servers.com BY NEXCESS | Sipfront

CASE STUDY

How Sipfront found a hosting partner that understands its workload

Table of contents

Sipfront provides automated quality assurance for telecom systems: carrier networks, contact centers, and emergency phone lines that have to work every time. Testing these systems means generating enormous volumes of call traffic on demand, a workload most hosting providers would misread as an attack, and one few can even support at all. Sipfront needed a partner that could custom-build server capacity for exactly that job and deliver it in hours.

3 **The Story**

How custom-built servers help Sipfront handle demanding test traffic, expand into new regions, and reduce costs.

5 **The Technical Companion**

A closer look at the workload, architecture, and hardware behind Sipfront's tests of up to 50,000 concurrent calls, its performance results, and how GPU servers support its voice AI testing.

THE STORY

Sipfront keeps critical phone systems working, with test capacity custom-built in hours

Andreas Granig has stood in a data center at 2am, testing a software upgrade case by case, two phones in hand, blowing into one to check that the other side hears it, racing a 6am deadline before real callers flood the system. He knows the risks of what can happen when exhausted engineers forgo this full QA testing effort. If the updated system belongs to emergency services, for example, the cost of even a single missed test can be measured in lives.

CUSTOMER
Sipfront, automated quality assurance for voice AI and telecom systems

INDUSTRY
SaaS: voice AI and telecom testing

HQ AND REACH
Austria; test traffic simulated across the US, Europe, and Southeast Asia

SERVERS.COM SERVICES
Custom-built bare metal, GPU servers for AI testing workloads, global footprint, dedicated account team

> **"The age of manual testing has to be over"**
>
> ANDREAS GRANIG
> CEO & FOUNDER, SIPFRONT

Andreas Granig, spent years manually testing software upgrades to telecommunication systems. He was often running QA throughout the night to ensure that the new software would work as required, since emergency services companies were dependent upon the quality and uptime of their telecom network. He was convinced that there had to be a better way to test new software.

That conviction became Sipfront, the company he founded in 2020 after 25 years spent building the plumbing of modern telephony, including open source software that now runs inside VoIP systems worldwide, including at AWS Connect and at Google. "If you've made a phone call today, chances are it already ran over software we developed."

Today Sipfront tests it all automatically: routine functionality, crushing concurrent call loads, end-to-end call quality, and, since 2026, voice AI agents now answering many of those calls. Their tests help businesses address scenarios they fear most, sudden spikes in demand for service. When a surprise storm threatens to leave rail passengers stranded and cause thousands of anxious callers to hit a voice bot at once, Sipfront proves the system will hold before the storm arrives. When a large German telco moved its contact center onto new core systems, Sipfront and its integration partners verified that 5,000 contact center agents could go online at 7am the morning after the migration.

The challenge

Sipfront started on AWS deliberately, to avoid managing servers while the product found its footing, knowing the economics of simulating tens of thousands of calls on a hyperscaler solution would prove costly as the company grew. Once bills exceeded pain tolerances, they sought out a more cost-effective approach. Their first bare-metal provider offered no capacity in the US or Southeast Asia, where customers needed test traffic to originate, and the vendor’s automated security systems kept mistaking Sipfront's legitimate test traffic for an attack. Nobody there took the time to learn the difference, and other hosts were worse. Granig said, "One provider in the US just off-boarded us due to our load. When we opened support tickets, they were like, we can't do anything about that."

Sipfront needed a partner that would build for its workload instead of flagging it as an attack, put capacity wherever its customers needed it next, and answer a support call when it mattered.

4

data centers handle workloads across the US, Europe, and Southeast Asia

5,000

call center agents verified online at 7am after a migration

The solution

Servers.com by Nexcess custom-builds Sipfront's servers for the workload's exact demands and delivers them at project speed. "We get custom-built servers in a matter of hours mostly, and sometimes days depending on the availability in the region, but they are very much tailored to our workload," Granig says.

The flexibility of this bespoke arrangement unlocks a key benefit for Sipfront. Its control plane and orchestration stay on AWS, running exactly as before, while the heavy test traffic runs on Servers.com bare metal. The team kept the workflow it already had and moved only the workload that was breaking the budget.

"Servers.com understands what we do. They understand our traffic patterns, and they can accommodate them," Granig says. The relationship is a partnership, with dedicated support channels and regular check-ins designed to anticipate needs rather than respond to them. As Sipfront's testing extended into voice AI, the same partnership supplied GPU servers for the specialized quality-analysis models it runs in-house.



