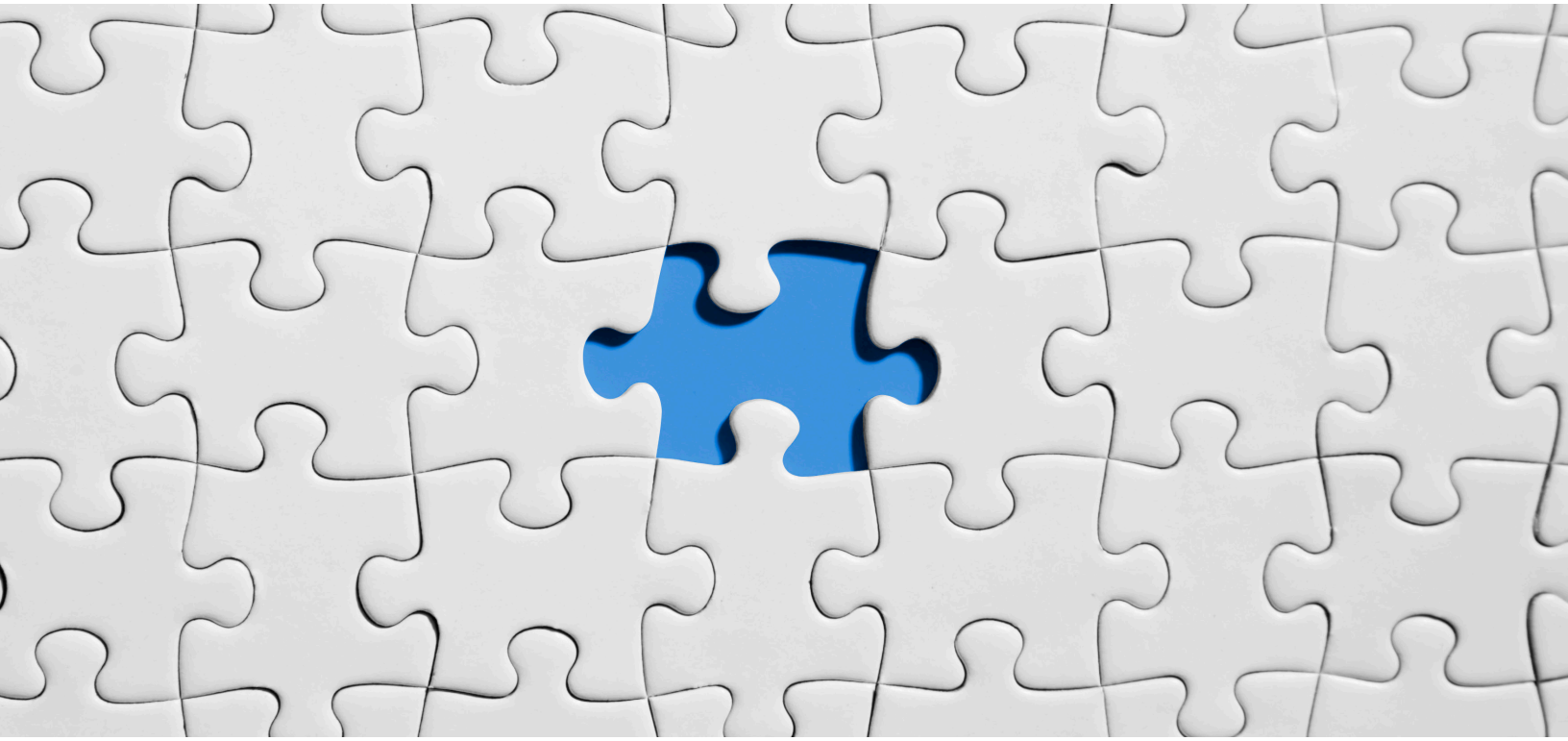
DELL Technologies

DATA PROTECTION FOR CONTAINERS



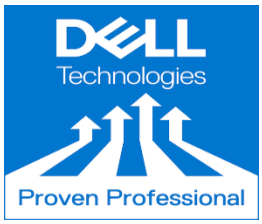
Naveen Chandrashekar

Director, Business Process Management

Naveen.chandrashekar2@dell.com

Knowledge Sharing Article

DELL Technologies Proven Professional



The Dell Technologies Proven Professional Certification program validates a wide range of skills and competencies across multiple technologies and products.

