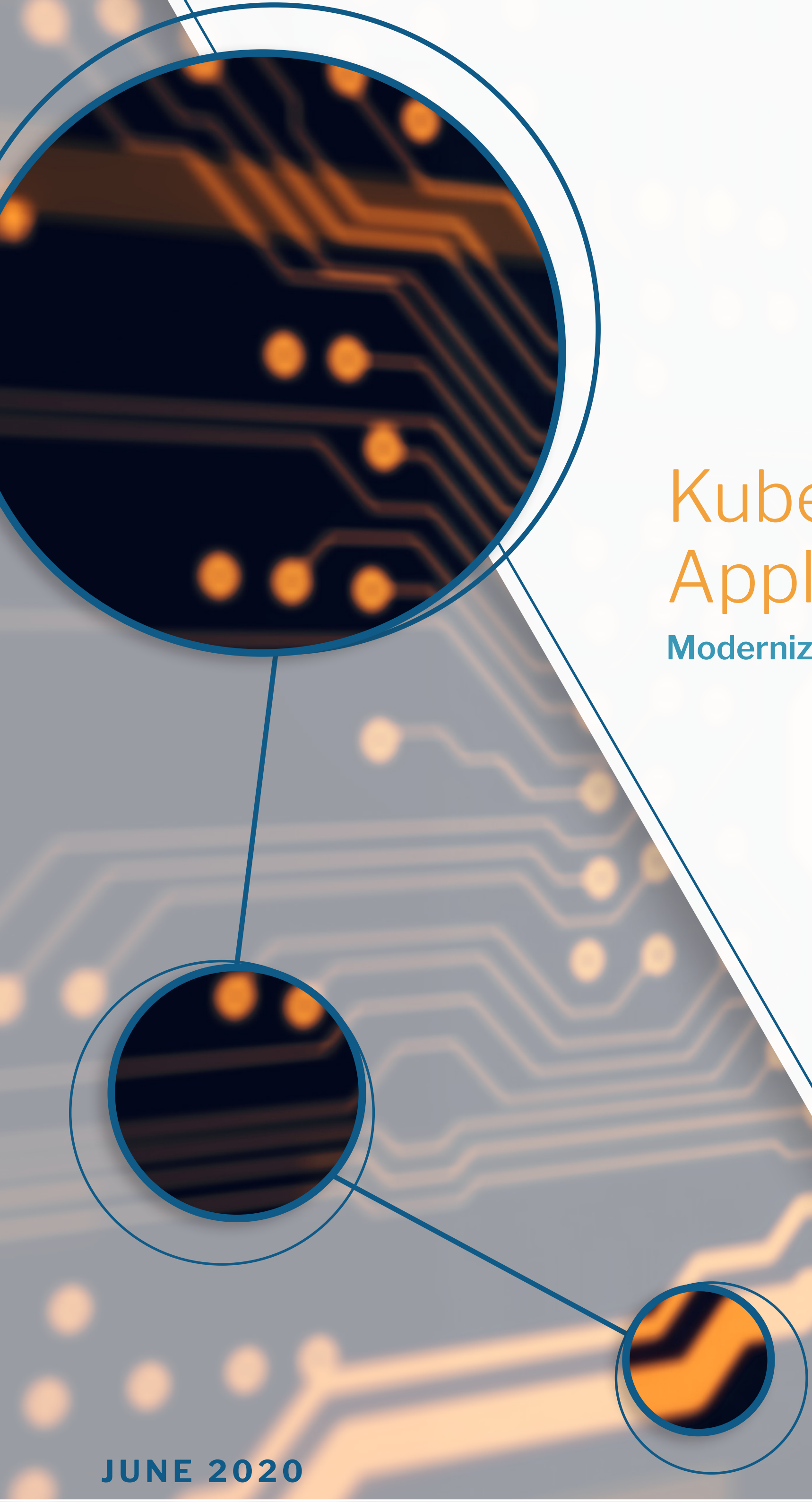
451 Research
Now a Part of
S&P Global Market Intelligence

PATHFINDER REPORT



# Kubernetes Backup and Application Portability

## Modernizing Data Protection

COMMISSIONED BY

kasten

JUNE 2020

# About this paper

A Pathfinder paper navigates decision-makers through the issues surrounding a specific technology or business case, explores the business value of adoption, and recommends the range of considerations and concrete next steps in the decision-making process.

