NVMe SSDs
- After GA, WS2025 will bring up to 90% more IOPS
- This innovation in full stack NVMe
 - Lowers host CPU consumption = More VMs per node
 - Increases IOPS = increased performance per VM

NVMe-oF Initiator

- Built-in NVMe-oF client for connecting to NVMe SANs
- Enables additional storage topologies
- TCP for traditional workloads
- Roadmap: RDMA support for lowest latency workloads



25% of Enterprises will deploy NVMe-oF solutions by 2027

Gartner - Top Trends in Enterprise Data Storage 2023

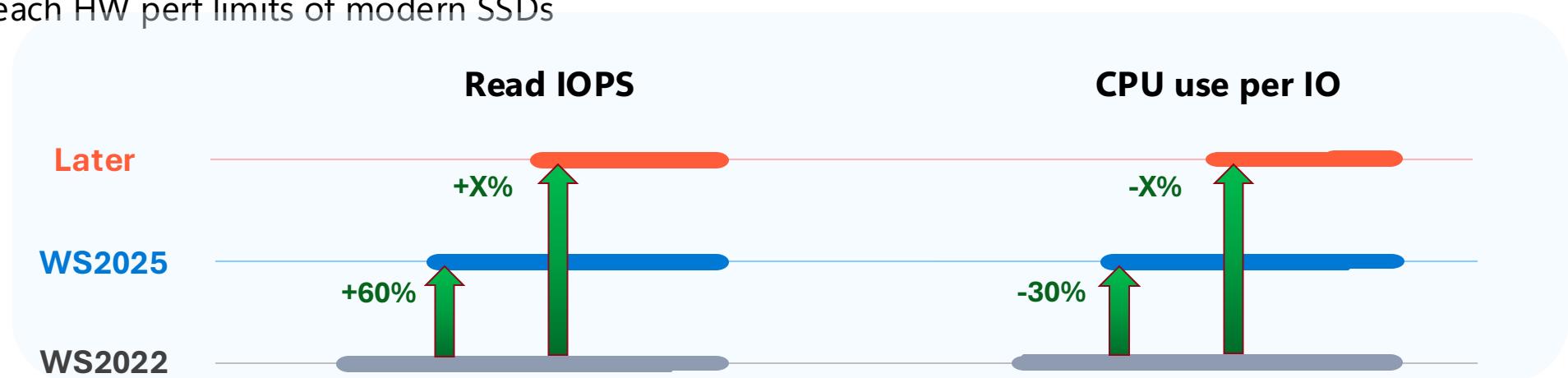
NVMe Stack Optimization



Modernize NVMe storage stack to deliver improved IO performance on NVMe devices

WS2025 introduced improved performance on PCIe NVMe devices.

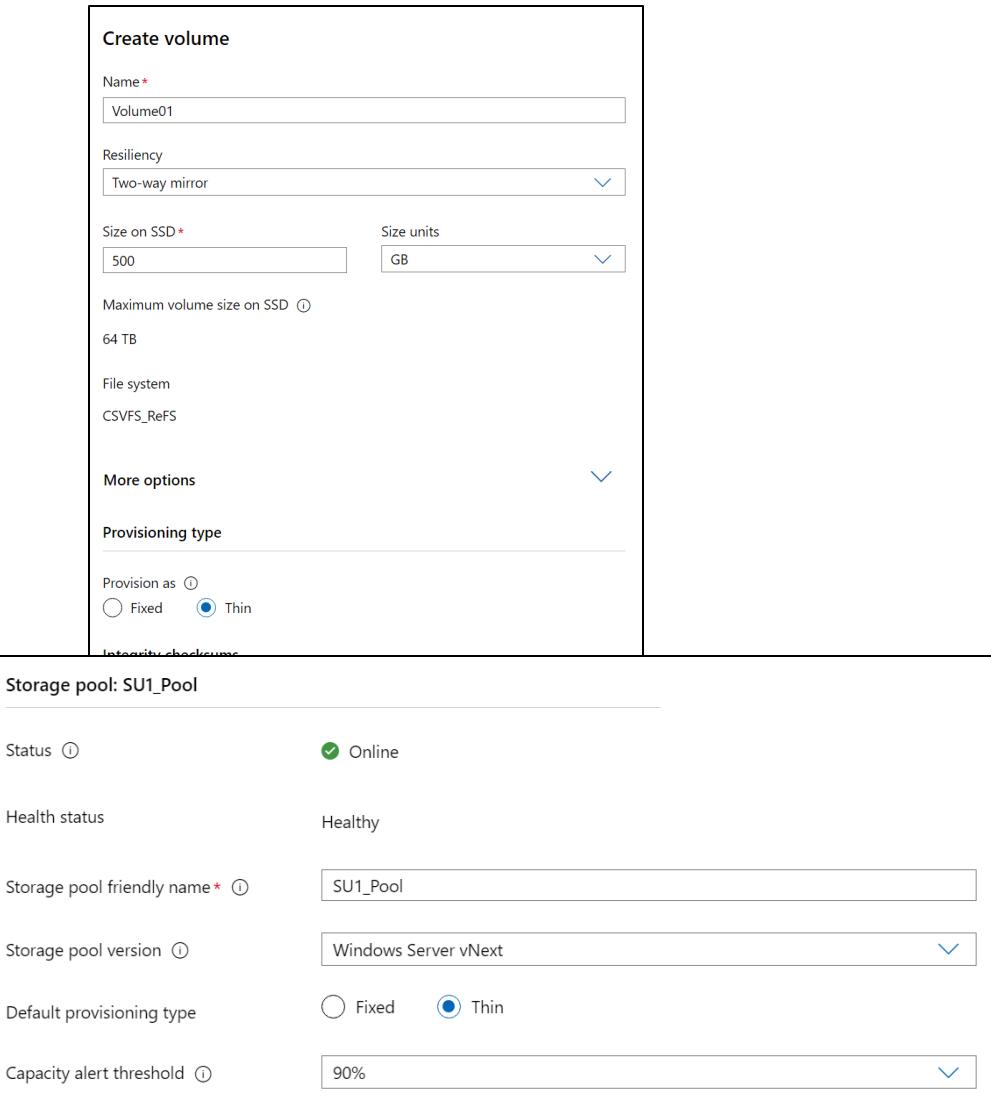
- Further optimizations underway
 - a.k.a. 'Native NVMe stack' or 'StorMQ support'
 - reach HW perf limits of modern SSDs



Thin Provisioned Cluster Volumes

Thin Provisioned Volumes

- In Windows Server 2025 S2D supportet
- Ermöglicht Überprovisionierung
- Extents werden erst bei Belegung zugeordnet
- Kann von „Fixed“ zu „Thin“ konvertiert werden
- Kann bei der Volume Anlage ausgewählt werden
- Kann in den Storage Pool Einstellungen als Default gesetzt werden



The screenshot displays two windows from the Windows Server 2025 Storage Pool configuration:

- Create volume** window:
 - Name: Volume01
 - Resiliency: Two-way mirror
 - Size on SSD: 500 GB
 - Maximum volume size on SSD: 64 TB
 - File system: CSVFS_ReFS
 - Provisioning type: Thin (selected)
- Storage pool: SU1_Pool** window:
 - Status: Online
 - Health status: Healthy
 - Storage pool friendly name: SU1_Pool
 - Storage pool version: Windows Server vNext
 - Default provisioning type: Thin (selected)
 - Capacity alert threshold: 90%

Neue ReFS Deduplication and Compression



ReFS Deduplication and Compression

Speicherplatz Ersparnis

- Über 60% Platzersparnis in Virtualisations-, Backup- und Fileserver-Workloads

Einfaches Management

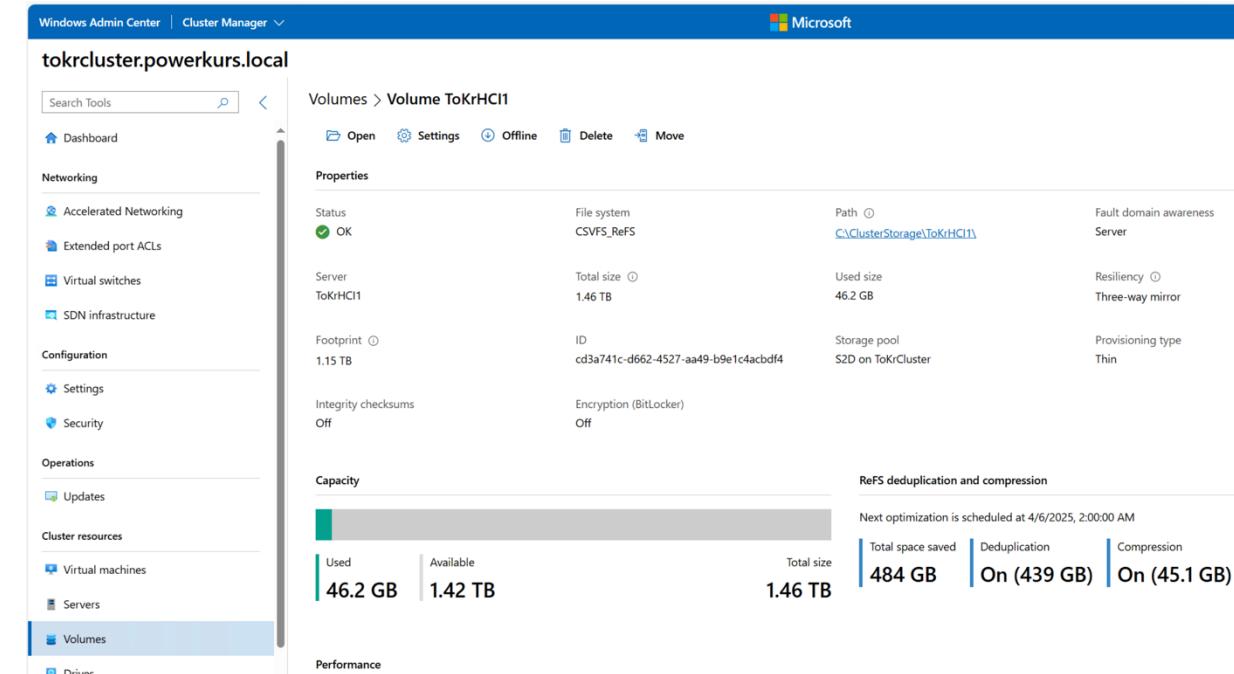
- Windows Admin Center oder PowerShell
- planen, monitoren und starten der Optimierung

Smart und Effizient

- Clusteraware, geringer Overhead (<1 ms)
- Dedupliziert nur neue oder geänderte Daten

Mehrere Modi

- Deduplication only, Compression only, oder Beides (default)
- Kompression hat zwei unterschiedliche Algorithmen

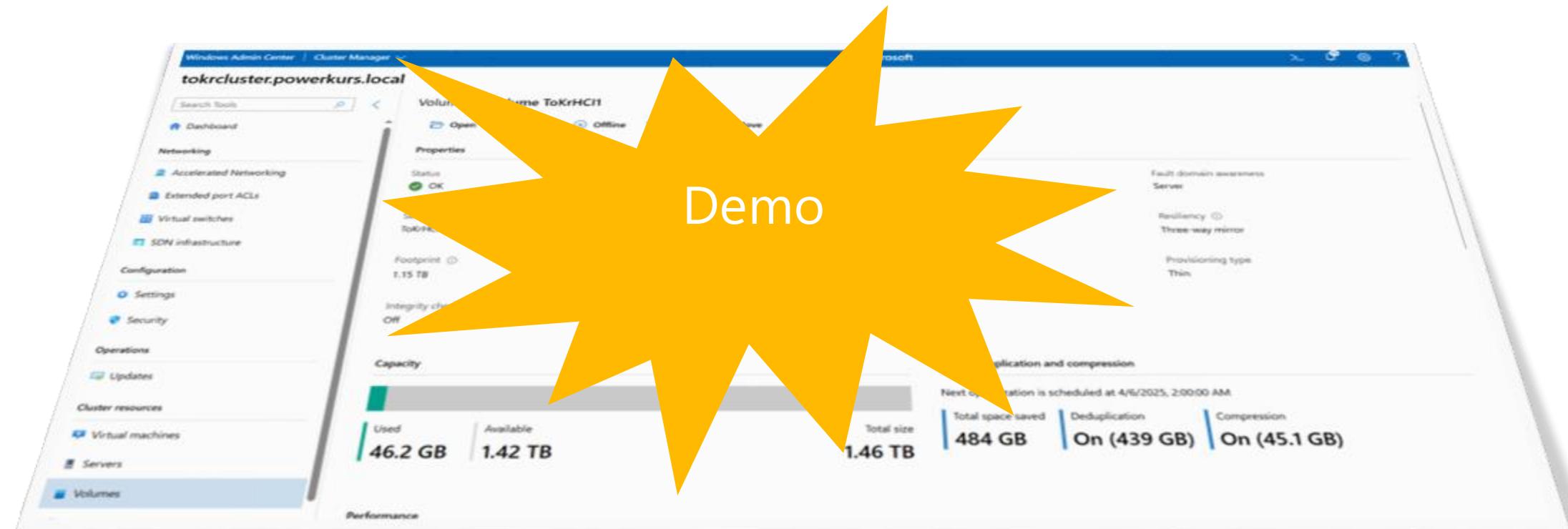


Properties

Setting	Value	Notes
Status	OK	File system: CSVFS_ReFS, Path: C:\ClusterStorage\ToKrHCI1
Server	ToKrHCI1	Total size: 1.46 TB, Used size: 46.2 GB
Footprint	1.15 TB	ID: cd3a741c-d662-4527-aa49-b9e1c4acbd4, Storage pool: S2D on ToKCluster
Integrity checksums	Off	Encryption (BitLocker): Off
Capacity	Used: 46.2 GB, Available: 1.42 TB, Total size: 1.46 TB	ReFS deduplication and compression: Total space saved 484 GB, Deduplication On (439 GB), Compression On (45.1 GB)

