DoK Community

RESEARCH REPORT

Data on Kubernetes 2024

Beyond Databases: Kubernetes as an AI Foundation

November 2024

Executive Summary

The transformative impact of Data on Kubernetes (DoK) accelerated in 2024 with a notable revelation: organizations increasingly view Kubernetes as one of the most important parts of their infrastructure for accelerating their AI strategy. This finding underscores how Kubernetes has evolved from its initial role in container orchestration to become an important foundation for next-generation data and AI infrastructure.

Our 2024 survey of 150 technology leaders reveals that DoK adoption has reached a new level of maturity, while simultaneously pushing into more advanced use cases. The ecosystem is expanding beyond traditional data workloads to embrace AI/ML, streaming, and advanced analytics, all while databases remain the backbone of DoK deployments for the third consecutive year.

The data tells a compelling story: organizations are not just experimenting with DoK—they're standardizing on it. This increase in advanced workloads on Kubernetes reflects both the maturing ecosystem and the growing recognition of Kubernetes as a foundational platform for modern data infrastructure.

The Data on Kubernetes Community would like to thank its sponsors, as well as members of the CNCF's Cloud Native AI Working Group (CNAI) who helped develop questions for the survey.

Key Findings

AI/ML Acceleration: Organizations are increasingly viewing Kubernetes for AI/ML workloads as a competitive advantage, with a majority agreeing that it will serve as a key platform for accelerating their AI strategy.

Database Dominance: Databases remain the #1 workload on Kubernetes for the third year. Even as the ecosystem expands into more advanced use cases like AI/ML, streaming, and analytics, databases continue to serve as the backbone of DoK deployments. This consistency demonstrates the platform's reliability for mission-critical workloads.

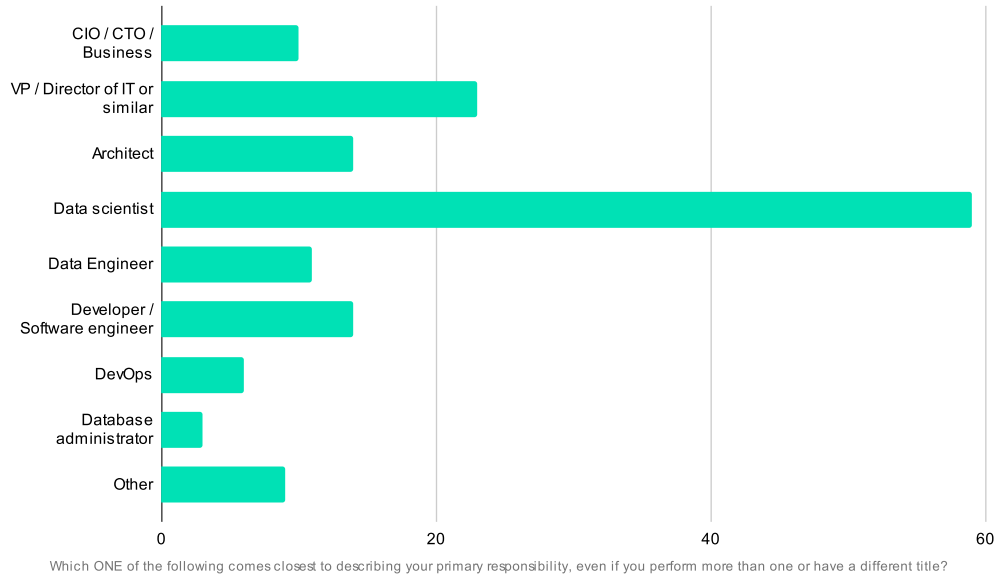
A Maturing Ecosystem: Organizations aren't just testing DoK—they're committing to it. Almost half are running 50% or more of their DoK workloads in production, with the most advanced organizations running 75%+ in production environments. Paired with the growth in more advanced use cases, this shows growing confidence in Kubernetes for diverse data workloads.

Challenges Remain: While acceptance and adoption is growing steadily, organizations still cite a lack of maturity of certain open source Kubernetes features as the #1 barrier. For organizations specifically running AI/ML workloads, additional hurdles emerge, including high compute costs and insufficient Kubernetes expertise within data teams.

Who We Talked To

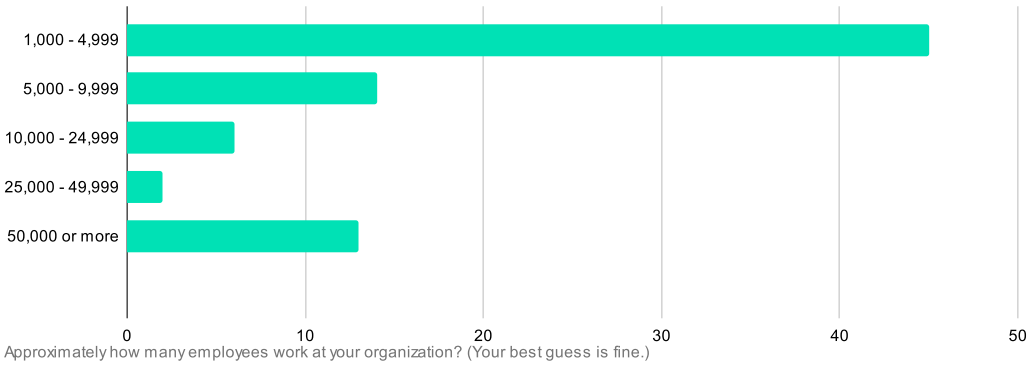
For the purpose of this report, the research firm only surveyed individuals whose organizations are currently running data workloads on Kubernetes. This included an international audience of 150 respondents.

