emma × AUGUR

User Story

When GPU Demand Outpaces Existing Cloud: How Augur Secured Scalable AI Infrastructure

Overview

AI companies move fast. But building AI infrastructure can quickly become a bottleneck when access to the right GPUs is limited, costs get out of hand, and engineering teams spend more time managing infrastructure than building models. Augur found itself in this exact situation.

Augur is a UK sovereign AI company designed to prevent threats to national security, public spaces and critical national infrastructure. The platform transforms existing sensor infrastructure into real-time intelligence networks that identify threats, enable rapid response and accelerate post-event investigation. By enabling sensors to understand space in 3D, Augur runs advanced ML and AI capabilities to detect perimeter breaches, track hostile reconnaissance and coordinate incident response.

As such, their advanced machine learning systems depend on powerful GPU infrastructure for both training and inference workloads. As demand for compute grew, the team needed reliable access to high-performance accelerators without introducing operational complexity or runaway costs.

Challenge

GPU Availability, Rising Costs, and Limited Visibility

At the time, Augur was running workloads on AWS using L40S GPUs. While these instances supported their initial infrastructure needs, the team soon faced two growing challenges.

First, they needed access to RTX Pro accelerators, which were not available in the AWS region they were operating in. Due to data privacy and regulatory concerns, Augur could not move workloads to regions outside of the UK. As their workloads evolved, the lack of GPU variety limited their ability to scale and experiment with different compute options.

01 Customer pain

- GPU access constraints
- Vendor dependency risk
- Limited cloud flexibility
- AI workload scalability challenges
- Cost pressure on training & inference

02 What they got from us

- Immediate GPU access with cloud credits
- Cost-optimized GPU infrastructure
- Multi-cloud deployment flexibility
- Single control plane for AI workloads
- Centralized orchestration layer
- Reduced vendor lock-in

03 Technical value

- AI/ML training & inference infrastructure
- Production-grade GPU compute
- Cross-cloud workload portability
- Scalable accelerator access

04 Strategic gains

- AI infrastructure freedom
- Future-proof multi-cloud strategy
- AI enablement
- Infrastructure optionality
- Operational cost control

Second, the team lacked clear visibility into how GPU resources were being used. The ML team was burning through AWS credits, with 50,000 credits remaining from an initial 100,000. The lack of visibility meant development instances were frequently left running, creating unnecessary costs and making it difficult to forecast budgets and control spending.

The operational overhead of managing GPU infrastructure was becoming significant and would have grown even further if Augur had added the additional instances they required.

The Journey

From Infrastructure Constraints to Operational Freedom and Control

At this point, Augur discovered emma through its Cloud Credits program, which provides AI teams with immediate access to high-demand GPUs and infrastructure credits across multiple providers. The program allowed Augur to test alternative GPU infrastructure without the financial risk of committing to a new environment upfront.

Augur began conversations with the emma team to explore how a more flexible, centralized approach to GPU infrastructure could address their challenges.

Instead of being tied to a single cloud provider, Augur gained access to GPU infrastructure across multiple providers and environments through emma. Crucially, this enabled immediate availability of emma's own NVIDIA H200 accelerators within the EU, which got the job done for Augur.

During the proof-of-concept phase, Augur tested inference workloads on the newly available GPUs. The tests validated both performance and stability, confirming that the infrastructure could support their workloads in production.

At the same time, the emma cloud operations platform provided something the team had previously lacked: **clear visibility and control over GPU usage.** Idle resources could be identified instantly, and AI-powered actionable recommendations ensured development environments didn't continue running unnecessarily.

> "It has been absolutely key for us to have access to this hardware in such a low friction way. The team has done model optimization, tuning and experimentation in a way that wouldn't be as possible, or nearly as fast, if we were worried about unbounded cost"

Imran Lone
Co-Founder & CTO, Augur.

The Result

Infrastructure Freedom for AI Workloads

With emma, Augur gained access to a more flexible infrastructure model suitable for demanding, dynamic AI workloads.

Rather than relying on a single provider, they could now deploy GPU workloads across multiple providers through a unified control plane. Within the platform, the team can define specific requirements, such as preferred GPU types, location or provider constraints, and emma automatically surfaces all available infrastructure options that meet those criteria, ranked by cost. This makes it easy to compare alternatives and choose the most suitable setup depending on workload needs, performance requirements, regulatory compliance, and cost considerations.

Once selected, deployments can be launched directly from the platform without needing to manage cloud-specific configurations or infrastructure differences across providers. The platform also introduced centralized orchestration and resource visibility across all infrastructure environments, allowing the team to manage training and inference infrastructure more efficiently.

Now Augur plans to run their infrastructure through the emma platform long term, making it part of their operational foundation.

Looking Ahead

For Augur, the partnership with emma provides more than just access to GPUs. It enables a future-proof infrastructure strategy where AI workloads can scale without being constrained by provider limitations or operational complexity.

By combining multi-cloud flexibility, centralized orchestration, and scalable GPU infrastructure, Augur can focus on advancing its AI systems while maintaining control over performance, cost, and infrastructure strategy.

As Augur continues to grow, we’re hopeful emma will ensure their infrastructure evolves alongside their AI ambitions.

"emma is already part of how we develop, and as we scale our training efforts, we see it becoming increasingly central to how we operate”

Imran Lone
Co-Founder & CTO, Augur.

About emma

Founded in 2021 in Luxembourg, emma is a cloud operations platform designed to help organizations run data-intensive workloads efficiently across hybrid and multi-cloud environments. It enables distributed cloud operations through a single, policy-driven operating model.

emma provides a unified operating layer spanning hyperscalers, regional European cloud providers, AI-optimized infrastructure, and on-prem environments — enabling centralized deployment, governance, and cost control from one interface. By combining comprehensive multi-cloud capabilities with AI-driven automation and real-time analytics, emma gives teams full visibility and actionable control over their cloud operations.

With emma, organizations can simplify complexity, reduce vendor lock-in, and meet sovereignty and regulatory requirements while operating at scale.

emma is a Series A company backed by Smartfin, RTP Global, CircleRock Capital, and AltaIR Capital.

LEARN MORE AT EMMA.MS

Contact us

info@emma.ms

Luxembourg,

19-21 Rte d'Arlon 8008 Strassen,
Luxembourg

United States,

8605 Santa Monica Blvd West
Hollywood, CA 90069