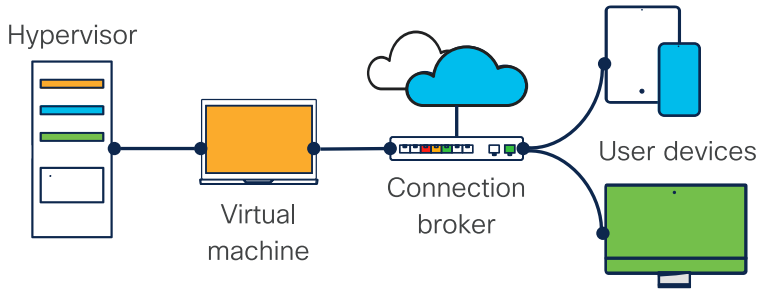
VDI Overview

The majority of U.S. companies expect their employees to spend part of their time working remotely. To manage this, will the cloud be a long-term sustainable solution for everyone? That's the question many IT teams are considering as they look for a cost-effective hybrid workspace model that will deliver a seamless transition between the office and remote workspaces.

VDI Drives Success

The combination of Cisco servers and NVIDIA GPUs are specifically designed to deliver high performance compute for Virtual Desktop Infrastructure (VDI) to support complex workstreams including Tier One workloads and enterprise applications in remote environments.



Graphic: VDI Simplified Concept

The benefits of VDI include:

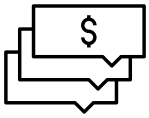
- Simplified image management
- Anywhere, anytime, any device access
- Fast application deployment and upgrades
- Reduced IP loss risk
- Enhanced data security
- Simplified backup, patching, and management
- Instant provisioning of new desktops

Next-Gen GPUs to Power Highly Graphical and AI Applications

NVIDIA software solutions, which includes NVIDIA Virtual GPU (vGPU) for graphics and NVIDIA AI Enterprise for AI, incorporates tools to help you proactively manage and monitor a virtualized environment and provide continuous uptime with support for the live migration of GPU-accelerated VMs. Nvidia GPUs will address your current VDI challenges today, but also position your organization to address future AI challenges.

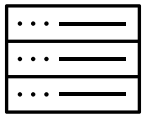
Server Solutions: On-Prem Vs. Cloud

The cloud has helped many companies as they have adjusted their business environments to remote work conditions, and there are benefits within a cloud environment. But as the work landscape settles and what was once an anomaly (remote work) becomes the norm, customers are reassessing their cloud options. What they're finding is that on-prem servers may have an edge over the cloud.



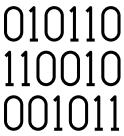
Cost Savings

Intersect360 Research confirmed an analysis* that showed a customer with 2,500 on-prem desktops with the latest Cisco UCS X-series Modular System could save $2M over 3 years vs hosting these desktops on a public cloud.



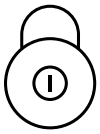
Incorporating Latest Server Technology

Individual companies can purchase the exact technology for their VDI requirements but cloud vendors, due to their size and data center complexities, are forced to incorporate new technology at a more moderate pace due to cost and related factors.



Managed Data Control

VDI enables remote access to mission-critical data points but keeps the data safely on the premises, behind all necessary firewalls.



Improved Security Opportunities

Cloud-based servers provide anywhere-access for employees but this comes with the loss of direct data management control for IT teams.

Intersect360 Research analysis

Accelerate Your Business Growth

With VDI you'll experience:

Security
Centrally managed infrastructure, tools, and processes for deploying and securing virtual desktops to employees on multiple endpoints.

Ease of Install
Simple integration between the hardware and software of multiple vendors, in addition to proven converged infrastructure solutions. All fully tested and documented in Cisco Validated Designs.

Scalability
Scalable infrastructure that ensures high availability of IT services.

Remote Access
The ability to work from any location, on any endpoint, at any time creates a satisfied and more productive workforce.

Bottom Line Success
Higher profitability comes from lower IT infrastructure and management costs along with streamlined productivity solutions.

Opportunities in AI
NVIDIA AI Enterprise software stack makes the transformative power of AI and innovation accessible everywhere.

NVIDIA GPUs

Facilitate greater workplace mobility and enhance productivity across the entire workforce with NVIDIA GPUs which can:

- Provide accelerated computing, maximizing efficiency with a simulation first approach
- Ensure remote workflows that can support growth
- Accelerate remote workloads across industries
- Enhance productivity, mobility, and security
- Future-proof servers with an easy-to-scale flexibility
- Disaster proof with always-on accessibility
- Enhance productivity, mobility, and security
- Enable GPU computing to support enterprise growth in a scalable, cost-effective way

With NVIDIA GPUs to power your complex graphical workloads, you can get stronger compute power with M7 based systems using less hardware, especially if you currently have M4 or M5 servers.

