Technical Insight Report

Protecting Kubernetes Applications with Dell PowerProtect Data Manager

By Krista Macomber

March 2022

Evaluator Group

Enabling you to make the best technology decisions

intel®

DELL Technologies

Challenges with Protecting Kubernetes Environments

An ever-growing number of applications, including data-intensive applications, are being rearchitected for, and being built on, container architectures. In Evaluator Group's recent study, Hybrid Cloud Matures: Pragmatism in a Post-COVID-19 World, a large majority of respondents (84%) are adopting or plan to adopt containers, with 29% having a goal of moving most, if not all, of their workloads to containers, and 18% already having moved most if not all of their workloads to containers. Only 16% had no interest in transitioning to containers.

Clearly, for today's enterprise, if containers are not a part of the present, they are a part of the near-term future. This means that they need to be protected, just like their physical and virtual counterparts. However, they do have a number of unique considerations beyond traditional data protection requirements that need to be factored in.

Firstly, the shift to container architectures drives a shift from few large scale-up databases to multiple scale-out databases – making centralized oversight and protection by IT more difficult. At the same time, protection must happen in a way that does not impede the agility of DevOps teams when developing and updating software. Backups must happen in an automated and seamless way, and DevOps teams do not have time to wait for IT if they need to roll an application back to a previous point in time.

Another challenge facing IT operations is that while Kubernetes is established as the most popular container orchestration tool, the Kubernetes distribution, management and storage ecosystems are inconsistent and very much still maturing. Enterprises can choose from a variety of open-source and vendor-specific implementations, making it difficult to be definitive about data protection requirements.

Enterprises are taking a variety of approaches to working with Kubernetes. Some are using PaaS platforms that support Kubernetes, such as Red Hat OpenShift, or they are retrofitting PaaS platforms on Kubernetes, such as CloudFoundry, which does not yet support Kubernetes natively. Cloud-based Kubernetes services such as AWS Kubernetes Service (EKS), Microsoft Azure Kubernetes Service (AKS), and Google Kubernetes Service (GKE) are also being adopted. Meanwhile, other enterprises still are using the Cloud Native Computing Foundation (CNCF) ecosystem itself to build their own. This includes:

- Building on top of their choice of Kubernetes distribution, including an open-source or a vendor-specific option such as VMware Tanzu, SUSE Rancher and RedHat OpenShift
- Creating source code and application lifecycle management capabilities
- Tapping CI/CD automation pipeline tools such as Jenkins.
- Creating IT infrastructure automation or infrastructure-as-code (IaC), which is the managing and provisioning of infrastructure through code instead of through manual processes.

- Using monitoring and logging frameworks such as Prometheus/Grafana.

Adding another layer of inconsistency, IT has several multi-Kubernetes cluster management tools such as SUSE Rancher and VMware Tanzu to choose from.

