Simplifying Storage Operations
By David Strom (published 3.15 by VMware)

Introduction

There are tectonic changes to storage technology that the IT industry hasn't seen for many years. Storage has been evolving since the early days when those miniscule 5 MB hard drives were first put inside a personal computer in the mid 1980s. But these latest changes are nothing short of revolutionary. In this eBook, we will describe where we have come from, how software defined storage and data centers have become important, and how IT teams need to evolve their staffing and skills to handle these new technologies. We'll also look at the newer approaches to storage and how modern storage management methods are being created and deployed to be more cost-effective.

A brief history of storage

Today it is hard to purchase a desktop or laptop computer with anything less than a terabyte of storage, and we take it for granted that we will have plenty of room for our files. And there is a wide range of network-based storage options too. But back in those early days of 5 and 10 MB hard drives, storage decisions for most IT staffs were fairly simple. There were rotating magnetic disks and files were stored according to the way the operating systems of the time dictated. There were no shared storage volumes, other than on mainframe computers. There was no Internet, no cloud, no storage arrays or storage appliances.

But almost from the beginning, users wanted to share their hard drives, or at least the files on these drives. The earliest form of this was network-attached storage (NAS), which were popularized by the first network file servers from 3Com and others. Then came storage area networks (SANs), which recorded files in blocks of storage and made use of higher-speed networking technologies. SANs were a breakthrough at the time because a collection of hard drives could be shared across a variety of applications, operating systems, servers and desktops.

As the growth of SANs and NAS arrays continued, storage blossomed into a variety of sub-species beyond the simple magnetic rotating disk. Now there are storage units using server DRAM memory, flash devices and high capacity low cost media. And with the rise of the Internet, storage moved beyond the local area network and into offsite and cloud-based storage units too. Today's storage devices are very specialized and often very expensive tools, with some products combining different technologies to improve performance and add redundant operations.

Complicating factors with network-based storage hardware

Today it is hard to find an IT shop that doesn't have at least one SAN as part of its storage solution. But SANs have had their limitations too. By abstracting the physical file to a logical storage unit, they removed the ability to have end-to-end visibility of how data gets stored on the disk, making performance improvements more complicated. Storage frameworks were constructed back when desktops ran their own OS and took control over storage resources, making the transition to virtual desktop infrastructures with their specialized storage needs more difficult.

Managing storage was often a **very targeted skill set**, and often IT departments segregated storage staffs from desktop, overall computing services and networking support teams. This resulted in a lack of integration of storage management tools with the remainder of IT operations. However, today's IT departments are learning to bring compute, storage, and networking teams together. Virtualization is removing the differences among these "silos" and making them leverage and work with each other.

While virtualization has taken hold in larger enterprises, a recent Gartner survey found that virtualization has even permeated mid-sized enterprises too. "On average, midsize enterprises have approximately 110 servers, 80% to 90% of which are virtualized. These same organizations average 50TB to 60TB of data storage across myriad storage arrays, NAS devices and appliances." And the larger enterprises have even more VMs and petabytes of storage.

As VMs become more pervasive, we need storage tools that allow IT to copy and migrate them from one hypervisor to another, and to do so quickly and effortlessly. We also want to incorporate the ability to make quick snapshots and to provide continuous data protection while the VM is running too. Another complication is that storage needs to **scale up quickly**, from hundreds to thousands of servers, and from thousands to multiple thousands of VMs. This means that the old memory constraints of just a few gigabytes are no longer going to work, and the older one and ten gigabit networks are also looking pretty slow for these tasks too.

Virtualization is just one of many changes happening that impact storage. Another is the cloud, which has been slowly becoming more legitimate as another storage tier for enterprises. In a survey for Information Week magazine last year, they found private clouds in production or testing at 77% of the companies surveyed. Some of them are using the cloud exclusively, while others are building extensive Software as a Service applications that leverage cloud-based storage. This is also changing the nature of what is stored, how often it is backed up, and how it works with existing storage repositories and management applications.

Another challenge is that as more workloads migrate to the cloud, it gets harder to use local network-based storage resource management tools. This means you need different tools to manage your entire collection of physical, virtual and cloud-based storage.

