GIGAOM

REPORT

Understanding and Managing Data Protection for Kubernetes in Enterprise Environments v1.0

The End-to-End Approach of Red Hat OpenShift

ENRICO SIGNORETTI | JAN 25, 2021 - 1:25 PM CST

TOPIC: DATA PROTECTION FOR KUBERNETES

GIGAOM

Understanding and Managing Data Protection for Kubernetes in Enterprise Environments

The End-to-End Approach of Red Hat OpenShift

TABLE OF CONTENTS

1 Summary

2 The Challenge

3 The Solution

4 Red Hat OpenShift and Data Protection

5 Conclusion

6 About Enrico Signoretti

7 About GigaOm

8 Copyright

1. Summary

Over the last year we have moved past the initial enthusiasm and adoption phases of Kubernetes. Enterprises are moving their applications to production, and that means that every single application now has the same characteristics of any other enterprise application, including:

- **Service Level Agreement (SLA):** The business doesn't really know how or where the application is deployed, but that doesn't change the fact that the IT team has to commit to respecting the SLA for that application, especially when it is critical for the business. This includes assuring the right recovery time objectives (RTO) and recovery point objectives (RPO) for business continuity.
- **User expectations to satisfy:** Even when this is not part of the SLA, user experience is utterly important. Any service disruption or data loss not only damages the business but also degrades the user's trust in the service. Long periods of downtime due to data retrievals, migrations, or day-to-day operations are likewise unacceptable.
- **Data protection and management:** Above everything else, data is the most important asset for every organization and it must be properly protected. In a dynamic environment such as the one provided by Kubernetes and containers, the ability to make and manage copies of data and applications quickly is fundamental to supporting the speed of change required by modern software development methodologies.Compared to what was envisaged a few years ago, if we look at the reality of Kubernetes, there are very few stateless applications in enterprise environments today. Stateful applications are the standard, and there are several reasons for that.

Table 1: Data Protection and Management Scenarios for Kubernetes

Scenario	Context	Need
"Lift and shift" and application refactoring	While not a best practice, it is common in enterprise environments to start adoption of new technologies with lift and shift migrations to standardize quickly on a new platform and then start the refactoring process later, depending on budget and business requirements.	This approach and the implied evolving scenario bring several challenges to application and data protection that can't be managed in a traditional way. Some applications will become hybrid, mixing modern and legacy technology, meaning that data protection is even more challenging and needs to be orchestrated correctly.
Application and data mobility	To provide application portability and mobility across on-premises and cloud infrastructures, developers can't afford to build applications that rely on external data services that may not be available on other platforms (for example, Amazon AWS RDS is not available in other clouds).	In this context, it is important to keep the application and its data together and manage them as a whole. This includes the ability to understand all the individual components and take all the necessary actions to ensure data consistency through the entire process.
Data ownership and governance	Most organizations are not ready to separate data from application ownership and manage them discreetly. Having data and applications together simplifies data governance processes and data protection, making it easier to identify application owners, manage tenants, and audit to enable quick action.	Multi-tenant data protection and self-service become key factors in this scenario. Security, of course, is a priority. While we always want to build for flexibility, we need all the necessary mechanisms to identify threats, limit risks of data theft, and create air gaps to mitigate attack.

At the end of the day, the adoption of Kubernetes in production environments creates new organizational and operational challenges around data that can't be solved by traditional data protection methods. From this point of view, a new innovative approach is necessary. It has to bridge the gap between modern applications deployed on Kubernetes and traditional enterprise data storage practices and processes.

Red Hat, a leader in enterprise Kubernetes thanks to Red Hat OpenShift, has developed the right set of tools to face these challenges and respond to the most demanding business needs related to data protection for Kubernetes environments. Red Hat's end-to-end approach to Kubernetes and data

management enables users to protect data across multiple environments while providing a consistent set of enterprise-grade data services to accelerate data mobility, migrations, and disaster recovery.

2. The Challenge

In traditional environments, data protection methods are focused on protecting physical or virtual machines (VM), their OS, and the application installed in them. The VM is the atomic unit, and the entire environment is very static. Most applications are deployed in one monolithic VM or a few of them, separating the application at the macroscopic level (for instance, a three-tier application with web front-end, application server, and database back-end). For Kubernetes, it is the exact opposite. Applications are organized as microservices, deployed in containers, with new container instances spun up and down continuously, depending on the needs of the moment.

That said, the concept of "cattle vs. pets" can't be applied to the enterprise. Data must be consistently protected whether the applications are installed in the cloud or on-premises, and data and applications must live together in the same environment. In fact, not all containers are the same. Some are stateless while others have data volumes attached to them; hence they become stateful. What's more, some applications will remain hybrid with a mix of microservices interacting with VMs.

