

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

Image credit: Golden Sikorka

[Joep Piscaer, Enrico Signoretti](#)
Dec 30, 2021 -- Market Radar

GigaOm Radar for Kubernetes Data Protection^{v 2.02}

Backup Solutions for Kubernetes-Based Applications

Table of Contents

- 1 [Summary](#)
- 2 [Market Categories and Deployment Types](#)
- 3 [Key Criteria Comparison](#)
- 4 [GigaOm Radar](#)
- 5 [Vendor Insights](#)
- 6 [Analyst's Take](#)
- 7 [About Enrico Signoretti](#)
- 8 [About GigaOm](#)
- 9 [Copyright](#)

1. Summary

Kubernetes is the de facto standard for container orchestration, and it's being used by born-in-the-cloud startups and cloud-native enterprises alike. In 2021, Kubernetes was in production on-premises, in the cloud, and even at the edge for many different types of applications, including those that Kubernetes wasn't initially built for.

Kubernetes was never really built for stateful applications, and by default, it lacks features for data protection. However, we see many organizations building and running their stateful applications on top of Kubernetes, indicating there's a gap in functionality between what Kubernetes offers and what the (enterprise) market wants.

Unfortunately, existing data protection tools, mostly built for legacy technologies such as virtual machines, do not fit well into the container paradigm. Vendors are adapting existing solutions or creating new products from scratch that are often better aligned with the cloud-native and container paradigms.

Many of these data protection solutions include other data management features, such as data security, disaster recovery, or heterogeneous data migration capabilities. There's some overlap among data storage solutions, data protection solutions, and data management solutions in the cloud-native space, with each solution offering some adjacent features.

This is a GigaOm Research Reprint: Expires Dec 30, 2022

functionality we expect to see in this space, as summarized in **Figure 1**.

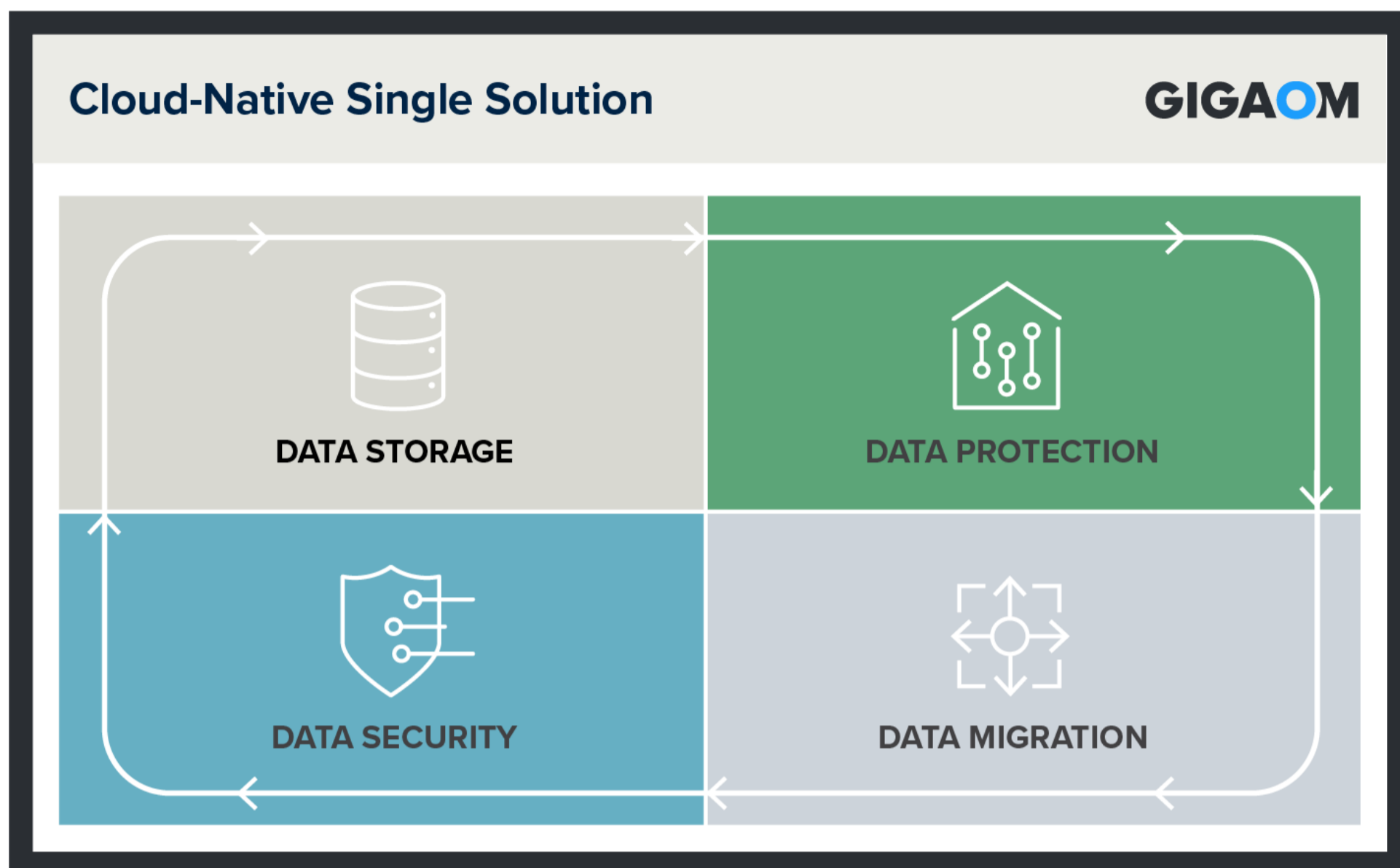


Figure 1. Cloud-Native Data Storage, Protection, Migration, and Security Converge Into Single Solutions

The market for cloud-native data protection is growing rapidly, with both incumbent vendors and challengers in the market competing for completeness of features across the four key pillars. Differences can be observed between those targeting existing, more traditional infrastructure alignment, and those targeting fully cloud-native environments.

In any case, we see the growing need for flexible, adaptive solutions that can follow the ever-changing requirements of their customers. Multi-platform, multi-cloud, multi-environment (including edge), multi-team, and self-service features are quickly becoming differentiating features that ensure successful adoption not for just one use case but for continuously changing use cases across the entire enterprise.

HOW TO READ THIS REPORT

This GigaOm report is one of a series of documents that helps IT organizations assess competing solutions in the context of well-defined features and criteria. For a fuller understanding consider reviewing the following reports:

Key Criteria report: A detailed market sector analysis that assesses the impact that key product features and criteria have on top-line solution characteristics—such as scalability, performance, and TCO—that drive purchase decisions.

GigaOm Radar report: A forward-looking analysis that plots the relative value and progression of vendor solutions along multiple axes based on strategy and execution. The Radar report includes a breakdown of each vendor's offering in the sector.