From Associate, entry-level courses to Expert-level, experience-based exams, all professionals in or looking to begin a career in IT benefit from industry-leading training and certification paths from one of the world's most trusted technology partners.

Proven Professional certifications include:

- Cloud
- Converged/Hyperconverged Infrastructure
- Data Protection
- Data Science
- Networking
- Security
- Servers
- Storage
- Enterprise Architect

Courses are offered to meet different learning styles and schedules, including self-paced On Demand, remote-based Virtual Instructor-Led and in-person Classrooms.

Whether you are an experienced IT professional or just getting started, Dell Technologies Proven Professional certifications are designed to clearly signal proficiency to colleagues and employers.

Learn more at www.dell.com/certification

Table of Contents

Preface

We are moving quickly toward infrastructure transformation that is converged, software-defined and virtualized. This transformation spawned new technologies such as Docker and Kubernetes which run applications on stateless containers. Even with this rapid change in technology, there will be some unforeseen, random events would cause data loss or data unavailability.

Data Protection requirements are growing rapidly and every technology needs to be future-ready. Today, while data protection is largely done in the traditional way with a backup server and backup storage, software-defined solutions are emerging, wherein data is protected in both the virtualized environment, as well as in the cloud.

Thus, one should have very good backup and disaster recovery plans to protect the containers from disasters.

Introduction to Virtualization

Today's IT runs on virtualization. However, due to server limitations, resources are not able to fully utilize each of the servers causing them to operate at a fraction of their volume. This leads to increased cost, ineffective management and monitoring, and protecting the data from these physical servers.

Virtualizing these servers enables shared server resources from a single server to run multiple Operating Systems and multiple Applications.

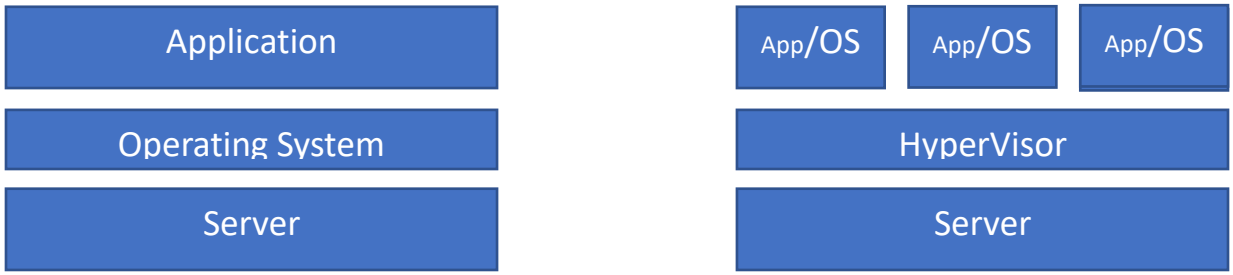


Figure1: Traditional vs Virtualization

Why Virtualization?

Virtualization helps to increase flexibility and scalability, efficiently manage infrastructure, and significantly reduce cost.

Additional benefits include:

- Reduced investment and operating costs.
- Minimized interruption.
- Increased IT productivity, efficiency, agility and responsiveness.

Types of Virtualization

Below are some of the most common virtualization types and how they help today's businesses.

Hardware Virtualization

The most common type of virtualization today, hardware virtualization is made possible by a virtual machine (VM) manager – the "hypervisor". The hypervisor creates virtual versions of computers and operating systems and consolidates them into one large physical server so that all the hardware resources can be utilized more efficiently. It also enables users to run different operating systems on the same machine simultaneously.

