Dell EMC PowerProtect Data Manager with Data Protection for Kubernetes on VMware

Silverton Consulting, Inc. StorInt™ Briefing

Introduction

The cloud-native containerized applications that have emerged over the last decade were originally used for stateless web services. But today, cloud native applications have matured, and along with advanced container orchestrator capabilities, we are seeing customers deploy both stateless and stateful workloads in containers.

VMware has a few container solutions that integrate with the VMware vSphere® ecosystem. These solutions often leverage open source projects and, where successful, mature into baseline functionality. Dell EMC™ has participated in many of these projects in an effort to make their own infrastructure systems and software better suited for running cloud native applications.

In the spirit of collaboration, VMware and Dell EMC™ have been contributing to the VMware-originated open source project, **Velero**. Velero is designed as a tool to safely backup and restore, perform disaster recovery, and migrate Kubernetes™ cluster resources and persistent volumes.

The mission for PowerProtect Data Manager has always been to protect data and deliver governance control for business-critical workloads across physical, virtual and cloud environments. As such, with the rapid adoption of Kubernetes in the enterprise, Dell EMC has incorporated best in class, data protection functionality into project Velero and leveraged that in their, **PowerProtect Data Manager with data protection for Kubernetes on VMware** solution.

In addition to Kubernetes workload data protection, Dell EMC's newest release of PowerProtect Data Manager enhances VMware support (discussed later), adds a new public cloud environment to Cloud DR, facilitates management for multi-Data Manager environments, and provides other improvements to make it easier for customers to protect data.

In this paper, we will review the rise of cloud native enterprise applications on premise and in cloud environments; the rise of Kubernetes for automating deployment, scaling and management of containerized applications; and the benefits of using Dell EMC PowerProtect Data Manager with data protection for Kubernetes on VMware.

Container applications

Containers were originally developed as a method to package, deploy and isolate applications from each other to improve density. As a result, DevOps teams were able to build, test, deploy and run containerized applications in cloud environments with limited operator engagement. Through DevOps automation, container-based applications can be updated and released into production daily or even at higher frequency. This level of developer agility has never been previously achieved.

Not surprisingly, containers have become the norm for modern applications and cloud-native applications. We are seeing an unprecedented growth of workloads and data being refactored, rearchitected and re-implemented to utilize containers. But, the challenge to protect data at scale increases as applications move to this new container paradigm.

In the beginning, container applications were **state-less**, meaning they had no need for storage or persistent data. Indeed, containerized applications could be started or restarted multiple times without a problem because they held no state information. Migrating an application from one node to another simply meant to end it and restart it on the new node.

However, as more enterprises are migrating mission-critical applications to containers, storage of state information or persistent data becomes a necessity. As a result, Kubernetes has evolved support for **container persistent volumes** that enable system vendors to provide durable storage for container applications.

Containers enable development teams to adopt more modern, cloud-native application development practices like standardized CI/CD (continuous integration/continuous deployment), with frequent deployments that automate everything, that can be built-once and deployed-anywhere. Enabling developers to operate in a self-service way is another very key benefit of Kubernetes container applications. IT is adopting container applications and orchestrators like Kubernetes, for many additional reasons, including the following:

- **Faster development and deployment** – container applications can be built from existing or new microservices, deployed across large infrastructure farms and updated automatically without operator intervention.
- **Cyber resiliency** – container applications can be quickly patched and rapidly redeployed across infrastructure to address any vulnerability.
- **Portability** – container applications can be developed once and run just about anywhere (e.g., on bare metal servers, on virtualized infrastructure and in private or public cloud environments).
- **Higher resilience** – container applications can be run as multiple container instances, making it easier to deploy high availability without having to add other infrastructure.
- **Scalability** – container applications can automatically be scaled-out in response to service demand.
- **API-driven capabilities** – Kubernetes as an orchestrator is designed to allow container applications to be deployed, run and updated under fully automatic API control to minimize operator intervention.

The rise of Kubernetes

Data centers can run containers in a number of ways and, depending on the infrastructure, run them on many different container engines. **Kubernetes** has emerged as the dominant container orchestration and scheduling solution across every enterprise and cloud provider, from the edge to the data center to the public cloud.

Kubernetes is an open-source system for automating the deployment, scaling and management of containerized applications. It supports a cluster of nodes, some of which are used for scheduling, routing and orchestration. Every cluster has at least one worker node. The worker node(s) run the workloads.

Multiple containers can share the execution environment of a single pod, and multiple pods can be run to provide all of the functionality of a containerized application, service or deployment. Container instance executables reside as **images** in **registries**.

Kubernetes also supports **namespaces**, which refers to a logical abstraction of virtual or physical cluster resources. Namespaces can be used to define a limited scope for Kubernetes control activities. Data centers can have one or more Kubernetes clusters in operation across their infrastructure.

Although, originally developed by Google, Kubernetes is now managed by the **Cloud Native Computing Foundation® (CNCF®), a subsidiary of the Linux Foundation™**. CNCF Certified Kubernetes products are currently offered in public cloud environments, such as Amazon Web Services™ (AWS™) EKS, Microsoft Azure™ AKS, Google Cloud Platform™ GKE™, IBM Cloud™ Kubernetes Service and many private cloud providers. In addition you can find CNCF Certified Kubernetes products in packaged offerings like VMware Enterprise PKS.

Dell EMC PowerProtect Data Manager

Dell EMC PowerProtect Data Manager is a software-defined data protection solution for the enterprise, that has been redesigned and re-implemented to provide an **architecture** that makes it easier to deploy, scale and update than prior-generation solutions. What this means for data center customers is that innovative cloud native services such as Velero will continue to be developed and enhanced to meet the growing business needs of enterprises.

