GIGAOM

MARKET RADAR

GigaOm Radar for Kubernetes Data Protection v1.0

ENRICO SIGNORETTI | NOV 12, 2020 - 11:03 AM CST

TOPIC: DATA PROTECTION FOR KUBERNETES

GIGAOM

GigaOm Radar for Kubernetes Data Protection

TABLE OF CONTENTS

1. Summary

Enterprises began developing next-generation container-based applications a while ago; now they are deploying them in production with Kubernetes. Deployment models are many and varied and depend on business and operational needs, from small on-premises clusters to multi-cloud scenarios. What they all have in common is the need for data protection.

Containers and Kubernetes are dramatically different from legacy technologies, such as virtual machines and hypervisors, and traditional data protection solutions aren't up to the task. To meet growing demand, many vendors are working to adapt existing products, while others are building new solutions from the ground up. Moreover, there is a general trend toward providing more than just data protection, to expand the scope of these products with data management features. The goal is to take advantage of the backup process and enable application and data mobility, improve security, and simplify DevOps processes with copy data management. In this context, and as explained more in detail in the Key Criteria companion report, users will find that more and more cloud-native data storage solutions are providing similar services, as shown in **Figure 1**.

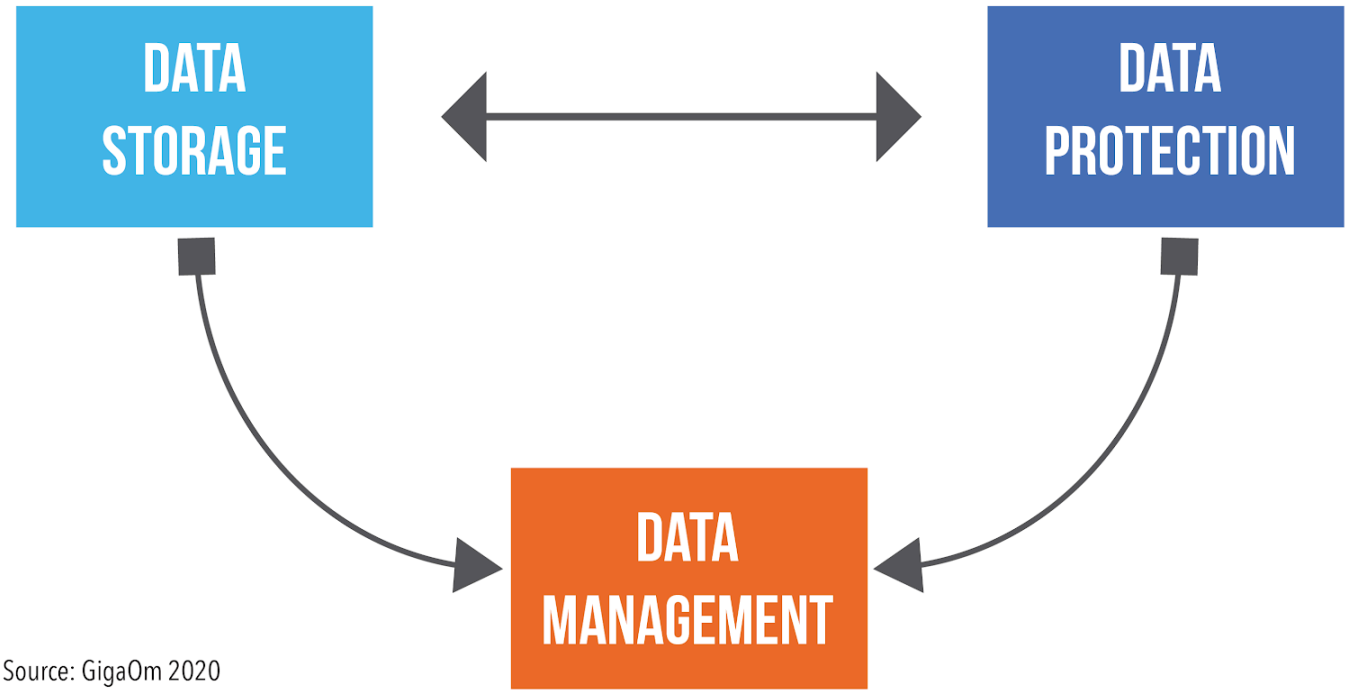

Figure 1. Cloud-Native Data Storage and Protection for Kubernetes Converge in Data Management Solutions