Role



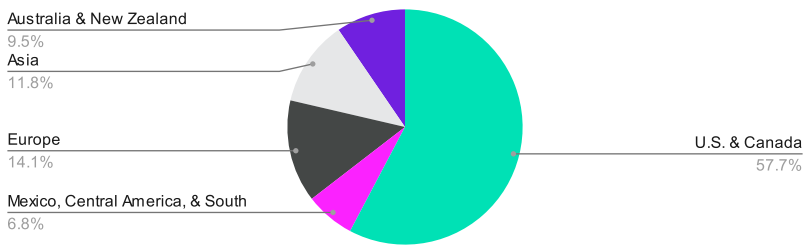
Which ONE of the following comes closest to describing your primary responsibility, even if you perform more than one or have a different title?

Org Size

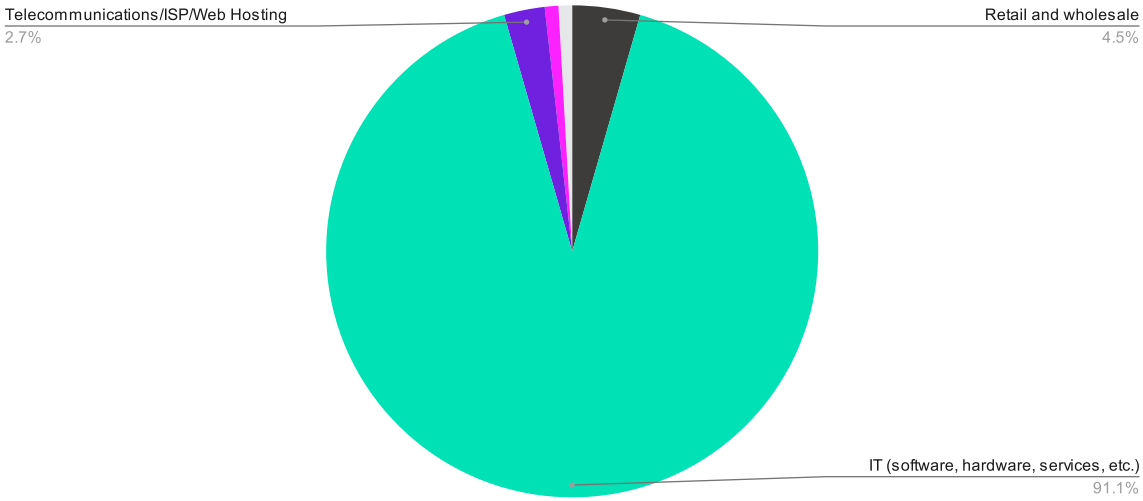


Approximately how many employees work at your organization? (Your best guess is fine.)

