



Research by  adience



The Spectro Cloud 2024 State of Production Kubernetes

Building resilient futures

As Kubernetes enters its second decade, new opportunities await the bold — but there are new headwinds to navigate, too. How do we build a cloud-native future strong enough to weather the storm?

Foreword: A decade of change

On June 6th, 2014, Joe Beda checked in the **first** Kubernetes code commit. Now, exactly 10 years on, it's a great opportunity to pause and look back at how far we've come as a community and industry.

Change can sneak up on you. It's easy to forget that ten years ago there was no CNCF, no KubeCon, no meme-spawning landscape of logos.

The next ten years will be just as interesting, as today's hot trends play out: from the explosion of AI as the **keynote topic du jour**, to shifting market dynamics and consolidation of the CNCF and OSS landscape, and the ever-present buzzword of platform engineering.

Since 2021, the **State of Production Kubernetes report** has been tracking the evolving landscape of K8s adoption. This year, we surveyed our largest ever group of Kubernetes practitioners and decision-makers — 416 — in partnership with Adience, a specialist market research firm.

To illuminate the stories behind the numbers, we also conducted six deep-dive interviews, and you'll see quotations from those interviews throughout this report.

You hold the result of this research in your (digital) hands, and we hope you find it a useful guide as you embark on Kubernetes' next ten years!



What do you think the next ten years has in store for Kubernetes?

We'd love to hear your feedback on our findings, your wishlist for our next survey, and your perspective on the world of K8s.

Drop us an email: research@spectrocloud.com

Section 1: Going all-in on Kubernetes

Adopters are strategically committed to Kubernetes – and this year they're running even more clusters, in more environments, with more complex software stacks.

Section 2: Familiar challenges, increasing pressure

Kubernetes may be 'mature', but companies find its complexity, security and skills challenges are worse than ever, having significant impact on adoption and outcomes.

Section 3: Brace for stormy seas ahead

Rising cost pressures and increasing instability in the vendor landscape are spooking organizations. They're hunkering down and curtailing their growth expectations.

Section 4: Maturing strategies for success

Organizations are adopting platform engineering practices and using commercial management platforms for multicluster automation, pursuing cost savings, security and stability.

Section 5: The new growth engines

AI hype is translating into action across the board, including at the edge. The Broadcom/VMware tsunami is driving containerization and Kubernetes momentum.

The typical K8s adopter has
20+ clusters running
11+ software elements

77% say K8s complexity
has '**inhibited their adoption**'

82% of senior decision-makers
say the cloud-native ecosystem
is **poised for consolidation**

38% of the biggest K8s
users have **fully adopted**
platform engineering

73% of edge K8s adopters are
already **deploying AI workloads**

Section 1

Going all-in on Kubernetes

Organizations are deepening their commitment to Kubernetes at scale

There's always a hot new thing in the world of IT. But even after ten years of adoption, organizations that use Kubernetes say they are strategically committed to it, reaffirming a question we asked back in 2022.

- In 2022, 60% said that Kubernetes was their preferred or only way to deploy applications to production.
- This year, 75% agree that "my organization has a strategic commitment to Kubernetes for our future infrastructure needs"

That ongoing commitment is playing out in their investments. Overall adoption of Kubernetes within businesses is up significantly from last year, and there's been a step change in how many clusters organizations are running in production.

As you'll see in Figure 1, back in 2022 it was most common for organizations to say they had 2-5 clusters in production. Now more than half (57%) say they have over 20 clusters in production.

Figure 1: Most organizations run more than 20 production clusters

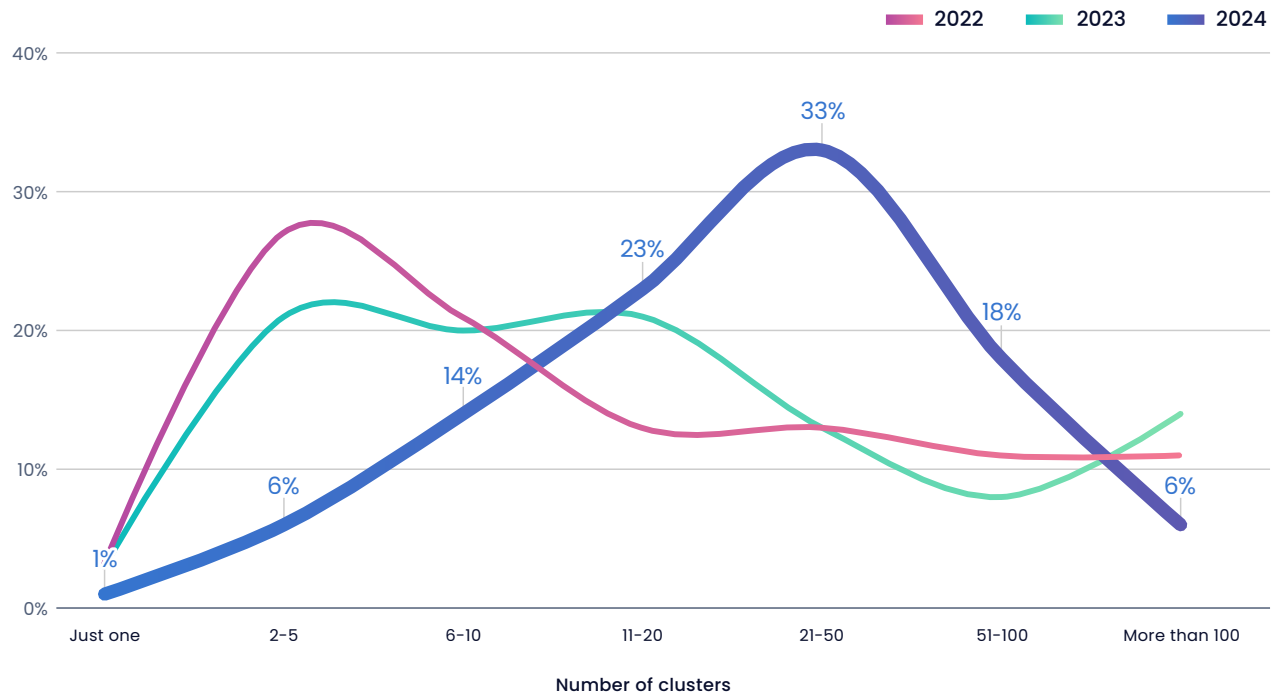


Figure 1: How many Kubernetes clusters does your organization currently run? 2024 survey versus 2022 and 2023.

"It's been around for about 10 years, but it seems like Kubernetes has only really come into vogue in the last two to three years."

Platform Engineer, Healthcare

Figure 2: Bigger organizations run more clusters – lots more

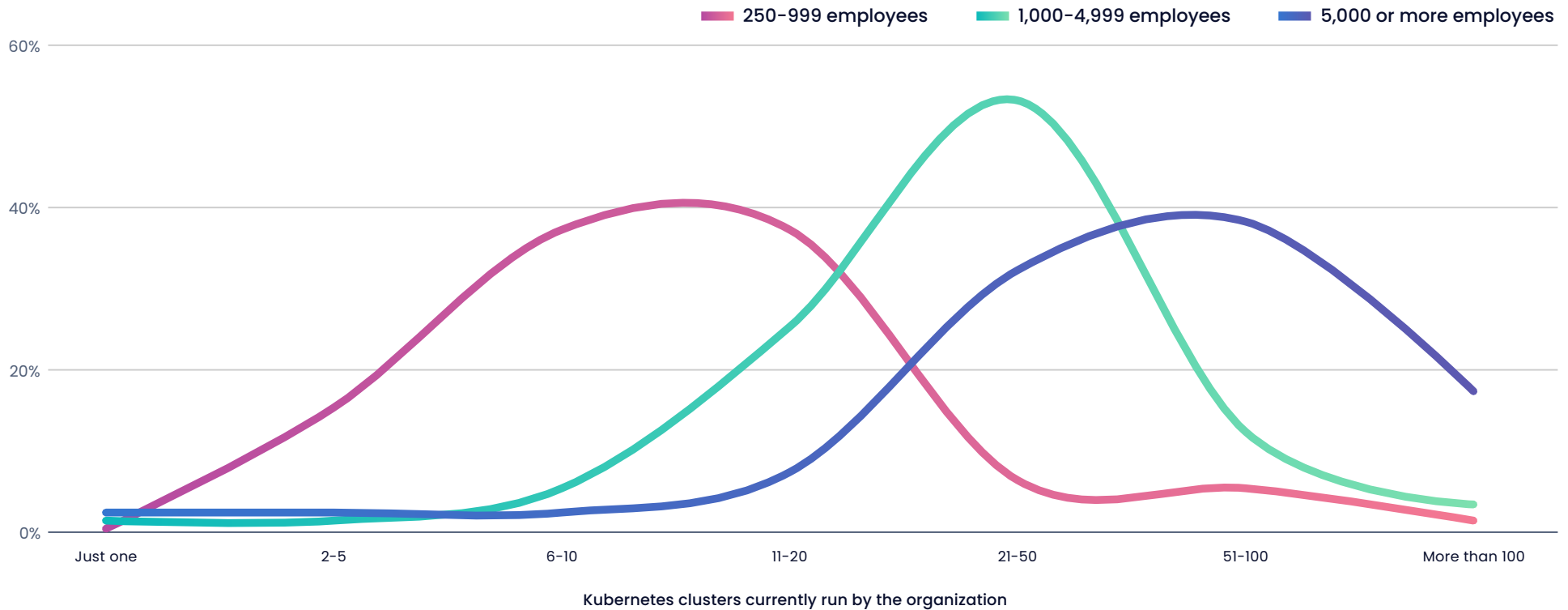


Figure 2: How many Kubernetes clusters does your organization currently run? By company size.

Bigger organizations have considerably more clusters, as you might expect. The typical large enterprise has more than 50 clusters.

Education and manufacturing sectors skew low with only 6% and 9% running more than 50 clusters, respectively.

Healthcare and retail skew high, with 23% and 28% running more than 50 clusters, respectively.

The technology sector, predictably, skews highest of all. 49% have more than 50 clusters.



Wherever you run applications, Kubernetes is there

The organizations we spoke to run their Kubernetes clusters in many different environments, across cloud, data center and edge.

In fact, 48% of respondents said they have clusters in four or more different environments.

As you can see in Figure 3, over the page, on-prem continues its growth, with a huge 85% having Kubernetes in **virtualized data centers**.

The progress towards 'cloud-inspired' **data center modernization** is clearly gathering momentum, potentially driven by **spiking cloud costs** or demand for AI workloads that require specialist on-prem investment. Financial services firms were most likely to run K8s in the virtualized data center (89%)

Bare metal is a long way off from being as popular as virtualized, but it is growing — perhaps driven by, again, the demands of intensive AI workloads, and by Broadcom's handling of VMware, as we explore in section 5. It has risen from 25% in 2023 to 31% today.

The biggest users? The healthcare sector (36%). Large enterprises were twice as likely to be using bare metal as smaller companies, at 40% vs 19%.

For the first time this year we asked about **air-gapped environments**, which are common in certain security-conscious or distributed industries. Overall, 10% say they are running air-gapped today, although some sectors hardly use it at all — education and financial services both stood at 3% adoption, while retail took the lead at 18%.

When it comes to **public cloud**, you can see the broader hyperscaler marketshare rankings play out. AWS is in the lead, Azure second, Google third.

Of course, there are industry differences — the retail sector much prefers Google over AWS, and financial services is relatively more likely to prefer Azure.

This year we broke our cloud options down to separate out IaaS products (like EC2) from managed K8s services (like EKS).

Across all three clouds, the self-managed options are more widely used than managed Kubernetes services. For example, 45% say they use EC2 for Kubernetes, 40% use EKS.

And there's strong overlap: the common pattern is for organizations to use both the IaaS and managed services from their preferred hyperscaler.



"Why have we grown our use of Kubernetes? It's the ability to scale seamlessly. It's agnostic, at least from a cloud platform perspective."

