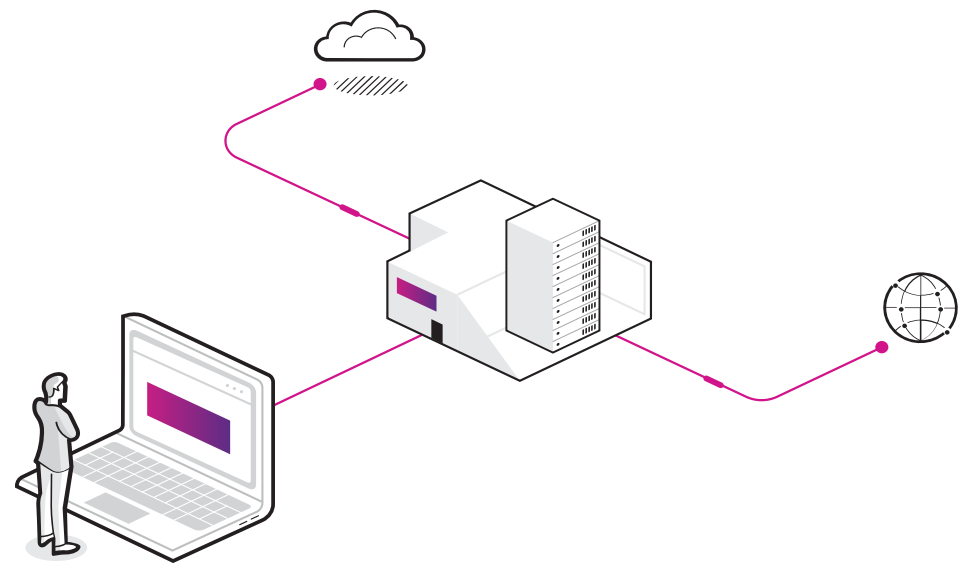


# AI/ML Compute as a Service with NVIDIA DGXA100

Accelerate your AI/ML-powered workloads on your own dedicated infrastructure

Have your artificial intelligence (AI) and machine learning (ML) workloads outgrown your existing infrastructure? Not all data centers are ready to scale AI/ML-powered workloads. But we are.

At Cyxtera, we've raised the bar by becoming the first global data center operator that can deliver access to subscription-based NVIDIA DGX A100 systems via our landmark compute as a service offering.



## Scaling Enterprise AI is Challenging

Whether you're training neural networks, or running inference engines for artificial intelligence processing, you may eventually find that your infrastructure is slowing down critical workloads, making it hard to successfully bring AI/ML projects to the finish line.

While both cloud-based and on-premise AI infrastructure typically work well for proof of concept, AI has many challenges that do not become evident until an organization truly operationalizes and scales its AI efforts.

### AI/ML Compute-as-a-Service

Cyxtera has recognized that organizations have had to compromise on either security, scalability or costs when making infrastructure buying decisions to support AI/ML powered workloads. We understand that enterprises, developers, data scientists, and researchers need a new platform that unifies all AI workloads, simplifies infrastructure and accelerates AI implementation.

That's why we've partnered with NVIDIA to offer AI/ML Compute as a Service, powered by DGX A100 systems, the universal system for all AI workloads - from analytics, to training, to inference.

### Benefits

#### Predictable Costs and Accelerated ROI

No longer rely on testing economies of scale in the cloud as you ramp your AI/ML programs. Manage budgets and upfront cost by shifting from a CapEx to a predictable OpEx model you control.

#### Quickly Scale Up and Down

Avoid over-provisioning, and quickly adapt to market changes and workload demands. Maximize productivity, with powerful automated remote configuration and deployment tools.

#### Maximum Control

Retain full control from hardware up through applications. Provisioning via Cyxtera offers the security and control provided by your own dedicated infrastructure, combined with a cloud-like buying experience.

#### Reduced Operational Costs

With automated remote configuration tools, there is no on-site staff required to rack, stack, configure, or support your NVIDIA DGX A100 hardware.

#### Faster Time-to-Solution

Consolidate training, inference, and analytics into a single unified AI infrastructure that can be deployed in a single business day, speeding up the time-to-solution for new AI/ML workloads.










#### Exceptional Performance at Massive Scale

Experience unprecedented compute density, performance, and flexibility with the world's first 5-petaFLOPS AI system. Massive GPU-accelerated compute with state of the art networking hardware and software optimizations means NVIDIA DGX A100 can scale to meet the biggest challenges, such as conversational AI and large-scale image classification.

## Options for Managing AI/ML Workloads

	On Premises	Traditional Colocation	Public Cloud	Cyxtera AI/ML Compute as a Service
Dedicated hardware	X	X	-	X
Control of data?	X	X	-	X
Nominal upfront investment?	-	-	X	X
Avoids Risk of Vendor Lock-in?	X	X	-	X
Hardware Maintenance Included?	-	-	X	X
Flexibility to experiment without additional cost?	X	X	-	X

## NVIDIA DGX A100 System Highlights

-  The world's first 5 petaFLOPS AI system.
-  Robust security posture for your AI enterprise.
-  Replaces legacy compute architecture with a single, unified system.
-  Fully optimized for end-to-end NVIDIA data center solution stack.
-  Ideal for both single node deployments and large scale Slurm and Kubernetes clusters with the combination of dense compute power and complete workload flexibility.
-  Fastest I/O architecture of any NVIDIA DGX system.
-  Universal for all AI workloads, from analytics to training to inference.
-  Unprecedented compute density.
-  GPUs can be used for large jobs and divided into as many as 56 separate/ fully isolated instances, with dedicated high bandwidth memory, cache, and compute cores.

**Need flexible, robust infrastructure to scale AI/ML projects fast?**

Reach us at [1-855-699-8372](tel:1-855-699-8372) or [Sales@Cyxtera.com](mailto:Sales@Cyxtera.com).

## System Specifications

<b>GPUs</b>	8x NVIDIA A100 Tensor Core GPUs
<b>GPU Memory</b>	320 GB total
<b>Performance</b>	5 petaFLOPS AI 10 petaOPS INT8
<b>NVIDIA Switch™</b>	6
<b>System Power Usage</b>	6.5KW max
<b>CPU</b>	Dual AMD Rome 7742, 128 cores total, 2.25GHz(base), 3.4 GHz (max boost)
<b>System Memory</b>	1TB
<b>Networking</b>	8x Single-Port Mellanox ConnectX-6 VPI 200Gb/s HDR InfiniBand 1x Dual-Port Mellanox ConnectX-6 VPI 10/25/50/100/200Gb/s Ethernet
<b>Storage</b>	OS: 2x 1.92TB M.2 NVME drives Internal Storage: 15TB (4x 3.84TB) U.2 NVME drives
<b>Software</b>	Ubuntu Linux OS
<b>System Weight</b>	271 lbs (123 kgs)
<b>Packaged System Weight</b>	315 lbs (143 kgs)
<b>System Dimensions</b>	Height: 10.4 in (264.0 mm) Width: 19.0 in (482.3 mm) Length: 35.3 in (897.1 mm)
<b>Operating Temperature Range</b>	5°C to 30°C (41°F to 86°F)