



# Kubernetes in the AI era

The Spectro Cloud 2025 State of Production Kubernetes

AI's irresistible gravity, edge in bloom, and the K8s powerhouses carving a path to operational effectiveness beyond human scale.

Research by  
adience

# Foreword

Last year we shone a spotlight on the first 10 years of Kubernetes. Now it's time for another milestone: 5 years of the State of Production Kubernetes report (you can see how the covers have evolved if you look to the right).

Every year we try to balance continuity and change. Asking the same question for five years is a unique opportunity to identify long-term trends — but only by asking new questions can we examine the emerging issues in our community.

So this year we have continued to track the scale of K8s adoption, challenges and growth expectations. But of course, we have also dialed up our coverage of AI, edge and KubeVirt, and expanded our coverage of the realities of Kubernetes operational practices and outcomes, following the footsteps of DORA.

The result is, for the first time, a breakdown of the K8s adopter community into four very distinct personas. Which one matches your business?

As always, we and our research partner Adience poured blood, sweat, tears and cold hard cash into making this research as good as it can be. We hope you enjoy the result. Grab a coffee and dig in.

Ant Newman  
Director of Content  
Spectro Cloud



*The previous four years of State of Production Kubernetes reports. Yes, we rebranded since the last one.*



# Contents

## Section 1

A return to confident scale, whatever the cost

## Section 2

AI driving everything, especially edge

## Section 3

Workloads on K8s and legacy modernization with KubeVirt

## Section 4

Operations and outcomes

## Section 5

The four types of K8s adopters

**Want to give us some feedback, or  
make a request for next year's report?**

Reach me at [ant@spectrocloud.com](mailto:ant@spectrocloud.com)

### A note on responsible AI use

This research was created by real humans using deterministic software: from writing the survey questions, to conducting the interviews, analysing the data, writing the draft and designing the chart and doc layouts.

We used some AI tools for supporting functions. We used ChatGPT o3 to perform some exploratory analysis on the raw data, which was subsequently verified by humans (yes, it made mistakes). We used AI transcription software to help wrangle 6+ hours of interviews into usable quotations. All em dashes were carefully placed by hand.

# Executive summary

Here's your TL;DR.

## Adoption matures, diversifies

- 65% say their employer has been using K8s for over five years.
- 51% of K8s adopters have >20 clusters and >1,000 nodes.
- The average organization has clusters in more than 5 different clouds and other environments.
- Multicloud strategies are the biggest driver of cluster placement.
- Two-thirds of orgs have >15 different software elements in their K8s stacks.

## A viable home for legacy VM workloads

- 68% of adopters have more than half of their app workloads in K8s today.
- 31% plan to migrate their remaining legacy VMs to Kubernetes — the most popular strategy.
- 86% have heard of KubeVirt and 26% use it in production today.
- Those who are 'all in' on Kubernetes are more likely to be using KubeVirt and planning migrations.

## AI in the spotlight, driving edge

- Cost is the #1 K8s challenge. 88% say their K8s TCO has grown in the last year.
- AI is the #1 opportunity to tackle cost. 92% said they were investing in AI cost optimization tools.
- Growth expectations are back. 90% expect their use of AI workloads on K8s to grow in the next year.
- AI is driving edge K8s adoption. 50% deploy to the edge today, with on average >50 sites.

## Operations: progress, not perfection

- 80% say they have a mature platform engineering function.
- >50% admit their clusters are snowflakes and their K8s ops work is highly manual.
- 59% say they could patch their entire K8s infrastructure within 24 hours.
- Organizations perform much better on DORA metrics when a central platform team manages tooling/pipelines for deploying apps to clusters.

## Section 1

# A return to confident scale, whatever the cost

# Do you remember your first time?

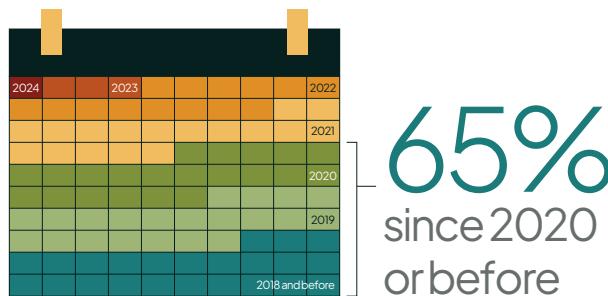
The State of Production Kubernetes research has always focused on those using Kubernetes in production in the enterprise, not students, tirekickers or homelabbers. So you'd expect our respondents to be experienced professionals.

Even back in our first report in 2021, 40% of respondents said they'd been running K8s in production for at least two years.

But over the past five years, our audience has grown into truly seasoned veterans, matching the evolution of Kubernetes itself.

This year, two thirds say their company has been using Kubernetes since 2020 or earlier.

**Figure 1:** Most orgs we surveyed have used K8s for 5+ years



**Figure 1:** How long has your company been using Kubernetes?

The takeaway? If you've been running K8s in production for five years, it's not new, it's not experimental — and we shouldn't be treating it as such.

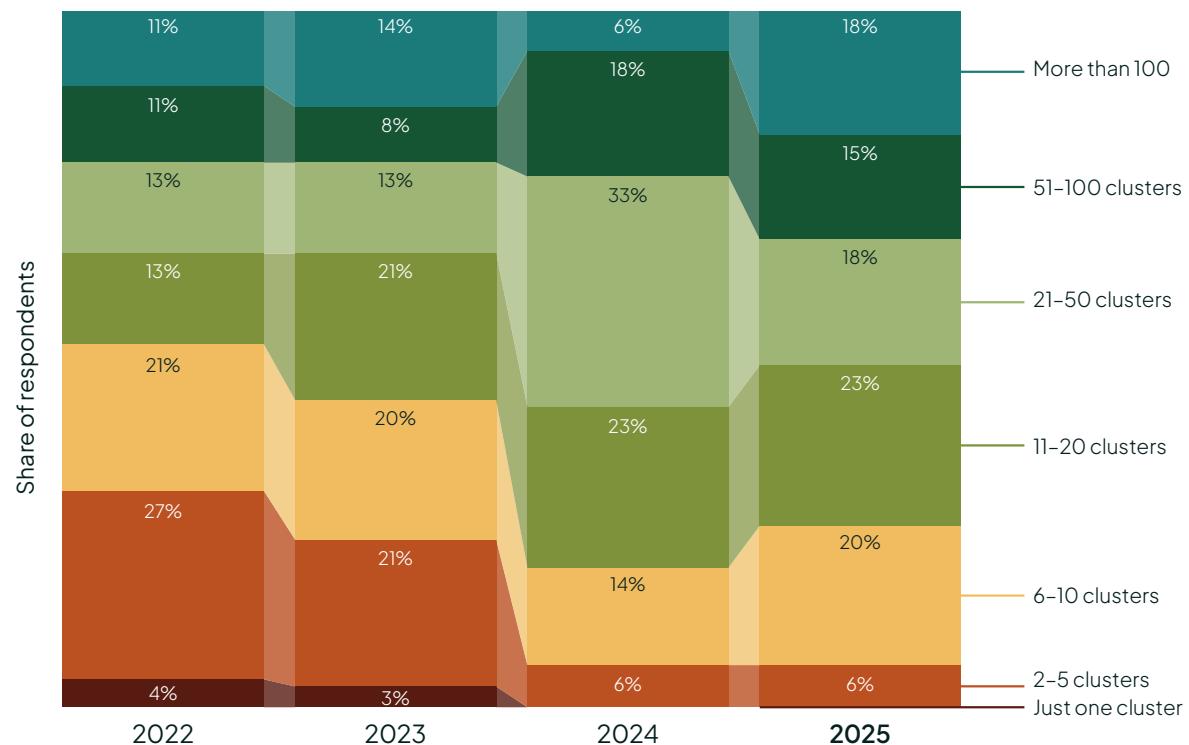
# Lots of clusters, lots of nodes... and lots of workloads

Our first task in the State of K8s report is always to gauge the scale of enterprise adoption.

In the early days, we used cluster count as a proxy for growing adoption: every year, we expected to see each adopter using more clusters, as they went from their first baby steps to increased scale.

By this stage of market maturity, we believe this growth factor is settling down, and that the count of clusters now reflects a more nuanced conversation around how to carve up resources: cluster architectures, multitenancy models, and the tradeoffs between running lots of small clusters or a few larger ones.

**Figure 2:** A third of orgs have more than 50 clusters



**Figure 2:** How many Kubernetes clusters does your organization currently run across all environments from cloud and DC to edge?

Often, the decision comes down to factors like balancing factors like blast radius, management overhead, organizational structures, and utilization. So we also asked about node count, to give a second measure for scale.

This year, half of all respondents have more than 20 clusters in production, as Figure 2 shows. 18% have more than 100. Just 7% run five or fewer clusters. It's clear that the burst of scale we saw emerge last year remains a reality.

Half have more than 1,000 nodes, 9% have more than 10,000. As you'd expect, the number of nodes correlates strongly with the number of clusters, as Figure 3 shows. If you have more than 50 clusters, you've probably got more than 5,000 nodes.

The large volume of nodes and clusters is understandable when you learn that 38% of organizations run at least three quarters of their total application workloads in Kubernetes today.

Last year we covered the overwhelming commitment that enterprises feel towards Kubernetes — we had a section titled “All in on Kubernetes”. It looks like nothing has changed.

Figure 3: Nodes per cluster scales pretty linearly



Figure 3: Approximately how many total Kubernetes nodes do you have running at any given moment?

## Three views of scale

“We're running eight clusters right now. The cloud group wants us in multiple zones, and we also separate by workload needs. Right-sizing beats one over-utilized monster.”

Lead DevOps and cloud engineer,  
US manufacturer

“We run at least 200+ clusters across the globe. Given we have more than 50,000 developers, and we have 1,000s of applications that are being developed across the globe, it's needed. It's not like we can simply say we'll have one big cluster with 10,000 nodes in it. We want to make sure there's enough isolation.”

Senior technology transformation executive,  
US financial services

“I can't give you an exact number, but we have 4,000–6,000 clusters. The installation has grown like 50% in four years, and I think we will increase our consumption by minimum, 15 to 20% year on year.”

Senior director DevOps and cloud infrastructure,  
UK telco

# The ‘full stack’ turns into a jenga tower

One of our central beliefs here at Spectro Cloud is that just firing up a K8s distro with CNI and CSI doesn't mean you have a functional enterprise Kubernetes cluster: there's so much more that you need to install in order to be production ready, and that's what we call add-ons, integrations, or platform apps.

As always, we asked how many of these software elements make up a complete stack. The growth trend we saw last year continued dramatically this year.

Now 41% say they have more than 20 software elements in their K8s software stacks.

When we consider add-ons like GPU operators as well as traditional components like ingress and monitoring, it's perhaps no surprise.

In our interviews, respondents consistently brought up how the pace and variety of tool innovation in the CNCF landscape is both a blessing and a curse.

Figure 4: Cluster ‘stacks’ are getting bigger every year

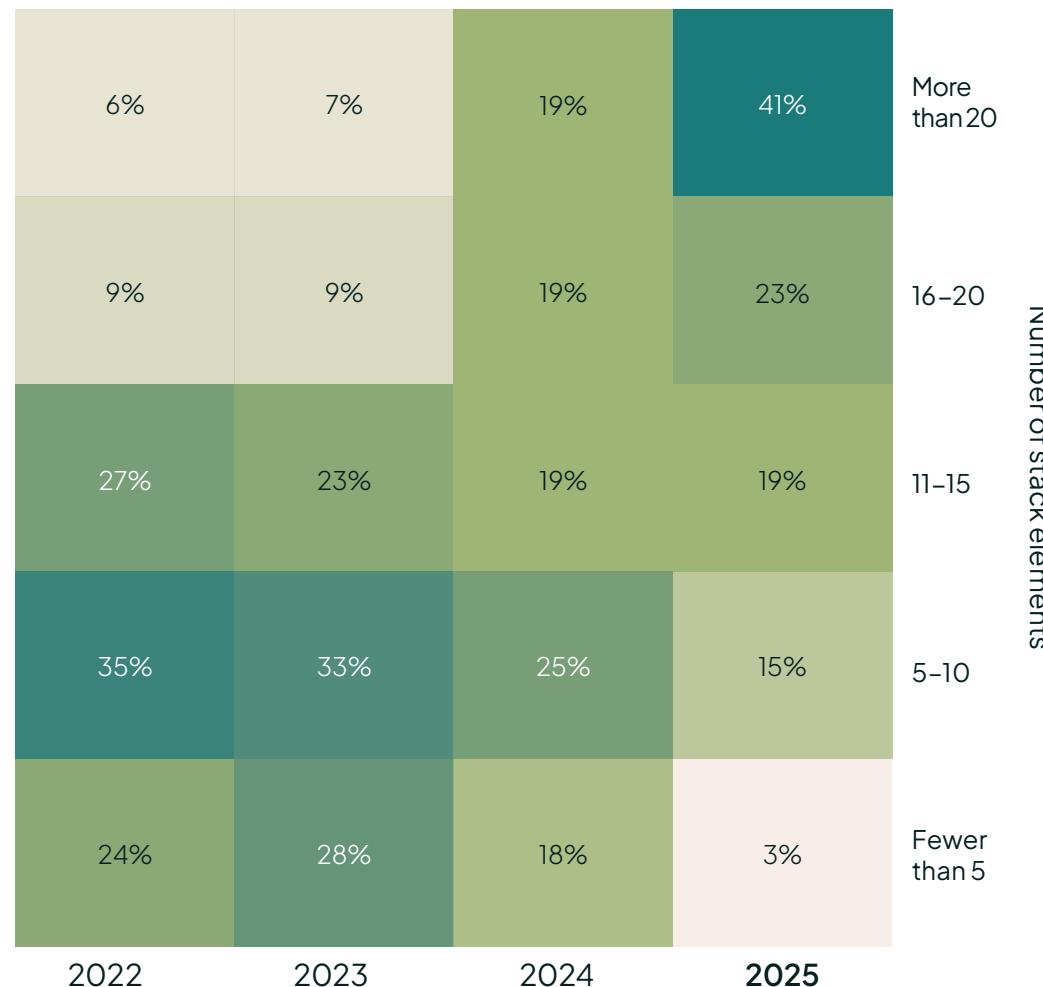


Figure 4: Approximately how many distinct software elements (open source or commercial) make up the complete platform ‘stack’ in your production Kubernetes infrastructure? E.g. monitoring tools, logging, service mesh, CI/CD, secrets management, ingress, etc.

