DoK
Community

RESEARCH REPORT

Data on Kubernetes 2022

Insights from over 500 executives and technology leaders on how data on Kubernetes has a transformative impact on organizations, regardless of size or tech maturity

Overview and Key Findings

Data applications empower organizations to create new business models and transform industries, and Kubernetes is increasing the velocity with which they can be deployed.

The appetite for running data on Kubernetes (DoK) is strong. That's what [our first report](#) found, with 90% of respondents believing that Kubernetes is ready for stateful workloads and a large majority of respondents (70%) running them in production.

This year's report found that running data on Kubernetes benefits the whole organization, not just engineering. Leaders believe DoK is transformative for business growth.

In the spring of 2022, the [Data on Kubernetes Community](#) engaged the research firm Clearpath Strategies to survey over 500 Kubernetes users to understand the types and volume of data workloads deployed in Kubernetes, benefits, and challenges, and the factors driving further adoption.

Key Findings

- Data on Kubernetes has a transformative impact on organizations. Respondents see a direct link from running DoK and making big gains: A majority (83%) attribute over 10% of their revenue to running data on Kubernetes. One-third of organizations saw their productivity increase twofold.
- DoK benefits are reported by all organizations, no matter their tech maturity. While Leading organizations tend to get the lion's share, the gains are felt by Lagging and Leaders alike.
- Respondents experience high satisfaction running DoK, however, Day-2 operations remain a significant challenge, with respondents relying heavily on Kubernetes operators. When evaluating an operator for adoption, security is the number one concern. The value in overcoming these challenges is high, as they recognize the transformative impact that standardization can have for their organizations.
- Organization's increased adoption of DoK requires more talent. Lack of qualified talent ranked second on the list of top challenges, and 41% of respondents plan to upskill or hire DoK talent within the next three to six months.
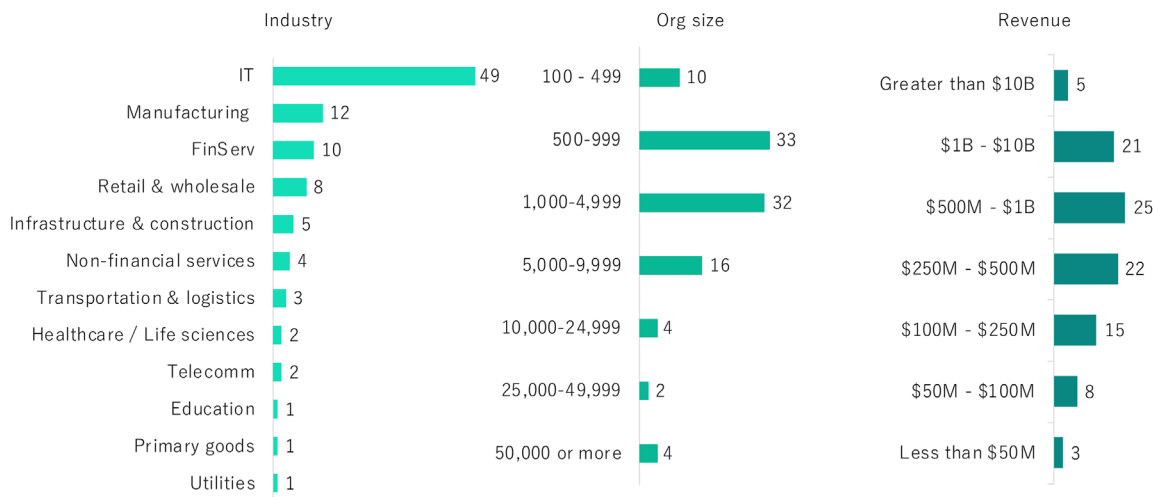
Demographics

For this report, the research firm only surveyed individuals whose organizations are currently using data on Kubernetes. This included an international audience of 501 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 (~42%), executives (~37%), and managers (~20%).
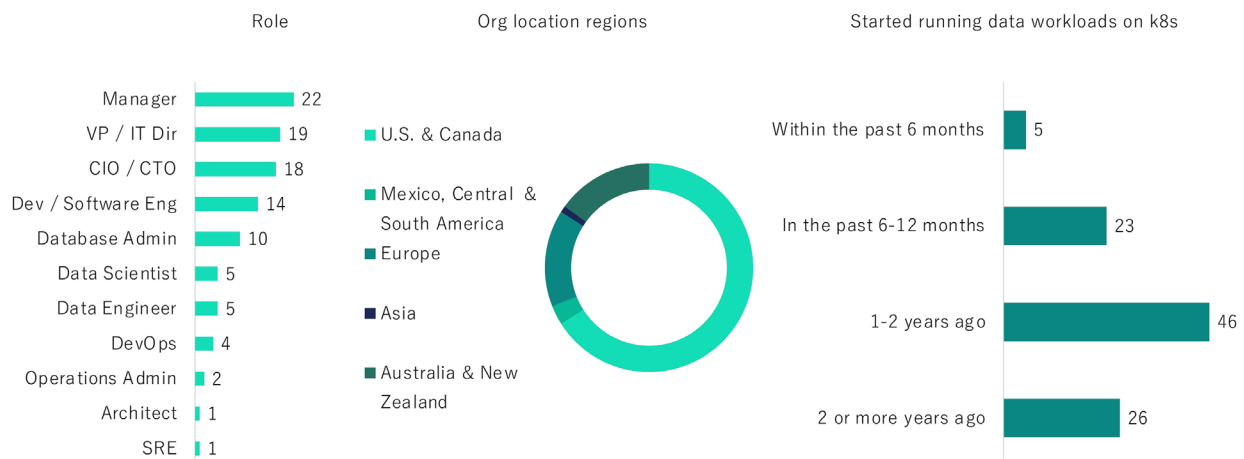
The majority of respondents (49%) came from Technology organizations (software, hardware, services), followed by Manufacturing (12%), FinTech (10%), and Heavy Industry (8%).

Respondents were from companies that started running data on Kubernetes 2 or mere years ago (25%), within 1 to 2 years ago (46%), in the past 6–12 (months), and within the past 6 months, providing a wide view of the current data on Kubernetes landscape.

Industry
- IT: 49
- Manufacturing: 12
- FinServ: 10
- Retail & wholesale: 8
- Infrastructure & construction: 5
- Non-financial services: 4
- Transportation & logistics: 3
- Healthcare / Life sciences: 2
- Telecomm: 2
- Education: 1
- Primary goods: 1
- Utilities: 1

Org size
- 100 - 499: 10
- 500-999: 33
- 1,000-4,999: 32
- 5,000-9,999: 16
- 10,000-24,999: 4
- 25,000-49,999: 2
- 50,000 or more: 4

Revenue
- Greater than $10B: 5
- $1B - $10B: 21
- $500M - $1B: 25
- $250M - $500M: 22
- $100M - $250M: 15
- $50M - $100M: 8
- Less than $50M: 3

(Left-hand side) Q. Which of the following industries best describes your company?
(Middle) Q. Approximately how many employees work at your organization? (Your best guess is fine.)
(Right-hand side) Q. What is your organization's annual revenue?