The market is burgeoning for Kubernetes data protection and for data storage in general. There are several new startups competing for technical leadership against established vendors that may have a more conservative approach, but whose offerings are familiar and aligned with the infrastructure in place.

With the fast pace of Kubernetes development and the multiple ways it can be deployed or consumed (on-premises and in the cloud), these solutions must be flexible, multi-platform, and readily adaptable.

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.

Vendor 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 able to serve the following market segments:

- **Small-to-medium enterprise:** In this category we evaluate solutions on their ability to meet the needs of small to medium-sized companies. We also look at departmental use cases in large enterprises, where ease of use and deployment are more important than extensive management capabilities, data mobility, and feature set.
- **Large enterprise:** For this category, offerings are assessed on their ability to support large, business-critical projects. Optimal solutions will have a strong focus on flexibility, performance, multi-cloud capabilities, and features to improve security and operations in large-scale environments. Scalability is another big differentiator, as is the ability to deploy the same service in different environments.

Data protection solutions for Kubernetes can also be categorized according to their architecture:

- **Traditional solutions that support Kubernetes:** These include solutions with traditional backup architectures that offer data protection services for Kubernetes and other platforms.
- **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

In addition, we recognize two deployment models for solutions in this report: SaaS (cloud-only) or hybrid and multi-cloud:

- **SaaS:** Available only in the cloud and as a 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.
- **Hybrid and multi-cloud:** These solutions are meant to be installed both on-premises and in the cloud. Integration with a single cloud provider might be limited compared with the other option, and more complex to deploy and manage. But this approach is more flexible, and typically provides more control over the entire stack with regard to resource allocation and tuning. These solutions can be deployed in the form of Kubernetes operators or specialized pods in the cluster.

Table 1: Vendor Positioning

	MARKET SEGMENT		ARCHITECTURE			DEPLOYMENT MODEL	
	Small-Medium Enterprise	Large Enterprise	Traditional	Cloud-Native Storage	Cloud-Native Data Protection	SaaS	Hybrid and Multi-cloud
Arrikto	+++	++	-	+++	++	-	+++
Commvault	+	+++	+++	+	++	+	+++
Druva	+++	+++	-	-	+++	+++	-
HYCU	+++	+	+++	-	+	+	+++
IBM	+++	++	+++	-	+	-	+++
Kasten by Veeam	+	+++	-	-	+++	-	+++
Maya Data	++	+++	-	+++	+++	++	+++
Portworx	++	+++	-	+++	+++	-	+++
Robin	++	+++	-	+++	-	-	+++
Trilio	++	+++	-	-	+++	-	+++
Velero	++	++	-	-	+++	-	+++
Zerto	++	+++	-	-	+++	-	+++

+++ Strong focus and perfect fit of the solution
++ The solution is good in this area, but there is still room for improvement
+ The solution has limitations and a narrow set of use cases
- Not applicable or absent.

Source: GigaOm 2020

3. Key Criteria Comparison

Building on the findings from the GigaOm report, "Key Criteria for Evaluating Kubernetes Data Protection," **Table 2** and **Table 3** summarize how each vendor included in this research performs in the areas that we consider differentiating and critical in this sector. 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 2. Key Criteria Comparison

	KEY CRITERIA				
	Multi-Cloud	Environmental Awareness	Disaster Recovery	App and Data Migration	System Management
Arrikto	+++	+	++	+++	+++
Commvault	+++	+	++	++	++
Druva	-	+++	++	+	+++
HYCU	++	-	+	+	+
IBM	++	+	++	++	+
Kasten by Veeam	+++	+++	++	+++	+++
MayaData	+++	++	++	++	++
Portworx	+++	+++	+++	+++	++
Robin	+++	+++	+++	+++	+++
Trilio	+++	++	+++	++	+++
Velero	+++	++	++	++	+
Zerto	+++	++	+++	+++	++

+++ Strong focus and perfect fit of the solution
++ The solution is good in this area, but there is still room for improvement
+ The solution has limitations and a narrow set of use cases
- Not applicable or absent.