Application Virtualization

This is a process where applications are virtualized and delivered from a server to the end user's device, such as laptops, smartphones, and tablets. So instead of logging into their computers at work, users will be able to gain access to the application directly from their device, provided an Internet connection is available. This is particularly popular for businesses that require use of their applications on the go.

Desktop Virtualization

Like the application virtualization mentioned above, desktop virtualization separates the desktop environment from the physical device, configured as a "virtual desktop infrastructure" (VDI). A big advantages of desktop virtualization is that users can access all their personal files and applications on any PC, meaning they can work

from anywhere without the need to bring their work computer. It also lowers the cost of software licensing and updates. Maintenance and patch management are simple, since all the virtual desktops are hosted at the same location.

Network Virtualization

Network virtualization combines all physical networking equipment into a single, software-based resource. It also divides available bandwidth into multiple, independent channels, each of which can be assigned to servers and devices in real time. Businesses that would benefit from network virtualization are those that always have many users and need to keep their systems up and running. With the distributed channels, your network speed will increase dramatically, enabling services and applications to be delivered faster than ever before.

Storage Virtualization

This type of virtualization is very easy and cost-effective to implement, since it involves compiling your physical hard drives into a single cluster. Storage virtualization is handy when it comes to planning for disaster recovery, since the data stored on your virtual storage can be replicated and transferred to another location. By consolidating your storage into a centralized system, you can eliminate the hassles and costs of managing multiple storage devices.

Now let's explore a different virtualization technology – containers.

Containers are not a new technology or advancement of Virtualization; it was there already present in UNIX and then in LINUX, which were part of process isolation.

Why Containers?

Imagine needing multiple versions of applications for testing or production. You would need multiple Virtual Machines which are running multiple iterations of applications with necessary binaries and libraries. This would be challenging as moving around large amounts of data limits VM mobility.

The Containers will run on the same base operating system with abstraction at the application layer that packages app version and dependencies together. Multiple containers can run on the same machine and share the Operating System Kernel with other containers. Each container will be running as an isolated process in the User Space.

Containers take up less space than Virtual Machines, which take up to tens of GB; containers take typically tens of MBs in size and handle multiple applications.

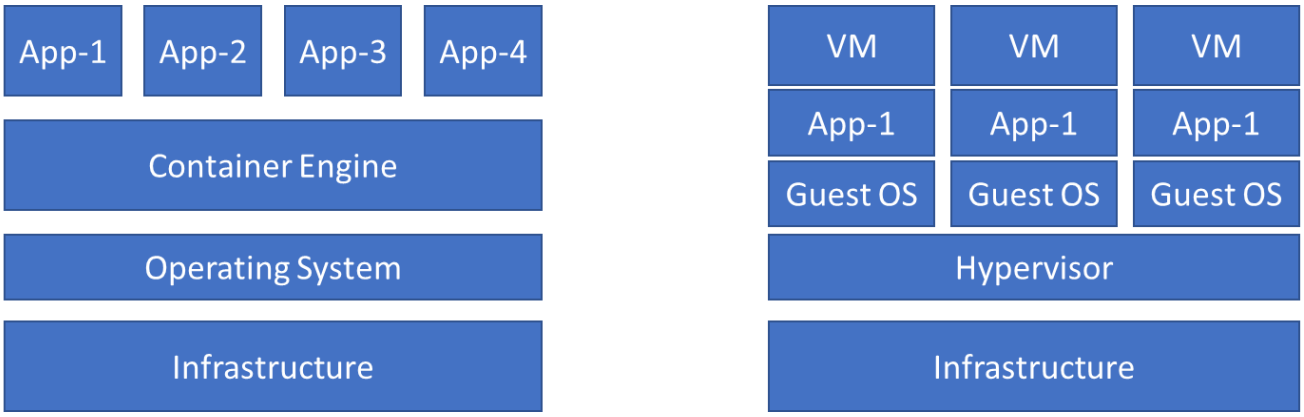
Containers vs Virtual Machines

Containers

- Shared Operating System
- Small Image Footprint in MBs
- Quick Start times
- Stateless
- Easily transportable

Virtual Machines

- Separate Operating System
- Large Image Footprint in GBs
- Full Boots
- Stateful
- Not easily portable, requires exports/conversions/etc.



