

# TRANSFORMING ENTERPRISE AI: CLEARML'S AI PLATFORM MEETS ZADARA'S SOVEREIGN AI CLOUD

Automate and monetize enterprise AI for CSPs with ClearML's end-to-end AI operations platform and Zadara's sovereign AI Infrastructure-as-a-Service

## A LAYERED APPROACH TO ENTERPRISE AI:

# INFRASTRUCTURE ISOLATION AND WORKLOAD ORCHESTRATION

## THE CHALLENGES OF THE MARKET

Enterprise Artificial intelligence (AI) is quickly maturing from isolated experimentation to mission-critical and an integral component of contemporary business processes. Enterprises now rely on AI for myriad use cases, such as in automating customer support, enabling natural-language enterprise "ask-my-data", or driving a growing range of operational and decision-making workflows.

However, it is a costly business to build your own AI factory, so enterprises are driven towards consumption-based service models, sourced from AI-as-a-Service and GPU-as-a-Service providers, such as neoclouds.

In response to growing enterprise demand for AI services, Cloud Service Providers (CSPs) are investing heavily in large-scale GPU clusters. Given the capital intensity of these deployments, achieving return on investment depends on maintaining consistently high utilization across GPUs, compute, network, and data infrastructure, requiring AI GPU clusters to be continuously shared among multiple tenant workloads.

In baseline deployments, and in the absence of AI-specific orchestration and scheduling, GPU utilization in AI clusters remains significantly below optimal levels. While standard Kubernetes-native stacks provide strong container orchestration, they offer limited capabilities for AI workload placement, pipeline-aware scheduling, and fine-grained resource coordination across GPUs, networks, and storage.

## THE SOLUTION

Sustaining high utilization in large-scale GPU clusters requires a two-layer architectural approach.

First, CSPs must partition shared GPU infrastructure into isolated, tenant-specific sub-clusters that can be independently MLOps-provisioned and managed. This is addressed by the Zadara Infrastructure-as-a-Service (IaaS) layer. Protected scheduling domains are carved accurately out of the large-scale GPU cluster, sketching accurate tenancy boundaries through network, compute and storage isolation.

Second, in order to maximize utilization of GPUs and associated resources within each isolated sub-cluster, CSPs must enable efficient AI development and operations environments with intelligent orchestration, namespace isolation, scaling, and accurate usage-based billing. This is addressed by the ClearML Platform-as-a-Service (PaaS) layer.

The combination of ClearML's PaaS over Zadara IaaS provides unprecedented capabilities.

Integrated with constant communication via Zadara's EC2-compliant API, elastic scaling is enabled to ensure optimal LLMOps performance with a secure underlying infrastructure layer to ensure the end customers' digital sovereignty.