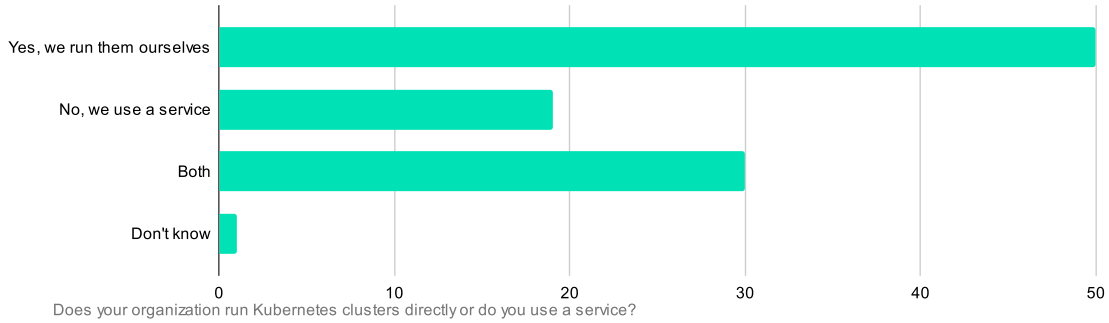
Region



Industry

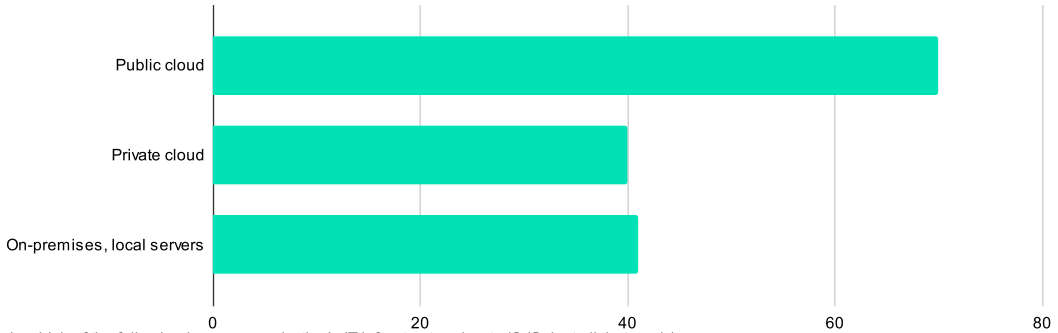


DIY or Hosted



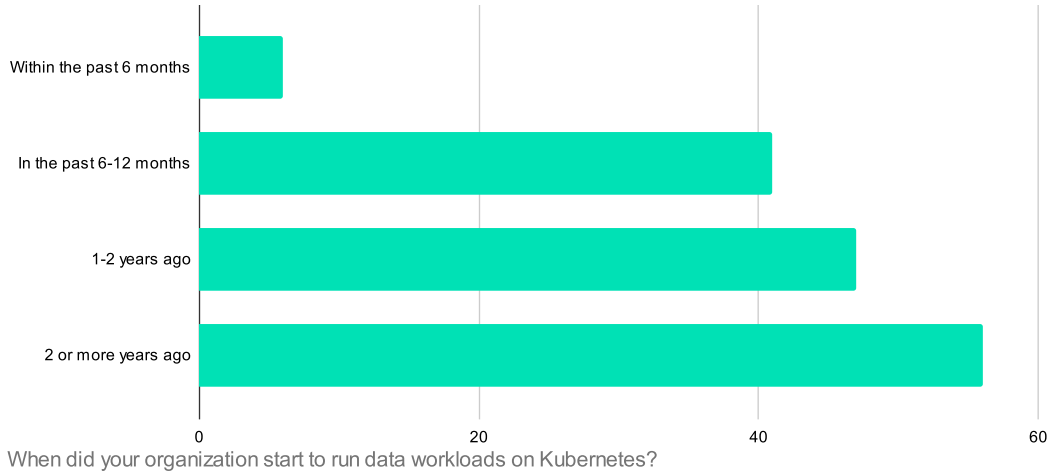
Does your organization run Kubernetes clusters directly or do you use a service?

Cloud



In which of the following is your organization's IT infrastructure hosted? (Select all that apply).

How Long Running DoK



When did your organization start to run data workloads on Kubernetes?

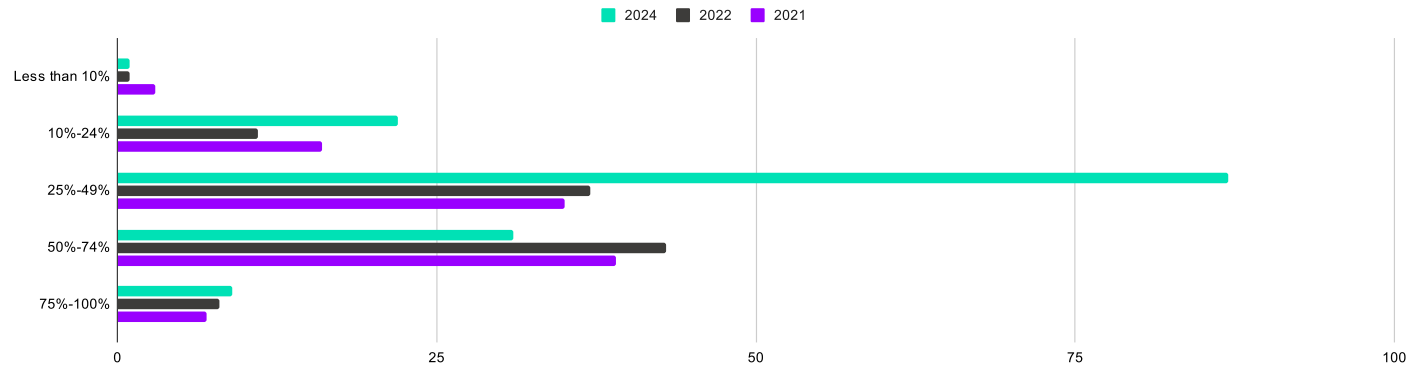
The Maturation of DoK

The journey of DoK from 2021 to 2024 reveals a clear pattern of maturation. What began as primarily a platform for basic stateful workloads (simple message queues, logging systems) has evolved into a comprehensive ecosystem supporting advanced data processing and AI/ML workloads. Almost half are running 50% or more of their DoK workloads in production, with the most advanced organizations running 75%+ DoK workloads in production environments.

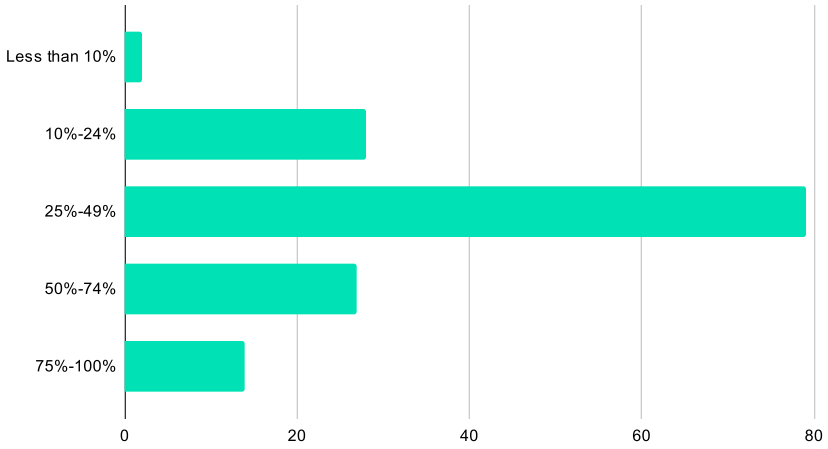
Watch: Database Patterns Whitepaper Panel Discussion

Percentage of DoK

What percentage of your organization's data workloads are running on Kubernetes? That is, of the total number of data workloads your organization is running, what percentage of them are running on Kubernetes?