The cost of AWS to run the same workload on Servers.com, Sipfront estimates

THE TECHNICAL COMPANION

How Sipfront runs 50,000 concurrent test calls

CUSTOM BARE METAL WITHOUT TRIPPING ABUSE ALARMS

Sipfront load-tests carrier networks, contact centers, and voice AI systems at concurrency scales of up to 50,000 simultaneous calls. This is the engineering story behind our partnership with Sipfront: the workload, the architecture, and the numbers.

A typical Sipfront test run pours simulated calls out of a handful of servers toward a single destination: a flood of very small packets from three or four source IPs, converging on one target. That is an orchestrated-attack signature, and automated security systems treat it as one. Sipfront's previous providers couldn't deal with it. One flagged legitimate load tests repeatedly, with alerts going upstream to other providers in the stack, where they damaged reputations at multiple levels at once. Another off-boarded the company for its load. Sipfront realized they weren't looking for the cheapest bare metal provider, but a partner who would take the time to understand their workloads, recognize that their network traffic shouldn't trigger security alerts, and tailor a cost-effective solution to fit.

"What looks like a denial-of-service attack to 90 or 95% of the typical hosting customers is actually legitimate traffic for us. It's real voice traffic."

ANDREAS GRANIG
CEO & FOUNDER, SIPFRONT

The workload

Telephony traffic flips most hosting assumptions on their heads. The packets are tiny, the volumes are enormous, and per-packet processing rewards raw clock speed over data-center pedigree.

"Typically you might get a Dell or an IBM server with Intel Xeon processors, which is usually a good thing if you run databases or web servers," Granig says. "Not so much if you need to handle and process a lot of real-time traffic. There you want consumer-grade AMD Ryzen CPUs with five gigahertz of clock rate, because you can then generate and absorb a lot of packets."

Availability requirements also deviate from norms. Test agents are ephemeral by design: workloads spin up, run for days, and are torn down, and if a node drops, the orchestration layer reallocates to another server. Sipfront deliberately trades five-nines hardware for clock speed, and the architecture absorbs the difference. That trade is only safe because the fleet is custom-built for it, which is the difference between a provider that stocks standard hardware configurations and one that builds it to order.

The financial piece of their puzzle is similarly unique. Egress pricing is what pushed Sipfront off a pure hyperscaler footprint in the first place. "These hyperscalers, they are not cheap," Granig says. "The way they charge for network traffic, when you do call simulations, 50,000 concurrent calls, point to point, that's going to be a lot of traffic and it's going to be expensive."

He was, however, willing to pay a premium for bare-metal servers from Servers.com because they could provision systems in global regions, including EMEA, US and Southeast Asia, and because they were able to configure the network and security components to treat the traffic as a specification instead of a violation.

The architecture

Sipfront runs a deliberate hybrid infrastructure. The control plane remains on AWS, managed through ECS Anywhere, so orchestration, APIs, and portals work exactly as they always have. The execution infrastructure is Servers.com bare metal, where every traffic-generating and traffic-ingesting workload lives. "You can still manage all of your workload through their APIs, their portals, but you actually plug your own infrastructure in there," Granig says. "That is a super efficient way to have the cake and eat it too." This bare-metal-behind-a-cloud-control-plane pattern allows infrastructure repatriation to be scoped to the workloads that need it, with the cloud operating model intact and the data layer untouched.

	CONFIGURATION AT A GLANCE
Compute	Custom-built servers on high-clock-rate consumer-grade AMD Ryzen 9 7950X CPUs (~5 GHz class), tailored per workload.
Compute (AI Workloads)	GPU servers running Sipfront's proprietary voice-quality inference models on NVIDIA Tesla L4s.
Orchestration	AWS ECS Anywhere control plane; Servers.com bare metal execution layer
Regions	Long-term points of presence plus project-based deployments in the US (Washington, Dallas), Europe (Amsterdam), and Southeast Asia (Singapore).