## ABOUT THE AUTHORS

### LIAM ROGERS

ASSOCIATE ANALYST

As an Associate Analyst in 451 Research's Applied Infrastructure & DevOps Channel, Liam Rogers covers technology and business-model innovation across the enterprise storage landscape. His coverage includes hyperconverged infrastructure, software-defined storage and persistent storage for containers.

### STEVEN HILL

SENIOR ANALYST, APPLIED INFRASTRUCTURE AND STORAGE TECHNOLOGIES

Steven Hill is a Senior Analyst of Applied Infrastructure and Storage Technologies at 451 Research. He covers the latest generation of software-defined systems, hybrid cloud storage, unstructured data management and business continuity/disaster recovery solutions for enterprise customers.

# Executive Summary

When Docker started the modern container revolution in 2013, the promise of a more lightweight and mobile model for business applications captured the imagination of developers and infrastructure vendors throughout the industry. Today, containers reach from the laptop to the mainframe, but as always, the biggest challenges lie not in the containers themselves, but in the ecosystem needed to manage, secure and protect these new workloads at an enterprise level. Containers began as a simple design for stateless applications that could be rapidly spun up and just as rapidly deleted, but today, we find an increasing number of containers are now hosting stateful business applications that need persistent, enterprise-class storage. The persistent storage challenge has been addressed in several ways by storage and cloud vendors, but what's been missing is a comprehensive model for providing data and application protections needed to meet enterprise backup and disaster recovery requirements.

The rise of Kubernetes as the preeminent support and orchestration platform for containers has established a common model for the container ecosystem, so the next challenge lies in building out Kubernetes-native data protection that can provide enterprise-class capabilities while facilitating application portability and embracing the flexibility and efficiency offered by Kubernetes. Reliable backup and recovery remain critical, but it's only the starting point for a business continuity and disaster recovery (BC/DR) model that combines the automation and flexibility offered by a Kubernetes environment. Regardless of platform, BC/DR is a critical challenge, and in a recent 451 Research Voice of the Enterprise (VotE): Storage study of end users, one in three respondents had experienced an outage or degradation in the previous 12 months due to a failure related to infrastructure resources. Of those affected, large companies tended to be more susceptible to failures than smaller organizations – 24% of organizations under 1,000 employees had suffered an outage compared with 41% for companies with 1,000+ employees.

We believe that containers and Kubernetes will continue to be instrumental technologies for the creation and operation of cloud-native applications well into the future, and meshing these technologies with strong data management and protection practices needs to play a key role in any critical applications operating in containers.

In this paper we discuss:

- The ongoing adoption of containers and Kubernetes and the growing trend of stateful applications being built on containers.
- How Kubernetes impacts day-to-day operations and the challenges it introduces to maintaining a backup and DR strategy that supports application portability.
- The importance of aligning cloud-native applications with a cloud-native and application-centric approach to data management.
- The need for common, policy-based protections for critical applications and data.
- The importance of application-awareness in protecting stateful applications.

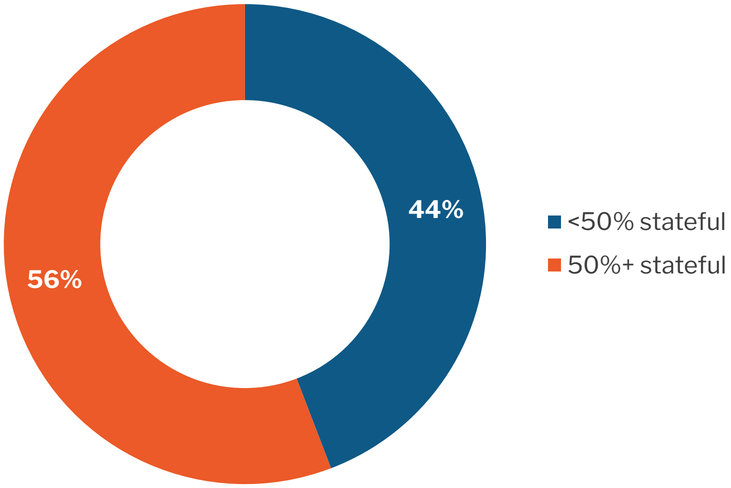
# The Rise of Stateful Applications Using Containers