VDI Benefits on Cisco UCS and Rack

Cisco UCS

Cisco UCS can help companies harness VDI for large scale deployments in mid-to-large sized companies. This growth potential allows one Cisco UCS instance to grow into a Converged Infrastructure that can accommodate up to 2,500 users per Cisco UCS chassis with Citrix or VMware Horizon deployments.

Rack

A rack optimized solution is particularly attractive to Power Users in companies with less than 1,000 desktops. It presents a budget-sensitive solution that offers companies a level of control without sacrificing compute performance.

Graphics Acceleration

NVIDIA virtual GPUs bring the graphics-accelerated experience to the virtualized data center. Our products include vApps for application streaming, vPC for standard office workers, and NVIDIA RTX vWS for professional graphics users.

AI Solutions

NVIDIA AI Enterprise accelerates every stage of the AI journey from data prep and model training through inference to deployment at scale. It includes over 100 frameworks, pre-trained models, and development tools optimized for building and running AI on NVIDIA GPUs.

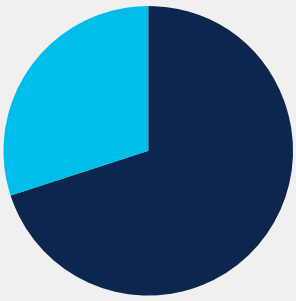
Sustainability Benefits

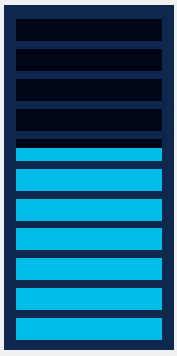
NVIDIA's accelerated computing solutions can meet the demands of your power loads while also improving their sustainability rates with:

- **30% lower power consumption** (NVIDIA liquid-cooled GPUs for mainstream servers)
- **66% less** rack space
- Use of under 1.2 PUEs
- Up to 20x more energy efficient for AI (using NVIDIA GPUs)

30%

lower power consumption (NVIDIA liquid-cooled GPUs for mainstream servers)





66%

less rack space used with NVIDIA's accelerated computing solutions

Finding the Right Solution

Choosing the right GPU for the job

UCSC/X-GPU-H100-80

H100

NVIDIA AI Enterprise

Single GPU AI and HPC secure multi-instance GPU

350W | 80GB | Gen5 | 2-slot FHFL

- Up to 2 GPUs per node
- Up to 28 MIG instances per node
- Up to 8 vCPU cores per MIG

AI training and interference

HPC and data analytics

Confidential compute MIG

Compute and confidential compute

UCSC/X-GPU-L40

L40

NVIDIA vWS

Highest universal AI and graphics

300W | 48GB | Gen4 | 2-slot FHFL

- 3 GPUs per node
- Fastest RT graphics
- Largest render models

Omniverse and cloud XR

Cloud rendering and vWS

Small scale AI

Universal compute and graphics

UCSC/X-GPU-L4

L4 *Not available on M6*

NVIDIA vWS, NVIDIA vPC

Entry-tier universal and edge AI

72W | 24GB | Gen4 | 1-slot FHFL

- 8 GPUs per node
- Video and graphics
- Compact and versatile

Graphics rich VDI with vPC

Inference and video

Edge video and AI

Mobile cloud games and vWS

Compute and confidential compute

Choosing the right GPU for the job

UCSC/X-GP100-80

A100

NVIDIA AI Enterprise
Highest performance compute

300W | 80GB | Gen4 2-slot FHFL

- Fastest compute
- FP64 up to 7 MIG instances
- AI, HPC, data processing

DL training

Scientific research

Data analytics

Compute

UCSC-GPU-A30

A30

NVIDIA AI Enterprise
AI inference and mainstream compute

165W | 24GB | Gen4 2-slot FHFL

- Versatile mainstream compute FP64
- Up to 4 MIG instances

Language processing

Conversational AI

Recommender systems

Compute

UCSC/X-GPU-A40

A40

NVIDIA vWS
Highest performing graphics visual computing

300W | 48GB | Gen4 2-slot FHFL

- Fastest RT graphics
- Largest render models

Cloud rendering

Cloud XR

Omniverse

Graphics

UCSC-GPU-A10

A10 *Not available on M7*

NVIDIA vWS, NVIDIA vPC
Mainstream graphics and video with AI

150W | 24GB | 1-slot FHFL

- 4K cloud gaming graphics
- Video with AI

Virtual workstation

Video conferencing

4K cloud games

Graphics

UCSC/X-GPU-T4

T4 *Not available on M7*

NVIDIA vWS
Small footprint data center and edge inference

70W | 16GB | 1-slot FHFL

- High density video and graphics
- Compact and versatile

Edge AI

Edge video

Mobile cloud games

Graphics

UCSC/X-GPU-A100-80

A16

NVIDIA vPC
High density virtual desktop

250W | 4 x 16GB 2-slot FHFL

- 4K resolution
- Max # of encode/ decode streams

Virtual desktop

Transcoding

Graphics

Let's connect.

If you are updating hardware or platforming a new VDI Cisco and NVIDIA solution, contact us directly to set up an appointment with a Cisco representative.