



DoK  
Community

RESEARCH REPORT

# Data on Kubernetes 2021

Insights from over 500 executives and technology leaders on how Kubernetes is being used for data and the factors driving further adoption

# Introduction

Regardless of which survey you read, the use of Kubernetes is on the rise in organizations of all sizes. While Kubernetes was initially designed for stateless workloads, the community has made major strides in supporting stateful workloads in the past few years resulting in more organizations running them in production.

As organizations become more data-driven and increasingly turn to real-time data for competitive advantage, their infrastructure needs to evolve to accommodate the collection, storage, and processing of data across different environments (edge, public cloud, and on-premise). In providing a standard way to run stateless and stateful workloads, Kubernetes is strategically positioned to be the platform organizations can leverage to build state of the art data infrastructures.

**[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. In September 2021, we engaged research firm Clearpath Strategies to survey over 500 Kubernetes users to understand the types and volume of data-intensive workloads being deployed in Kubernetes, benefits and challenges, and the factors driving further adoption.

## Key Findings

- Kubernetes has become a core part of IT – half of the respondents are running 50% or more of their production workloads on it, and they are very satisfied and more productive as a result. The most advanced users report 2x or greater productivity gains.
- 90% believe it is ready for stateful workloads, and a large majority (70%) are running them in production with databases topping the list. Companies report significant benefits to standardization, consistency, and management as key drivers.
- Significant challenges remain. As they seek to expand their data on Kubernetes footprint, enterprises find a lack of integration and interoperability with existing tools and stacks; skilled staff; quality of Kubernetes operators; and trusted vendors.
- Business demands are creating pressures for further adoption. The increasing importance of real-time data to competitive advantage will sharpen companies' need to run data on Kubernetes. A majority believe standards will improve data management and that data should become declarative.

# Data on Kubernetes Community

## Founding Sponsors

### PLATINUM

DataStax



### GOLD



### SILVER



# Who we talked to

For the purpose of this report, the research firm only surveyed individuals whose organizations are currently using, evaluating, or planning to use Kubernetes. This included an international audience of 502 respondents.

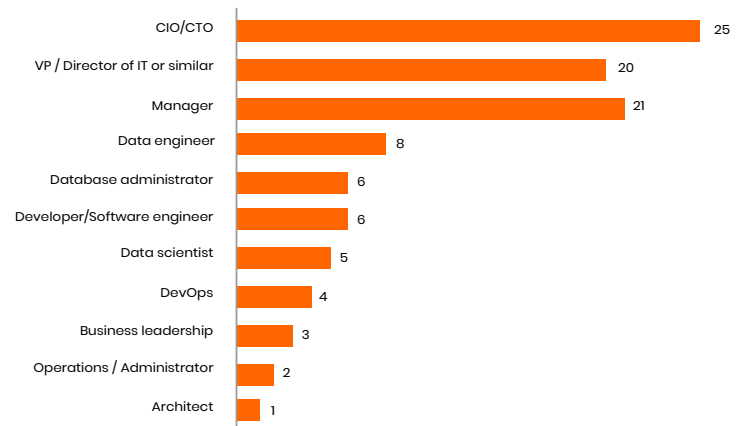
Because data on Kubernetes impacts a large part of an organization's IT team – from CIOs and CTOs, VPs and Directors of IT, Software and Site-reliability engineers, Database administrators, and Data engineers – the targeted demographic ratio is practitioners (~35%), managers (~20%), and executives (~45%).

The majority of respondents (49%) came from Technology organizations (software, hardware, services), followed by Financial Services (12%), Manufacturing and Heavy Industry (8%), and Telecommunications (6%).

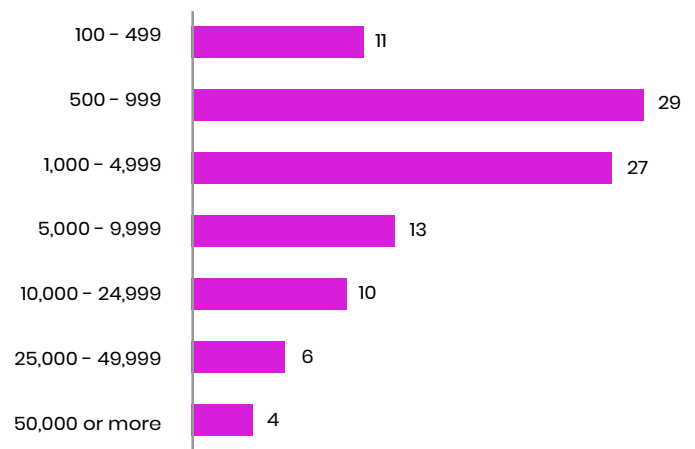
Respondents were from companies using or evaluating Kubernetes including a mix of enterprise (~60%), mid-market (~30%), and small business (~10%) providing a wide view of the current data on Kubernetes landscape across a diversity of company sizes and sectors.

Finally, we wanted to understand organizations' IT infrastructure hosting strategy which affirmed that overall, organizations operate in a hybrid, multi-cloud world.

