VERITAS

Veritas NetBackup for Kubernetes Data Protection

Optimize your Kubernetes environment with unified, enterprise-grade resiliency, choice and flexibility.

White Paper | November 2021

Contents

Executive Summary

Digital transformation has accelerated the need for enterprises to offer their services to customers easily over the web. This transformation is a culture change and a survival need that mandates more agility to improve the customer experience, reduce friction, increase productivity and elevate profitability. An integral part of an organization is its data, and it is vital the data is protected and recoverable from the edge to the core to the cloud. Organizations continue to choose the cloud as a storage or recovery target due to its flexibility, agile nature, easy deployment and scalability. Tailoring deployments for each cloud is inefficient and time-consuming. However, business demands fast deployments, native tools and the ability to tie into any data center and cloud at any time. Training different teams on the fine details and intricacies of multiple clouds is difficult and ineffective. And virtual machines (VMs) are over-provisioned, slower to deploy and more costly compared to application containers.

Kubernetes (K8s) provides the consistency teams need to work with multiple clouds by creating abstractions that bring all deployments into one environment. Kubernetes is an open source container orchestration platform that automates deploying, managing and scaling containerized applications' processes. Kubernetes clusters can span hosts across on-premises, hybrid and multi-clouds and is an ideal platform for hosting cloud-native applications that require scalability, simplicity, flexibility, platform- and storage-agnostic and application-centric protections (see Figure 1).

Figure 1. An overview of the features an effective Kubernetes data protection solution should provide.

As more Kubernetes workloads go into production, enterprises need to maintain compliance. The flexibility and scalability Kubernetes offers also introduces levels of data protection challenges. Solutions that are not designed to be Kubernetes-native cannot tie into the CI/CD pipeline and can negatively impact the cluster. In Kubernetes, applications are made up of many components, so application protection and recovery need to be well orchestrated or the applications may not be able to recover. With various distributions of Kubernetes running on-premises and in the cloud, organizations depend on individual cloud providers for data protection, which diminishes portability.

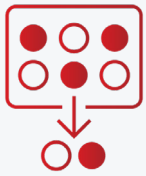
With Kubernetes, solutions must be application centric. One application workload could comprise hundreds of components and cross many containers. Containers are ephemeral and merely serve as a holding place. That's why Kubernetes data protection should focus on the application, not the container.

In general, a data protection strategy for Kubernetes should satisfy the following requirements:

- Preserve interdependencies between multiple containers during recovery.
- Use backup software that supports concurrent recovery of multiple containers.
- Detect the creation of containerized applications.
- Apply the right backup policy to each container, if necessary.
- Identify the dependencies that may exist between various containerized applications.
- Scale up or down to potentially handle millions of different backup-related activities.

Why NetBackup for Kubernetes?

Veritas designed NetBackup™ for Kubernetes to offer operational simplicity, enterprise-grade resiliency and choice and flexibility for Kubernetes workload protection (see Figure 2).



Operational Simplicity

Optimize

Designed to be cloud-native using Kubernetes custom operators and API's.

Designed to Allow the container architect or NetBackup admin the control to quickly and simply build protection for their K8's environment using native tools.

Enterprise-Grade Resilience

Fortify

Dicovery and protection of all component that make up an application in a namescape gives you the confidence to know that all objects required to run your business logical applications can be recovered quickly, easily and correctly.

Choice and Flexibility

Unlock

Key pillar of our design is a solution that can span across any Kubernetes distribution, on prem or in the cloud, for backup and recovery.

Figure 2. A summary of the features the NetBackup for Kubernetes solution provides.

Operational Simplicity

Veritas enables backup administrators to protect their K8s environment without becoming Kubernetes experts. Our backup and recovery solution is optimized by using Kubernetes-native constructs and a simple web UI.

Kubernetes-Native and API Integration

NetBackup uses Helm charts to enable customers to deploy all other Kubernetes resources. Helm is a tool for managing Kubernetes packages called charts. Helm can do the following:

- Create new charts from scratch.
- Package charts into chart archive (tgz) files.
- Interact with chart repositories where charts are stored.
- Install and uninstall charts into an existing Kubernetes cluster.
- Manage the release cycle of charts that have been installed with Helm.

We also provide a full suite of RESTful application programming interfaces (APIs) with role-based access control (RBAC) that enables self-service for users to align with their CI/CD pipeline and native tools. In NetBackup 8.2 or later, you can test out NetBackup APIs in your own environment with the Swagger interface at https://<master-server-name>/api-docs/index.html.

