IDC TECHNOLOGY SPOTLIGHT

Sponsored by: Pure Storage

Current VMware customers looking to leverage Tanzu to build and deploy containerized applications quickly and reliably need an integrated enterprise-class Kubernetes data services platform that is highly scalable; delivers container-granular, application-focused storage operations; and is optimized for hybrid, multicloud operations.

Increasing Containerization Highlights the Need for an Enterprise-Class Kubernetes Storage Platform

June 2021

Written by: Eric Burgener, Research Vice President, and Lucas Mearian, Research Manager, Infrastructure Systems, Platforms, and Technologies Group

Introduction

With most enterprises in the midst of digital transformation (DX) — meaning the evolution to much more data-driven business models — CIOs are moving in the direction of more agile, cloud-native approaches to application development and deployment. Cloud-native applications leverage new approaches such as DevOps, microservices-based design, and containerization to speed development and deployment, simplify software maintenance, enhance portability and scalability, and improve reliability. Most new applications will be built and rolled out using this cloud-native approach.

Historically, most enterprises have deployed monolithic applications in virtual machines (VMs). For production enterprise workloads, VMware virtualization technology is more widely deployed than any other. As organizations evolve in the direction of cloud-native applications, the application evolution will not happen overnight, and they will clearly need to run both VMs and containers side by side. For customers looking to add containerized applications to their VMware virtual infrastructure, VMware has introduced Tanzu, a portfolio of products and solutions that enable customers to build, run, and manage Kubernetes-controlled containerized applications in both in-house infrastructure and the public cloud.

AT A GLANCE

KEY STATS

According to IDC research:

- Today, more than 80% of enterprises are running hybrid cloud environments in production.
- By 2023, nearly 2 billion container instances will be running across more than 100 million hosts.

KEY TAKEAWAYS

To move containerized applications to production quickly and reliably, enterprises need a Kubernetes data services platform that is highly scalable; organized around granular, application-focused operations; and optimized for a hybrid, multicloud environment and that includes a comprehensive set of enterprise-class data services.

IDC primary research reveals that 84% of enterprises have repatriated at least one workload from a public cloud environment back into on-premises IT infrastructure. When containerized workloads are moved back in-house — a decision made easier by the data mobility that containerization supports — enterprises will need to have the right supporting infrastructure for them. Enterprises will need to be able to move containerized workloads freely among the different deployment models, and that requires the ability to run Kubernetes in-house as well as in the public cloud.

Today, organizations are considering how they will containerize both existing and new workloads that require persistent storage such as databases, file shares, collaboration platforms, and business analytics. In addition to persistence, that storage may also need to meet other performance, availability, data protection and recovery, and/or security requirements. Containers, which are more lightweight than VMs, are ephemeral by design. Kubernetes does enable persistent storage through a PersistentVolumeClaim (PVC), a storage request made by a user that decouples the data from the container and ties it to a local pod. If a container crashes, the data that is backed by volumes in its file system will persist on the pod. However, if the node or cluster on which the pod resides goes down, the data is not preserved. This is why the Container Storage Interface (CSI) — a standardized plug-in — is needed for storage that is not built as a cloud-native application (i.e., it is not designed to run natively as a Kubernetes-based application). CSI is the most basic requirement for container storage, but a Kubernetes storage platform that supports stateful, production enterprise applications must exhibit other essential attributes as well. Customers that want to leverage Tanzu to build and deploy containerized enterprise applications quickly for production use need to address these essential attributes requirements.

How Container Storage Is Different

Unlike VMs, containers do not have their own host operating system. Instead, they share a host operating system environment. This results in two primary advantages. First, containers can run in virtually any on-premises or cloud platform, meaning they're extremely portable. Second, containers are lightweight, meaning they require less storage capacity (i.e., the typical container is megabytes in size compared with the typical VM, which is gigabytes in size). DevOps teams will typically use public cloud resources to create and test containerized applications and then bring them back on premises to run them. While organizations have their preferred operating model — whether AWS, Azure, Google Cloud, or on premises — users want to be able to run applications across multiple environments and want a consistent experience. When containers running in production are deployed to support existing applications, those applications may be repatriated to an on-premises location to address security or regulatory requirements. As IT managers evolve toward more containerized, Kubernetes-based environments, they will need to take into account four key areas:

- **Scalability.** Cloud-native applications are architected very differently from legacy applications. While a traditional application will be a single large software component that includes many different "services" necessary for proper execution (all of which run in a single VM), a cloud-native application will actually consist of a collection of separate "services," each running in its own container. An enterprise application architected around a microservices-based design could easily have 100 or more separate containers, all of which are communicating through APIs, to provide what an end user views as the "application." An application running in a VM may have several (or more) volumes attached to it, but each container could have one or more persistent external storage volumes attached to it if its service is stateful. When containerized environments are being deployed, the storage will need to deliver on service-level agreements even as it handles thousands of concurrent operations per second and support management operations that can easily accommodate this level of scale.
- **Granular, application-focused operations.** A particular application service may be made up of many different microservices that are spread across nodes, namespaces, clusters, and/or geographic locations. To maintain consistency, an application needs to be able to ensure that storage operations such as snapshots, backups, recoveries, and migrations are applied at the container level and preserve the application state (i.e., the data + metadata + configuration). This feature moves administrators away from volume- and/or VM-focused operations and toward container- and/or application-focused operations.