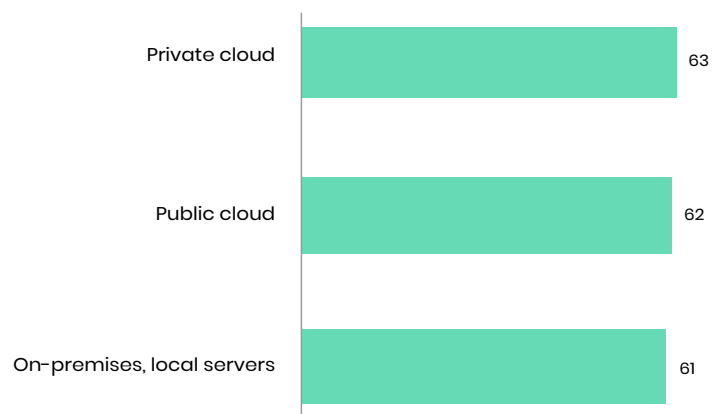
## ROLE



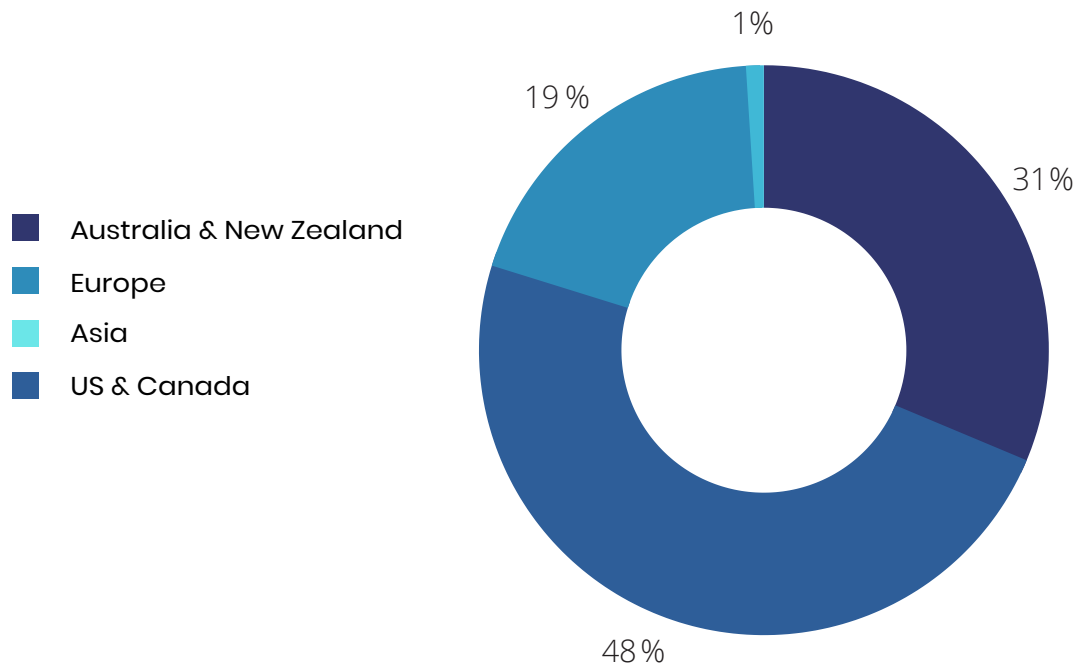
## ORGANIZATION SIZE



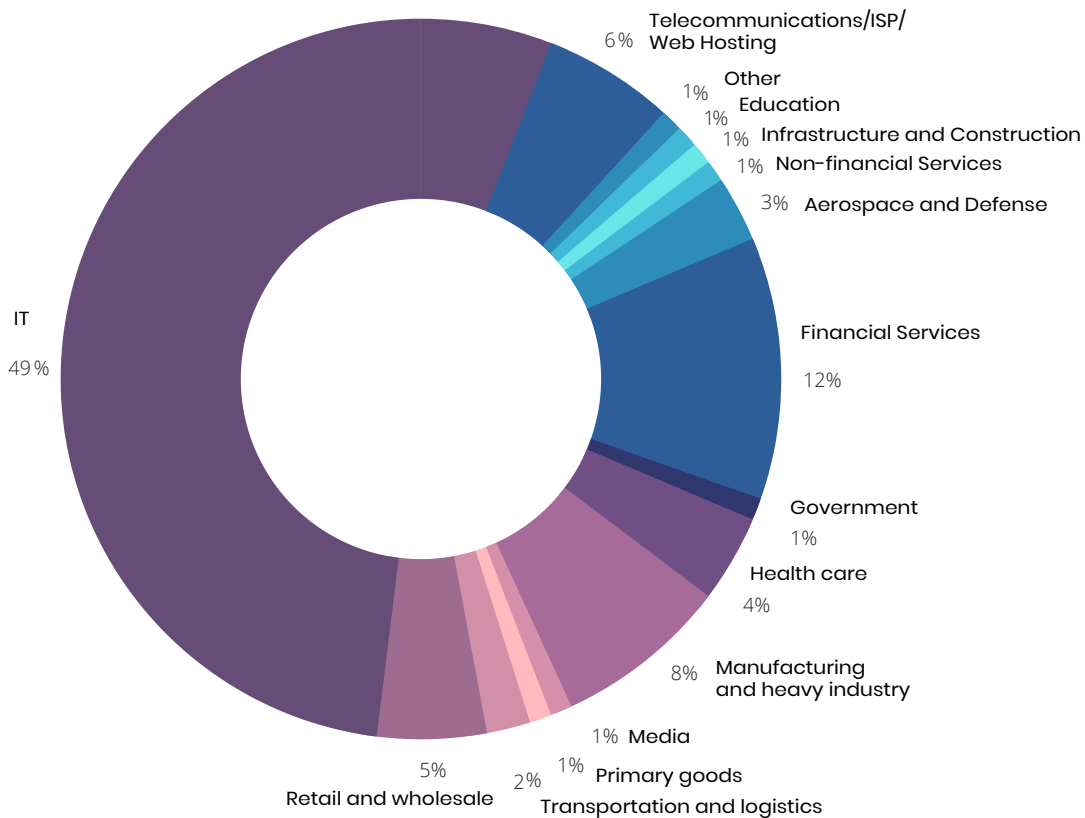
## HYBRID ENVIRONMENTS



## REGION



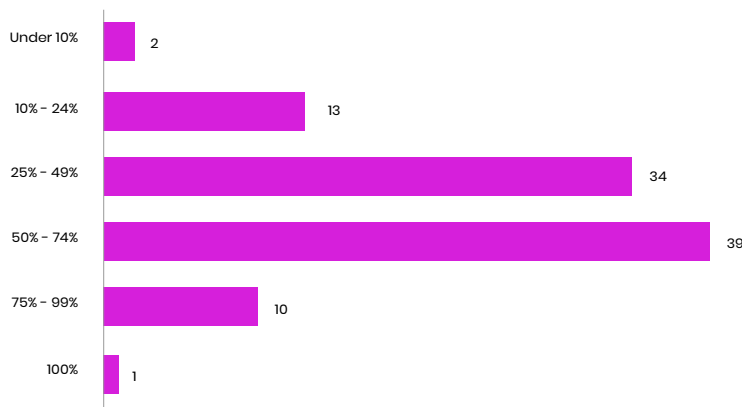
## INDUSTRY



# How Kubernetes is Used Today

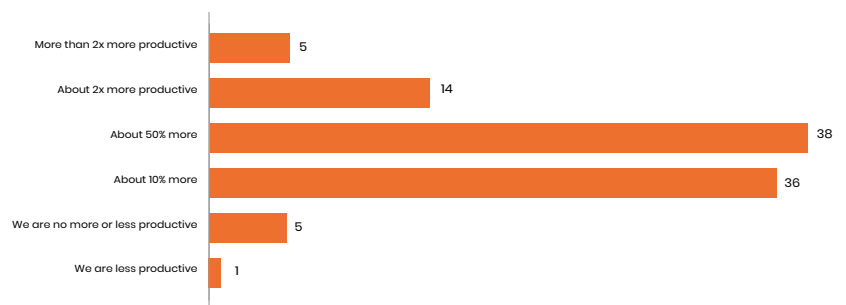
Before diving into the specifics of running data on Kubernetes, we wanted to understand how organizations are using Kubernetes across all workloads. We found that Kubernetes has become a core part of their infrastructure – half of the respondents are running 50% or more of their production workloads on it.

## What percentage of your organization's production workloads run on Kubernetes?



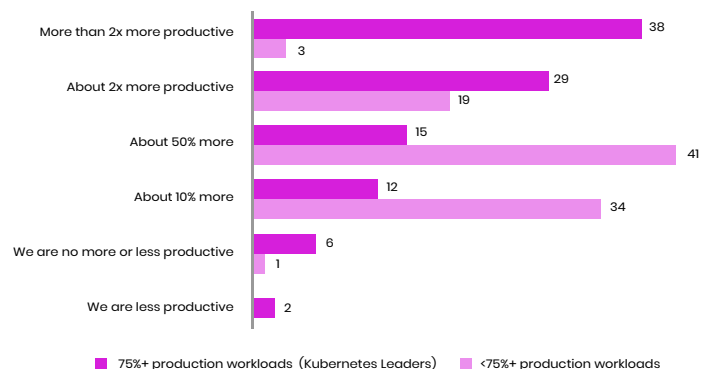
Those using it in production are happy with the results, with 69% being very satisfied. This satisfaction might be driven by the fact that over half of respondents report being 50% or more productive after adopting Kubernetes. Unsurprisingly 68% say they are very likely to increase their Kubernetes footprint.

