Orkestrator

Proven software built specifically to manage GPU clusters

EAI Orkestrator is a best-in-class tool which optimizes the use of compute and storage resources and bridges the gap between the needs of your AI practitioners and your compute infrastructure.

Optimize your GPU cluster

Actively manages the allocation of your centralized compute resources to get the best return on your investment.

Maximize your specialized resources

Lets your AI practitioners do what they do best - building quality models - by removing most of the engineering heavy-lifting required to efficiently use GPUs.

Increase overall productivity

Dynamically rebalances workloads to allow AI practitioners to process a higher volume of jobs and become more efficient in producing AI models

Future-proof IT infrastructure

Enables your IT infrastructure to meet your current and future ambitions and scale with minimal disruption.

Why EAI Orkestrator?

- Designed to work with GPU-based workloads including AI, simulated and virtual environments
- More efficient job distribution and increased fairness between users
- Better visibility into resource usage via management tools, enabling user proactivity
- Easier reproduction of jobs defined as containers
- Better user experience with dashboard view, customization of queuing, transparency and visibility on job allocation
- Allows AI practitioners to submit compute-heavy jobs, such as hyper parameter optimization, with minimum effort and friction and on a very large scale
- Docker images allow for a smoother path from research to production

CONTACT US sales@elementai.com

Benefit from more efficient job distribution

Fair share between users and more optimal and transparent resource usage

	Competitors	Orkestrator	Benefits
ARCHITECTURE	Most operate on programs	Operates on containers (self-contained)	Jobs defined as containers can be reproduced more easily More flexibility in terms of which technology/libraries are used (full control over what's in the container)
	Most run on bare-metal cluster	Runs on Kubernetes	Enables the use of a single cluster for both research workloads and production systems Docker enables a smoother path from research to production
QUEUE TYPE	Based on policy or other (sometimes) configurable conditions	Fair share between users, job type can be customized (preemptable, non-preemptable)	More efficient job distribution and increased fairness between users Users can set the priority level of their jobs
MANAGEMENT TOOLS	Limited or no built-in reporting	Reports include: job resource usage, efficiency report, user reports and maintenance dashboards.	Better visibility and control leads to better user experience and increased productivity Management and maintenance of clusters in real-time, including automatic healthchecks of nodes
SUPPORT	Limited to no support offered	Full support offered	User training offered and deployment support

</> Technical overview

- Designed for GPU-based workloads including AI, simulation and virtual environments
- Built on top of Kubernetes and Docker
- Can run on-premises or in the cloud
- REST API for flexible job control
- CLI to streamline job control
- Job resource requirements
  - CPU, RAM, GPU, Tensor Cores, GPU RAM, CUDA version, and more
  - Special case: X11 server sidecar (CPU or GPU)
- Multiple job types: preemptable, non-preemptable and "interactive"
- Meta job: Process agents launch jobs on behalf of users according to configurable settings
- Dashboards including overall cluster view, GPU status, efficiency report, per job view (CPU and GPU)

Case study

How Element AI uses the Orkestrator on premises to optimize resource usage

Element AI fundamental researchers, applied research scientists and developers require solid infrastructure to efficiently run their models. As the organization scaled and GPU-based workloads became increasingly heavy, distribution and management of computing resources became a challenge.

BEFORE EAI ORKESTRATOR

By tracking allocation manually (spreadsheets), Element AI averaged GPU usage of 25%

AFTER EAI ORKESTRATOR

With the tool, Element AI averages GPU usage at > 90%, and has increased GPU volume 36x without requiring additional staff

CONTACT US sales@elementai.com