Source: GigaOm 2020

Multi-cloud capability is one of the most requested features, and all vendors are working to satisfy users in this regard. At the same time, the most complete solutions focus as well on disaster recovery and application migration and mobility. We found Robin, Kasten, Portworx, and Trilio to be the most advanced solutions at the moment, with a series of runners-up that are working to close the gap or concentrating on specific functionalities. In this context, it is worth mentioning Arrikto, which has a strong focus on data management and collaboration, thanks to a brilliant mechanism for handling copy data management.

Table 3. Evaluation Metrics Comparison

	EVALUATION METRICS				
	Flexibility	Scalability	Performance	Ease of Use	TCO and ROI
Arrikto	++	+++	+++	+++	++
Commvault	++	++	++	++	++
Druva	+	+++	+++	+++	+
HYCU	+	+	++	+	+
IBM	+	++	++	++	++
Kasten by Veeam	+++	+++	+++	+++	+++
MayaData	++	+++	+++	++	++
Portworx	+++	+++	+++	+++	+++
Robin	++	+++	+++	+++	++
Trilio	+++	+++	+++	++	+++
Velero	+++	+++	+++	++	++
Zerto	++	+++	+++	++	++

+++ Strong focus and perfect fit of the solution
++ The solution is good in this area, but there is still room for improvement
+ The solution has limitations and a narrow set of use cases
- Not applicable or absent.

Source: GigaOm 2020

Cloud-native solutions usually show higher scores because they are designed specifically for the Kubernetes environment. They usually present more advanced features, better application discovery functionality, and more robust data management capabilities, resulting in better operation scalability, TCO, and ROI. Performance is not an issue across the board and most of the architectures are converging toward similar approaches. Traditional solutions working on Kubernetes data protection show potential, even as they suffer the most while adapting their backup models to container-based applications.

By combining the information provided in **Table 2** and **Table 3**, readers should be able to get a clear idea of the market and the available technical solutions.

4. GigaOm Radar

This report synthesizes the analysis of key criteria and their impact on evaluation metrics to inform the GigaOm Radar graphic in **Figure 1**. The resulting chart is a forward-looking perspective on all the vendors in this report, based on their products' technical capabilities and feature sets.