Containerization can provide enterprises with myriad benefits including a path away from monolithic applications to microservice-based architectures, application mobility via the lightweight nature of containers, and infrastructure resource efficiency by limiting the overhead traditionally required with virtualization. The drive to modernize applications and implement DevOps practices in the enterprise continues to fuel the adoption of containers and the usage of cloud-native applications. In a recent VotE: DevOps study, 73% of organizations indicated they have adopted containers in some capacity (22% have adopted them across the IT organization, 33% have adoption at team level and 18% are in POC) and only 6% of organizations do not have container adoption in their two-year plan.[1] Going hand in hand with the uptake of containers is the use of an orchestration platform such as Kubernetes, and currently, 56% of organizations have Kubernetes in use.

While containers were initially intended for use with ephemeral and stateless applications where data does not persist past the lifespan of the container, they are increasingly being used for stateful applications. In the same DevOps study, 56% of organizations indicated that half or more of their container applications are stateful (see Figure 1).[2] Containers are ideal for use with stateless applications, but many of their strengths are equally appealing for the development of applications, such as databases, where there is considerable value in retaining data in the event of a container's termination.

Figure 1: Stateful applications are prevalent

*Source: 451 Research's Voice of the Enterprise: DevOps, Workloads and Key Projects 2020*
*Q: What percentage of your container applications are stateful versus stateless?*
*Sample Size = 430 Base: Containers users*



[1] *451 Research, Voice of the Enterprise: DevOps, Workloads and Key Projects 2020*
[2] *451 Research, Voice of the Enterprise: DevOps, Workloads and Key Projects 2020*

Now, organizations must not only contend with the learning curve of implementing containers in production, but also the challenges of having data persist for key stateful workloads. The development of the Container Storage Interface (CSI) has been instrumental in standardizing the approach to block and file storage in Kubernetes. However, CSI continues to be a work in progress, and enterprise-grade data management capabilities are far off from being a part of upstream Kubernetes. Managing applications and data persistence can be a complex task because stateful applications in a Kubernetes environment are composed of many pieces including block and file constructs, various Kubernetes objects that include cluster state (common objects being Secrets and ConfigMaps) and often multiple databases (such as SQL or NoSQL). This requires an application-centric backup that accounts for and understands the relationships between these different constructs within Kubernetes. This is juxtaposed with other approaches to backup, which can focus solely on the physical or virtual infrastructure layers. It's imperative that enterprises deploying stateful containerized applications into production implement an effective strategy for data persistence and management so that valuable data can be retained and so that application state can be preserved on a granular level that will enable restores and migrations without disruptions.

