

The economics of edge Kubernetes: A practical guide to ROI and risk



Risk and reward at the edge

Edge computing has moved from experiment to mainstream. In Spectro Cloud's 2025 State of Production Kubernetes research, half of K8s adopters say they're already running clusters at the edge.

For some, it's small deployments of IoT gateways or local processing nodes; for others, it's fleets of thousands of sites supporting industrial automation, healthcare imaging, or retail analytics.

Regardless of your industry or use case, there's a good chance your chosen infrastructure is Kubernetes-based. K8s stacks are a natural choice for edge deployments, providing edge-to-cloud portability, native resilience, cloud-native agility, and access to a huge ecosystem.

The value of edge is easy to understand: edge lets you keep data close to where it's generated, reducing latency, improving resilience, and cutting cloud egress costs. It also unlocks new

(often AI-based) applications that promise to generate revenue and improve the customer experience — meaning there's often a measurable business case driving the investment.

But edge is a very different beast to cloud or data center, and it comes with unique economics founded on operational complexity. Unless you can account for these economics, manage the costs and mitigate the risks, your edge project could end up in the red.

In this paper we'll help you get a sense of the costs and risks associated with each aspect of your edge Kubernetes project... and suggest some ways that Spectro Cloud Palette can help you solve them.

Real data to support your strategy

Each year we invest in deep research into the state of enterprise edge, giving you hard numbers you can use to benchmark your progress and identify best practices. Get your copy of the 2025 State of Production Kubernetes report at bit.ly/sopk2025.



Edge winners and losers

Gartner predicted that 50% of edge computing solutions deployed by 2025 **without an enterprise edge strategy** would fail to meet goals for deployment time, functionality, and/or cost. Gartner also warned that 50% of edge computing solutions that “worked fine” as proof of concepts would fail to scale into production. Which side will you be on?

50%

Why the economics matter

Edge computing changes the financial model of IT. In the cloud, you can rely on hyperscale providers’ economies of scale and global automation. At the edge, you’re responsible for it all: the hardware, the network, the operations, and the lifecycle management.

Each additional site introduces both capital cost and operational drag. Scaling from ten pilot locations to hundreds — or tens of thousands — means multiplying your management overhead, risk exposure, and compliance surface.

That’s why any serious edge rollout needs a clear return-on-investment (ROI) framework. A spreadsheet might show per-site hardware costs, but true ROI comes from understanding the total cost of ownership (TCO) over time: deployment effort, ongoing maintenance, downtime risk, and the opportunity cost of operational inefficiency.

At Spectro Cloud, we’ve developed a simple but comprehensive edge ROI model, which we work through with our customers to map out the inputs and outputs — and highlight how automation and governance can make or break financial outcomes. Two identical hardware rollouts can have vastly different economics depending on how efficiently they’re managed.

Your personalized calculation

We’re here to help: get in touch with us and we’d be happy to workshop your personalized ROI model, 1:1. Our experts have helped edge customers large and small, so we know the benchmarks that matter and the keys to success.

Building your edge ROI model

The starting point is identifying the key cost, risk and return levers that shape your financial outcomes.

On one side of the equation, you probably already have a sense of the value your edge project is expected to deliver, and how you quantify it — whether you're expecting process efficiencies, new revenue opportunities, or improvements in safety, business agility, or environmental impact.

But the flipside is the cost to make those returns real, and that's where we can help. Our ROI template includes several major categories. Let's walk through them.

Hardware and connectivity

Edge infrastructure is often bespoke — industrial PCs, gateways, or embedded systems. Purchase costs vary widely based on performance, redundancy, and network requirements. As you start doing more AI projects at the edge, you may be looking at beefier accelerated servers too.

As well as the known procurement costs like per-server pricing, and the monthly cost of connectivity billing, there are softer costs to consider too: spend to validate configurations for different environments (even down to things like mounting points and accessories), to build software images for staging, unpacking and repacking cycles, warehousing.

Don't forget installation, shipping, and replacement cycles, including maintaining spares stocks at staging sites. You'll need a buffer to account for natural hardware failure and other accidents.

Software licensing and support

From Kubernetes distributions to observability, security, and management tools, licensing can add up quickly, especially when priced per site or cluster. Support contracts cost money too.