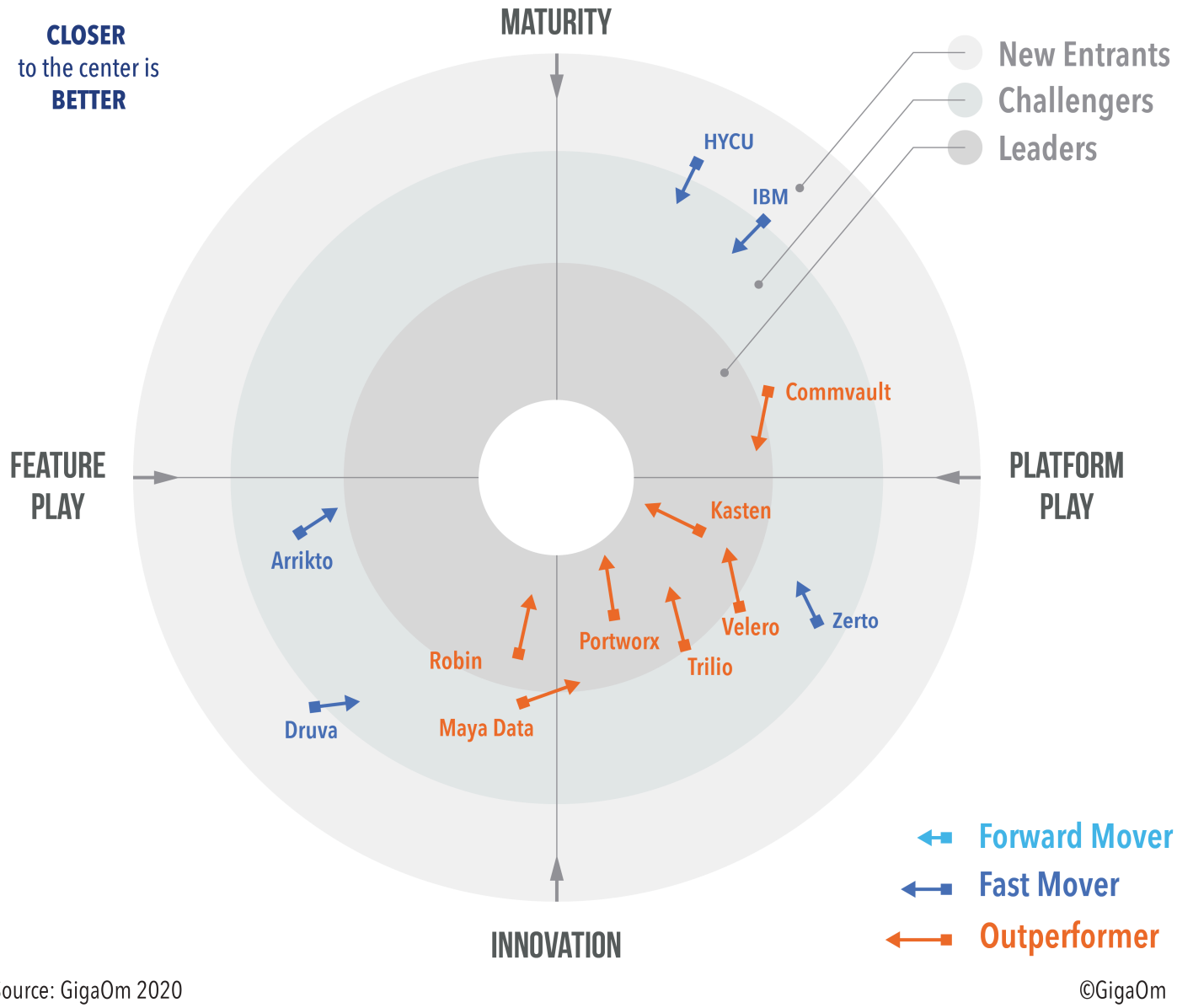

Figure 2: GigaOm Radar for Kubernetes Data Protection

The GigaOm Radar plots vendor solutions across a series of concentric rings, with those set closer to center judged to be of higher overall value. The chart characterizes each vendor on two axes—Maturity versus Innovation and Feature Play versus Platform Play—while an arrow projects each solution's evolution over the coming 12 to 18 months.

As you can see in **Figure 1**, 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 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 active and all the vendors are striving to build a consistent experience across multiple clouds while providing advanced application and data mobility.

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 **Platform Play** displays a broader platform focus and commitment to a comprehensive feature set.

The closer to center a solution sits, the better its execution and value, with top performers occupying the inner Leaders circle. The centermost 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

Arrikto

Arrikto Rok is an innovative data management solution for Kubernetes that includes data storage and protection functionality. Rok is designed to provide a CSI-compatible flash memory storage layer that can be installed on premises or in the cloud. It is optimized to work with NVMe devices and is well integrated with object storage to offload clones and snapshots. Its counterpart, Rok Registry, is a SaaS-based solution that enables users to manage application and data clones for disaster recovery, migrations, collaboration, and copy data management tasks from an easy-to-use web UI, APIs or a CLI. The two products work together to provide a consistent experience across different environments, with optimizations designed to minimize the impact on network traffic and speed up data recovery.