Efficient Management

NetBackup's self-service agentless management with RBAC and API-driven workflows lets you delegate tasks, so you can reduce the time spent on administrative activities and free up resources for higher-impact functions.

Simplified Administration

NetBackup provides a simple web UI to back up and recover Kubernetes workloads. With a few simple clicks, you can manage Kubernetes protection jobs and you can automate the protection jobs with labels.

Enterprise-Grade Resiliency

With NetBackup for Kubernetes, you can protect your environment without impacting or disrupting mission-critical applications. NetBackup is built with resilient Kubernetes constructs that deploy and work seamlessly with your environment.

Application Centric

Kubernetes consists of many abstractions that wrap applications and their data and provide interfaces for container orchestration services. For example, the underlying storage is provided to Pods via Kubernetes Persistent Volumes (PVs). They allow allocating a specific amount of storage to an app and configuring write/ read access permissions, I/O limits and storage security. Similarly, there are many other objects—such as Secrets, Service Accounts and Jobs—that control how containers within a Pod communicate and how the data is accessed by various microservices. Traditional data protection tools do not know how to interact with these abstractions to make true application backups that include Kubernetes objects, application configuration and data. NetBackup is Kubernetes aware and protects your applications seamlessly. NetBackup discovers and coordinates the snapshot and recovery of all components that comprise an application, including all PVs, config files and custom resources.

Snapshots

The Container Storage Interface (CSI) is a standard for exposing arbitrary block and file storage systems to containerized workloads on Container Orchestration Systems (COs) like Kubernetes. Kubernetes has a generic CSI plug-in that should work with any storage that is in beta. Many storage vendors are also writing their own CSI plug-ins to work exclusively with their storage arrays.

NetBackup leverages Kubernetes-native snapshots with CSI plug-ins. The native snapshots do not impact performance and provide the availability of always-on cloud operations.

Recovery at Scale

With NetBackup, you can recover one to thousands of namespaces with a single click, ensuring rapid recovery with the flexibility and scale to recover any workload.

Flexible Recovery

NetBackup provides the recovery flexibility to back up once and recover from any level of disaster. With NetBackup, you can restore an entire workload within a namespace, individual resources or just the associated PVs to the same or an alternate Kubernetes cluster.

Security

Organizations don't want to deploy a component that requires privileged access in their environment. The NetBackup for Kubernetes operator is designed on least-privileged access, so you do not need full Kubernetes admin privileges to deploy and run the Kubernetes operator.

Choice and Flexibility

NetBackup spans any Kubernetes distribution—on-prem or in the cloud—to enable easy backup and recovery.

Support Any Storage

NetBackup leverages CSI plug-ins to support any storage that has a production-ready CSI plug-in. NetBackup has completed storage validation for VMware vSphere storage and Google Persistent Disk.

Adopt Any Cloud

NetBackup is designed to use Kubernetes-native APIs so it can support any on-prem or cloud distribution. NetBackup is designed to be able to back up anywhere and recover anywhere.

Use Infrastructure As Code

NetBackup includes a full suite of APIs and Helm chart integration, providing the choice and flexibility you need to build custom workflows to fit into your CI/CD pipelines. You can use swagger to check the format and examples of these APIs. Build your own custom automation workflows to share with our GitHub community or take advantage of workflows already shared by others.

NetBackup for Kubernetes Architecture and Components

A Kubernetes cluster is a set of nodes that run containerized applications. Containerizing applications package an app with its dependences and some necessary services. Kubernetes clusters allow containers to run across multiple machines and environments: virtual, physical, cloud-based and on-premises.

There are three main components in the NetBackup for Kubernetes solution: the Kubernetes namespace, the controller node and the NetBackup dedicated namespace (see Figure 3).