If your edge equipment is going to be generating log events that are to be ingested into observability platforms, check the impact on your data retention and processing fees. You may also have to pay for cloud ingress or egress.

You should also consider standards adherence, openness and other aspects of vendor lock-in. Proprietary systems are convenient but often limited; conversely pure DIY with open-source components adds a lot of overhead and leaves you with a maintenance burden.

Deployment and lifecycle operations

Manual provisioning and upgrades consume staff time. Field engineers, travel, and downtime all carry costs, and this can in fact be one of the largest cost elements in your overall ROI.

In our experience, it's common for a single edge site visit to cost \$400–1,200, so it's important to avoid truck rolls wherever possible: for site prep, troubleshooting, upgrades, or provisioning. Automation and centralized control can cut FTE hours dramatically, especially for software rollouts and upgrades.



If a physical visit can't be avoided, look to simplify the onsite work so a non-specialist member of staff can perform it: for example, shipping a new box to site pre-staged so it can be turned on by any employee, given power and ping, and activated remotely.

Adopters feel the pain of deployment costs

IT Pro's Edge Computing Trends survey found that the **cost of deployment** was the biggest and most frequently cited challenge in adopting edge computing.

Automation is a vital way to reduce the risk of human error: e.g., there are risks of delays or errors if on-site users have to manually type IP addresses or follow complex multi-step runbooks.

Also think about pure speed. The longer your deployment cycle from lab to production, the worse your time-to-value. And the reason you're doing all this is because edge initiatives — from real-time analytics to AI-assisted products — enable new revenue streams or lower other costs. Sure, you can save money by cutting corners, but that might be a false economy if it causes delays.

This particularly matters when you're planning for scale. We've worked with customers that have hundreds or thousands of sites, and yet with their existing tooling, processes and people, could only commit to deploying their edge infrastructure to a handful of sites each week. It's simple math to spell out the implications: **the business would have to wait literally years for any kind of 'full' rollout to be completed.** This is normally when leadership looks around for a better way to get the project done!

Downtime and SLA impact

Even short outages at the edge can disrupt production lines, point of sale, or patient care. Modeling downtime cost per hour helps quantify the ROI of investing in things like on-site spares, high-availability architectures, and backup connectivity.

We've worked with retail customers that balked at a \$40 per month per site backup LTE connection, yet a day's lost revenue per site could easily top \$2,000.

Big causes of downtime include not just hardware failure (which is a fact of life at scale), but connectivity outages for cloud-dependent systems, network configuration issues, and failures during software updates, which are frequent in a K8s-based stack.

Technical measures — such as backup connectivity — can help mitigate downtime, but regardless you need an incident response plan with common scenarios modeled out.

Handling the risk of offline operations

Ensuring that your edge devices can continue operating without connectivity is imperative. To meet the continuity demands of its restaurant chains, Yum! Brands worked with Spectro Cloud to create an **easily deployable edge system** that runs offline, syncs later, and can be deployed with low or no touch. Hardware is pre-provisioned, shipped to stores, plugged in, and enrolled automatically.

As Yum! Brands' Ryan Good put it, "You need something that's reliable, resilient, scalable," and "everything must have the ability to run offline."



Yum!

Compliance and risk

Noncompliance with data, privacy, or security standards isn't just a technical issue — it's a financial one. Factor in audit costs and potential penalties, and expect to plan for and invest in a wide range of security controls for each edge site, whether you need to meet PCI, HIPAA or other standards.

Regardless of your compliance burden, security must be a huge consideration. Edge deployments are a perfect storm of security risks: unattended devices, exposed to physical tampering, sitting on third-party networks, potentially without regular patches.

Controls can include physical security (such as a locked cabinet for each edge server), data encryption at rest and in transit, system access controls, audit logging, and tamper resistance in both hardware and software (eg trusted boot).

Risk in regulated sectors

GE HealthCare's Ben Beeman described the challenges his team faced in controlling costs. "The closer your compute devices get to the patient, the more costly they become," he explained, adding, "Incorporating a new version can create a whole new development cycle, and with the regulations and rigor we follow, this can be too costly just to roll out a bug fix or patch."

Who owns what?