Solution Profile: An in-depth vendor analysis that builds on the framework developed in the Key Criteria and Radar reports to assess a company's engagement within a technology sector. This analysis includes forward-looking guidance around both strategy and product.

2. Market Categories and Deployment Types

For a better understanding of the market and vendor positioning (**Table 1**), we assess how well solutions for Kubernetes Data Protection are positioned to serve specific market segments, deployment types, and architectures.

We recognized three market segments for this report:

- **Small-to-medium businesses (SMB):** In this category, we assess solutions on their ability to meet the needs of organizations ranging from small businesses to medium-sized companies. Also assessed are departmental use cases in large enterprises. In this category, ease of use and quick consumption models, like software-as-a-service (SaaS), are important factors, as are simple pricing models such as subscription-based.
- **Large enterprise:** Here offerings are assessed on their ability to support large and business-critical projects. Optimal solutions in this category will have a strong focus on flexibility, performance, data services, and features to improve security and data protection, as well as extensive data mobility features for various migration scenarios. Scalability is another big differentiator, as is the ability to deploy the same service in different environments.

This is a GigaOm Research Reprint: Expires Dec 30, 2022

- **Edge and other specialized use cases:** A new deployment type becoming popular is deployment in edge and IoT-like scenarios, including telco and retail deployments. In these cases, policy-based fleet management of data protection across many clusters is a key differentiator, as are dark deployments.

We also recognize two deployment models for solutions in this report:

- **SaaS (managed and hosted):** Available only in the cloud and as a managed service, this approach is usually based on a pay-as-you-go subscription model. Users do not need to manage the infrastructure or backup repositories, just backup policies and day-to-day operations. This deployment model includes first-party SaaS (run and managed directly by the data protection vendor) and third-party SaaS (run and managed by a managed hosting partner). Often, the SaaS deployment model still requires an agent or component to run on each protected cluster.
- **Self-hosted (on-premises or in-cloud):** These solutions are meant to be installed both on-premises and in your cloud environment. While they are more complex to deploy and manage, they are more flexible in terms of where and how they are deployed. This deployment type is more suitable for those with stricter requirements for operational control or who have specific deployment requirements. The architecture of this deployment type can vary among hub-and-spoke, fully per-cluster, self-contained deployments, and a number of other architectures.

Table 1. Vendor Positioning

	MARKET SEGMENT			DEPLOYMENT MODELS	
	SMB	Large Enterprise	Edge and Other Specialized Use Cases	SaaS (Managed and Hosted)	Self-Hosted (On-Premises or In-Cloud)
CloudCasa by Catalogic	+++	++	++	+++	+++
Commvault	+++	+++	+++	++	+++
Dell Technologies	++	+++	+++	+	++
Druva	++	+	+++	+++	-
HYCU	++	++	++	++	++
IBM	++	++	++	-	+++
Kasten by Veeam	+++	+++	+++	++	+++
Portworx by Pure Storage	+++	+++	+++	+	+++
Robin	++	++	+++	+	+++
Trilio	+++	+++	++	+	+++
Veritas	++	+++	+++	+	++
VMware	+	+	+++	+	+++
Zerto	++	+++	+++	+	++

Source: GigaOm 2021

- +++ Exceptional: Outstanding focus and execution
- ++ Capable: Good but with room for improvement
- + Limited: Lacking in execution and use cases
- Not applicable or absent

In addition, data protection solutions for Kubernetes can also be categorized according to their main feature set:

- **Traditional data protection solutions that support Kubernetes:** These solutions have a mature feature set for VM-based environments, and have added support for Kubernetes.
- **Cloud-native storage with data protection capabilities:** These solutions offer data protection capabilities on top of a cloud-native data storage product. You can't use the former without adopting the latter.
- **Cloud-native data protection:** Data protection solutions specifically designed to work with Kubernetes.

Table 2. Architecture

	ARCHITECTURE		
	Traditional	Cloud-Native Storage	Cloud-Native Data Protection
CloudCasa by Catalogic	-	-	+++

This is a GigaOm Research Reprint: Expires Dec 30, 2022

Dell technologies	+++	-	-
Druva	-	-	+++
HYCU	+++	-	-
IBM	++	++	-
Kasten by Veeam	-	-	+++
Portworx by Pure Storage	-	+	++
Robin	-	+++	-
Trilio	-	-	+++
Veritas	+++	-	-
VMware	-	-	+++
Zerto	+++	-	-

Source: GigaOm 2021

- +++ Exceptional: Outstanding focus and execution
- ++ Capable: Good but with room for improvement
- + Limited: Lacking in execution and use cases
- Not applicable or absent

3. Key Criteria Comparison

Building on the findings from the GigaOm report, “Key Criteria for Evaluating Kubernetes Data Protection,” **Table 3** summarizes how each vendor included in this research performs in the areas we consider differentiating and critical in this sector. **Table 4** offers insight into evaluation metrics—the top-line characteristics that define the impact a solution will have in an organization. The objective is to give the reader a snapshot of the technical capabilities of different solutions and define the perimeter of the market landscape.

Table 3. Key Criteria Comparison

	KEY CRITERIA						
	Multi-Cloud & Multi-Distribution	Multi-Cluster, Multi-Tenant	Environmental Awareness	Disaster Recovery	App and Data Migration	Ransomware Protection	Analytics
CloudCasa by Catalogic	+++	+++	+++	+	+	++	++
Commvault	+++	+++	+++	++	+++	+++	+++
Dell Technologies	+++	++	++	++	++	+++	++
Druva	-	+	+	+	-	-	+
HYCU	+	++	++	++	++	+	++
IBM	+++	++	++	++	++	+	++
Kasten by Veeam	+++	++	+++	++	+++	+++	+++
Portworx by Pure Storage	+++	+++	++	+++	+++	+	+++
Robin	+++	++	+++	+++	+++	++	++
Trilio	+++	++	+++	++	+++	++	+++
Veritas	++	+++	+	+	+	+++	+++
VMware	++	+	+	+	++	+	++
Zerto	+++	+	++	+++	+++	+	+++

Source: GigaOm 2021

- +++ Exceptional: Outstanding focus and execution
- ++ Capable: Good but with room for improvement
- + Limited: Lacking in execution and use cases
- Not applicable or absent

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

evolution over the coming 12 to 18 months.

Figure 2. GigaOm Radar for Data Protection for Kubernetes

As you can see in the Radar chart in **Figure 2**, this report shows the typical characteristics of a new market with a number of startups leading the pack and with a series of converging ideas that differ in their implementation but are actually similar in their high-level vision. However, a few vendors are pursuing different paths and alternative solutions. Established vendors are still far from the bull's-eye but are working quickly to bridge the gap with the leaders.