Table 2: Applications and Data Volumes: Traditional vs. Kubernetes

	Traditional	**Kubernetes**
Application layout	One or few VMs	Many containers
Rate of change	Static	Very dynamic
Application type	Stateful	Stateless + stateful data volumes
Application status	Incorporated in the VM and its data store	Managed by Kubernetes

With this premise, it is clear that protecting Kubernetes applications with standard methods is impractical and not scalable.

- **Automation** is impossible and the risk of getting a partial view of what to protect is high. Traditional backup focuses on protecting small static groups of VMs. VMs include the application, its data, and its status. On the contrary, in a Kubernetes environment, the components to protect change over time depending on the status of the application, which is externally managed by the orchestrator. Traditional data protection solutions are not designed to operate efficiently with a continuously evolving environment like this and many of them rely on a series of manual operations to discover all the application components.
- **Data restore** is another big challenge, especially when the target environment is different from the source. Without having a full picture of the source environment, application status, and initial requirements, it is quite difficult to restore the application correctly, especially when the target Kubernetes cluster is different from the source. In fact, minor configuration parameter changes may be necessary during the restore process to ensure the correct application behavior, another aspect that is not covered by traditional data protection solutions.The last point is particularly important. One of the main benefits of Kubernetes is a level of flexibility that wasn't possible before, especially

in terms of application management and portability. Users want to take advantage of Kubernetes-native data protection to overcome some of the complexity of Kubernetes migrations and risks associated with it. By having access to tools designed to perform cross-cluster backup and recovery operations, the user can dramatically speed up testing and migration operations as well as implement disaster recovery procedures with minimal impact on production environments. To take advantage of this flexibility, it is necessary to adopt the right data protection tools, a solution that can also be of help in managing multiple data copies created for development and testing purposes.

Business continuity is another area to take into account. Remote data replication is a standard function in every enterprise storage system, and enterprises want the same functionality for Kubernetes. Unfortunately, standard data replication doesn't work—a storage array replicates only data volumes while a Kubernetes application needs additional data and information to be synchronized across clusters. In fact, the application's status and other information needed to support data replication are kept separate and must be moved and synchronized using a different method. What's more, if the user wants to deploy Kubernetes in hybrid cloud scenarios, there are additional challenges such as the lack of a cloud version of the physical storage array, making it impossible to use native replication mechanisms. These factors all contribute to an increase in both complexity and cost.

Last but not least, ease of use is a priority. While developers and Kubernetes practitioners favor APIs and CLIs, most enterprise data center operators are not ready to manage complex data protection operations in this way. They need a graphical user interface to set up and configure protection policies and to provide visual feedback on the workings of the environment.

3. The Solution

Complete end-to-end Kubernetes data protection should always be designed around three pillars:

- Application awareness
- Integration and data services
- Operations and ease of use

This is a major departure from previous approaches, where all the data protection effort was application agnostic and concentrated on the virtual machines instead of their content. In particular, application awareness is now critical, and understanding its layout and how all the components interact together is crucial for effective data protection. The discovery process is not complicated, however. Kubernetes applications can be deployed in several ways, sometimes using multiple mechanisms. The data protection solution must include tools to simplify and automate this activity, making it more accessible for backup teams not yet confident with Kubernetes.

As previously described, it doesn't make sense to protect single containers in Kubernetes. It is not enough, as they are only a part of the application and containers change over time. The right approach is to understand how the application is built, how it works, and all the components necessary to reproduce its state. To get a full picture of the application and its data, it is necessary to understand:

- Initial configuration (declarative state)
- Current state (runtime state)
- Additional information stored in external repositories and necessary for the correct functioning of the application including parts of the etcd database, CRDs, and other resources.

Additionally, to protect the application correctly, the data protection solution must understand how data is organized and managed within the application, including databases, to ensure that data is stable before being copied. This means that the backup process must include mechanisms (hooks) to put the database in a consistent state before performing the backup, and release it afterwards. Every database works differently though, and these tools must be designed to be highly configurable and scriptable.

Application discovery must be automatic. Leaving this operation to the user can lead to several issues, including the risk of backing up only part of the mandatory components needed to reconstitute the entire application safely. Kubernetes provides several mechanisms to group containers and other resources, but only a combination of them can provide a complete understanding of the application and its behavior.

