

AI Retail Analytics Platform Scaled Affordably

The company improved GPU performance and lowered data transfer costs on Akamai Cloud



Reduced overall costs 30%–40%



Doubled AI processing speed



Lowered data transfer costs 10x–15x

Powering smarter retail through real-time AI insights

This AI-powered platform helps retailers turn CCTV video into actionable insights that improve store performance and boost sales. As the company scaled to thousands of stores globally, its GPU-intensive workloads and massive data transfer requirements grew. This drove up cloud costs and strained efficiency. By migrating to Akamai Cloud, it improved performance and reduced infrastructure costs. It also created a scalable foundation for next-generation AI innovation.

Scaling video analytics across thousands of stores

This company's growth has been rapid. What began with a handful of retail brands expanded into a global platform supporting more than 7,000 stores and more than 200 clients.

Its CTO explained that the platform processes enormous volumes of video data. "We capture about 40 terabytes of data per day from stores across the world. After processing that using GPU-based video analytics, we deliver insights to retailers through our SaaS platform."

Its platform combines GPU-powered computer vision with CPU-based processing layers to extract metrics like footfall, dwell time, and customer journeys. These insights help retailers identify operational gaps and improve conversion rates.

Rising costs and limits with legacy cloud infrastructure

As platform usage grew, so did infrastructure demands. The platform provider initially ran nearly all workloads on one of the hyperscalers. While it was stable, cost pressures became unsustainable — especially for a data-intensive, GPU-heavy environment.

"The transfer costs became prohibitive," the CTO said.

Higher costs directly impacted the company's ability to grow. "Retailers are extremely cost conscious. We lost some customers because we couldn't reduce our subscription prices," he continued.



AI Retail Analytics Company

Location

Chennai, India

Industry

Software and SaaS

Solution

Cloud Computing

Products

Block Storage

CPU

GPU

LKE: Managed Kubernetes Engine

Managed Databases

NodeBalancers

Object Storage

Private Networking



The challenge wasn't just storage or compute. It was the combination of high requirements for GPUs, massive daily data ingestion, and continuous video streaming and downloads.

Evaluating alternatives and validating performance

The company evaluated multiple cloud providers but found gaps in cost efficiency, performance, or reliability. The turning point came during a proof of concept (PoC) with Akamai, comparing performance against the existing service provider using equivalent GPUs.

The results were decisive: 2x–3x better GPU performance, 10x–15x lower data transfer costs in some cases, and the ability to process the same workloads on fewer servers. “We could run double or triple the workload on a single server compared to our existing service provider,” the CTO said.

Migrating AI workloads to a more efficient platform

Impressed by the strong PoC results, the platform provider began migrating workloads to Akamai, starting with data ingestion, GPU inference, and storage pipelines. “Migrating our Dockerized environment has been very straightforward,” said the CTO.

Key platform workloads migrated included:

- Real-time CCTV video ingestion from retail stores
- GPU-based computer vision for customer behavior analysis
- Event processing and metadata enrichment pipelines
- Analytics dashboards powered by structured event data
- Operational AI for compliance, staffing, and merchandising

“We are migrating gradually, validating performance and consistency as we go,” he continued.

Leveraging Akamai Cloud for AI at scale

The company deployed a range of Akamai Cloud solutions to support its platform:

- **Compute Instance (CPU):** 100+ high-performance virtual machines featuring optimized networking dedicated to real-time video analytics and AI workloads
- **GPU:** 60+ cost-efficient, high-performance NVIDIA RTX™ 4000 Ada GPUs for real-time and batch inference
- **LKE: Managed Kubernetes Engine:** 30 autoscaling groups to support efficient resource utilization for dynamic workloads
- **Block Storage** and **Object Storage:** Reduces storage and data transfer costs while supporting large-scale video pipelines
- **NodeBalancers:** Manages traffic spikes and ensures reliable performance during peak usage
- **Managed Databases:** Supports scalable analytics and dashboarding infrastructure
- **Private Networking:** Lowers costs for high-volume internal data transfer



**Akamai delivered
2x–3x better
performance
while lowering our
costs up to 40%,
changing how we
scale AI.**

CTO

Combined, these capabilities allowed the company to optimize both performance and cost simultaneously. “Akamai does not experience any slowness when downloading data and streaming video,” the CTO explained.

Improving performance while reducing infrastructure footprint

Akamai’s performance gains translated directly into operational efficiency. In addition to reducing costs, the platform provider is reducing the number of servers it needs. According to the CTO, “We can process the same volume of data and workloads with fewer resources.”

This efficiency is especially critical for real-time analytics, where speed directly impacts customer value. “Better performance improves our real-time analytics,” he added.

Driving growth through lower costs and better scalability

By combining higher performance with lower costs, Akamai has become a strategic platform for the company’s growth. With projected cost savings of 30%–40%, the platform provider expects to improve both margins and market competitiveness.

“Akamai gave us better performance and significantly reduced our costs,” the CTO said. “That means we can lower our subscription pricing, onboard more customers, and drive higher revenue.”

Lower infrastructure costs enable the enterprise to invest in next-generation AI capabilities. Moving forward, the company is expanding beyond video analytics to deploy large language model (LLM) retail insight tools on Akamai Cloud.

“One of our next big projects is building an LLM specifically for retailers,” he said. “We plan to deploy that on Akamai.”

This model will allow retailers to query store performance, customer behavior, and operational metrics through natural language, powered by the company’s deep data insights. “Akamai infrastructure will power our next phase of growth,” he concluded.