CTO, Healthcare

Figure 3: Multi-environment Kubernetes is ubiquitous

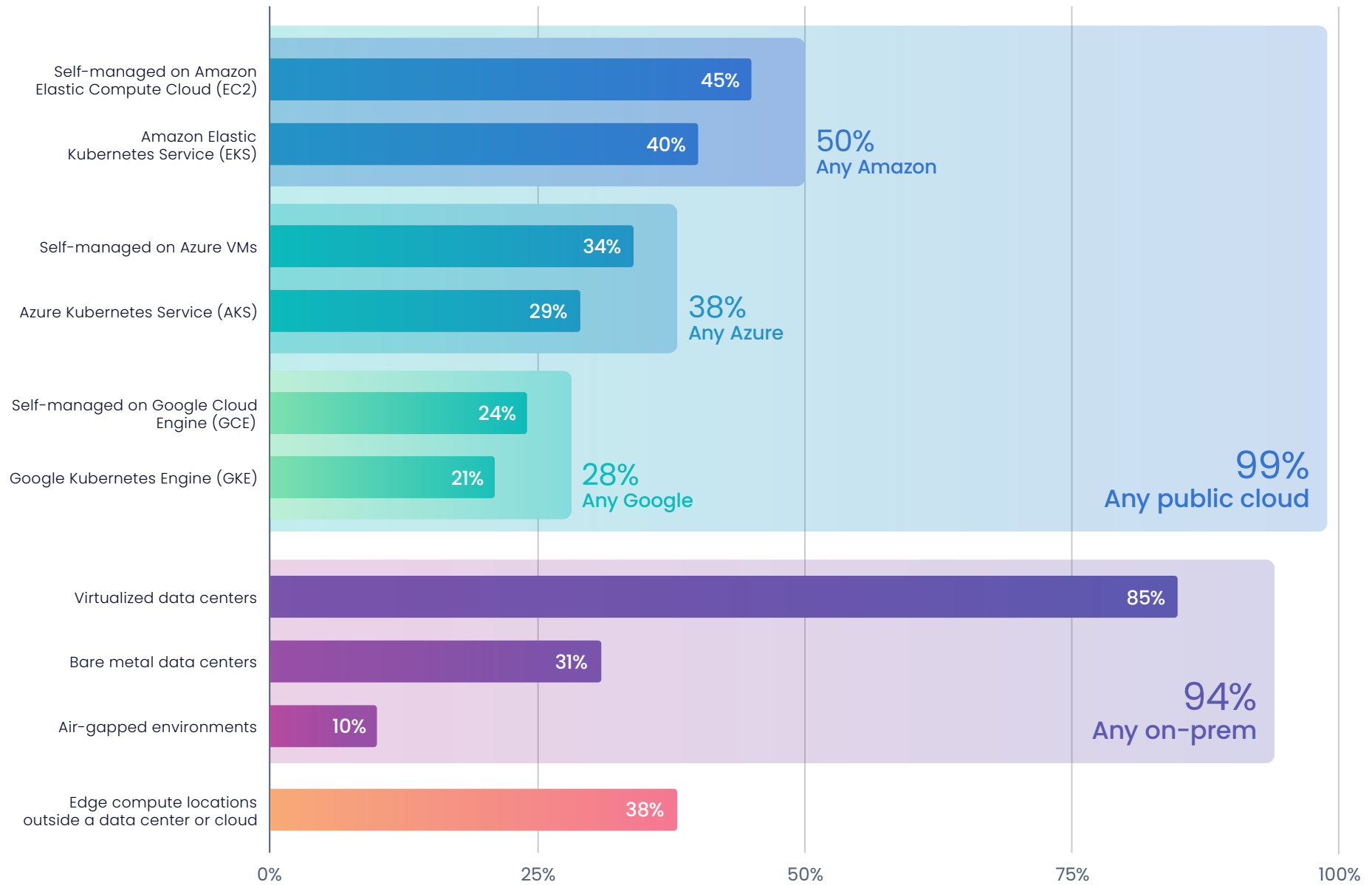


Figure 3: Which of the following environments are part of your production Kubernetes infrastructure?

What's in your cluster? Spoiler alert: lots

We use the number of clusters as our main metric for tracking the growing adoption of Kubernetes within organizations, but we also look at what's **in** those clusters.

This year we saw a big increase in how many distinct "software elements" our respondents said they were running in their Kubernetes stacks – the cloud-native integrations underpinning application workloads, such as observability, logging, CI/CD, service mesh, security tools, management agents, and so on. This is what Dynatrace calls "**auxiliary workloads**".

As Figure 4 shows, this year, 38% of respondents said they had more than 15 different software elements in use, up from 16% last year.

The more clusters an organization runs, the more different software elements it is likely to have across its full Kubernetes infrastructure.

If you're new to Spectro Cloud, this is why we're so big on the concept of managing the "**full stack**" – there's way more running in the cluster than just the core Kubernetes distribution and your application workloads.

Figure 4: Most have more than 10 different software elements in their clusters

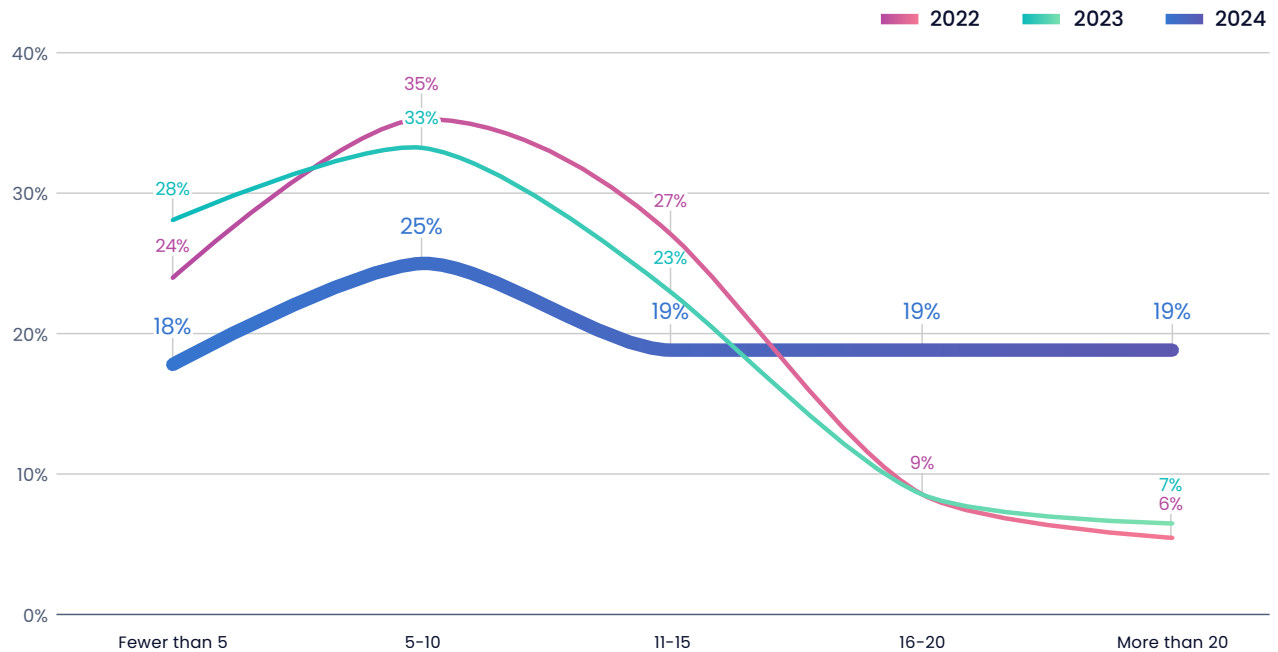
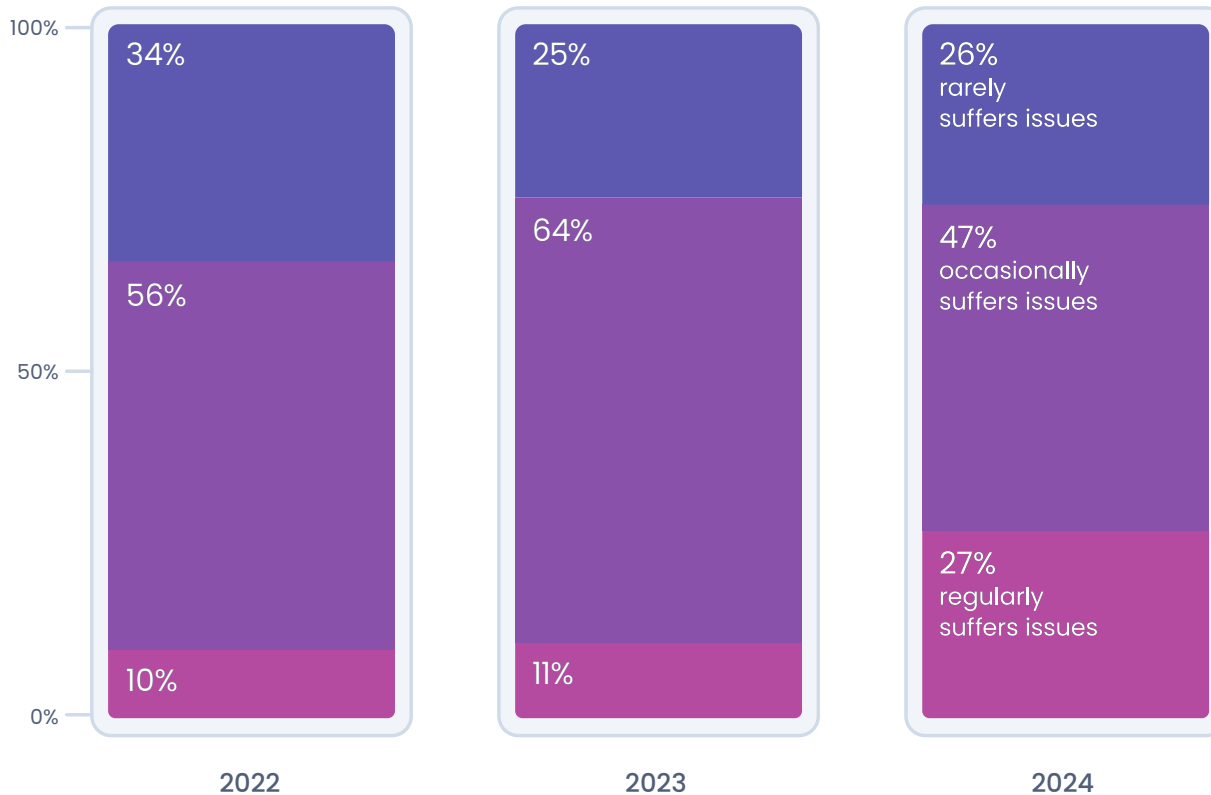


Figure 4: How many distinct software elements make up the complete 'stack' in your Kubernetes infrastructure?

More software elements = more problems

Figure 5: A quarter of organizations 'regularly' suffer issues affecting the running of their clusters



We repeated a question we've asked for the last couple of years, digging into the impact that the complexity of Kubernetes software stacks has on the most important KPI of all, cluster availability.

As Figure 5 shows, around three quarters of respondents say they occasionally or regularly suffer issues that affect the running of their clusters. That is holding steady – but the proportion regularly suffering issues has more than doubled year on year, to 27%.

This is far from scientific (what constitutes an 'issue'? What does 'regularly' mean in your business?), but as a measure of sentiment, the story is clear: as production environments grow more complex, it's harder to prevent problems from cropping up.

Through an industry lens, manufacturing performed worst: 44% suffer regular issues. The best performers were healthcare (18% regularly suffer issues) and technology (15%).

Figure 5: How does the interoperability of the elements in your stack affect the running of your production K8s clusters?

"If I showed you our architecture, you would probably get overwhelmed, because there's just so much that's interconnected, externally and internally."

Director, Government

Section 2

Familiar challenges, increasing pressure



'Mature' doesn't mean 'painless'

Kubernetes grows up

A majority (78%) of our respondents this year say that Kubernetes is a mature, production-ready enterprise tech, after a decade of development. In fact, 46% agree 'strongly' with that statement.

Most (70%) also say they themselves are 'highly mature' in their use of Kubernetes within their organization.