Percentage of DoK in Production



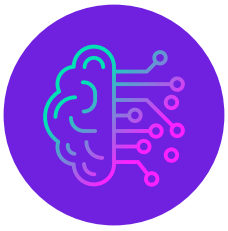
And what percentage of your data on Kubernetes workloads are running in production?

Evolution of DoK Workload Types 2021-2024

Year-over-year growth shows consistent expansion in both breadth and depth of workload types:

2021
Early focus on basic stateful workloads, with databases leading adoption

2022
Emergence of analytics and early AI/ML adoption

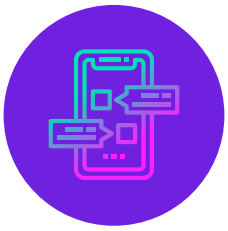


2024
Full spectrum of data workloads, with AI/ML growing

Our data shows the continuing evolution of workload types on Kubernetes:

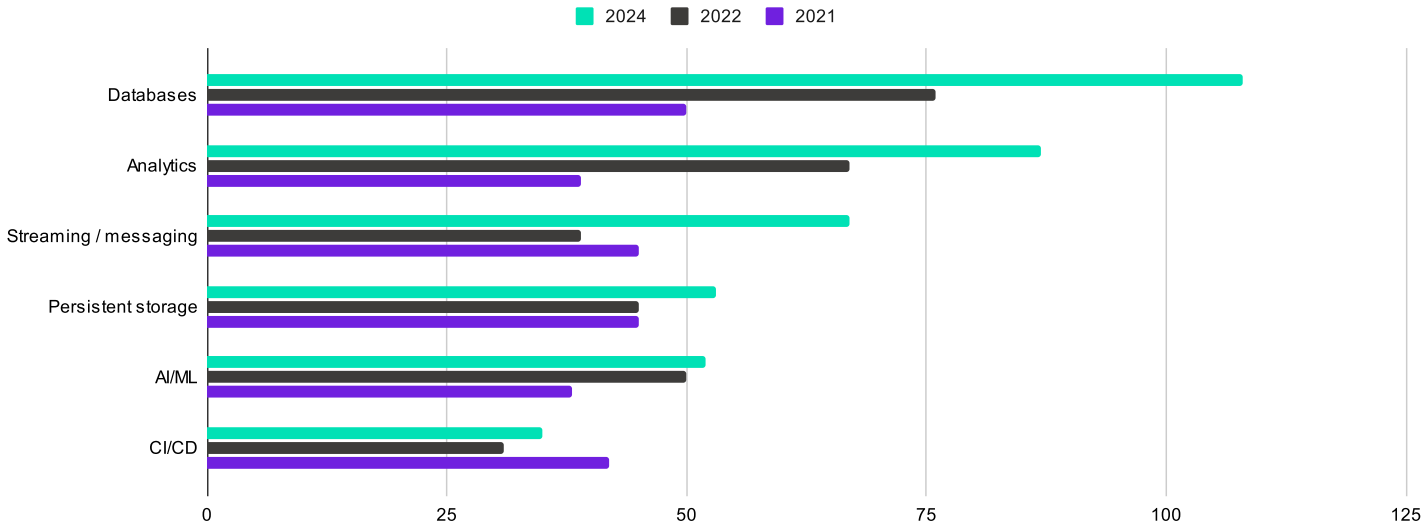
Databases
Maintaining their #1 position as the most commonly deployed workload across all three years

Analytics
Moving up to #2 in 2024 from a lower position in previous years



Streaming/messaging
Now #3, showing significant growth

DoK Workload Types



Which of the following data workloads does your organization run on Kubernetes?

Database Workloads: The Steady Foundation

Databases continue to be the cornerstone of DoK deployments. For the third consecutive year, databases remain the most common DoK workload, demonstrating the platform's reliability for critical data services. The consistency in database workload adoption demonstrates:

1. **Platform Reliability:** Organizations trust Kubernetes for critical data services.
2. **Operational Standardization:** Growing comfort with running databases on Kubernetes.
3. **Deployment Confidence:** Increased willingness to run production database workloads.

The AI/ML Revolution on Kubernetes

The emergence of AI/ML workloads on Kubernetes represents more than just a trend—it's a shift in how organizations approach their AI strategy. Our data shows that organizations are increasingly viewing Kubernetes as the foundation for their AI/ML infrastructure with many viewing Kubernetes as an important piece for AI strategy going forward (see Future Outlook section).

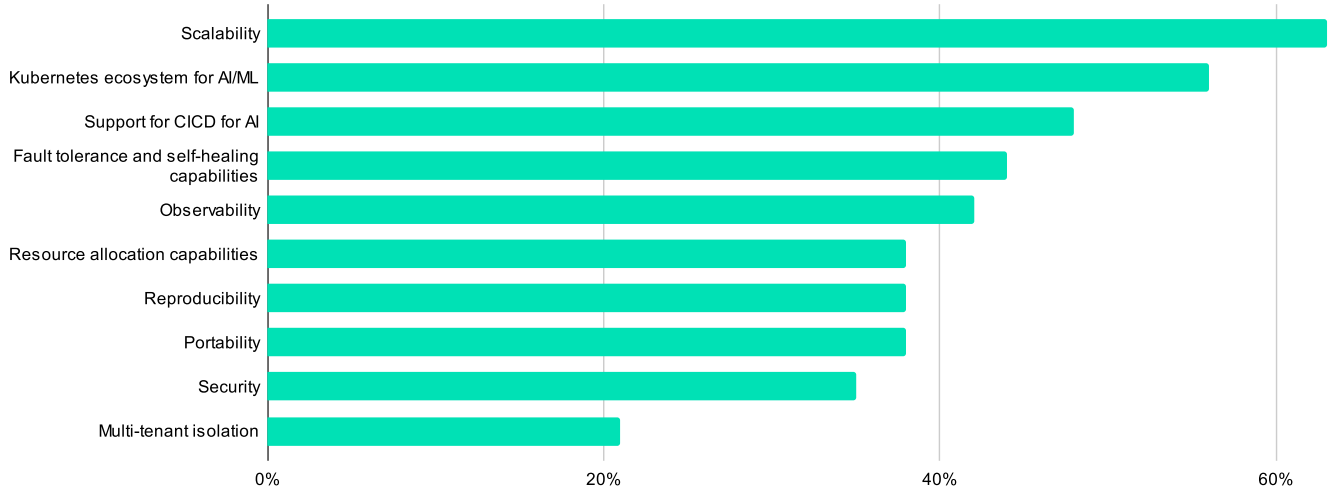
Watch: Adding Zonal Resiliency to Etsy's Kafka Cluster