In general, the market is very dynamic, and vendors are striving to build a consistent experience across multiple clouds while providing advanced application and data mobility. We see different approaches in this market, ranging from simple SaaS with a great user experience, to more complex and feature-complete solutions aimed at bigger enterprises, all-in-one solutions that include (and are exclusive to) primary or secondary storage, and solutions that specifically integrate into certain platforms.

In this context, a number of solutions are doing very well, including Kasten, Portworx, and Trilio. Some more established, traditional vendors like Commvault are also doing very well, by combining solutions for SaaS applications, on-premises (VM-based) infrastructure, and containers efficiently. Some solutions, such as CloudCasa, while basic, tick all the right boxes for those buyers seeking a simpler solution with a great user experience.

Additionally, we see some vendors combining data protection solutions with other functionality, like data storage. Robin receives a notable mention, as its data protection features are great, but they are exclusive to Robin's storage platform.

All in all, the data protection field is moving quickly and is accelerated further by Velero's growing importance in the CNCF ecosystem as the underpinning technology for many of the solutions discussed in this Radar. In the coming year, we'll see well-established data protection and management paradigms from more traditional architectures make it into the Kubernetes and broader cloud-native world, helping to improve the quality, resilience, and availability of our applications, regardless of where and how they run.

INSIDE THE GIGAOM RADAR

The GigaOm Radar weighs each vendor's execution, roadmap, and ability to innovate to plot solutions along two axes, each set as opposing pairs. On the Y axis, **Maturity** recognizes solution stability, strength of ecosystem, and a conservative stance, while **Innovation** highlights technical innovation and a more aggressive approach. On the X axis, **Feature Play** connotes a narrow focus on niche or cutting-edge functionality, while

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

circle is almost always empty, reserved for highly mature and consolidated markets that lack space for further innovation.

The GigaOm Radar offers a forward-looking assessment, plotting the current and projected position of each solution over a 12- to 18-month window. Arrows indicate travel based on strategy and pace of innovation, with vendors designated as Forward Movers, Fast Movers, or Outperformers based on their rate of progression.

Note that the Radar excludes vendor market share as a metric. The focus is on forward-looking analysis that emphasizes the value of innovation and differentiation over incumbent market position.

5. Vendor Insights

CloudCasa by Catalogic

CloudCasa is a SaaS service that enables you to backup, restore, and secure Kubernetes-based applications. The SaaS service runs on Kubernetes inside AWS but also can be self-hosted on request. It's available in the public cloud application catalogs, as well as in Rancher's application catalog, which simplifies adoption. Its pricing model is freemium with capacity-based pricing. The free tier allows customers to back up Kubernetes resource metadata to CloudCasa and manage PV (CSI and EBS) and RDS snapshots with a retention period of 30 days. Customers pay if they want to enable immutable backup repositories, or want to store PV backup data in CloudCasa-managed backup repositories, which the company claims are offered "at cost" of the object storage provided. The free tier allows users to back up up to 100GB of PV snapshots to CloudCasa, and pricing starts at \$199 per month for 1TB of cloud storage. Restores from CloudCasa repositories are free, without additional egress costs.

CloudCasa creates a gRPC connection from an agent in its own namespace to its SaaS components from customer clusters and runs the CSI data processing as CRDs inside those customer clusters, deployable via kubectl, Helm charts, or as Operators.

Its key users are developers and DevOps engineers who are familiar with, but not experts on, data protection, making it easy to integrate data protection into their day-to-day workflows and eliminating the heavy lifting of deploying and managing the lifecycle of the data protection solution.

CloudCasa strongly focuses on the Kubernetes ecosystem, offering support for many Kubernetes-based platforms including Azure AKS, AWS EKS, GKE, and DigitalOcean, but also has support for OpenShift, Rancher, Tanzu, and other platforms. In addition, CloudCasa supports backing up services in the AWS ecosystem, including RDS and Aurora-based databases and EKS metadata, as well as EBS-based storage for snapshots.

CloudCasa supports bring-your-own backup repositories (which works with any S3-compatible target, including on-premises appliances, and they're working on a certification program for most common on-premises and public cloud object stores) but also has a flexible selection of Azure and AWS object storage locations, including in different regions of the world. In the future, it will also support bring-your-own-keys for any backup repository. For any repository, they support data-at-rest and in-flight encryption, as well as immutability (which they call SafeLock), preventing backups from being deleted or retention policies from being weakened by any cryptolocker attempts.

CloudCasa supports autodetection of CSI and EBS-based PVs. In addition, CloudCasa inventories AWS accounts to detect newly added EKS clusters. It allows backup of applications based on labels, namespaces, and policies.

CloudCasa is in an early access stage with some of its security features, including container, networking, configuration, and best practices benchmark scanning. These features will cover not just Kubernetes but also scan AWS services, detecting misconfigurations and security vulnerabilities across AWS IAM, EKS, KMS, S3, RDS, VPC, and more.

However, the "AWS well-architected" accreditation means its SaaS service meets AWS' strict security, efficiency, and reliability demands, which include multifactor authentication for the SaaS Web UI, suspicious IP throttling, fraud and brute force attack detection, and SOC2/ISO27001 compliance.

The product supports various databases using pre- and post-backup scripts for creating consistent backups, including quiescing or stunning MySQL, MongoDB, and PostgreSQL databases with pre-built templates. The company plans to add more templates for popular data services, including git-based source version control systems in the future.

Application migration support requires users to use backup and restore workflows to migrate applications or data services between platforms. This makes CloudCasa less suitable for lift-and-shift migrations, application transformation, or cloud migration use cases. It has added features recently to enable heterogeneous restores between different storage classes, allowing the mapping of different storage classes within the CSI and EBS realm, and enabling DR-to-the-cloud migration scenarios.

Finally, its web-based UI lacks RBAC and self-service capabilities for limiting access of certain individuals or groups to specific clusters or namespaces. Users can see all resources as they are added to their accounts, but they have no way of delegating access; nor is there a way to manage multiple users using the concept of an organization for delegation. CloudCasa plans to release initial RBAC support in 2022.

Strengths: CloudCasa is a free, very complete data protection solution with support for both Kubernetes clusters and AWS services like RDS. As a SaaS solution, it's easy to deploy and manage. It has flexible backup repository options ranging from CloudCasa-provided to bring-your-own (on-premises) repository. It has support for backing up cloud-native databases in AWS and has plans to deepen its integration with AWS, Azure, DigitalOcean, and GCP.

Challenges: CloudCasa is not an application migration solution, making it less suitable right now for application transformation projects (and cross-cluster or DR-to-the-cloud scenarios). Its limited support for developer self-service and RBAC makes it less suitable for enterprise environments with stricter compliance and security requirements. CloudCasa is focused on deepening its capabilities in AWS but lacks the same support for Azure and GCP, currently.

