

CUSTOMER STORY

How Palo Alto Networks Takes Control of Its High-Stakes Cloud

Results

$ 3.5 M
Pre-discount savings from Sedai

90K
Autonomous actions in production

Zero
Incidents caused by Sedai's actions

Key Integrations

GOOGLE KUBERNETES ENGINE (GKE)

GOOGLE CLOUD DATAFLOW

AWS LAMBDA

AWS EKS

Palo Alto Networks needs no introduction. Its customers include nine of the Fortune 10, eight of the ten biggest US banks, and all ten of the world's largest utilities. Simply put, the company protects our planet's most critical infrastructure from cyber attacks.

Suresh Sangiah is responsible for keeping its cloud online — no matter what.

"Palo Alto Networks is the preeminent cyber security company in the world," he said. "We have the responsibility to make sure that our cloud is always on."

Because if ever our service goes down, it would mean the companies that we serve are not able to get their work done.

As Senior Vice President of Engineering, Suresh oversees Prisma SASE, the industry-standard cloud service that provides networking and security to distributed workforces around the globe. Under the hood, Prisma SASE runs on a huge array of its own cloud-based resources. Each of these resources requires careful management and constant optimization.

The result is a job that sounds impossible. Somehow, Suresh needs to control a cloud that is hosted by third-party vendors, configured by dozens of engineering teams, and stressed by millions of end users. And of course, he also faces pressure to lower costs — by eliminating every source of wasted spend.

Ultimately, the Palo Alto Networks team went looking for a new way to handle the complexity and costs of its cloud.



Suresh Sangiah delivering the keynote presentation at autocon 25

A Better Approach to the Cloud

This is the part of the case study where we tell you that Palo Alto Networks got Sedai — which solved every challenge with its cloud operations, overnight.

But that's not the truth.

The reality of enterprise software is much more nuanced, especially when you're dealing with such complex environments. Automated tools fail precisely because they ignore the nuance of the real world. These automated tools use fixed rules and scripts, without understanding context or adapting to change, which means things inevitably break.

The Palo Alto Networks team evaluated the leading solutions to manage its cloud, including automated tools like Cast AI and Kubecost. It decided to partner with Sedai, the first autonomous cloud management platform. The team found that only Sedai could "enhance production safety" at Palo Alto Networks, through "continuous learning."

Sedai didn't start making autonomous changes to the company's cloud, right away. Instead, our patented deep reinforcement learning technology began by understanding how its cloud functioned: from the overall architecture, down to individual microservices.

This big-picture perspective is critical to manage a complex, multi-cloud environment like Palo Alto Networks'.

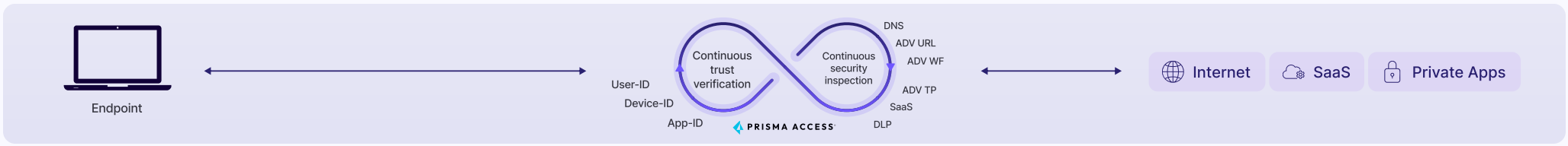
Whereas the cloud is often fragmented by provider or by team, Sedai provides a single control plane for the company's entire cloud footprint.

"Our cloud is extremely complicated," Suresh said. "We've got hundreds of services interacting with each other to deliver the outcome our customers expect. The only way to make sure our cloud is available all the time is for humans and AI to work together."

> "This is where AI becomes critical. Sedai performs autonomous actions that we wouldn't be able to handle without it."

That's exactly what happened. Sedai began making recommendations to reduce costs, improve performance, and fix availability issues across Palo Alto Networks' cloud — recommendations that its engineers would review and approve. At the same time, Sedai's platform continued to build its understanding of the company's environment.

The result was greater and greater trust in the platform.



Products like Prisma SASE require complex interactions across many cloud resources.

Recommendations vs. Real Impact

Importantly, Sedai is not another monitoring tool, and it goes beyond recommendations. Once the Palo Alto Networks team had established trust, it enabled Sedai to operate in Autopilot mode, in its production environment. In this mode, Sedai takes incremental, autonomous actions to achieve the team's specific SLOs.

