

Creating Beyond Limits:

How Pika Powers AI Video Generation
with Lambda

lambda flexible cloud infrastructure |



Challenge

+ Pika's Mission to Unlock People's Creativity

Pika, a fast-growing AI startup, is on a mission to democratize video creation by lowering barriers to entry and expanding creative opportunities for anyone with an imagination. While their goal is ambitious, the impact they hope to make is to bring more small, everyday creative ideas to life.

Pika empowers creatives to transform their ideas into videos that push the bounds of storytelling, with captivating realism and special effects.

Whether Pika's users want to create a funny video to share with friends, extend a family video, or make promotional content for their small business, they have the creative freedom to instantly turn their ideas into a reality. Pika's easy-to-use Pika 1.5 platform uses advanced generative AI technology to make it so that artists no longer need technical skills to make videos, democratizing access and allowing people to have fun.

"We are a very technologically advanced and focused company. We have the most brilliant minds in the world working on our models. What we're trying to do with it now is move AI video creation to a place that is much more fun, much more accessible, and much less serious".

-Lindsay Brillson, Pika's Head of Brand and Content

+ Scalability and Agility Imperative

Scalability and Agility Imperative

Operating in an AI market where major innovations and breakthroughs are often a daily occurrence, Pika needs infrastructure capable of supporting the resource-intensive nature of training large-scale deep learning models and running real-time inference with them. Pika works with several cloud providers to get the computational power that they need, but they were looking for an additional partner to help

them fill gaps beyond their multi-year contracts. Ashley McEnery, Pika's Head of Finance and Operations, said they were hesitant to sign multi-year contracts given the constantly changing nature of the industry. Uptime and responsiveness were also non-negotiable, necessitating a partner that could guarantee high uptime and provide responsive support for Pika to maintain their pace of innovation and the quality of the experience they provided.

"As a startup we prioritize optionality, and that was hard to find in some of the other players in the market,"

-McEnery said

Solution

+ Lambda's Elastic, AI-Optimized Infrastructure

Pika turned to Lambda for their service and flexibility delivered through Lambda's Reserved Cluster, leveraging NVIDIA H100 Tensor Core GPUs for high performance compute. Lambda's infrastructure is optimized for speed and performance with NVIDIA Quantum-2 InfiniBand for low-latency and high-bandwidth networking and software integration for automated resource allocation and self-managed node operations (including node reboots and health checks). Unlike general-purpose cloud providers, Lambda's infrastructure is specifically designed for AI and machine learning workloads.

+ Customized Support and Services

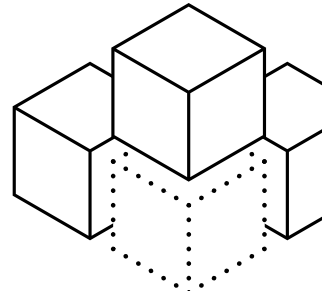
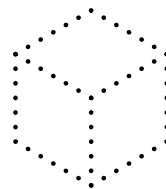
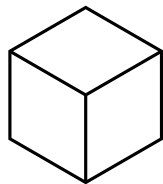
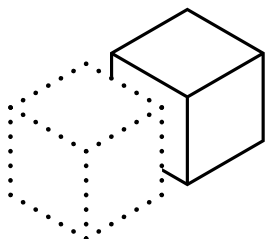
Lambda is committed to providing exceptional customer support and prioritizing customers like Pika with a more hands-on, tailored approach that meets them where they are. For example, Lambda provided Pika with real-time Slack support—Pika's preferred communication tool—for quick issue resolution and direct access to an AI-specialized support team with extensive expertise in deep learning and GPU optimization. This enabled Lambda's AI engineers to quickly get familiar with and resolve Pika's distinct use cases without the need to escalate through different support tiers, allowing Pika to focus on innovation rather than downtime and service outages.

+ Adjustable Engagement

Flexibility is crucial for a fast-moving startups like Pika: it enables them to mitigate risks like overprovisioning expensive resources. Lambda's shorter-term contracts were a key factor in Pika's decision, offering the freedom to adjust their usage as needed without being locked into a lengthy contract.

"The two really big things that have been wonderful for us are flexibility and optionality, and the way that that has been so much better set up from Lambda for an industry that is fast moving and very young and needs to adjust constantly and can't be locked into something because we don't know what's going to happen in the next three days, much less three years. That's a huge differentiation point for Lambda."

-Brillion said



Results

+ Focus on Core Business and Innovation

With great service and reliable infrastructure, Lambda and Pika accelerated AI video model techniques, reducing training times and enabling rapid iteration of new models for real-time applications.

"User count means that there is more inference demand. More inference demand means that compute demand will increase. That's just simple logic. As GPU clusters grow in size, managing the dynamic nature of user demand and efficiently scaling resources becomes increasingly complex. Support is crucial to deliver a great user experience, and this is what Lambda is good at."

-founding engineer Karli Chen said.

Improved performance enhanced Pika's product and attracted more users, thereby also increasing the need for additional AI compute. Fortunately, as Pika continues to grow, Lambda's expert support and scalable infrastructure will ensure their end customers get an outstanding experience, even in times of high demand.

"We have customers that want to make amazing, cool, stupid, funny things—and they want to make them now. They don't care what's going on in the background, or what's happening with the cluster; they just want their videos fast, and we want to be able to give that to them."

-Brillson said.

Try Pika [here](#)

Learn more about Lambda Private Cloud [here](#)