Commvault

Commvault Backup & Recovery is a backup solution that supports more than Kubernetes workloads, making it suitable for hybrid applications that run across Kubernetes, VMs, and cloud services, consolidating backup operations on a single platform.

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

instances, and Kubernetes-based containers. Additionally, there is support for backing up CI/CD pipelines including code repositories in Azure DevOps and GitHub. Its flexible deployment architecture and single interface across multiple deployments means the solution is multi-cluster and multi-cloud by default.

The solution is compatible with all CNCF-certified distributions and is integrated with CSI for snapshot-based backups. It can provide application consistent backups, thanks to the ability to use pre- and post-backup execution scripts to quiesce data on storage before taking the snapshots, minimizing risk of data loss. Many scripts for MySQL, Cassandra, MongoDB, and PostgreSQL are included. Furthermore, data and applications can be backed up and restored in an alternate Kubernetes environment, simplifying migrations and disaster recovery operations.

Its solutions are available across various cloud marketplaces. The Commvault solution works with a specialized access node (deployed as a VM) that interacts with the Kubernetes API server to discover and protect applications. Data movers are deployed on a cluster only during a backup, with no other parts of the backup application running on-cluster. An Operator is in the works.

The interface includes 200+ built-in reports and a custom reports builder, and shows infrastructure requiring attention and “at a glance” recovery readiness status across clusters.

There are self-service capabilities for developers through the Command Center UI, thanks to integration with the Kubernetes RBAC system. Applications are discovered automatically using label selectors or entire namespaces, covering new applications with a blanket backup policy.

Security and ransomware controls hint at Commvault’s long history in data protection. Its interface security is well taken care of with MFA support, data encryption (at-rest and in-flight), bring-your-own-keys and support for various Key Management Systems, air-gapped backups (optionally via Commvault’s Metallic Cloud Storage Service), backup target immutability, and deep support for anomaly detection for both operations in the UI and deletions/changes at the storage target level for malware detection. Backup storage targets include HyperScale appliances, Metallic Cloud Storage Service, and object stores (including S3-compatible on-premises, NAS/SAN, and tape), while functionality encompasses support for deduplication and compression.

Additionally, Commvault’s use of machine learning algorithms to optimize operational tasks is notable. For instance, these algorithms tune backup operations to keep backups within their set SLA automatically.

Disaster recovery has various options, including integration into Commvault’s storage system for (a)synchronous replication and snapshot-based disaster recovery across cloud regions or on-premises data centers. The UI includes mature workflows for heterogeneous restores (like cross-clusters, cross-cloud), remapping of resources, and dev/test pre-seeding.

In edge use cases, Commvault has a physical appliance-based solution, HyperScale X, that supports Azure Stack, AWS Outposts, and EKS Distro.

Commvault provides several purchasing options ranging from traditional perpetual licensing to subscriptions. Licenses are transferable between VMs and containers. Commvault Metallic, a SaaS solution, is aimed at small and medium size organizations and includes support for Kubernetes. It is integrated with HyperScale X, offering quicker on-premises restores using the physical appliance in disaster recovery scenarios.

Strengths: Its broad support for VMs, containers, data services, and cloud services in a single platform make it a great choice for hybrid and complex applications. Its security and ransomware controls are very extensive, making it suitable for larger enterprises.

Challenges: With flexibility comes complexity. Commvault’s solutions require time and attention to implement correctly, and they are not as easy to implement or maintain as other solutions.

Dell Technologies (PowerProtect Data Manager)

Dell PowerProtect is a container protection solution built on Velero, with appliance-based deduplication and compression using PowerProtect DD series appliances, which can be physical or virtual (running on top of cloud object storage).

The VM-based Manager deploys the data movers automatically on protected Kubernetes clusters. It also deploys as a VM in public cloud environments to protect cloud-based Kubernetes environments. There is no SaaS version available currently, but future releases will include an Operator-based deployment and will be available through various marketplaces.

The Dell solution can connect to multiple on-premises or cloud-based Kubernetes clusters from a single Manager.

PowerProtect uses Velero components to capture the Kubernetes-level objects, while it leverages its own data movers (running in their own namespace). Backups run at the namespace level. Restores include remapping of storage classes.

It supports various cloud-based Kubernetes services, including AKS, EKS, and GKE. In the on-premises world, it supports VMware Tanzu Guest Clusters, Tanzu Kubernetes Grid Integrated Edition, Rancher, OpenShift, and Diamanti.

PowerProtect includes at-rest and in-flight encryption, as well as target immutability with retention locks. Optionally, it can be integrated with Cyber Recovery Solution for air gapping and advanced machine learning and anomaly detection.

PowerProtect lacks mature RBAC and self-service capabilities, currently offering a CLI-based workflow for self-service users to restore applications using a custom YAML-based format, but improvements are planned for 2022.

PowerProtect Data Manager comes with built-in support for MySQL, PostgreSQL, and non-sharded MongoDB and Cassandra. Future releases will include support for shared databases. Currently, PowerProtect supports backing up only cloud-based databases like Amazon RDS via an additional product, Cloud Snapshot Manager. This functionality will be integrated in the future.

PowerProtect Data Manager includes workflows to support the migration of applications between clusters, including the underlying storage across disparate storage classes.

Strengths: PowerProtect protects more than just containers, making it a good solution for protecting Kubernetes-based applications when you’re already using PowerProtect. It offers efficient replication of data between DD appliances across on-premises environments or cloud providers for DR purposes.

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

Druva (Data Resiliency Cloud)

Druva recently announced a solution designed to protect Kubernetes workloads in Amazon AWS, in both EKS (managed) and EC2-based (self-managed) clusters. The solution is extremely easy to set up and use; it is application-aware and is able to protect data stored in RDS databases as well. Druva gives AWS users a solution capable of protecting complex applications that take advantage of both containers and the Amazon AWS ecosystem. At the same time, Druva can protect common databases such as MySQL, PostgreSQL, and MongoDB with the mechanisms necessary to ensure end-to-end application consistency.

The solution is simple, well-designed, and scalable, with the core Backup Controller module installed as an operator inside the cluster. Each individual backup job is instantiated separately for improved scalability and parallelism.

Druva allows backing up and recovering data in different AWS regions for disaster recovery and application mobility, and integration with Kubernetes RBAC lets application owners use the CLI to perform day-to-day operations in a self-service fashion.

Additional cloud services and on-premises Kubernetes distributions will be added later, allowing companies to protect and move applications and data across multiple environments.

Strengths: An easy-to-use and well-integrated solution for the Amazon AWS ecosystem with the potential to evolve into a credible multi-cloud solution. Data protection for DRS can simplify application protection dramatically in some circumstances.

Challenges: The lack of multi-cloud support is a showstopper for most Kubernetes scenarios, with applications deployed on-premises and in multiple clouds.