In the past, traditional storage operations tools often focused on sending alarms or noting particular events around faults, failures or connectivity issues rather than predictive analytics or understanding the actual root causes of why an application couldn't access its particular storage repository. And of course each storage vendor had their own toolset that was specific to their own arrays, which made for extreme specialization and prevented other IT staff from becoming facile with these tools.

New storage approaches are also required to handle different kinds of data that is being collected. Now that Big Data, machine-generated data and other unstructured data types have become more important and useful to enterprises., they are changing the way we consume and build out our storage systems. This means that for many IT applications, the notion of storing files is less important that just having access to particular data elements, so storage management has to evolve to understand this too.

Speaking of applications, the old ways of tying data services to a particular storage container are also evolving as our application needs become more complex. The notion of SANs originated with having static and pre-allocated storage blocks that were used to create large amounts of storage before any application was used. Often the SAN was purchased without any regard to the ultimate application, because all apps were considered to have similar storage needs and usage patterns. Those days are over.

When these patterns change, storage containers had to be adjusted, or massive amounts of data had to be migrated to a new container. These adjustments usually had to be done manually, which took IT staff away from other duties and required specialized skills. Or worse, the changes never happened, which often resulted in overprovisioning storage arrays as a cushion in case you needed the extra room.

This meant that you would reserve large blocks of storage that would remain mostly unused. Indeed, overprovisioning used to be considered an industry "best practice" – so a storage array wouldn't run out of room as applications added new data to a volume. However, this just adds to the overall cost of a storage solution, since you are purchasing basically empty space.

In the past, many IT departments made storage decisions based on two factors: they looked at which product would perform the best and which product offered the best solution for protecting their data. Often, these were mutually exclusive requirements, and hardware choices were a compromise between the two factors.

Finally, as enterprises virtualized more and more of their workloads, storage was still setup the same way it was when more physical servers filled the data center. Many storage arrays weren't aware of VMs or how VMs could share a common set of OS and applications-related files that wouldn't change and couldn't recognize them as such. This also wasted a lot of space storing the same file collections.

What is Software Defined Storage?

But storage is evolving to move beyond those early days where large hardware arrays dominated the data center. Today we need more dynamic delivery of storage and more granular control of storage resources. Apps matter, and have different storage consumption profiles. Overprovisioning isn't acceptable; especially as the storage management tools have gotten better at storage allocation.

With all these new approaches and requirements for storage, the old notions of hardware-centric storage are just outdated. There is a need to move the control plane from hardware-centric to more app-centric, and transform storage by better aligning it with the demands of today's applications. And just as we virtualized the compute and network resources and workloads more than a decade ago, we have to do the same with virtualizing the storage workloads to increase throughput and make allocating storage more efficient and effective.

Welcome to the new world of software-defined storage (SDS). This frees up our storage arrays and puts control back in the hands of the very applications that consume storage. And just as software-defined networking adds a layer between the compute and physical infrastructure to make VMs more flexible and useful, SDS adds a layer between the applications and infrastructure so that storage can be more flexible and useful too. SDS is a key component of the software defined data center, and a natural way that applications teams can leverage the skills and work together with compute and networking staffs.

