

ClearML Unveils New Hassle-Free Solution for Secure GPU Sharing Across Multiple Teams in Enterprise Environments

Enterprises can now broaden AI compute availability across departments & business units with secure multi-tenancy, maximizing cost efficiency & GPU performance

SAN FRANCISCO, CA, UNITED STATES, October 10, 2024 /EINPresswire.com/ --

[ClearML](#), the leading solution for unleashing AI in the enterprise, today announced the launch of its new multi-tenant [GPU-as-a-Service for Enterprise](#).

Now, enterprises, research labs, and public sector agencies with large on-premise or cloud computing clusters can offer and manage shared computing on fully isolated networks for multiple teams, business units, or subsidiaries.



ClearML Unveils New Hassle-Free Solution for Secure GPU Sharing Across Multiple Teams in Enterprise Environments

Enabling secure multi-tenancy on GPU clusters provides access and availability of compute to a greater number of AI builders to address more use cases for the organization. By increasing the utilization of GPU machines, ClearML helps enterprises improve ROI and reach break-even faster on their latest state-of-the-art AI infrastructure investments. That's important because enterprises are actively acquiring more computing power to address GenAI use cases and applications, with the global graphics processing unit (GPU) market valued at 65.3 billion U.S. dollars in 2024, according to Statista (<https://www.statista.com/statistics/1166028/gpu-market-size-worldwide/>).

For organizations struggling with efficiently provisioning and tracking compute consumption by stakeholder and issuing invoices or internal chargebacks, ClearML provides real-time tenant-based reporting on computing hours, data storage, API calls, and other chargeable metrics that can be integrated into a customer's financial system.

For business and AI leaders, ClearML serves as a single pane of glass for monitoring and

managing AI resources, offering full visibility and granular control over who and how resources are accessed and used. With this vital transparency, IT teams and AI builders can better control budgets and performance as well as ensure compliance with AI governance and data security requirements.

Previously, enterprises, research labs, and public sector agencies juggling multiple stakeholder access to shared compute could either manually manage provisioning for a single user at a time or through a specialized point solution that needed to be added on top of their already cumbersome tech stack.

However, the existing solutions in the market that aim to tackle this issue fall short on offering a robust, unified, enterprise-grade offering. They lack centralized dashboards and controls, support for complex infrastructures such as multi-cloud setups, legacy HPC environments, or environments with a mix of Kubernetes and bare metal clusters; as well as fully secure multi-tenancy. IT and DevOps teams managing infrastructure allocation manually, or even relying on tools that only solve part of the problem, must still address a significant amount of maintenance and overhead that fails to solve the problem of underutilized resources, resulting in a lost opportunity to improve the AI team's time-to-production.

ClearML's hardware-agnostic approach provides a streamlined way for infrastructure leaders to view their clusters on a single network for complete visibility and seamless management while growing their computing power and future-proofing their investments with GPUs from different chip manufacturers. Our broad support for AI orchestration also enables organizations to combine their existing GPU / HPC computing clusters of CPUs.

Modern infrastructure teams are also finding themselves managing multiple different types of computing clusters (on-premise, cloud, hybrid, Kubernetes, bare metal, etc.). By using ClearML's flexible, hardware- and silicon-agnostic architecture to unify and make accessible all of their different computing clusters together on a single infrastructure network, IT and AI infrastructure leaders can extract the most value from their existing compute investments. They gain visibility and the ability to closely monitor and control which resources are used, ensuring the most expensive computing clusters are always fully utilized.

ClearML supports hybrid infrastructures for any deployment type, enabling both the control plane and compute to be either in the cloud or on-prem, (including air-gapped systems) without compromising on any features or functionality. This seamlessly provides end users with an efficient and secure cloud experience whether accessing compute on-premise or in the cloud.

KEY BENEFITS FOR ENTERPRISES INCLUDE:

- Improved ROI and cost management as organizations get more mileage from their compute investments.
- Optimized availability for AI builders to access computing resources for AI development and

deployment along with optimized usage via fractional GPUs.

- Frictionless scalability with the freedom for customers to build additional clusters using any hardware or cloud of their choice. Our support for hybrid infrastructure helps optimize costs and preserve the cloud experience.
- Enterprise-grade security with SSO authentication and LDAP integration that ensures access to data, models, and compute resources are restricted to approved users and teams.
- Full operational oversight with a single pane of glass on an infrastructure that supports diverse workloads, including AI and HPC, by monitoring and controlling job queues.

“Our flexible, scalable, multi-tenant GPUaaS solution is designed to make shared computing a working reality for any large organization,” said Moses Guttmann, Co-founder and CEO of ClearML. “By increasing AI throughput for AI model and large language model (LLM) deployment, we’re helping our clients achieve a frictionless, unified experience and faster time-to-value on their investments in a completely secure manner.”

ClearML’s solution for secure multi-tenancy and shared computing is part of its AI Infrastructure Control Plane (<https://clear.ml/ai-infrastructure-control-plane>), a unified, end-to-end infrastructure management for AI development and deployment. AI teams leveraging ClearML benefit from a single AI workbench that powers AI development and deployment and increases GPU throughput with optimized utilization.

Built for the most complex, demanding environments and novel enterprise use cases, our open source, end-to-end architecture offers a frictionless and scalable way for businesses to run their entire AI lifecycle – from lab to production – seamlessly on shared GPU pools.

To learn more about ClearML’s GPU-as-a-Service for Enterprise, visit our website or [request a demo](#).

About ClearML

As the leading open source, end-to-end solution for unleashing AI in the enterprise, ClearML is used by more than 1,600 enterprise customers to develop a highly repeatable process for their end-to-end AI model lifecycle, from product feature exploration to model deployment and monitoring in production. Use all of our modules for a complete ecosystem or plug in and play with the tools you have. ClearML is an NVIDIA partner and is trusted by more than 250,000 forward-thinking AI builders and IT teams at leading Fortune 500 companies, enterprises, academia, public sector agencies, and innovative start-ups worldwide. To learn more, visit the company’s website at <https://clear.ml>.

Noam Harel

ClearML

PR@clear.ml

Visit us on social media:

X

[LinkedIn](#)

[YouTube](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/750397160>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2024 Newsmatics Inc. All Right Reserved.