Arrikto Rok does not provide application-consistent snapshots yet, but can be easily integrated with third-party application frameworks to simplify operations and provide crash-consistent backups. In this regard, the integration with KubeFlow is remarkable.

The solution is installed as a Kubernetes operator and users can work directly with their data and applications in a self-service fashion. Arrikto Rok also supports synchronous data replication alongside a snapshot-based remote replication capability that takes advantage of object storage. That data replication also can be automated in order to provide a copy of the application ready to start in a remote Kubernetes environment.

Strengths: Innovative, fast, and, scalable solution focused on data management, collaboration, and self-service operations. Well-integrated with Kubernetes and easy to use.

Challenges: Arrikto solutions are very innovative, but this could be an issue for traditional IT organizations that are not ready for the data-centric approach proposed by this startup. The product needs improvements around restoring granularity and application awareness for application-consistent backups.

Commvault

Commvault added support for Kubernetes in its flagship product, Backup & Recovery, very early in 2017. Even though Backup & Recovery is not specifically designed for Kubernetes, the product and roadmap show potential and users will be able to operate in a familiar environment while consolidating backup operations on a single platform. Currently, the Commvault solution works with a specialized access node (deployable in a VM), that interacts with the Kubernetes API server to discover and protect applications. A CRD operator offers visibility and control over applications and resources.

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. Furthermore, data and applications can be backed up and restored in an alternate Kubernetes environment, simplifying migrations and disaster recovery operations.

A complete API set is available to automate backup activities even further. This solution offers self-service capabilities for developers through the Command Center UI thanks to the integrations with Kubernetes RBAC system. The web UI is consistent with the rest of the product and easy to use. Applications are automatically discovered using labels, with additional discovery mechanisms that will be added in future releases.

Commvault provides several purchasing options ranging from traditional perpetual licensing to subscriptions. Commvault Metallic, a SaaS data protection solution aimed at hybrid cloud deployments for organizations ranging from small businesses to the enterprise, now includes support for Kubernetes as well.

Strengths: An easy-to-use and familiar environment to operate, helping the user to consolidate all backup operations on a single platform

Challenges: Some aspects of the product are still immature, but the roadmap shows that product development is heading in the right direction.

Druva

Druva recently announced a solution 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 necessary mechanisms 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 dramatically

simplify application protection 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

HYCU aims to simplify multi-cloud data protection, but it is still in a transition phase with regard to Kubernetes support. The product provides specialized data protection modules for major public cloud providers and on-premises virtualized infrastructures. The solution includes a SaaS-based management layer (Protégé) to coordinate, consolidate, and simplify operations across different environments. A similar approach is now applied to Kubernetes. In fact, currently supported platforms with a Kubernetes option can be protected in the same way as VMs and other resources are.

The solution lets you group container pods using basic search operations on Kubernetes tags, and the system can back up all resources necessary to restore the application locally or in a different environment. The user interface is simple and easy to use, but most operations still need to be managed manually and there are no application auto-discovery mechanisms.

HYCU doesn't provide a way to manage application-consistent backups for common databases yet, and Kubernetes support is limited to Kubernetes services available from the service providers supported, VMware and Nutanix. Data management is not yet mature and HYCU is still working on advanced features to improve copy data management and data migration capabilities. Most of these issues should be solved with upcoming product releases in the first quarter of 2021.

Strengths: A promising roadmap that adheres perfectly to HYCU's philosophy of building easy-to-use-and-optimize solutions for every environment while providing plenty of options for users.

Challenges: The product is still immature and it is not clear how platforms like Red Hat OpenShift or virtualized Kubernetes clusters will be supported.

IBM

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 soon added. 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 easily search the Kubernetes inventory 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.

IBM SPP boasts several licensing options, including perpetual licenses and subscriptions per terabyte of persistent storage under protection. The product is already available in the AWS and Red Hat marketplaces and will be on hand soon in other public clouds. At the moment, SSP supports only on-premises clusters, with multi-cloud support planned for the end of the year.

