Managed Kubernetes for Superclusters

Lambda's fully-managed Kubernetes offering builds on the power of our Superclusters. We install and maintain a Kubernetes installation on your cluster, run most daily operational tasks, and handle cluster and hardware failures - allowing you and your team to focus on using the cluster rather than operating it.

LEARN MORE

Operational Ease /

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, node pool adjustment and failed hardware replacement

+ Gathering chassis and cluster metrics and proactive monitoring

In return for allowing us to take full control of the hardware, 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 optional high-end storage appliances

+ 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 solutions and operations engineers will work with you before and during handoff to ensure the cluster is tailored to your specific type of workload, and you can get up and running as soon as possible.

Single-Tenant Security /

Lambda Superclusters are single-tenant and all Kubernetes and fleet management components will remain local. Nothing inside of the cluster will be exposed to the Internet unless you choose to - by default, all access is over a secure VPN, including to the Kubernetes API and dashboard. Lambda maintains access in order to run monitoring, alerting, and perform proactive fixes when we notice issues.

Single-tenancy also means you get flexibility as to when we do upgrades and to which versions; we'll work with you to establish maintenance windows and your preferred upgrade cadence.

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 8x5 continuous presence and 24x7 on-call availability.