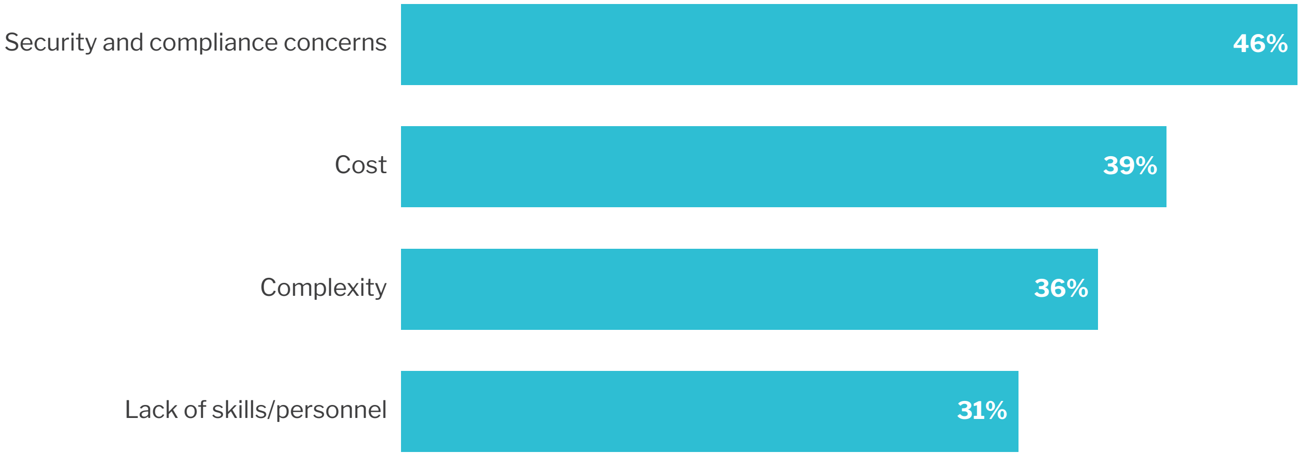
## Kubernetes 'Day 2' Challenges

While Kubernetes makes developing and deploying containers simple and efficient, it's only the beginning of the 'ops' aspect of DevOps. Stateless containers make things really simple: if a stateless application fails, it can simply be restarted without consequence other than its temporary absence. But stateful applications like databases make things a bit more complex when it comes to protecting these applications on an ongoing basis. These 'Day 2' challenges commence immediately after deployment and are ongoing as long as the application remains in production as with any other legacy environment, but this is where the flexibility and automation of Kubernetes can really shine if planned for from the beginning. Running a Kubernetes environment offers remarkable flexibility and freedom of choice, but along with that comes a certain amount of complexity. Fortunately, a key value proposition of the cloud-native approach lies in its support for automation, combined with the simplicity of APIs for tapping into a diverse set of microservices that can effectively mitigate that complexity.

The learning curve with Kubernetes is not limited to the initial adoption and deployment of clusters and containers, but also lies in understanding the ongoing needs of stateful applications, as well as the many flexible options that a cloud-native approach offers beyond the development phase. If challenges like security, compliance and BC/DR are planned for from the start as part of a Kubernetes ecosystem, protection and compliance can be ensured as a simple matter of course, rather than an external process with a separate set of challenges. Integrating data protection adds to the benefits of using containers, such as application portability and the ability to distribute workloads across clusters, regions or between types of infrastructure such as cloud and on-premises infrastructure. Beyond deployment, scaling stateful applications can also add to the complexity challenge, but again, this is where a Kubernetes-native approach to security, data and application protection can reduce or eliminate some up-front challenges by using an automated, policy-based model.

Figure 2: Top four hurdles for use of cloud-native technology

*Source: 451 Research's Voice of the Enterprise: DevOps, Workloads and Key Projects 2020*
*Q: What are the primary hurdles/challenges of using cloud-native technology such as containers, Kubernetes or serverless in your organization? Please select all that apply.*
*Sample Size = 512, Base: All respondents*



According to our data, 46% of organizations are leveraging public cloud as the primary approach to data persistence for stateful applications running on containers, followed by 33% of organizations using existing on-premises infrastructure.[3] Given that hybrid and multicloud deployments are increasingly prevalent, both data container storage and data management platforms must be able to span multiple environments. This includes providing deep integrations with storage and the ability to discover and protect applications whether they are in the cloud or on-premises, as well as facilitate data movement between them for data protection and application migration purposes. By leveraging safety measures such as end-to-end encryption, users can also address some of the security concerns about portability because data will be protected in flight and at rest regardless of the ultimate destination or secondary location.

[3] *451 Research, Voice of the Enterprise: Storage, Data Management and Disaster Recovery 2020*

## Enabling Application Portability

An organization's application may ultimately be destined for an alternate location from where it was developed or initially deployed. With increasing adoption of hybrid and multicloud deployments, there is also a need to ensure application portability. Portability can mean different things: it could be the ability to migrate an app between clusters in the same environment or between clusters deployed in different clouds or regions with different underlying infrastructure, even between different versions of Kubernetes itself. This movement can be desirable for myriad reasons such as cost, performance, security and to avoid vendor lock-in. While individual containers can be inherently portable due to their lightweight design, a Kubernetes application made up of many pieces must account for all these elements and the associated persistent data to ensure application portability across clouds. By keeping applications portable, organizations can exert greater control over their ability to manage cost and performance optimization.