- Organizations increasingly see Kubernetes as a distributed platform for their AI workloads
- Focus on optimizing resource utilization and reducing infrastructure costs
- Growing recognition of Kubernetes as key to providing scalability for AI applications

This shift is driven by several factors:

1. The need for scalable, efficient infrastructure for AI workloads.
2. Desire to standardize AI/ML operations alongside other workloads.
3. Growing ecosystem of AI-focused tools and operators.

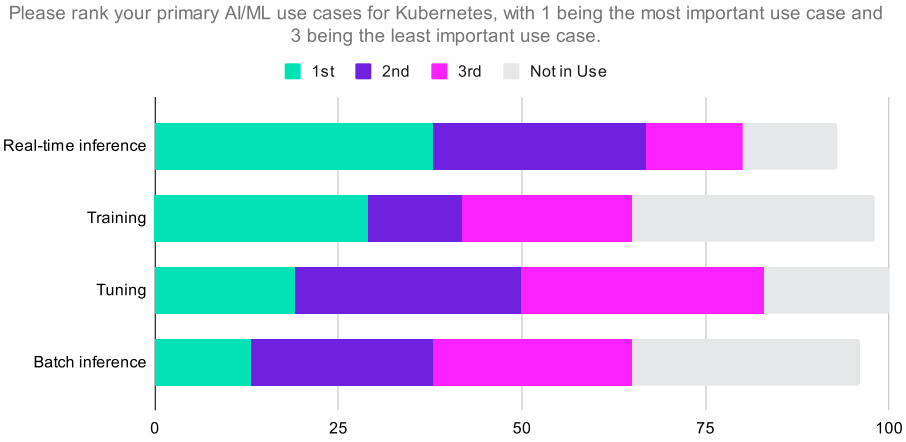
Why AI/ML and Kubernetes



Which of the following best explains the decision to use Kubernetes for AI/ML workloads at your organization? (Select all that apply).

Top AI/ML Use Cases

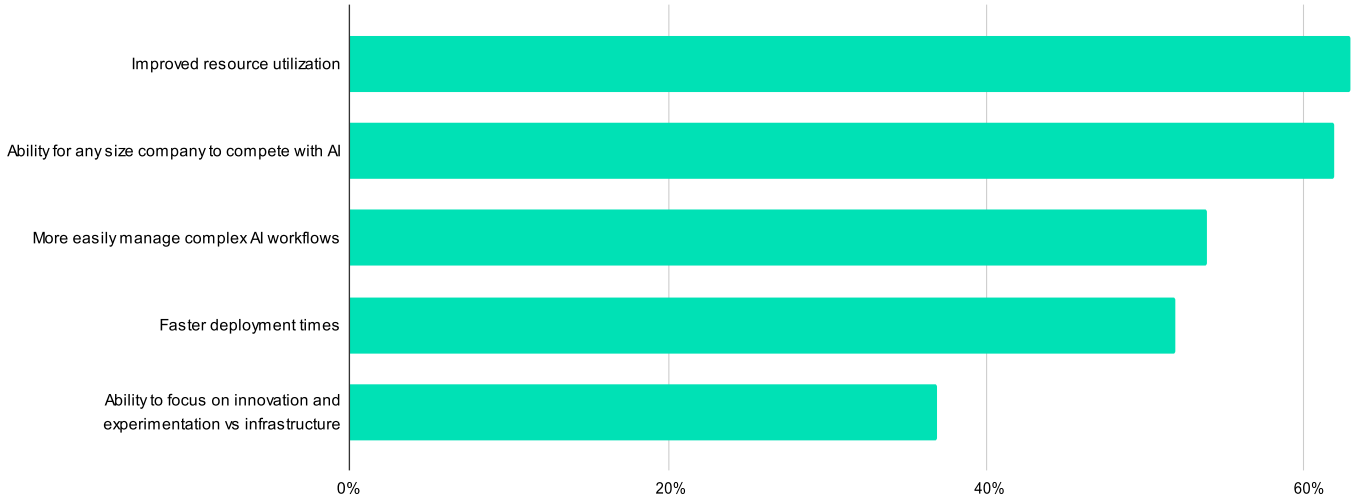
AI/ML and Kubernetes Use Cases



Key Benefits for AI/ML Workloads

The following key benefits—from improved resource utilization to faster deployments—reveal how Kubernetes is democratizing AI capabilities, allowing organizations to focus on innovation rather than infrastructure management while optimizing costs through better resource usage. This is particularly significant as the data show organizations increasingly view Kubernetes as a competitive advantage for their AI strategy and an important platform for AI acceleration.

