Lambda

Managed Kubernetes for 1-Click Clusters

Lambda's industry-leading 1-Click Clusters allow you to spin up a full, NVIDIA Quantum-2 InfiniBand-linked training cluster in whatever size you need and for however long you need it - from weeks to months. Our Managed Kubernetes offering lets you ramp up on using those clusters on day one - giving you familiar, managed tooling and the ability to quickly change and deploy whatever you like without worrying about low-level administration.

GET GOING ON DAY ONE

Lambda's world-class engineering team will deploy, run and operate our reference Kubernetes architecture designed for AI, machine learning and GPU workloads. This includes:

- Kubernetes installation and upgrades
- Control plane maintenance and high-availability
- NVIDIA Kubernetes operators and stack installed and configured
- Detecting node failures, cordoning nodes, and expediting node repair
- Gathering workload and cluster metrics and proactive monitoring

In return for allowing us to take low-level control of the cluster, you will get a fully-managed experience, with access to the Kubernetes API, graphical dashboard, and other components pre-configured:

- Storage drivers and classes, including to our scalable shared storage
- Workload monitoring and metrics
- Flexible pod networking including NVIDIA Quantum-2 InfiniBand
- Optional installs of Kubeflow, Ray, Volcano, and various other high-level schedulers and interfaces

Our goal is to get you up and running day one, and with the ability to quickly set up your environment, get working, and transition easily between multiple clusters or different cluster sizes thanks to a common Kubernetes platform.

PERFORMANCE & FLEXIBILITY

1-Click Clusters run on our industry-leading, dynamically segmented, Infiniband-linked clusters, with your primary workloads running on dedicated, single-tenant GPU machines. Clusters can be booked in intervals ranging from one week through to one year.

We perform any required critical upgrades to cluster components, and we proactively monitor your cluster 24/7/365 and attempt to fix issues as soon as possible - resolution times vary by issue type and severity. All clusters are located in datacenters where Lambda have 8×5 continuous presence and 24×7 on-call availability.

Questions? Call **(866) 711-2025**, email **enterprise@lambdalabs.com**, or visit **lambdalabs.com**

©2024 Lambda. Updated October 29, 2024