## Skills Gaps and Operational Challenges

While the technology itself is one facet of deploying Kubernetes, there are important organizational considerations as well. Infrastructure teams are already faced with having to do more with less, but the advent of Kubernetes means that these same teams must also contend with mastering new skill sets. According to our 2019 data on organizational dynamics, 52% of organizations indicated that they are currently experiencing a skills shortage among their infrastructure-based personnel (compared to 42% in 2018).[4] We also see that the areas of expertise where skills gaps are occurring are not those related to traditional core infrastructure skills. The areas where skills are most in short supply include DevOps (47% of organizations experiencing an infrastructure skills gap), container administration (39%) and database administration (35%). These categories are some of the most pertinent to cloud-native applications. Our data shows that training existing staff is the most prominent way to contend with these skills gaps, but in addition to improving the skill sets of current staff, there is also value in providing better tooling, including automation, to mitigate this challenge.

While Kubernetes' flexibility can lead to complexity, this can be countered by using tooling that emphasizes ease of setup, a simplified management experience and integrated automation to reduce the burden on staff and make up for Kubernetes-related skills gaps. Such tooling might include GUIs that abstract away the underlying complexity of Kubernetes, as well as policy-based management to automate routine but critical data management tasks. For example, the provisioning of storage in Kubernetes is done based on a system of 'persistent volume claims.' This provides a high-level, self-service model for allocating on-demand, persistent storage for developers or application owners without the need for deep storage management experience. This same self-service concept provides a model that can be leveraged alongside role-based access controls to include specific security and data protection requirements.

[4] *451 Research, Voice of the Enterprise: Storage, Organizational Dynamics 2019*

# Gaps in Current Offerings

Kubernetes continues to evolve along with the ecosystem of tools and services that surround it. However, as enterprise adoption is accelerating, organizations have not kept up with implementing tooling around Kubernetes so that cloud-native applications are as robust as the rest of their workloads. As an enterprise determines the best method for storage and data persistence, this transitions naturally into the need to protect that data and manage the full extent of its lifecycle. In our VotE: Storage study, 43% of organizations with containers in use indicated that they are relying on legacy data protection tools as their primary data protection strategy for containerized applications and associated data volumes.[5] While enterprises are always faced with getting the most value out of existing purchases, the strategy of using legacy tooling means that many organizations will face a mismatch between their Kubernetes-based applications and their backup and DR plans, which solidifies the need for Kubernetes-native backup.

## Dynamic Patterns

Legacy data protection tools tend to be geared toward backing up an entire VM or disk and not toward the dynamic nature of cloud-native applications. While it is true that containers are still commonly run on top of VMs, a Kubernetes application may be distributed across many VMs. A VM-based protection strategy will, therefore, not guarantee that containerized applications are being effectively backed up just because the VM may be. Further, given that there are multiple Kubernetes resources that are integral parts of a Kubernetes-based application, Kubernetes-native backup and restore needs to account for these various components, and the only logical approach to doing so is to make the data management process application-centric and aware.

To take an application-centric approach to data management, it's important to have visibility across environments and automated discoverability to adapt to portable and potentially distributed applications. Users should also be aware that even capabilities that are part of Kubernetes or storage platforms geared specifically toward containers and Kubernetes may not necessarily offer backup and DR capabilities optimized for container-based applications that can be used for application-consistent restores. In addition to offering better data protection, tooling that is Kubernetes-native can also add an element of portability that will simplify application migrations.

## Application-Centric Consistency

In particular, snapshots have become commonplace, and volume snapshots have become a part of Kubernetes' functionality, but the ability to restore from snapshots can be limited because they are usually not durable, and snapshot-based backup does not guarantee application-consistent recovery. Snapshots will only capture certain parts of the workload, such as the volumes, but will not capture other relevant Kubernetes objects that are critical components of the application, which provides an incomplete picture of the application's state. Currently, 14% of organizations are relying solely on snapshots as their primary data protection strategy for containerized applications and data volumes. Where snapshots fall short, implementing a Kubernetes-native and application-centric approach to backup provides a way to protect the entirety of an application.

