E-Guide

cloudcasa
by Catalogic

DevOps Guide for Choosing the Right Kubernetes Data Protection Strategy

The Safe Home for Cloud-Native Backup
www.cloudcasa.io

Introduction

As Kubernetes continues to grow in popularity, the shift to microservices is changing the way organizations develop, run and protect applications. Microservices architectures make it easy to develop, deploy, and decommission containerized applications quickly. But as adoption of container orchestration increases, specifically Kubernetes, organizations are using them to handle more critical workloads. Today, in a world where application data is often the lifeblood of these organizations, this brings up the very important question:

Do you need to backup the application and configuration data from these workloads? And how do you do that?

It is no secret that data protection and security are not strengths of the Kubernetes platform, at least not natively. This gap leaves developers operations teams searching through third-party backup providers, trying to decide which option would be best for them and their specific architecture. Because of the flexibility and portability of containers, every environment is different, which not only makes choosing a backup vendor very difficult, but often makes recovery almost impossible if you chose the wrong backup solution.

Security and data protection for Kubernetes environments can be difficult to setup and manage.

E-Guide

cloudcasa by Catalogic

This E-Guide will cover 9 topics to consider when trying to choose the best data protection strategy for your environment.

1

Are your Kubernetes workloads stateless or stateful?

2

Does your data reside on-premise or is it hosted using a cloud provider?

3

Does your environment contain multiple clusters?

4

Do you require a disaster recovery solution, or are local snapshots good enough?

5

Does your environment contain databases that need to be protected?

6

Are you concerned about ransomware protection?

7

Do you need to recover individual resources such as secrets or config maps?

8

Do you require data to migrate between different Kubernetes environments?

9

Is open source alternative Velero, good for your Kubernetes data protection strategy?

9 considerations before investing in a data protection solution

1. Are your Kubernetes workloads stateless or stateful?

Originally, one of the arguments in favor of containers was that no backups were needed. Since developers were only using the environment to rapidly spin up and spin down applications with very short working life, this data did not have to be protected. These stateless workloads do not need to persist and do not require a data protection strategy.

However, more organizations are beginning to see the value that containers can provide and are using them for more critical, stateful workloads. There are important datasets that the user is storing which need to be backed up, a lot of logs and configuration details are stored in clusters.

When Kubernetes contains production data that is critical to keep the organization running, a data protection strategy should be a requirement.

2. Does your data reside on-premise or is it hosted using a cloud provider?

Another result of the growth of Kubernetes is a wide-range of options when it comes to hosting Kubernetes distributions. Early adopters may have started with an on-premise architecture, using Red Hat OpenShift VMware Tanzu, and SUSE Rancher. But more popular these days is the option to use the managed Kubernetes environments of the major cloud providers. These include Akamai, AWS, Azure, Google, OVHcloud and others.

If your Kubernetes environment is hosted by one of these cloud providers, make sure that the data protection solution has integration with the cloud provider itself. For example, CloudCasa™ has direct integration with AWS EKS, AKS, and Google GKE, which allows for automatic discovery of available Kubernetes clusters, cloud account protection, full-stack and Any2Cloud recovery, and other features.

3. Does your environment contain multiple clusters?

If you are just beginning to use Kubernetes or you have not yet expanded to a multi-cluster environment, you may not need to pay for a third-party backup solution. Velero is an open-source backup and restore tool for Kubernetes, that allows users to protect single clusters or parts of one using namespace and label selectors. For small, less complex workloads that do not require a lot of management or restoration policies, Velero can be a great option.

However, Velero begins get difficult to use when Kubernetes workloads consist of multiple clusters. It complicates management and the CLI makes it difficult to manage multiple Kubernetes clusters, especially when data migration is involved. Overall, the performance of the CLI and of backups degrades when workloads become more complex, and applications increase in size.

Plan your Kubernetes backup strategy **to scale** as the company grows!

4. Do you require a disaster recovery solution, or are local snapshots good enough?

Kubernetes platforms allow users to take local snapshots of data. Developers can quickly save their work by taking a snapshot of their persistent or stateful data. Kubernetes users can then use these copies to bring a volume back to a point in time, or to provision a new volume. Third-party backup providers have the ability to schedule and manage these local snapshots, many even allowing more granularity when it comes to what is being protected.

But is this truly data protection? Since these snapshots are stored locally, what happens if the data center goes down? Or what if the cloud-hosted environment is unreachable? You lose all of those snapshots. That is why you must consider your plans for disaster recovery when choosing a data protection strategy. The backup provider must be able to store backups in an alternate location, preferably in the cloud using object storage. CloudCasa, for example, has the ability to copy and store Kubernetes data in either Amazon S3, or Azure, or in any other S3 compatible storage already owned by the organization.

5. Does your environment contain databases that need to be protected?

Kubernetes is application-centric, so it is not uncommon for Kubernetes users to utilize a cluster to host organizational databases containing critical data. In order to provide true data protection, you need application awareness to back up the application in a quiesced state. You need to capture the database in a particular state in time and recover the application in a usable form. Another challenge associated with protecting databases is Kubernetes Drift, a result of organizations transitioning from small, piloted application builds, and then expanding to the point where they have many developers building and deploying applications across different, often self-managed, clusters. Similar to the old “VM Sprawl,” this creates challenges when it comes to deploying consistent, secure applications with proper resource requirements.