Next optimization is scheduled at 4/6/2025, 2:00:00 AM

ReFS Dedup / Thin Provisioning Demo



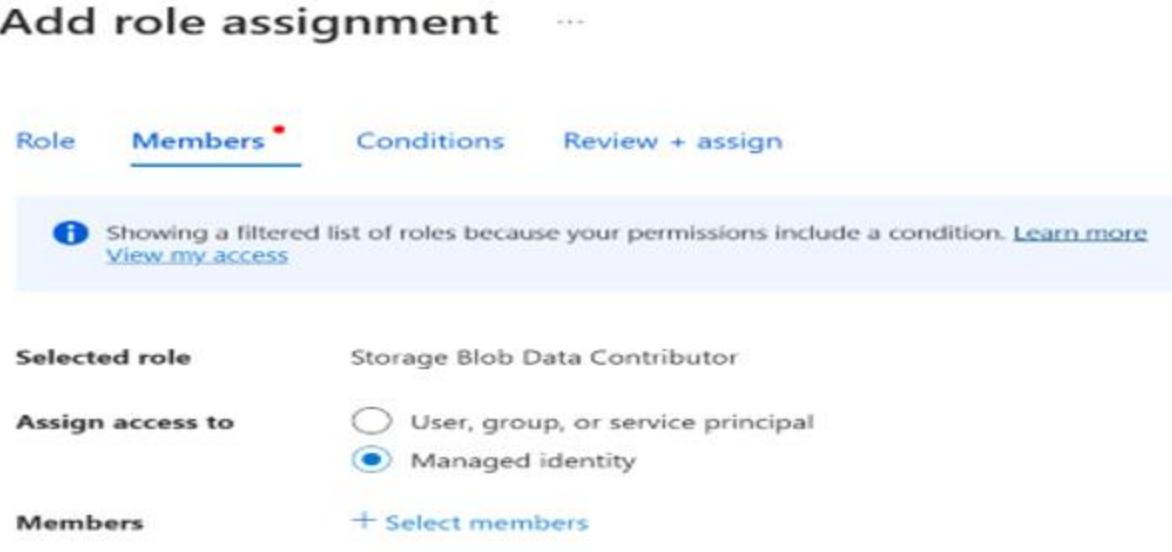
Coming Soon



Cloud Witness with Managed Identity

Home > cloudwitnesstestentra | Access Control (IAM) >

Add role assignment



Showing a filtered list of roles because your permissions include a condition. [Learn more](#)

Selected role Storage Blob Data Contributor

Assign access to User, group, or service principal Managed identity

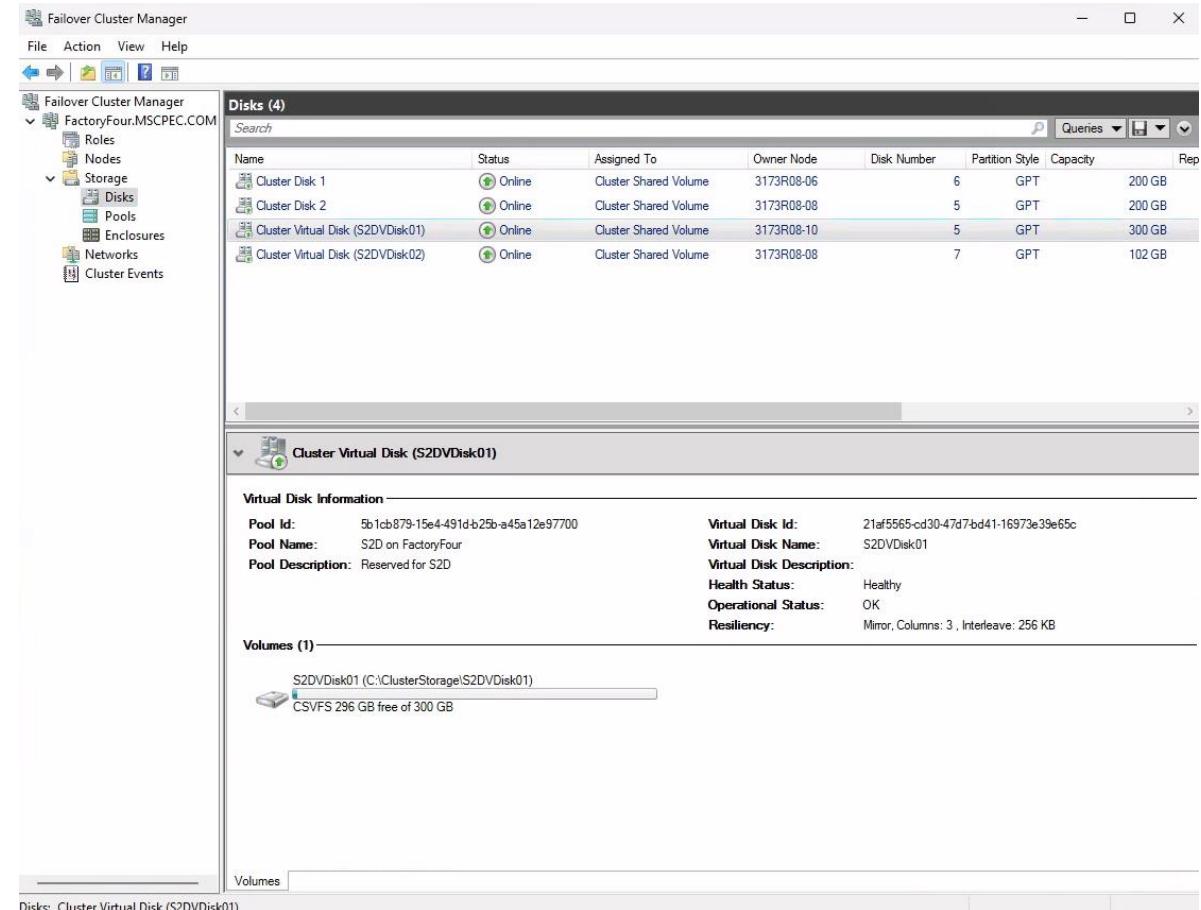
Members + Select members

1. Connect the nodes to Arc for Servers.
2. In the Azure Portal, assign the Storage Blob Contributor role to machines for the storage account used for the Cloud Witness.
3. Configure the Cloud Witness with the Managed Identity using PowerShell:

**Set-ClusterQuorum -CloudWitness -
AccountName <> -UseManagedIdentity**

S2D Storage + SAN Coexistence on WS 2025

1. SANs can be connected to Clusters running S2D.
2. S2D CSVs and SAN CSVs “Coexist” peacefully.
3. SAN disks are **COMPLETELY SEPARATE** from S2D disks.
4. SAN disks **MUST NOT** be used in the S2D storage pool.
5. Fibre Channel SANs, iSCSI SANs, and iSCSI Target will be supported.
6. SAN volumes should be formatted as NTFS before being added to CSV.
7. S2D volumes should be formatted as ReFS before being added to CSV.

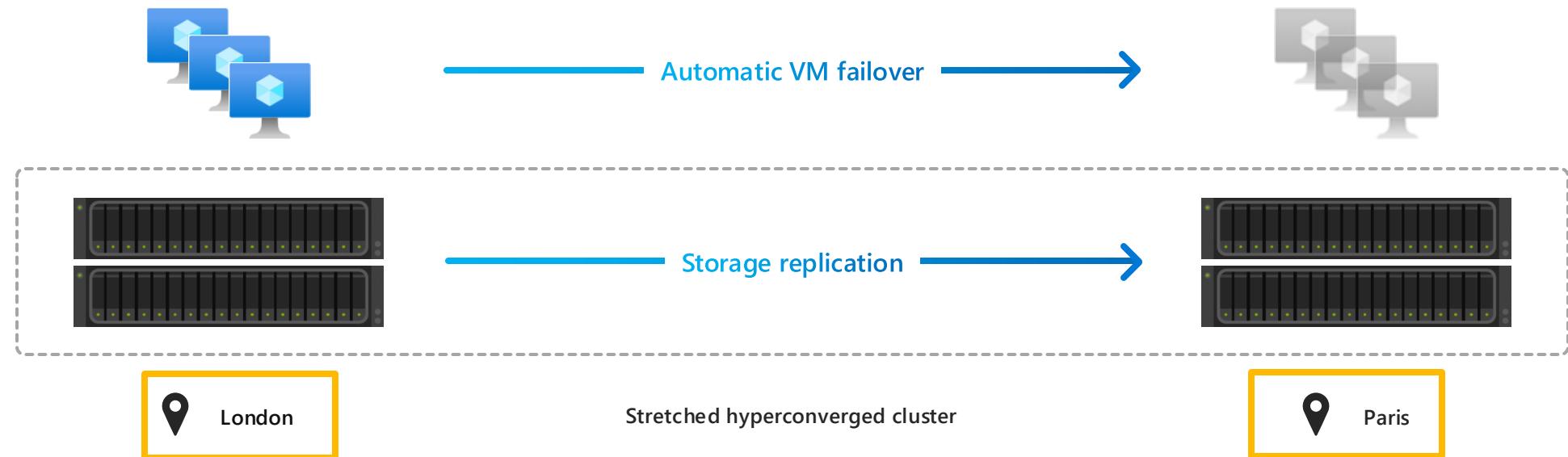


Stretched Cluster



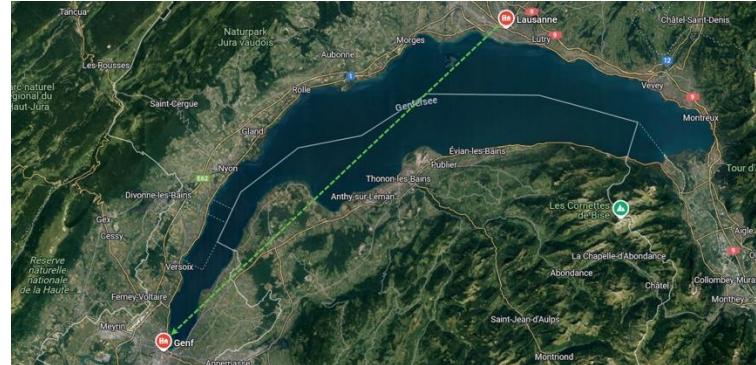
Stretched Cluster

- Was war ein Stretched Cluster für Microsoft bisher



Stretched Cluster

Metro Cluster



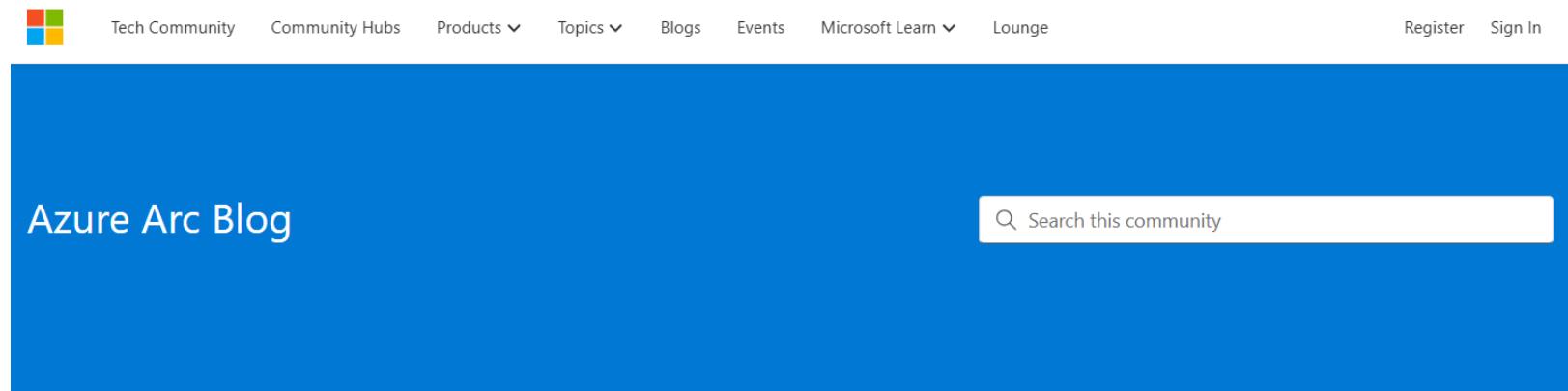
- Entfernung der Sites viele manchmal sogar hunderte Kilometer
- Netzwerk zwischen den Sites meist über WAN oder angemietete Verbindungen (dark Fiber)
- Netzwerkgeschwindigkeit zwischen 1 Gbit/s und 20 Gbit/s
- Meist L3 Verbindungen
- Wenig oder keine Redundanz bei Netzwerkwegen
- Verschieden IP-Bereiche für die Workload

Campus Cluster



- Sites auf dem gleichen Campus einige Meter entfernt (normal zwischen 50 und 800 Meter)
- Eigene Netzinfrastruktur, oft mit dedizierten Switchen
- High Speed Ethernet (=> 20 Gbit/s) Netzwerk zwischen Sites
- Meist L2 Verbindungen
- Viele redundante Netzwerkpfade
- Gleich IP-Bereiche für die Workload

Azure Local is no longer supporting S2D Stretch Cluster Topology...



The screenshot shows the Azure Arc Blog page. At the top, there is a navigation bar with links: Tech Community, Community Hubs, Products ▾, Topics ▾, Blogs, Events, Microsoft Learn ▾, and Lounge. To the right of the navigation bar are 'Register' and 'Sign In' buttons. Below the navigation bar, the page has a blue header with the text 'Azure Arc Blog' on the left and a search bar on the right containing the placeholder 'Search this community'. The main content area features a large blue banner with the title 'Evolving Stretch Clustering for Azure Local' in white text.

AZURE ARC BLOG 3 MIN READ

Evolving Stretch Clustering for Azure Local



BarbaraPei  MICROSOFT

Dec 05, 2024

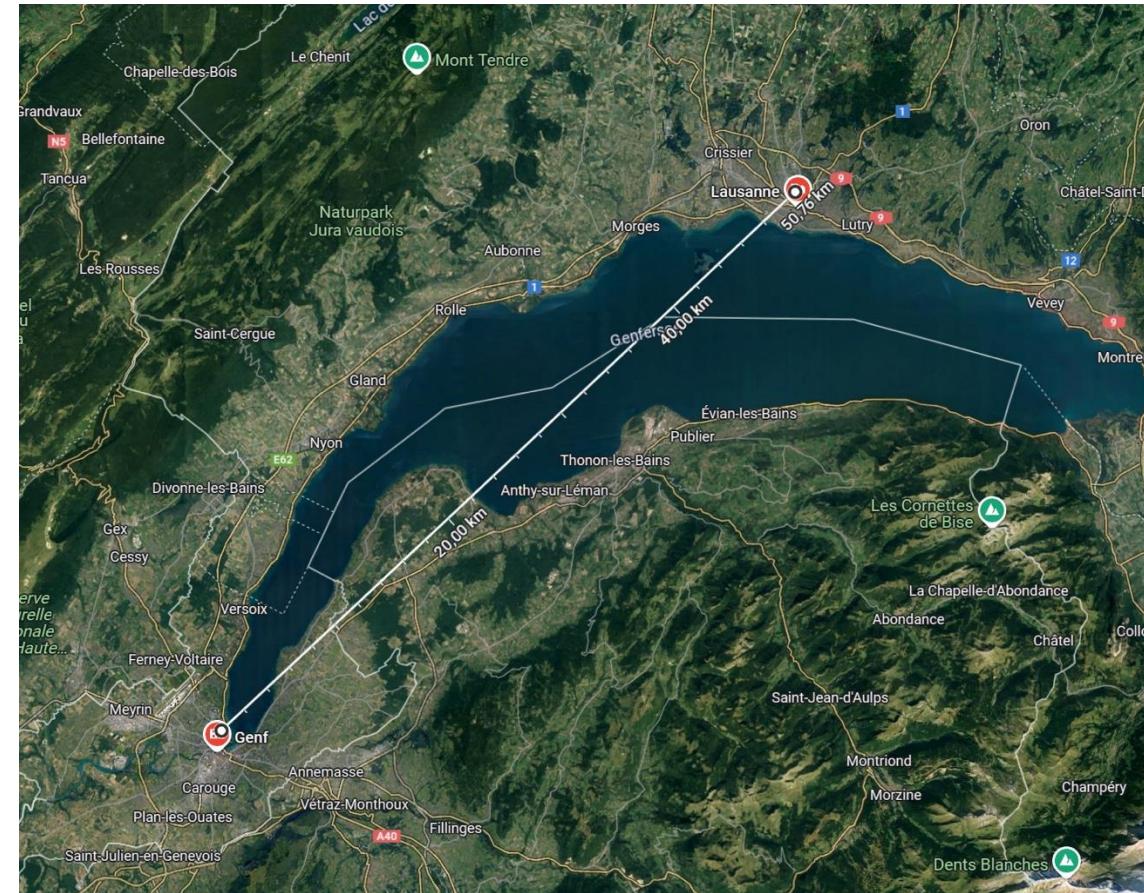
Stretched clusters in Azure Local, version 22H2 (formerly Azure Stack HCI, version 22H2) entail a specific technical implementation of storage replication that spans a cluster across two sites.