## How much more productive is your organization/are your developers after adopting Kubernetes?



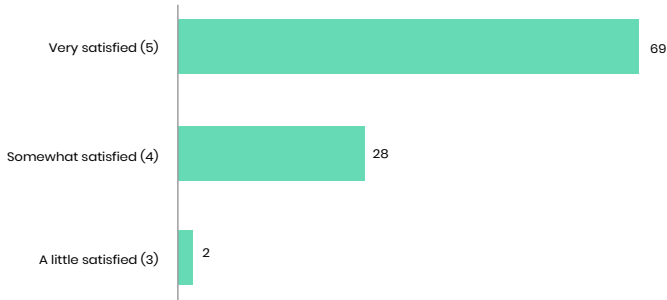
The more Kubernetes is in use; the more productive an organization is: when we look at Kubernetes Leaders – the 11% of respondents running 75% or more of their production workloads on Kubernetes – they report even higher levels of productivity with a majority achieving an impressive 2x or more productivity.

## A majority of Kubernetes Leaders are 2x or more productive

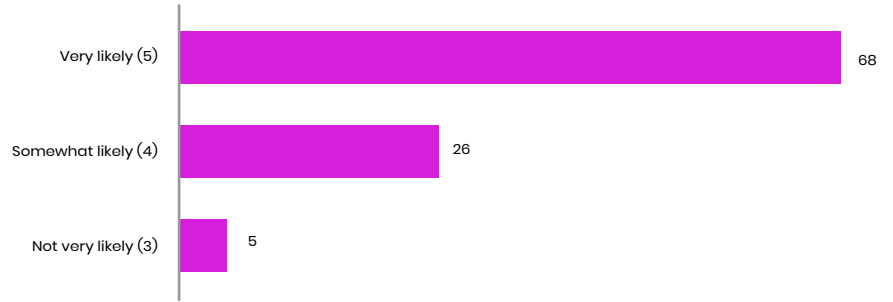


Kubernetes adoption is recent for many respondents with 86% adopting in the past two years with many in the past six to 12 months – a likely correlation with the digital transformation accelerated by the pandemic.

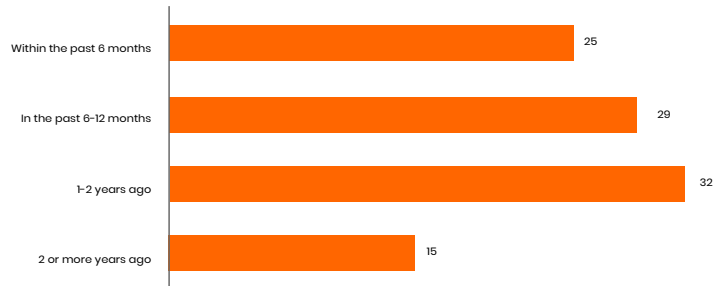
### In general, how satisfied are you with the use of Kubernetes for production workloads in your organization?



### How likely are you to migrate additional production workloads to Kubernetes in your organization?



### When did your organization start to use Kubernetes?



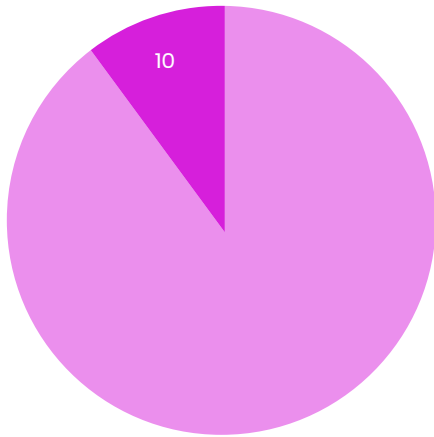
# The State of Data on Kubernetes

## Data on Kubernetes is widely adopted

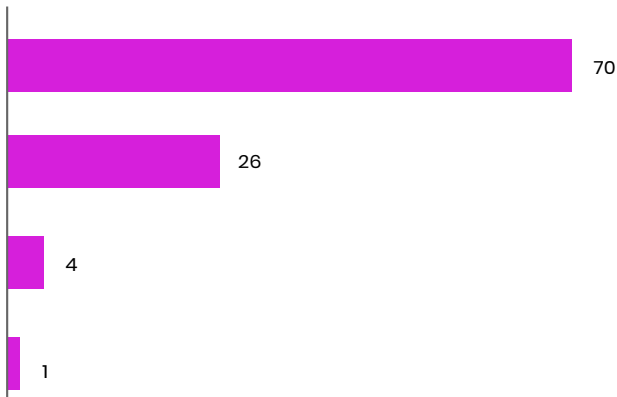
The data is clear: Enterprises are confident that Kubernetes is ready to run their organization's stateful workloads in production with 90% believing it and 70% currently doing so. (We'll cover the 10% of "non-believers" in the next section).

For those currently running stateful workloads in production, we see broad usage that maps closely to overall Kubernetes usage in the previous section. Over half of this cohort (51%) intends to increase the volume of workloads by 30% or more in the next 12 months. More Kubernetes = more productivity = high satisfaction. See a pattern here? Kubernetes is here to stay.

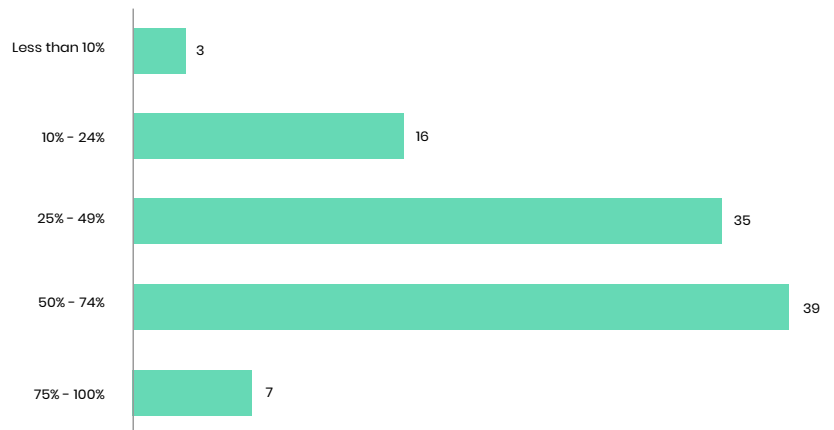
Do you believe Kubernetes is ready to run your organization's stateful workloads in production?



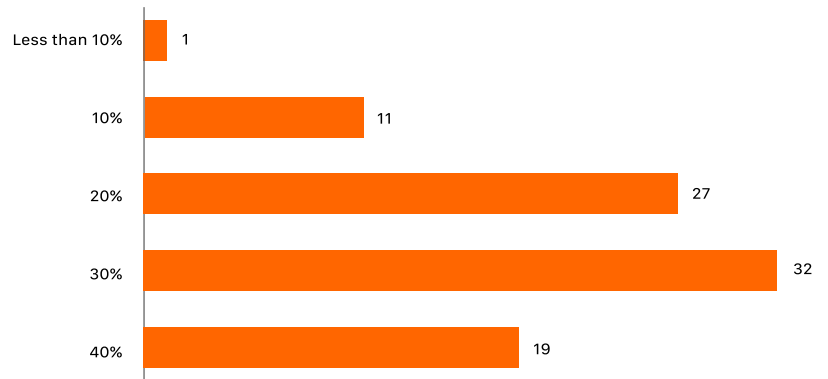
Does your organization run stateful workload(s) on Kubernetes?