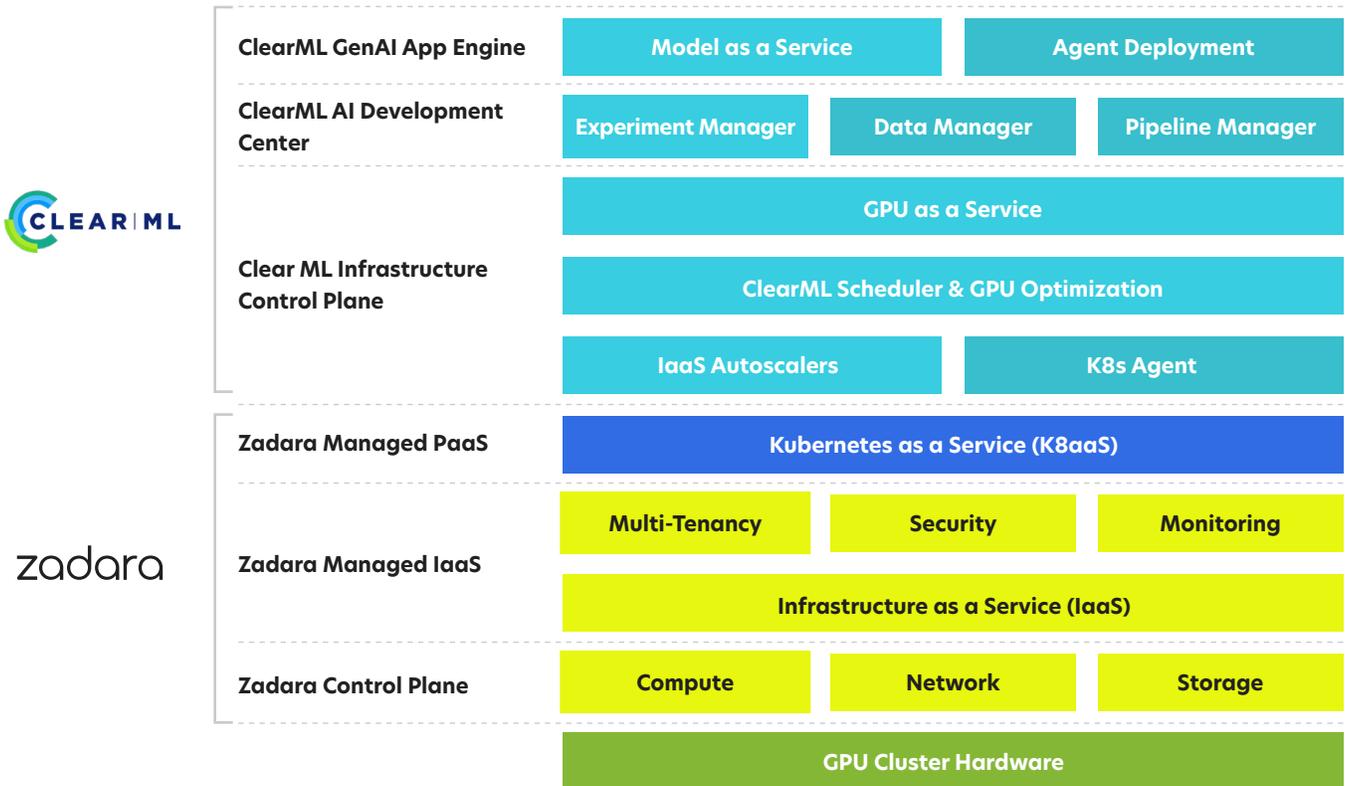


Fig. 1: A Full AI Cloud Stack by Zadara and ClearML

## ZADARA AND CLEARML PARTNERSHIP OVERVIEW

ClearML and Zadara are proud to announce a strategic partnership built specifically for Cloud Service Providers.

This unified solution allows CSPs to deploy sovereign, multi-tenant AI clouds with minimal operational overhead.

ClearML's AI and MLOps PaaS and Zadara's Sovereign AI IaaS combined, create an integrated solution for cloud service providers. ClearML provides secure GPU-as-a-Service orchestration, a multi-tenant AI workbench, LLM endpoints and GenAI application deployment, RAG and pipeline workflows, distributed training orchestration, and intelligent autoscaling.

Zadara contributes secure VM-level multi-tenancy, GPU virtualization, elastic compute and storage, and integrated management and billing.

Together, these platforms deliver secure GPUaaS with hardware-assisted, multi-tiered multi-tenancy across VM, Kubernetes, and application layers. The solution also provides infrastructure elasticity combined with AI workload automation, resulting in a white-labeled, fully operational AI Cloud that CSPs can resell immediately.

Through deep integration, ClearML's autoscalers communicate directly with Zadara's elastic infrastructure to deliver a seamless AI Cloud experience. The result is an automated, monetizable AI cloud that is achievable only through this combined solution. This partnership sets a new standard for sovereign AI performance, automation, security, and commercial viability.

AI/ML OPERATIONS	
ClearML PaaS	
<b>LLM deployment and GenAI application hosting</b>	
<b>Agent and RAG pipeline orchestration</b>	
<b>Cloud-like AI development environment</b>	IDE sessions, pipelines, model registry, datasets, experiment management

MULTI TENANCY			
ZADARA IaaS		ClearML PaaS	
<b>Infrastructure level Tenant VM isolation assisted by hardware</b>	Per-tenant cloud compute, storage and network allocation	<b>Platform level Secure dynamic multi-tenancy</b>	Kubernetes, and workload isolation
<b>GPU pass through allocation</b>		<b>Hardware Infrastructure abstraction</b>	Tenants don't have to use SSH or raw GPU nodes
		<b>Dynamic fractional GPUs and optimized GPU workload assignment</b>	
<b>Network VPC segmentation</b>			
<b>Per-tenant cloud storage allocation</b>			
<b>Tenancy control over high-performance low-latency AI storage</b>			

SECURITY	
ZADARA IaaS	ClearML PaaS
Access control management	Per-tenant Identity Provider (IdP)
Security Groups management and control	
Infrastructure and cloud security	

PROVISIONING	
ZADARA IaaS	ClearML PaaS
Single pane of glass cloud and storage control	Cluster-level MLOPs dashboards (GpU jobs, pipelines, models, endpoints, etc.)
EC2 compliant API - auto scaling support	EC2 API access for auto scaling
Disaster recovery	
Dynamic cloud provisioning and automation	

METERING & MONITORING	
ZADARA IaaS	ClearML PaaS
Hardware-level monitoring	Per-tenant and cross-tenant quotas, over-quotas, spillover, and access control
Metering tools that track usage consumption tracking; supports chargeback/usage-based billing models	Cluster-level dashboards (jobs, pipelines, models, endpoints, etc.)
Billing data export for or invoicing or customer billing purposes	Per-tenant usage metering and billing support at both the VM and workload level

Take the next step toward launching your own multi-tenant sovereign AI cloud, powered by ClearML and Zadara. Contact us to begin your deployment.



## ABOUT ClearML

As the leading infrastructure platform for unleashing AI in organizations worldwide, ClearML is used by more than 2,100 customers to manage GPU clusters and optimize utilization, streamline AI/ML workflows, and deploy GenAI models effortlessly. ClearML is an NVIDIA partner and is trusted by more than 300,000 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>.

Follow ClearML: [LinkedIn](#), [X](#), and [YouTube](#).

# zadara

## ABOUT ZADARA

Zadara (<https://www.zadara.com>) is a leading provider of sovereign AI edge clouds offering a revolutionary distributed cloud platform that simplifies operational complexity and enables seamless multi-tenancy through automated end-to-end provisioning of compute, storage, and networking. With over 500 edge cloud locations worldwide and with the ability to deploy a cloud in any location, Zadara's clouds are uniquely positioned to meet the unique demands of various cloud-based use cases, including sovereign cloud and AI inference at the edge for service providers and the modern enterprise. Zadara operates worldwide with a highly skilled team that provides 24/7 follow-the-sun support and services. Zadara is headquartered in Irvine, California.

Follow Zadara: [LinkedIn](#), [X](#), and [YouTube](#).

# zadara

Enterprise Edge Cloud Services Provider.  
Any data type. Any protocol. Any location.

Contact us at:  
[www.zadara.com](http://www.zadara.com)  
[info@zadara.com](mailto:info@zadara.com)