Azure Local, version 23H2 has evolved from a cloud-connected operating system to an Arc-enabled solution with Arc Resource Bridge, Arc VM, and AKS enabled by Azure Arc. Azure Local, version 23H2 expands the requirements for multi-site scenarios beyond the OS layer, while Stretched clusters do not encompass the entire solution stack.

Based on customer feedback, the new Azure Local release will replace the Stretched clusters defined in version 22H2 with new high availability and disaster recovery options.

Stretched Cluster

- Metro Stretched S2D Cluster
 - Stretched Support wie in Azure Stack HCI
 - Storage Replica als Redundanz Technologie
 - Synchronous oder Asynchronous
 - Verbesserte Performance durch neues Raw Log
 - Storage Spaces Direct
 - Herausforderndes Netzwerkdesign
 - Aller Traffic wird geroutet
 - Support für Cluster mit 4 bis 16 Knoten
 - Gleichmäßig verteilt auf 2 Sites (2-2, 3-3, 4-4, 5-5, ...)
 - Witness in dritter Site oder in Azure



Kommt in Windows Server 2025!

Stretched Cluster

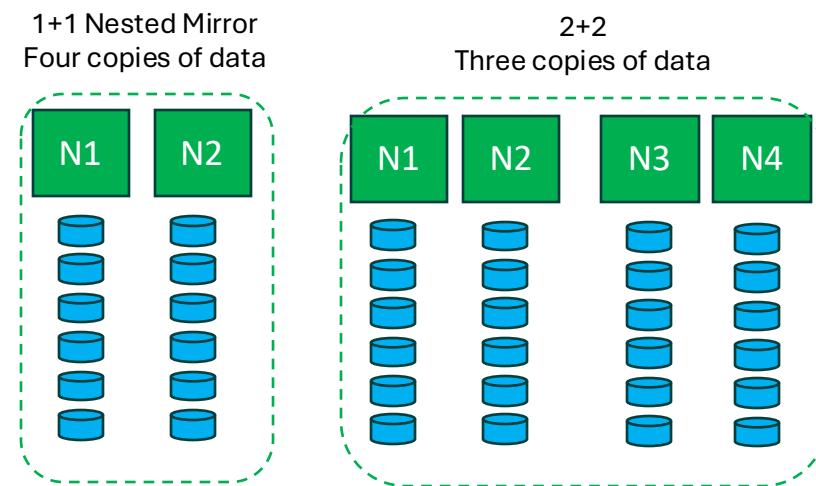
- S2D Campus Cluster
 - Nutzt Extend Verteilung als Redundanz Technologie
 - Immer Synchronous
 - Hervorragende Performance über den SBL
 - Storage Spaces Direct mit 4 Kopien
 - Einfacheres Netzwerkdesign
 - Storage Traffic meist im Gleichen Subnetz
 - Kein Routing notwendig aber möglich
 - Support für Cluster mit 2 und 4 Knoten (vermutlich auch 6 und 8 Knoten)
 - Gleichmäßig verteilt auf 2 Sites (1-1, 2-2, 3-3, 4-4)
 - Witness in drittem Raum oder Azure



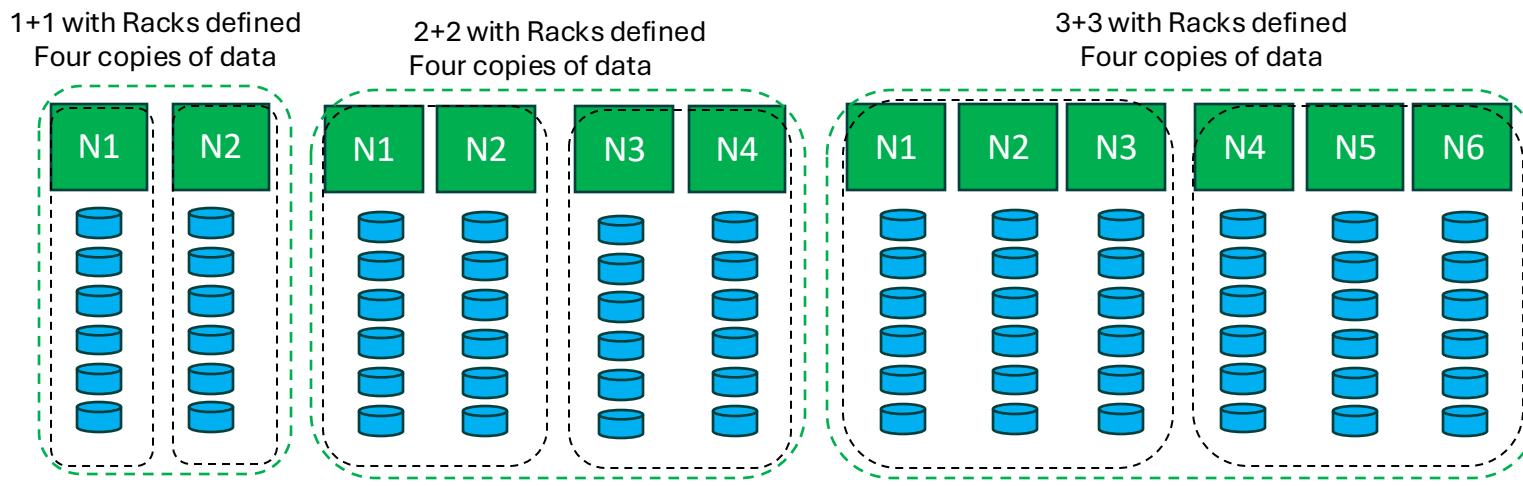
S2D Campus Cluster Support

Coming Soon

Phase 1: Existing S2D Resiliency



Phase 2: Forthcoming: S2D Template Tier for 4-way Mirror



- Existing resiliency is currently being validated
- Soft Anti-affinity rule is implemented (corresponds to VMware SHOULD rule): **New-ClusterAffinityRule, Set ClusterAffinityRule -SoftAntiAffinity**
- 4-way Mirror S2D Template Tier is being evaluated / estimated for Windows Server 2025

Cluster Improvements



Cluster OS Rolling Upgrade

- Upgrade your Failover Clusters without downtime!
- Windows Server 2022 → Windows Server 2025
 - The Windows Server 2025 Feature Update is downloaded from Windows Update (FU via WU)
 - CAUs RollingUpgrade plugin will work when we publish WS 2025 Feature Update

```
Invoke-CauRun -EnableFirewallRules -Force -CauPluginName Microsoft.RollingUpgradePlugin
```

There are many small improvements in Windows Server 2025 Failover Clustering

- **Soft Anti-affinity** for antiaffinity rules – support for Campus Cluster, similar to VMware SHOULD rules
- **Repair-ClusterNameAccount** – cmdlet to automate functionality that was previously only available through FCM UX
- **CAU**
 - Invoke-CauRun from any clusternode
 - Custom Reboot option – CAU plugin determines if a reboot is required
- **Node Fallback Status** - parity with node Drain Status
- **Add-ClusterExcludedAdapter**, **Remove-ClusterExcludedAdapter**, **Get-ClusterExcludedAdapter**, **Set-ClusterExcludedAdapter** to simplify management of excluded adapters, which is becoming more important as more customers have BMC/IDRAC



Q&A

Closing Subtext



Danke an unsere Sponsoren

PLATINUM



GOLD



INFOTECH
[IT & Communication]





Closing

Closing Subtext



Azure Virtual Desktop auf Azure Local

**Carsten Rachfaß – Microsoft Azure und Cloud and
Datacenter Management MVP**





AVD Überblick

Closing Subtext



Vorteile von AVD auf Azure Local

Management Plain komplett in der Azure Cloud

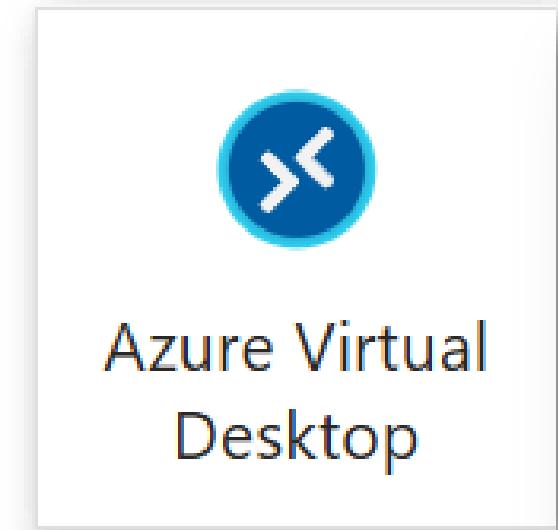
- Microsoft kümmert sich um Bereitstellung, Skalierung und Updates

Bereitstellung von Azure Marketplace Images

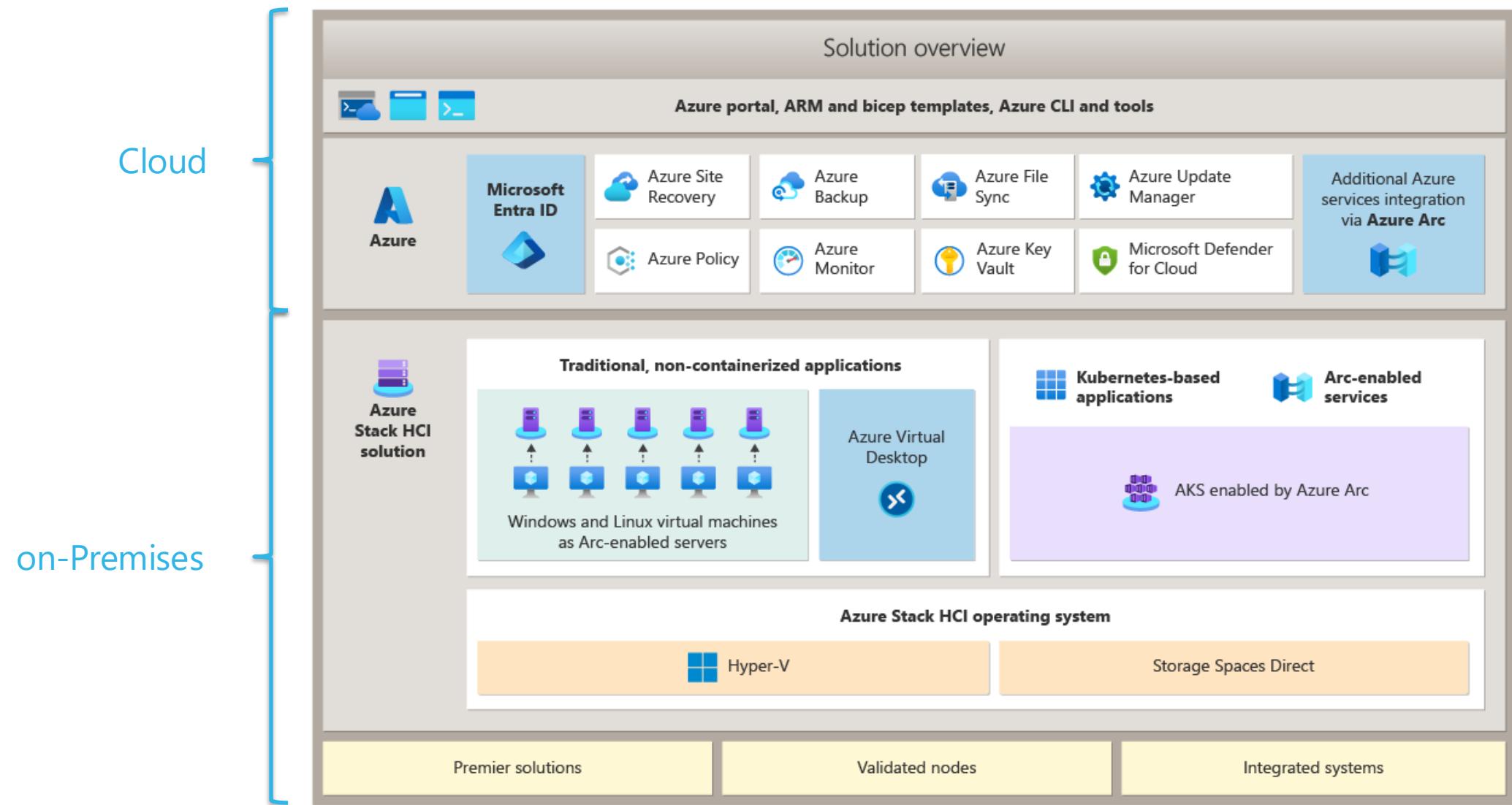
- Vorbereitete und optimierte Images

Windows 10/11 Multisession Host

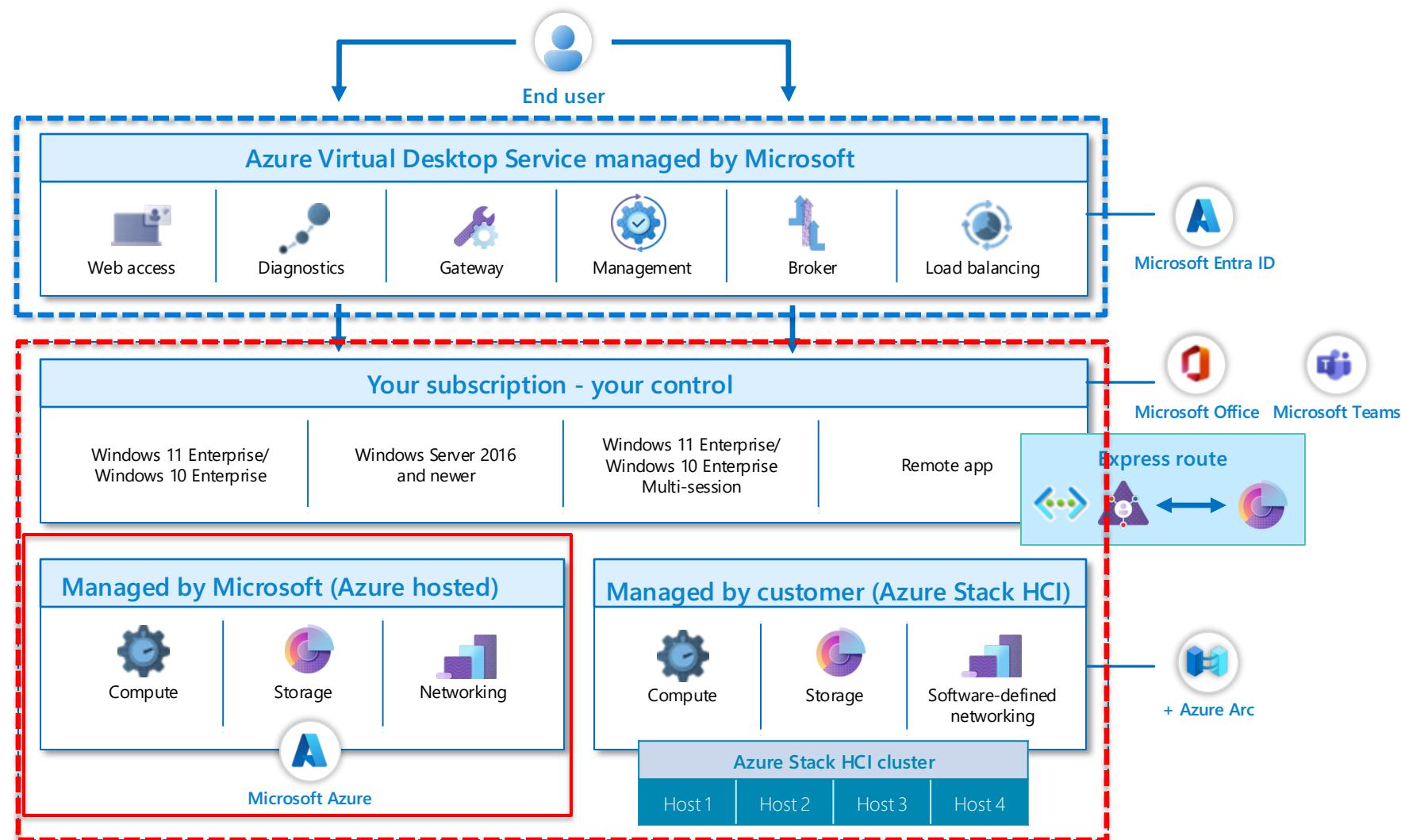
- Betrieb nur in Azure und Azure Local möglich