What percentage of your organization's stateful production workloads are running on Kubernetes?



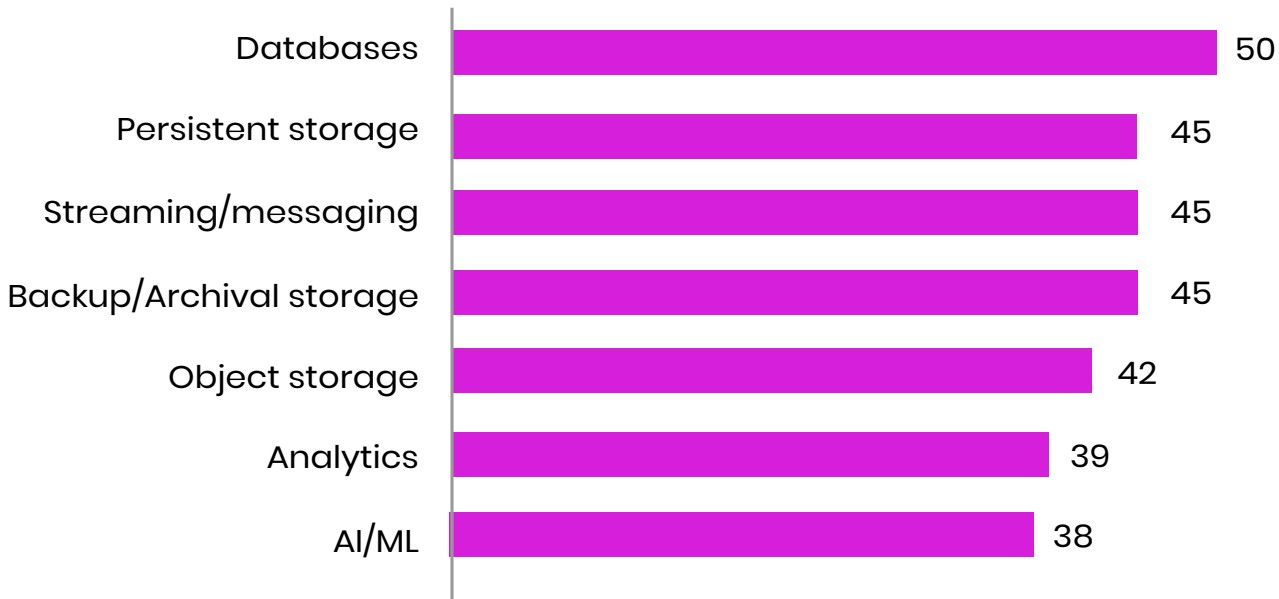
By how much do you expect your organization's percentage of stateful production workloads run on Kubernetes to increase in the next 12 months?





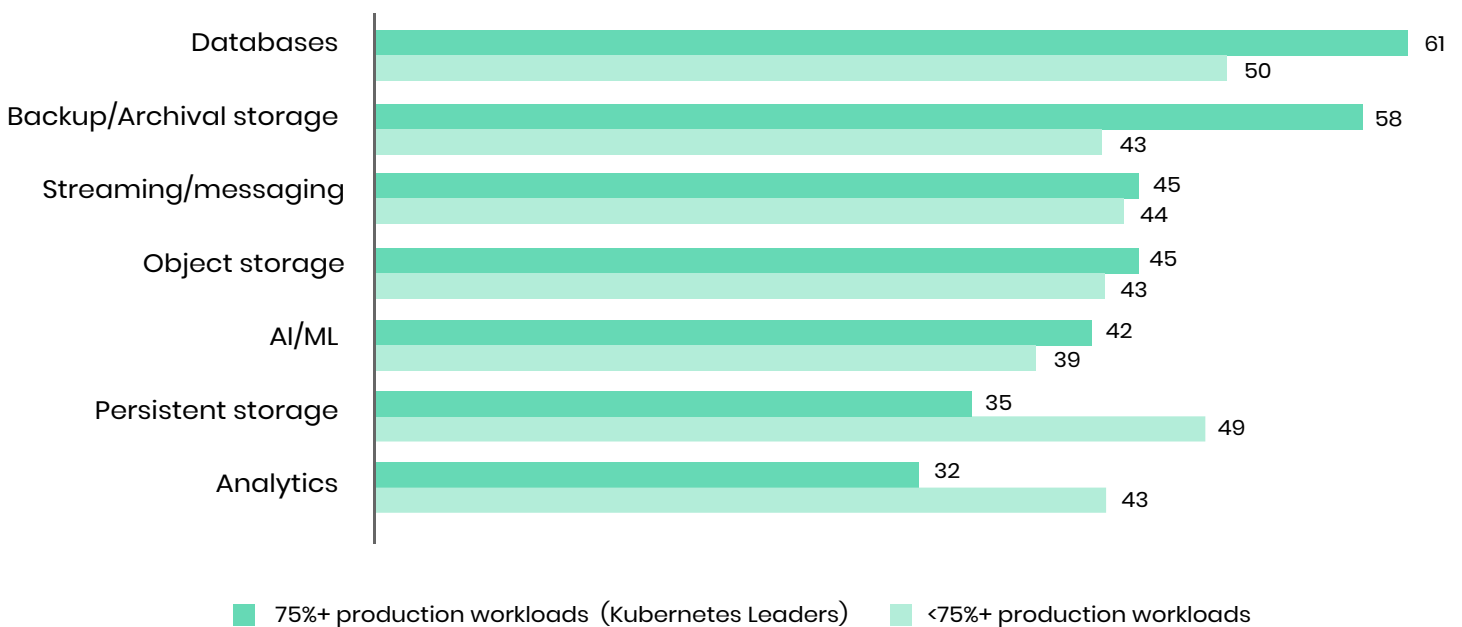
Respondents are running a wide range of stateful workloads on Kubernetes with the Databases in the top spot followed by a three-way tie including Persistent Storage, Streaming/Messaging, and Backup/Archival Storage.

## Q. Which of the following stateful workloads does your organization run on Kubernetes?



When we look at Kubernetes Leaders, Databases remain in the top spot but become more important – jumping 11% – while Persistent Storage becomes less important to this group, a reduction of 14% compared to all respondents.