# What's driving deployment choices?

Every cluster lives somewhere, of course, and as last year, we asked about where our respondents were placing their K8s clusters.

The answer is almost uninteresting, because it's '**everywhere**'.

All three major hyperscalers. On prem. Edge. Airgap. GPU clouds. Sovereign clouds. The average respondent has clusters in more than five different environments this year, as Figure 5 shows.

Statistically, every extra environment ticked in our survey strongly correlated with increased cluster count and increased node count. Sprawl and scale go hand in hand.

We asked respondents to pick from a list of factors that might be driving their choice of environments, and the answers explain a lot. Figure 6 shows the percentage of respondents that picked each factor as "very important", focusing on the top three.

6 in 10 said their **multicloud** strategy was very important, which drives up adoption across all cloud providers.

Indeed it was striking in our qualitative interviews this year just how many companies, even in regulated sectors, were cloud first, massively growing their cloud use particularly in managed K8s services like EKS and AKS, and aware of the value of doing so — but also acutely aware of the risks and tradeoffs.

Turn the page to read in their own words how they feel about it!

Figure 5: A typical org now runs clusters in 5+ clouds and other environments

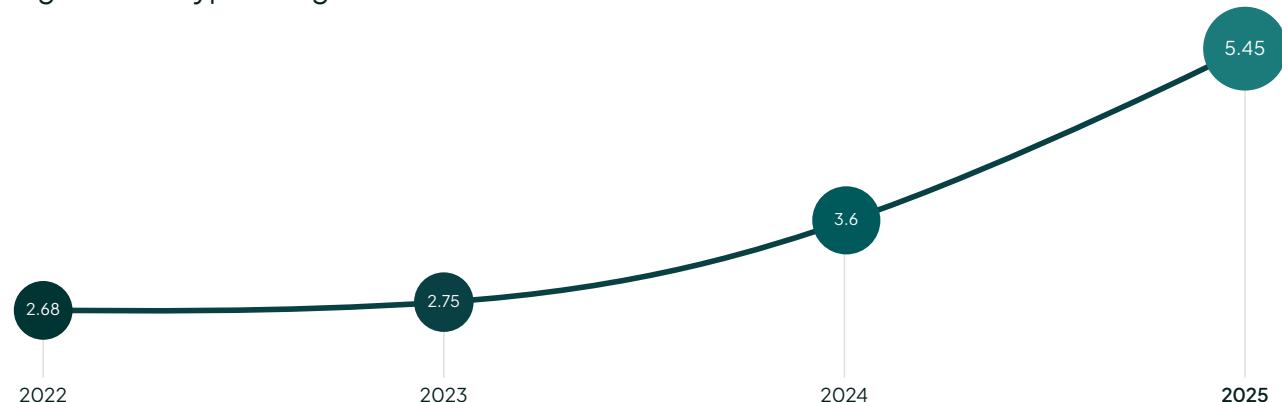


Figure 5: Which of the following environments are part of your organization's production application infrastructure for Kubernetes? Count of environments chosen.

The second most popular driver was, ironically, on-prem **repatriation** of cloud workloads for cost reasons — which drove up the adoption figures for edge, bare metal and airgapped environments in our data this year. We'll look more at the cost challenge a little later in this section.

And **AI** makes its first appearance in our report, driving the homing of workloads perhaps at the edge, in specific cloud services, GPU cloud, or on-prem DCs. Again, we will discuss AI at length starting in Section 2.

Figure 6: Cluster placement is driven by multicloud, on-prem repatriation, and AI



Figure 6: To what extent are the following factors important in your decisions about where to deploy your Kubernetes clusters? "Very important" only.

## Multicloud perspectives

**"We primarily use a mixture of Google Cloud and Microsoft Azure. It's an active choice: the relationships that we have with these cloud providers allows for significant discounts that we are able to obtain as a larger organization. But also, I would say that security has been one of the key reasons why we consume a lot of cloud type infrastructure, especially in our highly regulated space. Azure, for instance, provides a lot of native tools to integrate with your legacy authentication systems."**

Cloud engineering manager, US health and pharma company

**"I've been using Kubernetes almost since it started. In my last company we were on bare metal, we went through OpenShift, Cloud Foundry/Tanzu, then EKS. And now at my current company we're all EKS, managed Kubernetes — but now we are looking at edge Kubernetes, which will be bare metal again."**

Lead DevOps and cloud engineer, US manufacturer

**"We require more and more multicloud adoption. The drawback is that to port across different clouds is a challenge. Within Google, everything works like magic, but if I do anything on AWS or Azure, there are a lot of portability challenges. And I think slowly, that is pushing us to make a choice between one or the other."**

Senior director DevOps and cloud infrastructure, UK telco

**"I'm the CIO and CTO; every bit of tech rolls up through my 70-person team. We're moving from on-prem to the cloud because I don't want to run data centers anymore — the HVAC, power, fire-suppression, hiring server folks — better to pay a monthly fee and let the provider handle it. Cloud also gives us stronger security, cybersecurity, and backup. We're two years in, about 30–40% cloud-based, and since we're heavily Azure, using Azure's native Kubernetes and security tools just makes sense."**

CIO, US public transit agency

## The boom is back, baby

Last year we found our respondents were cautious about expectations for growth. This year, the enthusiasm has bounced back — despite all the global economic and geopolitical uncertainty happening while our survey was open in May 2025.

Figure 7, over the page, shows what percentage of respondents predict 'some' or 'strong' growth in each aspect of their organization's Kubernetes use, compared to similar lists going back to 2022. (Not all options were asked in each survey, so if you see a stubby line, it's not a design error).

Predictably, 90% of respondents said that their use of AI workloads on K8s is going to grow — how could they not? But we saw a return to stronger growth expectations across many other areas too. Look at all those upward-trending lines!

Even after several years of production adoption, our respondents still have work to do to bring more applications and dev teams into K8s, which necessitates expanding cluster infrastructure accordingly.

Figure 7: AI workloads are the fastest-growing use of Kubernetes

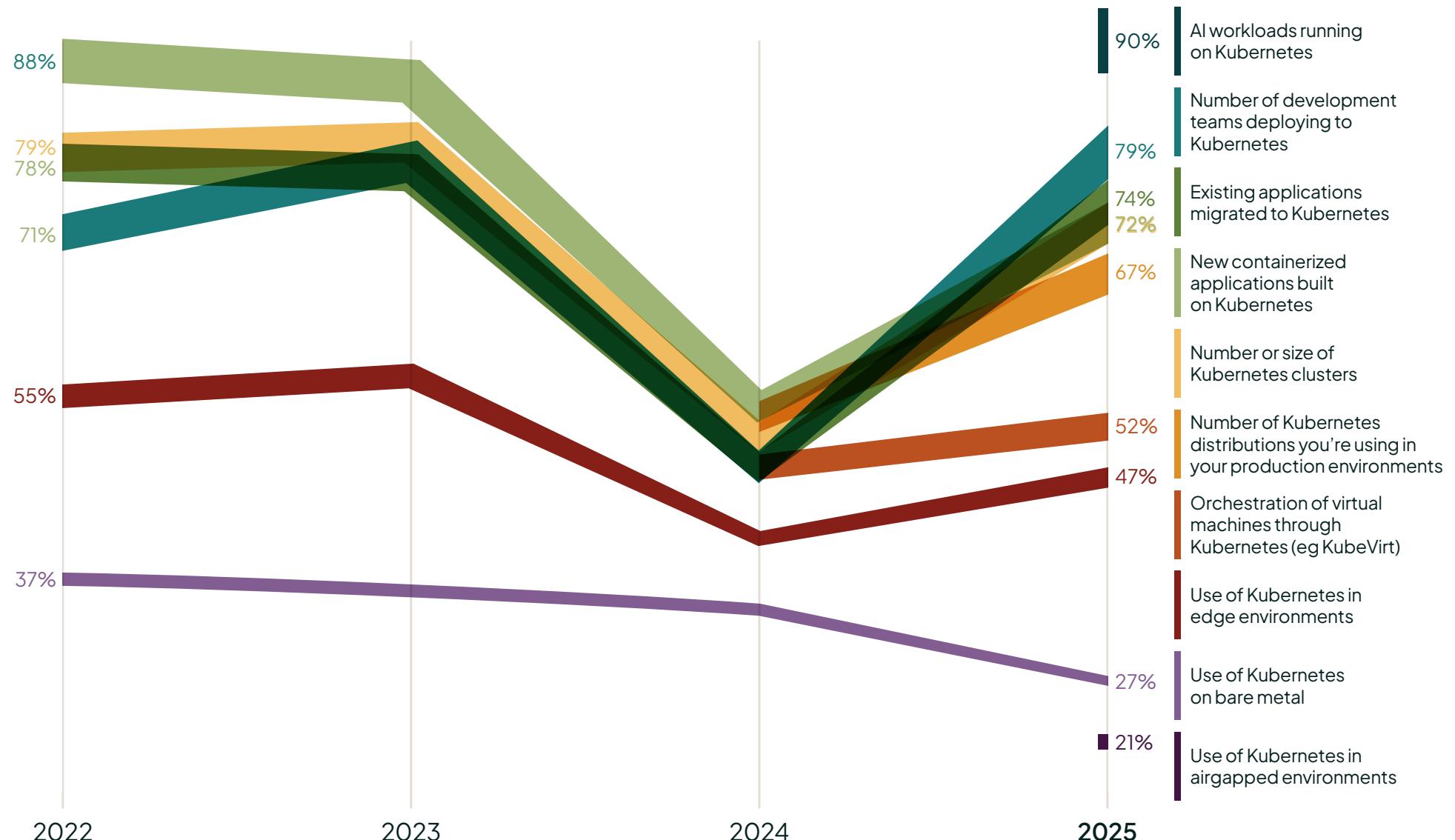


Figure 7: For each of the following, please indicate how you expect your organization's Kubernetes use to change over the next 12 months. Percentage of respondents picking some or strong growth.

# With all this growth and confidence, what about pain?

Mirroring the more enthusiastic growth story, the number of challenges our respondents reported has dropped significantly this year.

As Figure 8 shows, only three challenges this year were picked by more than 40% of respondents:

## Tell me where it hurts

"We run 200-plus clusters across the globe, and as that number grows the challenges grow too. The biggest issue is consistency and control: how do you keep a single pane of glass for the control plane and a holistic view of every cluster? We have to stay sharp on FinOps, keeping spend under control."

Senior technology transformation executive, US financial services

"Security is where we feel we're behind: multi-cluster, multi-tenancy, RBAC audits, GDPR — there's a steep learning curve and we rely on third-party tools right now."

Senior director DevOps and cloud infrastructure, UK telco

"My team of five is the only one at the entire company with any K8s experience. We're spread thin, so it was easier to go through a cloud-managed service. The problem is, it was all so new to the application teams too, so we've slowly expanded as we build more comfort with the development teams."

Lead DevOps and cloud engineer, US manufacturer

- **We struggle with the cost of our Kubernetes environments and managing it** — which makes sense, since usage is up, and cost unfortunately scales with usage.
- **We struggle to keep multi-cluster deployments consistent across diverse environments** (bare metal, cloud, edge, etc.) — which makes sense, since more environments are in use, as we discussed in Section 1.

- **We struggle to properly protect ourselves against security breaches** — which makes sense, because everyone always picks security, let's be honest.

Skills and headcount came in fourth. While K8s may not be 'new' by any definition, if we move outside the cloud-native bubble it's still seen as a challenging and relatively niche technology, and not only are there few real 'experts' out there, they command a premium salary. A tale as old as time.

Figure 8: Cost and multi-environment consistency are the top K8s challenges

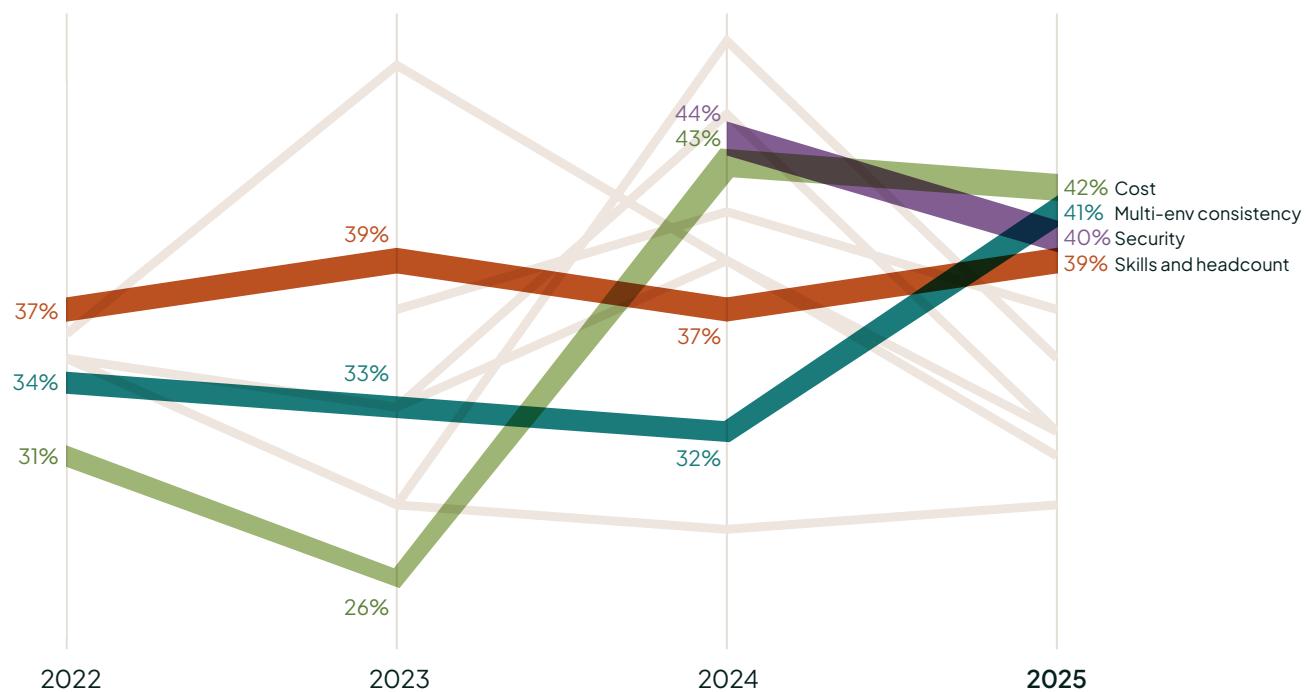


Figure 8: What challenges does your organization face with using Kubernetes in production? Check all that apply.