In other words, Suresh's team sets the "directions" for its cloud, and Sedai drives it there.

"As you can imagine, we've got a very, very large volume of alerts coming in from our monitoring tools, which is simply impossible for humans to process," Suresh said. "This is where AI becomes critical. Sedai [performs] autonomous actions that we wouldn't be able to handle without it."

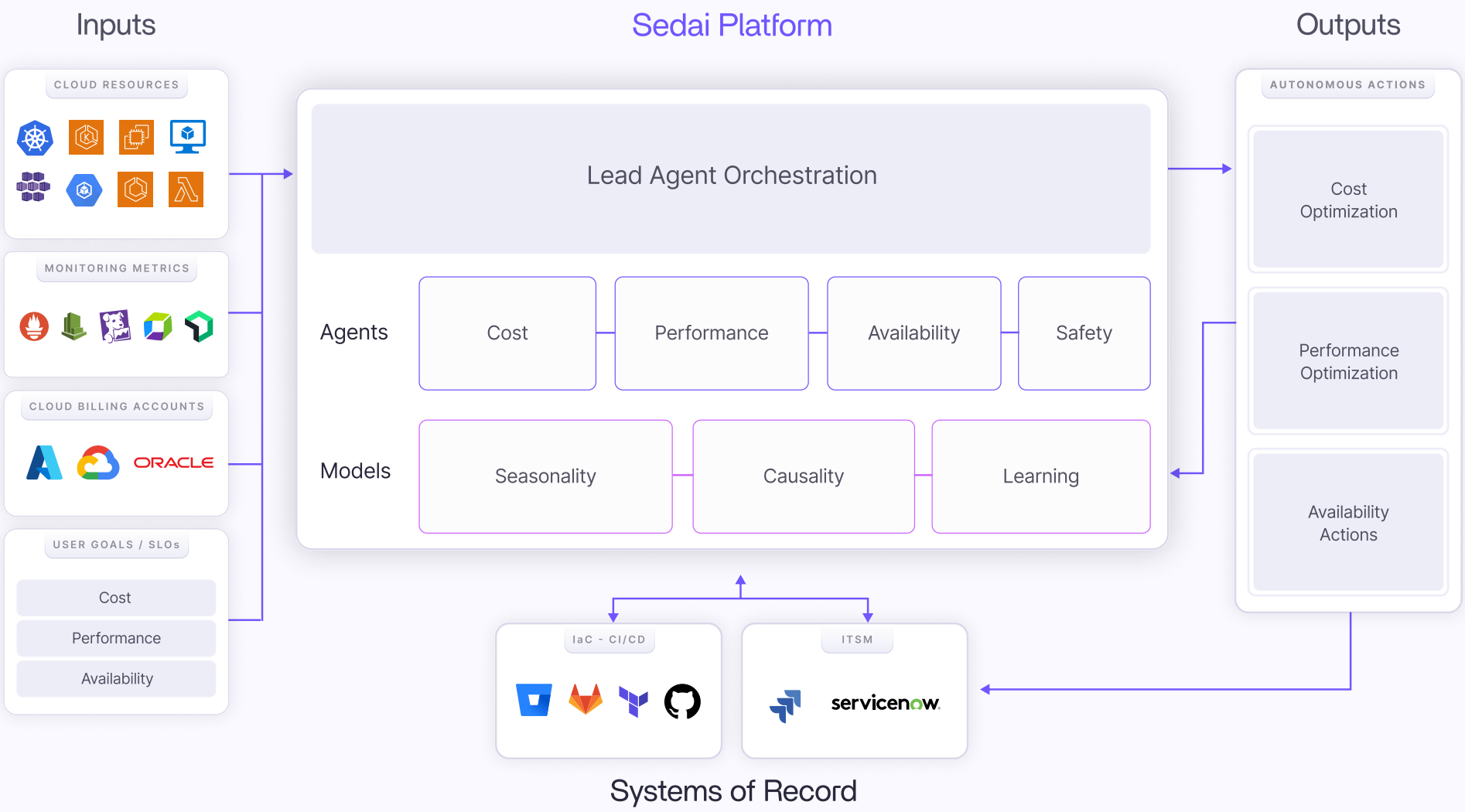
Sedai's unique approach relies on multiple AI agents, each focused on a specific objective such as cost, performance, or availability. For example, when Sedai detects an overprovisioned resource, the cost-optimization agent can autonomously reduce the memory allocated to that resource. It stops at the precise point where other agents predict that further reduction would impact performance or availability.

