servers.com BY NEXCESS | HasData



CASE STUDY

How HasData moved half a billion monthly requests to bare metal

Table of contents

HasData's APIs deliver web data to AI agents, sales tools, and SEO software. As demand climbed, the company moved to a new hosting setup. This case study follows that change in two parts: the business story, and a technical companion for readers who want the engineering details.

3

The Story

Why HasData moved its scraping APIs off shared virtual machines, and how dedicated bare metal helped it reduce latency and grow without changing how its team works.

7

The Technical Companion

A closer look at the workload, architecture, and hardware behind HasData's half a billion monthly requests, plus the results of its move to bare metal and what comes next.

THE STORY

Why HasData runs half a billion web requests a month on a bare metal cloud its own team operates

HasData scrapes the web for data so its customers don't have to. AI agents, sales tools, and SEO software need information from complex web pages. HasData gets it for them. Customers make simple API calls and get back clean, structured data. HasData handles the hard parts: managing requests and proxies, rendering pages, and parsing data from search results, property listings, and product pages. Its 50-plus ready-made APIs handle roughly half a billion requests per month. Their customers' integrations demand reliability and speed. To deliver at that scale, HasData needed a cloud setup its own team could run, with the steady compute and private networking a real-time API needs.

Roman Milyushkevich, HasData's co-founder and CTO, spent most of his career extracting data from websites. At every new job, he built tools to handle similar workloads. Browsers had to run by the hundreds, and the parsers that turned pages into data broke every time a site changed its layout.

CUSTOMER
HasData, web scraping and data extraction APIs

INDUSTRY
SaaS: web data for AI agents, go-to-market software, SEO tools, and real estate

HQ AND REACH
United States with a global customer base

SERVERS.COM SERVICES
Enterprise Bare Metal servers, cloud instances, and load balancers in Dallas-Fort Worth; global private network; a direct line to the account team

> **“I built HasData based on my own experience, to solve my own problems, and it appears others have the same problems.”**
>
> ROMAN MILYUSHKEVICH
> CO-FOUNDER AND CTO, HASDATA

HasData now has more than 1,300 customers, and the largest group of them is building AI agents. An agent that needs fresh information from the web has to search for it and turn the results into data its software can use. HasData provides that information through its APIs.

Outside of AI agents, HasData’s customers use its APIs for a variety of other purposes. Some use the company and contact data they retrieve through HasData to add more information to records in their own sales tools. Other customers use the data in SEO rank trackers and real estate investment software. Some customers use HasData’s APIs to track how brands show up inside AI answers.

Every one of them trusts HasData to deliver the data they need, when they need it. That means HasData needs a cloud environment it can rely on.

The challenge

HasData initially used another cloud provider's virtual machines for its infrastructure. They were cheap and good for the price, but the company outgrew them as request volume climbed. HasData needed more CPU and more private network bandwidth, and the provider's outages were frequent. Growth exposed other problems, too.

First, latency began spiking on the network in ways the team could not trace to its own code. Roman suspected some of the cloud provider’s other customers were causing the problem. Their virtual machines shared the same hardware and could take CPU time away from HasData's parsers.

Then Google began blocking the cloud provider's IP addresses based on reputation, meaning the addresses had carried enough bad traffic in the past that Google treated any request from them as suspect. The blocking broke HasData's own calls to critical services like Google Cloud Storage and Firebase. To restore access to those services, the team had to route that traffic through other providers.

Then, the cluster setup created another challenge. HasData ran its own Kubernetes cluster on the cloud provider. But its nodes could only communicate through a VPN over the public internet.

The most obvious solution to the problem was working with a hyperscaler. Roman spent a couple of days comparing prices from AWS and Google Cloud. "I just opened their pricing calculators and it was a very fast decision that we were not going to use them," he says. A mid-tier cloud provider cost less, but was still too expensive at HasData's scale. The company was handling roughly half a billion requests per month.

Since HasData was already running its own Kubernetes cluster, the decision shifted from price to what kind of servers the cluster should run on. "If we were hosting our own Kubernetes cluster on [our previous provider], why are we looking for a managed solution?"

“If we were hosting our own Kubernetes cluster… why are we looking for a managed solution?”

ROMAN MILYUSHKEVICH
CO-FOUNDER AND CTO, HASDATA

The solution

Bare metal was the answer, and Servers.com by Nexcess was where HasData found it. The company migrated in November 2025.