5

Role
- Manager: 22
- VP / IT Dir: 19
- CIO / CTO: 18
- Dev / Software Eng: 14
- Database Admin: 10
- Data Scientist: 5
- Data Engineer: 5
- DevOps: 4
- Operations Admin: 2
- Architect: 1
- SRE: 1

Org location regions
- U.S. & Canada
- Mexico, Central & South America
- Europe
- Asia
- Australia & New Zealand

Started running data workloads on k8s
- Within the past 6 months: 5
- In the past 6-12 months: 23
- 1-2 years ago: 46
- 2 or more years ago: 26

(Left-hand side) Q. Which ONE of the following comes closest to describing your primary responsibility, even if you perform more than one or have a different title?
(Middle) Q. In what region(s) is your organization located?
(Right-hand side) Q. [IF "Currently running data workload(s)" (1) IN DATAWKLD] When did your organization start to run data workloads on Kubernetes?

6

Data-Driven Enterprise Scale

We leveraged Clearpath's methodology to define the technical maturity of a company. The data-driven enterprise (DDE) scale is a validated tool that allows us to measure the tech maturity of organizations, primarily based on their use of and ability to operationalize and derive value from data.

Based on the DDE scale, our respondents were categorized into 4 main categories:

Leading users: represent 13% of the respondents

Advancing users: represent 37% of the respondents

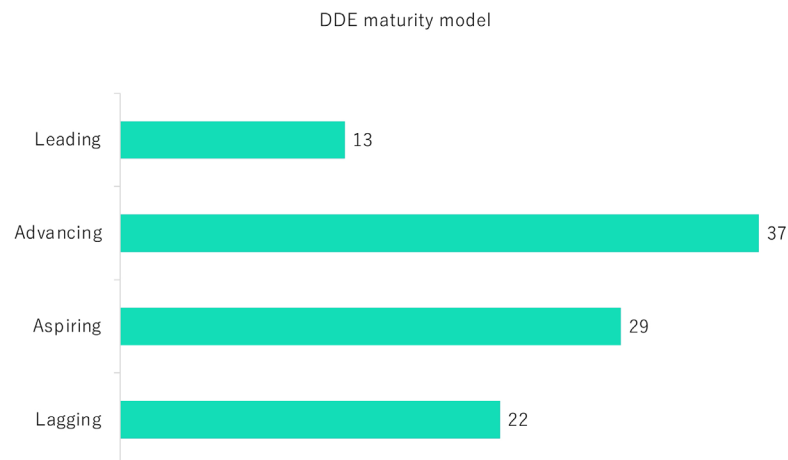**Aspiring users:** represent 29% of the respondents

Lagging users: represent 22% of the respondents

Respondents lean more mature by design

Higher scores predict greater tech maturity

DDE scale

Our data is highly scalable	48
Our data strategy is discussed at the Board level	46
We use data to improve the reliability and scalability of our IT/cloud infrastructure	46
We have clearly defined data product owners responsible for turning data into products	46
Someone has the title or responsibility of Chief Data Officer	46
We have high data quality	45
We share real time insights across business units	45
Our apps are highly scalable	44
We are increasing our use of open source software	43
Accountability for creating value from data is at the business unit level	42
We use data effectively to create net new revenue streams	42
Our data infrastructure is aligned with our data strategy	41
We have cross-functional teams responsible for deploying data products	40
We have a unified, modular data platform	39

Population skews toward "Leading" org tech maturity

DDE maturity model

Leading	13
Advancing	37
Aspiring	29
Lagging	22

The Transformative Impact of DoK

One of the highlights of our findings is that the majority of respondents (83%) attribute over 10% of their revenue to running data on Kubernetes. These data show that the benefits of DoK go beyond engineering to benefit the whole business, with a small cohort (9%) attributing more than 20% of revenue to DoK.

Most attribute over 10% of their revenue from running DoK

Revenue from running DoK

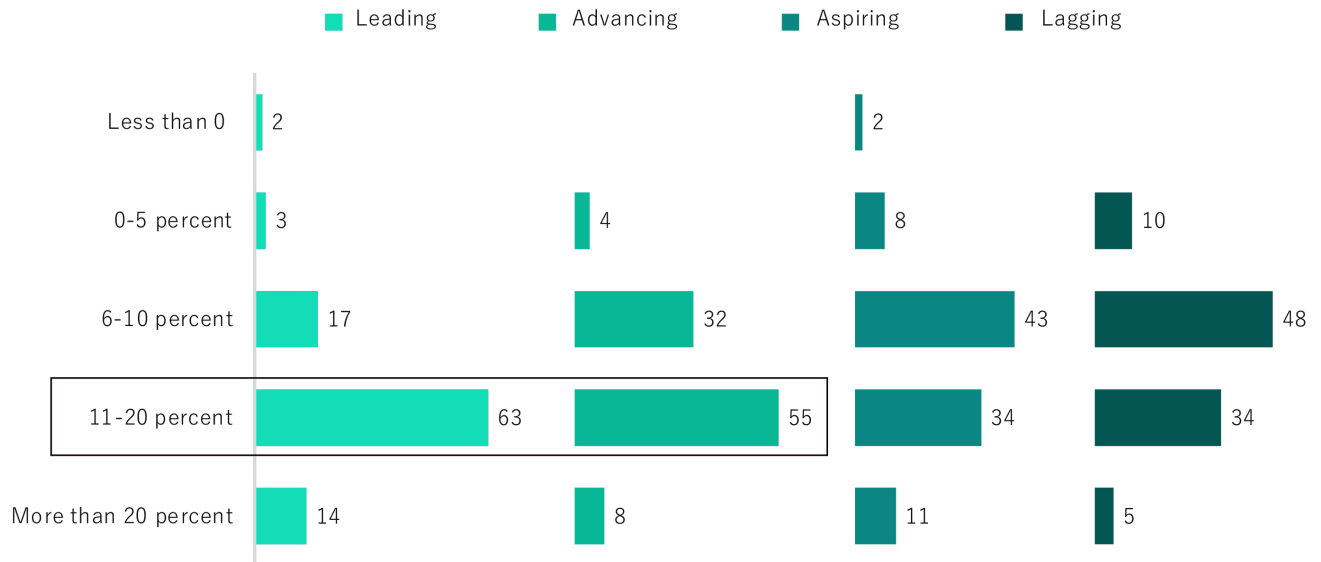|	
Less than 0	1
0-5 percent	7
6-10 percent	37
11-20 percent	45
More than 20 percent	9
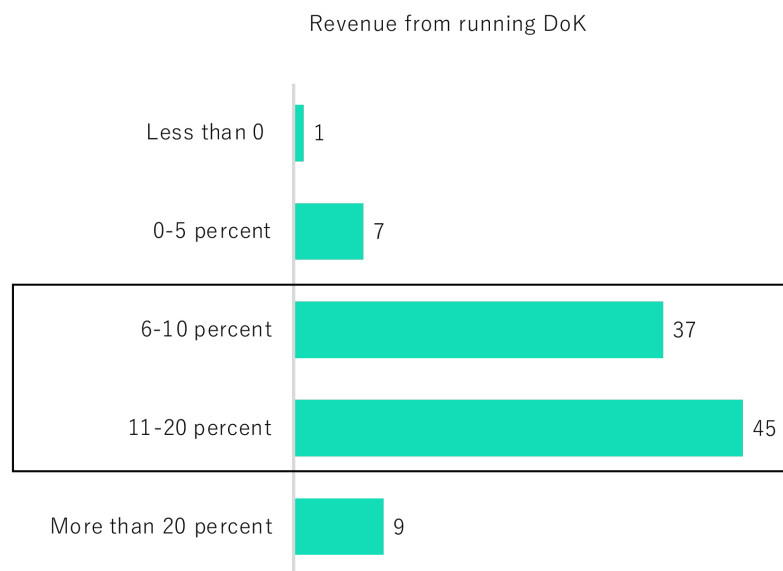
Q. Approximately what percent of revenue does your company attribute to your ability to run data workloads on Kubernetes? Your best guess is fine.