# Kubernetes cost: a solved problem or an AI daydream?

So cost is the most commonly cited challenge this year. That's the headline, but that's not where the story ends. Let's stitch together some data points.

- Cost is the #1 challenge, picked by 42% of respondents.
- More businesses than ever (88%) say their K8s TCO has increased in the past 12 months. The delta here is eyebrow-raising: a 26pp year on year increase.
- More businesses than last year (64%) say the pressure on them to reduce costs has increased in the past 12 months.

Notably, C-suite and VP-level respondents were much more likely to agree that they face cost pressure and see cost increasing. They are the ones in the firing line from the CFO or board every quarter, fighting for budget.

But apparently organizations have been working on this issue since we last surveyed them: the group saying that they lack cost visibility has plummeted to 22%. This is the biggest year-on-year change of any question in our data set, and if we didn't trust our research agency so completely we would be tempted to write it off as an error.

Figure 9: Cost, and cost pressure, are up significantly



Figure 9: Please indicate your agreement with each of the following statements. Percentage of respondents agreeing.

## The cash vs time tradeoff

"The biggest drawback to Kubernetes is cost – especially with managed services. If you run your own cluster with something like an open source tool, whether on AWS, Azure, or anywhere else, you pay only for compute and storage. The catch is people and time: you need a team to handle upgrades, configuration, and day-to-day ops. Managed Kubernetes erases that operational burden, letting you move faster, but the service premium drives the bill up. It's a straightforward trade-off between paying more cash versus investing more headcount and effort."

Lead DevOps and cloud engineer, US manufacturer

# Betting on AI to drive K8s ops efficiency

So what's going on? Well, the evaporation of cost-visibility blindspots suggests our respondents have invested in some tooling from the burgeoning cost-optimization landscape.

And certainly when we ask about future plans, shown in Figure 10, only 36% say they have a real opportunity to invest in finops tools, down from 48% last year. It looks like plenty have already bought some.

What's left in the toolbox to fix the growing cost challenge? Well, many of the options we presented in Figure 10 saw a significant drop in interest... they're the ones marked in red.

But it's worth calling a few things: autoscaling remains important, despite 80% of our respondents saying they use it 'fully'. And notice that cloud spend and repatriation is hanging on in there in fourth place — remember that cloud repatriation was the #2 driver in cluster placement.

In fact, it seems that everyone is betting on AI to solve their cost problems: it was the only majority strategy picked for efficiency gains, squeaking in at 51%.

And in a followup question, a staggering 92% said they were already looking to invest in next-generation AI-powered optimization tools.

Figure 10: AI presents the single biggest opportunity to optimize K8s operations

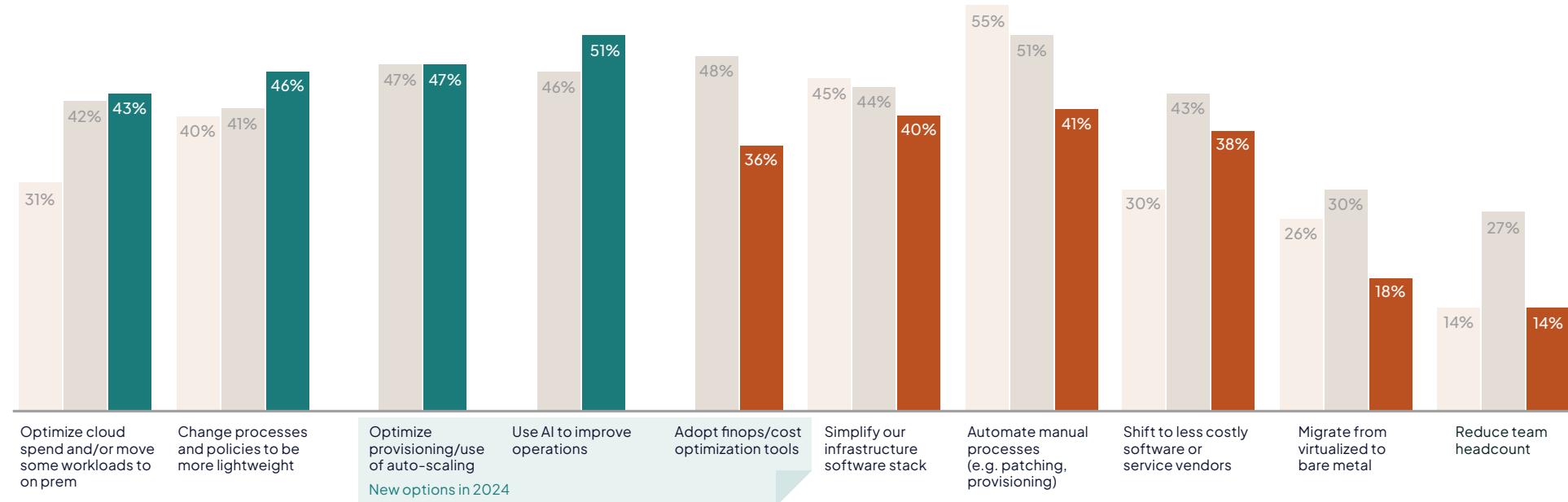


Figure 10: Where do you think your organization has the biggest opportunities to improve Kubernetes infrastructure operational efficiency? Choose all that apply. 2023, 2024 and 2025 data.

## AI pros and cons

But for all the AI excitement, notes of skepticism came out in some of our interviews, whether based on technical feasibility or socio-economic factors.

**"The biggest challenge is that application teams think they know what they want, but we end up with clusters that are not fully utilized. AI can definitely help. With natural-language prompts you could ask, how should we reconfigure and optimize it? AI could even generate the YAML manifests: describe your app — how many clusters, nodes, sizes — and it returns the recommendations. You'd chat with it and get the information you need. Could that become a Kubernetes co-pilot like today's AI coding assistants? Yes, absolutely — 100% yes."**

Senior technology transformation executive, US financial services

**"I don't see an AI co-pilot for Kubernetes happening any time soon. The platform has too many variables that sit outside system control."**

Senior director DevOps and cloud infrastructure, UK telco

**"We're heavily unionized, and my developers would see an AI Kubernetes co-pilot as a threat to their jobs. They might tinker with it at their desks, but if I said, 'We'll lay off half the team and let AI handle it,' they'd push back hard. A staff-replacement play is for a for-profit company, not a public agency like ours."**

CIO, US public transit agency

## Recommendations

As we close out Section 1... what can you take away?

Multicluster across diverse environments remains the reality for almost everyone. If that includes you, it's a foundational requirement for you to have tooling and processes that allow you to manage consistently in the face of diversity and scale.

The state of the art in cost optimization is always changing, whether 'traditional' finops and autoscaling, or AI-powered next-gen tools.

But don't expect miracles: you will need to assert your strategic perspective on the substrates you use (particularly pricey managed K8s services), your multitenancy approach, how you use autoscalers, how much overprovisioning you allow. These are matters of strategy and policy, not technology.

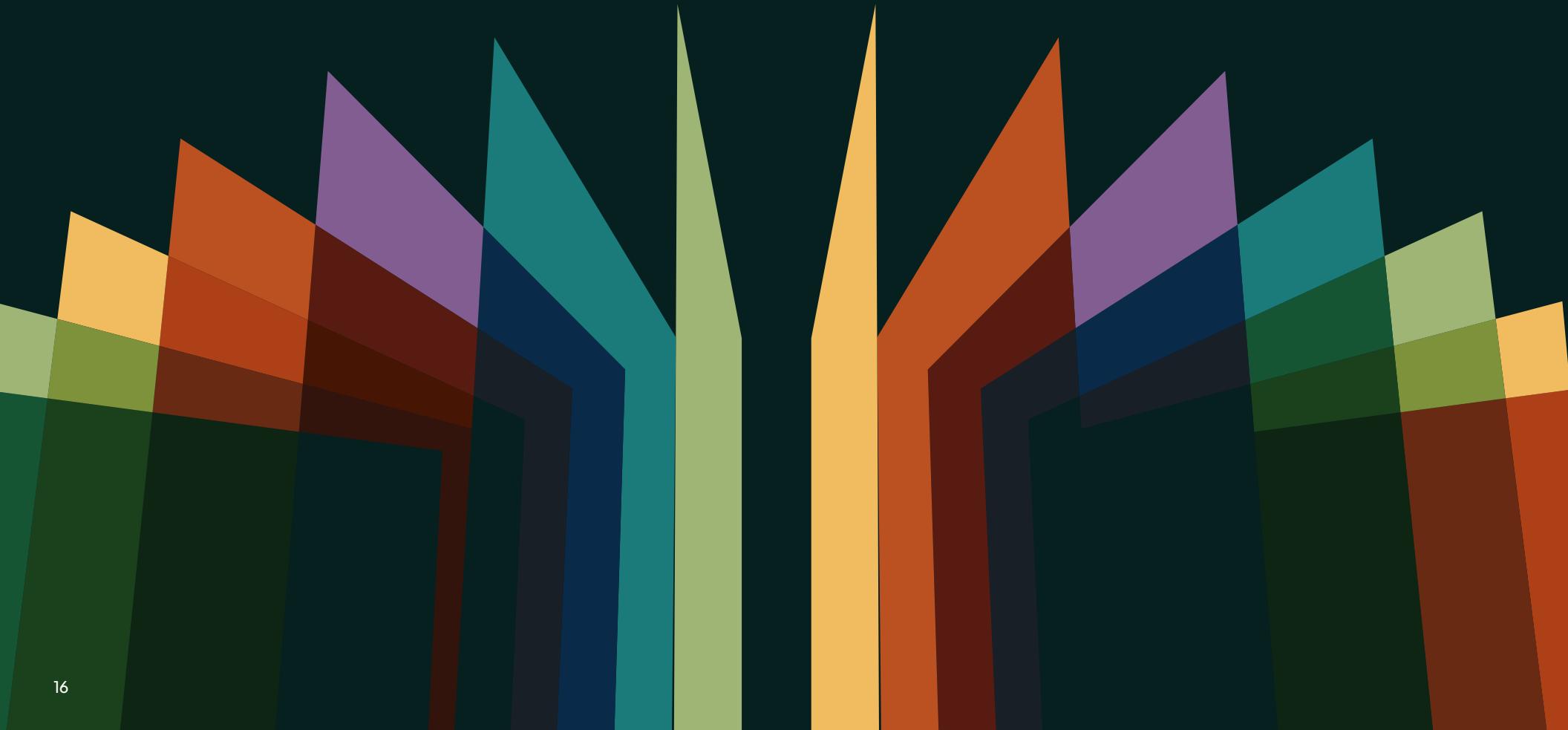
And remember: a huge portion of your true infrastructure TCO is in the human cost, time and salary. How can you show your business that you're both using your time and skills efficiently, but also working to multiply value across the business, eg by accelerating developer velocity?

If you're looking to justify investment to your CFO, don't just look at your cloud bill and license costs.



## Section 2

# AI driving everything, especially edge



# AI from hype to reality

We called it in last year's report, when 70% of our respondents said that AI is the "biggest trend in Kubernetes this year".

And now that's manifested into reality. We've already seen it in Section 1: **AI is the new center of gravity for K8s.**

- AI is the third most popular factor dictating cluster placement, with 28% already using K8s on dedicated GPU clouds.
- It's seen as the top solution for managing growing infra costs, and real budget is going toward AI-powered tooling.
- 90% expect to run more AI workloads on K8s in the next year, the strongest growth vector of any we asked about.

And it's these AI workloads that we're going to explore in this section.

Several of our interviewees talked about using AI workloads in their business, and the huge expectation on IT to make them happen. One called it a 'gold rush':

*"I've seen this first-hand: our CEO, board members, and the rest of the leadership team are staking the company's future growth on large-scale adoption of AI tools — deployment that delivers actionable, profitable business outcomes and opens new avenues for revenue. It genuinely feels like a modern gold rush. With more use cases emerging — especially in healthcare — I'm confident we'll keep seeing a steady rise in AI-driven workloads."*

*Cloud engineering manager, US health and pharma company*

While enthusiasm for AI is high in even regulated sectors like healthcare, not every organization is completely all-in:

*"Transit is a conservative industry, so we're wary of handing real control to AI — we don't want an algorithm turning every traffic light red. AI could draft board memos or procurement documents, sure, but I don't see it running day-to-day transit operations any time soon."*

*CIO, US public transit agency*

What many AI workloads have in common is that they are latency-sensitive and data-intensive, making them a perfect fit for... you guessed it, edge computing environments.