- **Optimization for a hybrid, multicloud environment.** For most enterprises, the move to the public cloud provides their first exposure to containerization. As they develop their own cloud-native applications using this technology, they want to take advantage of the portability that containers support to spread their IT infrastructure across multiple deployment models, including both on-premises and off-premises (i.e., public cloud) locations as necessary. The IT operations of more than 80% of enterprises are already spread across in-house and off-premises infrastructure (i.e., hybrid IT) in one or more public clouds. Storage solutions for these environments need to take into account this reality, supporting a consistent set of data services and a unified management dashboard that provides "observability" across all these different deployment models.
- **A comprehensive Kubernetes storage stack.** To support "enterprise" use cases, a storage platform must deliver a comprehensive set of data services: persistent storage provisioning, high availability, security, capacity management, data migration, backup, and disaster recovery. To effectively support stateful containerized applications, a Kubernetes storage stack must also include these data services.

Trends

With the increasingly dynamic business environment, the need for IT agility will only grow over time. The innovations upon which cloud-native computing is built — DevOps, microservices-based architectures, and containerization — are driven in large part by this need for agility (in addition to other requirements such as improved software reliability and better application portability).

Containerization is expected to grow rapidly. According to IDC, nearly 2 billion container instances will be running across more than 100 million hosts by 2023. Increased deployment of cloud-native applications, as well as the refactoring of legacy workloads into containers, is driving this growth. Another driver is the growth in containerized applications that have already been deployed (a scenario that leads to the need for additional containers when the performance of certain functions such as microservices must be scaled to support growth).

Primary research performed by IDC in December 2019 found that 55% of containers running in production were deployed to support existing applications, a majority of which required persistent storage. The primary reason to move legacy workloads to containers is to speed time to market by improving the agility with which applications can be deployed, upgraded, and moved to new deployment models. Of the applications running in containers, a good percentage of them require persistent, enterprise-class storage (regardless of whether they are legacy or cloud-native applications). Given this broad need for persistent storage, there is a clear and long-lived need for Kubernetes storage platforms that support four essential attributes: scalability; granular, application-focused operations; optimization for a hybrid, multi-public cloud environment; and a comprehensive Kubernetes storage stack with support for key data services. Most enterprises performing application modernization will require such a platform, and most of those organizations currently have legacy workloads running on top of VMware virtualization.

Considering Pure Storage

Pure Storage is a $2 billion enterprise storage company that provides block-, file-, and object-based storage infrastructure delivered both as appliances and as cloud-based services. It was one of the first to introduce the use of all-flash storage for mixed enterprise workloads and has been one of the most successful of all the all-flash array (AFA) start-ups. A large number of Pure Storage's customers use virtualization technology from VMware, and interest in and use of containers are growing rapidly in the AFA vendor's installed base. In September 2020, Pure Storage announced the acquisition of Portworx, a software vendor whose Kubernetes storage platform is widely deployed in the enterprise (i.e., Global 2000).

The Portworx Enterprise Kubernetes Data Services Platform provides a comprehensive data services platform for persistent storage, data availability, security, capacity management, data migration, backup, and disaster recovery in container-based environments. Portworx is a Kubernetes-native platform supporting container-granular, Kubernetes namespace-aware data services that includes six different software components:

- **PX-Store** provides scalable, highly available and persistent storage that is container optimized, enables single-writer or multiwriter shared volumes across multiple containers, provides storage-aware class of service, and supports availability configurations that are node, rack, and availability zone aware.
- **PX-Secure** includes clusterwide AES 256-bit encryption; role-based access control for authorization, authentication, and ownership; and Active Directory and LDAP integration.
- **PX-AutoPilot** provides comprehensive capacity management with automatic resizing of individual container volumes and entire storage pools, a rules-based engine with customization capabilities for application performance optimization, and integration with Amazon Elastic Block Store, Azure Block Storage, and Google Persistent Disk.
- **PX-Migrate** provides multicloud data mobility, simplifying the migration of applications across clusters, racks, and clouds. It supports application-consistent snapshots and enables simple, snapshot-based backup to any cloud.
- **PX-Backup** allows administrators to back up and restore entire applications, including data, application configuration, and Kubernetes objects, with a single click. It supports continuous backups across global datacenters; backup and recovery of cloud volumes from Amazon, Microsoft, and Google; and the observability needed to meet compliance and governance requirements within a single pane of glass in hybrid cloud environments.
- **PX-DR** extends the data protection already included in PX-Store with Zero RPO Disaster Recovery in multisite configurations, providing synchronous and asynchronous replication options as well as the ability to set all disaster recovery policies at the container-granular level.