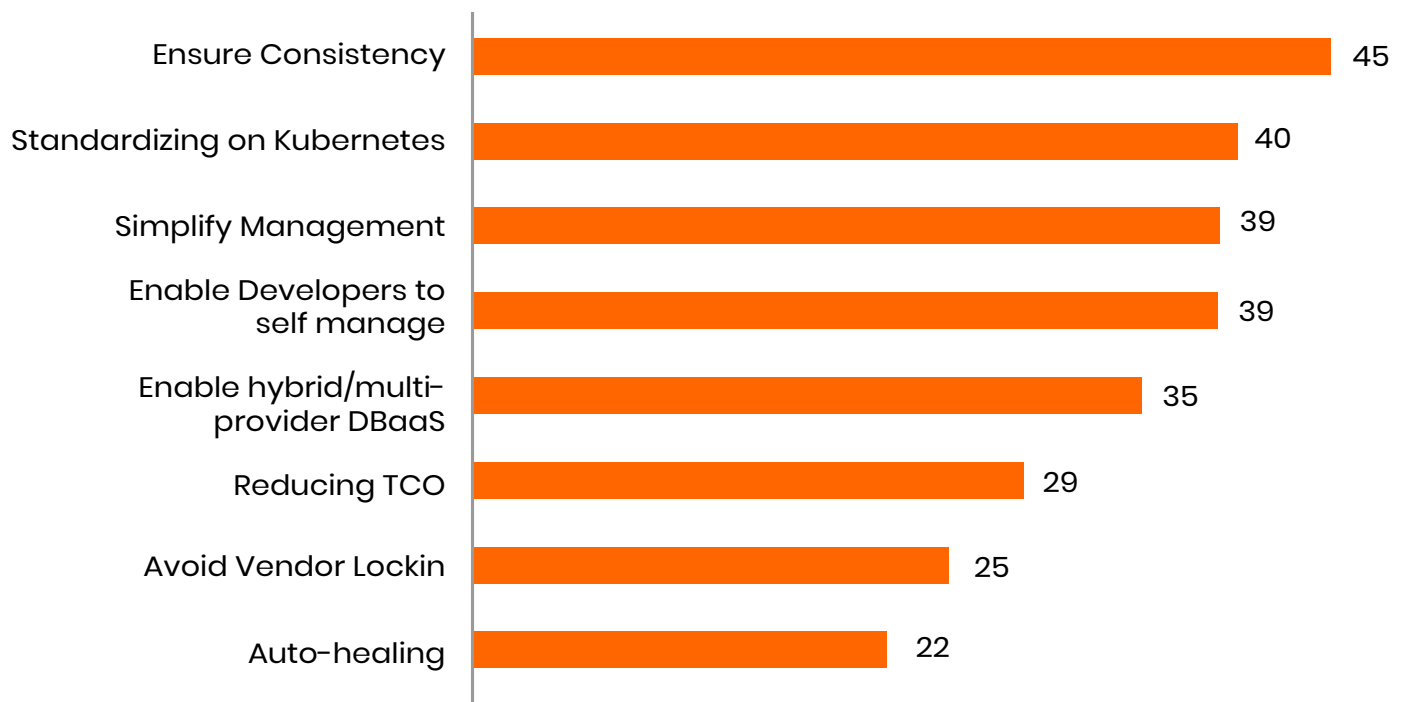
## More databases and storage for Kubernetes leaders



## Standardization drives DoK adoption

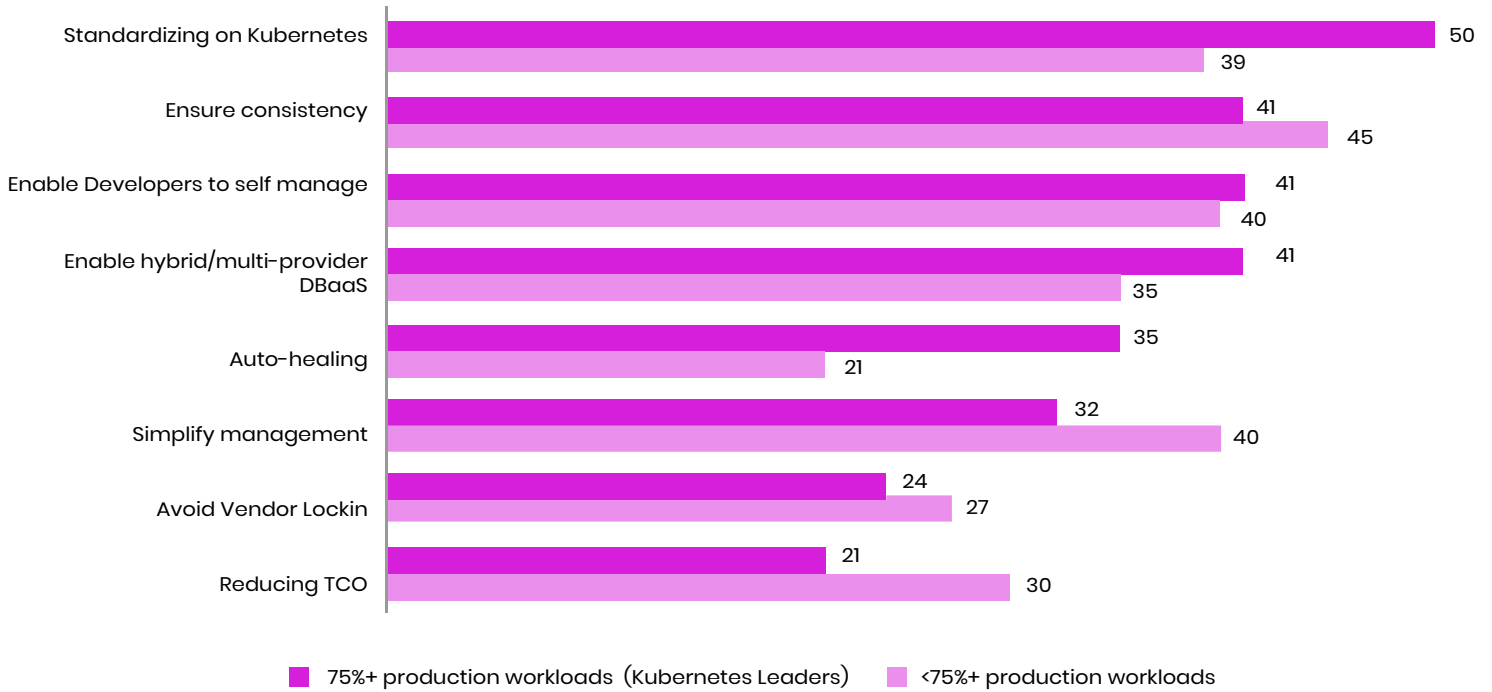
In spite of the difficulty of running stateful workloads on Kubernetes (more in this in the next section), organizations have embraced it. They do this not only because Kubernetes makes it easy to scale – a Day 1 benefit – but also due to the ability to manage all workloads in a standard way, generally considered a Day 2 benefit. The massive productivity gains and high satisfaction we see in the first section may be linked to the ability for organizations to standardize across hybrid environments with Kubernetes – a data point we will explore in our next survey.

From the following, which are the **THREE** most important factors in your organization’s decision to run stateful workload(s) on Kubernetes?



This becomes even more pronounced when we look at responses from Kubernetes Leaders. Not only does standardization jump 10 points, we also see security jump ahead of scalability, deployment, and other Day 1 benefits of Kubernetes. The more workloads an organization runs on Kubernetes, the more it can capitalize on the standardization advantage.

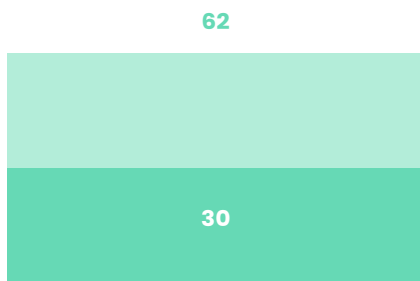
## Standardization is the key driver for Kubernetes Leaders



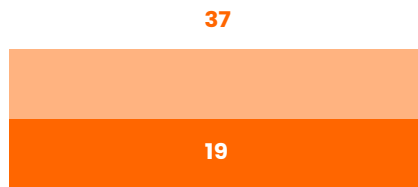
Looking ahead to a future where organizations can more seamlessly react to data in real-time, a majority would like to see data become declarative, just like Kubernetes. Our next survey will delve deeper into what will be required to create a language for declarative data on Kubernetes.