You might expect that a more mature technology, managed by more mature organizations, would result in a generally easier experience for operators.

But organizations continue to report facing many of the same challenges that we've been tracking every year... and in fact, those challenges are growing more severe.

More challenges than ever

We asked our respondents to look at a list of challenges and pick whichever ones they face when running Kubernetes in production.

Figure 6, over the page, shows the shifting results over the past three years.

Overall, the numbers are up.

Last year, any given challenge was picked by 34% of respondents; this year it's 40%.

So even if Kubernetes is getting more mature and therefore easier, other factors are clearly at work.

Let's explore the details.

Choice isn't always good

A few challenges saw steep rises year on year.

48% said it's difficult to choose their stack from the **broad cloud-native ecosystem** – given the large increase in software elements in the cluster we saw in section 1, this is not a shock. We explore the implications in section 3.

45% felt the pain of **configuration drift**. With more 'stuff' in each cluster, and more clusters running, conventional cluster-by-cluster imperative management approaches struggle to keep up.

Security was a new entrant at 44% (we previously made this part of our 'enterprise guardrails' option, which was our top ranking option last year).

And up from last place in 2023 to fourth this year at 43%, **cost management**. We explore that more in sections 3 and 4.

Only 1% said they don't face any challenges at all. Lucky them.



Figure 6: Production users report a shift in challenges

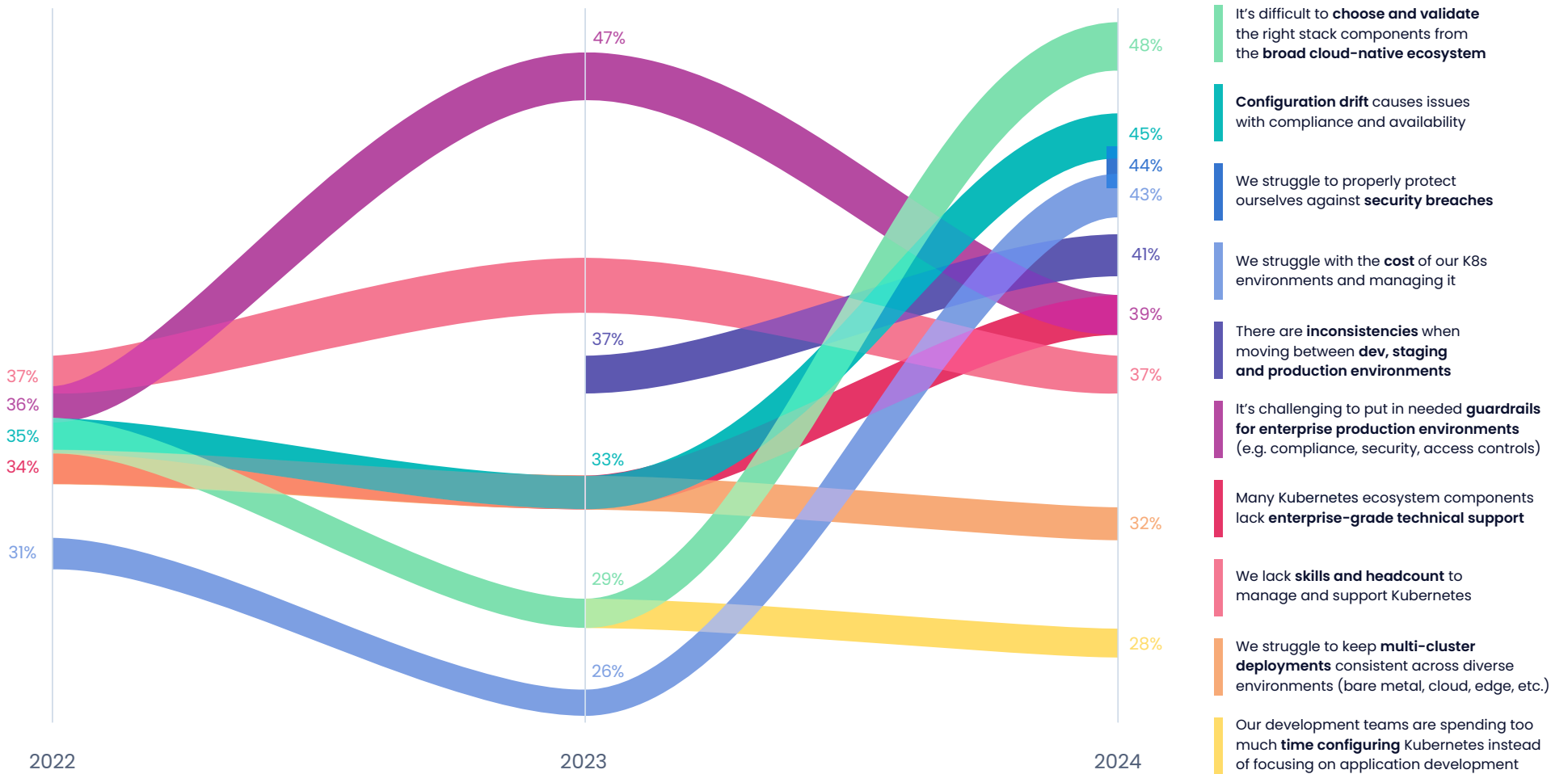


Figure 6: What challenges does your organization face with running Kubernetes in production?

Complexity, security and skills have inhibited growth

As well as asking which challenges organizations are experiencing, we went a step further this year and asked respondents whether some of the key challenges they've experienced in their use of Kubernetes have actually **inhibited** their adoption.

The answer was a resounding yes. More than a third of organizations say they have been "inhibited a lot" by:

- The complexity of managing multiple clusters
- Concerns about Kubernetes security
- Access to expert talent and **skills**

See Figure 7 for the full, alarming, numbers.

None of these issues are new – that's why we asked about them – but within the cloud-native community there is an overwhelming narrative about Kubernetes 'eating the world', becoming the de facto computing platform, growing rapidly and transforming businesses.

This may be true, but it's vital for the community to recognize the challenges that adopters continue to face, and have a productive conversation about how to address them.

And one more important point: remember that for this survey we only interviewed those who are running K8s clusters in production today, which means they persevered and adopted despite the inhibitors. There are potentially thousands of organizations out there who may have looked at the hurdles and decided to go with a different technology instead.

Figure 7: Three quarters say K8s complexity, security and skills have inhibited their adoption

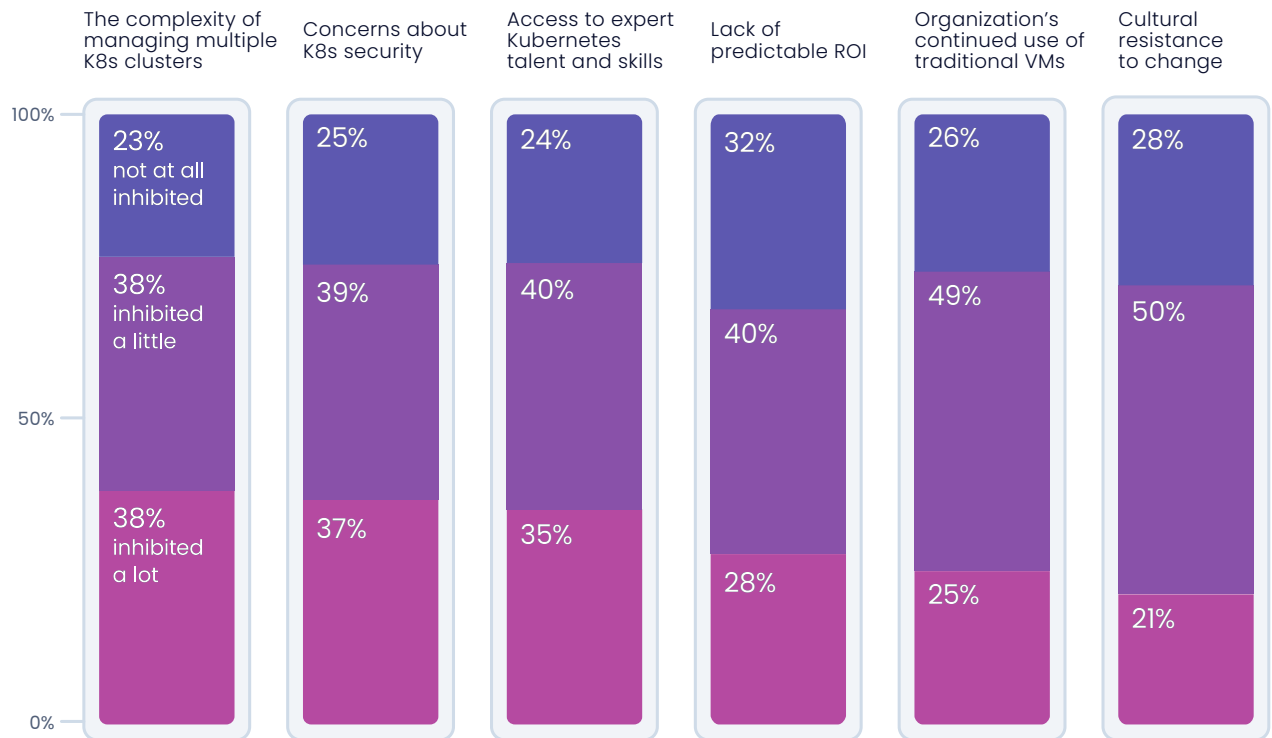


Figure 7: To what extent have the following factors inhibited your adoption of Kubernetes for production workloads?

"Our biggest issue is how complex it is to the layperson. That hasn't stopped us; we are slowly but surely moving more and more stuff into production and Kubernetes. But it slowed us down dramatically. We could have been at least twice as far as we are now."

DevOps engineer, Manufacturing

Section 3

Brace for stormy seas ahead



Ecosystem consolidation is on the cards

73% of our respondents said that Kubernetes will still be an important part of the enterprise technology stack in 10 years.

Few technologies earn the right to become part of the bedrock of IT for 20 years, so that's cause for celebration.

But just because users see a bright future doesn't mean that K8s, and the cloud-native ecosystem that's grown up around it, has smooth sailing ahead in its second decade.

Already, we have several data points pointing to a looming market adjustment, as Kubernetes graduates into a later stage of what HBR calls "the consolidation curve".

The headline: 70% of our respondents said that the cloud native ecosystem is poised for market consolidation, as Figure 8 shows.

That figure rises to 82% of senior decision-makers, with 47% agreeing 'strongly' with the statement.

There are several different issues at play here — let's explore.

Figure 8: Most K8s users are concerned about the future of their projects and vendors

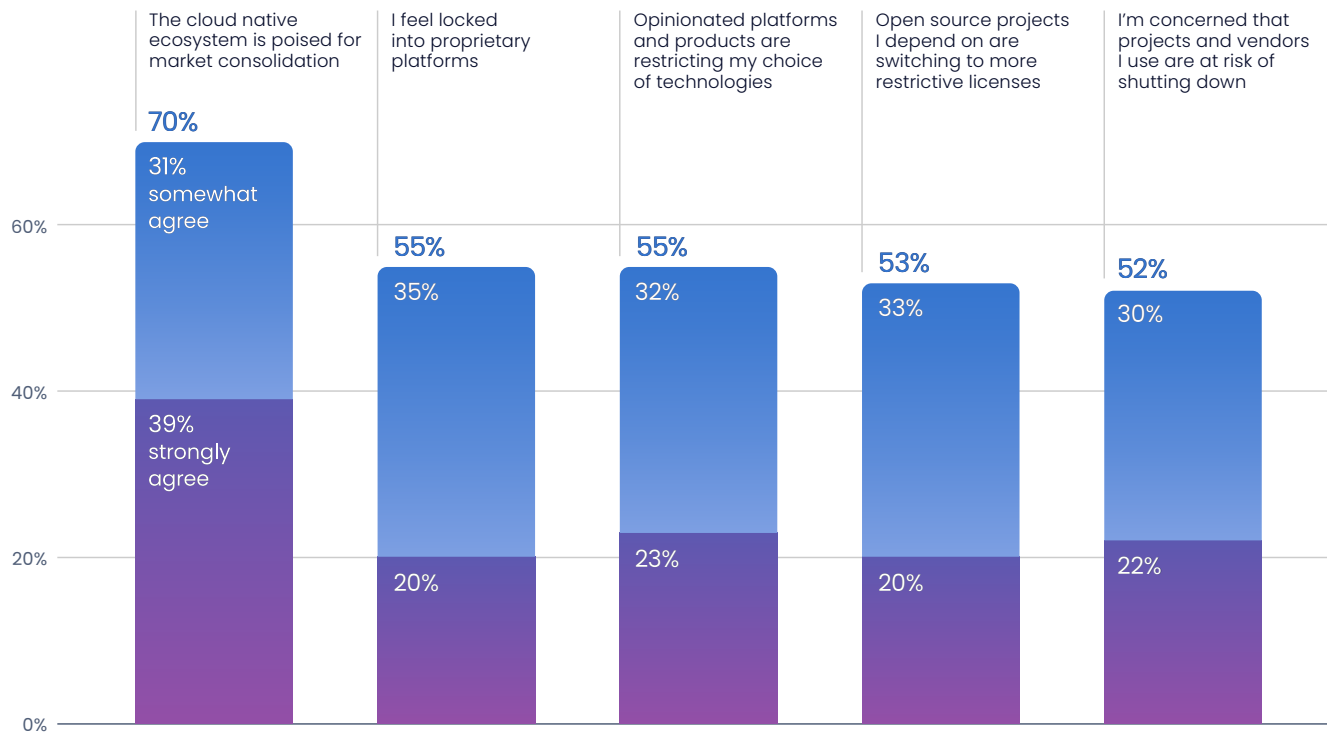


Figure 8: To what extent do you agree with these statements?