11

While Leading organizations see the most gains, all organizations benefit from DoK regardless of tech maturity. Most Leading (63%) and Advancing (55%) organizations attribute 11% or more of their revenue gains to running data on Kubernetes. While 43% of Aspiring and 48% of Lagging (48%) organizations attribute 6% and more. This probably explains why, as we will discuss later in the report, organizations are planning on moving more stateful workloads to Kubernetes.
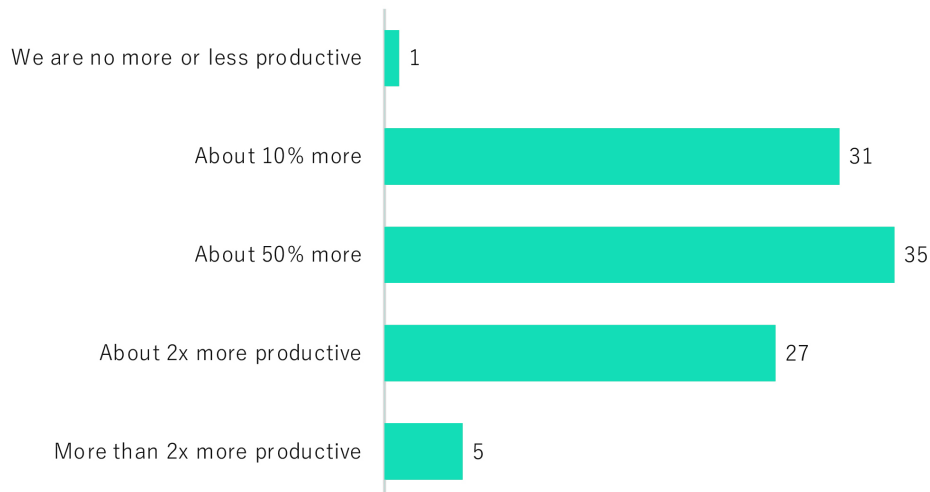
And while leaders still benefit most, revenue gains are felt by all

Leading ■ Advancing ■ Aspiring ■ Lagging

	Leading	Advancing	Aspiring	Lagging
Less than 0	2		2	
0-5 percent	3	4	8	10
6-10 percent	17	32	43	48
11-20 percent	63	55	34	34
More than 20 percent	14	8	11	5

As we saw in last year's report, productivity gains remain a key benefit: one-third of respondents saw their productivity increase 2x or more – and this is true across organizations at any maturity level. Most Leading (30%) and Advancing (39%) organizations see their productivity increase 2x or more. While 39% of Aspiring and 35% of Lagging organizations experience 50% or more productivity gain.
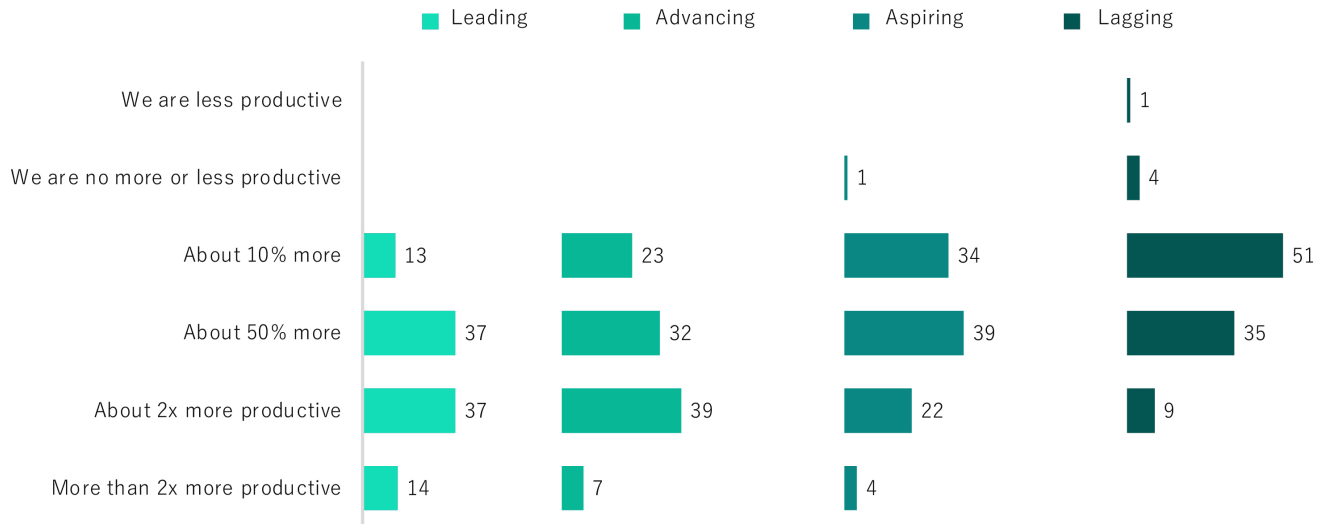
A third say running DoK workloads has at least doubled their productivity

Productivity after k8s

We are no more or less productive	1
About 10% more	31
About 50% more	35
About 2x more productive	27
More than 2x more productive	5

And these productivity gains aren't limited by tech maturity
Though, the more advanced you are, the greater the gains

■ Leading   ■ Advancing   ■ Aspiring   ■ Lagging

	Leading	Advancing	Aspiring	Lagging
We are less productive				1
We are no more or less productive			1	4
About 10% more	13	23	34	51
About 50% more	37	32	39	35
About 2x more productive	37	39	22	9
More than 2x more productive	14	7	4	

Q. How much more productive is your organization/are your developers after adopting Kubernetes to manage data workloads, if at all?
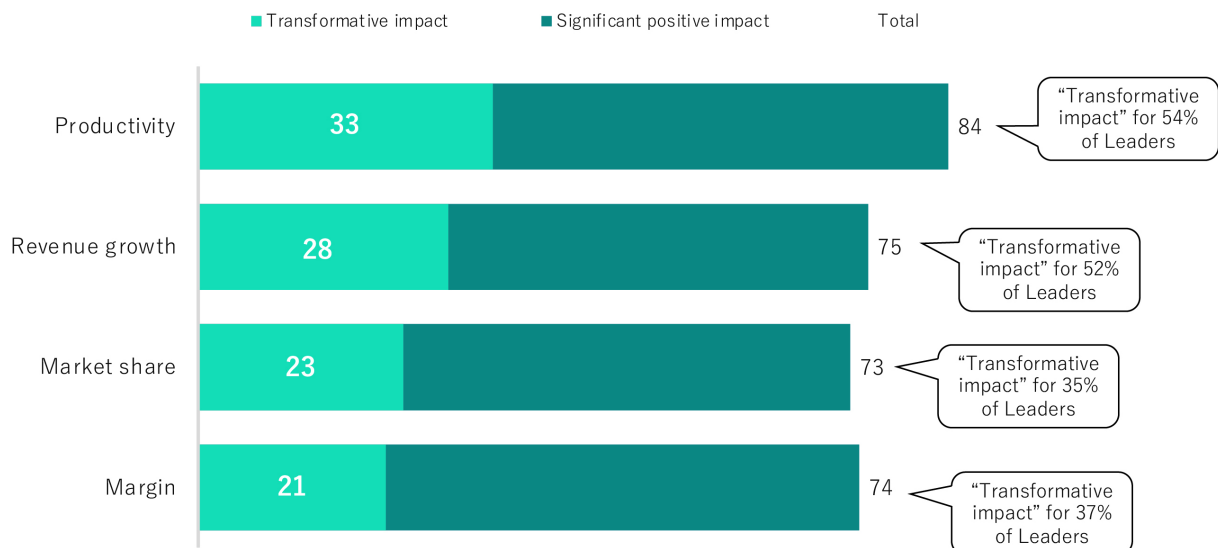
14

Organizations see data on Kubernetes having a transformative impact on other key business metrics including revenue growth, market share, and margin. As expected, leading organizations experience the greatest impact. The more technically mature organizations are, the more transformative impact they can expect from running data on Kubernetes.