Azure Virtual Desktop auf Azure Local



AVD Architektur



Provide your employees with a secure, remote desktop experience.

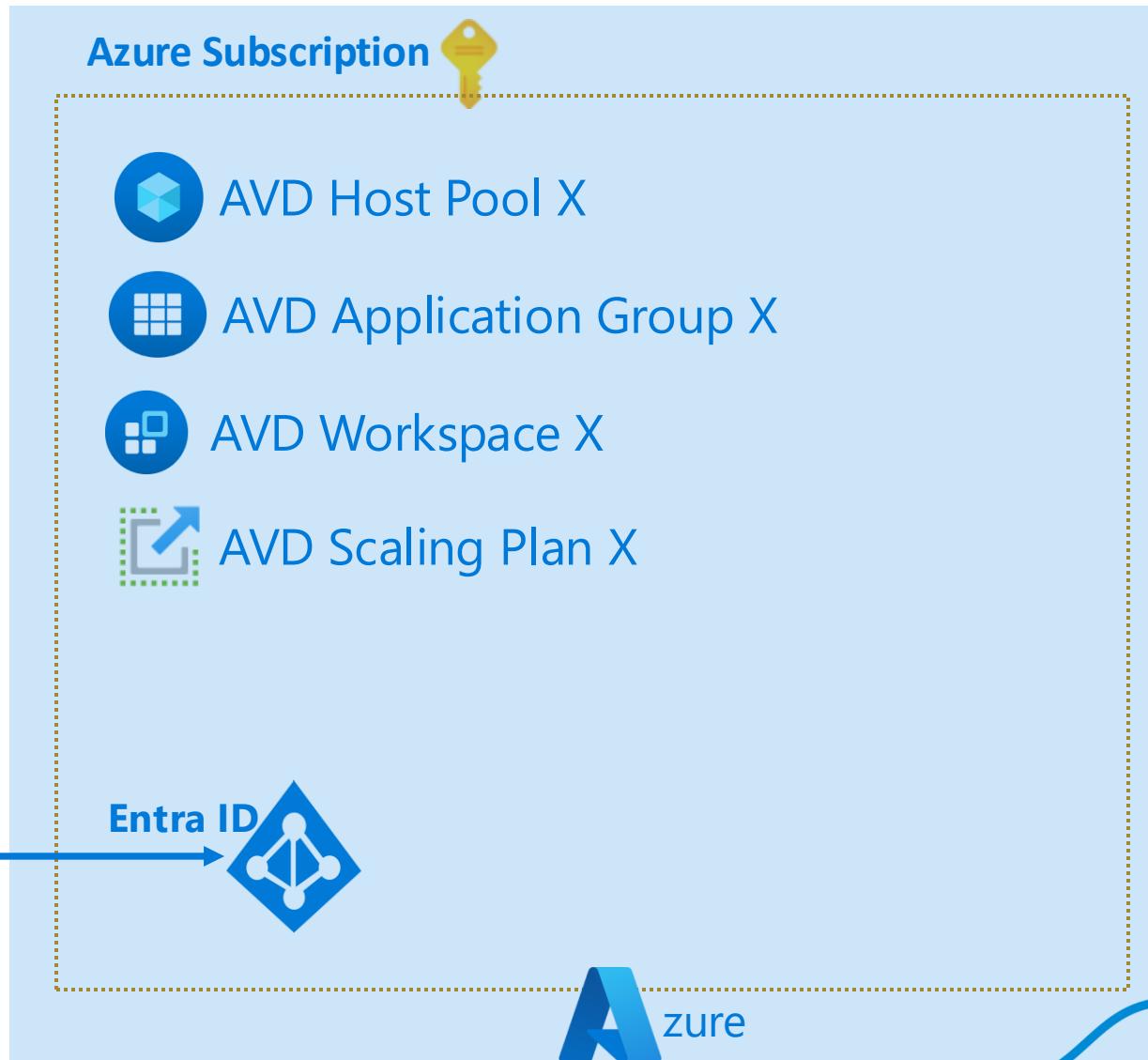
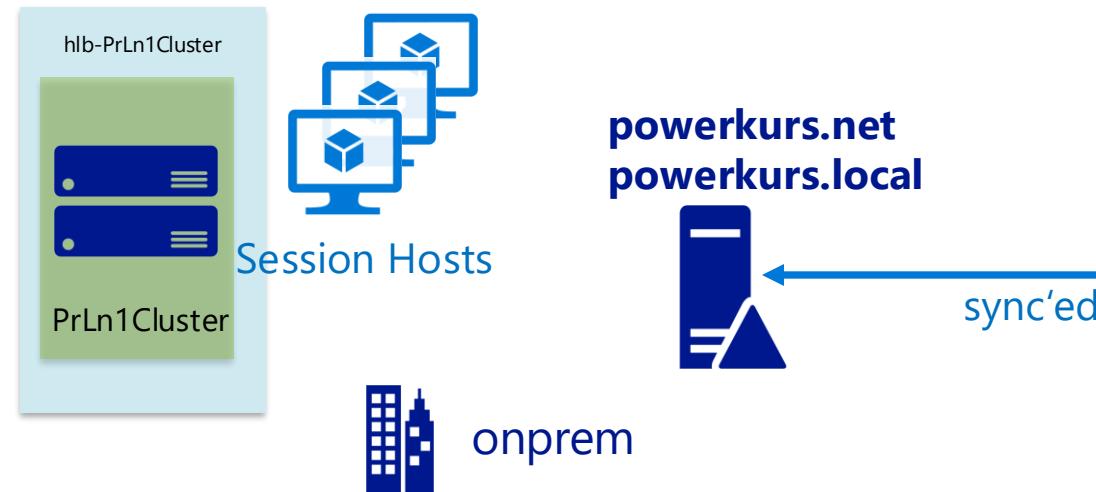
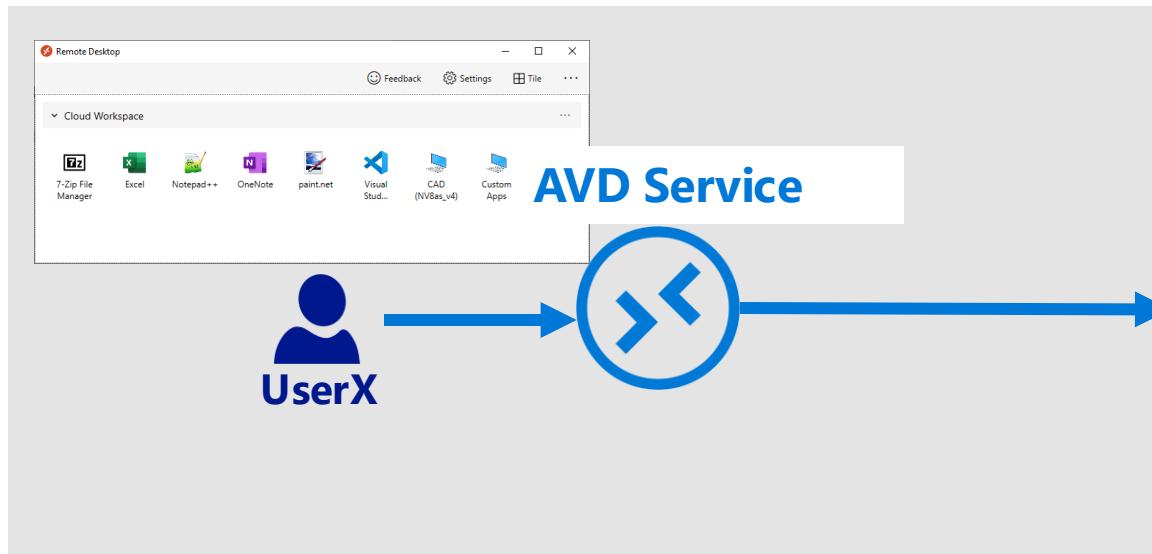


Connect from virtually any device of your choice.



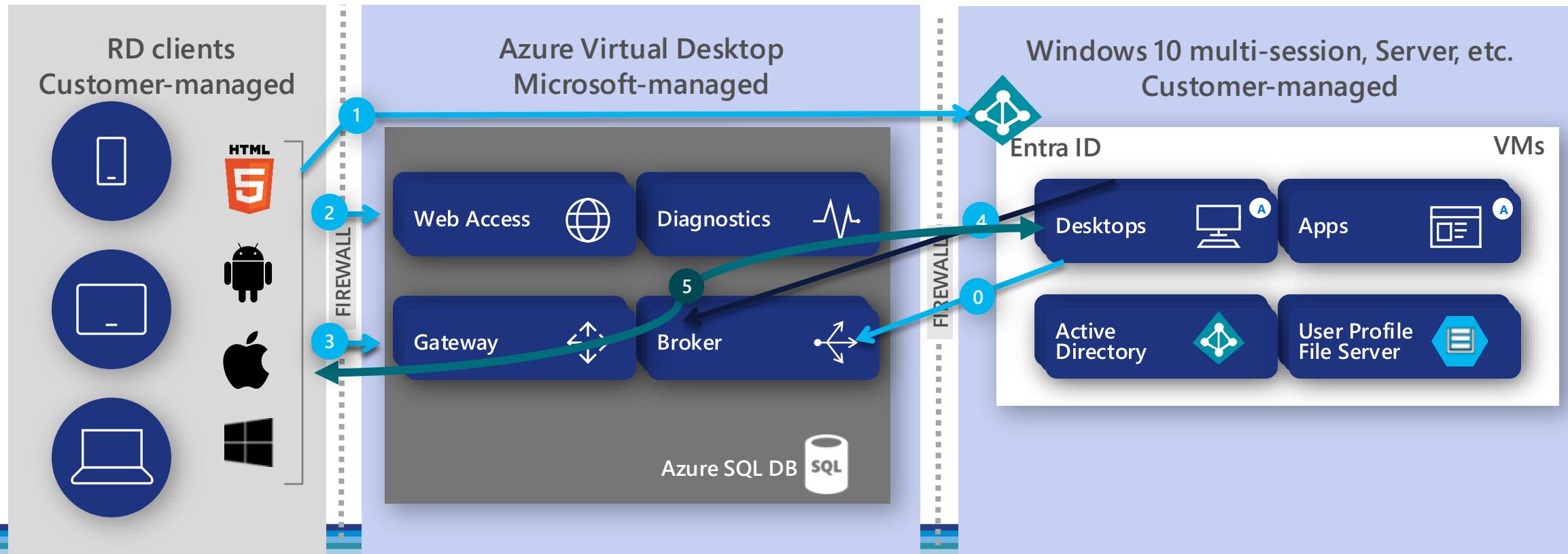
Focus on the right policies and controls rather than managing infrastructure.