The CNCF landscape is causing friction

In section 1 we talked about the many elements included in the typical Kubernetes stack, and in section 2 we found that the top challenge was choosing the right software from the cloud-native ecosystem. You don't have to look far for articles and conference sessions about the landscape and [how daunting it is](#).

Until recently, the vibrancy of the CNCF landscape was unequivocally a good thing – it showed a community in bloom. As adopters encountered a need, the community innovated to ideate and solve it. Everyone likes to have options, right?

But as an operations team wanting efficiency and stability in a maturing market? You're now facing the **"tyranny of choice"**. It's a real challenge just to effectively evaluate and pick a technology that has the right features, performance and stability, let alone keep on top of all the new projects being established.

And when you're no longer making purely greenfield architectural decisions, it can be difficult to act on choices, too. More than half of our respondents said they felt locked in by proprietary platforms, and that their choice of technologies was restricted by opinionated products.

"What we don't want to do is be caught in vendor lock-in where I give the original contract to you and now I'm stuck with you forever, whether you're good or bad."
Technology Director, Government

Is community support enough?

For enterprise production use, it's also critical to consider non-technical factors like support and longevity.

Many open-source projects are offered on a community best-effort basis. That's fine when you're exploring in your lab, or when you have a super strong technical team internally with a DIY mindset, that can take on the burden of keeping things running. But for the typical enterprise in production, the late adopters, this is not feasible.

39% of our respondents in our survey said they were already challenged by the lack of enterprise-grade technical support from OSS projects.

When something goes wrong with a production cluster at 2am on a Sunday, who can you call to get help? If you discover a bug or a vulnerability, what's the SLA for a patch to close it?

Survival of the fittest for OSS projects

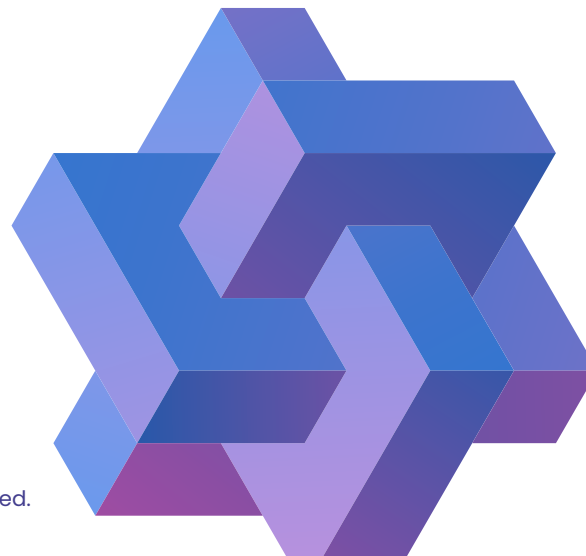
Many widely used OSS projects have a relatively small team of contributors, potentially even just a single maintainer. This of course is not new (and there's an [obligatory XKCD](#) about it...), but perhaps it's never been this bad.

The recent security scare around xv utils [brought the situation into the public eye](#), and [intensified community debate](#) about whether the current OSS landscape is sustainable. When the lead maintainer gets sick or burns out, gets laid off, or just finds they've had enough – what happens to the project?

Other projects face a different problem: they fail to break through and get enough adoption. This is natural selection: there are only so many tools that the user community can sustain. But the result is the same – the project stops getting updated, and the users are left wondering what to do.

More than half (52%) of those we surveyed are concerned about projects and vendors they actually use today shutting down, both open source and commercial.

Organizations that have 'regularly' suffered issues with their clusters are significantly more worried, with 40% 'strongly' agreeing.





Vendor backers are tightening up

Many OSS projects large and small are of course sponsored by vendors, which use open source as a funnel for commercial offers such as professional services and support.

In the past year or so we've seen a shift in this commercial dynamic as cash-strapped vendors decide they need to extract more revenue from users, and stay ahead of competitors.

RedHat, **Redis**, **Hashicorp** and many others have moved features behind paywalls or changed to more restrictive licenses, in many cases resulting in **a community backlash** and forks such as OpenTofu.

In our survey, more than half (53%) of organizations say they have actually already experienced 'restrictive license changes' from open source projects they depend on.

"I don't lose sleep over Kubernetes. I'll tell you that. I'm losing sleep over VMware and some of these other vendors out there... it's the strong arm tactics they're using to increase their revenue... I would probably go so far as to say it's almost predatory." CTO, Energy

The M&A cycle is accelerating

We've also seen acquisitions big (Hashicorp being bought by IBM and Isovalent by Cisco) and small (Civo buying Kubefirst, Dynatrace buying Runecast, to name just a couple).

And those watching the startup community will have noticed vendors making layoffs, shutting down or **pivoting**, leaving their customers and users with an uncertain future.

What can you do in the face of all these swirling challenges? Make careful choices, and choose vendors that give you an exit strategy in terms of open standards.

"So with the proliferation of Kubernetes lots and lots of vendors have popped up with cool ideas about Kubernetes. And they'll start a project and they'll get a user base, but then that tool disappears."

I've had this happen several times over the years, where I've implemented something that's a great idea and then they just completely disappear the next day and I'm having to support it myself.

That I think is the biggest risk I see with Kubernetes at this point is making sure you pick a technology vendor that you trust to survive more than a year or two.

I've dealt with several vendors that have shut down completely, or they've abandoned products just off the bat, just because I wasn't making them any money.

So it's this one thing you have to do extra in your vetting of products."

DevOps Engineer, Manufacturing





Cost pressures start to bite

In our 2023 survey, cost concerns ranked lowest in our list of challenges. Now the topic of cost has surged to the top 5. In some industries, such as education, cost is the overall number one challenge in operating production Kubernetes. Smaller businesses and those with fewer clusters were also much more likely to pick cost as a challenge.

As Figure 9 shows, more than 60% of respondents said their TCO has increased in the past year, that they don't have accurate forward visibility, and most important of all, they face more pressure to reduce infrastructure costs than they did.

44% of retailers said they "strongly agree", and those that "regularly" suffer issues in their clusters were 3x more likely to face cost pressure than those that don't.

68% also said their adoption of Kubernetes has been inhibited by lack of predictable ROI (and you can't calculate a return on an investment when you can't measure the investment itself).

"We're starting to really see the cost of the cloud as we move to it. And sometimes it's not as inexpensive as people are advertising. And it could be more expensive than what you were doing on-prem because now you don't own a thing and you're paying by the drink for everything."

Platform Engineer, Healthcare

Where are these cost pressures coming from?

The year of growth that we identified in section 1 is certainly part of it. Whether in the cloud or in the data center, more clusters running more software means more expenses, and we've seen all three hyperscalers [report double-digit revenue growth](#) in 2024.

New specialized workloads and initiatives — like edge and AI, which we'll cover in section 5 — may contribute too. GPUs don't come cheap.

In section 4 we'll examine some of the strategies that organizations are investigating to manage these pressures, but first: what does this all mean for the future?

Figure 9: Costs are up, pressure is intensifying

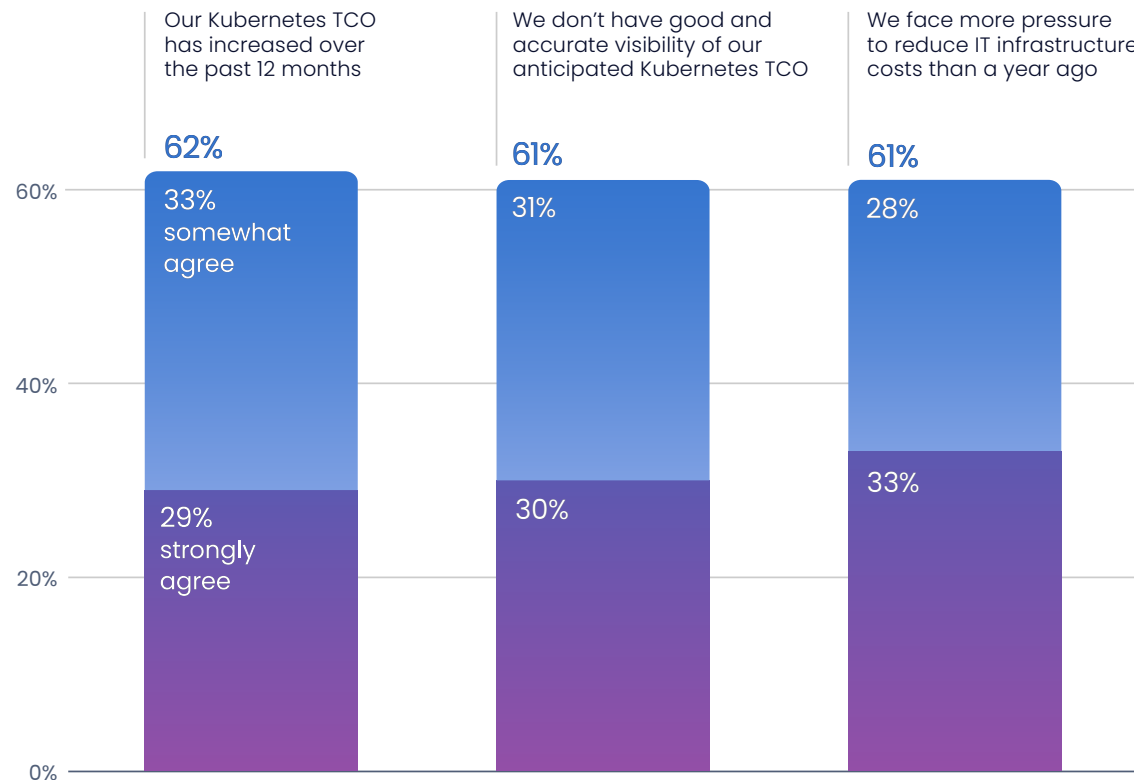


Figure 9: To what extent do you agree with these statements?



Adopters are growing cautious

Only half expect growth in next 12 months

Each year we ask our respondents to rate the growth they expect in several areas of their Kubernetes infrastructure, from the number of clusters to the number of applications they're bringing to Kubernetes.

In 2022 and 2023, around 80% of organizations said they expected their cluster estate and range of applications to grow, with around 20% expecting 'strong' growth.

This year, they're much more cautious, as Figure 10 shows, over the page.

Now only around half of organizations expect to grow their Kubernetes clusters or applications.

For the first time, many expect their K8s use to fall

In this question organizations are also allowed to report either no change, or a decline in each of these areas.