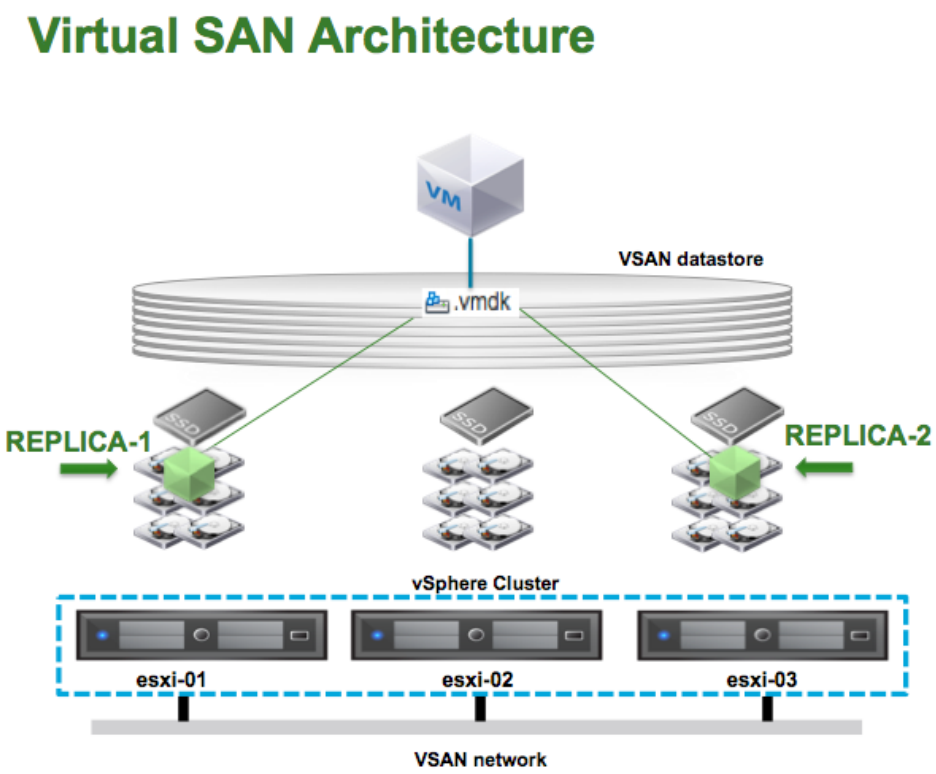
The elements of SDS

SDS has four important elements:

First, by **virtualizing the SAN array itself,** you remove the need for specialized storage hardware. You can repurpose existing magnetic and flash storage devices to handle a variety of application requirements. With the virtual SAN, all storage containers are virtualized. No longer are we locked into logical storage units that can't expand or contract as storage needs dictate. With virtual SANs, you can assign storage to particular VMs and handle them as you would the VM itself. Virtualization brings more elastic scaling to storage, more granular solutions, and more flexibility.

Second, you manage the virtual SAN by using the VM hypervisor. This means that you can have the shortest I/O path for your data to traverse from the storage repository to the VM that needs it. Some IT shops have managed to shave their database response times from minutes to seconds and reduce latencies to less than a millisecond. You can also remove storage appliance overheads, since the storage is being managed directly by the hypervisor itself. Having the storage within the VM hypervisor means you can handle VM migrations and storage migrations easily,

without the need to learn a new management tool. You can assemble a cluster of storage elements, and make replicas easily, as shown in the diagram below.



The hypervisor makes use of storage policies to assign storage to specific VMs, and can **automatically tune and balance storage blocks** to keep service levels as high as possible. With this configuration, you have more granular storage capacity and can still scale up your storage easily without having to provision a new SAN array that can be cost prohibitive.