We predicted last year that AI would be the driving force behind edge K8s adoption, when **68%** of respondents said "the popularity of AI is driving interest in edge computing". This year proved the point.

*"AI inference workloads that demand real-time decisions — think autonomous vehicles — belong at the edge, as close as possible to the data source, because millisecond-level latency is non-negotiable. Our regional cloud-hosted Kubernetes clusters handle most applications fine; a 1- to 2-second round-trip is usually acceptable. But for low-latency inference, edge-deployed Kubernetes clusters are still the best fit."*

*Cloud engineering manager, US health and pharma company*

# Edge: steady growth, fueled by AI

Indeed, it's been a good year for edge Kubernetes in general, according to our data.

More people say they're using edge in production than ever before: 50%, up from 38% in 2024.

And if we total up all stages of live adoption, including pilots, 62% of organizations are engaged in using edge in some form.

Scaling edge is a journey. The majority of adopters already have more than 50 edge clusters; 4% have more than 1,000.

As to the future, respondents are quietly confident.

Looking at those that use edge today, 81% expect growth in their edge environments over the next 12 months. 26% say that growth will be 'strong'.

For four years, we've asked what challenges edge adopters face when using it.

This year, the story has shifted a lot, as Figure 12 over the page shows.

Security is still prominent, of course, but many familiar old challenges have fallen away, and we see a new trifecta of challenges coming to the forefront.

Figure 11: Half of all orgs use K8s at the edge

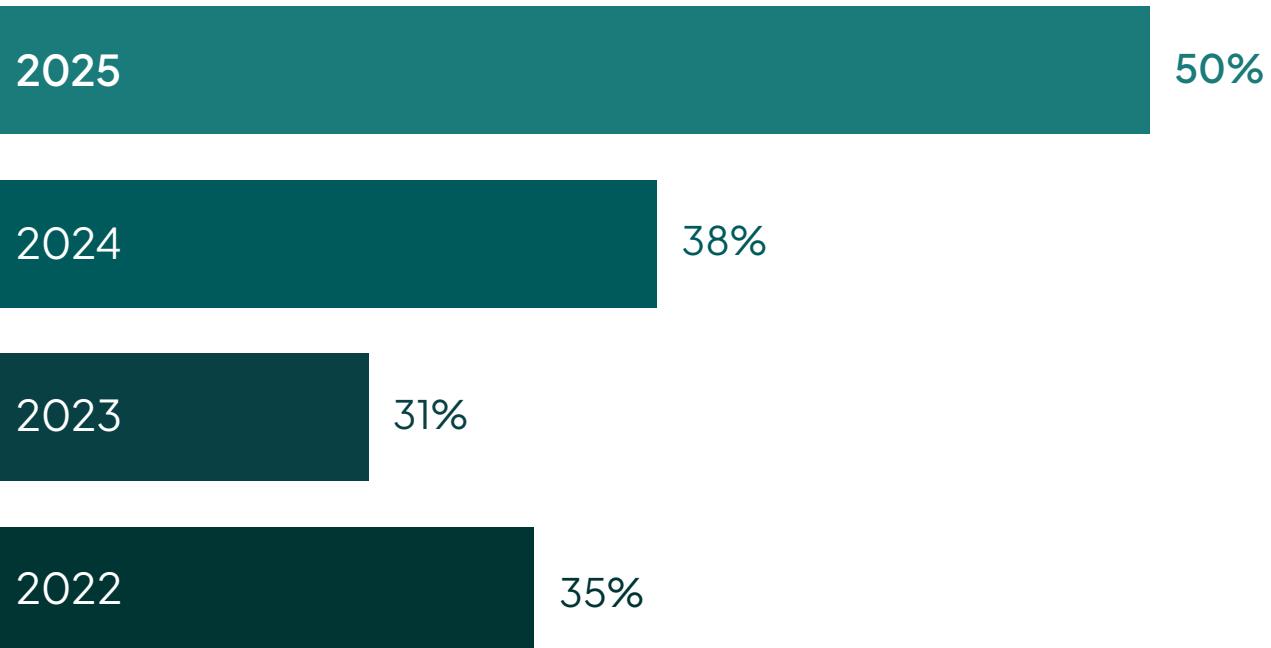


Figure 11: Which environments are part of your production K8s infrastructure? Choosing edge.

"Edge Kubernetes looks like the biggest opportunity.

Two years ago we even put a small cluster on our service trucks for the DOT, and now we're evaluating similar setups for factories.

Tiny, offline Kubernetes nodes are popping up everywhere to run local apps in tight spaces. In short, the edge is where Kubernetes will expand fastest."

Lead DevOps and cloud engineer, US manufacturer

Figure 12: Edge challenges evolve, reflecting the needs of AI workloads

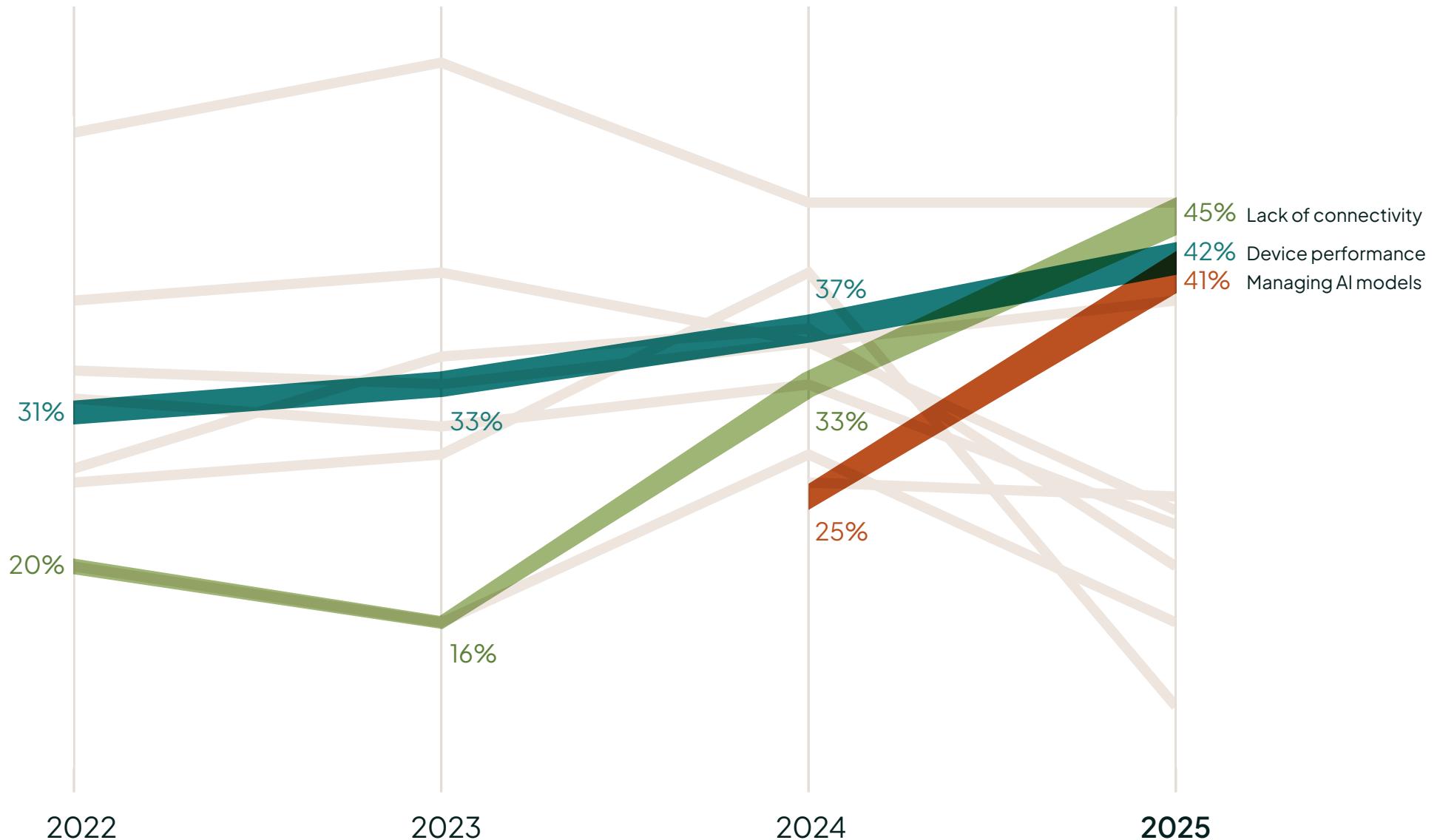


Figure 12: To the best of your knowledge, what is challenging about deploying and managing Kubernetes on edge devices?

# AI at the edge: what's the pain?

There's a relatively simple explanation for this big swing in edge challenges. When we talk about edge, we're no longer talking about miniature boxes for limited IoT telemetry; performance and connectivity are paramount when you're running beefy workloads like computer vision (which edge analysts [STL say is the #1 edge AI workload](#) — and that tracks with our own experience in the field).

So put those four top challenges together — security, lack of connectivity, limited performance, and managing AI models — and they create a perfect storm around the connected edge for intensive AI workloads. Our interviewee from a US manufacturer sums it up nicely in the box below.

We asked a followup question: what's challenging about running AI at the edge? And the answers here back up our thesis. Aside from security, it's about how you get models to the edge, and update them there.

“We already have an AI project — an LLM-based assistant — running on Kubernetes at every mill.

Because of what it does, we enforce much tighter security and more intricate networking than on our typical apps, and we have to dramatically update our nodes for it, loading them with GPUs and extra memory.

Performance is really the biggest issue that I see for AI and Kubernetes. With AI it's completely different from normal K8s workloads. It's not scale, it's power. LLMs need sheer horsepower, so we've spent a lot of time tuning container settings, node profiles, and even introducing a bit of unavoidable state to hit the speeds our users expect.

Until Kubernetes bakes that GPU awareness in by default, we'll keep investing engineering effort to make it work.”

Lead DevOps and cloud engineer, US manufacturer

Figure 13: Edge AI challenges are all about moving models

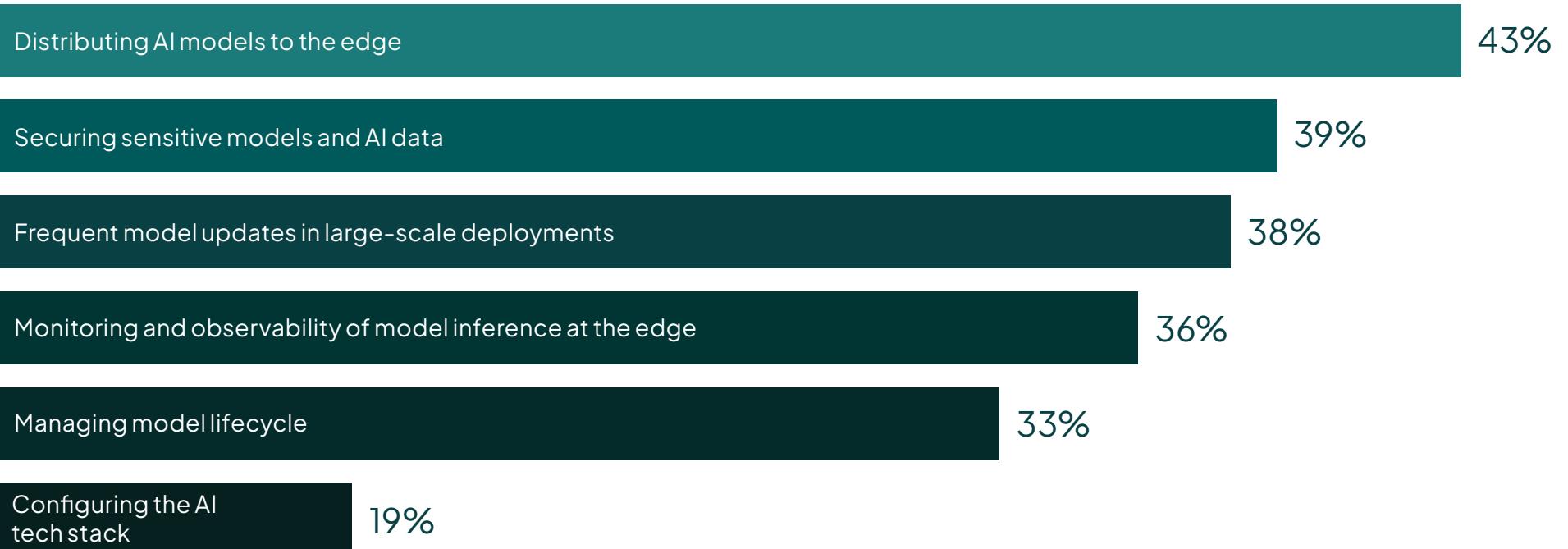


Figure 13: Thinking about your use of AI workloads in edge environments, to what extent have you faced the following challenges? Percentage saying major or minor challenge.

## Recommendations

It's easy to get swept up in AI hype. Whatever your personal perspective, one thing is sure: your business stakeholders will be applying pressure on you to deliver AI apps for users. And then it becomes an infrastructure problem of sizing, siting and optimizing clusters to suit the needs of each application.

As a platform team, your focus is on keeping pace: making sure the right infra options are available, whether that's a GPU cloud or edge boxes, while controlling costs. If edge is in the mix, don't fall into the trap of treating it as just another cloud. There are very real technical and economic differences in areas like scale, security, availability, the operational impact of degraded connectivity, and distributed large-scale deployments.