In previous years the percentage expecting a decline in any area was basically single digits – nobody expected their infrastructure to go backwards. We didn't even talk about it in our writeups.

This year, the numbers have changed, in a big way. Around a quarter of organizations said their number or size of clusters would decline, along with their application base and the development teams they work with.

Smaller organizations are more likely to be pessimistic about growth than large enterprises. For example, 30% of smaller organizations said they would shrink the number or size of their clusters, compared to 17% of large enterprises.

It's natural for adopters of a maturing technology like Kubernetes to see a leveling off in growth rates after a while. But these respondents weren't picking the 'no change' steady-state option.

Commitment and consolidation can coexist

To see some adopters reporting a decline suggests either that Kubernetes has failed (which we know from other answers that it most definitely hasn't), or more likely a cost-driven consolidation: retiring old applications, decommissioning unused clusters, consolidating environments, shutting down projects.

Certainly recent research ([such as this from cast.ai](#)) suggests that there is plenty of fat to cut.

We're not sounding a death-knell here: net out our data and the growth is still overall positive.

And what's more, remember that we surveyed those who already use Kubernetes in production.

For them, Kubernetes may be now a mature, mainstream part of the fabric of their infrastructure. But there's lots of potential for growth from new adopters.

Still 50% of KubeCon attendees each year are first-timers. Gartner estimates that fewer than 50% of organizations run containerized applications in production. And [according to Dynatrace](#), only 13% of mission-critical workloads are running on Kubernetes today.





From growth to consolidation

Figure 10: Only half expect to grow their K8s use; for the first time ever, a quarter expect to shrink

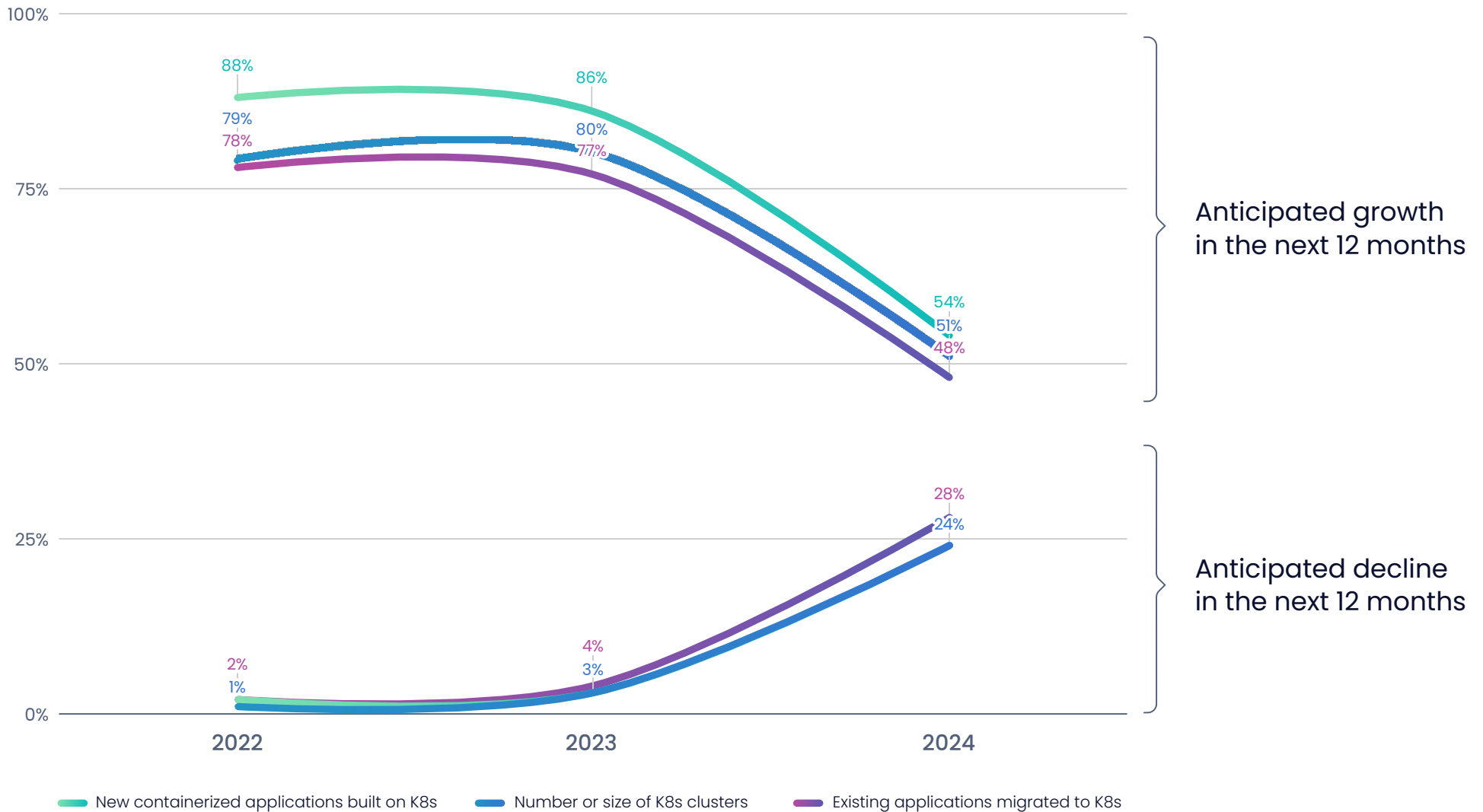


Figure 10: How do you expect your organization's Kubernetes use to change in the next 12 months? 2024 responses compared to 2022 and 2023

Section 4

Maturing strategies for success

Management best practices emerge

Responding to headwinds

In section 2 we looked at the challenges organizations face with using Kubernetes today.

In section 3 we looked at external factors, such as market uncertainty and cost pressures, that IT teams are predicting for the year ahead.

Now let's look at the operational approaches they're taking to position themselves for ongoing success.

Limiting experimentation

In response to the complexity of the cloud-native ecosystem, and the issues caused by complex software stacks, many organizations (58%) say they are sticking with 'tried and true' products instead of evaluating new ones.

This is particularly true of those that report regularly having suffered issues.

"A good reason for staying with tried and tested systems is that it is my assumption that it has gone through full QA and testing for it to be operationalized or at least certified for use. Bringing in any new system, I would always be wary." CTO, Healthcare

In other words the winners and losers in each CNCF landscape category are likely to stay much more stable than they were, with fewer new entrants gaining traction.

Multicluster for isolation

As organizations deploy more applications to Kubernetes, and as concerns about security mount, we're seeing more organizations change their approach to cluster sizing.

As Figure 11 shows, a majority are choosing to segment and isolate their applications across many clusters, instead of lumping different workloads together into fewer clusters and relying on namespaces to provide isolation (and yes, this is probably feeding into the growth in clusters that we saw in section 1).

"Any development group runs on a separate tenant. It's pure security for us." CTO, Energy

Larger organizations tended to have adopted strong isolation strategies more than smaller ones, but smaller companies are more likely to use technologies like vCluster to get the same benefits of hard isolation, without the overhead of managing lots of full host clusters.

This trend towards multicluster isolation is a step on even from the CNCF's 2023 survey, where 191 production Kubernetes users said they separated applications via clusters, 239 via namespaces.

Figure 11: Larger orgs use separate clusters for security isolation

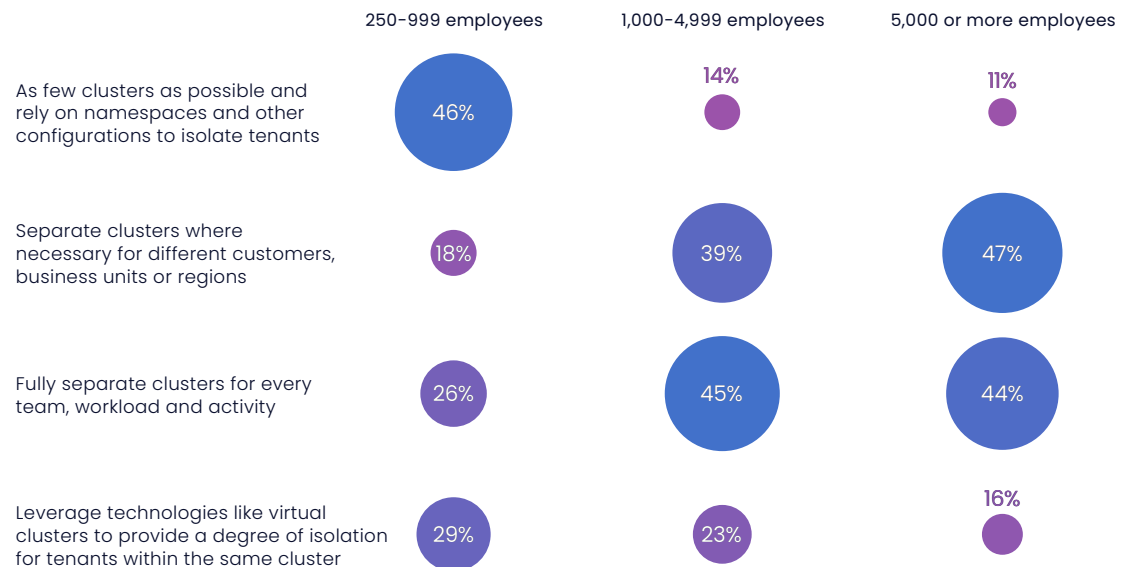


Figure 11: What's your current approach to multitenancy in Kubernetes?



Commercial management tools grow more popular

To manage the growing number of clusters, we're seeing organizations move away from DIY tooling and single-environment dashboards, towards a combination of commercial management platforms and outsourcing or managed services.

Now, as Figure 12 shows, commercial management solutions are the most popular solution for organizations, for the first time.

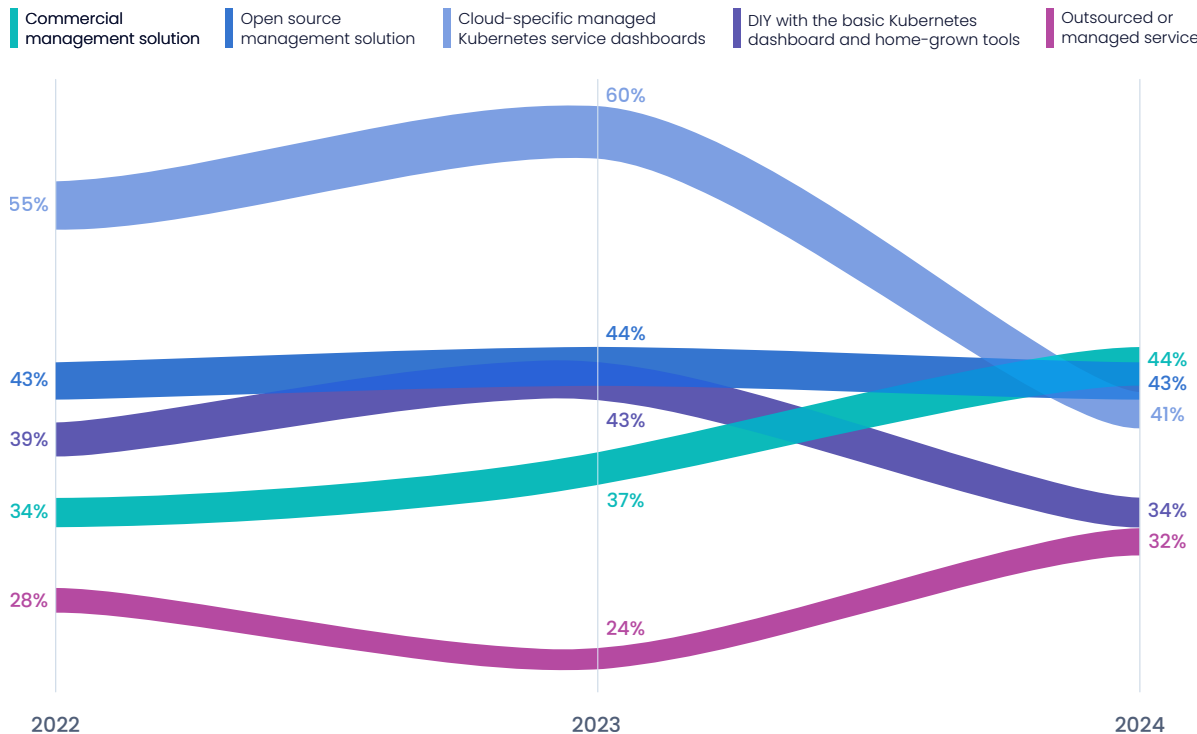
Organizations with the largest estates were more likely to use commercial solutions: 49% of those with more than 20 clusters picked this option.