[5] *451 Research, Voice of the Enterprise: Storage, Data Management and Disaster Recovery 2020*

# Conclusion

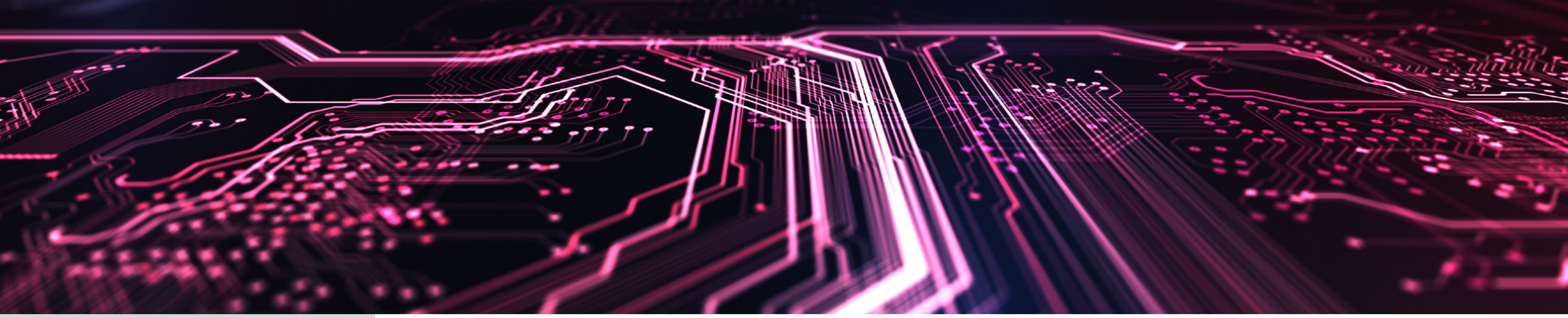
Containers and Kubernetes offer a remarkably flexible and efficient model for next-generation applications, but we also believe it's an imperative to provide consistent security and data protection for business- and mission-critical applications regardless of the production environment on which they reside. Often, the efficiency of these services depends upon leveraging the native capabilities of the production platform, so developing within and adhering to a Kubernetes-native backup strategy can be a key feature in protecting against or reducing the impact of system failures. Today's data management is not only about ensuring availability and resiliency but must also ensure regulatory compliance as well as address security concerns and enable workload portability. Providing all these attributes can be challenging in the context of cloud-native applications that may rely on multiple data services that can be spread across clusters and locations, which is why there is a need for data management tooling that is tailored to supporting these types of applications.

Portability is an additional major benefit of a containerized approach; it provides the ability to move applications across namespaces, clusters, regions, infrastructure providers or even different Kubernetes distributions. But making stateless applications portable is easy; it's data at scale that really presents a portability problem for multicloud, data-intensive use cases, and moving data is the key to efficiently restoring, cloning, upgrading and migrating applications. This is as critical for stateful, cloud-native applications as it is for traditional infrastructure systems, and both require comparable data management practices and tooling to protect data as well as resume production workloads quickly. These criteria should be considered in the context of both current and future needs as container use becomes more prevalent.

As the ecosystem around Kubernetes continues to grow rapidly, the potential for complexity can be daunting for users, but once the learning curve is over, the richness of this ecosystem also provides the opportunity to develop deep integrations and automate much of the busy work of IT management. Enterprises should explore tooling options that support integrations with their data sources in use but also abstract underlying management tasks in a way that makes day-to-day operations easy to accomplish. Taking a software-driven, application-centric approach to cloud-native BC/DR is a logical way of protecting next-generation production workloads in a way that could ensure global compliance, data protection and application resilience while reducing wasted effort and management burden on the staff tasked with doing so.