Containers **Virtual Machines**

Figure 2: Container vs Virtual Machines

Virtualization and Data Protection

Data is growing rapidly and exponentially and organizations are expected to store more data at less cost. Data availability is key; instant access of data from anywhere at any time is expected and should be possible. The data should be protected at any cost, as the environment is mostly data driven and each data is important.

Innovative business continuity software for virtual environments eases backup of virtual servers by adding software between hardware and OS.

The basic backup works by using the native snapshot and cloning technologies built into native hypervisor. The flexibility of virtualization is that it helps encapsulate the OS, applications, accelerating backup and restore operations. The entire encapsulated image can be backed up or restored, or only specific files can be backed up and restored.

This helps us in disaster recovery planning, which results in faster RTO and RPO. Organizations must spend more money in traditional environments to realize tighter RPO and RTO. Meanwhile, a virtual environment helps organizations perform testing and managing servers remotely and enables disaster recovery with minimal impact to the operations at affordable cost.

Containers and Kubernetes

Kubernetes will be used to manage and monitor Containers. Containers and Kubernetes gained popularity since Google released it to the open-source community half a decade ago.

> "58 percent of developers report that their businesses are presently using containers or plan to use containers in the next 12 months." Forrester. "Forrester Analytics Global Business Technographics Developer Survey," 2018.
>
> "90 percent of all applications will feature microservices architectures by 2022". IDC. "IDC FutureScape: Worldwide IT Industry 2019 Predictions," October 2018.
>
> "71 percent of more than 200 enterprise decision makers surveyed indicated they were using Kubernetes to manage their container infrastructure". 451 Research. "Hybrid Cloud Drives Growing Container Production Use and Disruption, May 2017

So, to provision, automate or scale out Containers we would need an orchestrator to help manage containers. Orchestration tools expand container functionality, helping to manage multiple containers with scheduling, cluster management and resource provisioning.

Kubernetes – the most widely used orchestration platform – provides:

- Service discovery and load balancing
- Storage orchestration
- Automated rollouts and rollbacks
- Automatic bin packing
- Self-healing
- Secret and configuration management

Data Protection for Containers

Containers are good at running multiple applications with less overhead than virtual machines. But what about protecting the data on containers? Do containers need data protection and management? The response is yes indeed; Containers needs to be protected.

But typically, a container does not have to have its running state protected, as most of the containers are stateless. Since High Availability is built into every part of the container infrastructure, containers are always spawned and killed as needed.

Many confuse High Availability and Disaster Recovery. The reason for backup would be cluster, container nodes and associated persistent storage.

There are several types of container backup such as mounting, plug-ins, traditional backup applications, volumes and scripting.

The 3-2-1 backup rule still applies to Containers as well. Hence, the data will be safe in almost any scenario. The rule is to keep at least three (3) Copies of data, store two (2) backup copies on different storage media and one (1) of them located offsite.

For Containers running database, store database outside of the container and store the database in the container volume, which would give crash-consistence backup. However, for application consistence backup, we must quiesce the database and take a snapshot.

Also deploying traditional backup software into the container and backup application data to persistent storage. would consume a lot of resources and compromise container efficiency.

Conclusion

Though it may seem that data protection is not needed for containers, critical parts of the Container infrastructure must be backed up to protect against any disaster in the data center.

References

https://www.snia.org/sites/default/files/DPCO/Data%20Protection%20BP%20White%20Paper%20Final%20v1_0.pdf

https://www.snia.org/sites/default/files/SDCIndia/2019/PDF/6%20-%20Application-Consistent-Backup-for-Containerized-Apps-JACOBA-SNIA-2019.pdf

https://www.networkworld.com/article/3511584/do-containers-need-backup.html

https://www.snia.org/sites/default/files/CSI/Containers_Best_Practices_and_Data_Management_Services_Final.pdf