"Since I first started looking at [Kubernetes], some amazing management products have come out to orchestrate and manage and report on all of your containerization. So the biggest change I've seen is the enterprise shift has been very fast." CTO, Energy

Management tooling is a powerful way to tackle the K8s skills and cost challenges, by automating away tasks and providing a paved-road experience that bakes in best practice. One interviewee told us:

"The tools that we have to spin up and make changes to clusters have improved a lot. It's fairly easy to build a Kubernetes cluster, if you design it right. You just need some specialized people to get it set up like that." DevOps Engineer, Manufacturing

Figure 12: Organizations are moving from DIY to commercial tools



Those who said they used a commercial management solution were:

- Less likely to be challenged by skills or headcount to manage Kubernetes (31% vs 42% of those that do not use a commercial tool)
- Less likely to say they face cost challenges (39% vs 47%)

Nevertheless, in the typical large enterprise with pockets of Kubernetes, standardization of tooling and practices is far from common.

"There's just a lot of different departments and divisions that are using different methods for managing Kubernetes." Director, Government

Figure 12: What approach does your organization take to managing its Kubernetes clusters?



Platform engineering makes progress

Hype or happening?

Unless you've been living under a rock for the past few years, you'll have heard the buzz about platform engineering.

Definitions of platform engineering vary.

But in short, most agree that platform engineering is a practice where a dedicated platform team builds and operates internal platforms that give their "customers" (application developers) the ability to consume infrastructure services such as Kubernetes clusters, often in a self-service or on-demand manner. In other words, Kubernetes as a Service (KaaS).

Green shoots

The driver behind platform engineering is to make life easier for developers to maintain feature velocity, without having to deal with the complexity of Kubernetes or 'go rogue'; for the enterprise and the platform team, it's a way to enforce standardization and governance.

Given all the hype, you'd be forgiven for thinking that platform engineering is everywhere. It's getting there, but like the future, it's unevenly distributed, and many of those we interviewed spoke about activities just beginning:

"We are now seeing within our organization, the opening up of roles associated to the build-out of platform systems." CTO, Healthcare

"It is coming into play. And we do look at these concepts and these new methods." Director, Government

On track for widespread use

29% of our respondents strongly agreed that "my organization has fully adopted platform engineering roles and patterns", with a further 41% somewhat agreeing.

This is absolutely in line with a prediction from the analysts at Gartner:

"By 2029, more than 80% of enterprises will adopt a centralized platform engineering and operations approach to facilitate DevOps self-service and scaling, from less than 30% in 2023" (How to Run Containers and Kubernetes in Production, 18 April 2024, by Stephanie Bauman, Arun Chandrasekaran).

Platform engineering also seems to be something that goes hand in hand with widespread Kubernetes adoption – although it's up for debate as to which is the cause and which is the effect.

38% of those with more than 50 clusters say they have fully adopted platform engineering, versus a quarter of the rest.





The old ways aren't gone

Only 10% of our respondents self-identified as platform engineers today, compared to 14% DevOps and 6% SRE. So much for **'DevOps is Dead'**: the 'old' roles haven't fully disappeared yet in a blaze of LinkedIn announcements.

Interestingly, those who identify themselves as DevOps and SREs were the least likely to say their organization had fully adopted platform engineering.

And from our interviews, there was still a degree of uncertainty and even skepticism about the term.

"I don't know if it's just another buzzword that we seem to create in engineering all the time.

I can only tell you from a job requisition perspective we're not using that expression, we still use our traditional SRE.

I think it's more like a mindset as to how to approach DevOps. The engineering roles have not changed, right? They're still doing the same type of job." Director of Cloud and Application Security, Retail

Remember that we've already seen organizations' adoption of Kubernetes inhibited by limited access to talent and skills, and cultural resistance to change – those same inhibitors are probably blocking adoption of platform engineering, too.

Better outcomes?

There is a correlation that suggests platform engineering practices are beneficial.

Those that have fully adopted platform engineering are half as likely to suffer regular issues with the running of their clusters, as Figure 13 shows.

Figure 13: PlatEng adopters half as likely to suffer regular issues

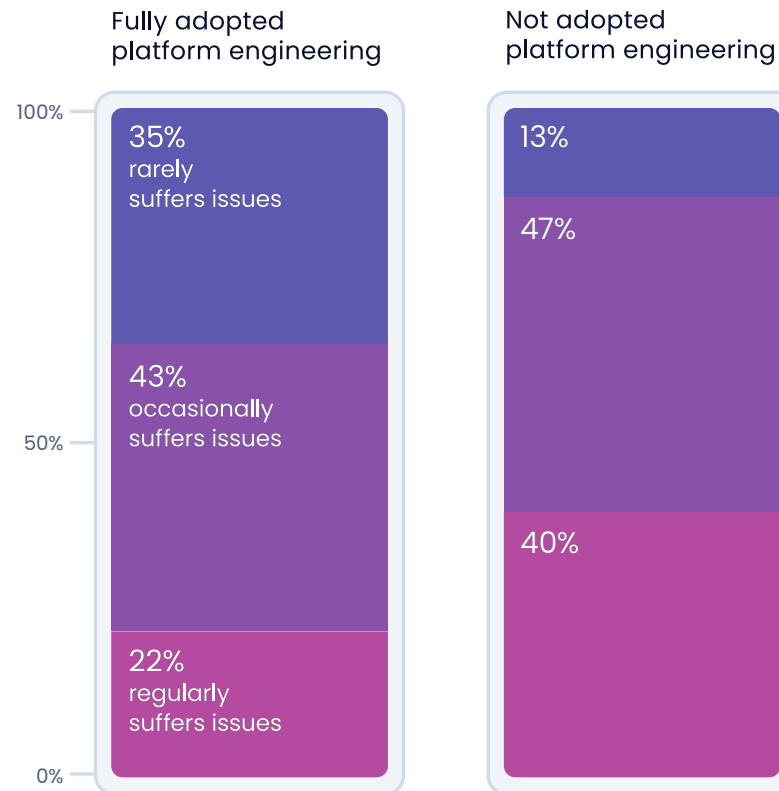


Figure 13: Strongly/somewhat agree that my organization has fully adopted platform engineering, vs how often my organization suffers issues related to the running of its production Kubernetes clusters



Cost reduction by any means possible

We spoke in section 3 about the cost headwinds facing organizations.

For the past few years we've asked what techniques organizations are using to optimize Kubernetes efficiency, and this year we asked an expanded version, shown in Figure 14. The responses are illuminating.

Let's walk through.

Figure 14: Organizations are trying a range of approaches in response to growing cost pressures

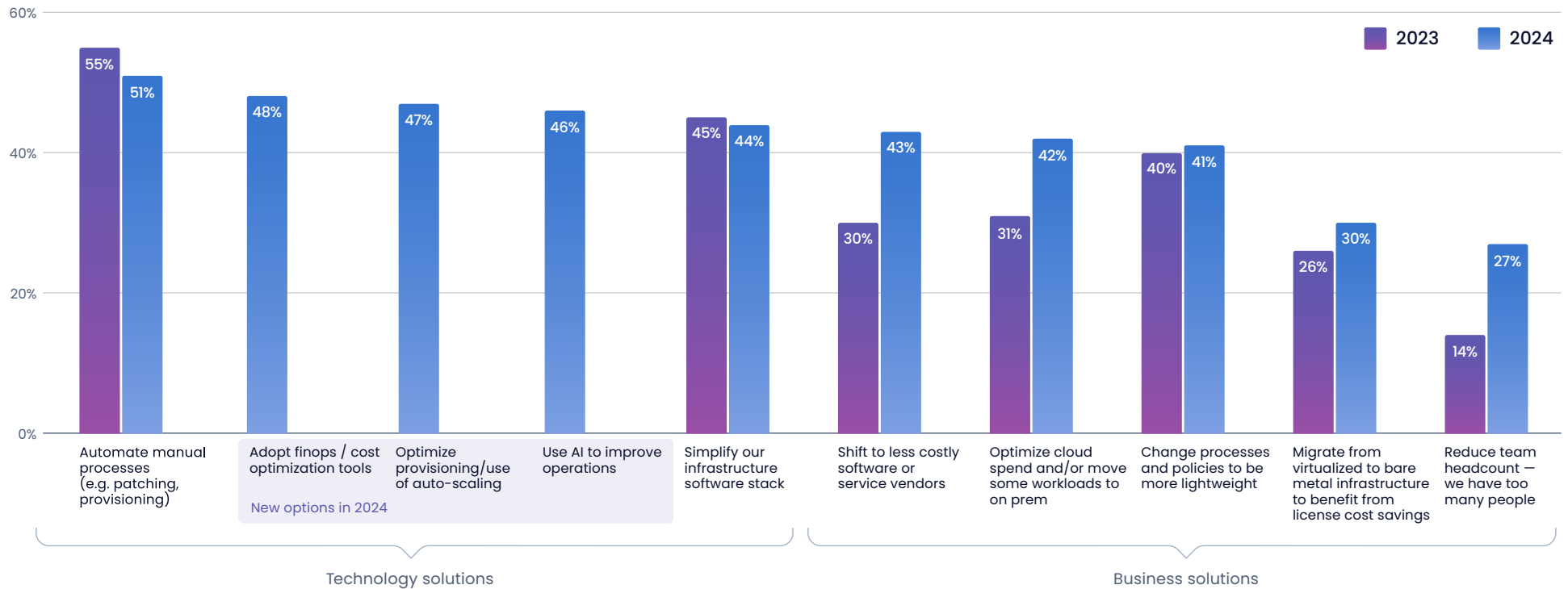


Figure 14: Where do you think your organization has the biggest opportunities to improve Kubernetes infrastructure operational efficiency?



Every tactic is in play

First: Organizations are investigating many different ways of driving efficiency.

Eight of our ten options sit within a 10 percentage point band; there's no one clear winning approach that fits everyone.

Automation remains #1

Second: the most popular solutions (on the left of Figure 14) are essentially **technology** levers. Can I flip a switch to save money?

Automation (such as using a management platform to deploy and perform cluster lifecycle operations) is #1 both last year and this, with high levels of interest from organizations running every number of clusters, and of every maturity level.

But we also see a lot of interest in dedicated finops tools, and use of autoscaling to reduce cluster overprovisioning.

These were new options added to the list this year.

Vendor strategy in the spotlight

Third: organizations are starting to pull the **business** levers on the right of the chart, too.

Compared to last year, we saw a big spike in interest for shifting vendors, up from 30% to 43%, particularly driven by senior decision-makers.

More than 40% of organizations also said they were looking to optimize cloud vendor spend or even move away from cloud vendors back to on-prem.

And there was some increase in moving those on-prem environments from virtualized to bare metal (perhaps prompted by hefty Broadcom renewals rolling in).

Cutting headcount when skills are short?

And finally: we see a big increase in interest for **reducing team headcount**, up from 14% to 27%.

This is worth unpacking a little.

You might expect this would be an order from the top, but actually individual contributors were nearly twice as likely to select this option as the C-suite (32% vs 18%).

You might also expect headcount cuts are the preserve of large enterprise, but all sizes of business were equally affected, although there was some trend towards those with more K8s clusters being more likely to show interest in cutting.

So what's the explanation? Headcount reduction sometimes goes hand in hand with automation, and we're perhaps also seeing a recognition that existing 'legacy' team members may not have the right mix of skills and culture needed for today's cloud-native stack.

What about AI, which comes in fourth in Figure 14? Good point. We'll discuss that in more detail in section 5.

Section 5

The new growth engines



Did somebody say AI?

Alops is making strides

70% of our respondents said that AI is the “biggest trend in Kubernetes this year” – which won’t be news if you’ve attended a KubeCon recently.

