Lenovo
TruScale

Unleash AI innovation across your business with TruScale GPUaaS

A rock-solid foundation for sharing GPU resources

For AI innovators and companies in industries such as life sciences, cloud gaming, entertainment, and others, there are likely to be multiple GPU-intensive initiatives active at any one time. Internal IT teams are tasked with supporting them all, but often there is not a perfect solution for doing so.

Public cloud can be launched quickly and carries little management overhead. However, prices have steadily risen and changes to pricing models can be difficult to keep track of — more than 70% of organizations have experienced higher than anticipated public cloud costs.[1] Moreover, orchestration and resource allocation in public clouds can be complicated and often require third-party tools.

On-site infrastructure provides security advantages and can be customized to your unique needs. But purchasing infrastructure requires a significant up-front investment of time and budget, and valuable ongoing resources to coordinate, manage, and maintain. Keeping infrastructure refreshed and ready for evermore demanding AI applications is also a challenge for many organizations — just 34% of IT leaders say their data infrastructure is AI ready.[2]

Adding to the risks, GPUs are sometimes difficult to procure, and long lead times to acquire them can delay projects. Lead times for NVIDIA GPUs that are purpose-built for AI applications have recently ranged between 3 and 11 months.[3]

Organizations managing multiple GPU-intensive applications require rock-solid orchestration capabilities to balance budgetary and resource constraints with shifting priorities.

70% of organizations experienced public cloud costs that were significantly higher than initially anticipated[1]

34% of IT leaders say their computing infrastructure is AI ready[2]

3-11 months
Recent range of lead times for AI-ready NVIDIA GPU[3]

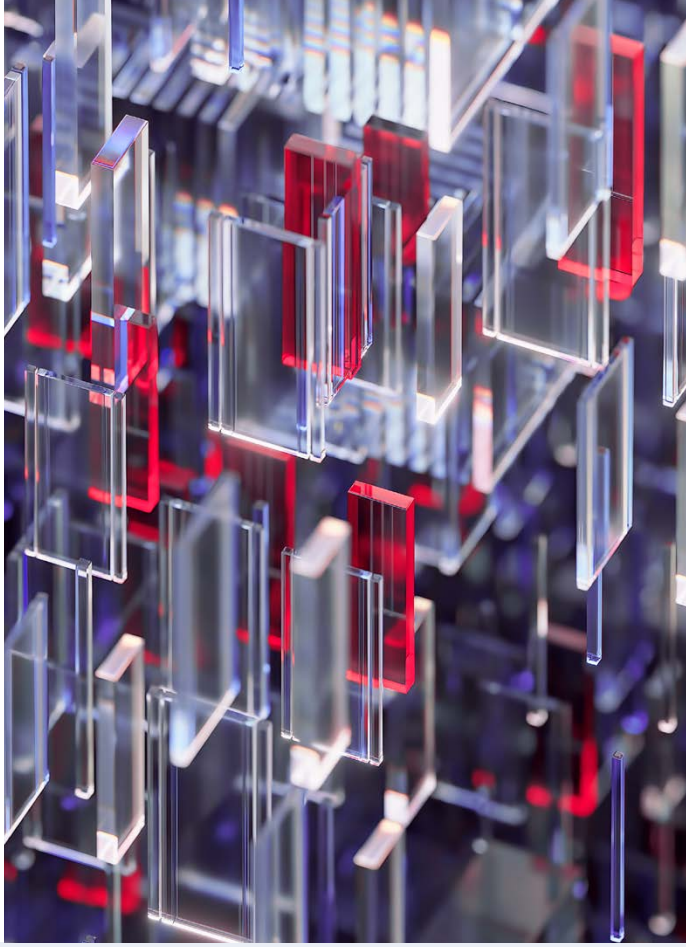
Smarter technology for all

Lenovo

Accelerate AI outcomes with Lenovo TruScale GPUaaS

Cost-effective, state-of-the-art AI infrastructure with robust multi-workload management capabilities

TruScale GPU-as-a-service (GPUaaS) is a patent-pending solution that pairs Lenovo infrastructure with NVIDIA GPUs, offered as an enhancement of **Lenovo TruScale for HPC**. TruScale for HPC provides fully managed AI-ready hardware, software, and services with predictable billing and flexible utilization, which with the addition of TruScale GPUaaS adds powerful orchestration to allocate workloads and facilitate metering. Together, these solutions comprise an economical alternative to public cloud that removes the resource barriers to hosting on-site infrastructure while delivering high-end performance, Lenovo reliability, and management flexibility for organizations with multiple high-GPU workloads.