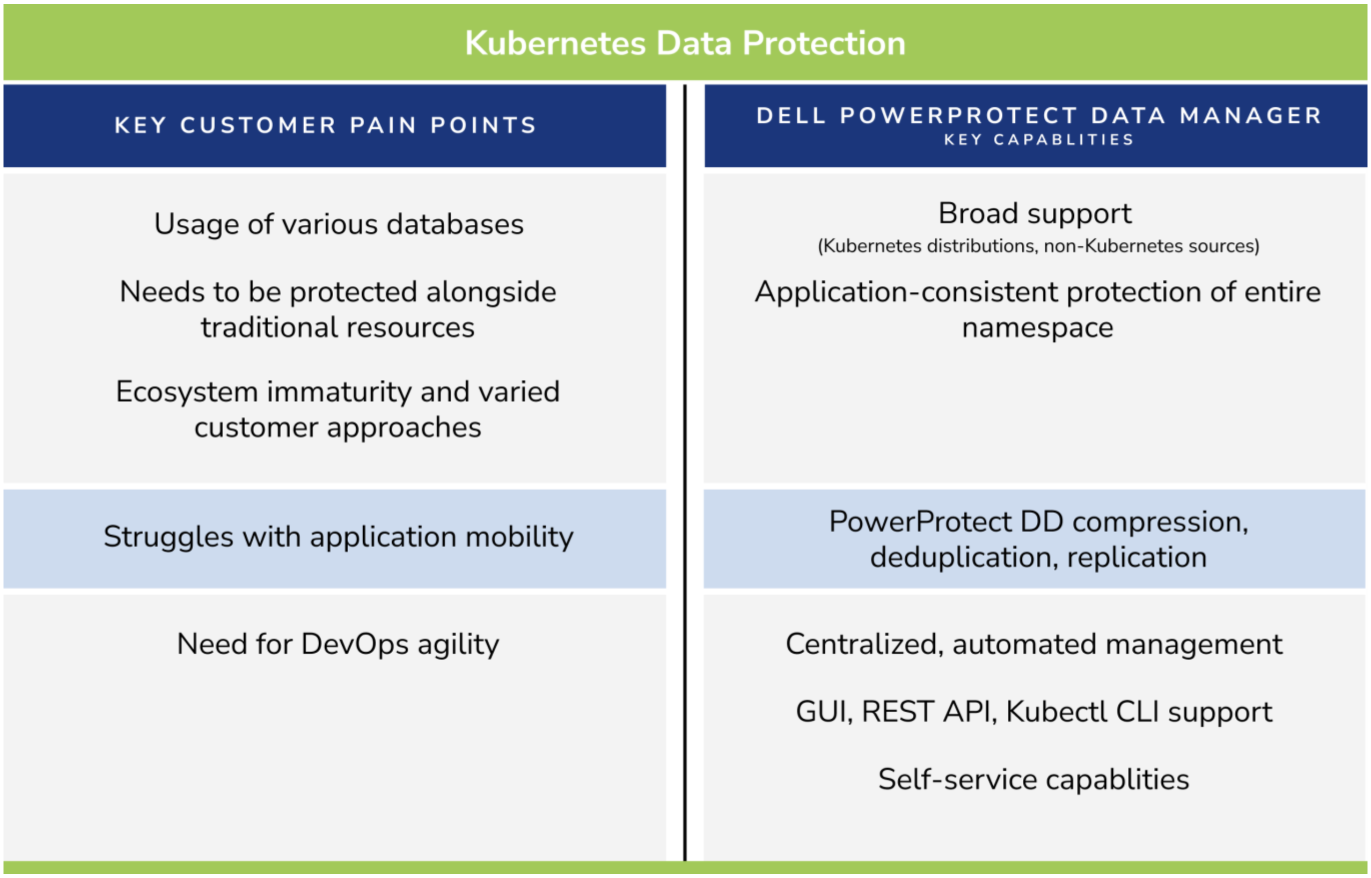
Similarly, there is a broad range of Kubernetes storage infrastructures available for IT to choose from. Kubernetes supports access to persistent storage via the Container Storage Interface (CSI), but CSI has numerous gaps that a number of storage vendors are working to address with their own container storage offerings. While some vendors, like Dell, are layering value-added capabilities on top of CSI, others, like Portworx, have developed their own container-native storage offerings.

There is no single-vendor Kubernetes solution addressing all of customers' platform, infrastructure and management requirements. From a data protection standpoint, this creates the challenge of supporting all of these varied systems. With all this in mind, a few core capabilities emerge for Kubernetes data protection. IT requires a complete but simple-to-use offering so that it can oversee and enforce protection policies without impacting the DevOps or business user experience. This necessitates a single platform for multiple apps, centralized management, self-service capabilities and support for the native Kubernetes command line interface (CLI). The data protection solution should enable the application to be migrated to different distributions and infrastructures. Table stakes in any data protection implementation is also optimizing efficiency. This paper explores Dell PowerProtect Data Manager as a tool for addressing these requirements.

Dell PowerProtect Data Manager for Kubernetes Protection

Dell PowerProtect Data Manager is a data protection software solution offering backup, operational recovery and disaster recovery for a variety of sources, including Kubernetes environments. Data Manager itself was built on a container-based, microservices architecture, and it is deployable on-premises as well as in the AWS, Microsoft Azure and Google Cloud Platform (GCP) clouds. It supports Dell PowerProtect DP series or PowerProtect DD series appliances as targets, as well as the ability to back data up, restore data and tier data to AWS, Microsoft Azure and GCP clouds for archiving and long-term retention.

Data Manager natively supports a number of major Kubernetes distributions, including Red Hat OpenShift in VMware vSphere, VMware Tanzu Kubernetes Grid clusters and Kubernetes distributions delivered as-a-Service (e.g., AWS/EKS, Azure/AKS, and GKE). Not only does it unify protection of these sources, PowerProtect Data Manager also supports a variety of non-Kubernetes sources. These include AWS, Azure and GCP infrastructure-as-a-service (IaaS), enterprise databases (Cassandra, Microsoft Exchange and SQL, MongoDB, MySQL, Oracle, PostgreSQL, SAP HANA), VMware environments (on-premises and VMware Cloud on AWS) and Linux and Windows filesystems.