## A majority believe that data should become declarative, just like Kubernetes.

■ Much  
■ Somewhat



In the future, the way data is managed should become declarative, just like Kubernetes.



Data management is imperative because it is too complex.

While many organizations have experienced success with Kubernetes and are running stateful workloads, challenges remain. Next we'll zoom in on how respondents are managing state in Kubernetes (there's an operator for that), and the key challenges they face.

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.

# The Stateful Challenge

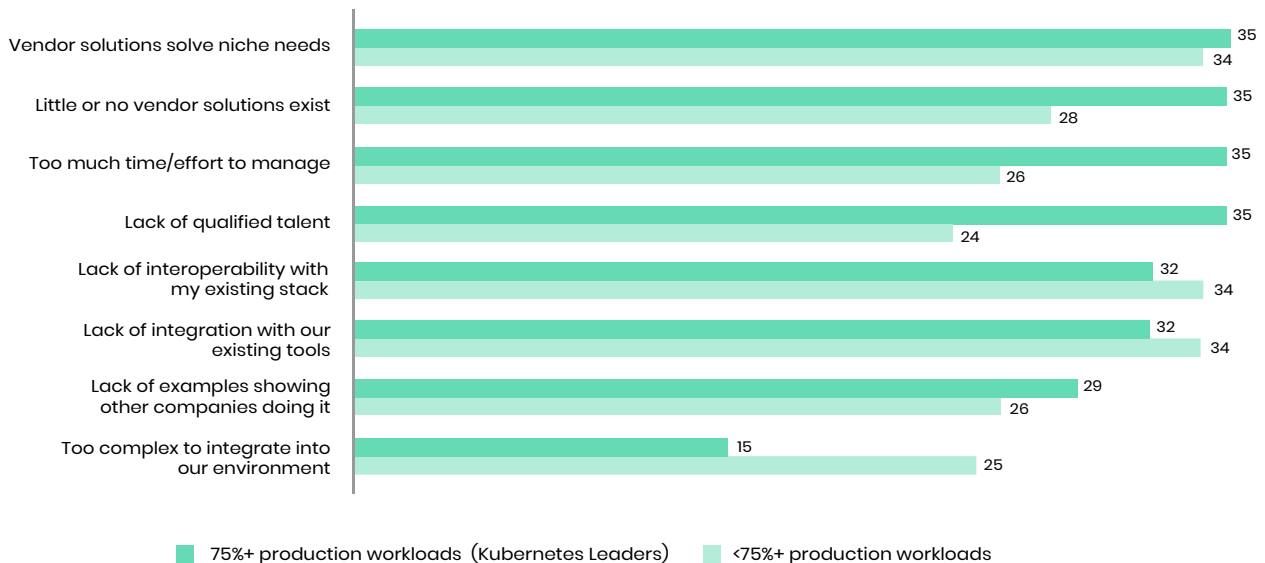
If more Kubernetes = more productivity = higher satisfaction, then what challenges do our respondents face when running data on Kubernetes? Here we see a mix of Day 1 and Day 2 problems emerge, with the primary challenge being the lack of integration with existing tools (35%) followed by a lack of interoperability with the rest of their stack (32%).

## What are the primary challenges of running data on Kubernetes?



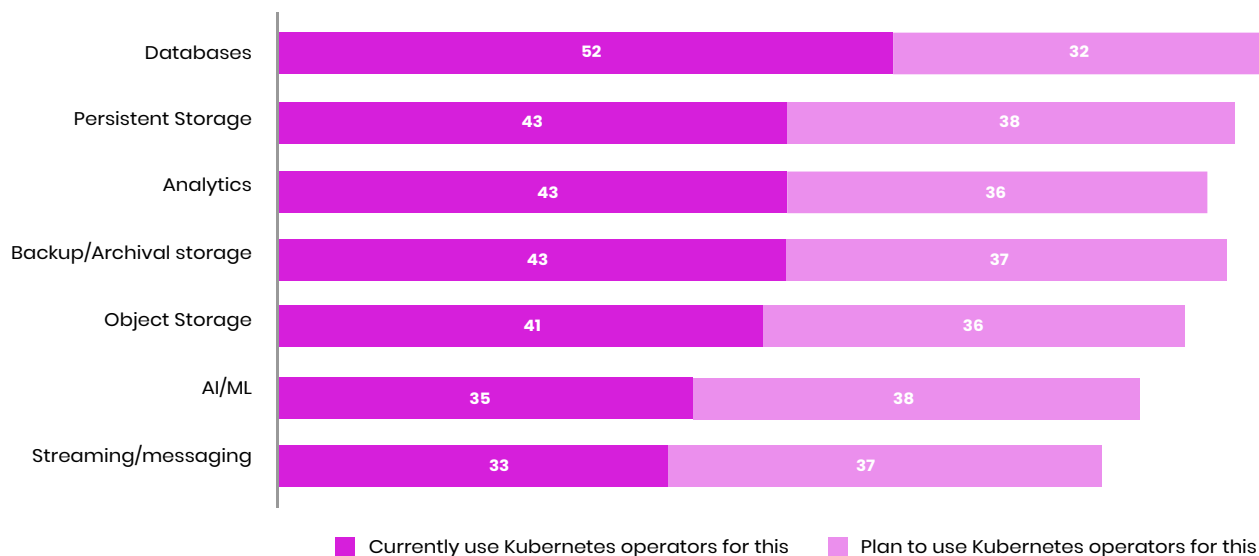
Kubernetes Leaders face a different set of challenges with a four-way tie for first place: vendor solutions solve niche needs, little or no vendor solutions exist, too much time and effort to manage, and lack of qualified talent. The talent gap was the most drastic difference when compared with all respondents, jumping 11%.

## Kubernetes Leaders face a lack of support and skills



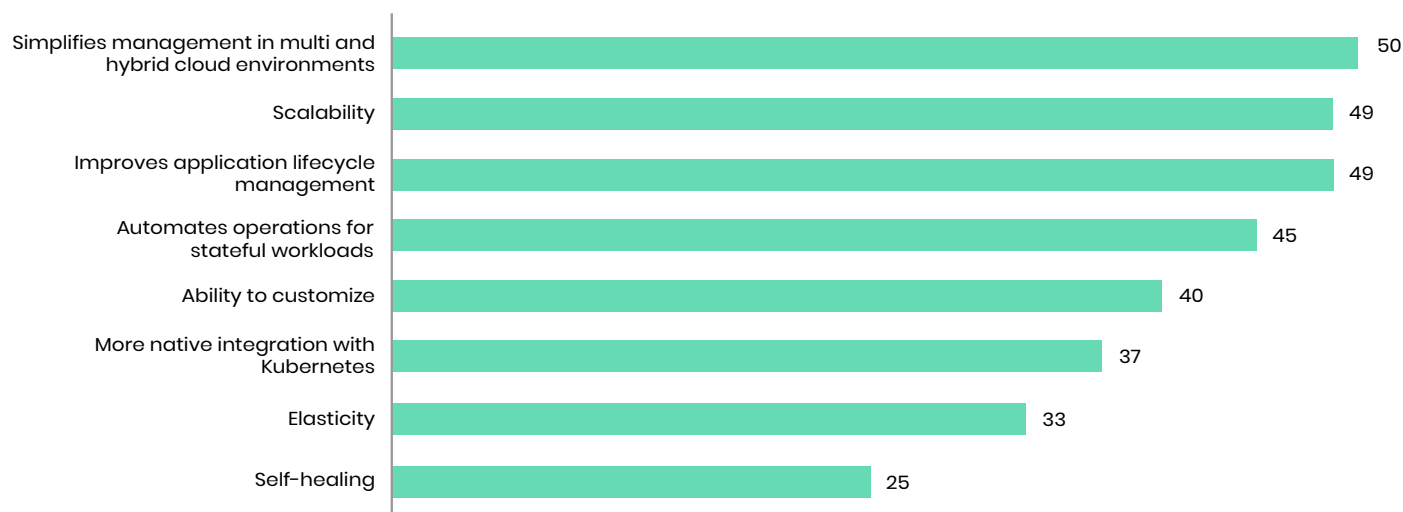
In addition to native Kubernetes features, like StatefulSets and the Container Storage Interface (CSI), the **operator pattern** extended Kubernetes' use to stateful workloads. It uses the Kubernetes API to create, configure, and manage instances of complex stateful applications on behalf of a Kubernetes user. Today, respondents are using operators to manage a wide range of stateful workloads from databases to streaming/messaging.