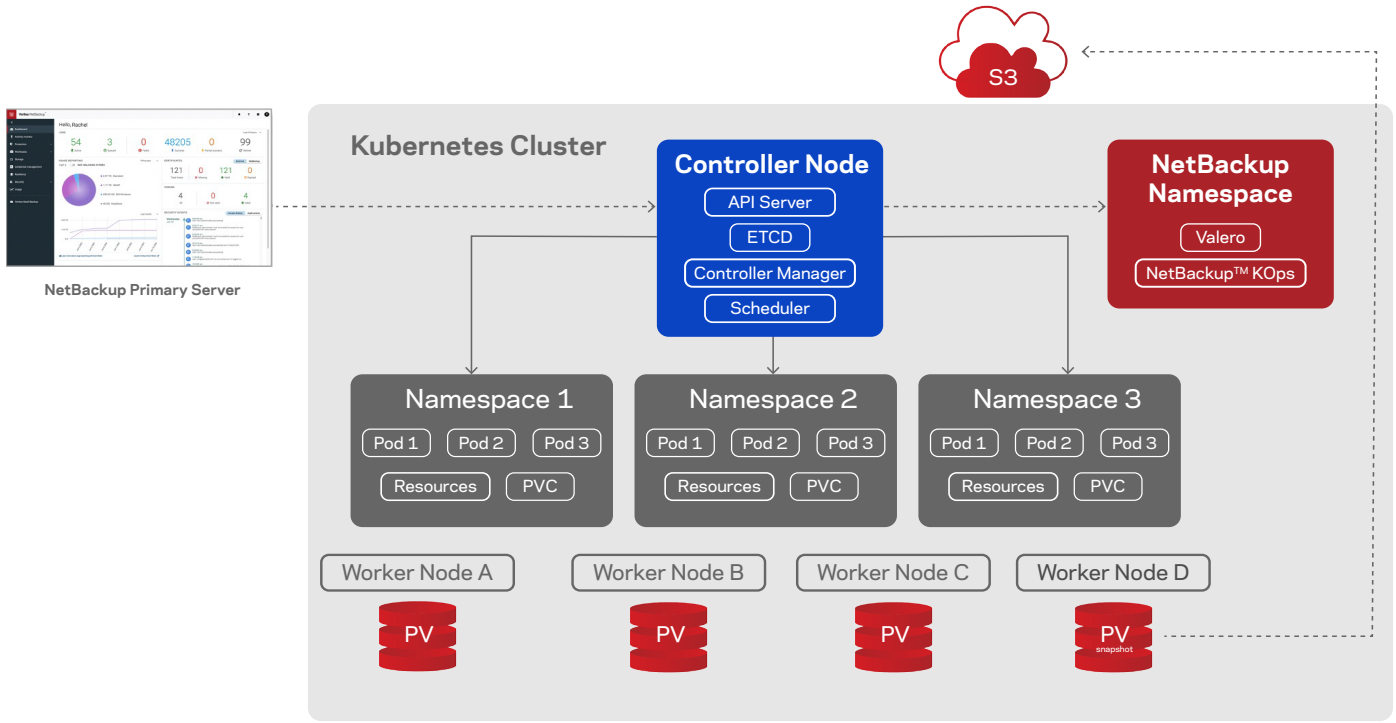


Figure 3. An overview of the main components of the NetBackup for Kubernetes solution.

Kubernetes namespaces include all the custom resources and configuration that comprise that namespace. Persistent Volume (PV) and Persistent Volume Claim (PVC) are two important concepts in Kubernetes storage.

An administrator can provision PVs or dynamically provision them using Storage Classes. Just like a node, a PV is a resource in the cluster. PVs are often presented to the names to hold any persistent data for an application. PVs have a lifecycle independent of any individual Pod that uses the PV. This API object captures the details of the implementation of the storage, whether it is NFS, iSCSI or a cloud-provider-specific storage system.

When a pod needs to store data, the storage resource is requested from Kubernetes, which is called PVC. The PVC also has capabilities such as capacity and read/write access modes. As Pods consume node resources, PVCs consume PV resources. Pods can request specific levels of resources (CPU and Memory). Claims can request specific size and access modes (for example, they can be mounted as ReadWriteOnce, ReadOnlyMany or ReadWriteMany or see AccessModes). Kubernetes matches the appropriate resource to the request, assigns it to the pod and mounts the volume to the node where the pod is running for the use of the pod.

The master node of Kubernetes listens for new requests for PVC generation, looks for and matches the PV and binds both together. Once the PV is bound to the PVC, the relationship between them is exclusive and unique, meaning the binding of PVC and PV is a one-to-one mapping. If there is no PV satisfying the requirements of the PVC request, the PVC will be in an unbound state indefinitely. Once there is a matching PV, the PVC will bind to the PV immediately.

The pod uses the PVC as a volume, and the Kubernetes cluster looks up the bound PV through the PVC and mounts it to the pod. For the volume that supports multiple access modes, you can specify the desired access mode when using the PVC as the volume, and the PV is exclusively owned by the user.

The **controller node** contains all the cluster information and access. Communication to and within the cluster is handled by an API server. Any communication to the NetBackup Custom Operator or calling of snapshots goes through the API server to make the request.