Processing power when and where it's needed

Orchestration of GPUs enables consumption and costs to be managed by workload. Provision, meter, and monitor GPU usage for resource sharing across teams and ensure GPUs are available when and where they are needed most.

More time for what matters

A fully managed onsite service, **TruScale GPUaaS** frees up valuable resources for strategic demands while Lenovo maintains your infrastructure. Reduce operational demands on data science teams so they can focus on analysis and AI development.

Control costs

Avoid surprisingly high bills, unexpected fees, and other public cloud pitfalls. **TruScale for HPC bundled with Truscale GPUaaS** provides predictable, subscription-based costs that make the up-front expenses of moving to onsite infrastructure more manageable while enabling you to manage infrastructure usage with visibility and control.

Improve performance

Powerful NVIDIA GPUs and the latency advantages of moving onsite can accelerate AI/ML and HPC workloads.

Access to AI innovators

More than 50 **Lenovo AI Innovator** partners participate in the **Lenovo AI Discovery Center**, where they evaluate and execute proofs of concept on more than 165 validated industry- and domain-specific business-ready solutions.

An ideal choice for data sovereignty

TruScale for HPC keeps your data behind your firewall. Comply with data sovereignty requirements, manage industry regulations, and gain control of your data.

Smarter technology for all

Lenovo

Future-proofed AI infrastructure that scales in lockstep with your needs

Increase business agility

Gain the agility to quickly scale up new projects or shift priorities. Add infrastructure on demand, quickly adjust orchestration parameters, and leverage TruScale consultants with deep AI experience.

Avoid GPU shortages

GPUs can be difficult and expensive to procure, including market-leading NVIDIA GPUs. **TruScale GPUaaS** allows your team to avoid the GPU procurement process and ensures a steady supply, even during periods of rapid growth.

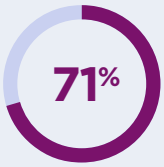
Permanently current

Frequent technology refresh cycles with the latest solutions from Lenovo and NVIDIA help ensure your infrastructure is future-proofed and ready to support your growth.

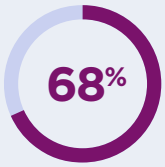
Improve data center sustainability

Lenovo experts are ready to assist with optimizing power consumption and cooling practices, improving data center sustainability with tailored planning, implementation, and maintenance services. Our Neptune power and cooling technology can reduce overall power consumption by up to 40% compared to air-cooled systems, helping to drive greater value from your infrastructure.

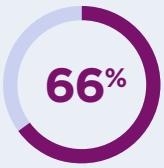
Concerns listed as one of the top two about digital infrastructure over the next 12 months:[4]



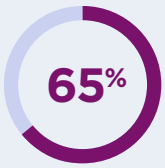
Lack of qualified staff



Procuring equipment



Forecasting future capacity



Improving energy performance for IT

Smarter technology for all

Lenovo

A rock-solid foundation for AI

High-end NVIDIA GPUs with ultrareliable Lenovo appliances deliver AI-ready performance you can depend on. Backed by the services expertise of Lenovo, a trusted partner to IT leaders worldwide for four decades, **TruScale GPUaaS** is a patent-pending solution that can help you maintain control of your data and efficiently manage all of your AI initiatives.

The vision is yours. Get there with Lenovo.

To learn more, contact your Lenovo sales representative.

#1
x86 server reliability 10 consecutive years[5]

#1
x86 security 5 consecutive years[5]

#1
Top-ranked supercomputer provider 6 consecutive years[6]

#1
Green500 most energy-efficient supercomputer[7]

Smarter technology for all

Lenovo

Sources
1 CFO, "Preventing runaway cloud costs in 2024," December 2023
2 Lenovo, "Lenovo Global CIO Report 2024," April 2024
3 Forbes, "Nvidia Stock Soars After-Hours On 265% Revenue Growth," February 22, 2024
4 Uptime Institute, "Uptime Institute Global Data Center Survey 2023," July 2023
5 Information Technology Intelligence Consulting, "ITIC 2023 Global Server Hardware, Server OS Reliability Report," September 2023
6 TOP500.org, "Top500.org June 2024 rankings," June 2024
7 TOP500.org, "Green500 June 2024 rankings," June 2024