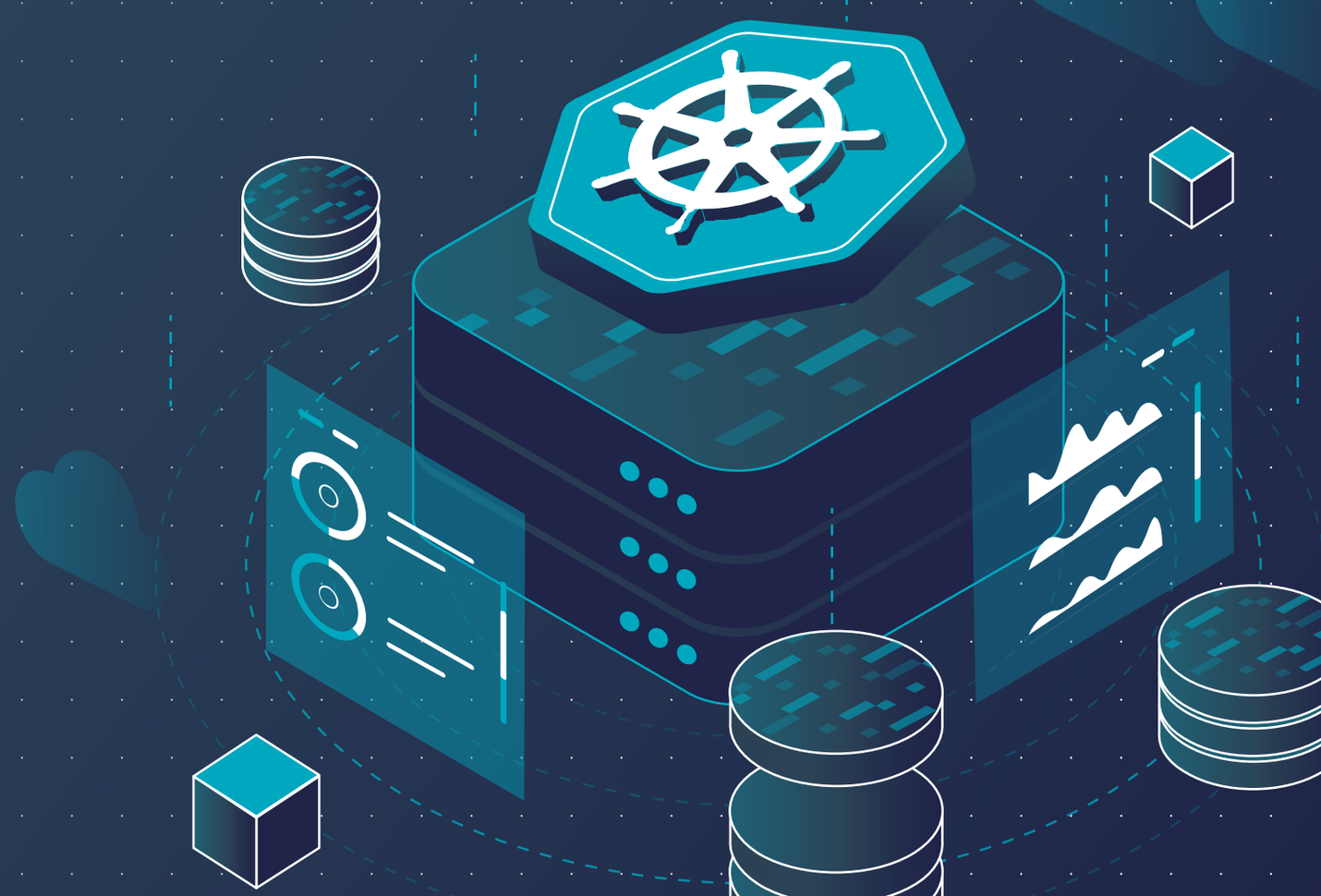
pepperdata

REPORT

2022 Big Data on Kubernetes Report

Exploring the Benefits and Challenges of Running Big Data Applications on Kubernetes



Background

This survey was conducted to identify which big data applications companies are migrating/ planning to migrate to Kubernetes (K8s) and what challenges they are having.

Cloud vendors have proliferated and promised users optimal performance and tight spend. Still, many of these vendors don't provide visibility into Kubernetes big data, resulting in performance issues, poor resource allocation, overspending, and ineffective tuning.

