HAMi

Open source, dynamic AI GPU sharing, scheduling and orchestration that accelerates AI throughput, delivers seamless auto scaling, and maximizes GPU utilization.

10x	10x	50%	0
GPU Availability	Workloads Running	GPU Utilization	Manual Intervention

WHY HAMi?

01 Maximize GPU Utilization

GPU on-demand sharing, scheduling optimization, and priority strategies eliminate waste, reduce costs, drive efficiency, and accelerate AI initiatives.

02 Seamless Auto Scaling

GPU on-demand auto-scaling: Seamless VPA scaling for AI workloads without restarts during GPU consumption surges.

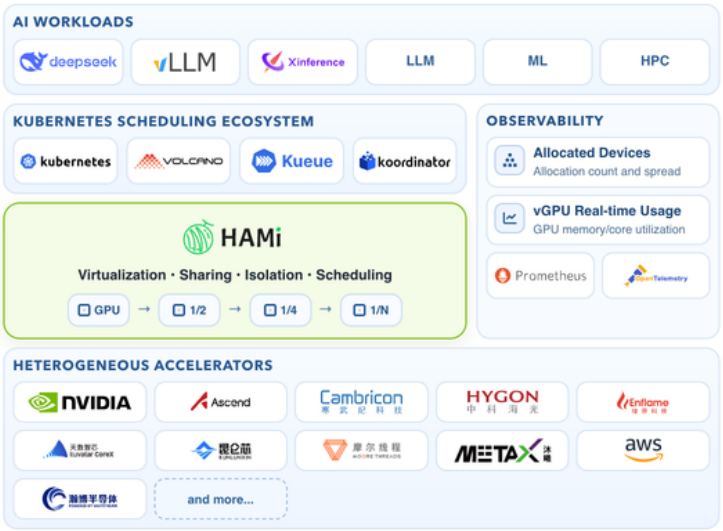
03 Unified Memory Out of the Box

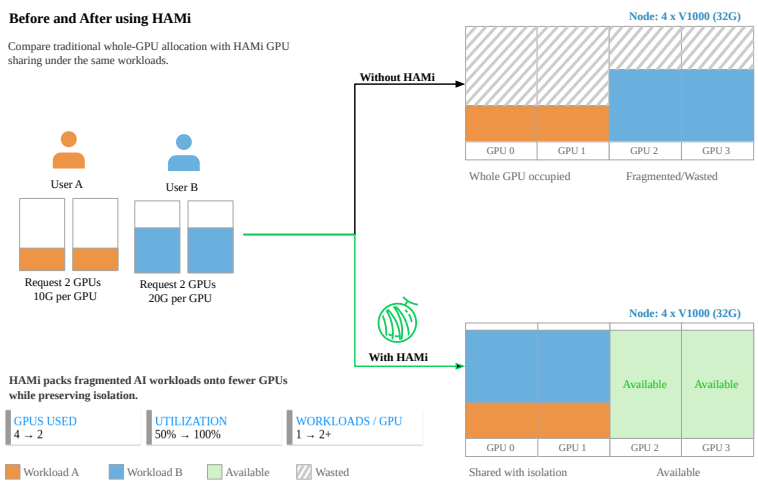
Use HAMi to integrate NVIDIA unified memory in Kubernetes and enable GPU memory oversubscription.

04 Flexible Integration

The open architecture integrates seamlessly with any toolchain (e.g., Kueue, Koordinator, Volcano) and infrastructure across public clouds, private clouds, or on-premises data centers.

ARCHITECTURE





Learn how HAMi improves GPU utilization in Kubernetes → https://project-hami.io

USE CASES

OCR and Translation

Scientific research (e.g., PyTorch, TensorFlow) with co-located, fragmented algorithms sharing GPUs

Colleges and Labs

Multiple small/medium models share GPUs seamlessly and flexibly to maximize utilization.

Hybrid Cloud AI Workloads

A lightweight, non-intrusive, centralized solution is needed to maximize GPU utilization across public clouds, private clouds, and on-premises data centers.

AI Training and Inference Platform

Colocation of training and inference tasks, flexible GPU sharing,seamless auto scale, over-allocation of GPU memory.

Cost Effective GPU Service for Startups

GPU sharing, oversubscription, and task priority enable flexible sales and leasing models, reduce costs.

Finance

Model fine-tuning, GPU virtualization, and precise scheduling.

ADOPTERS

AND MORE...

GROWING GLOBAL COMMUNITY

500+	300+	4K	1.4K+	323K
Contributors	Adopters	Stars	Commits	Image Downloads

https://project-hami.io

https://github.com/project-hami/hami

GitHub

Discord