



Simplify GPU Usage Across Any Kubernetes Cluster

**Serverless GPU – GPU provisioning made
easy, built for Seeweb by CLASTIX**

Machines That Need to Learn Are Hungry for Resources

COMPANY SUMMARY

- **Company name:** Seeweb
- **Website:** seeweb.it/en
- **Headquarters:** Italy
- **IT Environment:** Serverless GPU service built on Capsule, integrated into Seeweb's Cloud ecosystem
- **Industry:** Cloud Infrastructure / Hosting Provider
- **No. of employees:** ~50

HIGHLIGHTS

- **Serverless GPU-as-a-Service**, integrated into Seeweb's Cloud ecosystem
- **Launch a GPU workload in under 3 minutes**, with no dedicated hardware to buy or manage
- **Automatic multi-tenant policy enforcement** – network policies, quotas, and fair-use – via Project Capsule
- **99.99% uptime SLA**, delivered from Italy- and Europe-based infrastructure

Seeweb is a leading Italian cloud infrastructure provider, delivering European-based hosting, computing, and AI infrastructure services. As demand for GPU-accelerated AI workloads grew, Seeweb set out to offer a serverless GPU service that its customers could consume on demand, without buying or managing dedicated hardware.

CHALLENGES

AI adoption is accelerating quickly: Italy's AI market reached €760 million in 2023, with 6 out of 10 large enterprises already using the technology (source: EY Italy AI Barometer, Il Sole 24 Ore Radiocor). Demand for compute is following the same curve worldwide.

At the center of that growth sits the GPU – the hardware best suited to AI workloads, and some of the most sought-after and costly compute available today. Hosting GPUs on premises is expensive, and even renting them from a cloud vendor traditionally meant a slow, manual process of selecting, configuring, and provisioning a remote machine before a model could even run.



GPU shortages were slowing AI innovation across Europe as a result, and Seeweb wanted a different model: a public, serverless GPU service serving many customers from a shared pool of hardware, while still giving each one strict isolation, predictable performance, and a simple, standard Kubernetes experience. Doing this safely at scale meant solving multi-tenancy, control-plane isolation, and elastic GPU scheduling all at once.

SOLUTION

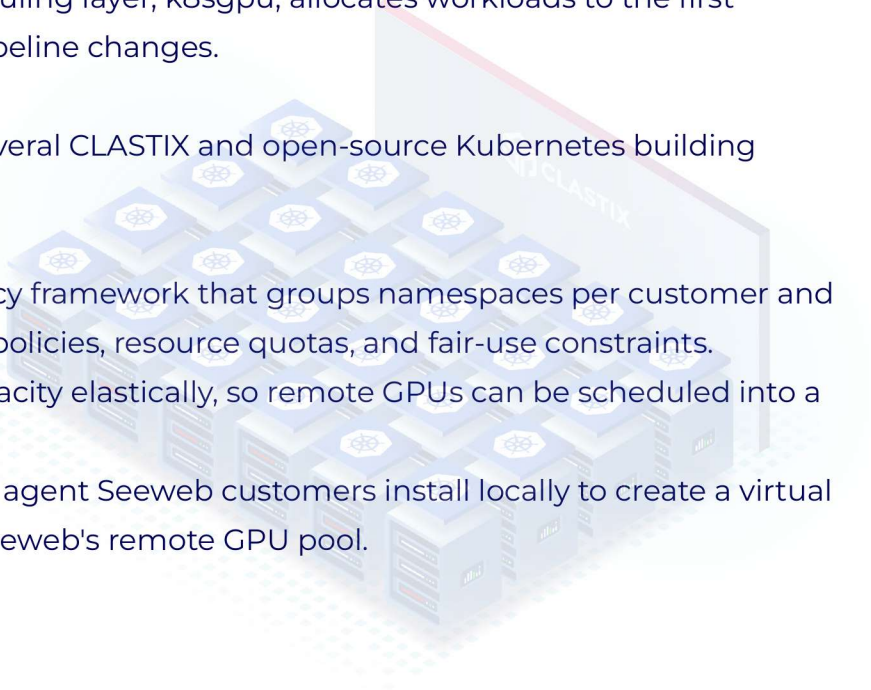
CLASTIX worked with Seeweb to build a fundamentally different kind of access to GPUs: instead of provisioning and managing GPU servers directly on the Cloud Provider machine, users access any remote GPU straight from their local Kubernetes cluster, integrating with the tools they already use and love. Installing a lightweight agent creates a virtual node presenting remote GPUs, immediately available for scheduling AI workloads.

Kubernetes virtual-kubelet made this possible: the tool enables a Kubernetes cluster to access any remote resource as a local resource in Kubernetes.

Built on that foundation, Serverless GPU abstracts the underlying architecture away entirely: users already runs a local Kubernetes cluster, can “mount” Seeweb's GPUs as "remote" resources inside their own infrastructure, with rapid spin-up, responsive auto-scaling, and hourly pay-per-use billing. An advanced scheduling layer, k8sgpu, allocates workloads to the first available GPU instance, requiring no pipeline changes.