## Section 3

# Workloads on K8s and legacy modernization with KubeVirt

# What about the VMs?

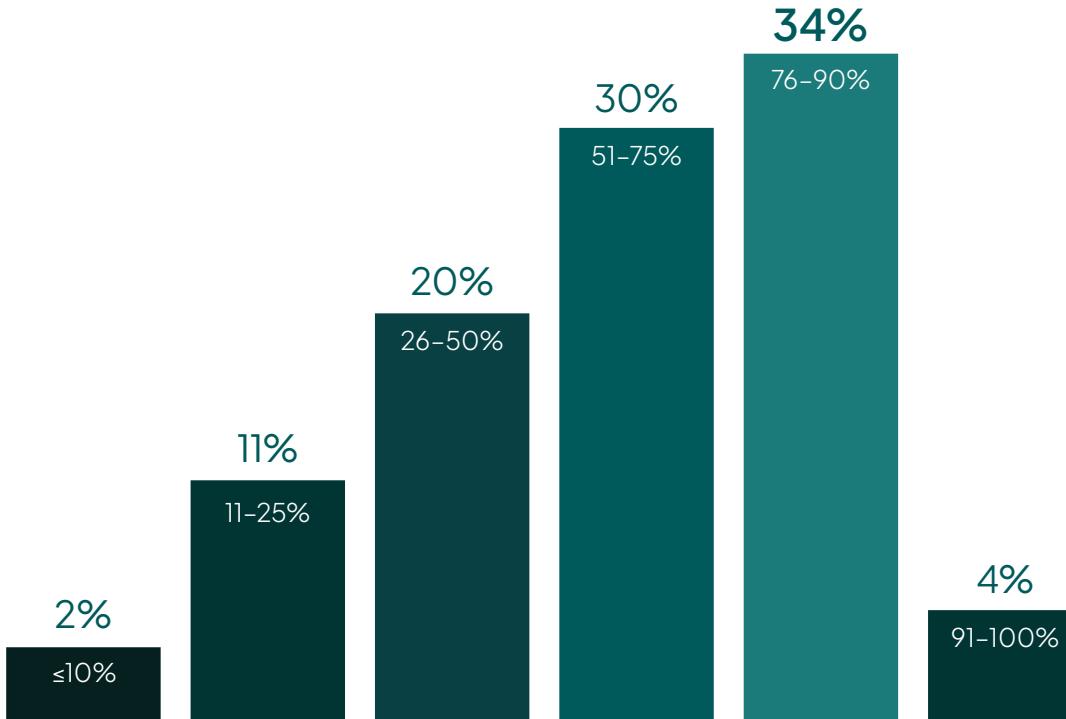
In Section 1, we noted in passing that our respondents already run the majority of their workloads in Kubernetes, with larger orgs leading the way. Figure 14 shows the detail.

From our earlier question about anticipated growth, we know that most organizations expect to be bringing more new apps, more existing apps, and more dev teams on board into this cloud-native nirvana in the next year.

But as the low-hanging fruit runs out after five+ years of adoption, there's one area that all our interviewees called out as a focus area: legacy applications.

Figure 14: Most orgs have more than half their workloads in K8s

Figure 14: What percentage of your workloads run in Kubernetes today?



So, if you really want to run almost all your workloads on K8s (as our interviewees below say they are heading towards)... well, what about those tricky legacy VMs?

Ever since Broadcom made waves with its acquisition of VMware through 2022 and 2023, we've been asking respondents about their strategies, in different ways.

In previous reports we investigated the pains and concerns our respondents felt, and we know that many are hurting and, frankly, not huge fans of Broadcom.

"We've been moving workloads to Kubernetes for about four years, with most progress in the last three. We're still early – only a modest slice of our estate has migrated, but it's a solid start. In principle we can migrate everything, but only when teams are ready and have bandwidth. Many of our developers come from a traditional engineering background, so adopting cloud-native tech like Kubernetes takes time. It's the legacy application teams that are harder to nudge."

Lead DevOps and cloud engineer, US manufacturer

"I lead our global cloud-transformation program – 5,000 people modernizing infrastructure and moving workloads off-prem. The target is about 80% of workloads on Kubernetes within the next two years. Net-new apps and anything with a clear ROI are already on, or moving to, K8s. But roughly 20% – legacy mainframes and other complex systems – will stay where they are because the effort outweighs the return. Realistically we'll land in the 75–80% range."

Senior technology transformation executive, US financial services

"At least half our workloads run in K8s today, and it's growing year on year because of telco digital transformation, edge computing, and now AI. There are a lot of containerized app workloads, and K8s is becoming the de facto. The industry – and we along with it – is shifting away from VMware. Our modernization projects now mean rewriting everything for Kubernetes, and my VMware spend has already fallen as we keep migrating off the platform."

Senior director DevOps and cloud infrastructure, UK telco

# KubeVirt becomes a mainstream option for legacy workloads

This year we've moved on to exploring organizations' strategies for solving the problem.

Are they going to stick with VMware regardless of the price hikes? Find a direct replacement like Hyper-V or Nutanix? Lift and shift to the cloud? Finally refactor all those VMs to containers? Or (and this should be no surprise to regular readers), migrate VMs to run in Kubernetes?

Here's the headline: as Figure 15 shows, 31% are planning to shift VMs to K8s (42% among larger organizations).

52% expect their orchestration of VMs on K8s will grow in the next 12 months — that rises to 71% for larger organizations.

In our view, these numbers are consistent with [other surveys from vendors such as Portworx](#), or even a little more cautious. But it reflects accurately what we're seeing in the market when we're meeting with enterprise CIOs. A migration like this is not something to leap into, and IT teams are moving thoughtfully.

**"Some workloads will stay on VMs because moving them to Kubernetes doesn't give us enough ROI. For those cases we use KubeVirt so we can manage VMs and containers through the same Kubernetes API — giving us a unified view. We'll keep that limited to a few legacy or low-priority apps; rolling it out more broadly isn't on the roadmap."**

Senior technology transformation executive, US financial services

If you're planning to migrate VMs to run directly in Kubernetes, there's really only one core technology in the CNCF landscape to enable it: **KubeVirt**.

So to support the strategic questions of how to handle legacy VMs, we asked specifically about awareness and use of this tool.

Most (86%) of the people we surveyed claim to have at least heard of KubeVirt (see Figure 16). 26% are using it currently in production in some form, with another 5% having used it in the past.

As you might expect, some kinds of businesses skew much higher on these questions: more than half of large enterprises say they use KubeVirt today, and it's 45% for those that have 76–100% of workloads already running in Kubernetes.

They're the ones with the fewest VMs left over, and they really want to get them dealt with!

Figure 15: Orgs choose a range of options for handling their VMs

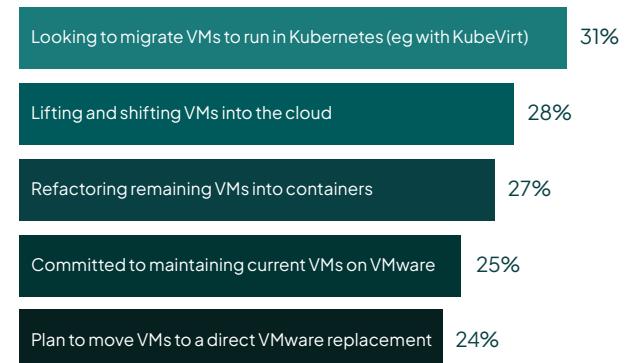


Figure 15: Which of the following best describe your business' strategy towards VM workloads?

Figure 16: Most have heard of KubeVirt, many have used it already

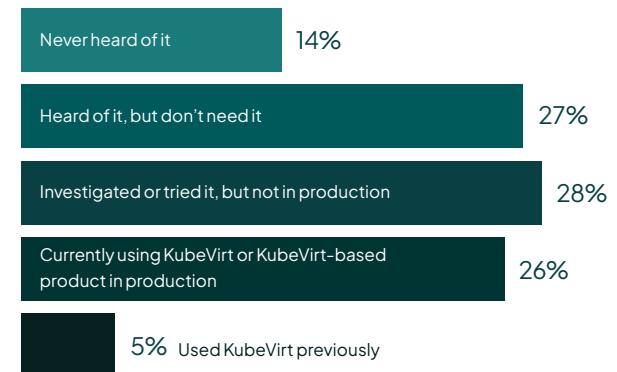


Figure 16: KubeVirt is an open source project for running VM workloads on Kubernetes clusters. What is your experience with KubeVirt?

# Different situations, different strategies

Figure 17 shows the answers to the two questions we asked on the previous page, and one more:

- Do you plan to migrate your VMs to Kubernetes?
- Do you already use KubeVirt in production?
- Do you expect growth in orchestration of VMs through K8s in the next 12 months?

...cut by large vs small organizations, and the share of their workloads already running in Kubernetes. The data is crystal clear (although the chart might take a little reading).

Those with most of their workloads already in K8s, and larger organizations, are much more likely to be using KubeVirt — the top-right quadrant shows 54% are already using KubeVirt (they're also more likely to expect 'strong growth' in their use of KubeVirt).

But it's actually those large enterprises with **less** of their app estate in Kubernetes that are most keen to migrate to Kubernetes: they're the ones likely facing the biggest pain from their VMware renewals. Perversely, those that have already slimmed down their VM estate by adopting containers in K8s over the past years are already much less dependent on VMware and feel less urgency.

Figure 17: The biggest K8s adopters are betting on KubeVirt

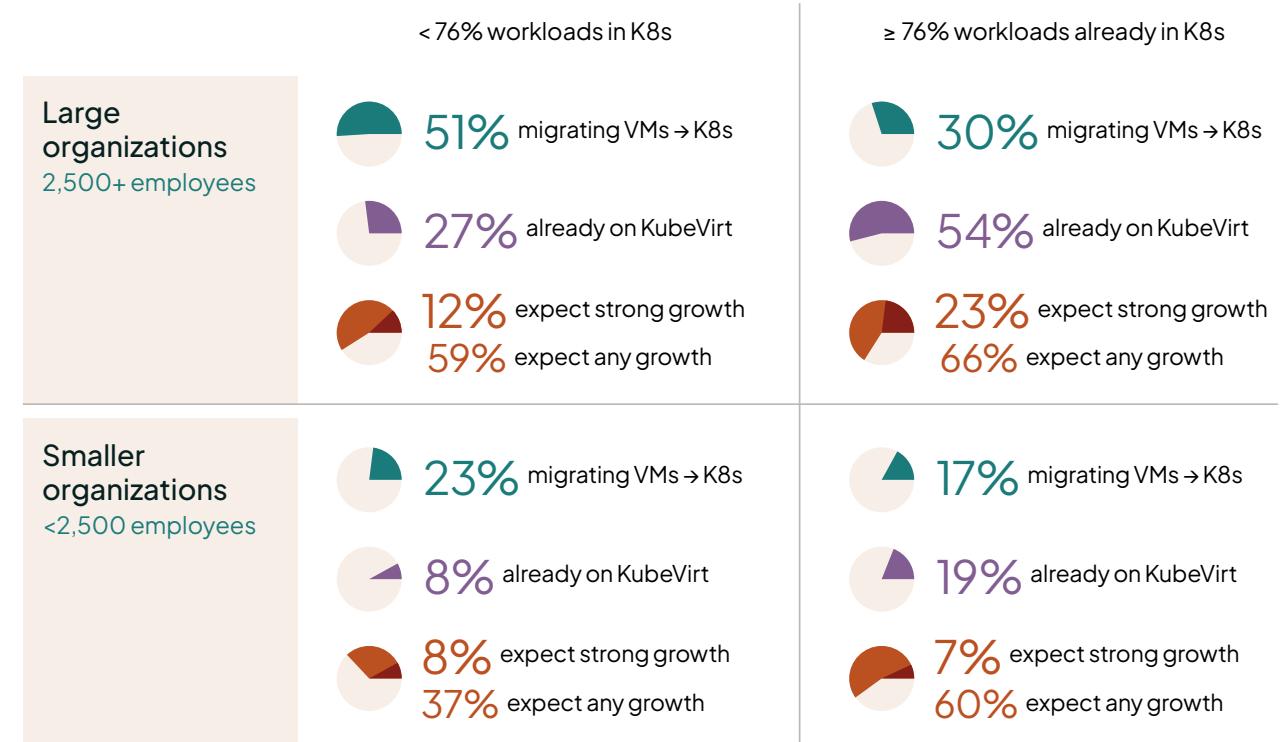


Figure 17: Comparing proportion of orgs migrating VMs to K8s, running KubeVirt in production and expecting growth in KubeVirt, cut by the % of workloads in Kubernetes and company size.

"When we do the assessment and say, 'Okay, this application cannot move from a VM to Kubernetes,' we normally have to pause. But recently a new possibility appeared — you can run the VM inside Kubernetes with KubeVirt. For us, that is a really big change, because even if migration isn't possible we now have the option to scale up and keep latency low. We are already using KubeVirt: when an application can't be moved from a VM to Kubernetes, we run the VM inside the cluster and it works fine."

# What's holding KubeVirt back?

The biggest obstacles to KubeVirt adoption are complexity with storage, and converting workloads, as Figure 18 shows.

This also tracks with our experience. Storage was one of the big focus areas when we built out our [KubeVirt reference architecture](#), and the workload conversion burden was why we built the [VM Migration Assistant](#), based on the open source Forklift project.

But let's not forget the factors of cultural resistance and lack of enterprise support, which came in third and fourth.

Many organizations have been using VMware for 20+ years, they have the certifications, they feel comfortable and safe — and their VM admins might dig their heels in about handing everything over to those new-fangled platform engineers. Especially when there's the enterprise agreement with VMware to fall back on in the event of an issue. VM workloads are seen as too important to leave to community support.

And indeed that feeling seems to follow through into KubeVirt adoption behavior. When trying KubeVirt, 76% use the open source project, to test the functionality — but only 32% take the open source through to production. Production deployments skew much more towards commercial KubeVirt-derived solutions.