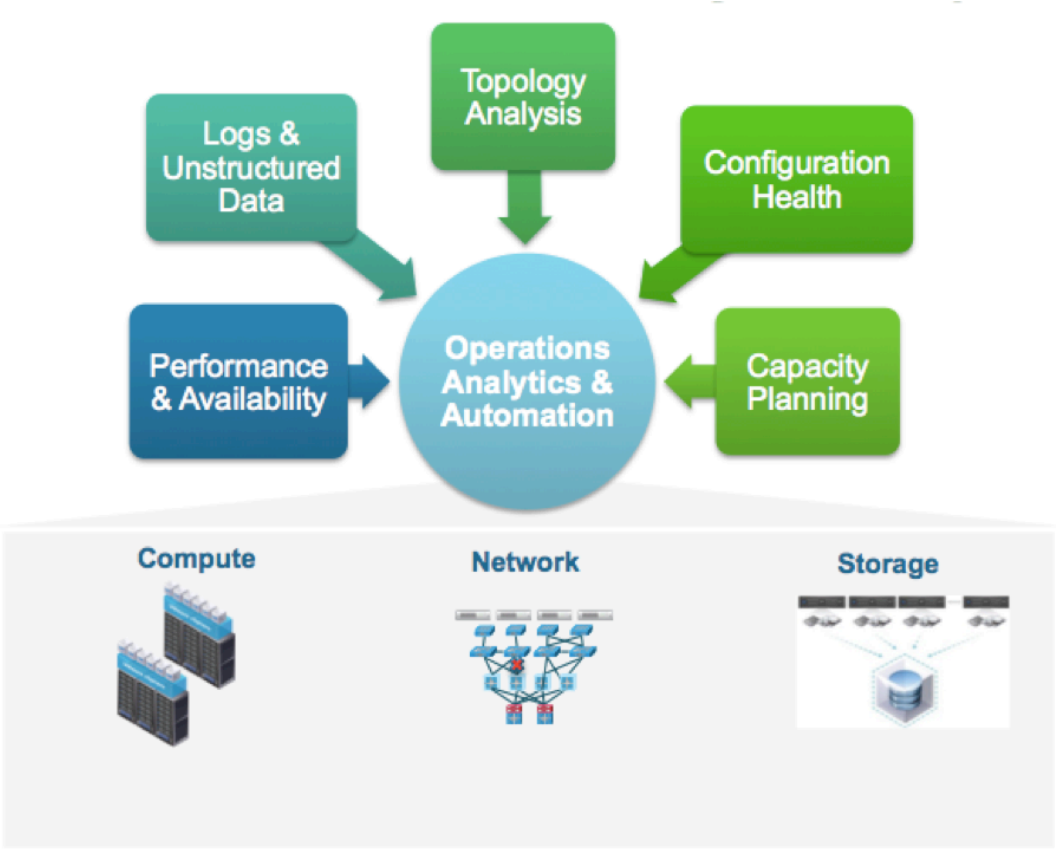
Third, you **combine storage host, fabric and storage repository** together. This has the advantage of leveraging common computing and other components and not requiring expensive and specialized storage hardware.

Finally, with SDS we can **automate storage management tools** and bring together a single solution for performance management, capacity optimization, and real-time log analytics. Using these types of automation tools means we can provide predictive analytics leveraging both structured and unstructured data. Automation can also make for more elastic and granular scaling of storage needs and lower capital investment.

Traditional storage operations tools deliver performance and availability information, capacity planning and topology analysis information. Some of them may have configuration data in addition to the above and very few may also be doing log analysis. Mostly though are just gathering and analyzing this data for

storage components. However, the lines separating issues between compute, storage or networking are quickly disappearing, and the tools are getting better at providing a complete view across all three arenas.

This means faster problem resolution and better root cause analysis to eliminate finger pointing and cut downtime too. Having this cross-domain expertise will make it easier to troubleshoot and maintain storage networks and provide staffing flexibility in how they are managed too.



How VMware can help

Given this move into SDS, VMware has a number of products that can help define the modern software defined data center. Let's look at a few of their solutions.

First is **vSphere v6**, which comes with loads of new features that are focused on storage management, including nearly instant VM cloning and long distance vMotion replication. It manages VMware's **Virtual SAN** so you can provision new storage repositories in just a few mouse clicks. You can snap, clone, encrypt, and troubleshoot your disk storage with the same tools that you use for doing similar operations on your VMs.

vRealize Operations Management Pack for Storage Devices has been available since December 2013 and gone through several revisions. It allows you to rightsize your storage environment, model and manage storage capacity, and gives you better visibility into your storage stack from end-to-end. This means you can track events from the originating VM through the storage fabric to the array and the ultimate consuming application. There is a single user interface to visualize all this information that is split among three categories: health, risk and efficiency. (See screenshot below.) And there are even new controls to handle user access rights too. You can download it from the VMware solutions exchange here.

vRealize Log Insight delivers real-time log analysis to help you better troubleshoot your virtual storage infrastructure along with other IT operations. It collects logs from a variety of sources and aggregates them for predictive and Big Data analysis.

Conclusion, calls to action, product links

The modern data center needs to evolve beyond hardware-centric storage to a more flexible and capable software-defined storage. VMware can help with this transition, just as it has done so for virtualizing computing and networking workloads over the past decade. Check out some of the links to these products and strategies so you can get started with SDS today.

https://www.vmware.com/products/vrealize-operations-insight/

http://blogs.vmware.com/management/2015/02/vsan-simplifying-sddc-storage-operations-with-vrealize-operations-management-pack-for-storage-devices.html

http://www.vmware.com/getthefacts/virtualization-management