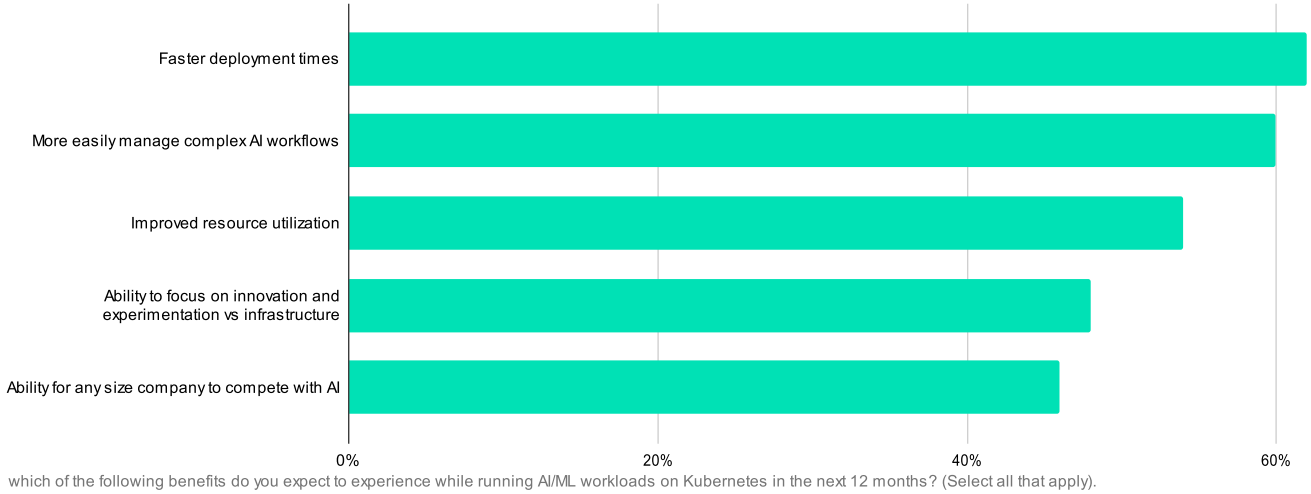
AI/ML on Kubernetes Benefits



Which of the following benefits have you experienced while running AI/ML workloads on Kubernetes?

Faster deployment times was cited as the number one benefit users expect to attain in the future. This signals a shift toward production-focused AI operations on Kubernetes and aligns with the report's findings of growing production deployments and the increasing view of Kubernetes as an important platform for accelerating AI strategy.

Future AI/ML on Kubernetes Benefits

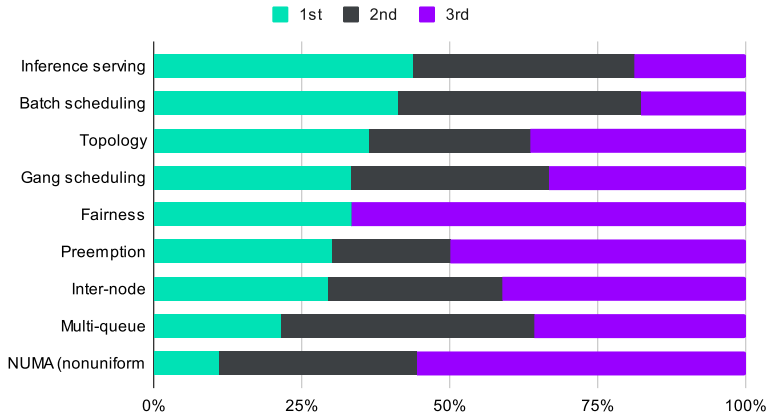


which of the following benefits do you expect to experience while running AI/ML workloads on Kubernetes in the next 12 months? (Select all that apply).

To achieve these anticipated benefits, organizations have identified specific technical capabilities needed in Kubernetes. The prioritization of features like batch scheduling, preemption, and gang scheduling reflects a sophisticated understanding of what's required to optimize AI/ML workloads at scale. These desired improvements focus on the fundamental mechanics of how Kubernetes handles resource-intensive AI workloads, particularly around scheduling and resource management—critical capabilities for achieving the faster deployment times and improved resource utilization that organizations expect. The Cloud Native AI Working Group is actively addressing these requirements.

AI/ML Desired Features

Please rank the top THREE features you would like to see Kubernetes improve for AI/ML workloads with 1 being the most important feature and 3 being the least important feature.