Integration with the underlying storage stack is as important as application discovery. Kubernetes

provides the Container Storage Interface (CSI). CSI is an abstraction layer that enables Kubernetes and applications to deal in a similar and consistent manner with storage systems from different vendors and types of storage, without knowing every single API or function. This integration is fundamental for taking persistent volume snapshots and helps simplify and speed up the entire process. With deeper integration, it is possible to enable more sophisticated services and provide additional functionality and options for the user, including the ability to reduce RPO and RTO in mission-critical scenarios. This is of particular importance for organizations that must minimize the risk of data or service loss.

We are living in a moment of transition, with traditional operational models joined by DevOps and different approaches to infrastructure management. In this context, we can't afford to introduce a new data protection system that adds complexity to the entire storage infrastructure and proves difficult to use. Although the internals of Kubernetes are completely different from virtualized infrastructures, enterprises want to find solutions that support a smooth transition between the two worlds and avoid disruption of existing processes. Combining different user interfaces and taking advantage of familiar methods to help system administrators manage new environments is the only way to preserve freedom of choice and assist users with this transition.

Data protection is a critical process, and some organizations prefer to leave it to consolidated teams instead of giving control to application owners or developers. Again, it comes down to a combination of the two approaches—with interfaces for both traditional operations and DevOps teams—that will be the most beneficial in the medium and long term.

In the end, the user should always favor open approaches with solutions that integrate well with existing data protection frameworks via APIs and plug-ins.

4. Red Hat OpenShift and Data Protection

Red Hat OpenShift is a unique solution in the market. Yes, it is synonymous with enterprise Kubernetes for many users, but OpenShift is more than that. It provides a comprehensive platform that includes storage and data services, among many other things. The result is a consistent user experience, whether OpenShift is consumed on-premises or in the cloud.

With the introduction of additional functionality to Red Hat OpenShift Container Storage, Red Hat added several capabilities to its storage layer and is now in a position to provide enterprise-grade data services to its users. Red Hat OpenShift Container Storage now provides enhanced data protection and data mobility followed soon by additional features to facilitate the integration of Red Hat OpenShift and the applications it runs into the disaster-recovery and business-continuity strategy at the enterprise level. These are all features in high demand among enterprise users with Kubernetes in production environments but are provided by very few vendors today.

The solution is solid and easy to use, helping users embrace Kubernetes while maintaining the same level of service and data protection. Red Hat OpenShift limits the disruption of Kubernetes adoption in traditional enterprise environments by providing functionality that is similar to those of enterprise storage systems already in place. The end-to-end approach and deep integration between Kubernetes and the storage components provides:

- **Simplified transition to Kubernetes in production:** Makes it more adaptable to data protection policies already accepted by the technical and business organizations.
- **Operational continuity:** Enables applications that have been moved to Kubernetes through a lift-and-shift migration but are not yet ready to work in a full Kubernetes-native fashion. This also helps to migrate and preserve existing data protection policies, meeting the required levels of service.
- **Steep learning curve:** System administrator and data center operators don't need extensive re-training thanks to the easy-to-use UI and familiar concepts used in the product.
- **Developer-friendly:** Developers can quickly take snapshots and build clones for test and development purposes through API, CLI and OpenShift Console UI.