Servers.com bare metal gave the Kubernetes cluster two things the virtual machines could not. The first was dedicated servers. No other customers share the CPUs, so nothing else takes processor time from the parsers. That removed the factor Roman believed was causing the latency spikes. Parsing latency dropped by roughly 30% after HasData moved the cluster to bare metal.

The second was a global private network. Every node in the cluster now talks to every other one over a private path, eliminating the problematic VPN setup. Before committing, Roman ordered two servers for a test. Confirming they could communicate without a VPN and seeing the network speeds settled it for him. Evidence showed the Kubernetes cluster would perform far better on Servers.com infrastructure.

Moving to Servers.com also eliminated the IP address reputation issues. Since the move, HasData's calls to Google Cloud Storage and Firebase have not been blocked.

Most importantly, HasData did not have to change how it operates. The team has kept its own Kubernetes cluster, managed with Rancher, and deploys to it the way it always has. Standing that cluster up on bare metal was the one real effort in the move, and HasData chose to hire a DevOps engineer to handle it. He still maintains the environment, but developers can deploy new pipelines themselves, with the DevOps engineer reviewing their work. The economics reinforce the decision.

The environment has run without an incident since the move. Support is also more direct. "We have a channel on Telegram where we can ask any questions at any time," Roman says. "I have direct contact with Travis [King, HasData's account executive at Servers.com]. He replies within minutes. And most importantly, we don't have any problems."



Parsing latency down roughly 30% after the move to bare metal

"What we are paying for servers plus [our DevOps engineer's] salary is less than the AWS bill would be for us at our scale."

ROMAN MILYUSHKEVICH
CO-FOUNDER AND CTO, HASDATA

0

Unplanned incidents since the November 2025 migration

1

DevOps engineer maintains the environment, and developers deploy themselves

Nearly a year in, Roman's report on the Servers.com environment is short and sweet. "It just works. We set everything up at once, and now we're just updating some stuff regularly." HasData was running 9 servers at first. It now runs 14, plus 3 cloud instances and 2 load balancers. It adds a server about every two months, and request volume has reached roughly half a billion a month. The environment has kept up without a major change in how the team runs it.

For the engineering story behind this partnership, including how a CPU-bound workload behaves once it stops sharing hardware, read the technical companion on page 7.

THE TECHNICAL COMPANION

Direct CPU access and a private network: the architecture behind HasData's move off virtual machines

HasData serves roughly half a billion web-scraping API requests a month from a Kubernetes cluster its own team runs on Servers.com bare metal. This is the engineering story behind the HasData case study: the workload, the architecture, and the numbers.

> **“With our Servers.com infrastructure all worker nodes now connect to each other through a private network, allowing us to get rid of our old setup that used a VPN over the public internet.”**
>
> **ROMAN MILYUSHKEVICH**
> CO-FOUNDER AND CTO, HASDATA

Each HasData request is a client asking for structured data from a webpage. Behind its APIs, HasData runs thousands of real Chrome instances. They render JavaScript-heavy pages while proxies rotate and parsers turn HTML into JSON. Parsing is CPU-bound, so when a noisy neighbor virtual machine takes processor time, HasData's customers see it as a slower response.

HasData originally ran that workload on a cloud provider's virtual machines, using a Kubernetes cluster it operated itself. The search for a better provider led to a foundational question. Does a team already running its own cluster need to pay a hyperscaler to run one?

If not, what does that team need to deliver reliability at scale? They need hardware that doesn't share cycles and a network that lets the cluster's nodes talk privately without relying on the public internet.

The workload

There are two ways to scrape a page: "The first way is to send a simple HTTP request and get back HTML," Roman says. "It's very simple, very fast, and it doesn't consume a lot of resources." The second method uses a browser to render pages that require JavaScript. Chrome is very resource intensive, especially with CPU and RAM. Running enough browsers to render hundreds of thousands of pages a day requires substantial compute capacity.

HasData runs real Chrome behind a virtual display, so each browser instance has the CPU and RAM demands of a full desktop browser.

The parsing step is where being on dedicated bare metal hardware creates big performance wins. HasData was experiencing latency spikes it couldn’t explain.

Roman had a hypothesis, "In the cloud environments your virtual machine can be on the same server with virtual machines from other customers, and they can sometimes steal your CPU resources. For our use case CPU is very important, because we are parsing data. Parsing HTML into JSON requires a lot of CPU resources." Moving to bare metal confirmed it. "On the bare metal with direct access to CPU we don’t face this issue anymore, and our latency on parsing since migration to bare metal dropped maybe 30 percent."