Ecosystem Challenges and Opportunities

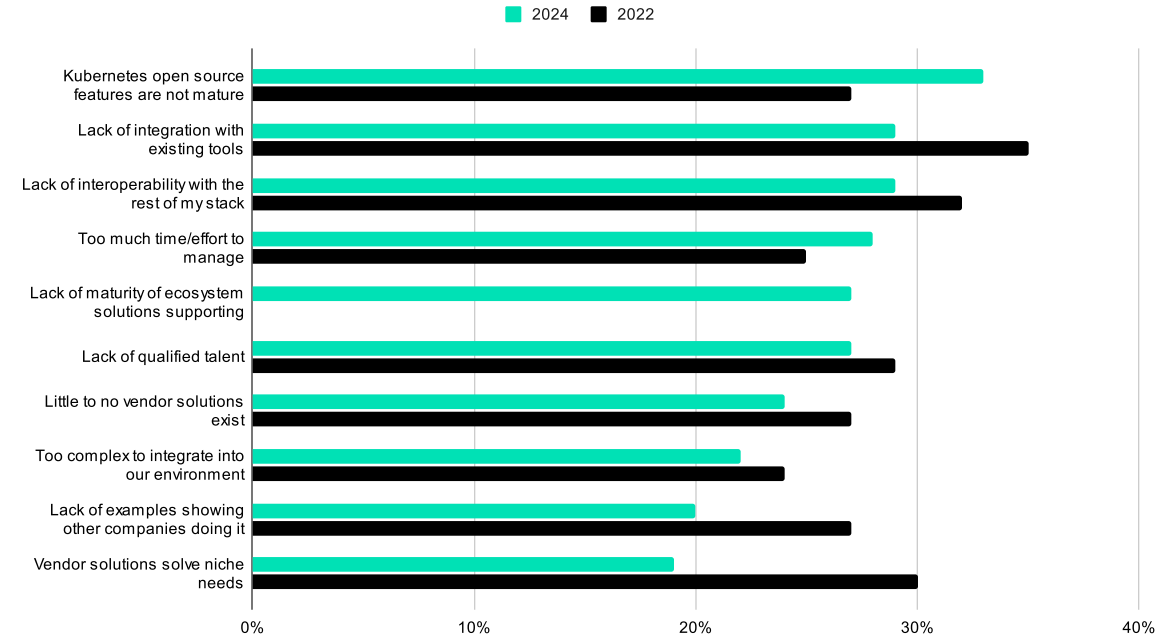
While DoK has matured significantly, several challenges remain that need to be addressed for the ecosystem to fully support advanced workloads.

Top 3 Challenges

1. Kubernetes open source features are not mature enough (↑ from #3 in 2022)
2. Lack of Integration with existing tools remains a significant hurdle year-over-year (steady at #2)
3. Interoperability with existing stack (↑ from #5 in 2022)

Watch: Kafka on Kubernetes: Reloaded for Fault Tolerance with Grab

All Challenges

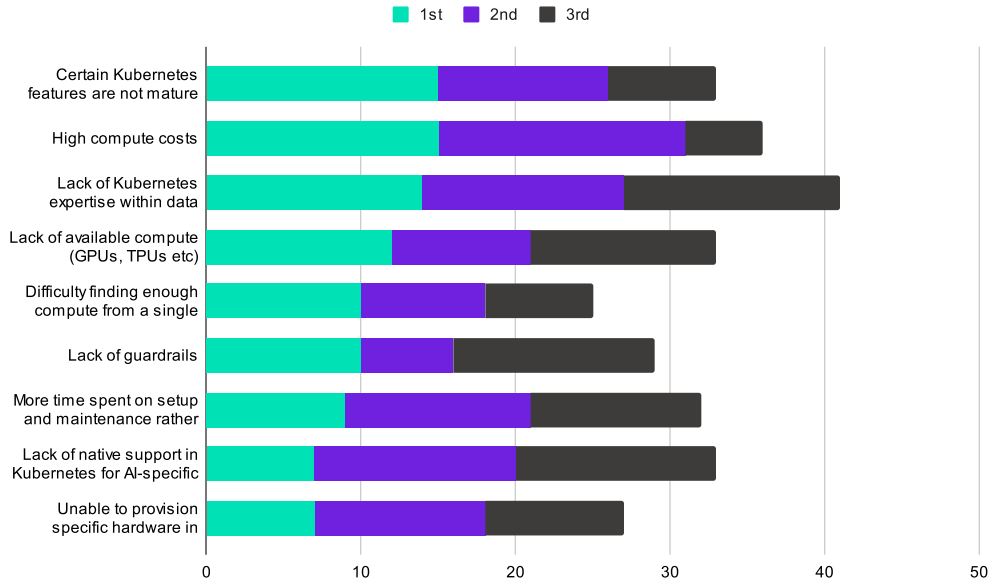


What are the top THREE challenges of running data workloads on Kubernetes? Select up to THREE.

AI/ML Challenges

AI/ML workloads on Kubernetes face a specific set of challenges. While feature maturity tops the list—consistent with the broader ecosystem—organizations running AI/ML workloads report two distinctive barriers: the high computational costs of these resource-intensive workloads and a significant expertise gap between data teams and Kubernetes operations.

