

A GPU cloud security review

Check data control, access and evidence before upload.

Updated October 29, 2026

Classify the data before you upload it. Include prompts, training examples, outputs, logs and checkpoints. A model artifact can contain sensitive information even if the input files are removed.

Access and network controls

Use scoped service credentials. Keep deployment credentials separate from personal access. Ask how the service records administrative access and how you export audit logs. Test credential revocation. Confirm whether the GPU worker needs a public address and whether outbound access can be restricted.

Data lifecycle

Record where each artifact lives. Ask how temporary disks, persistent volumes, backups and snapshots are deleted. Request the retention rules for logs and failed jobs. Verify the location of storage as well as compute. A regional endpoint alone does not prove where support staff or subprocessors can access data.

Evidence for approval

Ask for a data processing agreement and the scope of any audit report. Match the report's service and location to your order. A badge is not proof that your workload is compliant. Have your security and legal reviewers decide whether the evidence meets your obligations.

Kubernetes documentation explains how GPU resources are exposed to workloads. It is not an isolation certificate. Review the provider's tenant boundary and host administration model separately. Test the controls with harmless trial data before production. Save the approval, open issues and the person responsible for follow-up.

Sources

- [Kubernetes GPU scheduling](#) — fetched 2026-10-29.

Frequently asked questions

Is this a vendor quote?

No. Confirm current terms and capacity with the provider before you rent.