HYCU (Protégé for Kubernetes)

HYCU's solution for Kubernetes has matured since the previous report. Its specialized data protection modules for major public cloud providers and on-premises virtualized infrastructure still exist, but the SaaS-based Protégé has matured.

Storage integration is not CSI-based (nor are there any plans to support CSI in the short term), but HYCU leverages native cloud APIs (Google, Azure, AWS) and existing storage providers for snapshot functionality (Nutanix, VMware). No other Kubernetes distributions or services are supported.

Kubernetes applications are auto-discovered via the YAML metadata. The policy-based backups are assigned using Kubernetes labels. Policies include options to do snapshots, full backups, long-term retention, and storage tiering from a single policy. Additionally, a copy can be stored outside the local region for disaster recovery purposes.

Restores can be made to heterogeneous environments, including different clusters and different regions, with cloning functionality for test/dev. Workflows include mature remapping options for restores to disparate cluster configurations.

The web-based interface is simple and easy to use, but logins are cloud-provider specific; it re-uses the identity provider active in the cloud where the solution is deployed. Each instance of HYCU is limited to use in a single cloud (depending on where it's deployed) and is not natively multi-cloud. A "manager of managers" interface is available upon request for multi-cloud capabilities. The UI includes self-service capabilities based on the Kubernetes primitives.

HYCU does not support backing up cloud-native database and storage services, like RDS or S3. Backup targets include S3-compatible storage services; the on-premises product also supports existing SMB and NFS targets. Immutability features are supported only on S3. Data encryption in-flight and at-rest is enabled and includes bring-your-own-keys as an option.

HYCU has basic support for data management and application migration/transformation but requires re-use of existing backup and restore workflows.

The pricing model is based on allocated source capacity.

Strengths: Its integration into Nutanix and VMware-based environments make HYCU a strong option for those customers already running on these platforms. A promising roadmap shows the potential to make HYCU more broadly applicable for those on other platforms. The backup policies and restore workflows are very mature, supporting migrations, transformations, and disaster recovery alike.

Challenges: The product is immature in supporting other platforms, and its lack of CSI support will remain a challenge for the foreseeable future. Multi-cluster support is minimal but expected to mature quickly. The absence of RDS support will prove a disqualifier for some.

IBM (IBM Spectrum Protect Plus)

IBM Spectrum Protect Plus (SPP) recently added the ability to protect Kubernetes workloads. This solution is designed to protect all modern environments, including virtualized infrastructures, and can be used by organizations of all sizes, including service providers.

SPP provides a native Kubernetes CLI and leverages Kubernetes orchestration capabilities to allocate resources necessary for backup jobs. The product's current focus is on Red Hat OpenShift, but support for a broader range of Kubernetes distributions will be added soon. In this regard, SPP takes advantage of Red Hat OpenShift APIs for data protection (OADP) based on Velero open source technology, and IBM directly interfaces with Velero for other Kubernetes distributions, keeping the user experience consistent across distributions.

The SPP component for Kubernetes is deployed as a CRD operator. Administrators and developers can search the Kubernetes inventory easily and manually select the applications to protect. Comprehensive and automated discovery mechanisms will be added in future releases. Application consistency is not yet available, but IBM is working to implement mechanisms to ensure off-the-shelf application consistency for its IBM Cloud Paks, common databases, and custom applications.

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

clusters, with multi-cloud support planned for the end of the year.

Strengths: Strong focus on Red Hat OpenShift and IBM ecosystem (IBM CloudPaks), while providing protection for both traditional and modern workloads.

Challenges: SPP shows potential and a good roadmap for future product releases, but right now it presents several limitations for complex Kubernetes environments.

Kasten by Veeam (K10)

Kasten K10 is a cloud-native data management platform for Kubernetes. It includes backup, restore, disaster recovery, and migration functionality for container-based applications.

It's a self-hosted solution, designed to run on each cluster it protects. It has a minimal resource footprint on those clusters that is achieved by auto-scaling components like its data mover and instantiating parallel instances based on running jobs. It's installable via various cloud and Kubernetes marketplaces as a CRD-based install into its own namespace. Despite the lack of a SaaS version, Kasten piggybacks on Veeam's VCSP program, which allows cloud service providers to offer Kasten K10 as a service.

Kasten focuses on enterprise customers, and its key users are both backup admins as well as developers for self-service backups and restores. It is extremely easy to use, with a quality GUI, consistent APIs, and a handy CLI. Kasten K10 is designed to cope with large-scale environments, but it can be deployed in small infrastructures as well.

K10 supports a wide range of Kubernetes flavors, including OpenShift, Rancher, Tanzu, EKS, AKS, and GKE, across on-premises and cloud. It has special partnerships with K3s and AWS EKS Anywhere to support edge and retail deployments, including a newly announced partnership with AWS Containers Anywhere, aimed at on-premises edge deployments. K10 does not support backups of AWS resources like VPC or IAM configuration.

Kasten's business model is 100% channel-based, including the SaaS offerings by third-party providers. Kasten's license model is a node-based subscription. There is no capacity-based pricing. There is a free edition with a maximum of 10 nodes for smaller production, proof of concepts and labs.

Kasten K10 uses a policy-based approach to capture the dynamic nature of container-based applications, and backups include Kubernetes resources and metadata such as Secrets and ConfigMaps. K10 also natively discovers data services (like MySQL, MongoDB, PostgreSQL, Amazon RDS, Kafka, and Cassandra) as part of applications, and automatically applies the right data management policies (for things like quiescing) using Kanister, an open source data management framework initially developed by Kasten to create an industry standard for stateful data management. Additionally, K10 includes a blanket policy of protecting any new or previously unprotected applications automatically, and supports assigning backup policies based on labels and namespaces.

K10 natively supports CSI, including snapshots and backups for traditional enterprise storage arrays and Amazon EBS, VMware, and container-attached storage solutions, but has deeper CSI integration for a number of CSI providers offering additional functionality and backup performance. It includes features like changed-block tracking, fast incrementals, and transfers between repository regions. When it can use a more optimal underlying storage integration, such as OpenStack, CEPH, vSphere, or a public cloud's API, it will offload tasks directly instead of using CSI. Space and network efficiency are assured, thanks to deduplication and compression techniques.

Supported repositories include object stores, NFS shares, and even existing Veeam Backup & Replication backup repositories. For any repository, K10 supports data-at-rest and in-flight encryption, as well as immutability for S3, MinIO, Cloudian, and others, preventing backups from being deleted or retention policies being lowered in any cryptolocker attempts.

Backup data is always encrypted, both at-rest and in-flight. Encryption keys can be stored in the cluster or by using an external key management system like HashiCorp Vault or AWS KMS. Other security features include image vulnerability scans.