Figure 18: Orgs struggle with KubeVirt storage and manual migration

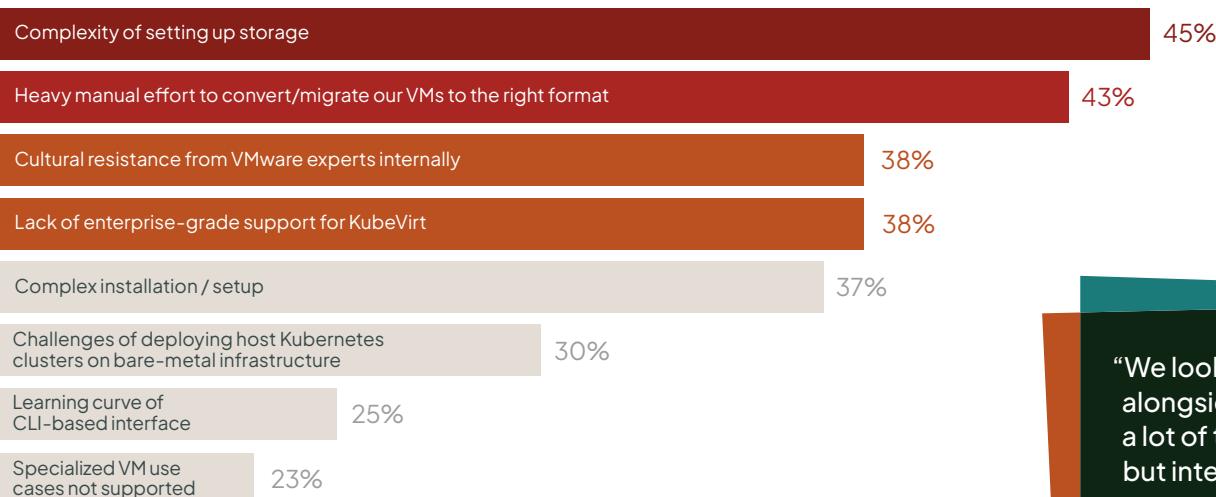


Figure 18: What challenges did you face when trying KubeVirt?

# Recommendations

It sounds trite, but it's true: the right strategy for handling legacy VMs depends on a lot of factors: technical, cultural, and contractual. Expect a marathon, not a sprint, as you triage your install base, and make decisions that limit future tech debt.

Our advice is a little biased here, but it's sincerely meant: think about whether you want to replace one proprietary VMware-style stack with another, or whether this is the opportunity for you to go cloud-native, KubeVirt-first — at least as much as your workloads allow. If you last checked out KubeVirt a couple of years ago, it's time to revisit and see what's changed.

And don't underestimate the people factor. We've had meetings with customers about KubeVirt migration where it was the first time that the VMware team and the Kubernetes team had been in the same room. If you're planning to change the foundations of your K8s clusters, expect resistance — you'll need to win hearts and minds if you want the change to go smoothly.

"We looked at KubeVirt to bring VMs alongside containers. It says it does a lot of things, and some do work, but integration still needs work."

Senior director DevOps and cloud infrastructure, UK telco

## Section 4

# Operations and outcomes



# What modern K8s operations looks like

## Not just what, but how

This year we wanted to look at not just what K8s infrastructure organizations operate, but how they operate it — and how well that goes for them. So we asked questions in a couple of key areas:

- We asked about behaviors: how apps are deployed, whether they had platform engineering and an IDP, how much they automate, what their upgrade strategy is, etc
- Inspired by the famous [DORA metrics](#), we asked about change frequency, error rate, recovery speed, patch speed, etc

By combining those questions, we can comment on how organizations are operating and ideally isolate repeatable behaviors that produce the best outcomes.

OK, first let's look at the inputs and behaviors. There are some interesting headlines here for those of you tracking the space or trying to benchmark your own organization.

## Platform engineering

In last year's study, 70% of our respondents said they had 'fully adopted' platform engineering roles and patterns, 29% agreeing 'strongly'.

This year we changed the wording a little, but the spirit was the same: **80%** said they 'have a mature central platform engineering function', with 32% strongly agreeing.

The traction that platform engineering is getting in the community is unmistakable, and it's clearly flowing through into organizational structure and hiring patterns.

## IDPs

As anyone who visits an event like PlatformCon will hear loud and clear, platform engineering is about much more than your internal developer portal/platform (IDP) — but nonetheless, an IDP is a centerpiece of how platform engineering works, giving a self-service, catalog-style experience for developers looking to consume platform resources.

90% of our audience said that their devs have self-service access to clusters through an IDP, 35% agreeing 'strongly'.

But note that not all platforms are created equal, and certainly just because a platform exists, doesn't mean it's widely used.

"Until about two years ago every request had to flow through a central team, but automation has since enabled an internal developer platform for self-service. Developers fill out a single interface — choosing expected traffic volumes, fail-over needs, latency targets, stateful components, and ingress/egress patterns. The system then offers a few recommended configurations; once they pick one, the required Kubernetes resources are automatically provisioned in 30–60 minutes."

Senior technology transformation executive, US financial services

# Cluster management practices: room for improvement

Perceptions don't tell the whole story, but we wanted to see how well our respondents thought they were keeping pace with the cloud native landscape.

We all know that K8s is updated three times a year, and every other piece of software in your clusters releases new versions, too.

78% said they stay close to the latest version of K8s and other cloud-native software – 38% agreed 'strongly'. Given how fast the typical release cycle is, and how challenging upgrades can be, this is impressive.

It's been well over a decade since we all became familiar with the analogy of '[cattle vs pets](#)' in modern infrastructure, and by now we hope it's well understood that there are many advantages to running consistent, interchangeable clusters instead of unique 'snowflakes'.

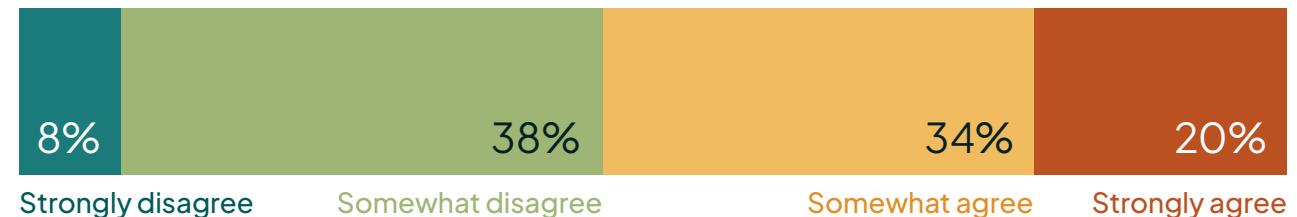
But certainly not everyone has reached that nirvana, as Figure 19 shows.

Your take on this data will really depend on whether you see the proverbial glass half full or half empty. But by the raw numbers alone, a majority of businesses still have snowflakes and manual, cluster-by-cluster management, by their own admission.

Does that seem like where we want to be after 5+ years of adoption, when small teams say they are stretched by growth and facing increasing cost pressure? Room for improvement, right?

Figure 19: More than half admit to manual snowflake operations

Our clusters are 'snowflakes' or 'pets', and have to be managed one by one



Much of our Kubernetes operational work is still highly manual

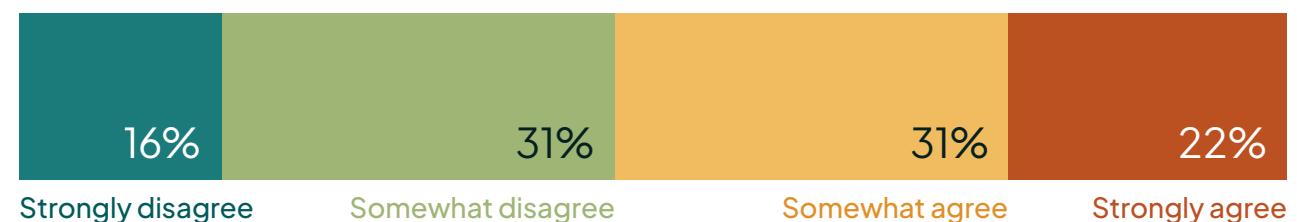


Figure 19: Which of these statements is true of your business's use of Kubernetes?

# Getting apps into the cluster: who does what?

We asked respondents to pick the most representative way that their organization deploys applications to Kubernetes clusters. As you can see in Figure 20, we got a wide spread of responses.

Note the groupings: in particular, do devs do the work or does a platform team do the work? Is the deployment automated or manual?

Figure 20: There's no consensus on getting apps into clusters

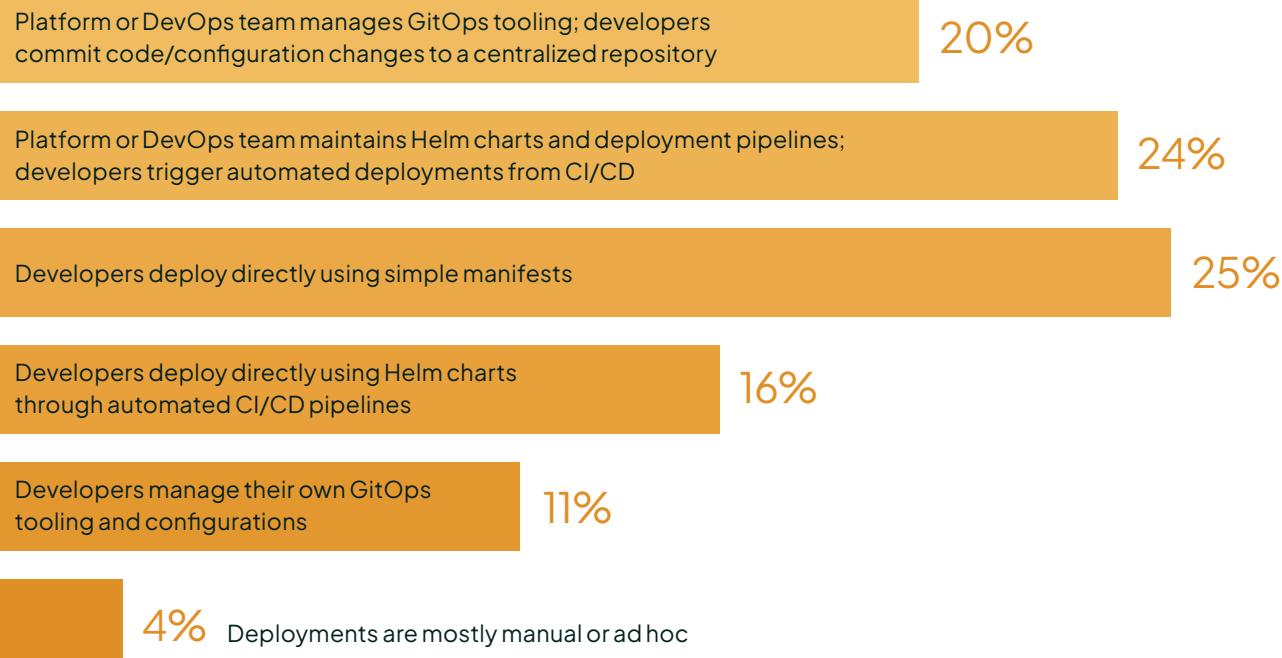


Figure 20: How are applications typically deployed to your Kubernetes clusters?  
Select the option that best describes your current practice.

# DORA metrics to the rescue

Now what about the outcomes?

This is where we bring out our version of the famous [DORA metrics](#). We didn't want to completely rip off the wording of the questions from the [State of DevOps reports](#), but hey, if it works and it's familiar, let's emulate, right?

We asked:

- **How often do you make changes in your production K8s environments?** We know that there's pressure to push new features and security patches — so we assume here that the more frequent the changes, the better the platform team is responding.
- **How long does it typically take from proposing an infra change to it being live in production?** Faster is better, we hope uncontroversially.
- **What percentage of changes result in incidents, rollbacks or degraded performance?** Lower is better, of course.
- **When a critical K8s infra issue occurs, how long does it take to fully restore service?** Faster is better.
- **If you were alerted today to an urgent patch needed to all your K8s clusters, how quickly could you have updated your entire infrastructure?** This was based on a real scenario that happened to one of our healthcare customers a couple of years ago.

Figure 21 over the page shows the data with a quick commentary.

Overall, these outcomes are clearly not perfect, but the modal answers to our questions sit generally within the standard deviation of the State of DevOps 2024 [industry baseline](#). That's a good thing.

**Frequency:** While we see a fairly standard distribution, half of respondents make changes to their production envs at least a few times per week, with nearly one in five changing every day. As you might expect, larger firms operating at greater scale make more changes.

**Time to prod:** Half take at least a week for a given change to go live. Larger firms with more distinct software elements in their clusters, and more clusters/nodes, claim to be able to go live quicker. This may be linked to the previous question — they're doing more changes in general, so either they're doing smaller/quicker changes, or they're doing it so frequently they have it down to a fine art.

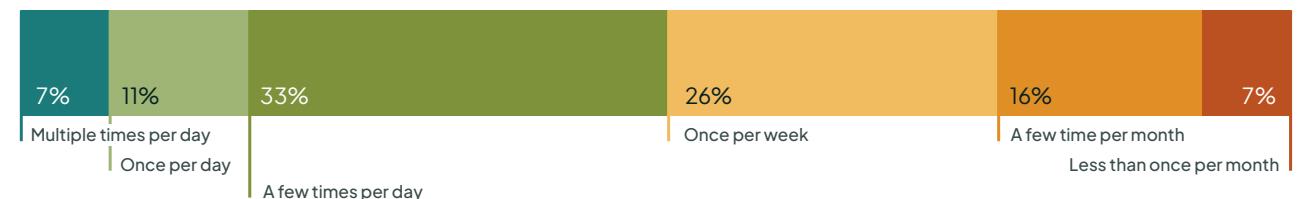
**Incident rate:** Nearly half of our respondents have incidents or must rollback more than 15% of their changes. Guess the staging environment isn't exactly like prod, eh? Again, larger organizations with more clusters/nodes claim better performance. Those staying close to the latest version of K8s are more likely to have a low number of rollbacks. (Looking at you, those folks who try to leapfrog multiple releases in one upgrade, despite [all the warnings in the release notes](#).)