One of the things that makes edge deployments unique (and problematic) is that you're often dependent on infrastructure you don't own or fully control. Your hardware may be deployed in a customer's building, reliant on their network configuration, power, and physical security. This not only leaves you vulnerable to downtime, but can also complicate access for maintenance and troubleshooting.

Checklist: evaluating the true cost of edge K8s

Use this checklist as a conversation starter with your stakeholders between IT, finance, and operations. Each row represents a measurable lever in your ROI equation.

Category	Key questions to ask	Example metrics
Deployment	How repeatable are site rollouts? Can you automate provisioning?	Average time per site
Operations	How many FTE hours are needed per site per month to perform admin, upgrades, troubleshooting? How much unproductive time is lost in travel and expenses?	Hours per cluster/month
Reliability	What's your tolerance for downtime? What would it cost to add another 9?	Cost per hour of outage, MTTR, spares holding cost
Compliance & security	Can you enforce consistent policies and updates?	% sites compliant
Scalability	How quickly can you add or retire sites to meet growth objectives?	Sites deployed per month
Vendor lock-in	How portable are your clusters across hardware and clouds?	Migration effort (person-hours)

How Spectro Cloud Palette helps de-risk edge operations



Spectro Cloud's Palette platform was built to make Kubernetes practical anywhere, but especially at the edge. Here's how it directly addresses the economic and operational pain points.



Automated lifecycle management

Palette automates and de-risks key operational tasks across the full edge lifecycle, for example with zero-touch provisioning and failover-enabled upgrading. This eliminates the manual effort and truck rolls that typically dominate edge TCO.



Policy and governance enforcement

Platform teams can define global policies for security, access, and resource usage, then automatically enforce them across all edge clusters. Palette includes a host of edge-specific security features, such as trusted boot.



Repeatable, immutable deployment

Every cluster is built from a centrally managed, version-controlled profile that captures desired state. That means zero-drift compliance and instant reproducibility — critical for regulated industries and fleets of edge sites. Through our immutable image-based deployment approach we can promise that your edge devices are always in line with policy.



Scale and flexibility

Palette scales linearly from a handful of clusters to tens of thousands, supporting your choice of Kubernetes distribution on a wide range of hardware. No vendor lock-in or forced stack choices.



Financial impact

Organizations using Palette report dramatic reductions in operational overhead — often 50–70% savings in lifecycle management effort — and faster rollout timelines.

For a 1,000-site deployment, those efficiencies can translate into millions in avoided costs annually.



Airgapped and disconnected operations

Edge often means unreliable connectivity. Palette supports fully airgapped deployments with offline management, local registries, and synchronized updates once connectivity resumes.

From pilot to production: turning edge complexity into ROI

Building an edge strategy is both a technical and economic journey. The winners will be those who treat edge as a long-term investment and design for scale, compliance, and efficiency from day one.

Before you scale, run the numbers. Use our ROI template to model your potential cost and return scenarios, from per-site deployment to fleet-wide management. Then test your assumptions against the checklist above.

When you're ready to move from modeling to execution, Palette provides the operational foundation to make it real. Learn more about Palette at spectrocloud.com.

Explore more edge insights and best practices



RapidAI: AI at the edge for life-saving care



Dentsply Sirona: taking Kubernetes to the dentist



YUM! Brands scaling edge platforms to thousands of locations



Edge Kubernetes resources and guides



Spectro
Cloud

About Spectro Cloud

Spectro Cloud is your trusted partner for edge Kubernetes. We've helped three of the top 10 global restaurant chains, who collectively operate more than 100,000 sites. We work with healthcare innovators like GE HealthCare, RapidAI and Dentsply Sirona. And our technology powers tactical edge projects in military units operating on land, air and sea.

GigaOm rated us a Leader and Outperformer in edge Kubernetes three years running. We're an STL top 50 edge computing company, a Gartner Cool Vendor in edge computing, and on the CRN top 100 edge computing list.

We partner with the edge players you care about: from NVIDIA to AWS, HPE and Supermicro to SNUC (SimplyNUC).

And we give back to the community, driving innovation as lead sponsor of the CNCF's Kairos project for secure edge.

In short: nobody knows edge like we do.

Find out more at [**spectrocloud.com**](https://spectrocloud.com).