AI/ML Barriers

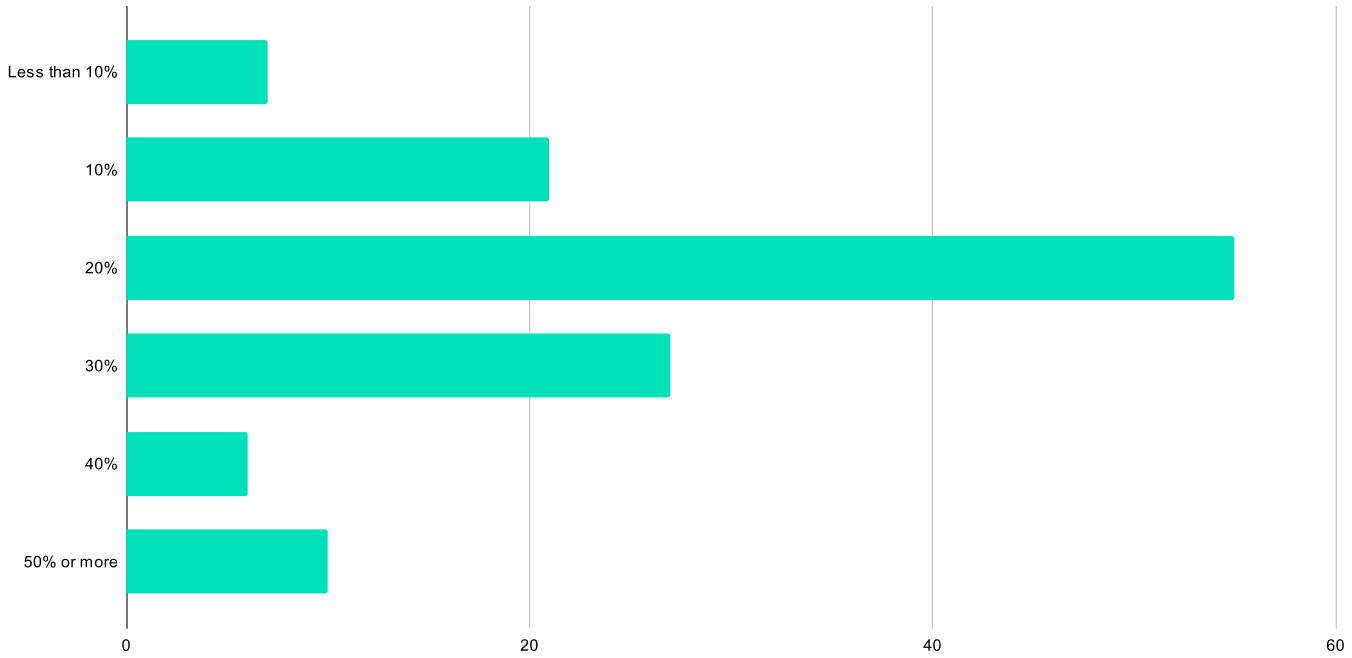


Please rank the top THREE barriers for deploying and managing AI/ML workloads on Kubernetes with 1 being the most difficult barrier and 3 being the least difficult barrier.

Future Outlook

The data shows clear trends for the future of DoK, with organizations making commitments to expand their usage.

DoK Increase Next 12 Months

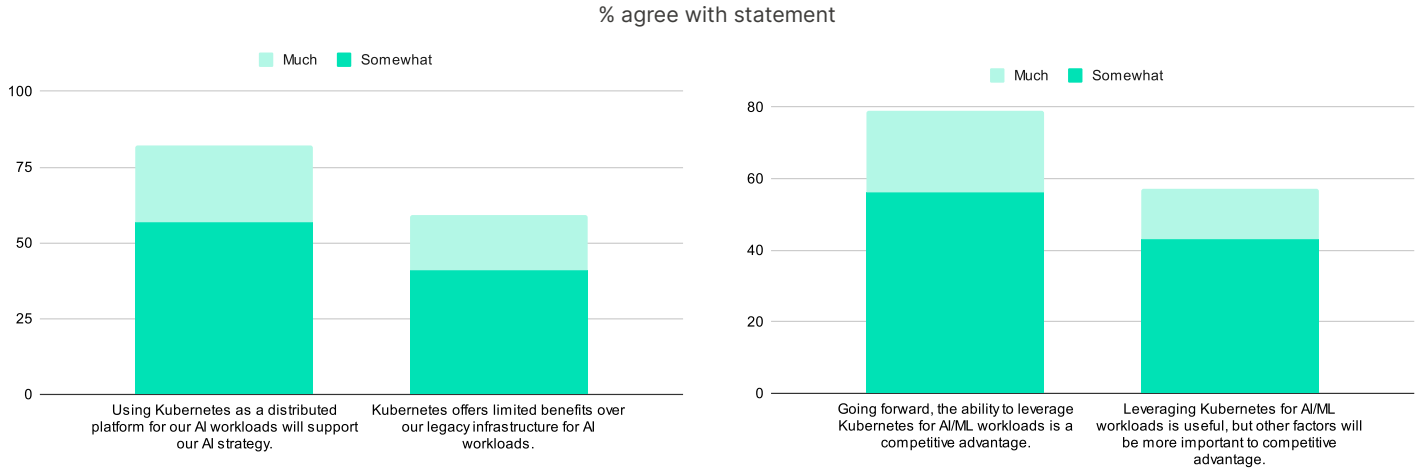


By how much do you expect your organization's percentage of data workloads run on Kubernetes to increase in the next 12 months? n=126

Key areas of future focus include:

Kubernetes Supports AI Strategy

K8s for AI/ML is a Competitive Advantage



Now you will see a series of pairs of statements. For each pair select the statement that you agree with more, even if you agree with both a little.

Methodology

This report is based on a survey of 150 technology leaders conducted in September-October 2024. Respondents represent organizations across various industries and sizes, with a focus on those actively using Kubernetes for data workloads.

The survey methodology included:

- Quantitative analysis of workload types and adoption patterns
- Qualitative assessment of challenges and opportunities
- Cross-reference with previous years' data for trend analysis

DoK Community

About Data on Kubernetes Community

The Data on Kubernetes Community (DoKC) is an openly governed group of practitioners sharing in the emergence and development of techniques for the use of Kubernetes for data. Founded in June 2020, DoKC exists to assist in the emergence and development of techniques for the use of Kubernetes for data.

Our community continues to grow and evolve, providing resources, best practices, and forums for knowledge sharing among practitioners. Through regular meetups, workshops, and online events, we foster collaboration and innovation in the DoK ecosystem.

https://dok.community