**MTTR:** When infrastructure issues occur, it typically takes less than a day, and maybe less than 8 hours, for many to restore full service (or so they say). Again, larger firms with more nodes/clusters, plus those who stay close to the latest version of K8s, claim better performance.

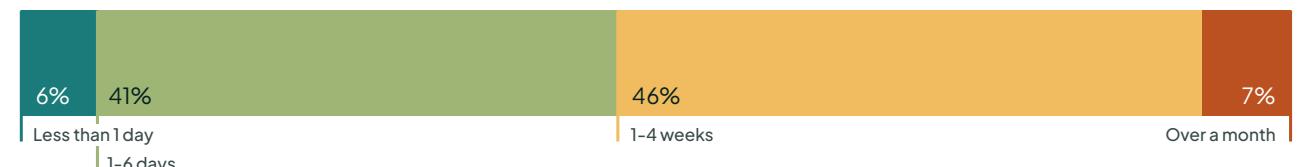
**Firedrill:** As to a major patch scenario, it's easy to read the data glass-half-full: 59% will be all done and dusted within a day. More pessimistically, we also see 15% would take weeks to months to patch their whole infra — leaving them vulnerable to exploit. Again larger, more complex businesses claim better metrics.

Figure 21: DORA scores show predictable performance

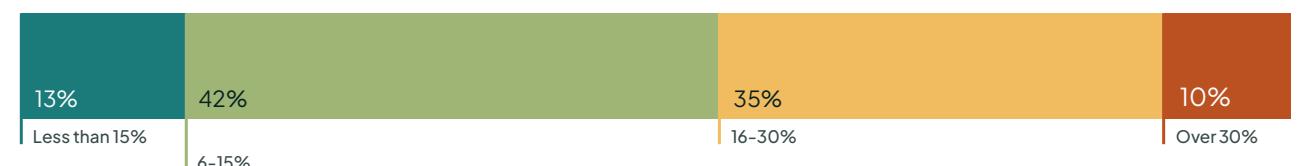
#### Frequency of changes in environments in production K8s



#### Typical time between proposing an infra change to live production



#### Percentage of changes resulting in incidents, rollbacks or degraded performance



#### Time to fully restore service after a critical K8s infra issue occurs



#### Time required to update entire infrastructure



Figure 21: Questions indicated within chart.

## Recommendations

It's hard to critically evaluate your own practices, even though we in IT pride ourselves on being dispassionate and data-driven. It's especially difficult in emerging areas like platform engineering, where not everyone truly knows or agrees what a 'good' IDP looks like, or how a 'mature' platform team functions.

Our advice: use the benchmarks in this section and compare them with your own outcomes and practices. Do they meet the needs of your business, your overall scale and appetite for speed? If the honest answer is 'no', maybe it's time for a deeper dive. Investing in something like a [\*\*'Path to Production' mapping workshop\*\*](#) may pay dividends.

Tooling is absolutely part of the solution — but as always, don't assume that implementing a new technology is going to solve your problem without work on culture, process, and team topologies.



## Which attributes produce better outcomes?

Now, let's get to the fun stuff. We can score answers to each DORA question out of 5 (fastest frequency, lowest failure rate, etc), and see how different behaviors affect the overall score — including the approach to how organizations deploy apps to their clusters. Figure 2 shows the heat map: teal is better, orange is worse.

It hopefully won't surprise you that those standardizing on centralized, pipeline-driven models win on engineering performance. Organizations where a platform or DevOps team curates the release artefacts (either Helm or GitOps) top every individual DORA metric.

GitOps managed by dev teams does not deliver the same results; it exhibits higher roll-back rates and slower MTTR, suggesting that central standards and guardrails matter more than the tool set itself.

Some other behaviors are correlated with higher DORA scores — in particular, being a large enterprise, having more clusters and more nodes, using K8s for more years, and committing more workloads to K8s — but nothing compares to the application delivery mechanic you choose.

Figure 22: Performance goes up when central platform teams own app deployments

Frequency of changes in environments	Time between infra change to live production	Changes resulting in incidents	Time to fully restore service	Time to update entire infrastructure	Average score	
3.1	3.8	3.9	3.8	4	3.7	Platform or DevOps team maintains Helm charts and deployment pipelines; developers trigger automated deployments from CI/CD
3	3.5	3.8	3.7	3.7	3.5	Platform or DevOps team manages GitOps tooling; developers commit code/configuration changes to a centralized repository
2.3	3.5	3.3	3.7	3.5	3.2	Developers deploy directly using Helm charts through automated CI/CD pipelines
2.3	3.4	3.5	3.5	3.4	3.2	Developers manage their own GitOps tooling and configurations
1.8	3.1	3.4	3.7	3.3	3.1	Developers deploy directly using simple manifests
1.9	3.2	3.5	3.6	3	3	Deployments are mostly manual or ad hoc

Figure 22: Average DORA scores across five measures plus mean DORA score for each answer to 'How are applications typically deployed to your Kubernetes clusters?'



Section 5

# The four types of K8s adopters

# Leaders and laggards

If you were paying attention in Section 4, you might have spotted that in this year's research, we're not just looking at the population of K8s users as a whole and reporting the headlines. There's diversity in the group, leaders and laggards, and they behave in very different ways. Population averages can only tell you so much.

We ran an analysis based on quantitative signals, like the DORA metrics and number of clusters, and found four clear, roughly equal groupings that bring nuance and meaning to the wider research. We (and our research agency) poked at the numbers, and they hold up.

We're not going to slice our data by these groups in all respects — this report is long enough already — but you'll see the trends as clear as day.

Meet the four personas:

## K8s Factories

These are the leaders in our research. They are experienced with Kubernetes, they operate at large scale across many environments, pushing into emerging areas like edge and KubeVirt. As you might expect, their biggest challenge is multi-environment consistency. They excel on all the DORA metrics and report fewer challenges than other organizations.

We tend to see the largest enterprises, finance, telcos and big technology firms in here; very few from healthcare or public sector.

## Mid-Scale Builders

These organizations are mid-size operators, working still at some scale — a median 36 clusters and 3,000 nodes — but across fewer environments, with generally lower DORA performance than the Factories. They are chugging along. They showed a lot of growing pains: they were more concerned with cost than anything else.

All industries are represented in here, but retail/CPG and media firms stand out.

**"In my field, local transportation and transit, I'd say we're on the cutting edge. I wouldn't compare us to, you know, a major technology company or a Wall Street bank or anything, but for transit and transportation, we're pretty advanced."**

CIO, US public transit authority.

## Modernizing Learners

Of all our respondents, these have been using K8s for the least time, an average of just under four years, and consequently are at much lower scale with, on average, 16 clusters and fewer of their workloads in Kubernetes. They're not yet using edge, or KubeVirt — they're investigating. Their DORA score outpaces the builders, perhaps reflecting their smaller, more agile scale and their lack of K8s legacy behaviors.

Healthcare, education and non-tech SMBs are featured here. 44% are smaller organizations.

## Old-School Statics

And lastly, the most problematic group, who we're calling Statics. They've been using K8s longer than learners and have put more of their workloads into it. But actually they have fewer, larger clusters, in fewer environments than anyone else. And their DORA metrics are worse across the board, particularly their incident rate.

These guys have got stuck, somehow — and maybe it's their strict regulatory environment or preponderance of legacy tech that holds them back. They called out skills and headcount as their top challenge, which may be a clue.

We see public sector, manufacturing and energy organizations in here.

Figure 23: Four types of K8s adopter have very different behaviors

	Median clusters/nodes	Years on Kubernetes	Median % of workloads on K8s	Mean DORA composite score	Median incident rate	Uses edge in production	Uses KubeVirt in production	Average number of environments where K8s is used
<b>K8s Factories</b> 	150 clusters 7,500 nodes	6 years	 83%	4.02 ••••○	6–15%	 85%	 75%	7.1
<b>Mid-Scale Builders</b> 	36 clusters 3,000 nodes	5.8 years	 63%	3.2 •••○○	6–15%	 56%	 38%	5.9
<b>Modernizing Learners</b> 	16 clusters 300 nodes	3.9 years	 38%	3.61 ••••○	6–15%	 7%	 9%	5
<b>Old-School Statics</b> 	8 clusters 300 nodes	4.6 years	 63%	2.64 •••○○	16–30%	 2%	 5%	4

Figure 23: Key distinguishing attributes of the four personas we identified, drawn from 12 different questions.

# What the four personas feel, believe... and do differently

Across our survey, these four personas answered questions very differently.

For example, when we asked about their VM modernization strategy (explored in Section 3):

- Factories push hardest on migrating VMs to K8s (KubeVirt) (chosen by 45%).
- Builders are split between K8s (34%) and refactoring lifts (28%), reflecting a dual-track strategy.
- Learners are the only group actively “still exploring”—11% haven’t committed to any path.
- Statics are wary of KubeVirt — only 15% plan to use it. Instead they will lift and shift to cloud.

The K8s cost challenge showed an equally broad spread. 96% of Factories said they had good visibility of K8s TCO, but only 66% of Statics and 61% of Learners said the same. Related, 92% of Factories use autoscaling, while only 45% of Statics do.

While everyone is interested in using K8s more for AI workloads, Factories were much more likely to expect “strong” growth over the next 12 months. 60% of Factories said so, vs 28% of Statics and 31% of Learners.

You get the picture: there’s a spectrum from Factories to Statics, and in all aspects the Factories are ahead.

We can get a clue about why Factories are able to achieve confident growth and strong results by looking at some of their capabilities and activities.

Factories are much more likely to say they have mature platform engineering, an IDP for devs, to stay close to the latest versions of cloud native tech, and to manage clusters as cattle.

This is also where the Learners vs Statics picture gets interesting. Statics have been doing K8s for longer, and they think they’re good at it — they’re more confident than Learners that they are automated and that they’re doing platform engineering, but as we saw above, it doesn’t translate into either better results, or greater scale and experimentation.

Figure 24: Each persona has a mix of different beliefs and behaviors

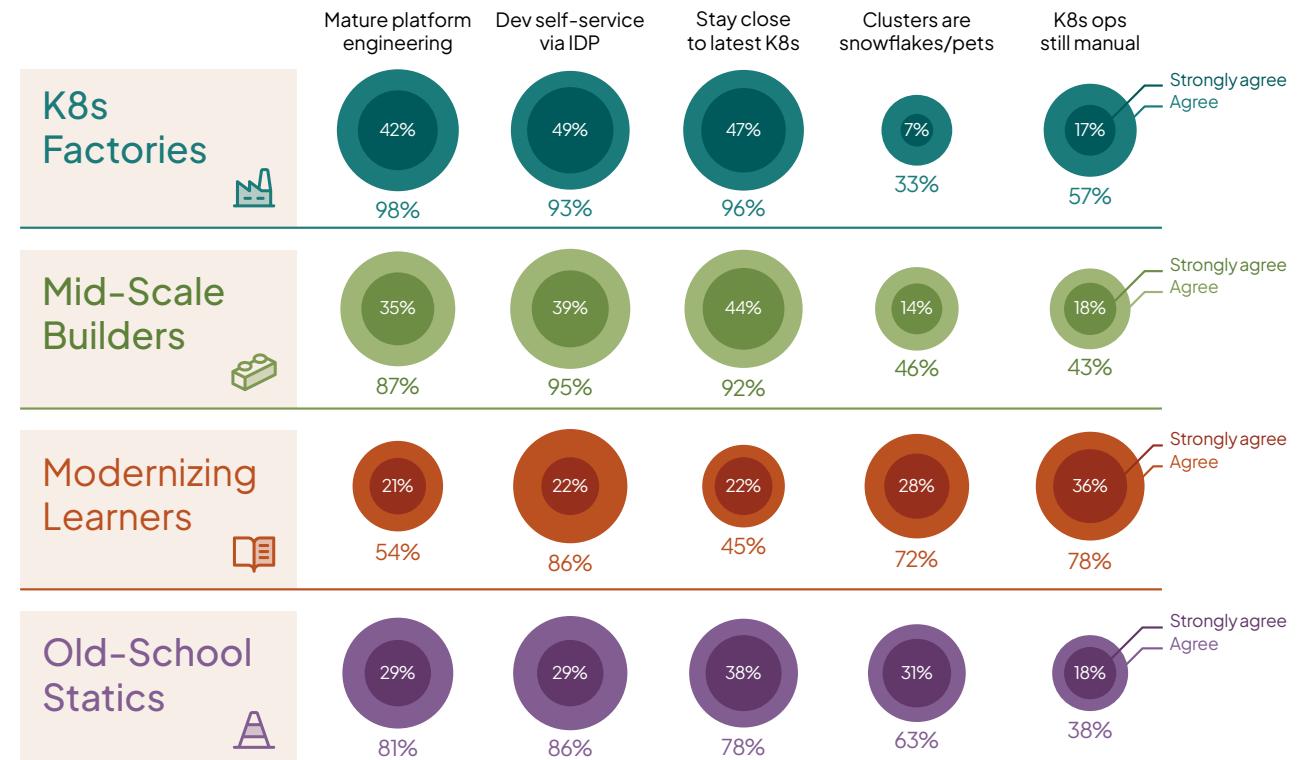


Figure 24: Which of these statements is true of your business’s use of Kubernetes? Cut by persona.

And what do we know about the way these organizations deploy applications? As you might guess, our top performers, the Factories, have overwhelmingly gravitated toward centralized, platform team-owned pipelines — whether that's Helm charts baked into a shared CI system or GitOps controllers managed by a dedicated platform team.

This is the killer best practice to take away, and it relates strongly to the DORA operating model: they're pushing many changes every week, so they can't afford manual hand-offs or per-team silos. By codifying every step into one golden path, they release on demand with minimal risk.

By contrast, the Old-school Statics still rely heavily on ad-hoc, decentralized workflows — think laptop kubectl commands and plain manifests. Their dozen or so clusters carry brownfield workloads where every deploy feels like a big event, so manual approvals and bespoke scripts persist.