This is why utilizing a data protection solution that provides application consistency is a necessity. CloudCasa has the ability to not only protect relational databases like Amazon Aurora, but also other databases built in Kubernetes clusters via pre-built application hooks. These include app hooks for MariaDB, MongoDB, PostgreSQL, MySQL, Oracle, and MS SQL. The intuitive scheduler of CloudCasa makes it easy to manage backups of these isolated applications, and schedule “gameday” or DR test recovery scenarios of these applications for testing purposes or for point-in-time recovery.

6. Are you concerned about ransomware protection?

Ransomware continues to be a major threat for organizations in all industries, and Kubernetes and associated cloud accounts data is no exception. Kubernetes data is particularly vulnerable in environments experiencing Kubernetes configuration drift. When multiple individual developers create and manage their own applications on separate clusters, it becomes increasingly difficult to validate security and resiliency in these siloed distributions.

Security in your backup strategy should start with protecting your backup copies, creating tamperproof copies, utilizing immutable snapshots, and S3 Object Lock. It should feature identity and access management controls like MFA, brute force prevention, and role-based access controls.

CloudCasa has a security-first approach, when it comes to protection of Kubernetes data. Not only does it include all the features mentioned above, but CloudCasa also provides security and vulnerability scanning on Kubernetes clusters, as well as associated cloud accounts. If CloudCasa identifies any misconfigurations, or vulnerabilities during cluster or cloud account scanning, it presents the issue to the end user in a readable report.

Security and ransomware protection should be an important part of any data protection strategy

7. Do you need to recover individual resources, such as secrets and configmaps?

Though most disaster recovery scenarios require recovery of entire clusters, there are several instances where backup or restore of a single resource, like storageclasses or secrets is required. Some tools, like Velero, are not able to provide the granularity needed to focus on individual resources. Instead, you are forced to backup and restore an entire namespace, performing more work than is really needed.

The Kubernetes backup solution must capture the data and application configuration at a granular level to ensure a fast recovery.

8. Do you require data to migrate between different Kubernetes environments?

As Kubernetes environments grow, especially in cases where developers create and manage their own applications on separate clusters, Kubernetes users may find themselves in scenarios where data from one cluster needs to be migrated or recovered to another cluster. There are many use cases where data migration between clusters might be required, including replicating applications to a staging or test environment, and migrating data from an on-premise Kubernetes environment to a cloud-hosted environment.

With its full-stack recovery feature, CloudCasa users can backup data from a cluster in any Kubernetes distribution and restore that cluster, namespace, or resource into any other Kubernetes cluster that is registered on their account within CloudCasa.io. To make migration to the cloud even easier, the Any2Cloud Recovery feature not only allows for data from any Kubernetes distribution to be migrated to EKS, AKS, or GKE, but also removes the need for a target "stand-by" cluster. Any2Cloud will instantly spin up a new cluster in AKS, EKS, or GKE on the fly to be used as the destination for the restore. This is something no other solution in the market can do and it and saves DevOps and Platform teams significant time and usage fees.

9. Is open source alternative, Velero, good for your Kubernetes data protection strategy?

Velero is the most popular and battle-tested open source data protection solution in the market. It is a powerful and versatile tool for backing up and restoring resources in a Kubernetes cluster, performing disaster recovery, and migrating resources and persistent volumes to another Kubernetes cluster.

However, if you are seeking additional insights into your backup and recovery operations, Velero might not suffice. Many times, you need a way to know where the backups are stored, what was included in the backup and their success/failure rates.

If you need a UI for your Operations team and want to be open source compatible, CloudCasa for Velero is the answer. It is fully compatible to Velero and addresses this important compliance and governance need for Enterprises. CloudCasa for Velero allows users to run Velero at enterprise scale with no disruption or migration. It guides the user through a user-friendly UI to recover from their Velero recovery points. CloudCasa for Velero enables monitoring, alerting and reporting for backup and recovery operations with its SaaS offering.

Pick the Right Data Protection Strategy!

Your data protection strategy should consider these 9 topics and ensure that the backup solution is built with the above considerations in mind. Since many Kubernetes environments grow organically, and have different owners with different requirements, you may think about using multiple backup tools designed for each use case.

Maybe you have a tool or script for backing up etcd, and another solution for the persistent volumes or databases. Or you have different tools for different Kubernetes distributions, or on-premise vs. the cloud. This may work fine until disaster strikes. Having multiple tools almost always causes issues with recovery. How do you coordinate restores from two or more tools and restore to a specific point in time? What if you need to restore an application to multiple clusters?

This is why it makes more sense to have a single backup solution that is able to manage each Kubernetes distribution, provide migration between those environments, the security required to give you peace of mind, and an easy-to-use interface to make management of these multi-cluster environments as simple as possible.

Choosing the best data protection for your company's needs is critical, as an inadequate solution might prevent you from recovering quickly.

Visit CloudCasa.io or please get in touch to review your requirements for a cloud-native backup solution.