Architektur

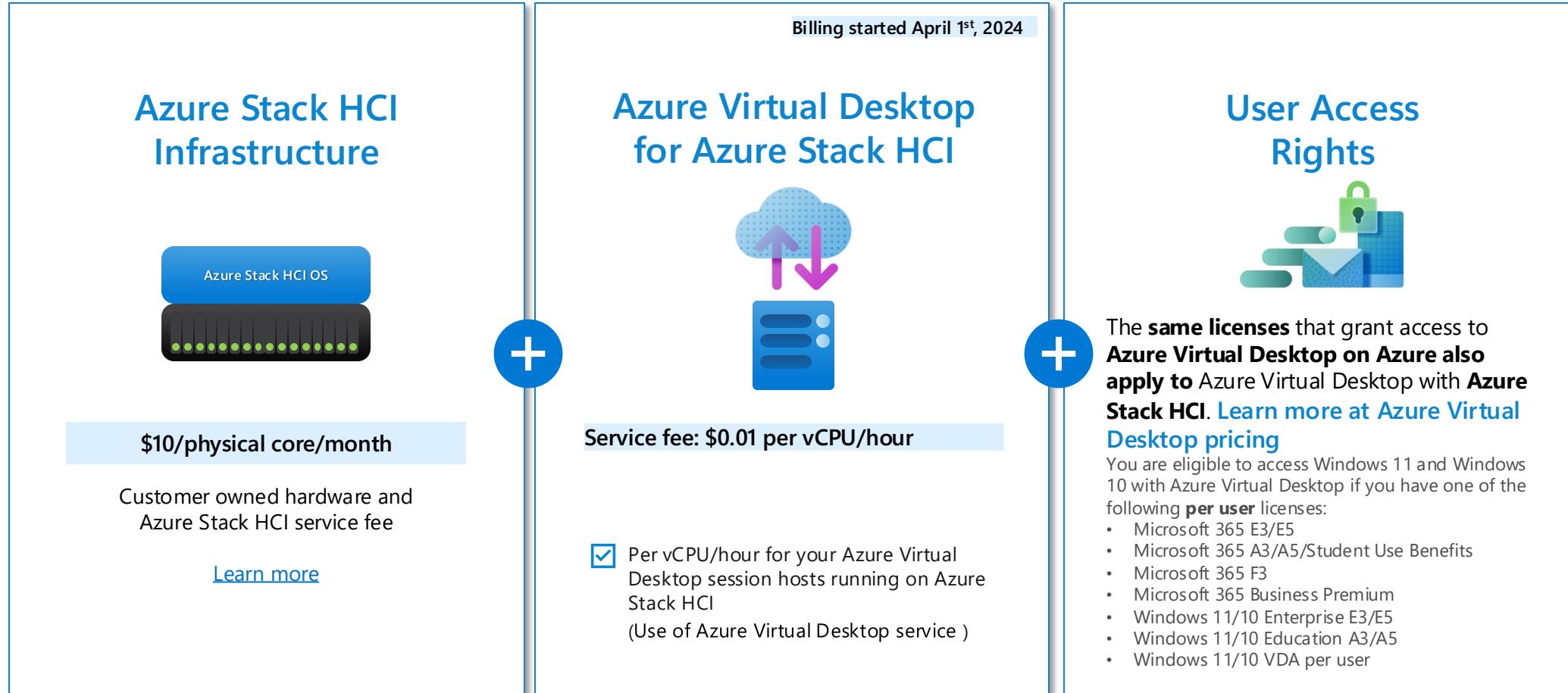


Benutzer Verbindungs Flow

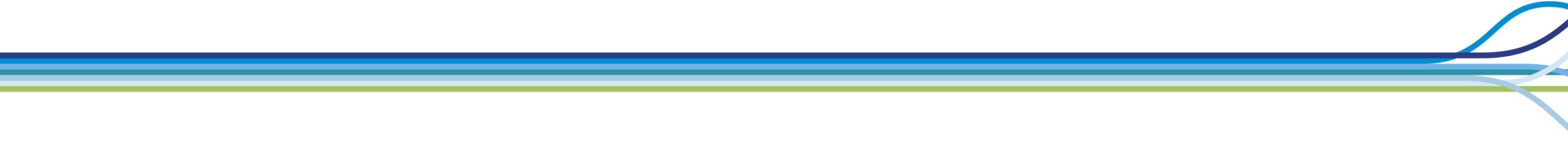
0. Wenn Session Host startet meldet er sich am Broker an
1. Benutzer startet RD Client und meldet sich an Entra ID an, welches einen Token zurück liefert
2. RD Client präsentiert Token dem Web Access, vorauf der Broker die DB nach Ressourcen für den Benutzer durchsucht
3. Benutzer wählt Ressource aus und RD Client verbindet sich mit dem Gateway
4. Broker orchestriert Verbindung from Host Agent zum Gateway
5. RDP Traffic fließt zwischen RD Client und Session Host VM über WebSocket Verbindung 3 und 4



AVD auf Azure Local Preise und Lizenzierung



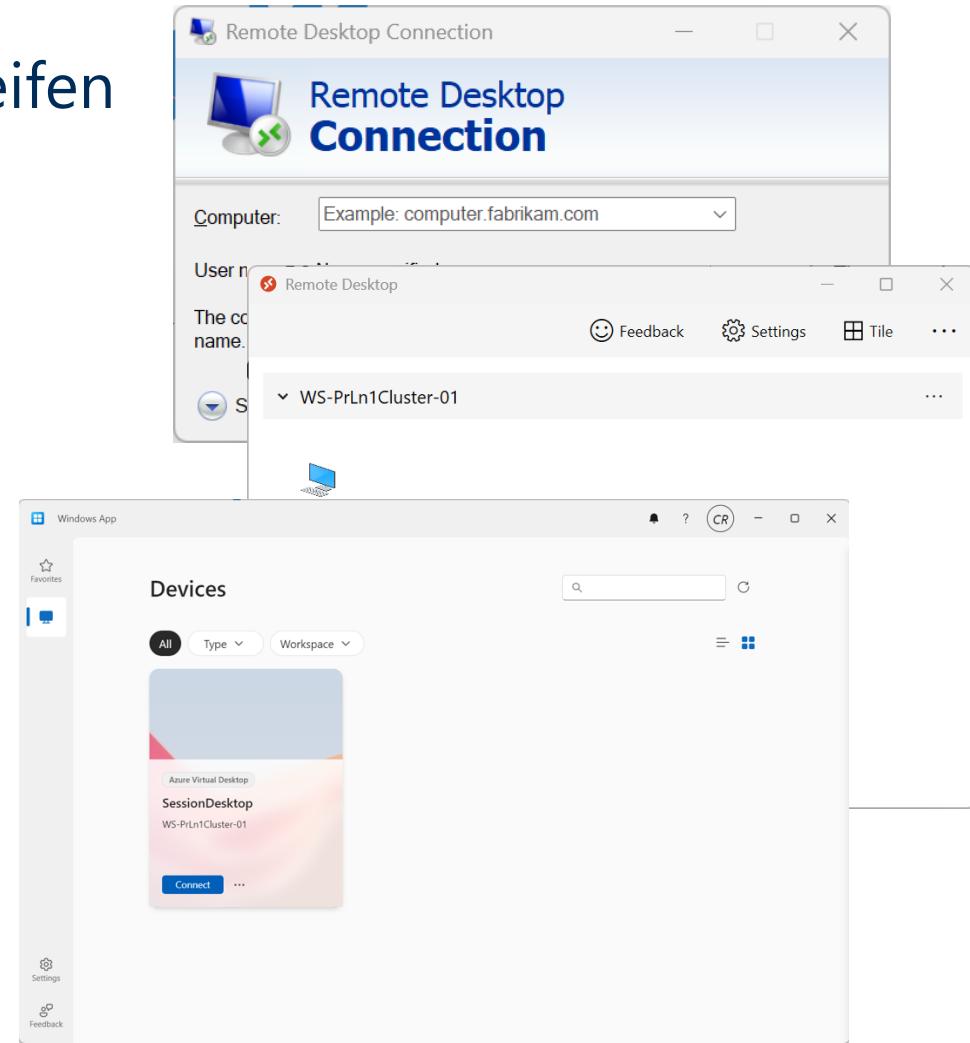
AVD Benutzer Zugriff



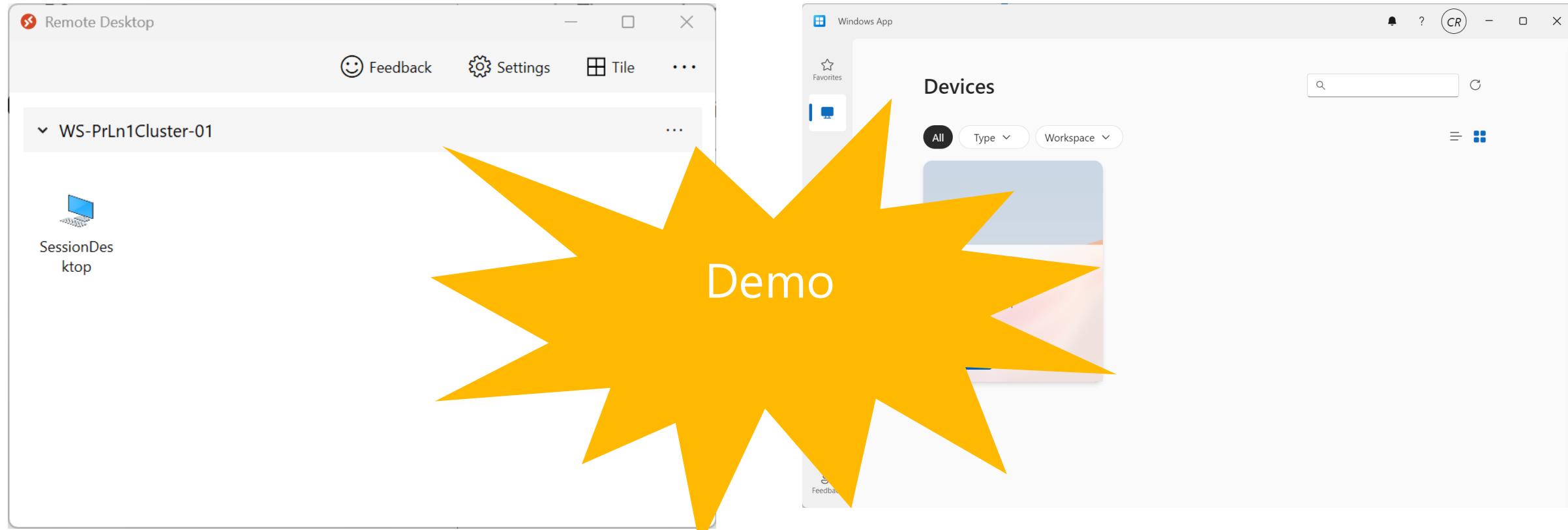
Benutzer Zugriff

Wie können Benutzer auf die AVD Hosts zugreifen

- **MSTSC (Microsoft Terminal Server Client)**
 - Nicht möglich da die Anmeldung mit einer Entra ID erfolgt
- **Remote Desktop client (support Ende Mai 25)**
 - Windows
 - Web browser
 - macOS
 - iOS/iPadOS
 - Android/Chrome OS
- **Windows App**
 - Windows
 - macOS
 - iOS/iPadOS
 - Android/Chrome OS (preview)
 - Web browsers
 - Meta Quest VR headset (preview)



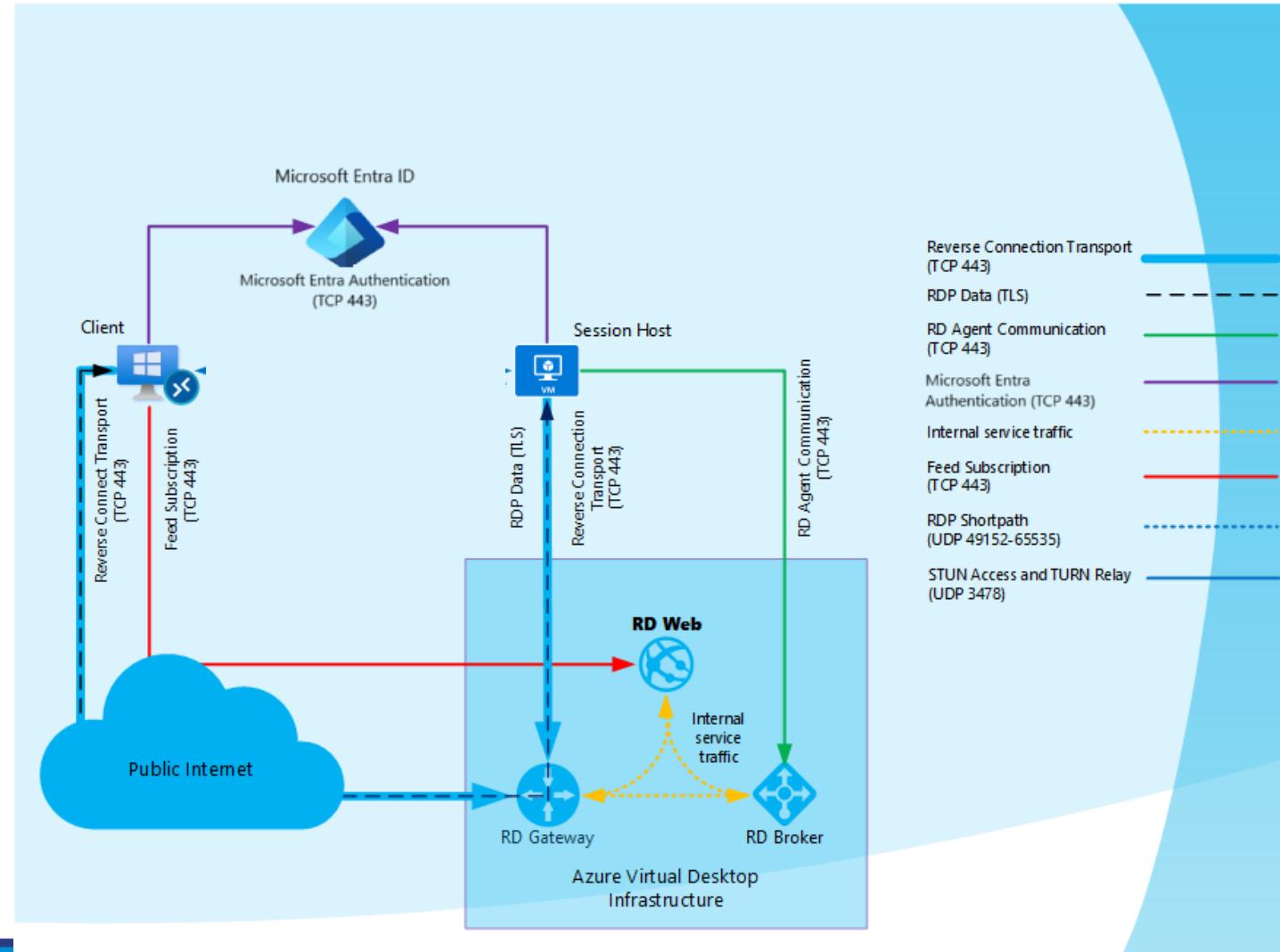
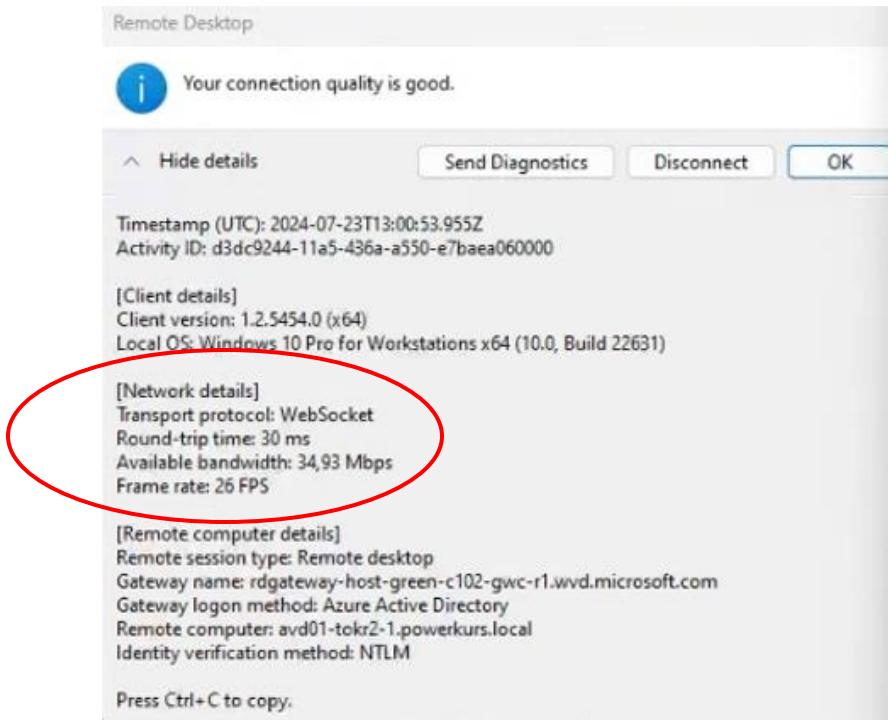
Benutzer Zugriff



