CUDO COMPUTE

Distributed cloud, purpose-built to overcome supply chain issues

Developers

Console UI   **CLI**   **Terraform**   **API**

Virtual machines   Automation

COMPUTE

Networking   Security

Storage   Billing

Bare metal suppliers
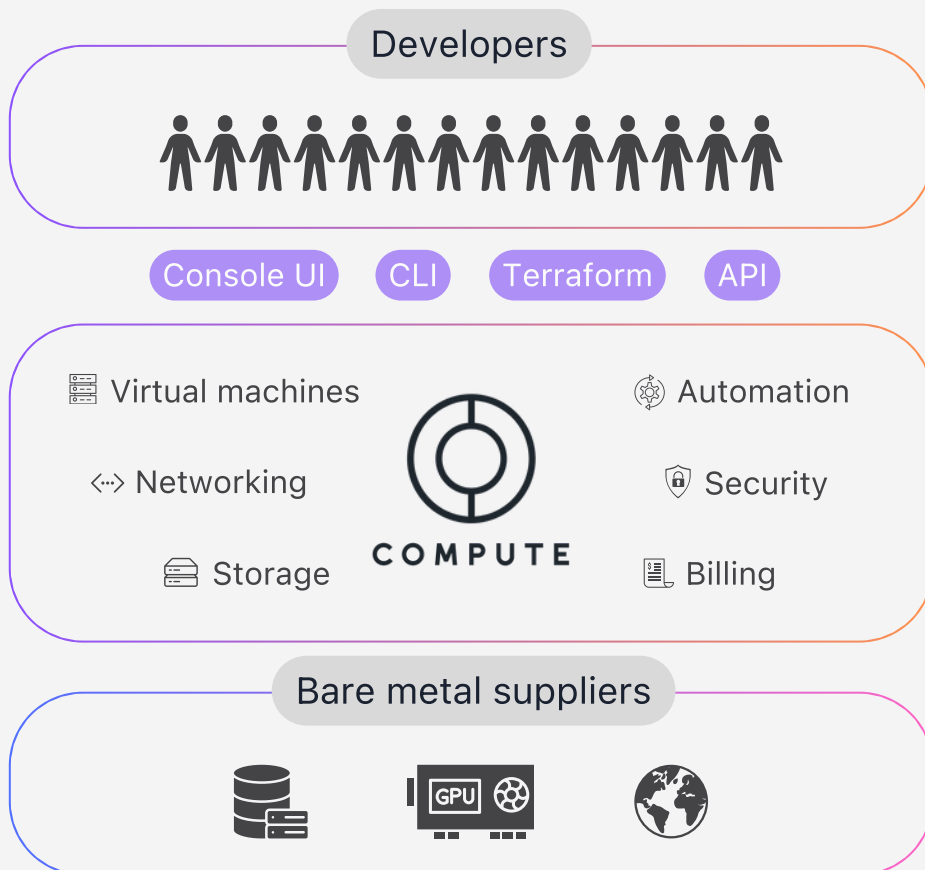
Cloud-based Nvidia GPUs for AI training and inference

Available

Cudo Compute solves GPU supply chain challenges by bringing together global HPC suppliers in one intuitive cloud marketplace.

Scalable

Businesses can scale up or down, ensuring they have the resources they need during demand spikes and saving money during low-demand periods.

Lower TCO

Cloud GPUs require no upfront investment, reducing total cost of ownership. Maintenance and upgrade costs are also lower since they occur in the cloud.

High-Performing

Cloud GPUs offer significant performance benefits over on-premise GPUs, with lower latency and the ability to handle large-scale training datasets.

Learn more at www.cudocompute.com

Cloud-based GPUs from Cudo Compute

Cudo Compute offers a range of cloud-based GPUs for AI training and inference. By leveraging our revolutionary global supplier network and on-demand marketplace, developers can overcome supply chain issues and enjoy the benefits of accessibility, lower TCO, scalability, and performance.

GPU	Memory & Bandwidth
NVIDIA RTX A40	48GB GDDR6 (696 GB/s bandwidth)
NVIDIA A100	80GB HBM2e (1,935 GB/s bandwidth)
NVIDIA H100	80GB HBM2e (2 TB/s bandwidth)
NVIDIA RTX A4000	16GB GDDR6 (448 GB/s bandwidth)
NVIDIA RTX 4090	24GB GDDR6X (1008 GB/s bandwidth)
NVIDIA RTX A5000	24GB GDDR6 (768 GB/s bandwidth)
NVIDIA RTX A6000	48GB GDDR6 (768 GB/s bandwidth)
AMD MI250/300	128GB HBM2e (3276.8 GB/s bandwidth)

Use multiple lower-spec cards

When there are tight deadlines or budgets, multiple Nvidia RTX A6000 cards can be used to provide similar performance to the A100 at a lower cost.

Run AI models slightly slower

By adjusting performance requirements, you can still achieve your training goals without relying on high-spec GPUs when budgets are tight.

Get started now

Start running your training and inference workloads on cloud-based GPUs instantly. With Cudo Compute, there are no supply chain issues and no need to invest in dedicated hardware.