These DoK impacts are "transformative"
Leading companies especially are winning BIG with DoK

Impact of data workloads on Kubernetes on…

■ Transformative impact   ■ Significant positive impact   Total

|	Transformative impact	Total	Leaders note
Productivity	33	84	"Transformative impact" for 54% of Leaders
Revenue growth	28	75	"Transformative impact" for 52% of Leaders
Market share	23	73	"Transformative impact" for 35% of Leaders
Margin	21	74	"Transformative impact" for 37% of Leaders

The DoK Journey

DoK's benefits are crystal clear, but how do organizations get there?

The data show that most organizations begin their DoK journey by moving Databases and Analytics workloads to Kubernetes, then continue to move more complex workloads such as AI/ML and streaming/messaging. In fact, Analytics jumped up to the #2 spot this year compared with #6 last year; AI/ML to #3 from #7; and Persistent Storage dropped from #2 to #4.

While a ton of people are storing their data in K8s, far fewer are manipulating it for a purpose
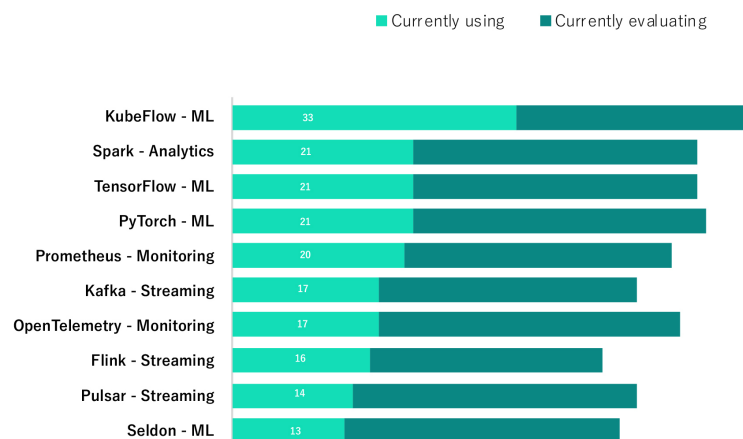
Which data workloads on k8s
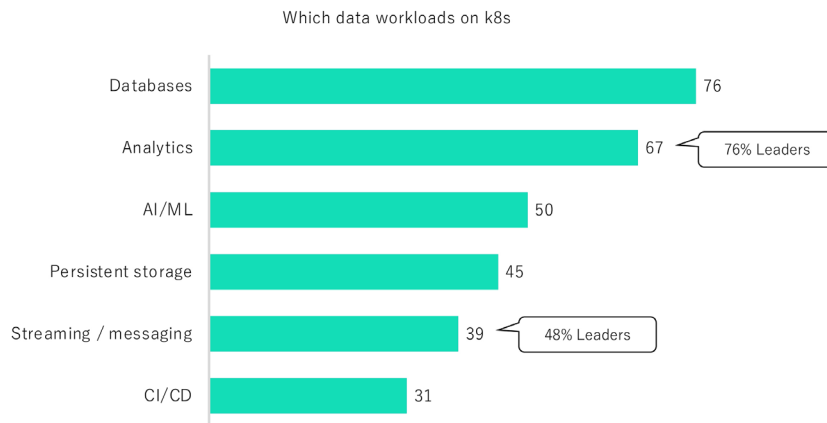
Q. Which of the following data workloads does your organization run on Kubernetes? Select all that apply.

60

The fact that AI/ML workloads are increasingly being run on Kubernetes is reflected in the tools that practitioners use to fulfill their DoK needs. KubeFlow, TensorFlow, PyTorch, and MLRun are all ML tools.

So, they turn to operators. Everyone is relying on them to run their DoK workloads

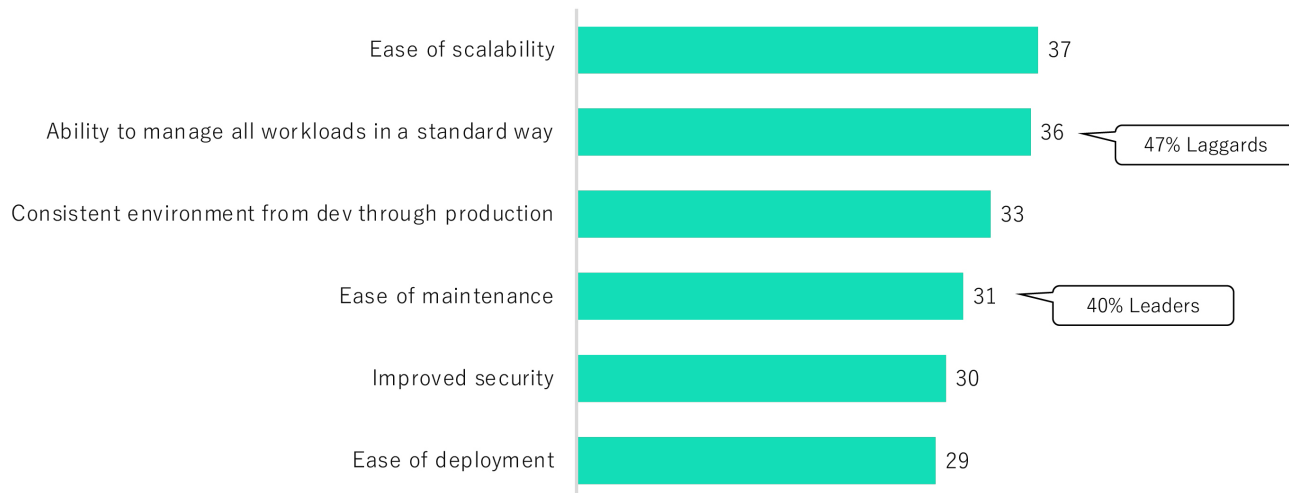■ Currently using   ■ Currently evaluating

Running data workloads on Kubernetes brings a number of benefits including ease of scalability (37%), the ability to standardize the way all workloads are managed (36%), bringing consistency from dev throughout the production environments (33%). Organizations want to standardize the way they run any type of workload and Kubernetes is the tool of choice for that.

While Laggards appreciate standardization (a key driver for all in last year's survey), Leaders experience even greater improvements to ease of maintenance.