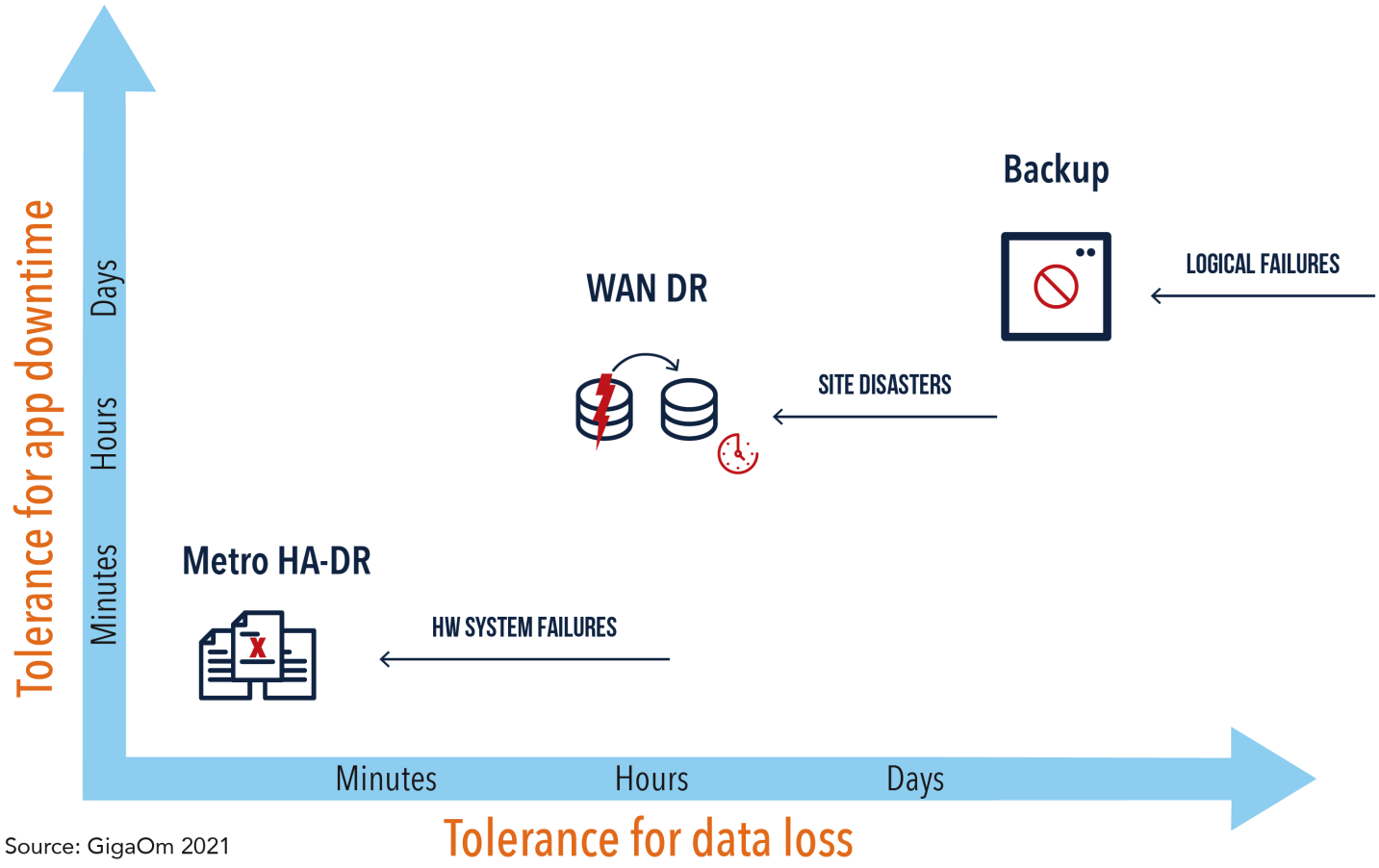
The work that Red Hat has done around OpenShift Container Storage and its integration with the rest of the stack enables users to provide first-class data service for Kubernetes, including efficient and application-aware integrated data protection. Thanks to integration with third-party solutions, it can manage local backup copies or deliver them remotely for workload migrations and create a separation gap for disaster recovery or security reasons. Furthermore, Red Hat OpenShift provides public APIs for further integration with third-party data protection solutions, giving additional options to the user.

The recent work done on Red Hat OpenShift Container Storage opens additional exciting opportunities for OpenShift users, including a series of upcoming features that include:

- Native asynchronous remote replication and automatic application failover capabilities, enabling users to provide the best service levels even in the worst scenarios.

- Highly available metro clusters—three or two site metro clusters, with synchronous data replication for zero RPO and RTO— to answer to the most demanding business requirements regarding service levels objectives (SLO).

Figure 1: Red Hat OpenShift Container Storage Data Protection Options



Source: GigaOm 2021

By looking at the entire picture, it is clear that Red Hat OpenShift Container Storage is going to offer complete end-to-end protection, with a broad feature set and data services covering a wide range of use cases, responding to the needs of organizations of all sizes deploying Kubernetes for business-critical applications (Figure 1). The solution is designed specifically for Kubernetes while keeping an eye on traditional enterprise needs, helping enterprises, and minimizing the impact of application migration to Kubernetes from the infrastructure point of view.

5. Conclusion

In general, enterprises moving from a development phase to production for Kubernetes should always think about data protection. Traditional data protection solutions and storage systems are not up to the task, and many Kubernetes-native solutions are not designed to work well with traditional enterprise approaches.

Red Hat brings decades of experience engaging enterprise customers and environments, which has informed its effort to create data services with OpenShift Container Storage. This approach enables enterprise organizations to move critical applications into production quickly and efficiently, while backed by the proper level of data protection and without disrupting existing processes around data management and protection. At the same time, its ease of use enables IT organizations to simplify operations and give access to the data protection tools to existing system administration and DevOps teams.

Red Hat OpenShift today is already in a leading position. The addition of storage and data services to the solution stack will serve to further improve the ROI and TCO that organizations can expect to see over the medium and long term.

6. 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.

7. 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.

8. Copyright