Kubernetes Data Protection

KEY CUSTOMER PAIN POINTS	DELL POWERPROTECT DATA MANAGER KEY CAPABLITIES
Usage of various databases Needs to be protected alongside traditional resources Ecosystem immaturity and varied customer approaches	Broad support (Kubernetes distributions, non-Kubernetes sources) Application-consistent protection of entire namespace
Struggles with application mobility	PowerProtect DD compression, deduplication, replication
Need for DevOps agility	Centralized, automated management GUI, REST API, Kubectl CLI support Self-service capablities

For all of these varied Kubernetes distributions and other sources that are supported, PowerProtect Data Manager offers consistent management procedures to streamline day-to-day tasks for IT operations. Management tasks and the user experience are further simplified through the ability to automatically discover resources to be protected and then to automatically back them up through a policy engine that is controlled by IT. Policy-based automation can also be applied to replication, retention and tiering to S3-compatible storage. A GUI is offered, and REST APIs and the kubectl CLI are supported to allow for management by both IT operations and DevOps teams. Reporting on adherence to various service level agreements (SLAs) is available for IT operations.

Data Manager is agentless and its protection spans the entire Kubernetes application namespace. This is important because just as persistent data residing outside of Kubernetes requires protection, so does all of the Kubernetes components in order to avoid configuration drift. All of this needs to be backed up together, in context, so that the application can be recovered successfully and in a timely manner. Data Manager protects Persistent Volumes (PVs) and Persistent Volume Claims (PVCs), Kubernetes Pods and StatefulSets, consistency groupings, the kube-scheduler, ClusterRoles and RoleBindings, the etcd metadata database, various Kubernetes Certificates, ConfigMaps, Custom Resource Definitions, and Secrets.

In addition to protecting the entire application namespace, Data Manager facilitates agentless application consistency through an extensible framework. For example, Data Manager integrates Kubernetes storage volumes and StatefulSets in order to meet even strict RPOs/RTOs. Data versioning and extensive backup cataloging provide a strong foundation for minimizing data loss in the event of a cyber-attack such as ransomware.

Data Manager obtains several important additional capabilities through integration with Dell PowerProtect DD series appliances. These include compression with source and target-side deduplication for faster backup jobs and increased cost efficiency, as well as in-place restores for reduced downtime. PowerProtect DD series replication allows for mobility of the Kubernetes application between sites. Integration with PowerProtect DD Virtual Edition (DDVE) abstracts replication (in addition to other capabilities including compression and deduplication) from an on-premises PowerProtect DD series secondary storage appliance to the cloud – allowing application migration to be orchestrated from on-premises to the cloud. Finally, development teams can use PowerProtect DD series to create a data pipeline architecture. For example, Data Manager allows them to insert multiple data sources into a common data warehouse, so that cloud analytics can then be applied to the data.

Conclusion

Today’s IT teams have quickly become faced with the need to protect Kubernetes applications and their data. This is a unique challenge, because Kubernetes applications utilize a variety of databases that are managed from outside of IT. They must be mobile across an array of distributions and infrastructures. And all of this must be achieved in a way that does not impede the agility of DevOps, and alongside the protection of more traditional physical and virtual resources. This is especially a tall order, because the entire ecosystem surrounding Kubernetes applications is fragmented and still evolving and maturing.

With PowerProtect Data Manager, Dell has brought to market an offering that addresses these key pain points, including through its support of a broad range of protection sources and targets under one umbrella, its streamlined and flexible management capabilities, its ability to protect the entire application namespace in an application-specific way, and its integration with Dell PowerProtect DD series.

About Evaluator Group

Evaluator Group Inc., an Information management and data storage analyst firm, has been covering systems for over 20 years. Executives and IT Managers rely upon us to help make informed decisions to architect and purchase systems supporting their data management objectives. We surpass the current technology landscape by defining requirements and providing an in-depth knowledge of the products as well as the intricacies that dictate long-term successful strategies.

About Dell Technologies

Dell Technologies helps organizations and individuals build their digital future and transform how they work, live and play. The company provides customers with the industry's broadest and most innovative technology and services portfolio for the data era.

About Intel

Today's organizations face strategic challenges as they modernize data centers and servers. Intel is driving platform innovation and next-generation capabilities across every infrastructure domain—from compute to storage to network to memory to accelerator technologies. With Intel® architecture-based platforms, you have a clear path forward for the data-centric era.