Strengths: Strong focus on Red Hat OpenShift and IBM ecosystem (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

Kasten K10, recently acquired by Veeam, is a cloud-native data protection and management platform that describes itself as “purpose-built for Kubernetes.” 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. It supports all certified Kubernetes distributions.

The deep integration with Kubernetes allows the product to have strong awareness of the protected applications and all resources involved. K10 is very efficient and allocates the necessary resources to perform backups and restores only for the time necessary to complete the job, while protecting applications and databases with native methods when possible.

Space and network efficiency are assured, thanks to deduplication and compression techniques. Disaster recovery and data migration features, based on the same core technology used for backup, are easy to configure and manage and provide full control over resources and application configuration on the target site. Fine-grained RBAC capabilities enable developers to take direct control of copy data management needs in a self-service fashion, making it possible to perform a partial data recovery, for example, and to minimize the data footprint for development and test purposes. User-defined policies are seamlessly ported across different environments, helping to simplify data and application mobility while keeping data protection policies consistent.

K10 also has a strong focus on security, monitoring, auditing, and authentication, due to its support for several identity management services, good encryption capabilities, and integration with Prometheus.

Kasten's business model is 100% channel-based and software subscriptions are available on a per worker-node basis.

Strengths: Designed for and well-integrated with Kubernetes, Kasten K10 is easy to use, scalable, and efficient for large-scale and multi-cloud infrastructures. Promising data management features.

Challenges: Backup-based disaster recovery does not provide the fastest RTOs and RPOs. Backup targets limited to S3-compatible object stores, with NFS support currently only available in preview. Multi-cluster system management is similarly still in preview, though general availability is expected soon.

MayaData

MayaData is building a compelling enterprise platform based on its open source projects that has been well-received by the community and backed by technical and traditional investors. The solution, Kubera, is a data storage and performance management solution that also includes data protection functionality aligned with the expectations of enterprise customers. The Kubera Protect core is based on Velero, another leading open source solution designed for Kubernetes data protection. It is worth mentioning that MayaData is among the leading contributors to the Velero project.

Kubera Protect brings together the solidity of Velero with a modern user interface and additional tools aimed at addressing GitOps needs. Kubera inherits most Velero functionality and can take advantage of the underlying storage to take CSI-based snapshots, perform manual and scheduled backups, and specify backup hooks to perform pre- and post-backup operations, putting components such as a database in a consistent state before the backup process begins. The same functionality can also be used for application and data mobility across different environments, including migrations. Unfortunately, the data protection component, Kubera Protect, is still tightly coupled with the data storage component, but will be released as a separate product at the beginning of 2021.

Kubera has an application-centric approach and can be used to back up data for disaster recovery purposes or migration to different on-premises and cloud environments.

Strengths: Strong open source commitment, compelling vision, and interesting roadmap. Kubera addresses several challenges of complex and multi-cloud Kubernetes environments.

Challenges: The product is still in an evolutive stage and some of its components need to be unbundled to provide additional flexibility and ease of adoption.

Portworx

Portworx, recently acquired by Pure Storage, is a technology leader in data storage for Kubernetes and its solution also includes data protection and disaster recovery options. The solution is designed to provide a scalable and consistent enterprise-grade data storage management layer across on-premises and cloud environments. Portworx boasts a rich and expanding ecosystem of technology partners and certified solutions. It focuses primarily on large enterprise and ISP/MSP markets, but the pay-as-you-go subscription model and a limited free version of the product contribute to ease adoption

for SMEs and small development teams. Furthermore, the backup solution can be acquired as a standalone product and it works with Kubernetes clusters both in the cloud and on-premises.

Portworx PX-Backup can take advantage of any CSI-compatible storage to perform and speed-up application consistent backups, leveraging a series of integrations and automation procedures. Every backup includes all necessary data and metadata to retrieve the application, or part of it, locally or in another environment. Additionally, PX-Migrate, a solution for workload migrations, helps manage and automate migration activities between different Kubernetes environments. Finally, PX-DR, a solution specifically designed for disaster recovery, dramatically reduces RPOs and RTOs, thanks to synchronous and asynchronous remote data replication.