The results

	OUTCOME
Concurrency scale	1,000 to 50,000 simultaneous calls
Abuse-system friction	Zero abuse-system incidents since switching
Provisioning turnaround	Custom-built servers in hours, sometimes days by region
Peak network load	Up to several million packets per second
Migration surface	Control plane unchanged; data layer never moved

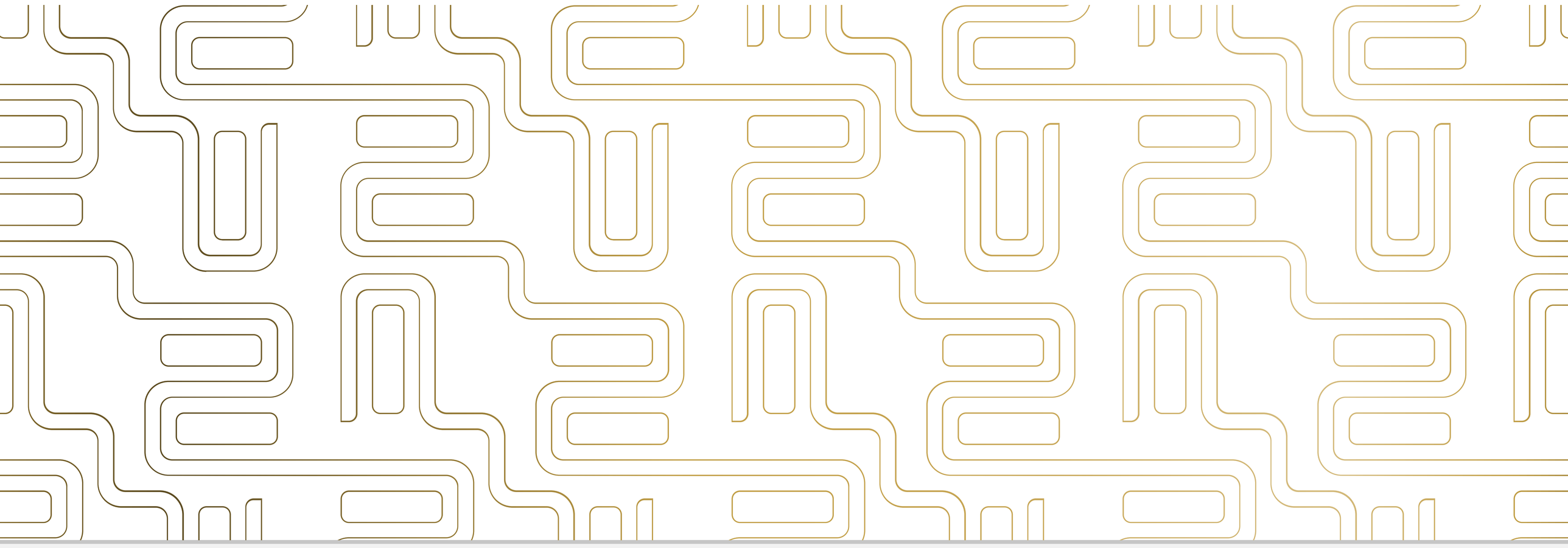
The future

The arrival of voice AI added a new dimension to Sipfront's call quality tests. Before deploying AI-driven voice agents, Sipfront’s customers used more primitive systems that detected callers’ prompts and selected from a predetermined set of actions and responses. In that paradigm, Sipfront could measure quality by comparing a call to a reference recording. Their customers’ new voice AI systems respond unpredictably by design, and the most effective way to test them at scale is to grade recorded test calls with AI models. However, off-the-shelf LLMs aren’t suited for the task, so Sipfront needs to use specialized models that are trained on hundreds of thousands of minutes of recorded calls. And GPU-equipped Servers.com bare metal allows them to do it. "Running your own AI models is technically not so complex," Granig says. "You just need to have the right hardware."

Sipfront must also test their customers’ AI-driven systems for compliance with new regulations. Provisions of the EU AI Act that took effect in August 2026 require Sipfront’s customers to disclose when an EU-based caller is talking to an AI. The Act also includes provisions for capturing consent and accommodating callers’ rights to human handoff and dialect fairness. These are all things Sipfront tests for today. As those workloads grow, the GPU footprint grows with them, on the same bare metal partnership that carries the call traffic.

"We go to Servers.com and say, look, we need these kind of servers with these kind of GPUs, can you provide them in these regions?" Granig says. "The support is always super fantastic. We get hardware we need at a price that is a no-brainer."

Thanks to Andreas Granig for sharing the Sipfront x Servers.com story.



Servers.com by Nexcess custom-builds bare metal around the demands of your workload.

Contact the team

servers.com BY NEXCESS | Sipfront