Please indicate whether you, professionally, currently use Kubernetes operators, plan to use Kubernetes operators, would like to use Kubernetes



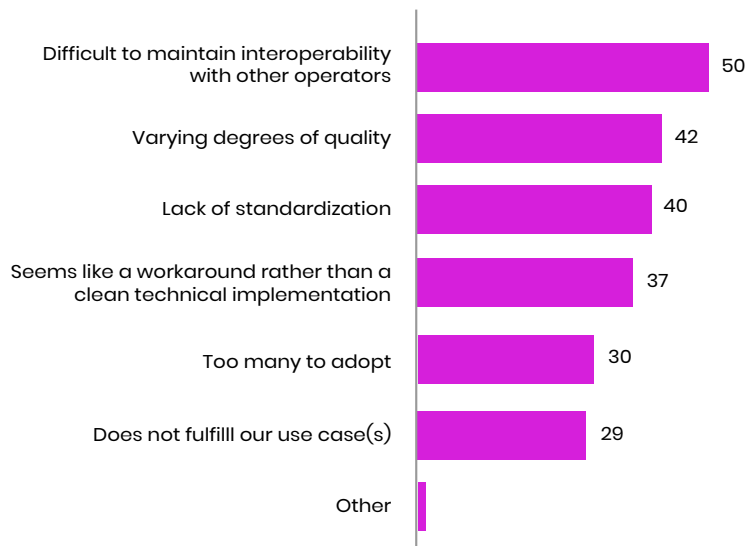
The primary benefit of Kubernetes operators cited by respondents is that they simplify the management of workloads in multi-cloud and hybrid cloud environments. Simplicity and scalability also rank highly.

What would you say are the primary benefits of using Kubernetes operators? (Select all that apply)

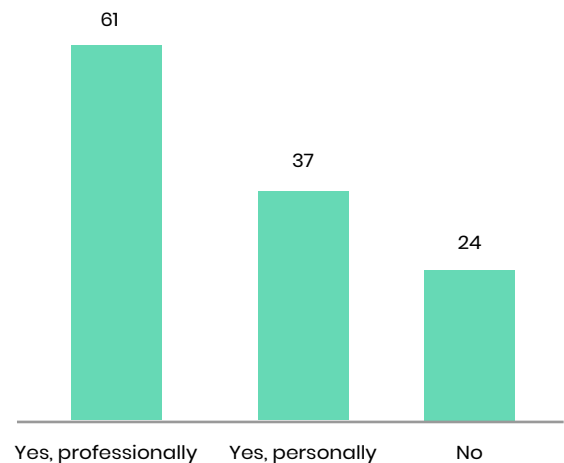


Without industry standards for the development of operators, key challenges remain and may hinder the broader adoption of Kubernetes for stateful workloads. Interoperability is cited as the primary challenge (50%), followed by varying degrees of quality (42%) and lack of standardization (40%). A consequence of this is that a majority of the respondents we surveyed are developing their own operators professionally.

### CHALLENGES WITH OPERATORS

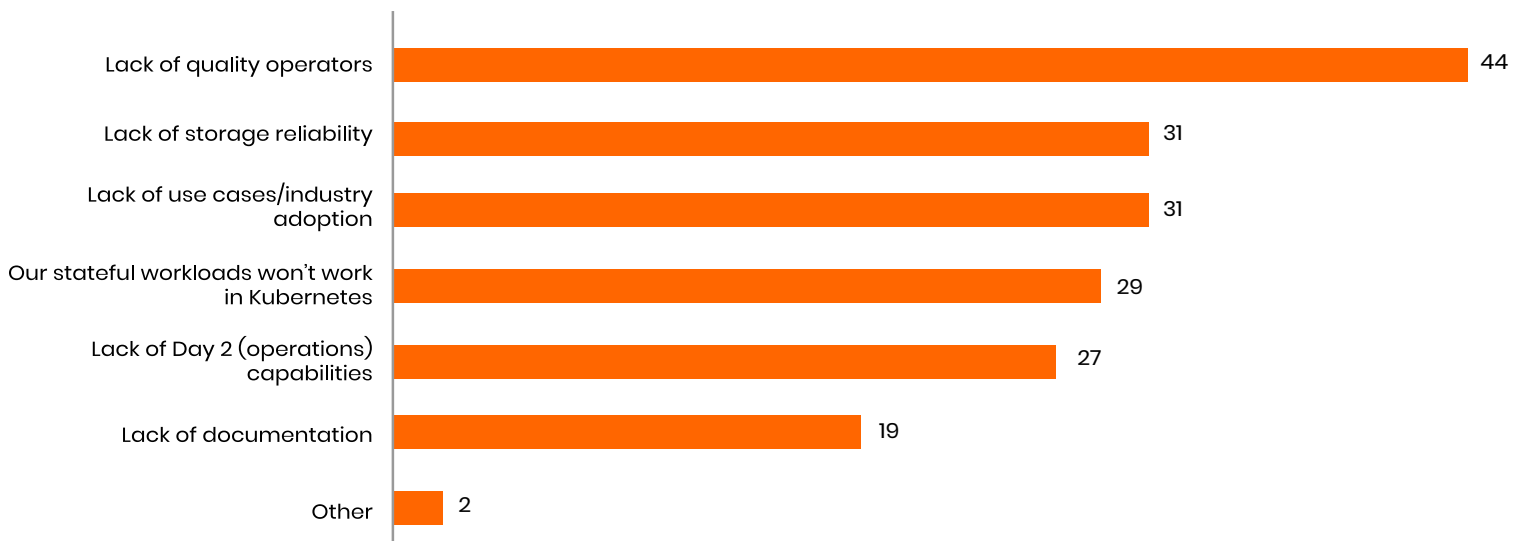


### DEVELOPING OWN OPERATORS



Further underscoring this point, lack of quality operators is cited as the number one reason preventing some from using Kubernetes for stateful workloads – the 10% “non-believers” who do not think Kubernetes is production-ready for running data.