While observability is crucial to maximizing the performance of big data applications running on Kubernetes, more than half (54%) of our polled users don't know which metrics to focus on. This survey was conducted using Pollfish in October of 2021, among 600 participants across a range of industries. A majority (62%) worked at companies with between 500 and 5,000 employees.

Key Findings

1. More than half (54%) of the respondents are moving big data applications to Kubernetes to reduce their overall spend.

2. Spark is the leading big data application running on K8s.

3. Over 77% of the respondents expect to migrate 50% or more of their workloads to containers with Kubernetes by the end of 2021.

4. Respondents cite "allocation and reallocation of resources" as the main challenge that comes with using containers and Kubernetes.

5. Over 42% of respondents say they use their cloud vendors' solutions to monitor their containers and Kubernetes, which often don't include big data application performance metrics.

What big data applications have you migrated to containers with Kubernetes?

Spark is a very fast and powerful processing engine designed for large-scale big data processing, making it already popular with enterprises that rely heavily on big data. Running Spark on Kubernetes presents numerous benefits for developers and big data analysts, such as eliminating dependency issues commonly occurring in Spark.

Enterprises need to first understand which apps they need to move to K8s and why. They also need to have a good grasp of the scope of migration prior to moving.

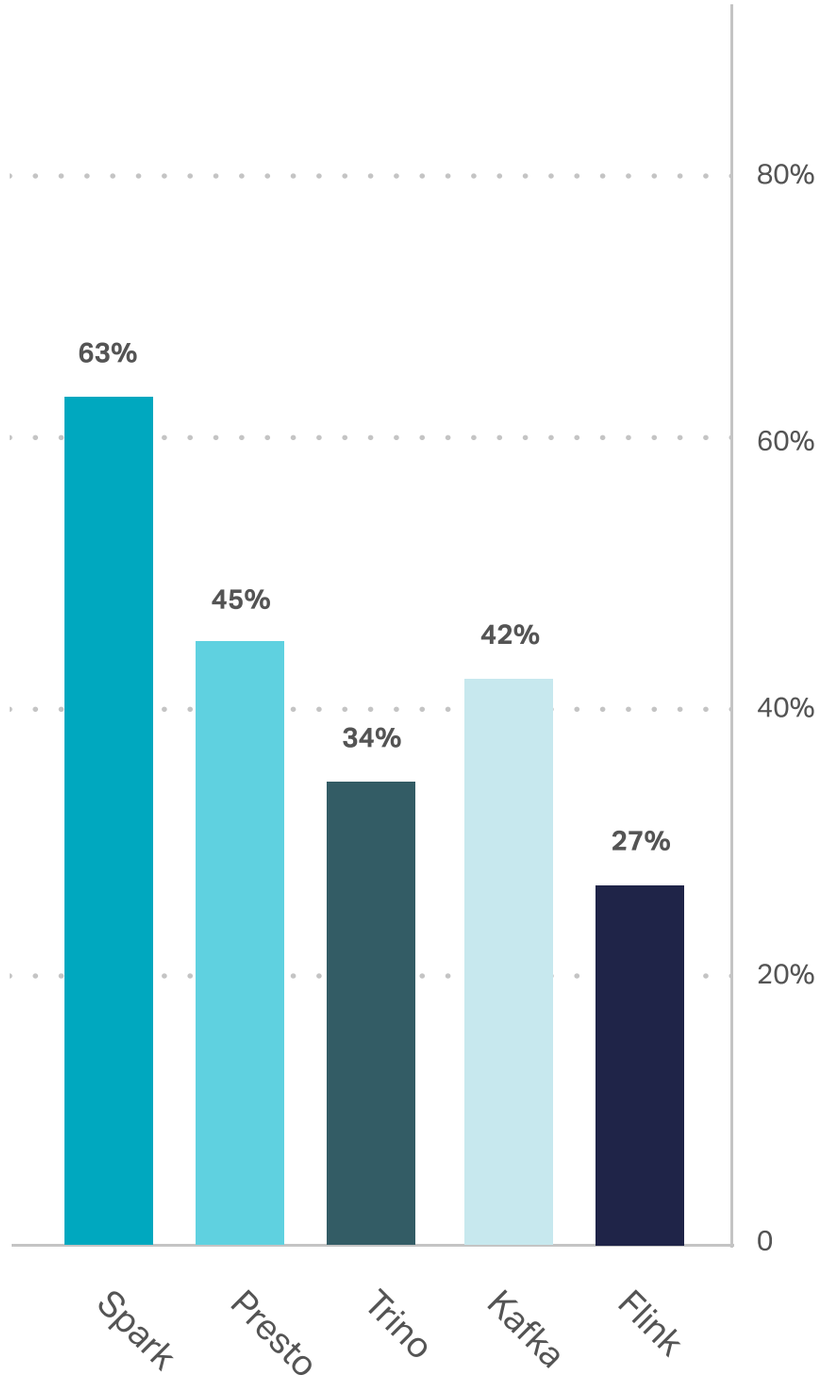
A. Spark

B. Presto

C. Trino

D. Kafka

E. Flink



What percentage of workloads do you expect to be migrated to containers with Kubernetes by the end of the year?

Results from our survey clearly indicate widespread adoption of Kubernetes across various industries. Over 77% of respondents expect to migrate 50% or more of their workloads to containers with Kubernetes by the end of 2021.

It's crucial for enterprises to know which apps to migrate to Kubernetes and which apps should remain on legacy platforms. While migrating apps to Kubernetes can result in better uptime, scalability, and vendor agnosticism, not all apps can or will be migrated.

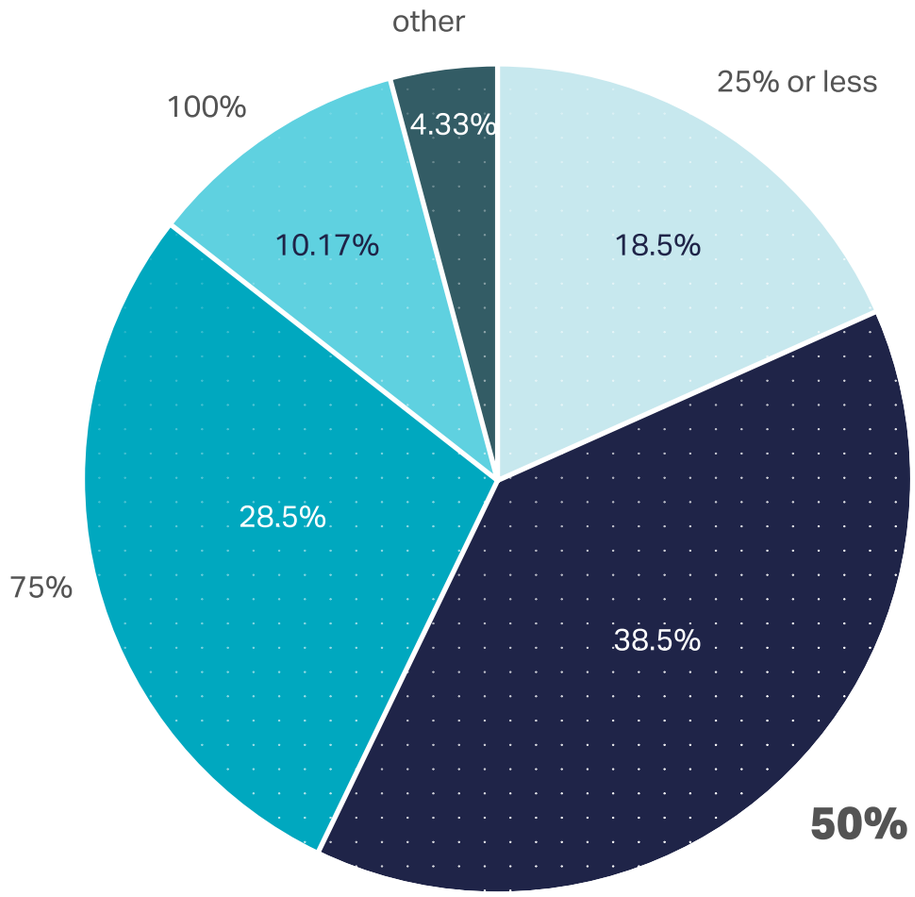
A. 25% or less

B. 50%

C. 75%

D. 100%

E. Other



What are your goals in migrating to containers with Kubernetes?

