

HYP\_Link

## Serverless Solutions Seek to Relieve Large Scale Computing Needs and Cost Demands in the Cloud

Tom Sorensen and Earl Joseph  
December 2024

### RECENT DEVELOPMENT

---

In an [announcement](#) made in late November, 2024, recent US-based start-up Kinesis Network announced the launch of their serverless feature targeted to enable users to fluidly run workloads in a multi-cloud environment. By leveraging underutilized GPU capacity in cloud environments, Kinesis hopes to help organizations reduce infrastructure costs by up to 90% while providing access to high-performance GPUs for AI and data-intensive tasks. Founded by industry veterans from AWS, IBM, and Microsoft, Kinesis is gaining attention as AI and HPC/AI integrated organizations grapple with the high costs of cloud computing, increased demands for advanced computing power, and a continuing need for flexible, scalable access to processing capability.

Accessing and leveraging underutilized GPU resources across cloud providers not only alleviates scaling challenges for users but also can better utilize costly hardware while maintaining the benefits of public cloud use. "The compute access shortage is a critical bottleneck for AI innovation and scientific research," said the company's COO, Hamza AK. "Our platform solves this growing issue by unlocking a vast, untapped reservoir of global computing power, making it accessible to any enterprise or organization."

### ANALYST COMMENT

---

As AI integration processes have swept through the worlds of research, academia, and enterprise, many groups are scaling these efforts up as they reach production levels. Thus far, cloud resources have been critical for many organizations in exploring, testing, integrating, and running their models to full potential. Hyperion Research studies indicate that for most users, training, fine-tuning, retraining, and the integration of new data is an ongoing, GPU-intensive process. Furthermore, inference demands are on the rise across nearly all verticals of advanced computing users. With hardware costs at an all-time high and availability low, the accessibility and flexibility of public cloud offerings has, for many, provided the heavy lifting for efforts to integrate and maintain AI and AI-supported applications. Kinesis is [not the only group](#) developing frameworks to squeeze more efficiency out of existing resources and relieve cost and demand pressures. Many organizations consider management of cloud costs as a priority, and any solutions to this need will be a valuable asset.

---

### Copyright Notice

Copyright 2024 Hyperion Research LLC. Reproduction is forbidden unless authorized. All rights reserved. Visit [www.HyperionResearch.com](http://www.HyperionResearch.com) to learn more. Please contact 612.812.5798 and/or email [info@hyperionres.com](mailto:info@hyperionres.com) for information on reprints, additional copies, web rights, or quoting permission.