Roman is direct about the tradeoffs. "For us it was a decision between paying more on AWS, where Kubernetes is built in and updates without our participation, or paying less but setting up everything ourselves and maintaining it ourselves," he says. His verdict after nearly a year is clear. "It requires some more effort in the beginning, but it pays off later."

The architecture

The cluster is operated by HasData's team. Every worker node is Servers.com bare metal, and the nodes communicate over the global private network rather than a VPN overlay on the public internet.

The one real effort in the migration was standing the cluster up. "On AWS you just write these Kubernetes manifests and deploy the app," Roman says. "When you're on bare metal, you have to set up Kubernetes first." HasData chose Rancher to manage the Kubernetes cluster running on Servers.com bare metal. "It is pretty straightforward to set up. It's scary when you do it the first time." The team hired a DevOps engineer to build and maintain the cluster on bare metal, but the day-to-day doesn't require him. "There are a lot of ready-made scripts on GitHub, and nowadays with the help of AI coding tools our developers don't bother our DevOps engineer to deploy something new," Roman says. "He just reviews stuff from us."

The result is self-managed Kubernetes on bare metal. HasData kept the same cluster and deployment model, but moved from shared virtual machines to dedicated hardware with private networking between nodes. The worker nodes run on single-socket AMD EPYC servers, and the databases run on separate blade servers with NVMe storage. Two load balancers spread traffic across the nodes, and everything sits in Servers.com's Dallas-Fort Worth data center.

CONFIGURATION AT A GLANCE

Compute	Enterprise Bare Metal, single-tenant. Worker nodes: AMD EPYC 9354P, 128 GB RAM, 960 GB SSD. Database nodes: Supermicro blades (H13SRD-F) with AMD EPYC 4564P, 96 GB DDR5, and 2 x 3.84 TB NVMe SSD. One small test node.
Footprint	14 servers, 3 cloud instances, and 2 load balancers as of October 2026, up from 9 servers at the November 2025 migration. About one server added every two months.
Browser fleet	Thousands of real Chrome instances rendering JavaScript-heavy pages.
Network	Global private network carrying all traffic between cluster nodes; VPN overlay eliminated. One server upgraded to 10G network cards in 2026.

Regions	Dallas-Fort Worth. All servers and cloud instances are in one data center.
Redundancy	Two load balancers split traffic across the nodes.
Orchestration	Self-managed Kubernetes on Rancher; DevOps engineer maintains, developers deploy. New servers come through the account executive, who finds the hardware, moves it to Dallas-Fort Worth, and sends an order link the team uses when it's ready.

The results

	OUTCOME
Parsing latency	Roughly 30% lower after migration
Incidents on Servers.com	Zero unplanned incidents since migration (November 2025 through early October 2026)
Cluster networking	Private network replaced a VPN overlay on the public internet
Product uptime	99.9%
Footprint	From 9 servers in November 2025 to 14 servers in October 2026

The proof

Before committing to bare metal, HasData bought two servers as a test. The deciding question was how the networking connections would perform. "In the very beginning we did not understand how the Servers.com global private network works," Roman says. "So we had to order two servers. We did a small test, tried this GPN between them, and we were happy with the results." The test is the reason the cluster's nodes run without a VPN overlay today. After the test, the team spent about two weeks settling on the server configurations it wanted.

The future

HasData's fastest-growing customers did not exist a year ago. "We have a growing segment of different AI visibility trackers," Roman says. "A lot of products for tracking AI visibility were created in recent months, and some of them are using our APIs to actually get this visibility data." Sales software and AI agents that need web access are still HasData's main customer segment. And HasData expects request volume to keep climbing from them as well.

To support the company's growth, the cluster grows with request volume. "We're updating, adding new servers, adding new nodes to our cluster," Roman says. HasData has added a server about every two months since the migration and recently upgraded one server to 10G network cards. When the team needs more capacity, it tells its account executive what the servers are for. He finds the hardware and moves it to Dallas-Fort Worth. Then he sends an order link, and the team brings new servers online on its own schedule.

To talk through a CPU-bound workload that shared virtual machines keep slowing down, contact the team.

Thanks to Roman Milyushkevich for sharing the HasData 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 | HasData