All of the Portworx Enterprise components are integrated into a complete solution, managed through a single pane of glass, that exhibits highly scalable storage management and is optimized for application-focused operations in a hybrid, multicloud environment. Portworx is integrated with the VMware Tanzu portfolio of cloud-native options to bring enterprise-class storage management functionality to container-based workloads that can be run and managed side by side with VMware VM-based workloads. Portworx provides Kubernetes-native storage functionality on top of any storage system backing Tanzu, including vSAN, vVols, or any VMFS data store, making it a flexible option for heterogeneous multicloud and hybrid cloud deployments.

Challenges

Although use of containers is expected to increase rapidly over the next several years, enterprises are still coming up to speed on this somewhat new technology. Administrators will have to overcome a number of challenges as they integrate the use of persistent storage into containerized environments. Portworx, which has focused exclusively on providing enterprise-class storage functionality to containerized environments, brings a wealth of expertise to this area.
Since the acquisition, Pure Storage has focused on creating more comprehensive integration between its AFA platforms (FlashArray//X, FlashArray//C, and FlashBlade), Portworx, and VMware Tanzu with the objective of simplifying the use of persistent enterprise storage in containerized environments. This integration provides an advantage to customers that want to leverage Tanzu to build and deploy containerized enterprise applications quickly in a hybrid, multicloud environment.

Conclusion

Newer technologies such as containers offer significant promise for enterprises seeking to improve IT agility. Deploying containerization will require that most IT organizations implement a persistent, enterprise-class storage solution that can meet stringent performance, availability, security, migration, data protection, and disaster recovery requirements — all while providing highly scalable and efficient capacity management. Given that most enterprises are highly invested in VMware virtualization technologies, enterprises want an approach that allows them to run VMs and containers side by side as they move more in the direction of cloud-native computing. In September 2020, VMware announced Tanzu as the strategic platform for VMware customers looking to add container-based workloads to their existing virtualization infrastructure in hybrid, multi-public cloud environments.

With a hybrid cloud environment built around VMware virtualization, Tanzu, Portworx Enterprise, and Pure Storage AFAs, enterprises will be able to move new container-based workloads to production quickly and reliably.

For VMware customers interested in quickly adding container-based workloads to their existing virtual infrastructure, Pure Storage offers Portworx Data Enterprise, a Kubernetes-based solution that provides the needed functionality and maturity. Portworx is highly scalable, application focused in its management paradigm, and optimized for a hybrid, multicloud environment. It also addresses the key data services requirements for enterprise-class container storage — persistent storage provisioning, high availability, security, capacity management, data migration, backup, and disaster recovery — with a platform that is more widely deployed in the enterprise than any other. With a hybrid cloud environment built around VMware virtualization, Tanzu, Portworx, and Pure Storage AFAs, enterprises will be able to add new, container-based workloads to their overall IT infrastructure quickly and reliably and without compromising on enterprise-class capabilities. IDC believes demand for containerization will continue to grow, and to the extent that Pure Storage can address the challenges described in this paper, the company has a significant opportunity for success.

About the Analysts

Eric Burgener, *Research Vice President, Infrastructure Systems, Platforms, and Technologies Group*

Eric Burgener is Research Vice President within IDC's Enterprise Infrastructure practice. Mr. Burgener's core research coverage includes storage systems, software and solutions, content infrastructure, quarterly trackers, and end-user research as well as advisory services and consulting programs. Based on his background covering enterprise storage, Mr. Burgener's research includes a particular emphasis on solid state technologies in enterprise storage systems as well as software-defined infrastructure. He was awarded the Alexander Motsenigos Memorial Award for Outstanding Innovation in Market Research in 2017 by IDC and is an active participant in the IT Buyer's Research Program at IDC.

Lucas Mearian, *Research Manager, Infrastructure Systems, Platforms, and Technologies Group*

Lucas Mearian is responsible for qualitative and quantitative research with a primary focus on Software-Defined Infrastructure within IDC's Enterprise Infrastructure practice. Topics covered include storage software, software-defined storage, and hyperconverged infrastructure. Mr. Mearian also collaborates with IDC analysts covering enterprise software and infrastructure markets such as data protection, storage systems, and servers. Prior to joining IDC, he was a writer and editor for Computerworld for 18 years covering storage and other topics.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com