Underneath, the platform combines several CLASTIX and open-source Kubernetes building blocks:

- **Project Capsule** – a multi-tenancy framework that groups namespaces per customer and automatically enforces network policies, resource quotas, and fair-use constraints.
- **Virtual Kubelet** – pools GPU capacity elastically, so remote GPUs can be scheduled into a customer's cluster on demand.
- **k8sGPU agent** – the lightweight agent Seeweb customers install locally to create a virtual node, bridging their cluster to Seeweb's remote GPU pool.



Ongoing Benefits: Under 3 Minutes to a GPU Workload

At the heart of the service lies a synergy between Seeweb's computational capacity, specially optimized for AI, and CLASTIX's innovative approach to Kubernetes – a shared vision of making GPU technology accessible and affordable for AI projects of all sizes.

With Serverless GPU, **teams can launch a GPU workload in under three minutes**, paying only for actual hourly usage rather than buying and managing dedicated hardware.[1] **Project Capsule handles control-plane isolation and policy enforcement automatically for each tenant**, cutting down on manual configuration work.[1]

The service works with standard Kubernetes environments – managed offerings like AKS, EKS, and GKE, as well as on-prem distributions such as vanilla Kubernetes, OpenShift, Tanzu, and Rancher – so customers keep their existing pipelines, delivered from Italy- and Europe-based infrastructure with an uptime SLA above 99.99%.[2][3]

One successful use case leveraging Seeweb's Kubernetes GPU technology enabled by CLASTIX is the collaboration with RedCarbon, a cybersecurity company that uses Serverless GPU to train and optimize its AI threat-detection models. **The switch has cut RedCarbon's model training time by 40%, improving its response speed to attacks.**[2]

NEXT STEPS

Seeweb and Clastix invite AI teams to stay tuned for further updates and prepare to unlock unprecedented flexibility and efficiency in their AI projects. For teams ready to accelerate AI code releases and scale up their available microservices without added effort, new GPU instances can be launched with Serverless GPU today. Talk to us to get started.

SOURCES

Primary announcements:

CLASTIX – "Simplify GPU usage across any Kubernetes Cluster":

<https://clastix.io/post/simplify-gp-us-usage-across-any-kubernetes-cluster/>

Seeweb – "Serverless GPU: il GPU provisioning facilitato" (Italian):

<https://blog.seeweb.it/serverless-gpu-il-gpu-provisioning-facilitato-2/>

Supporting (Ongoing Benefits/ROI figures):

[1] CLASTIX – "GPU-Based CaaS, Platform Engineering, and the Future of Multi-Tenancy" (17 Nov 2025, Dario Tranchitella):

<https://clastix.io/post/gpu-based-caa-s-platform-engineering-and-the-future-of-multi-tenancy/>

[2] AiThORITY – "Seeweb Launches Serverless GPU: The Solution to Overcome GPU Shortages and Accelerate AI Innovation":

<https://aithority.com/machine-learning/seeweb-launches-serverless-gpu-the-solution-to-overcome-gpu-shortages-and-accelerate-ai-innovation/>

[3] Seeweb – Serverless GPU product page: <https://www.seeweb.it/en/products/serverless-gpu>



About Seeweb

Seeweb is a European cloud computing and hosting provider founded in Italy in 1998. As an established market leader and part of the DHH International group, the company manages multiple proprietary, state-of-the-art data centers across Italy and Switzerland. Seeweb specializes in high-performance infrastructure solutions tailored for IT businesses, system integrators, software vendors, and web agencies. Its robust portfolio ranges from traditional colocation, managed databases, and scalable Cloud Servers to advanced GPU Computing as a Service designed to accelerate Artificial Intelligence (AI) and Machine Learning (ML) workloads. Built upon a foundation of absolute reliability and European data sovereignty, Seeweb guarantees 99.99% uptime SLA backed by proactive 24/7 technical engineering support. Furthermore, the company is deeply committed to ecological responsibility, running its entire infrastructure on 100% renewable energy as a verified gold partner of The Green Web Foundation.

About CLASTIX

Clastix builds enterprise-grade Kubernetes infrastructure through hosted control planes and hard multi-tenancy – solving cluster sprawl, fragile isolation, and runaway infrastructure costs at scale. With deep open source credibility – including the donation of Capsule to the CNCF – the kMetal platform is purpose-built for cloud providers, MSPs, and platform teams that demand real isolation, efficiency, and scale. Customers include NVIDIA and Mistral AI.

Contact us

Please contact us to request a quote or learn more about our support offerings.



CLASTIX SRL
V. Francesco Caracciolo, 17
80122 Naples, Italy

VAT No. IT11183320966
REA: NA-1152378
© 2026 Clastix | All Rights Reserved

hello@clastix.io
www.clastix.io
<https://github.com/clastix>