And ultimately, while most believe running DoK workloads leads to manifold benefits…

Benefits of running data workloads on k8s

Benefit	Value	
Ease of scalability	37	
Ability to manage all workloads in a standard way	36	47% Laggards
Consistent environment from dev through production	33	
Ease of maintenance	31	40% Leaders
Improved security	30	
Ease of deployment	29	

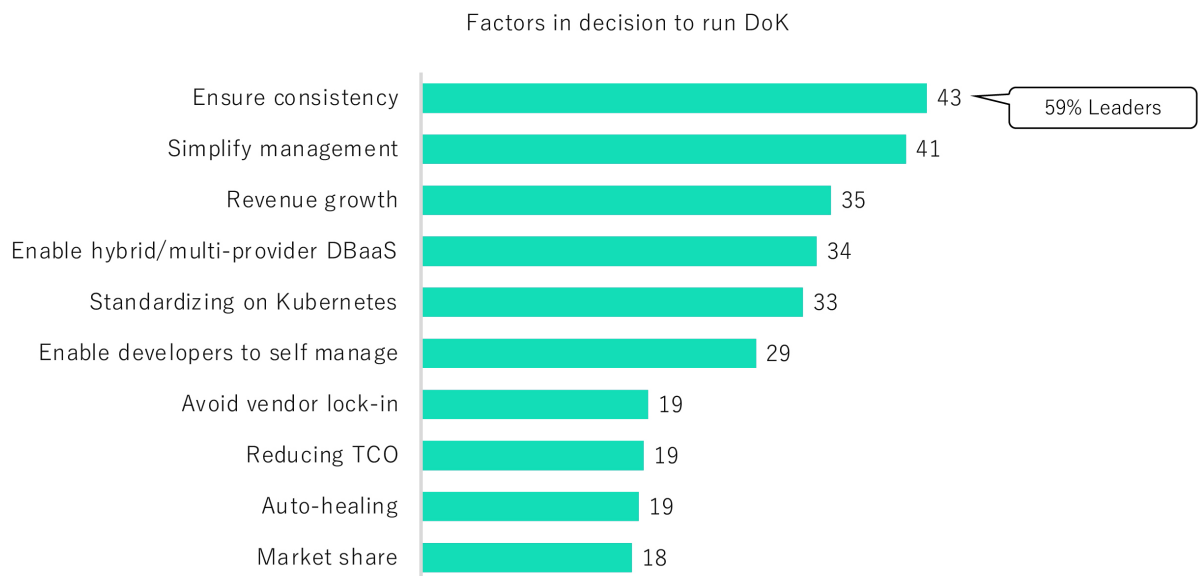Q. What in your opinion are the primary benefits of running data on Kubernetes? (Select up to TWO.)

35

Bridging the Gap

Organizations run data on Kubernetes to streamline their IT operations by ensuring consistency (43%) and simplifying management (41%). The impact isn't limited to engineering: 35% of respondents chose to run DoK to increase revenue growth – and successfully as previously discussed.

And here's why
Driven mostly by need for consistency and simplicity

Factors in decision to run DoK

Factor	Value
Ensure consistency	43 — 59% Leaders
Simplify management	41
Revenue growth	35
Enable hybrid/multi-provider DBaaS	34
Standardizing on Kubernetes	33
Enable developers to self manage	29
Avoid vendor lock-in	19
Reducing TCO	19
Auto-healing	19
Market share	18

But achieving those goals isn't without challenges. While the DoK toolset and ecosystem continues to grow, there's more work for us all to do. For the second year in a row, organizations cite a lack of integration with their existing tools as the number one challenge of running data workloads on Kubernetes.

Second is a lack of qualified talent – up from fourth last year – which likely correlates with the third challenge: too much time/effort to manage.

This is a common refrain for Kubernetes users, not just those running DoK: the more usage and benefits, the more talent required to grow, scale, and manage – and organizations are hiring aggressively, as we'll see later in the report.
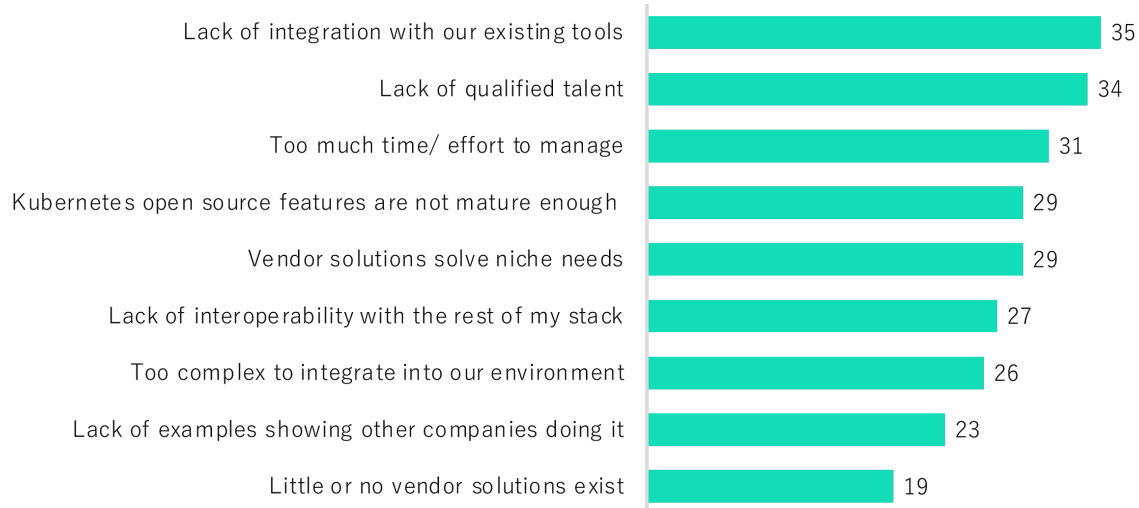
Interesting to note was that *Little or no vendor solutions exist* ranked second in last year's report, moving to the very bottom of this year's list. Perhaps this is an indicator that vendors have been equipping their customers with the right tools for their DoK needs.

DoKC staff were pleased to see the *Lack of examples showing other companies doing it* toward the bottom of the list. We hope we had at least a little something to do with that. 😎

But the promise is unmet
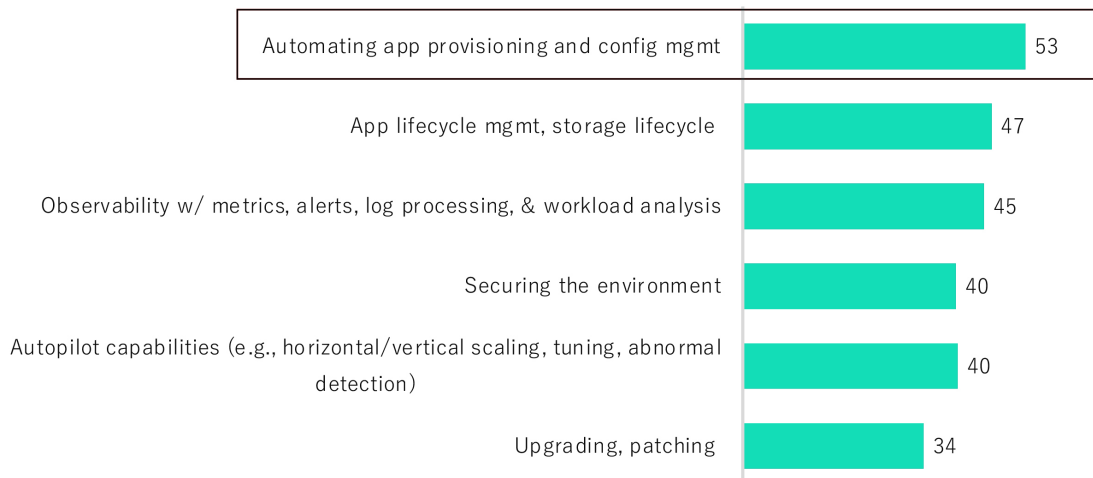## Running DoK workloads is not all rainbows and sunshine