The Kasten interface supports managing multiple clusters from a single interface and supports RBAC and role/scope limitations for self-service access for non-admin users to specific resources like a single cluster or namespace. It uses Kubernetes Roles and ClusterRoles, replicating the already defined roles in the cluster. Kasten enables admins to standardize (global) policies to tenant clusters, allowing policies to be centrally managed but distributed across clusters. In addition, local cluster admins can define local policies. In the future, Veeam Backup admins will be able to see Kasten jobs and policies from within the Veeam Console.

The Application Transform Engine supports application data and metadata transformations, migrations, and mapping, allowing use cases ranging from simple storage class mappings to cross-cluster, cross-region and cross-AZ, cross-distribution, and cross-cloud migrations. It also includes disaster recovery functionality to protect against cluster and availability zone failures, as well as storage system failures in on-premises scenarios.

An embedded instance of Prometheus is included with Kasten K10, storing metrics about the operational state of backups and the system, as well as an embedded instance of Grafana with various included dashboards.

Strengths: Kasten is a Kubernetes-native, mature solution that's very suitable for self-hosted, self-managed use cases. Its architecture scales well and is especially well-suited for edge deployments. Its RBAC features and centrally managed policy model are well aligned with large enterprise and self-service requirements. Its application-aware data management framework, Kanister, is promising and quickly maturing. It has good support for on-premises repositories.

Challenges: Kasten's native support for cloud services is limited, not leveraging deep integrations into AWS, Azure, or GCP. Its lack of first-party SaaS may be off-putting to some customers and its per-node subscription licensing may not work in very dynamic environments. Backup-based disaster recovery does not provide the fastest RTOs and RPOs.

Portworx by Pure Storage (PX-Backup)

Portworx PX-Backup is a data protection solution for Kubernetes. It's compatible with any Kubernetes cluster on-premises and in the cloud, including OpenShift, Tanzu, EKS, AKS, and GKE.

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

PX-Backup is deployed as a Helm-based application in a Kubernetes cluster. It's also available in the AWS Marketplace, but there is no SaaS version. It's a multi-cluster solution, requiring a single PX-Backup deployment in an administration cluster, able to protect any Kubernetes cluster across on-premises and cloud. Additionally, PX Central, a manager-of-managers, is able to manage multiple PX-Backup instances centrally.

For each cluster it protects, it deploys specific Portworx-components, as well as Stork, which bridges the gap between the PX-Backup server and the administration cluster. Backups are policy-based, and applications are assigned to a policy via namespace and label selectors.

Additionally, PX-Backup supports any CSI-compatible storage to perform and speed up application-consistent backups, Amazon EBS volumes, Google Persistent Disks, and Azure Managed Disks. PX-Backup supports S3(-compatible), Azure Blob, and Google Cloud Storage as backup targets. In recent versions, it added support to copy cloud provider snapshots (like EBS snapshots) into an S3 bucket, including storing that snapshot in a different region. Security is a critical aspect of the product, with data encrypted at rest and in transit. However, it does not include ransomware functionality like immutability or anomaly detection.

PX-Backup supports pre and post hooks, and it has built-in support for Cassandra, Elasticsearch, Jenkins, MongoDB, MySQL, PostgreSQL, and RabbitMQ.

It's multi-tenant aware, with mature RBAC functionality. It integrates with the Kubernetes RBAC controls and thereby adheres to the scope limitations and permissions set in Kubernetes so that users can interact only with their own namespaces or applications.

Portworx has mature data migration and transformation capabilities, based on Stork, allowing for application and data migrations between clusters, regardless of whether these are on-premises or in the cloud. These features do, however, require Portworx Enterprise to be installed on both the source and target clusters in addition to PX-Backup. In situations without Portworx Enterprise, it does support backup-and-restore workflows to migrate data to a different, heterogeneous cluster.

While there is a free trial of PX-Backup, there is no free version. Licensing options include annual subscriptions and pay-as-you-go models based on \$0.20 per node, per hour pricing with a 1,000 node-hours minimum.

PX-DR is an add-on solution to Portworx Enterprise, specifically designed for disaster recovery, leaning on synchronous and asynchronous data replication; however, PX-Backup does not include these features.

Strengths: A solid contender in the data protection market, highly integrated into Portworx Enterprise, making it a good fit for those looking beyond data protection at data storage, data management, and advanced disaster recovery.

Challenges: Even though PX-Backup works without Portworx Enterprise, key features like DR using (a)synchronous replications and more advanced application migration and transformation require the storage layer of Portworx Enterprise. Support for cloud data services like Amazon RDS and immutability features are planned for future releases. A SaaS version is not yet available but is in the works.

Robin (Cloud Native Storage)

Robin Cloud Native Storage (CNS) is a storage solution aimed at persistent, stateful container applications. Included in the product are various data protection features, including synchronous and asynchronous storage replication, as well as snapshot and backup functionality.

It can be installed on major cloud providers and distributions alike, including OpenShift, Rancher, EKS, AKS, GKE, and more. Some cloud providers offer it as SaaS, and it is available via various application marketplaces, including Google and OpenShift. Notably, VMware Tanzu is not supported. Robin has friendly, consumption-based, per-node-hour pricing, with discounts for annual subscriptions, similar to those for public cloud offerings.

Some features, including the (a)synchronous replication, require the storage solution to be installed on both the source and the target clusters. While able to lower RPO and RTO, this requirement limits CNS' practical use to those applications unable to leverage application-level resilience offered by microservice design approaches, while delivering enough business value to justify the additional cost and complexity of running an always-on Kubernetes cluster plus the Robin components. Similarly, CNS does not protect any data not stored on its storage volumes, including RDS databases, limiting the applicability of its data protection features.

A key benefit of the storage layer, however, is Robin's full control over, and visibility of, the Kubernetes objects and underlying storage. The data protection features incorporate Kubernetes metadata and configuration natively to protect and replicate all Kubernetes objects and constructs. The storage layer is CSI-compliant, making it easy to quiesce storage volumes during backups and snapshots, and CNS includes scripts for popular data services like MongoDB, Cassandra, MSSQL, MySQL, Oracle, DB2, and PostgreSQL.

CNS supports snapshots (local and remote), backups to a separate repository (local or remote), synchronous replication between storage volumes (using a stretched-cluster setup across two Availability Zones), and multi-site asynchronous replication (with multi-cloud or replication across availability zones supported).

One area where Robin shines is migration scenarios between disparate environments, including migrations from on-premises to cloud, or cloud-to-cloud. The storage layer's replication features are a clear benefit for migrations, making it easy to migrate an entire application, including its storage, between on-premises and public cloud. The same is true for cloning applications for dev/test scenarios. These scenarios do require Robin to be installed on the target cluster. Similarly, backups can be restored only to Robin's storage volumes.