This new PowerProtect Data Manager (Data Manager) also offers the following functionality:

- **Enhanced VMware data protection**, which includes granular file/folder level restores, support for VM application data protection in VMware Cloud™

(VMC) on AWS, instant restore of VM images, SQL application-aware backups and simplified management for VM admins.
- **Enhanced cyber recovery**, which provides a secure, "air-gapped" PowerProtect vault that can be used to recover systems in the event of security lapses, inadvertent corruption and other situations that can take down infrastructure.
- **Enhanced cloud disaster recovery**, with support for Azure and AWS environments as targets for PowerProtect Cloud DR or DRaaS.
- **Enhanced SaaS reporting**, which provides better analysis, reporting and monitoring of distributed PowerProtect environments.

In addition to the above, Dell EMC continues to add and update the PowerProtect REST API services, enhancing the level of automation and integration with other systems and tools.

When cloud native applications started to utilize persistent volumes or data, the need for data protection became a top priority for IT operations. By protecting container-persistent volumes, operations can be confident that they will be able to recover container applications in the event of infrastructure outages and other scenarios that cause data loss.

While operations was already familiar with data protection solutions used to protect virtualized workloads, what was missing was a software solution that offered data protection for both virtualized and containerized applications.

Dell EMC PowerProtect Data Manager with Data Protection for Kubernetes on VMware

As mentioned earlier, PowerProtect with Data Protection for Kubernetes on VMware includes a secure, enterprise-hardened and operations-enhanced version of Velero data protection functionality. Velero is an open source tool that can be used to back up and restore data as well as the configuration of workloads running in Kubernetes. By embedding and enhancing this functionality into Data Manager, Dell EMC has made Kubernetes data protection inherently more usable and a first-class citizen of PowerProtect offerings.

Prior to Velero, DevOps had to use scripts to copy container data to protect container applications from data outages. Even when using Velero, DevOps was left to configure and run protection activities on their own, as there was little to no operator centralized control, and little automation outside of backup scripts.

Data restoration was another problem. Persistent volume data restores often required DevOps to create tailor-made scripts for the data that was lost. Moreover,

script development usually occurred while container applications were down, thus extending service outages.

Dell EMC PowerProtect Data Manager with data protection for Kubernetes on VMware supplies automated protection for containerized application environments, offering hands-off backups of persistent volumes and easier, operator-specified restoration of persistent volume data.

Furthermore, PowerProtect data backups are **crash consistent** for Kubernetes persistent volumes, and readily stored on **Dell EMC Power Protect DD** (the next generation of Data Domain appliances available today), Using PowerProtect DD storage for container backups reduces storage footprint through highly efficient data deduplication and compression while providing a high-performance backup and restore platform that delivers the scalability needed to support growing application environments – on-premise and in the cloud.

Data Manager powered by Velero has access to Kubernetes structures and cluster resources and can easily scale up or down as backup workload demand changes.

As a native application, Data Manager can auto-discover all namespaces in a cluster and discover persistent volumes and persistent volume claims in the namespace that need to be protected. Backup administrators (IT Ops) can define policies in Data Manager to automatically back up all persistent volume data within a namespace. Further, PowerProtect for Kubernetes restore requests can do granular restores by selecting persistent volumes or namespaces rather than restoring all backed-up data.

Additionally Data Manager also supports self-service functionality allowing IT teams to work independently. A backup admin can create a generic policy to look for any name spaces labeled with something like “Backup_Policy: Gold” which developers or DevOps can use to label the namespaces using the Kubernetes console. Through the use of automated filters, Data Manager can discover these labeled namespaces and automatically protect all persistent volume data in use by those namespaces based on the policy definitions.

In addition, Data Manager is not limited to restoring persistent volume data to the same Kubernetes namespaces that were backed up. Instead, Data Manager restores can be made to a different namespace or create a new name space within that cluster. In a future release, Data Manager will be able to restore to a different Kubernetes cluster or namespaces within another cluster.

Data Manager can also be configured to support multiple Kubernetes clusters in the application environment. In this case, Data Manager deploys its container data protection application within each Kubernetes cluster.

Finally, as PowerProtect Data Manager is an enterprise-class data protection solution, it can easily recover or restore container application persistent volume data from any backup version that currently exists on secondary storage. Also this data can be tiered to a cloud Object store or tape for long term retention as the secondary storage becomes full. .[1]

Summary

Kubernetes clusters and containers are becoming an increasingly popular option for deploying enterprise applications into public cloud and on-premise environments . Most, if not all, enterprise applications require stateful containers that make use of persistent volumes. As these applications move to production, operations needs a way to provide enterprise-class data protection services in order to prevent major service outages.

Dell EMC PowerProtect Data Manager is the only data protection solution designed from the ground up to protect Kubernetes container applications, virtualized applications and traditional bare metal application environments. Moreover, PowerProtect Data Manager uses highly efficient secondary storage like Dell EMC Data Domain appliances to optimize your backup data footprint and throughput.

With Velero functionality integrated into Data Manager, DevOps teams gain fast, reliable, easy-to-use enterprise-class data protection for Kubernetes container application persistent volumes.

As cloud native adoption continues to grow due to business services demanding greater agility in application development and deployment, Dell EMC PowerProtect Data Manager offers the solution needed to make sure that all your modern applications and application data are compliant with your enterprise business continuity policy.

Silverton Consulting, Inc., is a U.S.-based Storage, Strategy & Systems consulting firm offering products and services to the data storage community.

[1] This summary highlights only some of enterprise-class data protection and recovery functionality Dell EMC PowerProtect for Kubernetes offers. For more information, talk to your Dell EMC sales representative.

Silverton Consulting
Strategy, Storage & Systems