Challenges of running data workloads on k8s

Challenge	Value
Lack of integration with our existing tools	35
Lack of qualified talent	34
Too much time/ effort to manage	31
Kubernetes open source features are not mature enough	29
Vendor solutions solve niche needs	29
Lack of interoperability with the rest of my stack	27
Too complex to integrate into our environment	26
Lack of examples showing other companies doing it	23
Little or no vendor solutions exist	19

Getting data workloads to run on Kubernetes is the first step. Next comes the arduous task of managing them throughout their lifetime (Day-2 operations). The number one challenge organizations cite is automating application provisioning and management.

This is inherently a structural challenge, as they look to automate provisioning and config management

Challenges of managing data workloads on k8s

Challenge	Value
Automating app provisioning and config mgmt	53
App lifecycle mgmt, storage lifecycle	47
Observability w/ metrics, alerts, log processing, & workload analysis	45
Securing the environment	40
Autopilot capabilities (e.g., horizontal/vertical scaling, tuning, abnormal detection)	40
Upgrading, patching	34

To handle Day-2 operations, organizations rely heavily on Kubernetes operators – 81% are currently using (66%) or evaluating (15%) operators to run DoK. The operator pattern uses the Kubernetes API to create, configure, and manage instances of complex stateful applications on behalf of a Kubernetes user.
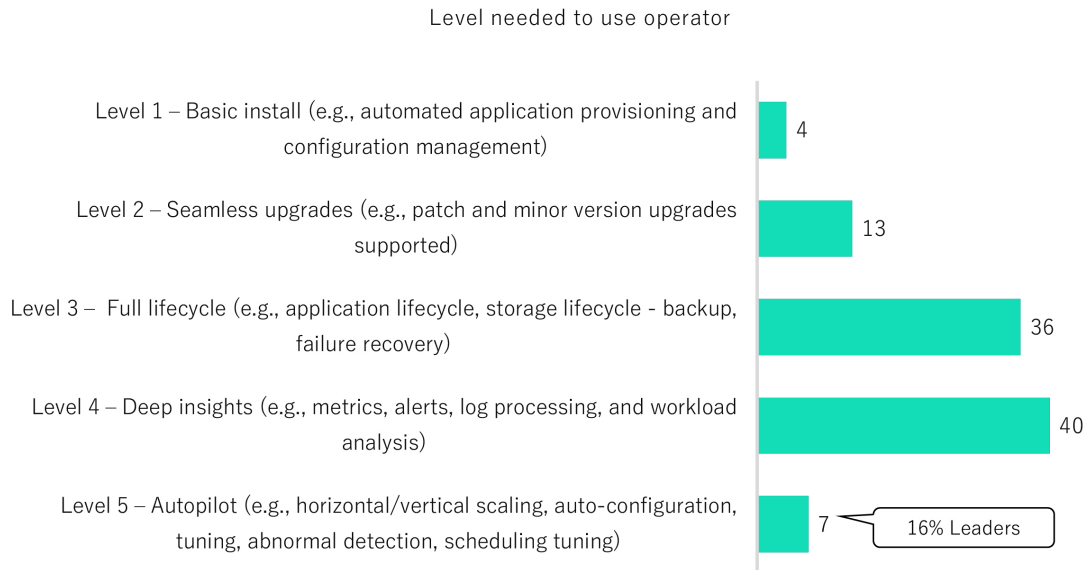
Kubernetes operators are a cornerstone of DoK. Most organizations (66%) use at least 20. The number of operators continues to grow, with the Operator Hub listing 278 of them.

Number of operators

Range	Value
1 to 10	10
11 to 20	36
21 to 30	30
31 to 40	16
41 to 50	4
51+	1
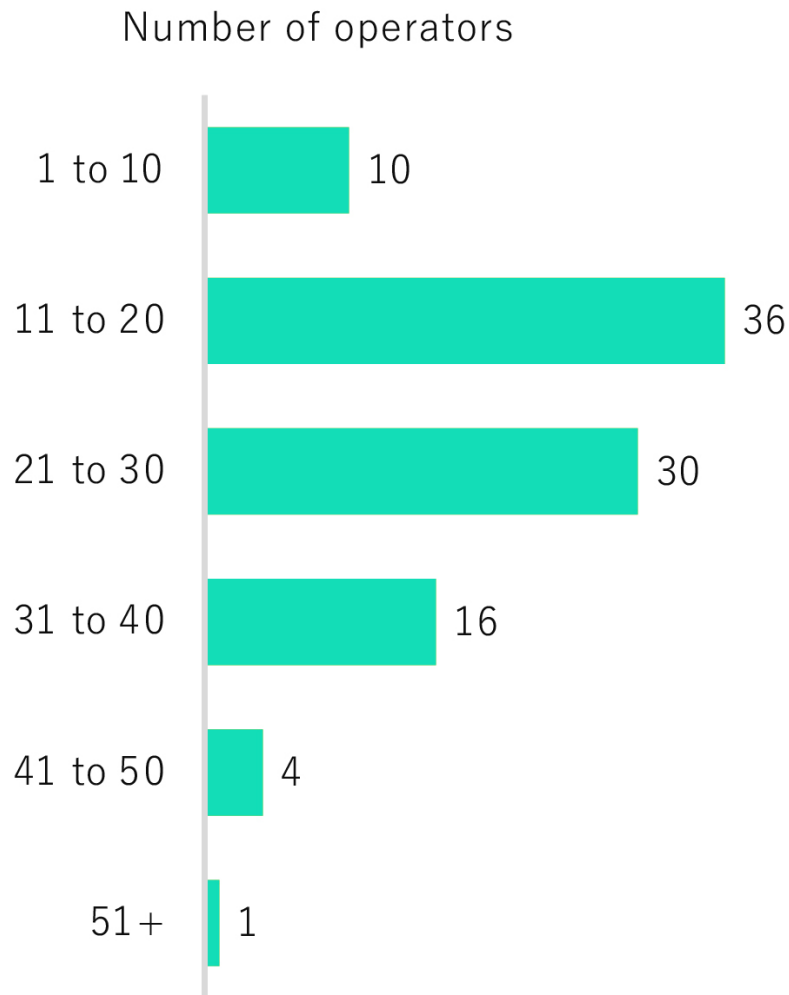
Organizations expect operators to perform at a very high level. Using the Operator Framework as a grading scale, we found that 36% of organizations expect operators to properly handle all Day-2 operations such as application and storage lifecycle, backup, and recovery failure. More than that, 40% expect advanced capabilities such as the ability to provide metrics, alerts, log processing, and workload analysis. Leaders require the highest level of automation from their operators, including horizontal/vertical scaling, auto-configuration, tuning, abnormal detection, and scheduling.
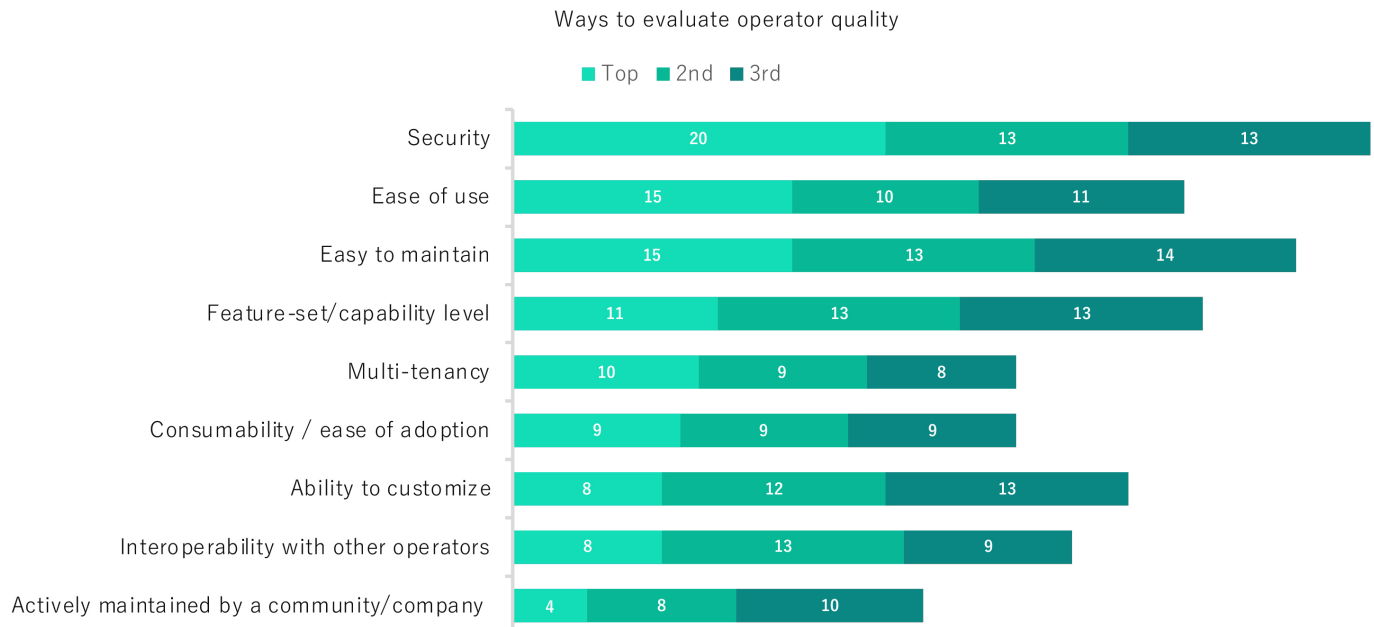
They expect operators to perform at a very high level
The bar is set at lifecycle automation, with a stretch goal of providing insights