Enterprises are embracing containers because they allow for better resource utilization. They become vendor agnostic as well. Containers support agile and DevOps initiatives, resulting in the significant acceleration of the development, test, and production cycles.

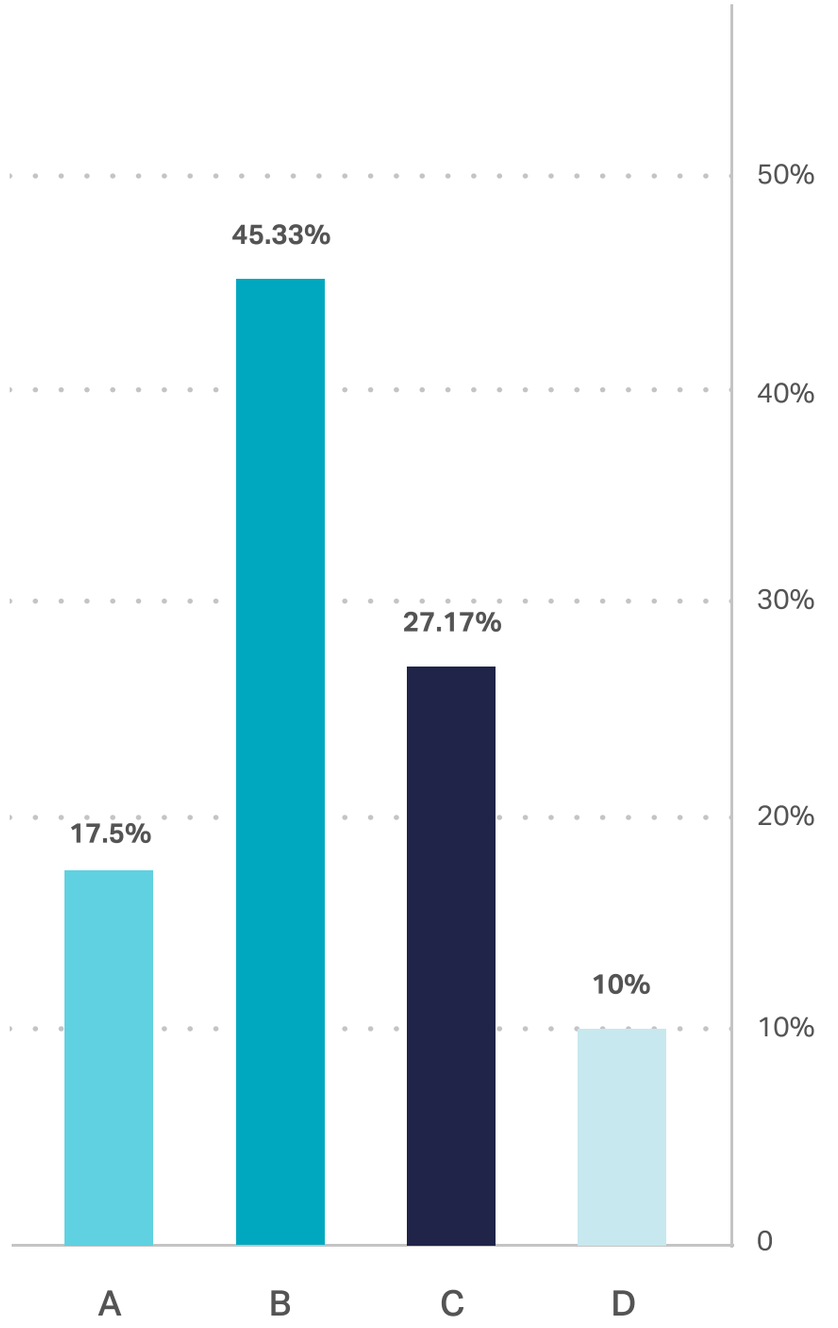
Running Spark on Kubernetes offers simpler administration, easier dependency management, and more flexible deployment. Kubernetes doesn't require a substantial amount of system resources, compared to conventional or hardware virtual machine (VM) environments. This leads to less overhead.

A. Decrease cost

B. Increase application performance/stability

C. Flexibility and portability of workloads

D. Leverage a multi-cloud solution to avoid being locked in with one cloud vendor



How are you monitoring your containers and Kubernetes today?