UPD Shortpath

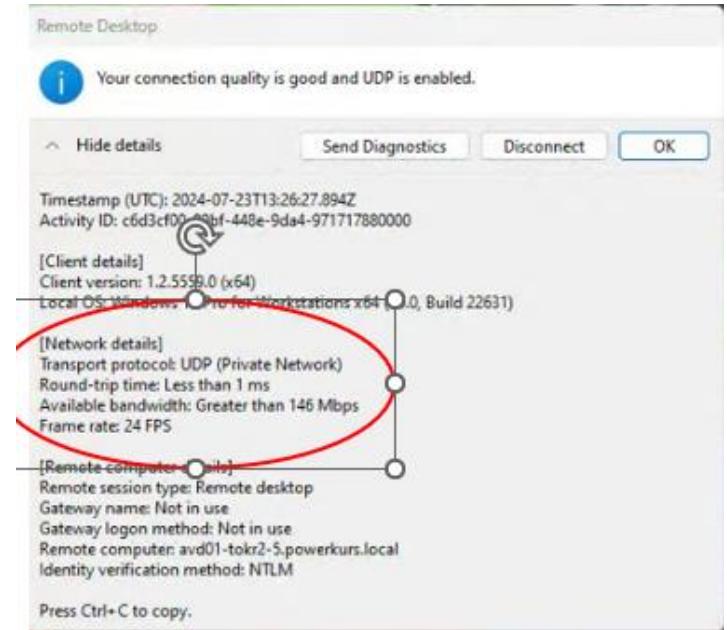
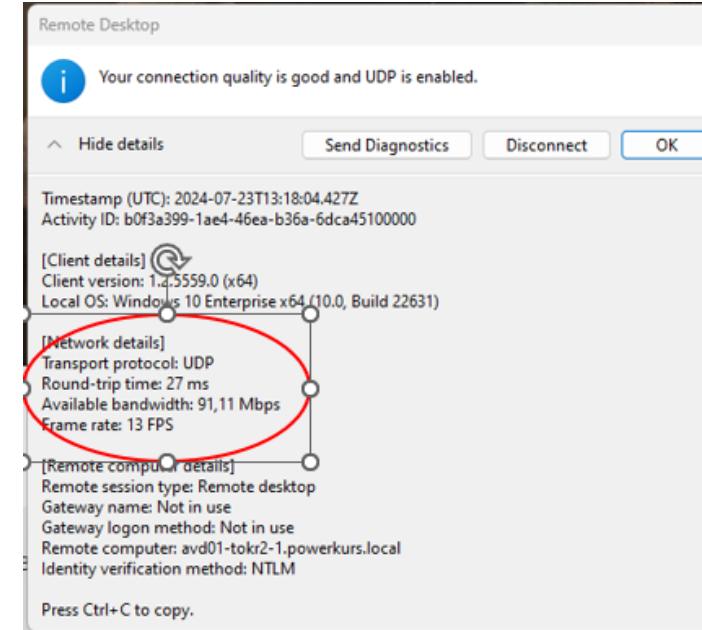
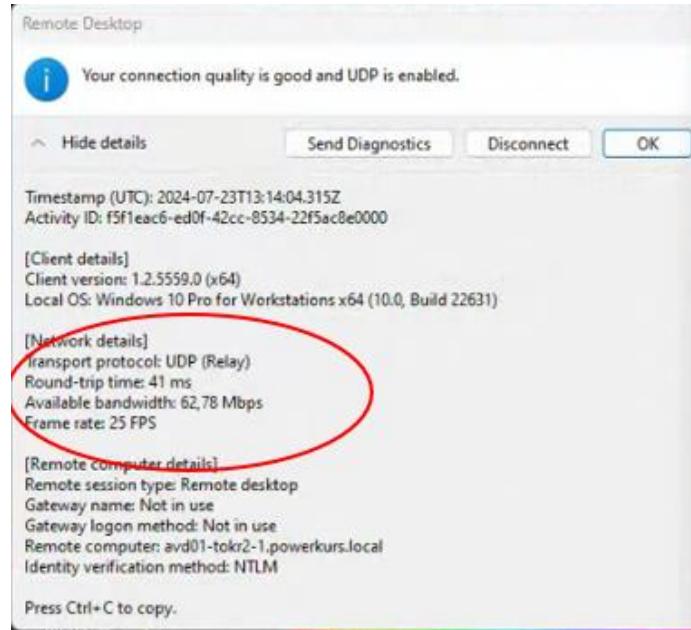


Azure Virtual Desktop Netzwerk Zugriff via https





RDP Shortpath Varianten

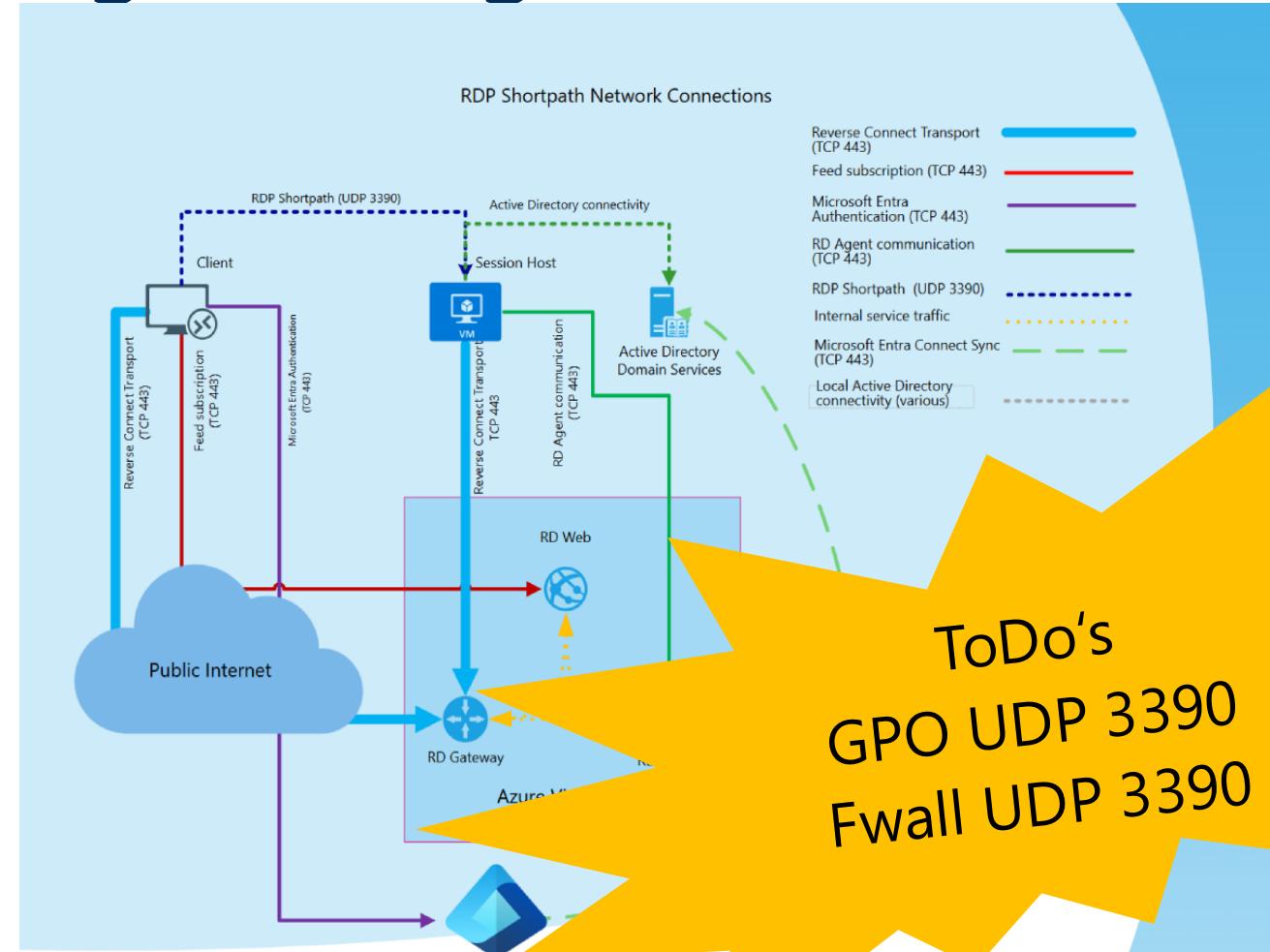
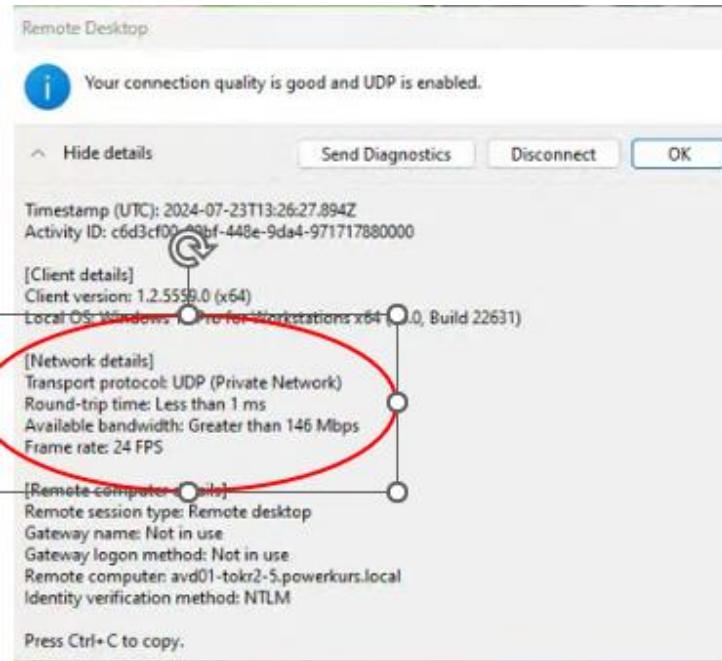


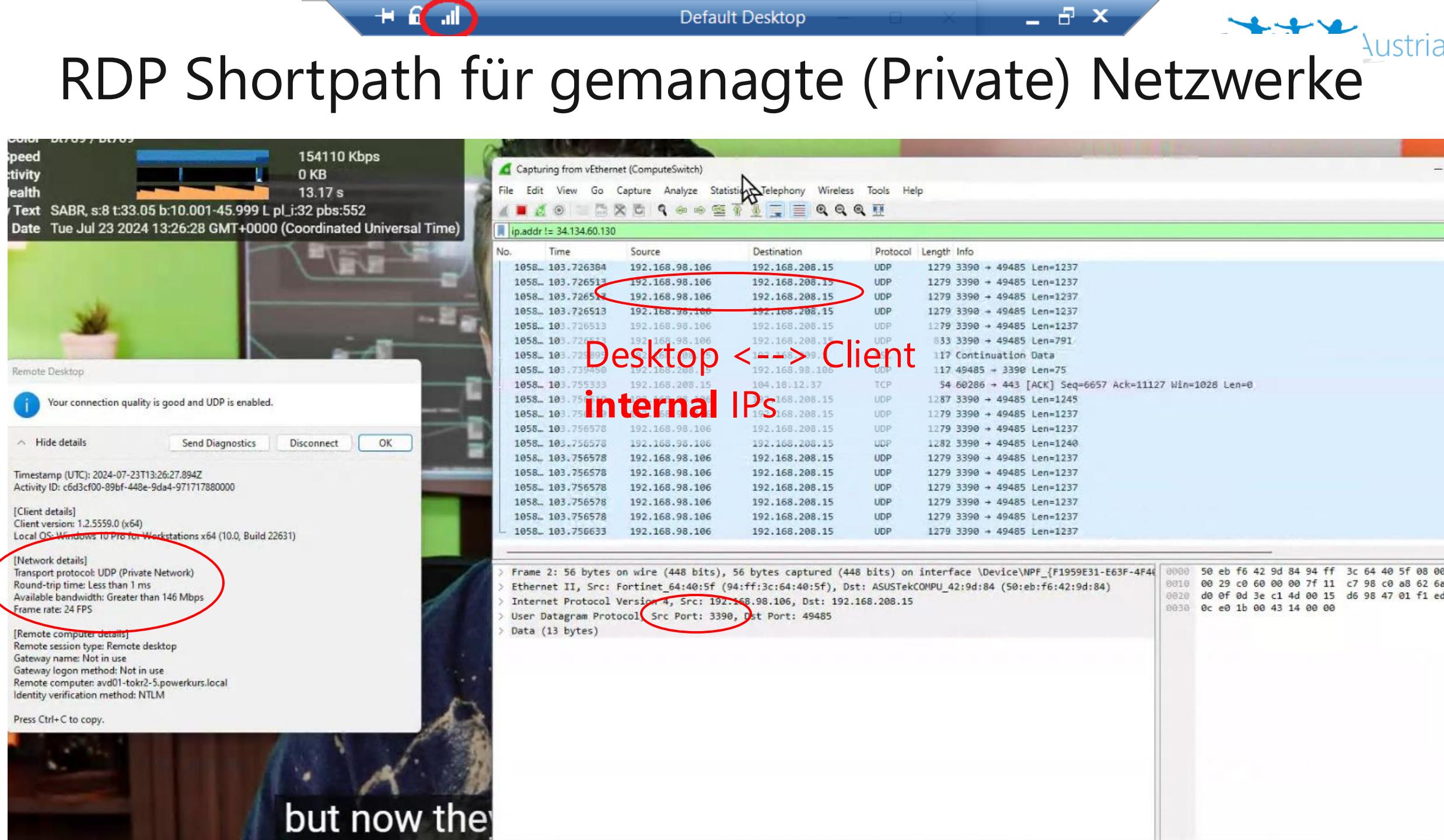
RDP Shortpath für Public
Netzwerke via TURN