But what does AI mean to our Kubernetes practitioners and decisionmakers? Are they doing anything, or just watching webinars?

One use case for AI is as an ops “copilot”, helping Kubernetes administrators manage their clusters, just as developers are increasingly using AI to help write and debug code.

Having an AI assistant make sizing recommendations or untangle obtuse error messages sounds appealing, particularly in the context of the challenges we’ve discussed above, around the complexity of Kubernetes and the difficulty of finding expert talent.

As Figure 15 shows, 44% of our respondents said they were using Alops today.

A recent survey by Dynatrace, not specific to Kubernetes, put the adoption of Alops at 72% – so 44% is actually a conservative figure.

A further 34% said they planned to be using Alops within the next year.

Note that ‘Alops’ doesn’t necessarily mean full, end-to-end AI-augmented automation of every operational activity. Our interviewees often gave examples of point solutions.

*“We’re using AI to scan for code vulnerabilities. A lot of different data breaches have occurred in the past. Sometimes those can be associated to code vulnerabilities, which AI tools deployed on our clusters have been invaluable in being able to identify.”
CTO, Healthcare*

AI for cost optimization

As we discussed, respondents also see AI as a tool to help hunt down cost savings.

46% said it was one of their biggest opportunities to improve operational efficiency, rising to 58% of those with more than 50 clusters.

Overall 40% of organizations say they’re already using AI cost optimization, with 34% planning to in the next year. Smaller organizations, and those with fewer clusters, are actually using AI cost optimization tools most today.

Figure 15: AI use is already widespread, both Alops and AI workloads

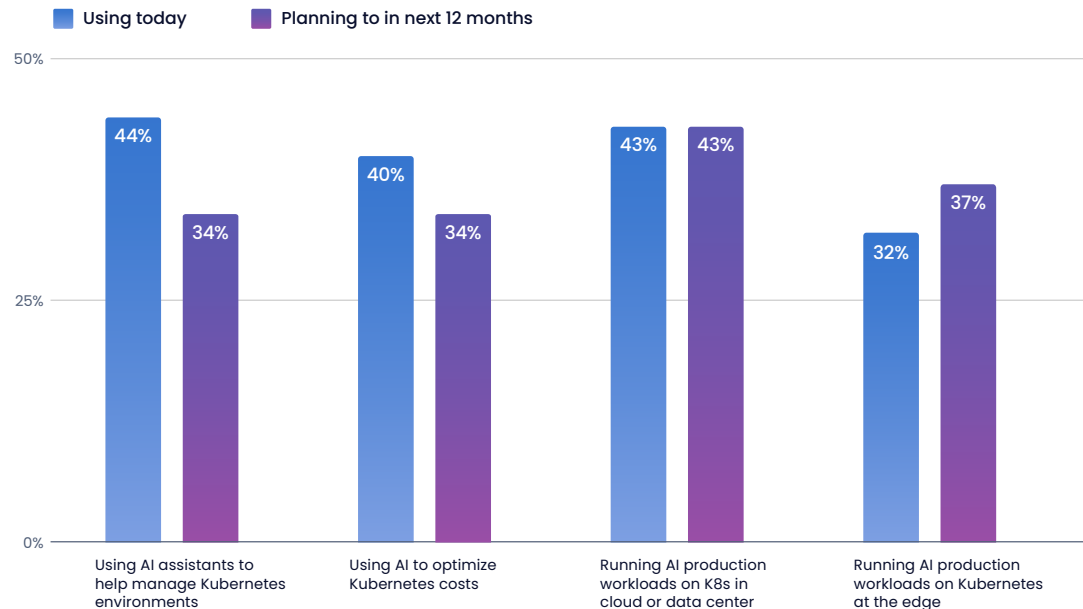


Figure 15: Which of the following are you doing today, or plan to do in the next 12 months?



Is AI delivering results?

Several of our interviewees expressed concerns about whether Alops and AI cost management tools are actually living up to expectations today.

"We have lots of tools that claim all this functionality with AI and we haven't seen anyone deliver on it. Yes, we have things like chat bots but nothing that can help us as far as actually orchestrating and managing clusters." Director of Cloud and Application Security, Retail

Others drew attention to the data privacy concerns of turning your operational logs into training data for Alops or finops tools (a challenge we at Spectro Cloud have been working on with **LocalAI**):

"If we can guarantee that it's our own AI instance running, then we're okay with using it. If it's not, then we do not use it inside any of our services... we don't want to share privileged data with the entire machine learning world." DevOps Engineer, Manufacturing

Kubernetes is the new home for AI workloads

The other side of the coin is using Kubernetes to orchestrate containerized AI workloads. It turns out, organizations see a natural fit for Kubernetes to perform this role.

68% said that "Kubernetes infrastructure is key to my company taking full advantage of AI in our application workloads".

Why? One interviewee explained:

"I definitely think our use of Kubernetes will grow. Some of what my company has embarked on in recent months is specifically because of the growth of AI and machine learning tools.

Those tools require elastic frameworks in order to accommodate the compute power that is required in order to operationalize these tools.

And so just given the capabilities that exist with Kubernetes, and the company's desire to consume more AI tools, we will use Kubernetes more in future." CTO, Healthcare

43% of organizations large and small report that they're already running production AI workloads on Kubernetes in the cloud or data center, with only a few sectors, such as manufacturing, dragging their heels.

If this number sounds high to you, remember that these aren't necessarily large-scale deployments. As Wiz's 2024 **State of AI in the cloud report found**, only 10% of AI adopters are 'power users' today.

While the cloud or data center is a natural home for intensive AI workloads like model training, we're seeing a lot of organizations investigating AI workloads in edge computing environments, closer to the source of data and the business process they're focused on.

Our survey bore this out. 68% of respondents said that the popularity of AI is driving interest in edge computing.

73% of those running Kubernetes at the edge today said they are already using AI at the edge, with a further 20% planning to do so.

AI and edge are a match made in heaven.



Edge is forging ahead

As we saw earlier, 38% said they were using Kubernetes at the edge. This average masks some massive differences.

First, by size of business: three quarters (74%) of the largest enterprises we surveyed said they were doing edge today, nearly 7x the rate of the smaller businesses we spoke to. See Figure 16.

Second, by maturity. Most computing environments (for example, AWS EKS) are used equally by mature and immature Kubernetes adopters. But for edge that isn't the case.

Those that rated themselves 'highly mature' were much more likely to be using edge, at 48% vs 31%.

Figure 16: Enterprises lead edge use

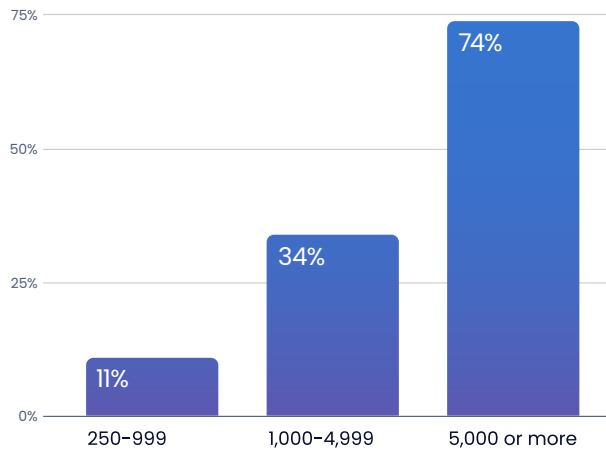


Figure 16: Edge K8s adoption by company size

As in 2023, we wanted to learn what "doing edge" really meant: just a pilot, or widely deployed in production?

In 2023, only 7% of our respondents said they had "full-scale production deployment" of edge Kubernetes clusters: the rest were some way through pilot or planning.

Today, 27% of our respondents said they had graduated to full-scale deployment, as Figure 17 shows. That's a big shift in a year, and it indicates that a lot of hard work and progress has been made.

Figure 17: Full production edge deployments are up 4x year on year

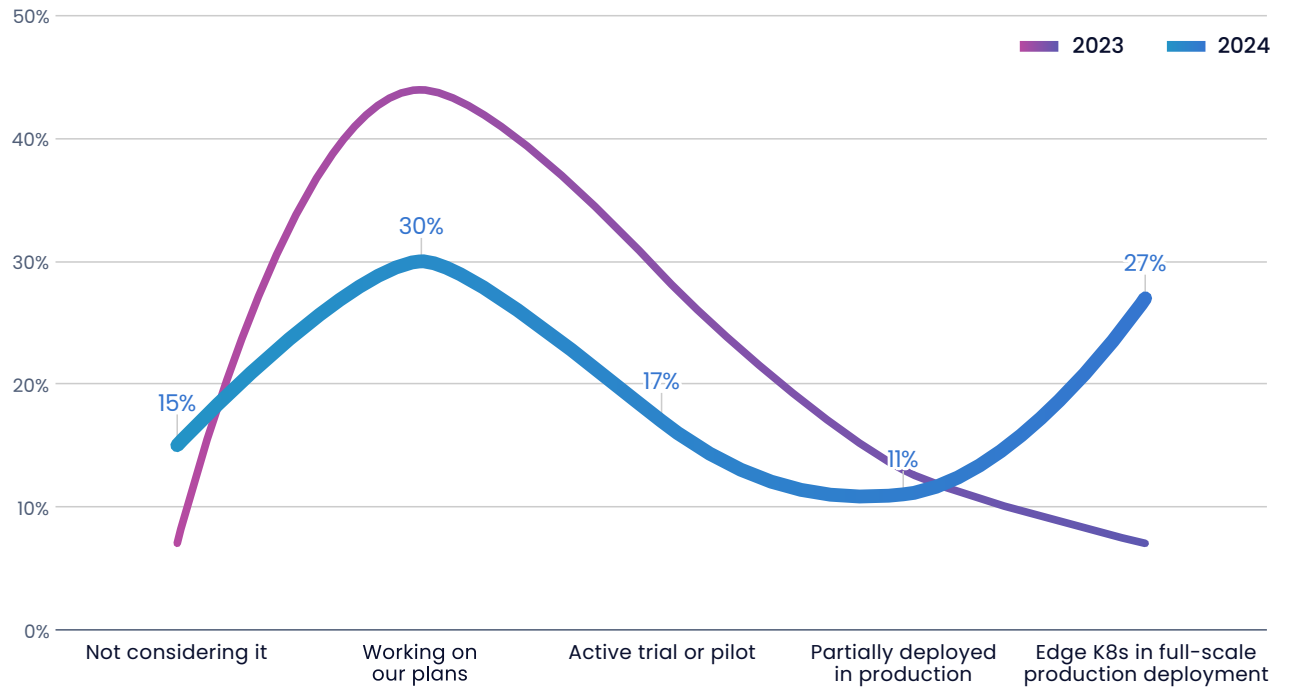


Figure 17: What is your current level of adoption of Kubernetes on edge computing infrastructure? 2023 vs 2024



Edge challenges persist but evolve

41% anticipate doing more with edge in the next year – and it's the largest organizations that are most bullish. While cloud is a leveler, accessible equally to small businesses, edge remains a domain where business scale seems to be a requirement.

Despite this confidence, challenges remain.

Note that in Figure 17, the number of those saying 'we're not considering edge' has more than doubled year on year, from 7% to 15% – clearly some of those who had previously said 'we're working on our plans' or were piloting decided not to proceed to production.

In last year's report, we explored the many difficulties of doing edge computing, from the sheer cost of field engineering visits, to concerns around the security of devices that are open to tampering.

This year, as Figure 18 shows, the situation is similar.

On a positive note, security concerns are down year on year – we'd like to take some credit for that after launching the **Secure Edge Native Architecture (SENA)** with Intel last year!

Operational, production challenges are up, though. Organizations are realizing that when they have devices out in the field, how do you perform day 2 operations like patching? That's up from 28% to 41%. The related challenge, around field engineering costs, is up slightly too.

"The biggest challenge for us is nobody on site that has any idea about the technology... we also have some use cases where they have no internet access. So we're working on solutions to mainly run these scripts on these clusters and see if we can get that to work." DevOps Engineer, Manufacturing

As interest grows in edge computing, and organizations start to push specialized workloads like AI to the edge, perhaps it's unsurprising that more are starting to be frustrated by device performance (37%) and finding the right hardware (up from 16% to 28%).