The GUI includes a monitoring dashboard and the tools necessary for fast troubleshooting, offering mature Kubernetes-based RBAC for self-service and multi-tenancy with multi-cluster features.

All backups are encrypted in-flight and at-rest at the application level, with support for external key management. Backup targets include S3-compatible object storage platforms as well as public cloud storage services such as Google Cloud Storage and Microsoft Azure Blob, as well as on-premises storage solutions via NFS.

CNS does not include many ransomware and security features. While immutable backups on WORM-compliant object storage are supported, the product does not include security scanning features.

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

Strengths: End-to-end solution with a friendly pricing model, well-integrated with Kubernetes. Multi-cluster and self-service capabilities fit well into large enterprises that consider the data protection features integrated into a storage solution acceptable. The (a)synchronous replication features add low RTO and RPO capabilities for those applications that need it. The storage layer adds flexibility to low-downtime migration scenarios. Easy-to-use GUI and CLI are very helpful for users who have limited experience with Kubernetes.

Challenges: Data protection is not a standalone product but part of the Robin storage solution, limiting its use of data protection features to applications running on Robin storage. Some of the features require CNS to run on the target cluster, thereby increasing cost, even though the pricing is competitive. Restores are limited to Robin-based volumes, and there are limited security features.

Trilio (TrilioVault for Kubernetes)

TrilioVault is a data protection and resilience platform for Kubernetes. Deployed as a Kubernetes CRD Operator, the solution allows you to back up data and applications on every supported platform and restore them locally or elsewhere for development, migration, or disaster recovery. Trilio can protect applications discovered via labels, namespaces, via Helm charts, or as Operators seamlessly. It boasts native support for Helm charts, which includes backing up its history and context, solving specific restore issues with Helm charts being overwritten by the default chart at restore time. Deployments can be air-gapped for dark or edge deployments.

Integration with tools like Prometheus and Grafana is possible for monitoring, and the product also takes advantage of Kubernetes RBAC, while it can provide self-service protection services for developers in their own namespace and applications. The UI generates predictive RPO and RTO metrics based on system performance.

Very well integrated into Red Hat OpenShift environments, it supports all certified Kubernetes distributions and cloud managed services as well, including Google GKE, Amazon EKS, Azure AKS, and Digital Ocean. Additionally, it supports SUSE Rancher, VMware Tanzu, and more. Furthermore, the solution is able to backup MongoDB, PostgreSQL, InfluxDB, MySQL, Redis, etcd, Cassandra, and AWS RDS-based databases.

Multi-cluster support is achieved by linking together multiple per-cluster deployments in the UI.

Backup targets can be S3-compatible object stores, Azure Blob, or NFS shared volumes, and the solution natively supports data compaction techniques for network traffic optimization. Backups are stored in the open QCOW2 format and use the open-source LUKS for data-at-rest and data-in-flight encryption on a per-backup file basis, not on a per-repository basis. It also applies immutability on a per-file basis but only on S3-compatible storage.

TrilioVault is built as an auto-scaling application, temporarily allocating the necessary resources for every backup job, ensuring scalability and performance for environments of all sizes. It spins up additional data mover pods as running backup jobs are started.

The Management Console includes workflows for disaster recovery plans and workflows to migrate applications to disparate clusters, while it also includes restore hooks for custom scripting during restores, transformations (like storage class mappings), and exclusions to granularly restore data. A data staging feature is in the works, which pre-seeds data continuously to one or more destinations to cut down on waiting time with large data sets in migration or copy scenarios.

Besides the enterprise subscription (licensed per worker node, vCPU, or cluster), the product is available for a free 30-day trial with an unlimited number of nodes. There is also a free Basic edition with a 10-node limit aimed at testing, small organizations, and developers.

TrilioVault can manage CSI snapshots to speed up backup operations and can run pre- and post-job scripts to synchronize data on disk to overcome current CSI limitations regarding consistency groups. This mechanism also can be used to prevent unwanted write operations from applications while taking the snapshot. Additional work to optimize and certify these processes for common applications is underway, and enhanced capabilities are expected with future product releases.

Strengths: Balanced solution with broad support of managed cloud solutions, distributions, major databases, and application platforms. Easy to install and manage with good application auto discovery capabilities to simplify operations. Heading in the right direction in terms of new features and roadmap.

Challenges: Proactive security measures like anomaly detection and security scanning are not yet available.

Veritas (NetBackup)

A well-known brand in the data protection industry is Veritas. In recent releases of NetBackup, it has included support for Kubernetes.

NetBackup's Kubernetes support is an extension of the regular NetBackup product, requiring a full NetBackup installation. Kubernetes-specific components are deployed to each cluster via an Operator, alongside Velero for CSI-functionality and metadata access. There is no SaaS version of NetBackup.

NetBackup currently supports backing up entire namespaces only, with support for more granular selection using labels coming up in a later release. There is no autodetection of applications currently.

Backup jobs are policy-based. While NetBackup supports much more than just containers, so far there is no logical construct to define an application across, say, an RDS database, a container, and a file share. NetBackup supports creating backups of Amazon RDS though, as part of its CloudPoint feature set, which is integrated into NetBackup and usable from the same UI as its Kubernetes data protection.

The product supports Red Hat OpenShift, Google Kubernetes Engine (GKE), and VMware Tanzu currently. The console supports protecting multiple clusters. However, heterogeneous restores are not yet supported, meaning there is no support for migrations across different distributions; for instance, from on-premise to the cloud. There are no built-in workflows for disaster recovery.

Backup targets currently include only S3-compatible storage. NetBackup supports many other targets for non-Kubernetes backups, and it aims to bring that support to Kubernetes data protection in 2022. Current support includes Object Lock functionality.

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

While there is no specific support for Kubernetes at the edge, Veritas does have a physical backup appliance aimed at edge use cases.

Strengths: NetBackup protects more than just containers, making it a good solution for protecting Kubernetes-based applications when you're already using NetBackup. It has wide support for cloud databases, virtual machines, and more. It has mature operational security measures, including role-based access control and MFA.

Challenges: NetBackup's Kubernetes support is not yet mature. Its roadmap shows promise, but needs more time to materialize.

VMware (Velero)

VMware acquired Heptio in 2018 to improve its overall Kubernetes strategy. Velero was one of the open-source projects part of that deal, and now it is the data protection layer integrated into Tanzu, VMware's Kubernetes portfolio.

Additionally, the Velero open source project is a core component used by other data protection vendors, including Veritas, Dell, and Red Hat, underpinning many of the commodity features of those solutions. In this context, VMware positions Velero more as a standard API-based framework to standardize backup operations in Kubernetes environments, both at the back end as well as at the front end, with APIs designed to simplify the interaction between different data protection solutions, orchestrators, management consoles, data movers and abstraction layers offered by third-party vendors.