NetBackup dedicated namespaces are deployed with the NetBackup Kubernetes operator (KOps). This operator is a custom resource Veritas developed to go into Kubernetes environment and facilitate all the calls by Velero through the API server. It communicates information and metadata back to the NetBackup Primary Server, defines the schedules and facilitates all communication between NetBackup and Kubernetes. The NetBackup KOps will communicate with Velero and the CSI plug-ins through the API server to take any snapshots and storage of those snapshots to any cloud-compliant storage and communicate that status back to the Primary Server. NetBackup has integrated with Kubernetes and provides protection policies, asset service and discovery, RBAC, APIs, a client library, credentials and a web UI.

- Interface between NetBackup and K8s
- Incremental discovery
- Event listener
- Create Velero CR
- Kubectl/K8s CR-based backup and restore
- Manage NB DM

Summary

Kubernetes is a portable, extensible, open-source platform for managing containerized workloads and services that facilitates both configuration and automation. NetBackup for Kubernetes is designed to optimize the Kubernetes environment for the backup admin. Its application-centric design provides unified, enterprise-grade resiliency for all Kubernetes platforms and delivers the choice and flexibility organizations have come to expect from Veritas.

Appendix

Kubernetes Terms

Key Term	Description
API Server	The Kubernetes API server validates and configures data for the API objects, which include pods, services, replication controllers and others. The API server services REST operations and provides the front end to the cluster's shared state through which all other components interact.
Controller Manager	The Kubernetes controller manager is a daemon that embeds the core control loops shipped with Kubernetes. In Kubernetes, a controller is a control loop that watches the shared state of the cluster through the API server and makes changes, attempting to move the current state towards the desired state.
CSI /CSI Plug-in	The Container Storage Interface (CSI) is a standard for exposing arbitrary block and file storage systems to containerized workloads on Container Orchestration Systems (COs) like Kubernetes. Kubernetes has a generic CSI plug-in that should work with any storage that is in beta. Many storage vendors are also writing their own CSI plug-ins to work exclusively with their storage arrays.
Jobs	A job in Kubernetes is a supervisor for pods carrying out batch processes, that is, a process that runs for a certain time to completion, such as a calculation or a backup operation.
Valero	Valero collects application resources and metadata, takes CSI snapshot, recovers PV from snapshots and recovers application metadata. NetBackup uses Valero as an SDK for snapshot management.
ETCD	Kubernetes uses etcd as the back end for service discovery and stores the cluster's state and its configuration.
Persistent Volume (PV)	Kubernetes makes physical storage devices like SSDs, NVMe disks, NAS and NFS servers available to a cluster in the form of objects called Persistent Volumes (PVs). A PV is an abstraction for the physical storage device you have attached to the cluster. If you are using Kubernetes on Google's or Amazon's cloud, you can make your Google SSDs or EBS volumes available to your containers in the form of PVs. Each PV is consumed by a Kubernetes pod (or pods) by issuing a PersistentVolumeClaim object, a PVC. The easiest way to create the PV/PVC pair for your Pod is to use a StorageClass object and then use the storageclass to create your PV/PVC pair dynamically whenever you need to use it.
Pods	Pods are the smallest deployable units in Kubernetes. As the official documentation states, "A pod (as in a pod of whales or pea pod) is a group of one or more containers, with shared storage/network resources, and a specification for how to run the containers." So in simple terms, a pod is the mechanism for how a container actually gets turned "on" in Kubernetes.
Storage Class	A storage class is a Kubernetes object that stores information about creating a PV for your pod. With a storageclass, you do not need to create a PV separately before claiming it.
CI/CD	Continuous integration (CI) and continuous delivery (CD) embody a culture, a set of operating principles and a collection of practices that enable application development teams to deliver code changes more frequently and reliably. The implementation is also known as the CI/CD pipeline.

Versions

Version	Date	Author	Key Updates
1.0	Sep 2021	Rachel Zhu	Original document.

About Veritas

Veritas Technologies is a global leader in data protection and availability. Over 80,000 customers—including 87 percent of the Fortune Global 500—rely on us to abstract IT complexity and simplify data management. The Veritas Enterprise Data Services Platform automates the protection and orchestrates the recovery of data everywhere it lives, ensures 24/7 availability of business-critical applications, and provides enterprises with the insights they need to comply with evolving data regulations. With a reputation for reliability at scale and a deployment model to fit any need, Veritas Enterprise Data Services Platform supports more than 800 different data sources, over 100 different operating systems, more than 1,400 storage targets, and more than 60 different cloud platforms. Learn more at www.veritas.com. Follow us on Twitter at @veritastechllc.

VERITAS

2625 Augustine Drive
Santa Clara, CA 95054
+1 (866) 837 4827
veritas.com

For global contact information visit:
veritas.com/company/contact

V1458 11/21