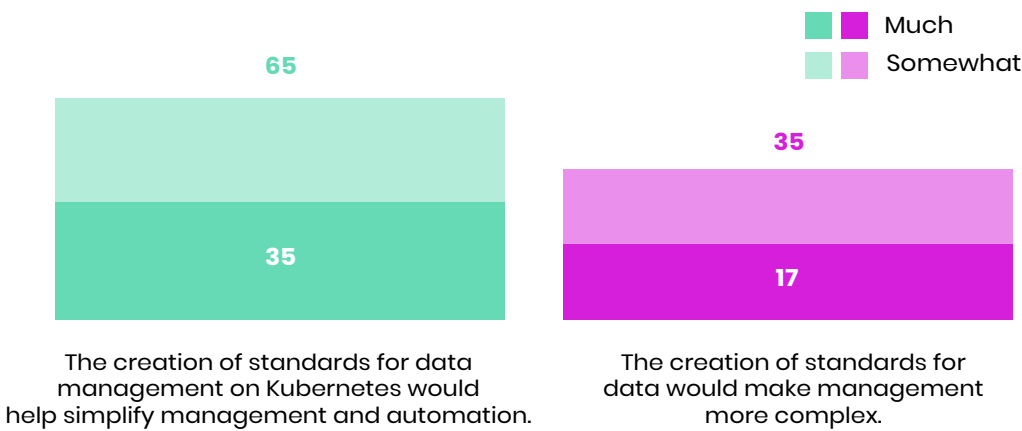
### Why do you not think Kubernetes is ready to run your organization's stateful workloads in production? (Select all that apply)



# The Future of DoK

Despite the challenges, respondents believe that running stateful workloads on Kubernetes is the way forward as evidenced by the productivity and satisfaction gains we see in the first section. When asked to envision a future that simplifies management and automation on Kubernetes, a 2:1 majority agree that the standardization of data management is important.

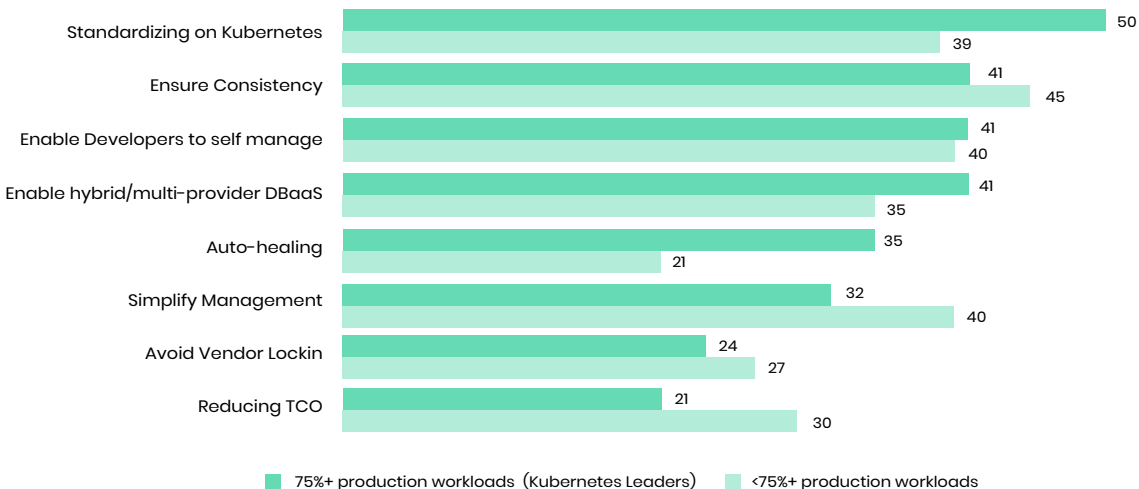
## 2:1 believe that standardization of data management is important



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.

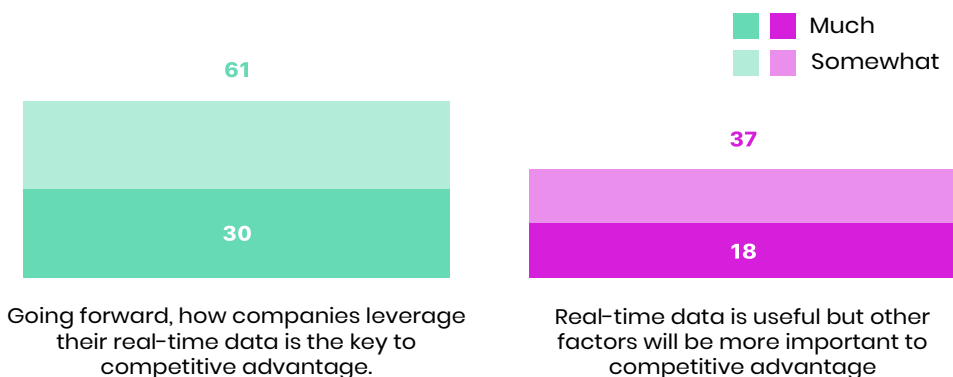
Standardization is a recurring theme for our respondents. When done well, it drives Kubernetes adoption; when absent, it slows it. This is even more pronounced for Kubernetes Leaders running 75%+ workloads in production – standardization jumps 11% to the top spot (up from number two for all respondents), followed by consistency and the ability for developers to self-manage.

## From the following, which are the THREE most important factors in your organization’s decision to run stateful workload(s) on Kubernetes?



A 2:1 majority also believe that how companies leverage their real-time data is key to competitive advantage. The rise of real-time is fueled by organizations' desire to quickly react to actionable insights that drive customer satisfaction and revenue. The ability for companies to standardize all workloads on Kubernetes puts it in a position to become the favored system for real-time workloads.

## Real-time data will increase the demand for running data on Kubernetes



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.

Kubernetes Leaders are also more interested in running AI/ML workloads than all respondents, jumping to the #2 position behind databases (up from #6) – an indicator of what the future of DoK may look like for all.

## Conclusion

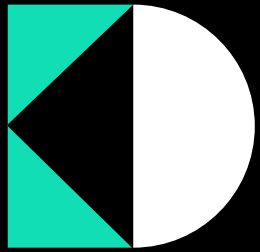
Stateful workloads are pervasive, and the most advanced Kubernetes users benefit from massive productivity gains thanks to the standardization of how they run stateless and stateful workloads. The operator pattern is beneficial, but not without challenges which is forcing organizations to build their own. Standards and/or best practices are needed to bring operators to a level of quality that will allow organizations to realize the benefits of running stateful workloads in a consistent way.

Signs point to a future where organizations can standardize, or further standardize, on Kubernetes for data-intensive workloads. This may be driven by industry standards and exemplified by declarative data and similar concepts. It will undoubtedly encompass the world of data technologies (persistence, streaming, analytics), data infrastructure (storage, security, networking), and data governance (policies, protocols, access) and require contributions from everyone; achieved with open communities, open standards, and open source. The future is ours to build.

## About Data on Kubernetes Community

Kubernetes was initially designed to run stateless workloads. Today it is increasingly being used to run databases and other stateful workloads. The Data on Kubernetes Community was founded in June 2020 to bring practitioners together to solve the challenges of working with data on Kubernetes. An openly governed community, DoKC exists to assist in the emergence and development of techniques for the use of Kubernetes for data. <https://dok.community/>





**DoK**  
Community