The user interface is modern and well designed, easing the work of sysadmins who are not yet familiar with the Kubernetes CLI and APIs. The product can be configured to provide self-service backup services to developers because of its integration with the Kubernetes RBAC system. Security is a critical aspect of the product, with data encrypted at rest and in transit.

Strengths: Highly scalable and flexible solution for Kubernetes data storage and management that works seamlessly across different on-premises and cloud environments.

Challenges: Even though PX-DR is a unique capability at the moment, it is only available for PX-Store users and it lacks a user interface and monitoring dashboard.

Robin

Robin is an end-to-end solution boasting integrated application management, data storage, and data protection for Kubernetes environments for organizations of all sizes. It is based on a Kubernetes-native and API-first approach and supports all CNCF-certified Kubernetes distributions, including Red Hat OpenShift and major public cloud provider services. Certification for VMware Tanzu is in progress.

Deployed as a standard Kubernetes operator, the solution is well-designed, with a good user interface for novices. It offers application-consistent, snapshot-based backup and disaster recovery, along with the ability to clone data and applications for development and test purposes. The GUI includes a monitoring dashboard and the tools necessary for fast troubleshooting.

Many operations can be scheduled and automated to simplify the day-to-day work of system administrators. Support for Kubernetes RBAC gives developers additional flexibility and control to protect and manage their applications. All backups are encrypted but the product doesn't support external key management yet. Backup targets include S3-compatible object storage platforms as well as public cloud storage services such as Google Cloud Storage and Microsoft Azure Blob. Scalability and performance are other key aspects of the product, with users now running the product on clusters of all sizes and in different scenarios.

Robin has friendly, consumption-based per-node-hour pricing, with discounts for annual subscriptions,

similar to public cloud offerings. The product is already available in the Google and Red Hat marketplaces, with others planned for the future.

Strengths: End-to-end solution with a friendly pricing model, well-integrated with Kubernetes. Easy-to-use GUI and CLI are very helpful for users with limited experience with Kubernetes.

Challenges: Even though the pricing is competitive, data protection is not a standalone product but part of the Robin storage solution.

Trilio

TrilioVault for Kubernetes is a new native data protection solution that adheres to the same key principles as other Trilio products for the OpenStack and Red Hat virtualized infrastructures. Very well integrated in Red Hat OpenShift environments, it supports all certified Kubernetes distributions and cloud managed services as well, including Google GKE, Amazon EKS, and Azure AKS. Backup targets can be either object stores or NFS shared volumes, and the solution natively supports data compaction techniques for network-traffic optimization. However, TrilioVault for Kubernetes temporarily allocates the necessary resources for every backup job, ensuring scalability and performance for environments of all sizes. Besides the standard enterprise subscriptions, 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.

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 seamlessly protect applications discovered via simple labels, via helm charts, or as operators. Integration with tools like Prometheus and Grafana is possible for monitoring, and the product also takes advantage of Kubernetes RBAC, and can provide self-service protection services for developers in their own namespace and applications.

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: Scalable and well-designed Kubernetes-native solution. Integration with Red Hat OpenShift UI is well-implemented, further simplifying operations within this environment.

Challenges: Encryption at rest is not yet available.

Velero

Velero is a popular open-source solution that provides a core against which providers can build strong data protection layers in their products. Backed by VMware, the technology is being integrated into solutions from VMware, like Tanzu Mission Control, and other providers, including Red Hat and MayaData. Velero boasts a growing and active development community that leverages its open architecture to contribute additional code, integrations, and plug-ins.

Velero has been designed from the ground up for Kubernetes environments and can be deployed both in the cloud and on-premises as a CRD. It uses object storage as primary backup targets and source of truth to keep the environment synchronized. The product allows users to run backup jobs manually, via APIs or by setting up a recurring schedule. The backup operations can be preceded and followed by scripts to put applications in a consistent state before taking snapshots and finalizing the backup. Velero offers several methods to identify applications and objects to protect, and integration with the Kubernetes RBAC system enables application owners to backup and restore their environments in a self-service fashion.