RDP Shortpath für
Private & Public Netzwerke mit
ICE/STUN

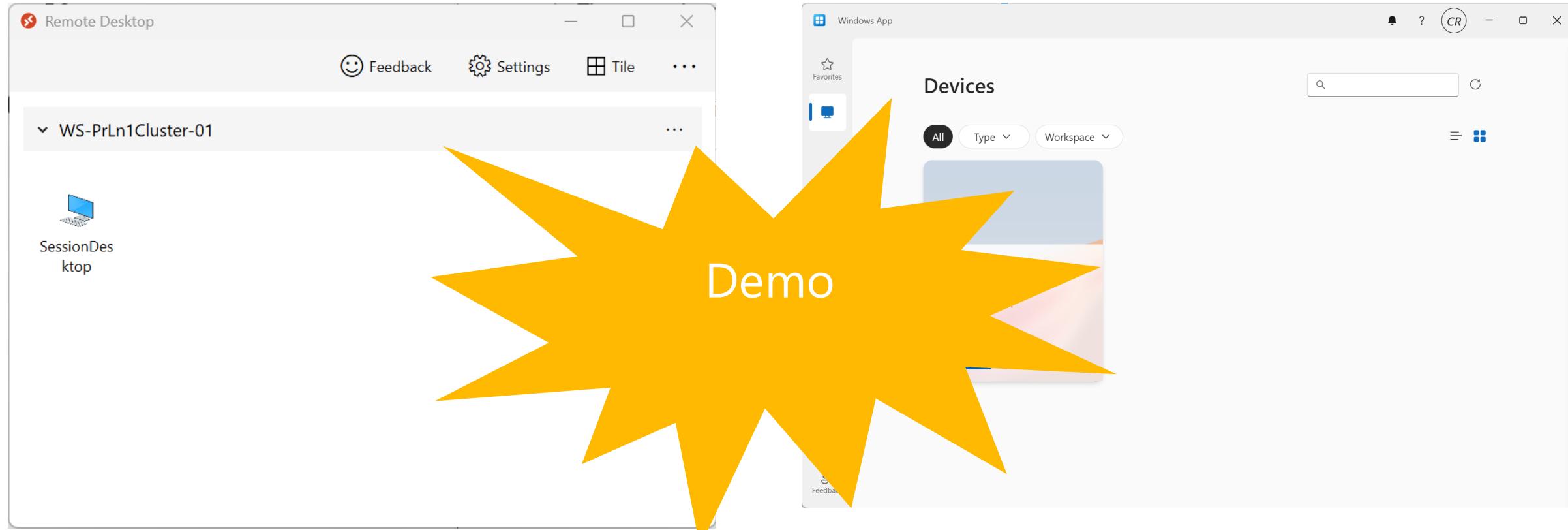
RDP Shortpath für gemanagte
(Private) Netzwerke

RDP Shortpath für gemanagte Netzwerke





Demo Short Path



Kosten Optimierung

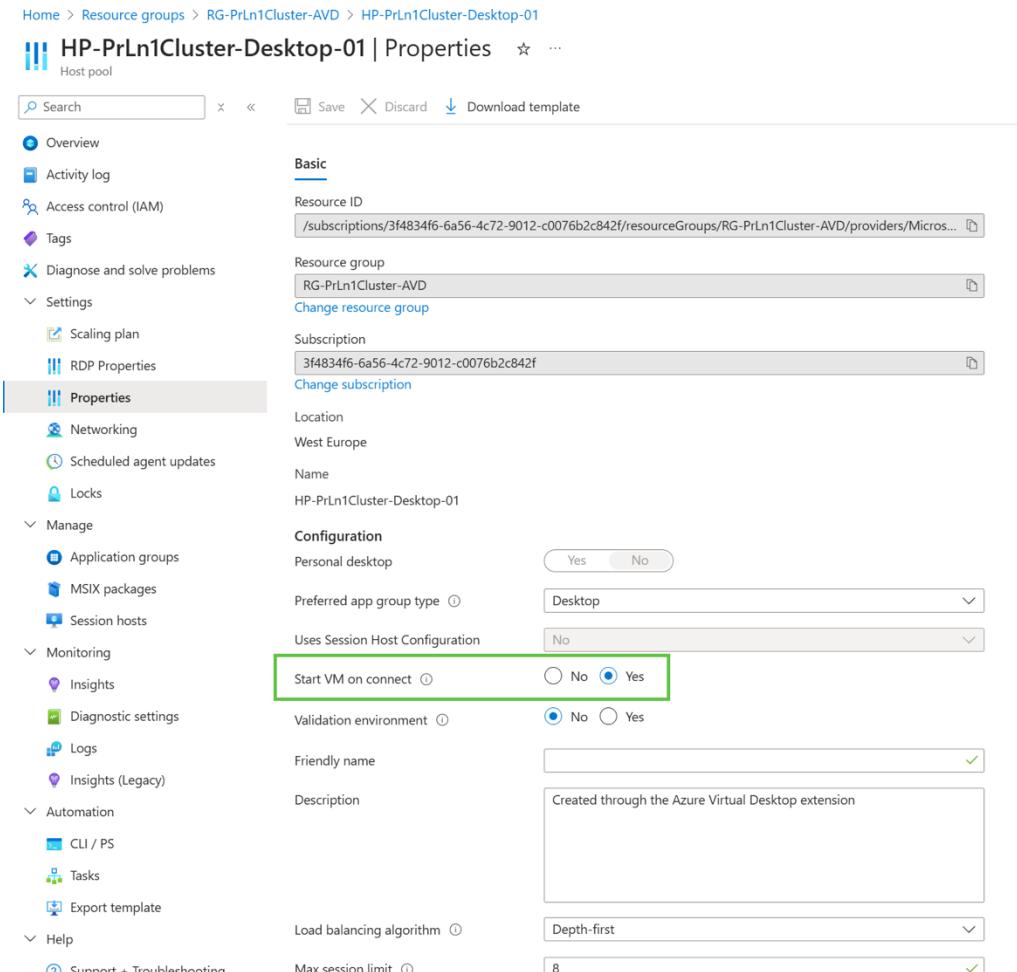


Start VM on Connect

„Start VM on Connect“ ermöglicht das:

- **Personal Desktops** nur laufen, wenn Benutzer damit arbeitet
- **Pooled Desktops** außerhalb der Arbeitszeiten heruntergefahren werden können

Start VM on Connect kombiniert mit einem **Scaling Plan** ermöglicht es, vermeidbare Kosten zu sparen



Scaling Plan

„Scaling Plan“ ermöglicht das:

- **Session Hosts** flexibel nach Benutzer aufkommen bereitgestellt werden
- **Scale Up** wenn mehr Session Hosts benötigt werden
- **Scale Down** wenn Benutzer sich abmelden oder konsolidiert werden können

Home > Azure Virtual Desktop | Scaling plans >
Create a scaling plan ...

[Basics](#) [Schedules](#) [Host pool assignments](#) [Tags](#) [Review + create](#)

Scaling plan enables you to apply schedules and preset conditions under which the autoscaling should occur for a host pool. [Learn more](#)

Project details

Subscription * [Select a subscription](#) Microsoft Azure Sponsorship

Resource group * [Select a resource group](#) Create new Select a resource group

Scaling plan name * [Scaling plan name](#)

Location * [Location](#) West Europe

Friendly name [Friendly name](#)

Description

Time zone * [Time zone](#) (UTC+01:00) Amsterdam, Berlin, Bern, Rome, Stockholm, Vienna

Host pool type [Host pool type](#) Pooled

Exclusion tag [Exclusion tag](#)

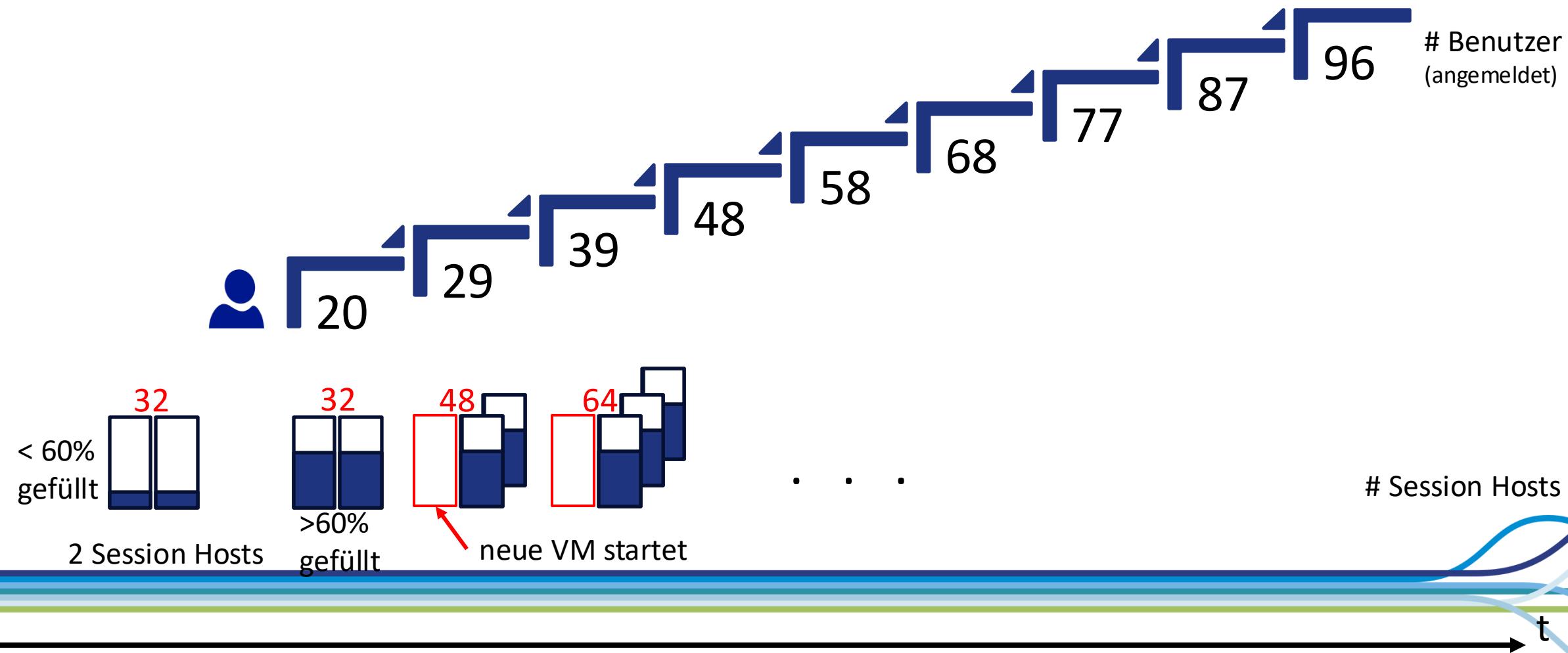
Scaling method *

Power management autoscaling
VMs will only be turned on or off to adjust available capacity. This is the only option available if your host pools are not using session host configuration. [Learn more](#)

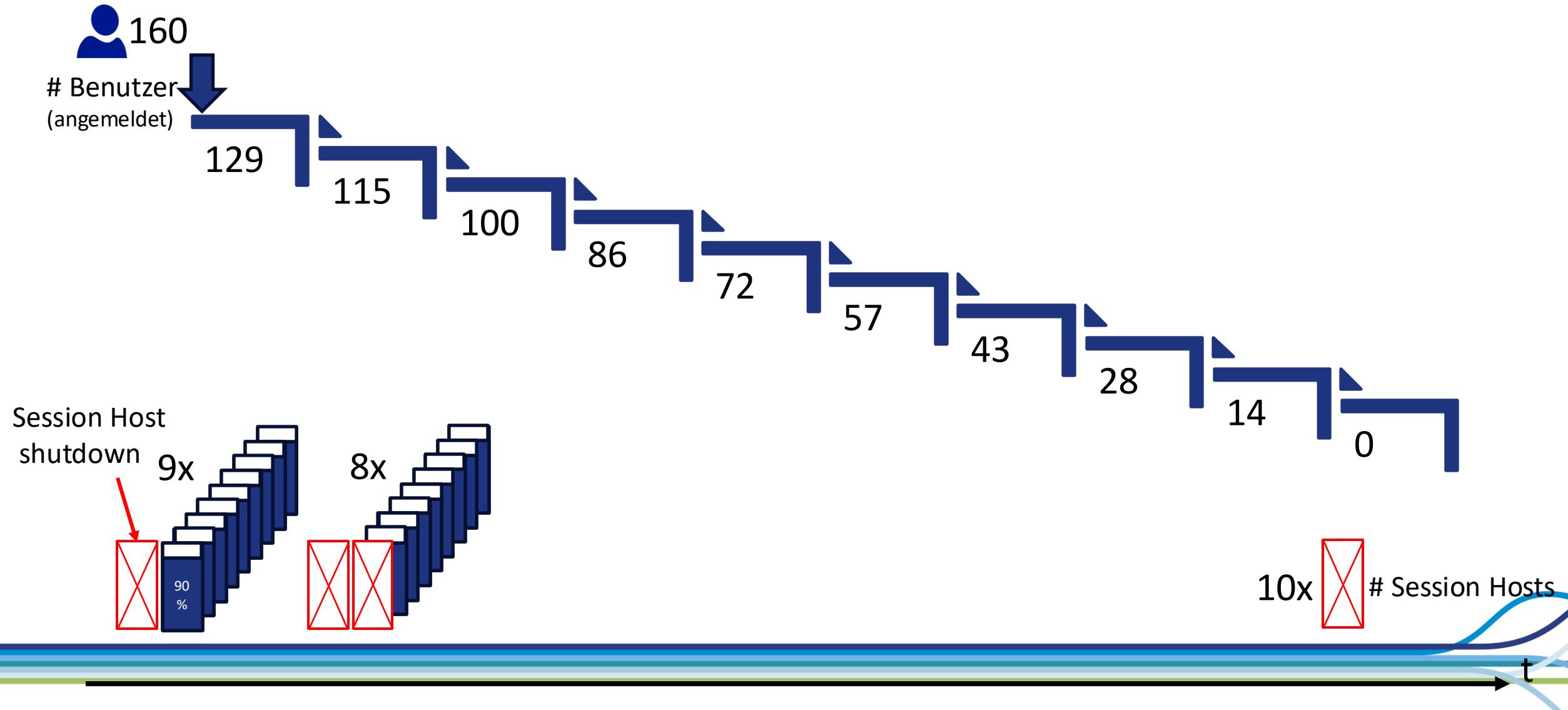
Dynamic autoscaling (preview)
Available capacity is managed by turning on/off existing machines and/or creating/deleting VMs. [Learn more](#)

[Review + create](#) [< Previous](#) [Next: Schedules >](#)

Ramp-Up Phase



Ramp-Down Phase



Video Scaling Pläne



Work A HP-PrLn1Cluster-Desktop-01 - Microsoft Azure

https://portal.azure.com/#@crachfahlrachfahl.onmicrosoft.com/resource/subscriptions/3f4834f6-6a56-4c72-9012-c0076b2c842f/resourceGroups/RG-PrLn1Cluster-AVD/providers/Microsoft.DesktopVirtualization/hostpools/HP-PrLn1Cluster-Desktop-01

Microsoft Azure Search resources, services, and docs (G+) Copilot powerkurs@crachfahlrachfahl.onmicrosoft.com RACHFAHL IT-SOLUTIONS GMBH

Home > HP-PrLn1Cluster-Desktop-01 Host pool

Search Registration key Refresh Delete Start Restart Stop

Overview Activity log Access control (IAM) Tags Diagnose and solve problems

Settings Scaling plan RDP Properties Properties Networking Scheduled agent updates Locks

Manage Application groups MSIX packages Session hosts

Monitoring Insights Diagnostic settings Logs Insights (Legacy)

Essentials

Resource group (move) : RG-PrLn1Cluster-AVD
Location : West Europe
Subscription (move) : Microsoft Azure Sponsorship
Subscription ID : 3f4834f6-6a56-4c72-9012-c0076b2c842f
Tags (edit) : Add tags

Host pool type : Pooled
Assignment type : ---
Uses Session Host Config... : No
OS disk type : -