# Recommendations

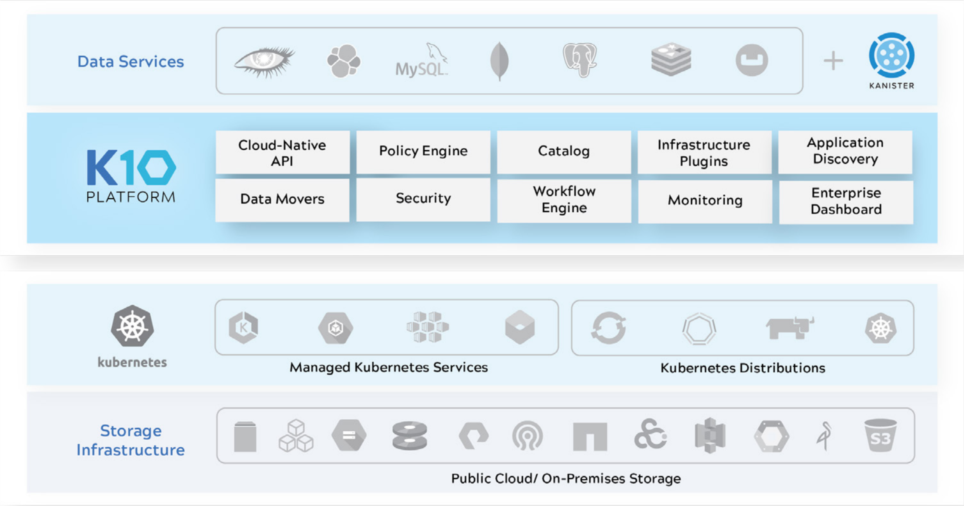
- **Take an application-centric approach to data protection and management** – The criticality of the application as well as its specific protection requirements should dictate the steps to ensure BC/DR needs are met.
- **Establish and regularly test a BC/DR plan** – BC/DR is a formula with a lot of variables, so it's important to have a plan in place that's easy to understand, documented, functional in the absence of key personnel and tested on a regular basis. A backup is only valid if it can be restored.
- **Be aware of dependencies between critical applications and supporting resources** – It's just as important to protect key resources as the applications they support. Restoring an application may be impossible without all of its supporting services.
- **Leverage automation** – Time is of the essence in any form of disaster, making a combination of system automation and human input a major factor in minimizing RTO/RPO estimates.
- **Adopt a clear set of data protection policies** – Critical applications should be bound by a common set of policies that can span all infrastructure options, on- and off-premises.
- **Maintain ongoing security** – It's important to ensure system security throughout a crisis event via role-based access controls and end-to-end encryption to protect applications and data.
- **Always have a failback strategy** – Ensure that a recovery scenario also simplifies a normal resumption of services, whether back to the original environment or to a different or new one.



CONTENT PROVIDED BY:

The Kasten K10 data management platform, purpose-built for Kubernetes, provides an application-centric approach to data protection and management. K10's capabilities of automatic application discovery, deep integrations with relational and NoSQL databases, Kubernetes distributions across on-premises, and clouds provide enterprises the freedom of infrastructure choice without sacrificing operational simplicity.

K10's policy-driven automation makes critical enterprise needs of Kubernetes Backup/Recovery, Application Migration, and Disaster Recovery a snap to set up and monitor. Critical enterprise requirements for a robust data management platform, including extensibility, end-to-end security, and self-service capabilities are built into Kasten K10.



You can:

- **Try Kasten K10** - free and in less than 10 minutes!
- Dive into Kasten K10 by visiting the Kasten **product page** and reading the **K10 datasheet** and **docs**.
- Contact Kasten via **email**, **web**, **Twitter**, or **LinkedIn**.

451 Research
Now a Part of
S&P Global Market Intelligence

## About 451 Research

451 Research is a leading information technology research and advisory company focusing on technology innovation and market disruption. More than 100 analysts and consultants provide essential insight to more than 1,000 client organizations globally through a combination of syndicated research and data, advisory and go-to-market services, and live events. Founded in 2000, 451 Research is a part of S&P Global Market Intelligence.

The terms of use regarding distribution, both internally and externally, shall be governed by the terms laid out in your Service Agreement with 451 Research and/or its Affiliates. The information contained herein has been obtained from sources believed to be reliable. 451 Research disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although 451 Research may discuss legal issues related to the information technology business, 451 Research does not provide legal advice or services and their research should not be construed or used as such.

451 Research shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve its intended results. The opinions expressed herein are subject to change without notice.

**NEW YORK**
55 Water Street
New York, NY 10041
+1 212 505 3030

**SAN FRANCISCO**
One California Street,
31st Floor
San Francisco, CA 94111
+1 212 505 3030

**LONDON**
20 Canada Square
Canary Wharf
London E14 5LH, UK
+44 (0) 203 929 5700

**BOSTON**
75-101 Federal Street
Boston, MA 02110
+1 617 598 7200

451 Research
Now a Part of
S&P Global Market Intelligence