The Mid-Scale Builders have been doing this a while, so they carry some baggage — not as much as the Statics, but enough to cause their poor DORA performance.

Finally, Modernizing Learners are still finding their way. They carry less excess baggage so they're more likely to be using CI/CD and have a central platform team, but there's still diversity and a lack of formal best practices.

So: which of our four personas do you see when you look in the mirror?

Figure 25: Top performers are more likely to rely on central platform teams to deploy

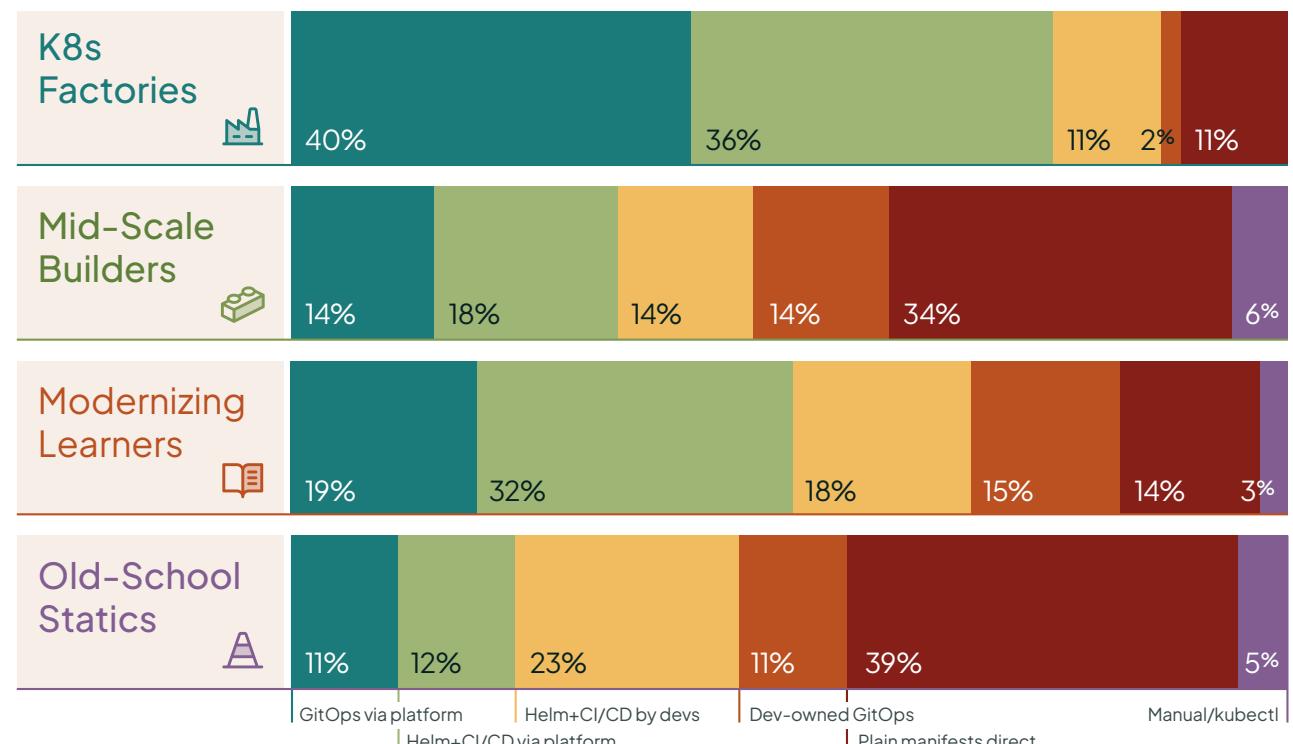


Figure 25: Which of these statements is true of your business's use of Kubernetes? Cut by persona

# Predictions for 2026



Every year, our CRO here at Spectro Cloud asks us to make some predictions in our research that we can test and come back to next year to validate. So this one's for you, Dave: eight predictions, with our logic for why it makes sense.

Agree or disagree? Share your own predictions with us on LinkedIn and we'll send the best suggestions some swag.

## Get in touch

If you want to discuss what the changing landscape of Kubernetes means for your organization, or how to get more value from it, our team is here to help.

Get in touch for a quick call to start the ball rolling at [spectrocloud.com/get-started](https://spectrocloud.com/get-started).

### 1. AI becomes the primary production workload on Kubernetes.

75% of enterprises will be running at least one mission-critical Gen-AI or predictive-AI service on K8s (many on GPUs/D GPUs). Why? 90% of respondents already expect growth in AI workloads on K8s in 2025, the single biggest “growth” signal in the study. Interviews confirm that execs are hinging their future growth on the AI goldrush, and are betting on K8s elasticity to host it.

### 2. The edge drives a cluster explosion.

By the end of 2026, 75% of enterprises will be using K8s at the edge. Why? Interviewees in every sector point out that edge is relevant to them, particularly in the world of AI. Each edge node is a tiny cluster, so fleet counts will soar.

### 3. 30%+ of operators run 100+ clusters overall.

Why? “Cluster-per-team” is already becoming normalized, and one in five already have more than 100 clusters. Teams and compliance love isolation, and the multi-cloud + GPU cloud + on-prem + edge trends mean that more clusters will sit in more places.

### 4. KubeVirt graduates from niche to mainstream (~50% live use).

Why? “Currently using” hit 26% in 2025 — half of large enterprises already use KubeVirt. VMware licensing turbulence and the imperative to unify VM+container ops keep the flywheel turning.

### 6. AI-powered copilot ops is in production at 1-in-3 enterprises.

Why? Ops leaders are already prototyping LLMs for cluster copilot use cases. The cost-pressure data (92% investing in next-gen AI optimization tools) proves the appetite. LLM-backed copilots will tune resources, troubleshoot errors, and build new clusters.

### 7. FinOps and policy engines get baked into every tool.

TCO blindness drops below 10%. Why? “We don’t have good visibility of K8s TCO” fell from 61% → 22% from 2024–2025. Vendors, cloud hyperscalers and open-source projects will finish the job with automatic cost attribution and guardrails.

# The end

At mid 2025, it's clear there are a few trends shaping how K8s is evolving in the enterprise.

The first thing to say is that the Kubernetes landscape continues to mature. Most respondents are seasoned pros and have the teams, processes and tooling to manage more clusters, in more environments, than ever before, with less perceived pain.

AI is everywhere you look: it's the driving force behind today's edge K8s requirements, it's the hoped-for solution to spiraling Kubernetes costs, and it explains the growth expectations our respondents shared.

At the other end of the spectrum, legacy workloads are becoming unignorable as platform teams run out of 'easy' workloads to put in Kubernetes. Many, particularly large cloud-native enterprises, are turning to commercial KubeVirt implementations to rehome their VMs.

Most are confident that they have adopted platform engineering, IDPs and other signs of modern operating models — but automation is still patchy, and we saw no clear consensus on how applications are deployed into clusters today, and that drove big differences in operational outcomes, as judged against DORA metrics.

And lastly, while the K8s community is seeing some population-wide trends, every business is running its own race. We found four distinct groups emerging, and it's not as simple as leaders vs laggards. While we wouldn't advise simply copying the high performers — their situation may be different to yours — you can certainly learn from them.

# Methodology

The 2025 State of Production Kubernetes report is based on an independent research study conducted by Adience, a specialist research agency. In May 2025, Adience surveyed 455 people via online survey and via phone, followed by in-depth phone interviews with six people. The survey and interviews were conducted in English, and we had respondents from the US, Canada, UK, Spain, Germany and France.

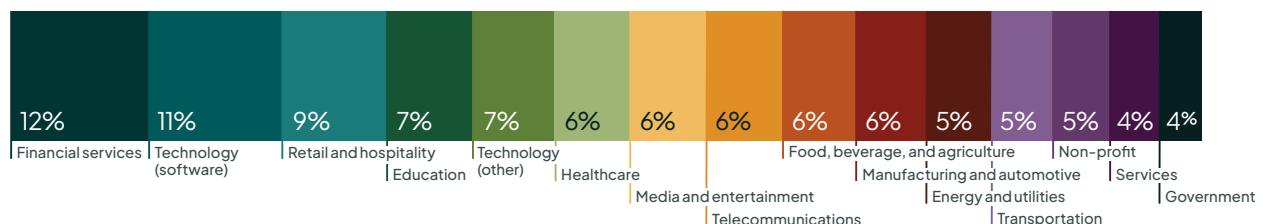
All respondents and interviewees were highly qualified as being directly involved in their organization's use of Kubernetes in production. We surveyed only organizations with more than 250 employees, but as always we were careful to represent different job functions, seniorities, sizes of business and industries in our research. Check out the chart to the right to see how our respondents break down.

## Our largest survey yet

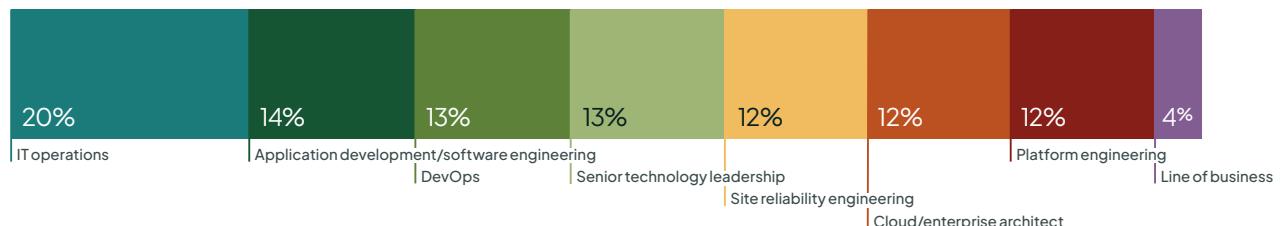
This year we surveyed 455 Kubernetes experts, making it our biggest survey yet.

Way back in 2021 we started with a survey of just 167 people – and since then we've gathered responses from 1,679 people, plus 28 hours of live one-to-one interviews.

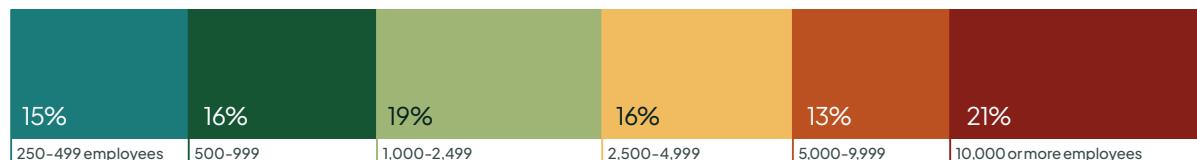
### Industry



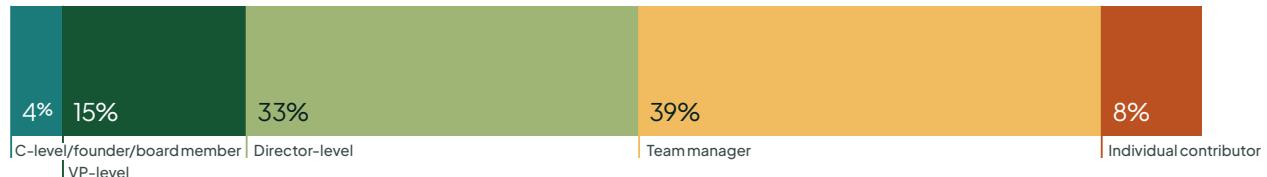
### Role



### Business size



### Seniority



# Five years of the State of Production Kubernetes reports

**2021: Operationalizing for scale.** Our first research tackled some familiar topics — including top K8s challenges, and the use of AI/ML workloads. But we mostly dug in to who owns this newfangled thing, and how standardized adoption and management are across the business. Only 40% had K8s in production for 2+ years.

**2022: Facing the new challenges of complexity.** We started tracking growth, and the complexity that results from it. Our deep dive this year was into the skills needed to manage K8s. 60% said K8s was their preferred or only way to deploy apps to prod — but 80% said it has a steeper learning curve than other technologies.

**2023: The new frontiers of Kubernetes.** While still tracking adoption, we branched out to new frontiers: from devex to edge and VM modernization. 82% said they struggled to serve their devs with clusters; 86% wished they could unify containers and VMs on the same platform.

**2024: Building resilient futures.** As K8s turned ten, we explored the headwinds facing users. While 78% said K8s is ‘mature’, 70% feared market consolidation and 76% say they had been held back by K8s complexity, security and access to talent. AI emerges as a key future driver: 68% say K8s is key to their business taking full advantage of AI.



# About Spectro Cloud

Spectro Cloud delivers simplicity and control to organizations running Kubernetes at any scale.

With its Palette platform, Spectro Cloud empowers businesses to deploy, manage, and scale Kubernetes clusters effortlessly — from edge to data center to cloud — while maintaining the freedom to build their perfect stack.

Trusted by leading organizations worldwide, Spectro Cloud transforms Kubernetes complexity into elegant, scalable solutions, enabling customers to master their cloud-native journey with confidence.

Spectro Cloud is a Gartner Cool Vendor, CRN Tech Innovator, and a ‘leader’ and ‘outperformer’ in GigaOm’s 2025 Radars for Edge Kubernetes and Managed Kubernetes.

Learn more and get started at [spectrocloud.com](https://spectrocloud.com).



# About Adience

Adience is a dedicated B2B market research agency that rebels against boring, cookie-cutter traditional research.

Instead, Adience’s experts use an arsenal of research methodologies to dive deep into the data, challenging assumptions with flexible thinking, total transparency, and smart analysis.

Drawing on decades of experience across SaaS markets, IT, and other B2B sectors, Adience delivers trusted answers to the questions that matter, so decisionmakers can act with confidence.

Learn more at [adience.com](https://adience.com).





# Spectro Cloud

Learn more at [spectrocloud.com](https://spectrocloud.com)

Research by 