PLATFORM9

Case Study

AI startup Norna keeps fashion retailers ahead of the competition with Kubernetes-as-a-service

Norna achieves 78% TCO reduction and 10x productivity gain for K8s AI workloads with Platform9 Kubernetes-as-a-Service

Summary:

- Norna freed up resources to focus on their mission of becoming the world leader in applied AI instead of having to spend valuable engineering cycles on Kubernetes platform operations "We are AI specialists, we cannot have in-house talent spending time becoming production-K8s experts." — Jonas Saric, Norna CEO/Founder
- Norna achieved a **10x** productivity improvement in Kubernetes management by leveraging a managed service model through Platform9's Kubernetes as a Service
- Norna reduced overall the total cost of operations (TCO) for Kubernetes by **78%** through a combination of workload repatriation from the public cloud (saving **50%** on cloud costs) and operations productivity gains by leveraging Platform9's hyperscale automation
- Norna improved customer service, reaching **99.98%** uptime for its Kubernetes clusters, with **90%** reduction in unplanned downtime
- "We wouldn't have survived if we hadn't transitioned Kubernetes operations to Platform9's hyperscale service" — Jonas Saric, Founder & CEO

NORNA

Background: Norna helps retailers compete with data-driven giants

Norna's unique AI-driven service helps fashion retailers with assortment planning and pricing through near real-time insights into changes in competitor pricing and offerings.

This powerful dataset allows retailers to understand, for example, the market price for a black cotton t-shirt of a certain quality and neckline in a given geography, which would be very different from the same shirt in cashmere. Retailers can adjust their prices or change items stocked based on these insights.

Norna captures data at a SKU and geographic market level, scraping millions of webpages each week, and applies ML techniques to create a standardized dataset with over 1,700 item attributes.

The data helps fashion retailers to compete on an equal footing with much larger data-driven retailers like Amazon, Europe's Zalando, and China's Shein.

Norna's use of AI in retail has been recognized many times for being both innovative and practical.

Situation: Rapid service growth

To deliver this service requires an extensive technology stack. Norna ingests over 10TB of raw data a month with the volume growing monthly. Norna also needs to process all this data using a proprietary AI approach that uses Kubeflow,

Tensorflow, and Pytorch. Norna's AI approach includes two types of artificial neural networks, LSTMs (long short-term memory) and CNNs (convolutional neural networks) for image analysis. Norna picked Kubernetes for its parallelism (which was a good fit for its ML needs) and resiliency (helpful when working with web data where failover and retry capabilities are needed). A typical Norna cluster configuration would involve 10 nodes (with a three master control plane) supporting 600+ pods. Norna's tech stack also includes continuous integration tools and Terraform automation.

To support the overall demands, Norna uses high-performance servers to provide the massive demands of its scraping process and GPU-based servers for model training.

Norna has been named Startup Partner of the Year by IBM, one of the Top 10 Artificial Intelligence Solution Providers in Europe 2020 by Outlook Magazine, and One of Europe's Most Interesting AI Companies by The European Innovation System.

Challenges: escalating costs, team workload, and uptime pressure

In its startup phase, Norna worked with a public cloud platform, like many startups, and found their service exceptional. As Norna's service grew, however, Norna ran into several challenges:

Rapidly escalating costs for their public cloud based infrastructure

High demands on the team's time to manage their production Kubernetes infrastructure with the high uptime that was critical to revenue growth

To address the cost of infrastructure, Norna leveraged a combination of their own lower-cost hardware as well as moving to a regional co-location service. The co-location service used the Platform9 Managed Kubernetes Freedom plan to partially manage Norna's K8s deployment. While the approach provided many capabilities out-of-the-box and local support, as Norna grew the burden of maintaining a larger production-grade infrastructure was becoming unsustainable.

During their early phase of trying to manage Kubernetes on their co-lo infrastructure, Norna's CTO Ying Liu could spend 100-hour weeks trying to patch things together, "with rubber bands and glue," as Liu put it. "We are AI and product specialists, we cannot have in-house talent spending time becoming production-Kubernetes experts," says Norna CEO/Founder Jonas Saric. It was time to make the shift from the Freedom Plan to a solution that provided more support.

"We are AI specialists, we cannot have in-house talent spending time becoming production-K8s experts." — Jonas Saric, Norna CEO/ Founder

Solution: Kubernetes-as-a-service

After quite a bit of research, as well as a recommendation from their co-lo service, they chose a managed Kubernetes-as-a-Service (KaaS) in the form of Platform9's Growth Plan. "Platform9 clearly suited us best, given its flexibility and our scale of operations." says Liu. The key decision factors included ease-of-use, strong support, and a fully managed service model which provided the safety and coverage that Norna needed to run their business.

Platform9 supports its managed KaaS entirely with certified Kubernetes administrators (CKAs) and provides the ease of public clouds with all the control of an in-house deployment. For Norna, it was the simplest and fastest path to running their production, cloud-native data harvesting and processing applications. They could quickly deploy K8s clusters with a rich set of pre-built, cloud-native services and infrastructure plug-ins.

"We had to move closer to the source of know-how, and that's why we went with the Platform9 managed service," according to Saric. "And the fee structure with Platform9 enabled cost-effective scaling."

"After considerable research, Platform9 suited us best, given its flexibility and our scale of operations." — Ying Liu, Norna CTO

The Results: reduced TCO, improved uptime, and peace of mind

Norna only makes money when its AI applications are running, and since they've signed on with Platform9 managed KaaS, their clusters are now operating 99.98% of the time.

Norna estimates their TCO has reduced by 78% through a 50% reduction in their cloud costs (which could grow to 75% as Norna steadily increases its server count, and time savings from the reduced burden of managing Kubernetes.

"Platform9 removed the production Kubernetes bottlenecks, and their support is outstanding. They are true partners," says Saric. The team feels like Kubernetes expertise is always available, so they're confident their production environment will be stable. This confidence is essential, because their core business, developing AI apps and customer implementations, is so demanding.

As Liu sees it, "Now, instead of going crazy for a week trying to make something work, I can talk to someone who actually knows what's going on and fix it in hours." Which means that the AI technology that is Norna's competitive edge and core to the company's mission is getting his full attention.

Book a demo at platform9.com/contact