Figure 18: Day 2 and field engineering challenges grow, security declines

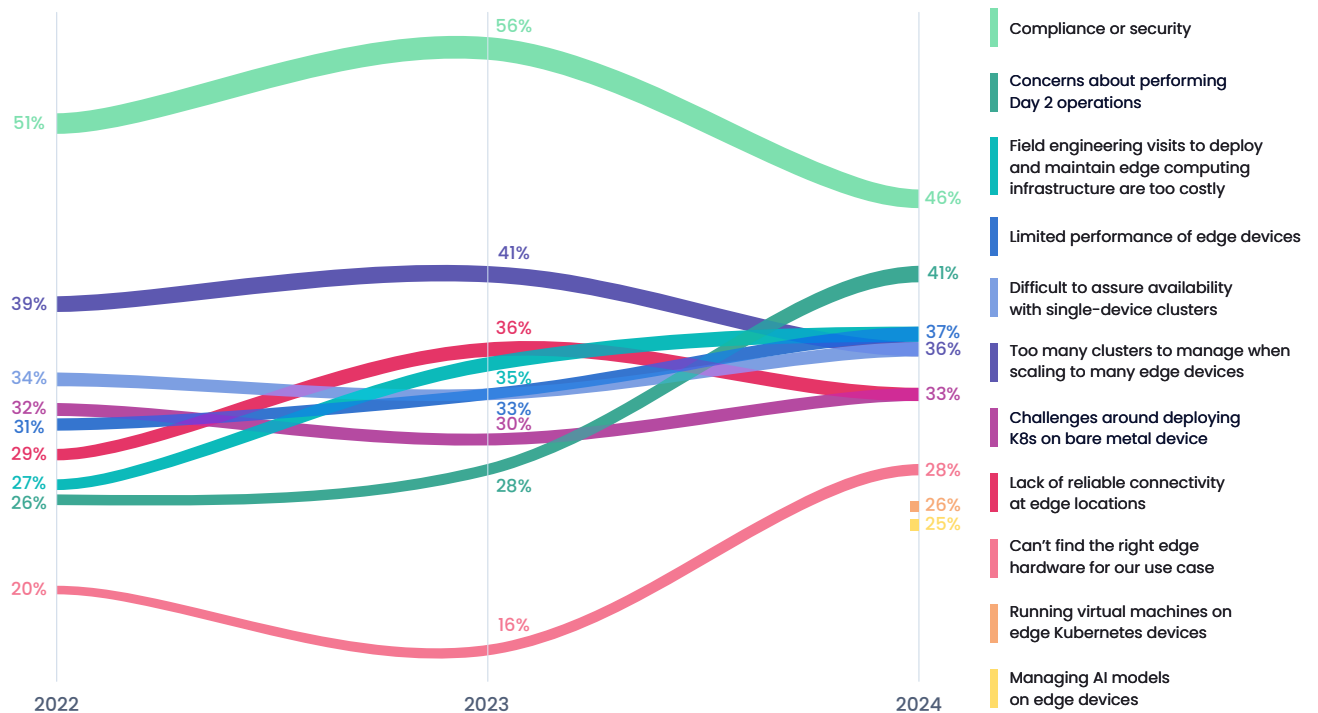


Figure 18: What is challenging about deploying and managing Kubernetes on edge computing devices? Data from 2022, 2023 and 2024



Modernizing the legacy VM tech stack

Legacy tech has slowed K8s adoption

While sexy new topics like AI get most of the keynote slots, it's legacy tech that's been keeping CIOs up at night over the last few months – in particular, the ongoing saga of VMware and Broadcom.

74% of organizations say their adoption of Kubernetes to date has been inhibited by their organization's continued use of traditional VMs.

And this is not only a pure infrastructure issue, but a skills and culture issue. 72% said their adoption of Kubernetes has been inhibited by cultural resistance to change.

"The big challenge is changing the culture, right? We've got an older workforce, they're not so pro making the changes. Some of them are near the end of their career. They don't want to learn new stuff, right? We've got more legacy expertise than we do newer expertise. So unfortunately, while we're going through this trend, we depend on contractors." Platform Engineer, Healthcare

Broadcom has catalyzed change

But this is changing. 59% told us that Broadcom's takeover of VMware has accelerated adoption of cloud-native technologies like Kubernetes. Senior decision-makers and organizations with more Kubernetes clusters were more likely to agree.

VMware renewals are exacerbating the cost challenges we explored in section 3, and as we discussed, we can see that playing out in our data:

- Adoption of Kubernetes on bare metal (versus on a hypervisor like VMware) is up from 25% to 31% of organizations
- 30% said they were investigating migrating from virtualized to bare metal infrastructure to benefit from license cost savings.
- 43% said they were investigating shifting to less costly software vendors.

Those data points cover running Kubernetes clusters on virtualized data centers. But there's another side to the story: what to do about traditional VM workloads if you're moving away from VMware.

Bringing VMs to K8s

It's possible to orchestrate VM workloads alongside containers, inside Kubernetes using cloud-native patterns, thanks to tools like KubeVirt.

In the past couple of years, KubeVirt has become much more widely adopted and is often **discussed in the tech press**.

"[The acquisition of VMware] is influencing my decision making right now, heavily. We know we're not going to move completely away from virtualization immediately. So we're looking at alternative virtualization, hypervisors, and things like that, to get us through. We are likely to be 80 to 90% containerized applications in three or four years. Everything we can move into containerization, we probably will." CTO, Energy

We asked organizations why they would start using Kubernetes to orchestrate virtual machines. While 48% cited the operational efficiencies of unifying VM and container management, 43% said it would be part of a strategic effort to reduce VMware dependence – that number rising to 51% of senior technology leaders.

In other words: there's a clear picture that VMware has been the comfortable default, but now it's a burning platform, and many organizations are looking to take this opportunity to jump to cloud-native with Kubernetes.

"Moving our legacy applications to a Kubernetes deployment is definitely a goal for us." Director of Cloud and Application Security, Retail



Conclusion

Here's to the
next ten years



Here's to the next ten years

Our research this year paints a picture of a technology and community in transition.

Through the answers that these 416 individuals gave, we can see that Kubernetes has crossed a chasm to a new degree of maturity.

Year on year organizations have dramatically increased how many clusters they run, and there appears to be an emerging consensus towards the adoption of platform engineering, commercial management tools and multitenancy practices that even just a couple of years ago were up for debate.

New growth vectors have emerged, with clear interest in AI, edge, and AI on edge — and a probable influx of VM workloads from those fleeing VMware.

For those that use it, Kubernetes has become a central pillar of IT, and it's not going away any time soon. 75% of users are strategically committed to Kubernetes, and 73% believe it'll still be around a decade from now.

But with growing maturity comes some difficult changes. Consolidation of the unwieldy CNCF landscape seems inevitable, and with it a phase of cost cutting and tough cultural change.

Let's see how the journey unfolds.

Methodology

About the research

The State of Production Kubernetes report is based on a primary research study conducted by Adience, an independent research agency.

416 respondents completed a quantitative research survey in April 2024. All respondents are directly involved in their organization's use of Kubernetes, and have at least one Kubernetes cluster in production.

Respondents by industry

Our top five sectors accounted for almost two thirds of respondents. They were technology, financial services, manufacturing, healthcare and retail.

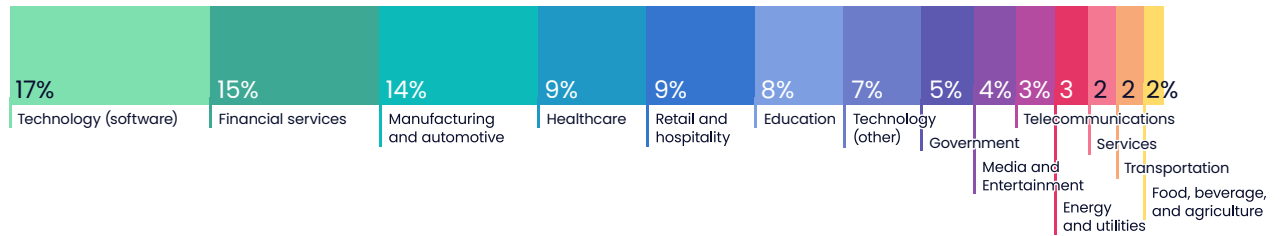
We ensured that the respondents represented a wide range of industries, company sizes, roles and seniority.

83% of the respondents were located in the USA, with the remainder in the UK, Canada, and other countries. The survey was conducted in English.

Adience also conducted six one-hour telephone interviews covering a mix of roles (CTO, platform engineer, DevOps engineer, etc) representing government, retail, energy, manufacturing and healthcare sectors. Quotations from these interviews have been used with only light editing for brevity.

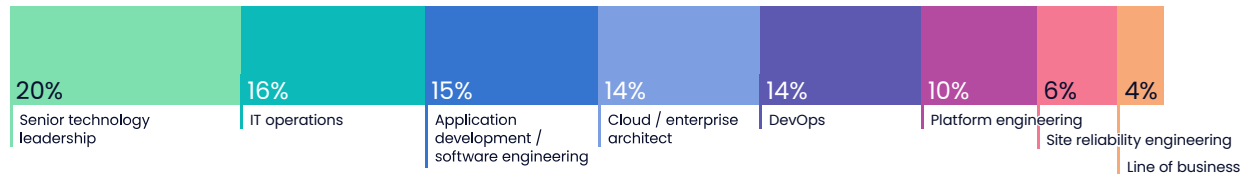
Respondents by role

20% of our respondents are in senior technology leadership positions. The majority of the rest self-identified in a variety of operational roles, including IT ops, devops, platform engineering and SREs. We also had representation from application development and line of business.



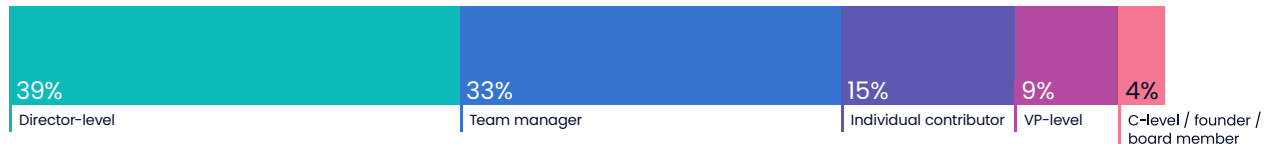
Respondents by seniority

The majority of our respondents were in team manager or director-level management roles, giving them a degree of both strategic and hands-on perspectives into the use of Kubernetes in production.



Respondents by business size

Our minimum threshold was set at 250 employees. We had even coverage all the way up to the largest enterprises, with 10% of our respondents working for organizations with 10,000+ employees.



About Spectro Cloud

Spectro Cloud uniquely enables enterprises and public sector organizations to manage Kubernetes in production, at scale.

Its Palette next-gen Kubernetes management platform gives IT and platform operations teams effortless control of the full Kubernetes lifecycle.

- Build the perfect stack for each cluster, with full choice over the OS, K8s distro, and integrations from across the cloud-native ecosystem, all declaratively managed. Both VM and container workloads are supported.
- Deploy to public cloud, virtualized and bare metal data centers, and to the edge, all from a single pane of glass, with a unique decentralized architecture for unbridled scale, optimized for each target location.
- Manage day 2 and beyond with powerful native capabilities for automating patching, upgrades, backup, security scans and more.

Spectro Cloud is a Gartner Cool Vendor, CRN Tech Innovator, and a 'leader' and 'outperformer' in GigaOm's 2024 Radars for Edge Kubernetes and Managed Kubernetes.

Learn more and get started at spectrocloud.com.



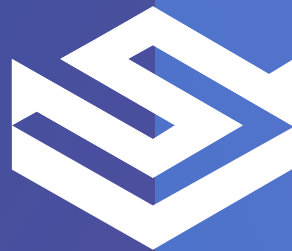
About Adience

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.





spectro cloud

Learn more at spectrocloud.com

Research by  adience