Despite the vast selection of premium monitoring solutions for Kubernetes available on the market, majority of users still rely on their cloud vendors' default monitoring tools to track and manage their containers and Kubernetes.

Most tools give you the details of the pods and containers, but no one offers the troubleshooting and waste-reducing data at the application level on Kubernetes.

As a Spark developer, you need in-depth visibility into your application. As a budget conscious CIO, you need to see the waste at the user and team level.

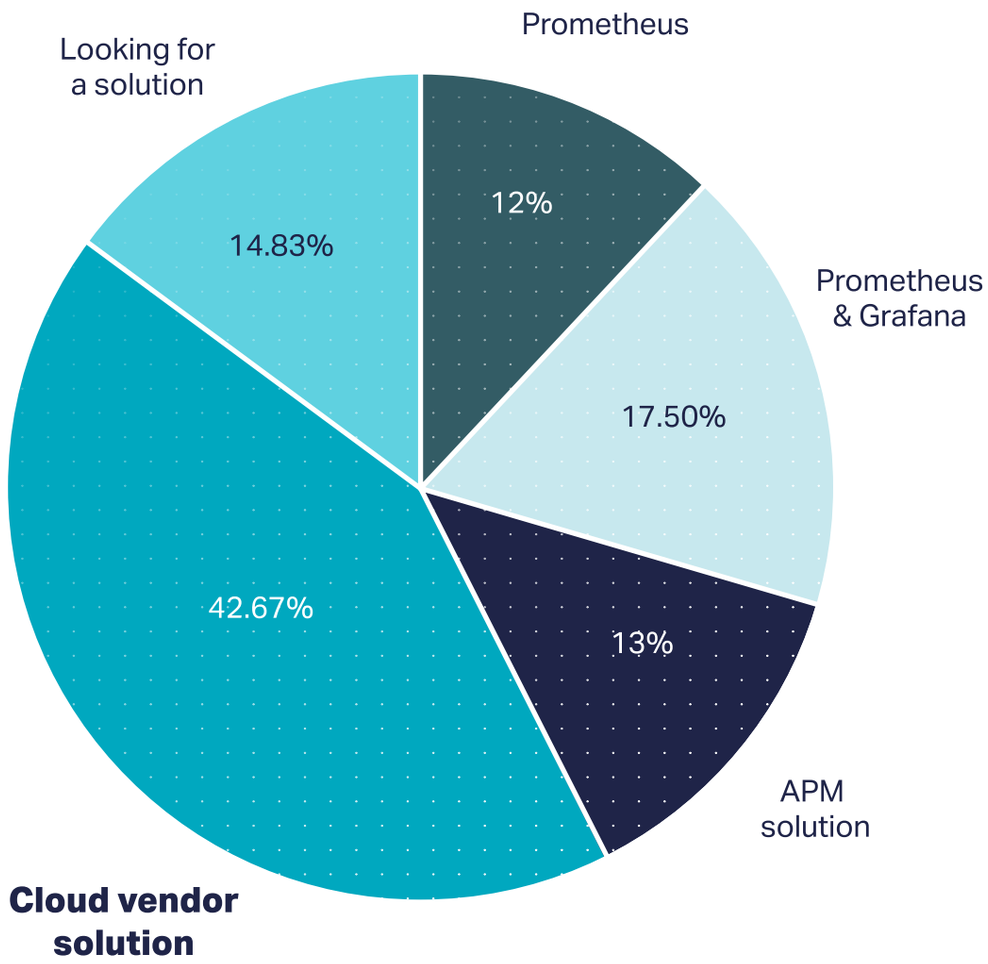
A. Prometheus

B. Prometheus and Grafana

C. APM solution

D. Cloud vendor solution

E. Looking for a solution



What are the metrics you monitor for Kubernetes today?

The majority of Kubernetes users just look at the container metrics—not the overall big data application performance. Only a few track other crucial indicators like resource and cost metrics.

The lack of comprehensive tools that provide granularity makes it difficult for enterprises to truly monitor and measure application performance on Kubernetes. While there are many tools that let users view nodes and pods, only few provide the depth required for monitoring big data application performance.

