

Data Center-Class GPU Cloud

Blackwell-ready compute for AI training, inference, rendering, media generation and enterprise private deployments.

Blackwell

Dedicated GPU

Bare Metal

Private Cluster

Built for enterprise AI workloads.

What customers can buy

SilicaDrive GPU Cloud is designed for teams that need secure compute without buying and operating their own GPU servers.

Dedicated GPU Nodes

Single-tenant GPU servers for training, inference, rendering and custom AI stacks.

Reserved GPU Capacity

Pre-booked capacity for predictable usage windows, pilots and recurring production jobs.

Private AI Clusters

Isolated environments for enterprise workloads, sensitive data and controlled access.

AI Studio Backend

GPU capacity behind image and video generation products.

Custom Deployment

Architecture support, workload testing and production setup assistance.

Data center GPU options

GPU availability depends on region, configuration and customer requirement. SilicaDrive focuses only on data center-class GPUs.

RTX PRO 6000 Blackwell SE

96GB GDDR7. Universal AI, visual computing, rendering, inference and video workloads.

B200 / GB200 Blackwell

Rack-scale Blackwell options for major training and high-throughput inference deployments.

H200 / H100

Hopper generation GPUs for LLM training, inference, HPC and large memory workloads.

L40S

Universal data center GPU for generative AI, graphics, rendering and video acceleration.

L4

Energy-efficient inference, video, visual AI and low-profile data center deployments.

A100

Mature AI, data analytics and HPC GPU platform for stable production workloads.

Choose GPU by workload

Recommended data center-class options for common AI and visual compute needs.

01 Training & fine-tuning

B200 / GB200, H200, H100, A100

02 High-memory LLM inference

H200, B200, RTX PRO 6000 Blackwell SE

03 Image / video generation

RTX PRO 6000 Blackwell SE, L40S, H100

04 3D rendering & simulation

RTX PRO 6000 Blackwell SE, L40S

05 Video AI & streaming

L4, L40S, RTX PRO 6000 Blackwell SE

06 Cost-efficient pilots

L4, L40S, A100

From one node to private cluster

01

Dedicated Node

Customer gets isolated GPU server access for direct workloads.

02

Reserved Capacity

Capacity block is reserved for planned runs and recurring jobs.

03

Private Cluster

Multi-node environment with controlled access and support.

04

Managed AI Stack

Portal, billing, usage visibility, APIs and customer support.

Enterprise-ready operating model

Isolation

Single-tenant GPU nodes and private deployment options for controlled workloads.

Visibility

Usage visibility, monitoring, support tickets and customer onboarding support.

Policies

Clear acceptable use, access rules and support process for business customers.

Recommended for teams that need predictable access, data control and practical deployment assistance.

Simple onboarding path

A clear process from requirement capture to production support.

01 Requirement capture

GPU type, quantity, workload, duration, data sensitivity and support needs.

02 Sizing & quote

Match workload to GPU class, node type, storage and network requirement.

03 Provisioning

Dedicated access, environment setup and initial workload validation.

04 Production support

Usage monitoring, support channel and capacity planning for scale.

Designed for business customers who need controlled GPU access and practical deployment help.

Build AI workloads on dedicated GPU infrastructure.

GPU Cloud | Bare Metal | Private AI Clusters | AI Studio

Website: silicadrive.com

Sales: sales@silicadrive.com

Support: support@silicadrive.com

Ask for a GPU capacity assessment to match your workload to the right data center GPU class.