From the VMware Tanzu perspective, Velero is tightly integrated within the Tanzu stack, and data protection operations can be managed through Tanzu Mission Control and take advantage of VMware's storage functionalities. Velero data protection is included in Tanzu and users have access to it at no additional cost.

Velero supports snapshot backups for major cloud providers and Kubernetes clusters hosted on vSphere, but support for CSI is still limited due to the lack of an internal data mover, an issue that will be solved in the near future. The product also misses several other enterprise features, including multi-cluster support, environmental awareness, strong encryption, self-service, data migrations, multi-region DR, and analytics functionalities. The roadmap shows that most of them will be added in 2022.

Strengths: Open source project sponsored by VMware with contributions from several other commercial data protection players. Compelling roadmap and new projects that aim to simplify, standardize, and automate data protection operations. Included and well integrated with VMware Tanzu.

Challenges: Velero, as a standalone backup solution, is currently missing several key features, like ransomware protection, multi-cluster support, environmental awareness, strong encryption, self-service, data migrations, multi-region DR, and analytics functionalities. These features will need to be developed further to make the solution fit for general-purpose use cases.

Zerto (Zerto for Kubernetes)

Zerto for Kubernetes is the cloud-native successor of Zerto's popular disaster recovery solution for virtual environments. Zerto for Kubernetes re-uses the same data replication technology, with the benefit of per-second journal-based recovery of persistent volumes.

The Helm-based installer deploys the application as a Stateful DaemonSet into a cluster. It supports VMware Tanzu, Red Hat OpenShift, and native Kubernetes. In the public cloud, Zerto supports Amazon EKS, Azure Kubernetes Service, Google Kubernetes Engine, and IBM Cloud Kubernetes Service. In addition to the per-cluster components, an instance of the Zerto for Kubernetes Manager must be deployed to either the cluster or a third site (like a separate VPC account). It has a minimal resource footprint on clusters by auto-scaling components and instantiating parallel instances based on running jobs. Licensing of Zerto for Kubernetes includes perpetual as well as subscription options.

Zerto's ZKM-PX components sit in the data path and intercept all I/O operations for replication. The product is designed to support local backups, remote disaster recovery, and long-term archival to S3, but it can be very effective for simplifying data and application migrations as well as the CI/CD process by adding tags to the Zerto journal during deployments of new application versions to capture that moment in time.

Zerto for Kubernetes does not include a UI but works via an extension of kubectl. It is limited to protecting the cluster it is deployed to. The Manager can manage multiple clusters. Access to the SaaS-based Zerto Analytics platform is included in the product, which provides some overview of multiple deployments.

Zerto for Kubernetes is able to discover applications and their resources to protect them as a whole, making both local and remote restores easy to perform. It currently supports backup selection through annotations. It backs up Deployments, StatefulSets, ConfigMaps, Secrets, PV Claims, CRDs, and Services, as well as the Persistent Volume data via the Zerto Replication Engine. It does not support any database-aware processing, nor does it support external databases like AWS RDS.

Zerto for Kubernetes is very efficient and replicates only those blocks necessary to the secondary replication pod, which can run in a separate cluster for DR purposes, or a local cluster for backup purposes. The journal-based approach enables a granularity of capturing checkpoints every couple of seconds and storing them for up to 30 days. Checkpoints can be copied to a long-term retention repository (such as an S3- or Azure Blob-compatible service) daily, weekly, and monthly, and include immutability features based on the underlying storage service.

Using the same journal-based approach, Zerto for Kubernetes can be used as an application migration tool with very short cut-over windows and includes failover and migration testing, deploying the application into a temporary new namespace, allowing verification. This is also the proposed workflow for creating temporary copies for testing and validation during application development processes.

Strengths: The remote replication capabilities of Zerto for Kubernetes enable users to achieve very low RTOs and RPOs. The product is easy to operate and offers good potential for enhancing data mobility and simplifying migration activities.

Challenges: Even though the core technology is rock solid for disaster recovery and application migration scenarios, the product is still rough around the edges for data protection needs.

6. Analyst's Take

This is a GigaOm Research Reprint: [Expires Dec 30, 2022](#)

models, but the technical and organizational complexity of enterprise applications put a wrench in the works.

With Kubernetes not natively capable of handling data protection of stateful applications, we see many vendors stepping in and offering a breadth of solutions to solve the data protection, and by extension, data management problem, ranging from table stakes backup and restore to more sophisticated multi-region disaster recovery and even fully-fledged data copy management solutions for test/dev environments, as well as data storage solutions.

No single solution is perfect for all use cases. That means that if you're looking for a data protection solution for Kubernetes, you need to assess your existing application and infrastructure situation, taking existing investments in data protection into account, considering what kind of data you need to protect across virtual machines, containers, cloud databases, and other cloud services, and then define what specific data protection features you're looking for. The best solution for you isn't the one that ticks the most or even all the boxes in our research in this radar; it's the one that ticks the **right** boxes, at the right price point.

7. About Enrico Signoretti

[Enrico Signoretti](#)

Enrico Signoretti has more than 25 years in technical product strategy and management roles. He has advised mid-market and large enterprises across numerous industries, and worked with a range of software companies from small ISVs to global providers.

Enrico is an internationally renowned expert on data storage—and a visionary, author, blogger, and speaker on the topic. He has tracked the evolution of the storage industry for years, as a Gigaom Research Analyst, an independent analyst, and as a contributor to the Register.

8. About GigaOm

GigaOm provides technical, operational, and business advice for IT's strategic digital enterprise and business initiatives. Enterprise business leaders, CIOs, and technology organizations partner with GigaOm for practical, actionable, strategic, and visionary advice for modernizing and transforming their business. GigaOm's advice empowers enterprises to successfully compete in an increasingly complicated business atmosphere that requires a solid understanding of constantly changing customer demands.

GigaOm works directly with enterprises both inside and outside of the IT organization to apply proven research and methodologies designed to avoid pitfalls and roadblocks while balancing risk and innovation. Research methodologies include but are not limited to adoption and benchmarking surveys, use cases, interviews, ROI/TCO, market landscapes, strategic trends, and technical benchmarks. Our analysts possess 20+ years of experience advising a spectrum of clients from early adopters to mainstream enterprises.

GigaOm's perspective is that of the unbiased enterprise practitioner. Through this perspective, GigaOm connects with engaged and loyal subscribers on a deep and meaningful level.

9. Copyright

© [Knowingly, Inc.](#) 2021 "GigaOm Radar for Kubernetes Data Protection" is a trademark of [Knowingly, Inc.](#) For permission to reproduce this report, please contact sales@gigaom.com.

[Data Protection for Kubernetes](#)



Stay on top of emerging trends impacting your industry with updates from our GigaOm Research Community.

Start Moving!