Sedai thinks like an expert engineer — but acts on a scale that is only possible with AI.

"The cost is really on us," Suresh said. "We're responsible for making sure we run our service in an optimal way. But as the scale goes up, it becomes harder and harder to balance optimal costs with performance and availability. The conventional, human-centered approach just doesn't work."

> "The only way to make sure our cloud is available all the time is for humans and AI to work together."

The results of autonomy have been remarkable — and that's not just marketing speak. Sedai has already saved Palo Alto Networks a staggering $3.5 million in cloud spend (before accounting for special discounts). And at the same time, Sedai has freed its engineers from thousands of hours of manual toil. The company can now redirect these resources, away from managing back-end systems and toward creating a better experience for customers.



Sedai is a single optimization plane for Palo Alto Networks' cloud environment.

99.999%

You can probably guess why we led with $3.5 million in savings: It's the type of big, flashy number that you'd love to show your CEO. But for Suresh, the most important number is 99.999%.

That is the availability standard that the entire engineering team at Palo Alto Networks works hard to deliver.

"To reach 99.999%, we can only afford five minutes of downtime per year," Suresh said. "It's an audacious goal, and we don't always get there. But we're getting better and better by using Sedai, which works alongside our engineers. It's this combination — engineering smarts and AI — that we need to handle issues."

"Five 9s" availability is difficult for any organization. But it's particularly challenging for Palo Alto Networks, given that the company uses all major public clouds, manages its own data centers, and runs tens of thousands of microservices. For Suresh, the recent AWS and Google outages have made this problem very real:

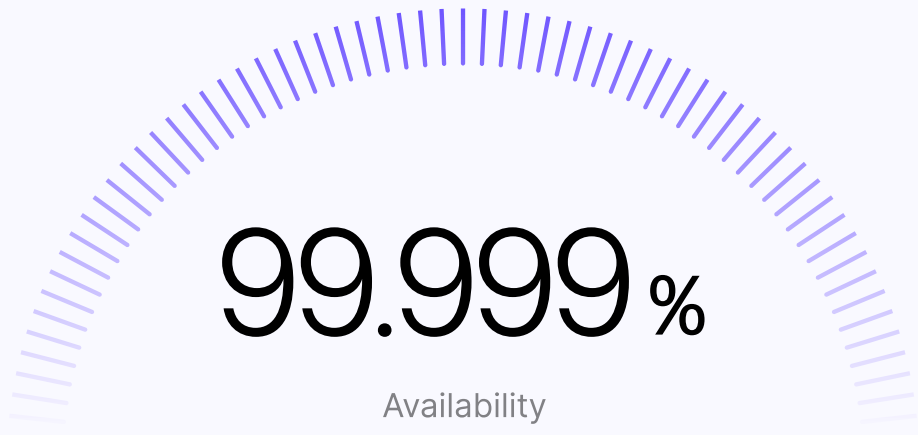
"How do you respond when a major cloud provider has an outage? Because disaster does strike. We need to design our cloud to be resilient. So when there is a failure, we're able to react right away and prevent disruption for customers."

Ramesh Nampelly, Senior Director Of Cloud Infrastructure and Platform Engineering, said that the sheer size of this environment had become overwhelming for his team:

"The engineers who managed our production operations were dealing with a huge amount of toil," Ramesh remembered. "They were working 24 by 7 to provide five 9s availability, which caused a lot of burnout. Our SREs were doing the same thing — again and again — without the ability to automate it. And we can't grow our SRE team linearly, as the number of customers and workloads increases."

Since then, Sedai has helped Palo Alto Networks truly take control of its availability. Because our platform has developed such a deep understanding of the company's cloud, it is able to detect anomalies in its typical behavior, including potential availability issues. This enables its SREs to spend their time on critical, meaningful work.

"Sedai has helped us save millions of dollars by optimizing and managing our own back-end services," Suresh said. "But most importantly, what Sedai has done very well is allow us to respond in real time when anomalies are detected."



AI That Puts Engineers in Control

For Palo Alto Networks, Sedai is more than just a fancy AI tool that reduces its costs. Sedai is the platform that allows engineering leaders like Suresh to understand, manage, and optimize its cloud — putting them in command.

"We've gone from what used to be automated, deterministic workflows to autonomous, with Sedai," Suresh said. "The human element is indispensable, and it always will be. But more and more engineering toil is being done by AI, so as humans, we can move up the value chain. That's how we can deliver what our customers expect of us."

See Sedai in Action

sedai.io