JSON View

Virtual machines

Total machines 10 Can connect 2 Can't connect 8

Active sessions 0 Disconnected sessions 0 Pending sessions 0 Total sessions 0

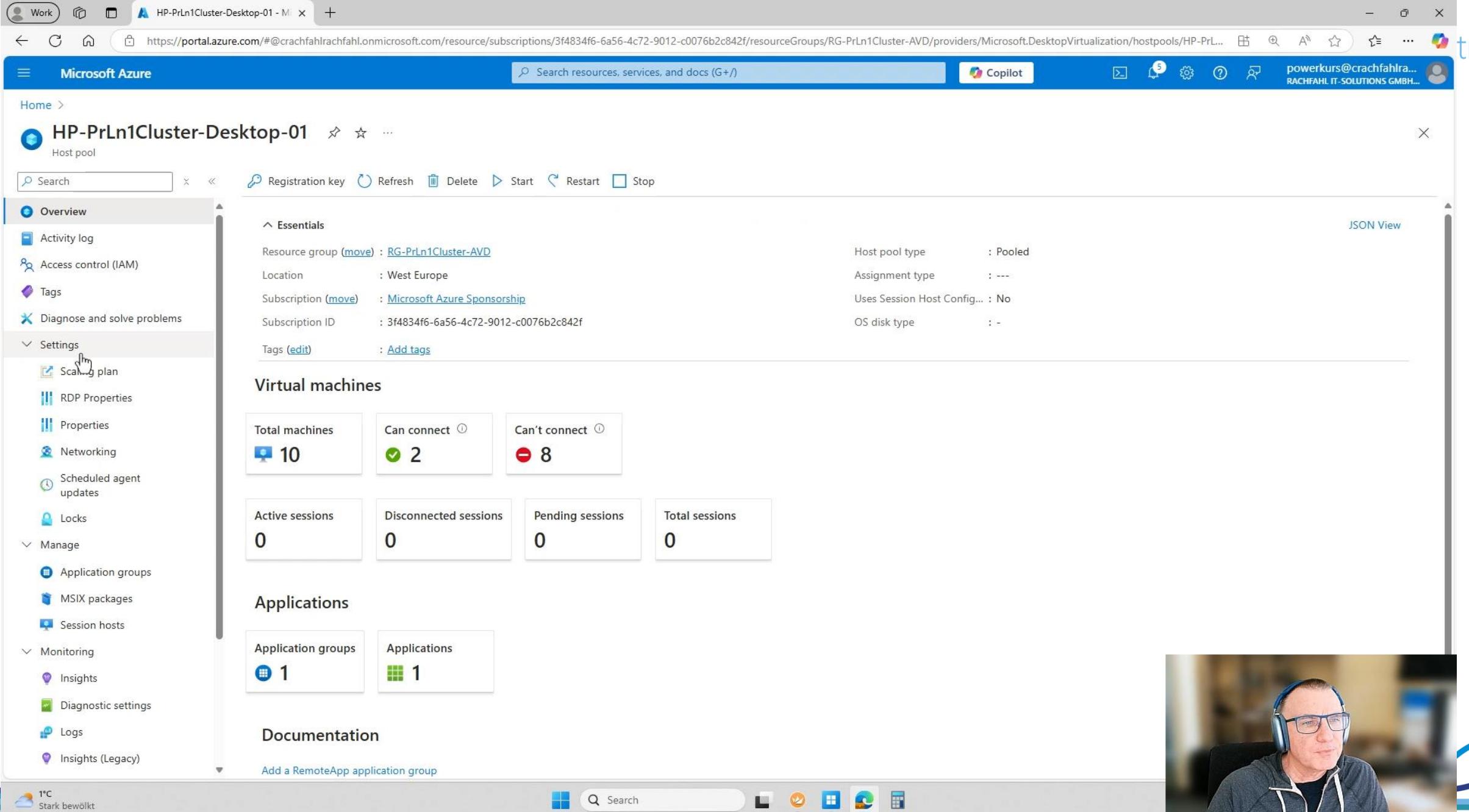
Applications

Application groups 1 Applications 1

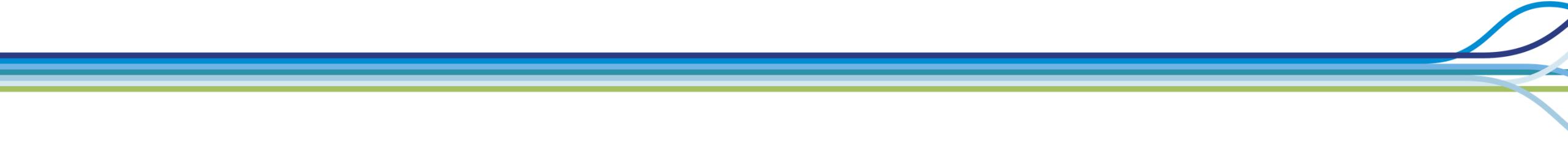
Documentation

Add a RemoteApp application group

1°C Stark bewölkt



GPU Unterstützung



GPU-Unterstützung in VMs

Viele Anwendung benötigen GPUs

- grafikintensive Anwendungen
- Multisession Hosts
- KI-Anwendungen

Supportete Technologien

- Discrete Device Assignment (DDA)
- GPU Partitioning (GPU-P)

Attaching GPUs on Azure Local

You can attach your GPUs in one of two ways for Azure Local:

- **Discrete Device Assignment (DDA)** - allows you to dedicate a physical GPU to your workload. In a DDA deployment, virtualized workloads run on the native driver and typically have full access to the GPU's functionality. DDA offers the highest level of app compatibility and potential performance.
- **GPU Partitioning (GPU-P)** - allows you to share a GPU with multiple workloads by splitting the GPU into dedicated fractional partitions.

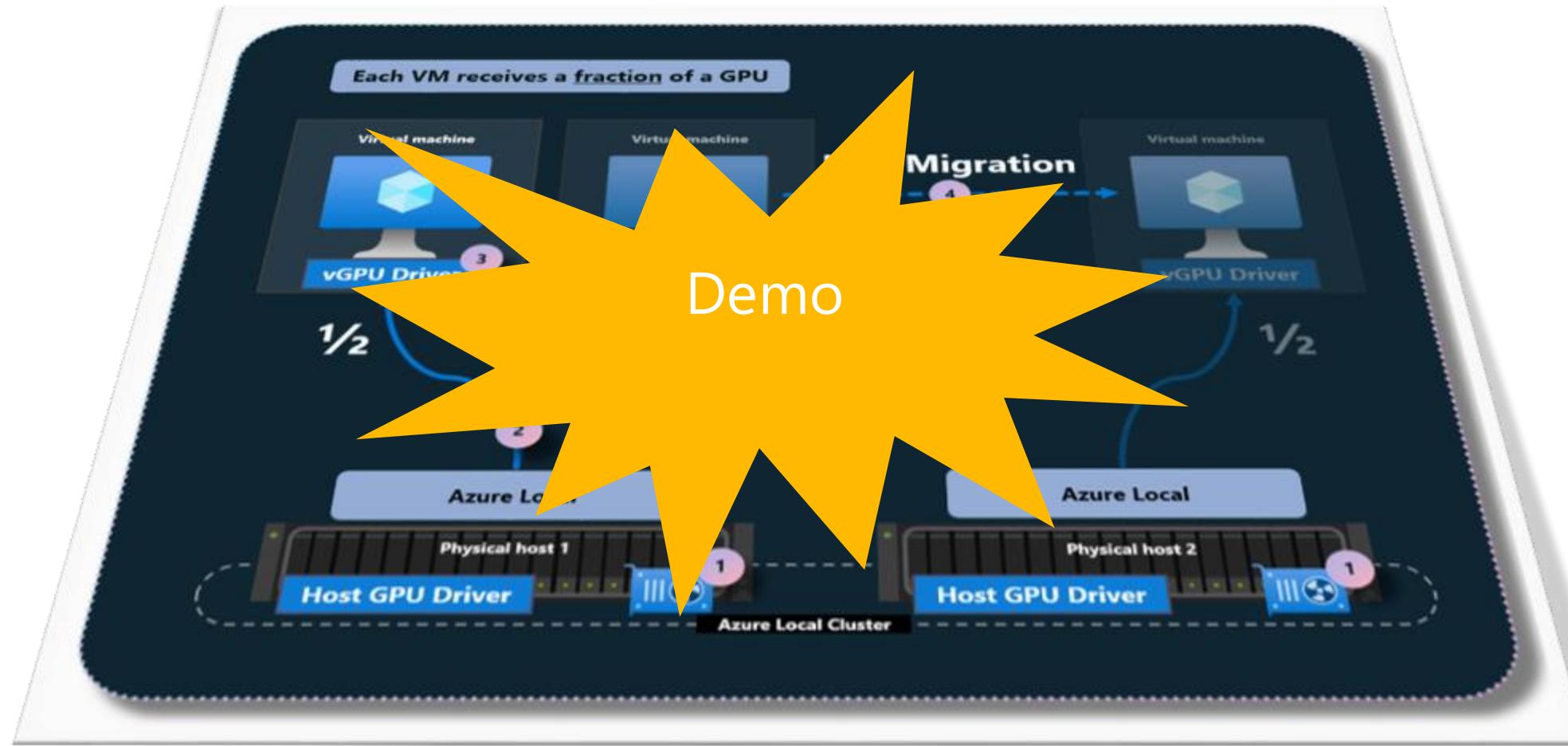
Consider the following functionality and support differences between the two options of using your GPUs:

 Expand table

Description	Discrete Device Assignment	GPU Partitioning
GPU resource model	Entire device	Equally partitioned device
VM density	Low (one GPU to one VM)	High (one GPU to many VMs)
App compatibility	All GPU capabilities provided by vendor (DX 12, OpenGL, CUDA)	All GPU capabilities provided by vendor (DX 12, OpenGL, CUDA)
GPU VRAM	Up to VRAM supported by the GPU	Up to VRAM supported by the GPU per partition
GPU driver in guest	GPU vendor driver (NVIDIA)	GPU vendor driver (NVIDIA)

Quelle: <https://learn.microsoft.com/en-us/azure/azure-local/manage/gpu-preparation?view=azloc-24112#attaching-gpus-on-azure-local>

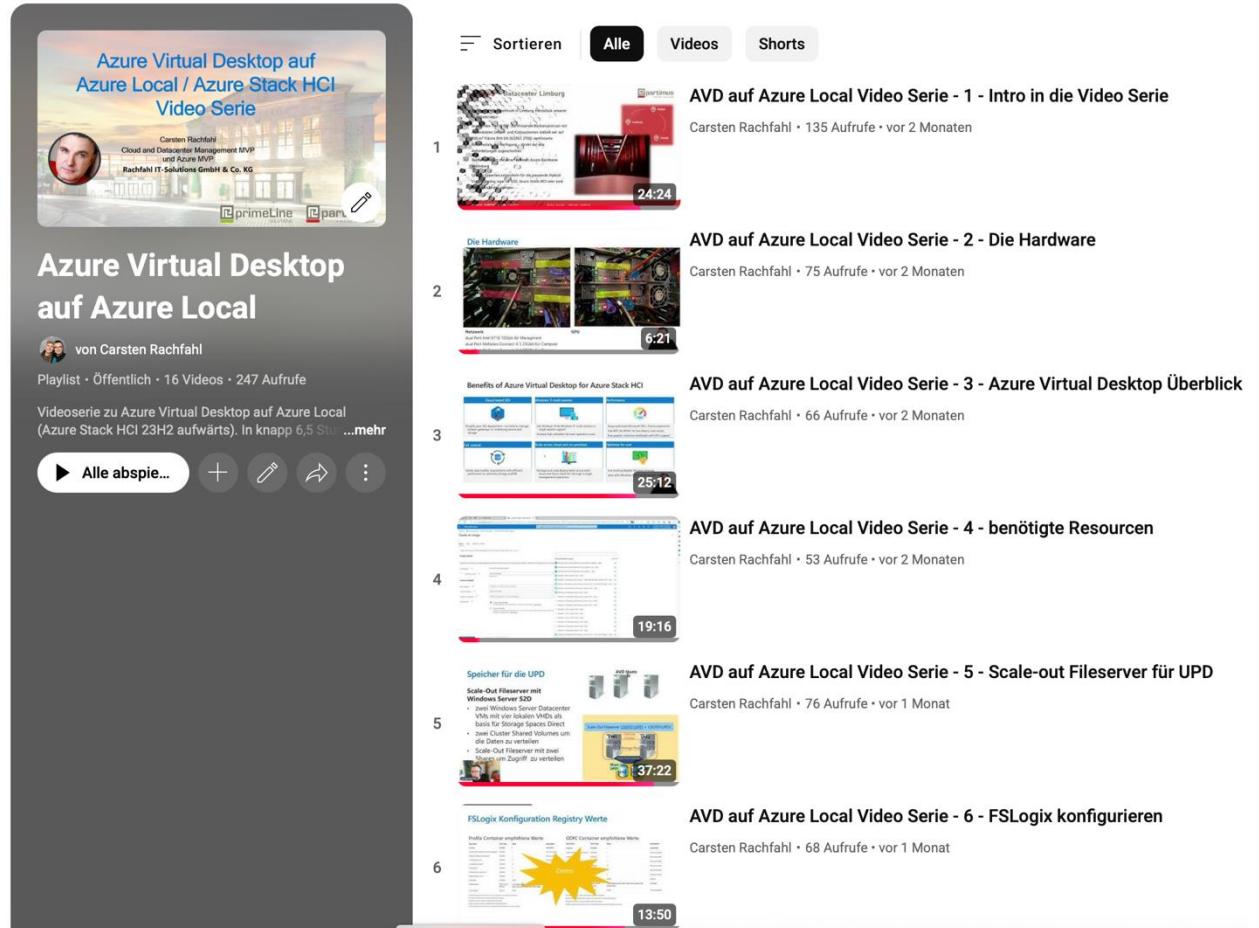
GPU-P Demo



Wer mehr wissen will

Video Serie mit mehr als 6 Stunden (16 Videos)

<https://bit.ly/3HLgomI>



The screenshot shows a YouTube channel interface. At the top, there are buttons for 'Sortieren' (Sort), 'Alle' (All), 'Videos' (Videos), and 'Shorts'. Below this, a thumbnail for the first video is shown, titled 'AVD auf Azure Local Video Serie - 1 - Intro in die Video Serie'. The video has 135 views and was uploaded 2 months ago. The channel name is 'Carsten Rachfahl'. The main content area displays a grid of 16 video thumbnails, each with a title, upload date, and view count. The titles include:

- AVD auf Azure Local Video Serie - 1 - Intro in die Video Serie
- AVD auf Azure Local Video Serie - 2 - Die Hardware
- AVD auf Azure Local Video Serie - 3 - Azure Virtual Desktop Überblick
- AVD auf Azure Local Video Serie - 4 - benötigte Ressourcen
- AVD auf Azure Local Video Serie - 5 - Scale-out Fileserver für UPD
- AVD auf Azure Local Video Serie - 6 - FSLogix konfigurieren

Each video thumbnail includes the channel name 'Carsten Rachfahl' and the number of views.



Q&A

Closing Subtext



Danke an unsere Sponsoren

PLATINUM



GOLD



INFOTECH
[IT & Communication]





Closing

Closing Subtext