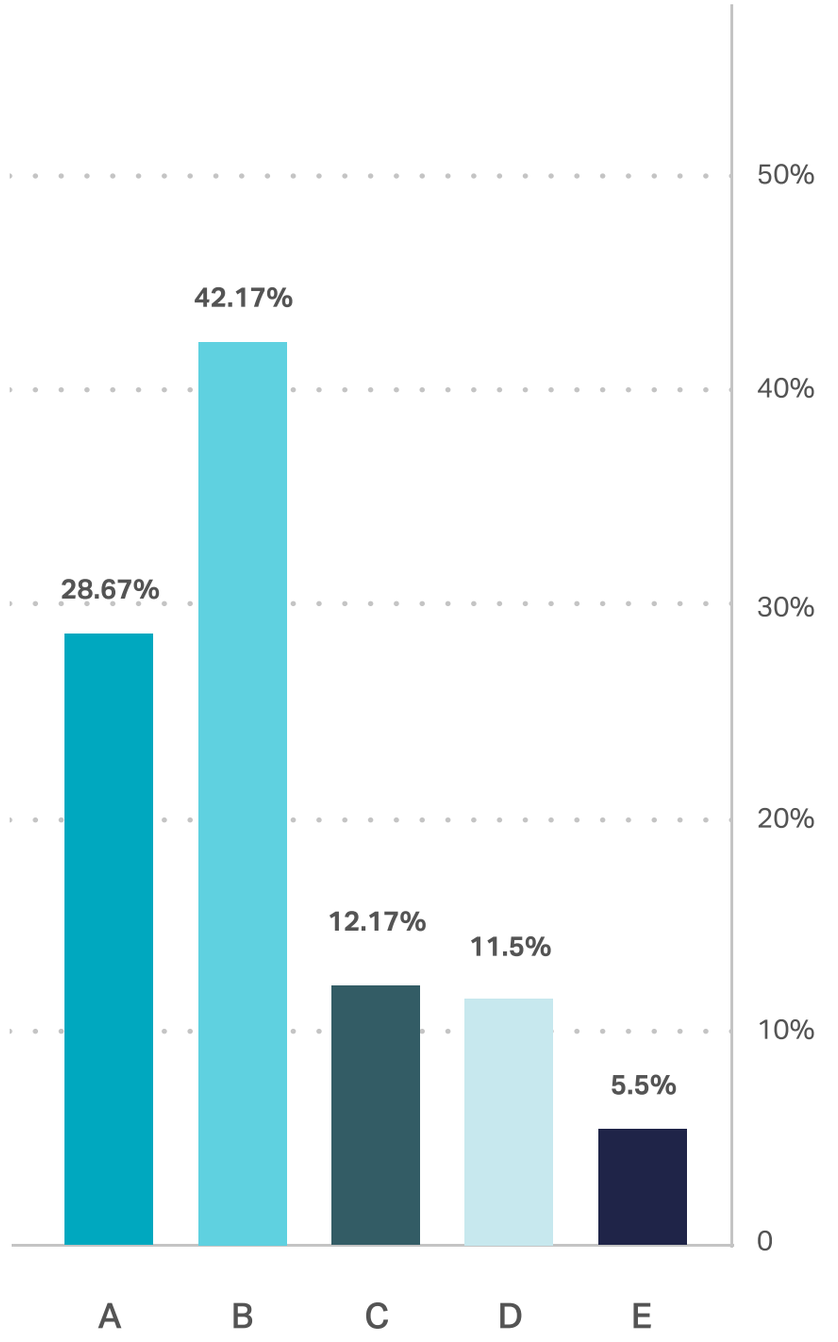
A. Resource metrics (CPU, memory, disk utilization, etc.)

B. Big data applications (e.g. Spark, Kafka, or Hive)

C. Cost metrics

D. Container metrics: resource utilization at the container level

E. None of the above



What is your main challenge with containers and Kubernetes today?

Kubernetes is a challenging technology to migrate to, and each challenge is connected to one another.

It's difficult to optimize resource allocation and perfectly tune Kubernetes for your big data applications. Tuning your clusters manually is not efficient and does not scale. There are a large number of applications you can use to monitor your K8s implementation. However, most look into only the most basic metrics. Combining multiple tools for a more complete view isn't practical. A single tool to manage resource allocation and application performance provides the best solution.