Level needed to use operator

Level 1 – Basic install (e.g., automated application provisioning and configuration management)	4
Level 2 – Seamless upgrades (e.g., patch and minor version upgrades supported)	13
Level 3 – Full lifecycle (e.g., application lifecycle, storage lifecycle - backup, failure recovery)	36
Level 4 – Deep insights (e.g., metrics, alerts, log processing, and workload analysis)	40
Level 5 – Autopilot (e.g., horizontal/vertical scaling, auto-configuration, tuning, abnormal detection, scheduling tuning)	7 16% Leaders

When organizations evaluate operators to adopt, security is the most important criterion. A possible explanation is that data is the most critical asset to protect, and Kubernetes isn't easy to secure. Recently, a cybersecurity firm found 900,000 Kubernetes clusters exposed to the Internet. The challenge is so steep and the risk so high that the US government got involved with the NSA and CISA releasing a Kubernetes hardening guide.

And they prioritize security and ease of use

Ways to evaluate operator quality

■ Top ■ 2nd ■ 3rd

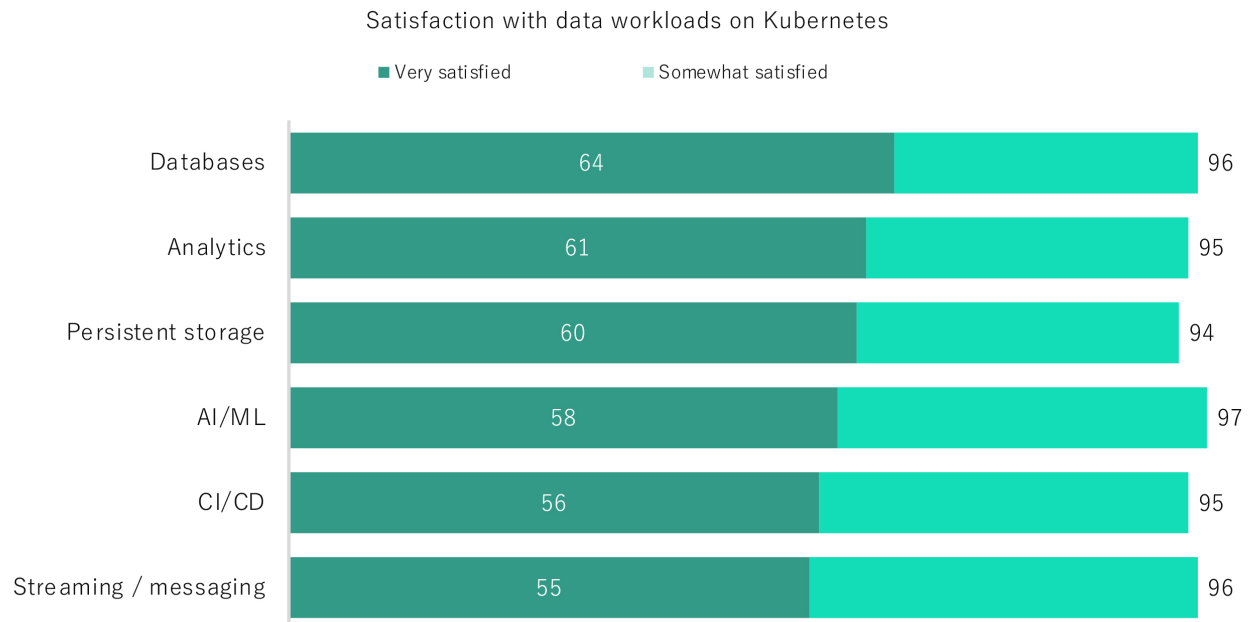|	Top	2nd	3rd
Security	20	13	13
Ease of use	15	10	11
Easy to maintain	15	13	14
Feature-set/capability level	11	13	13
Multi-tenancy	10	9	8
Consumability / ease of adoption	9	9	9
Ability to customize	8	12	13
Interoperability with other operators	8	13	9
Actively maintained by a community/company	4	8	10

Highly Satisfied Users are Standardizing on DoK

Organizations are very satisfied with running data on Kubernetes. The chart below mostly mirrors overall DoK usage, with satisfaction at its highest for the most widely adopted workloads: Databases (64%) and Analytics (61%). Emerging use cases such as CI/CD and streaming/messaging have room to grow; however, the 48% of Leaders running streaming/ messaging workloads also show a higher percentage of satisfaction at 55%.