Beyond data protection, common use cases for Velero include disaster recovery and application mobility. In fact, backups are self-consistent and can be retrieved from other Kubernetes environments in case of disaster or to create new development and testing environments.

Strengths: Open source project with a large development community and backed by VMware. Well documented, easy to use, scalable, and very flexible. The plug-in system allows third-party storage vendors to integrate the product easily with their solution.

Challenges: User interface is limited to a CLI and APIs. Security and end-to-end encryption management are aspects of the product that need some improvements to make the solution fit in demanding environments.

Zerto

Zerto for Kubernetes employs the Zerto Replication Engine (ZRE), the same data replication technology found in the company's flagship solution for virtualized environments. Deployed as a stateful daemonset in the worker nodes via a YAML file manifest with both a Helm Charts installer and a Kubernetes Operator in the works, Zerto sits in the data path and intercepts all IO operations to replicate them to a remote site or an S3 bucket (support for NFS is not yet available). The product is designed to support both local backups and remote disaster recovery, but it can be very effective for simplifying data and application migrations as well as the CI/CD process. The latter is not fully exploitable yet because there’s no RBAC support in this version. Zerto for Kubernetes is still in an early-access phase, but it will support all common Kubernetes distributions when it becomes generally available.

There’s no UI yet, but you can operate the product via a CLI or APIs, with an additional SaaS-based

analytics dashboard (Zerto Analytics) available as an included option.

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. Unfortunately, the product still isn't able to protect some elements in the etcd database and in other namespaces—issues that should be resolved in future releases.

Zerto for Kubernetes is very efficient and replicates only necessary blocks to the secondary site. Furthermore, the ZRE core is a mature, optimized component that is deployed on every worker node and scales with the size of the cluster. These characteristics help make the product scalable and efficient, consuming very few resources in the cluster nodes. Licensing of Zerto for Kubernetes will include perpetual as well as subscription options.

Strengths: The remote replication capabilities of Zerto for Kubernetes enables 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, the product is still rough around the edges and needs to be improved before reaching general availability status.

6. Analyst's Take

Data storage and protection for Kubernetes is a very hot topic for practically every organization with production environments based on Kubernetes. Many users tried to adapt existing technology and solutions to Kubernetes, but failed because of the complexity of the applications and the lack of necessary tools to help them understand the context in which they operate. The sheer number of operations requested by these applications and the increase in stateful applications in production require a different approach.

Even more, data management is becoming a pressing theme. More sophisticated users want to be able to replicate applications remotely, and create copies and clones for different purposes while giving the application owners and developers the necessary self-service tools to simplify and speed up operations.

In this context, a number of solutions are doing very well, including Kasten, Portworx, Robin, and Trilio. Note that Velero, the open source project now backed by VMware, is at the base of several solutions and is growing in importance in the Kubernetes ecosystem. Some, like Arrikto, chose a radically different path with a set of features that are particularly interesting for highly collaborative environments. MayaData, with Kubera, also deserves to be mentioned for its vision and the innovative solution stack it is building. On the other side of the spectrum, if we take a look at established vendors, Commvault is doing a good job of protecting traditional and Kubernetes-native workloads, with a good combination of functionality that can protect multiple types of environments and hybrid applications concurrently.

Other solutions already available in the market are evolving quickly. Most of their positioning is about maturity, not overall architecture or implementation, but it will take several months before we see whether they will be able to execute quickly and correctly on their roadmaps.

7. About Enrico Signoretti

Enrico has 25+ years of industry experience in technical product strategy and management roles. He has advised mid-market and large enterprises across numerous industries and software companies ranging from small ISVs to large providers.

Enrico is an internationally renowned visionary author, blogger, and speaker on the topic of data storage. He has tracked the changes in the storage industry as a Gigaom Research Analyst, Independent Analyst and 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