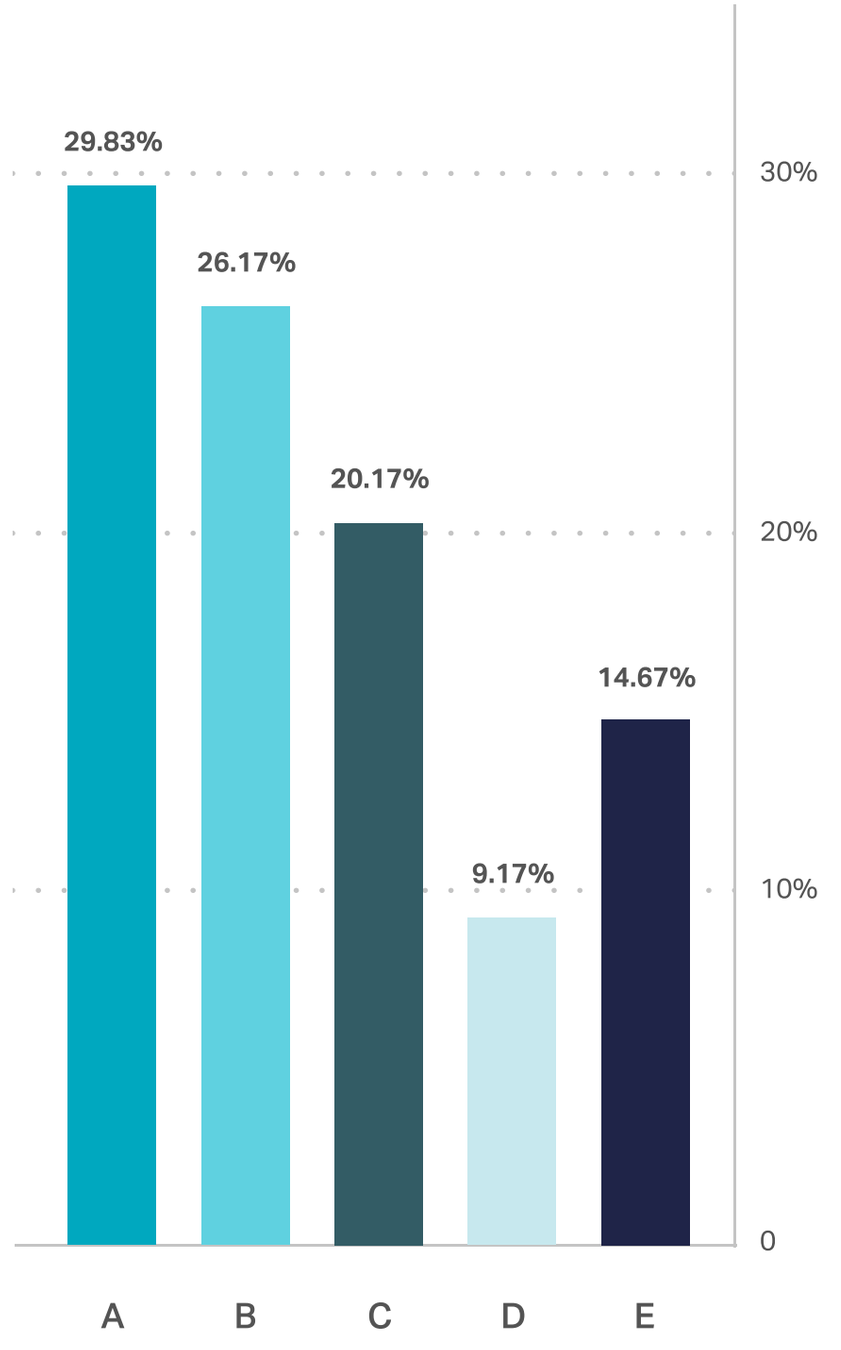
A. Allocation and reallocation of resources

B. Lack of comprehensive visibility and monitoring tools

C. Fragmented Kubernetes marketplace

D. Hard to manually tune Kubernetes

E. Complexity and scaling



pepperdata

Enterprises are migrating many of their big data workloads like Spark to Kubernetes. Achieving migration and post-migration success requires an observability tool that delivers more than just surface metrics.

Pepperdata provides enterprises with a comprehensive and powerful Kubernetes monitoring platform, built with superior visibility and comprehensive monitoring to closely track the performance and health of their Spark on Kubernetes workloads.

Start optimizing today.