Satisfaction = Reality - Expectations
DoK is winning the expectations battle

Satisfaction with data workloads on Kubernetes

■ Very satisfied   ■ Somewhat satisfied

Workload	Very satisfied	Total
Databases	64	96
Analytics	61	95
Persistent storage	60	94
AI/ML	58	97
CI/CD	56	95
Streaming / messaging	55	96

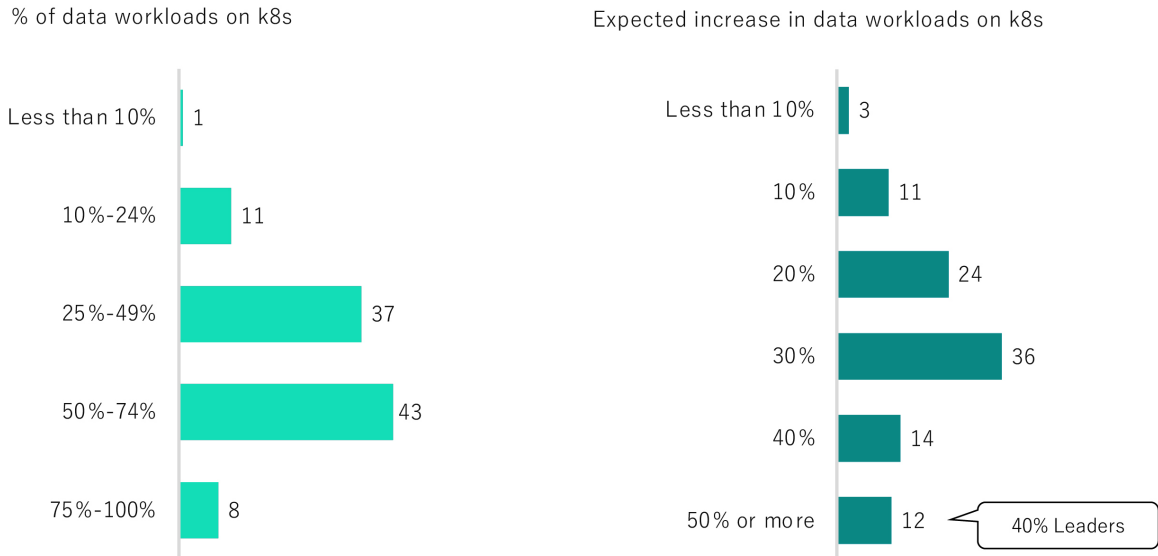Q. In general, how satisfied are you with using Kubernetes to run each of the following data workloads in your organization? Use a scale from 1 to 5 where 5 means "very satisfied" and 1 means "not at all

High rates of satisfaction lead to higher usage. Over half have 50% or more of their data workloads running on Kubernetes, with even more expecting to increase that by 30% in the next 12 months. Leaders are all-in: almost half plan to increase their DoK footprint by 50% or more in the next 12 months.

DoK workload %s are already high, and expected to increase

Leaders are chomping at the DoK bit

% of data workloads on k8s

Less than 10%	1
10%-24%	11
25%-49%	37
50%-74%	43
75%-100%	8

Expected increase in data workloads on k8s

Less than 10%	3
10%	11
20%	24
30%	36
40%	14
50% or more	12

40% Leaders

(Left-hand side) Q. 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?
(Right-hand side) Q. By how much do you expect your organization's percentage of data workloads run on Kubernetes to increase in the next 12 months?

19

A large volume of DoK workloads and high satisfaction is compelling organizations to upskill or hire talent with DoK expertise to meet growth. This is reflected in the [Dice Tech Job Report](#), which shows Kubernetes in job postings increasing by 99% year-over-year.

But the future of DoK is bright

Especially for Leading orgs, who plan to move quickly on DoK training/hiring

Planning to reskill/hire people

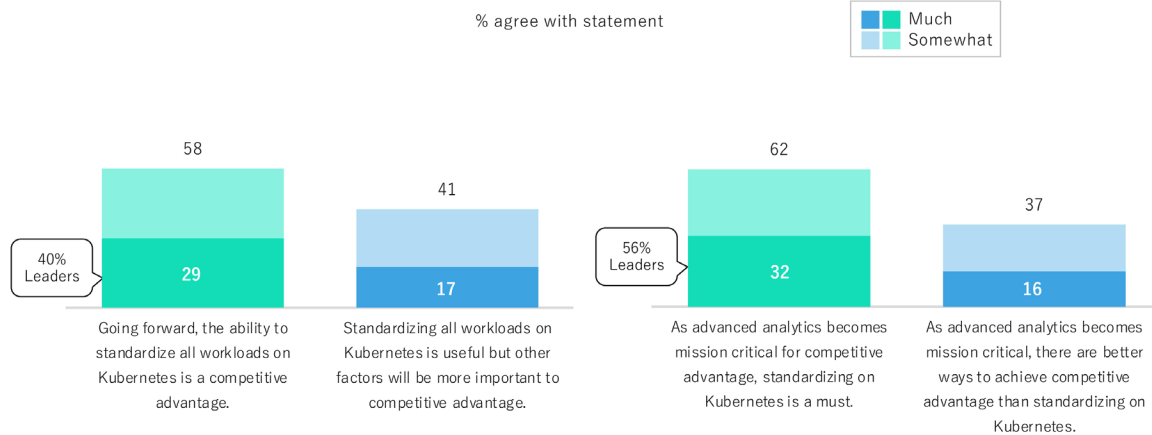|	
Yes, in the next 0-3 months	29
Yes, in the next 3-6 months	41
Yes, in the next 6-12 months	10
Yes, but our timing is uncertain	12
No	7

54% Leaders

Q. [IF ROLE = CIO / CTO / Business Leadership (16)] Are you/is your organization planning to reskill staff and/or hire people who have experience running data workloads on Kubernetes?

Organizations do not doubt that doing DoK is the right path forward, but they require the right tools, integrations, and workforce to do so.

Everyone is sold benefits of standardizing data workloads on k8s
Leaders need no convincing

% agree with statement

Much
Somewhat

58

40%
Leaders

29

Going forward, the ability to standardize all workloads on Kubernetes is a competitive advantage.

41

17

Standardizing all workloads on Kubernetes is useful but other factors will be more important to competitive advantage.

62

56%
Leaders

32

As advanced analytics becomes mission critical for competitive advantage, standardizing on Kubernetes is a must.

37

16

As advanced analytics becomes mission critical, there are better ways to achieve competitive advantage than standardizing on Kubernetes.

Q. 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.

39

Conclusion

Kubernetes is on its way to becoming the industry standard for managing cloud native data applications. Leaders understand the transformative impact DoK can have on the business by increasing revenue, market share, and margin. Those benefits are made possible by organizations standardizing and streamlining how they run workloads across environments, strongly boosting their productivity.

Running stateful workloads on Kubernetes isn't without its challenges. Operators are heavily relied upon to bridge the Day-2 operations gap that Kubernetes does not fulfill. And because data is an organization's most precious asset, organizations want to ensure it is secure.

While technical challenges remain, the lack of qualified talent will become a bottleneck unless organizations can fulfill their goal of upskilling and hiring. As organizations continue their adoption of DoK and of running more and complex workloads, the need will only become greater. And yet, the value in overcoming these challenges is high as respondents recognize the transformative impact DoK and Kubernetes standardization can have for their organizations.

About the Data on Kubernetes Community

The Data on Kubernetes Community was founded in June 2020 to bring together practitioners 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 using Kubernetes for data. https://dok.community/