

I D C E X E C U T I V E B R I E F

Identifying Storage Management Challenges When Deploying Server Virtualization

April 2008

Adapted from: *Storage Virtualization Comes of Age, Following Server Virtualization*, by Laura DuBois, Richard L. Villars; Doc #204175

Sponsored by Symantec

Introduction

Server virtualization market has been growing at a double-digit rate for the past two years, and is expected to continue growing in 2008 and beyond. The need for consolidation to reduce datacenter footprint and energy consumption, better resource utilization rates for existing resources, and improve time to market for new applications have been driving the adoption. However, the benefits gained through server virtualization have come at a cost.

Newly virtualized environments have encounter several of the following challenges that need to be addressed:

- Management of physical and virtual servers
- Managing storage resources for capacity and performance
- Enabling scalable data protection

These challenges are present in the physical environment and do not disappear when server virtualization is deployed.

The focus of this Executive Brief is on storage management and data protection.

Market Overview

Server virtualization market has been segmented into two major categories: operating system-based versus hypervisor-based virtualization. The most common implementations of the first category are AIX LPAR, Solaris containers and Logical Domains, and HPUX vPAR. The guest machines that are created in these environments are limited to running the operating system of the host, such as Solaris, AIX, and HPUX. In hypervisor-based virtualization, such as VMware VI3, Citrix XenSource, and Virtual Iron, the guest

machines typically run Linux or Windows; within a single physical host guest, machines can run various disparate operating systems. The challenges faced by hypervisor-based architecture will vary slightly at the technology layer, but have fairly similar management challenges.

From a market-share perspective, hypervisor-based virtualization has captured the largest share of the market with nearly 70% of the market. As such, let's look at the two primary architectures and the unique storage management and data protection challenges they face.

Challenges in Managing Storage

There are many challenges associated with managing storage such as reclaiming capacity, aligning I/O performance requirements with capacity consumption, reducing storage costs through storage/service tiering, and provisioning and migrating data in a timely fashion. Other challenges include data protection, application support for high availability, resource optimization, troubleshooting, and general management of assets. In a virtualized environment, these challenges, as well as some unique to server virtualized environments, are exacerbated.

Data Protection

The exponential growth of data and the increasingly shorter backup windows are creating new challenges for managers responsible for protecting an organization's information assets. Unless one considers the impact of virtualization on data protection strategies, these challenges can be made worse in a virtualized environment. A single physical server may have the necessary capacity to successfully complete a backup, but when this one, physical resource is distributed across multiple virtual machines, a backup job can severely hamper the performance of the whole environment. Data protection solutions that off-load backup operations from the server and the individual virtual machine are needed to address this challenge.

Currently, the following two approaches have been employed to resolve this issue:

- Use snapshots to off-load the I/O burden of backups from the physical host or the guest OS, thereby eliminating the impact on the environment. In this scenario, a proxy server would mount the snapshot and perform the backup job to disk or tape.
- Use of efficient backup methodologies such as continuous data protection products to reduce the impact data protection has on the physical host or the guest OS. These products are standalone solutions, however, and are commonly limited to the Windows platform.

Managing Virtual Machine Sprawl

An environment with multiple servers may experience an increase in storage usage for duplicate data sets represented by the operating system. This becomes more of a problem in server-virtualized environments where new virtual machines can be created quickly and easily, but require a substantial amount of disk resources on expensive shared storage.

This scenario poses a challenge, since an OS such as Windows may take up as much as 3 GB to 4 GB. Leaving additional 3 GB to 4 GB overhead for future patches and updates may bring the full allocation of capacity for the OS to 8 GB of capacity per virtual machine (vm). If there are at least 10 virtual machines running on this single server, 80 GB of capacity is required for the vm OS only. If 10 virtual machines on a single host is a conservative estimate, organizations are often running as many as 50 virtual machines on a single server host, then the capacity required to store the vm OS may be as high as 400 GB. To minimize impact of the vm OS on total capacity, features that reduce redundancy can help reign in the sprawl.

There have been a few approaches considered to resolve this problem. The first approach is to use deduplication technology to reduce capacity required to store operating system software. The issues associated with this approach include potential performance limitations that would impact the ability of an environment to support production applications.

The second approach is to use snapshot technology and virtualization to allow multiple machines to boot off the same image with the changes that are necessary for its operation. The downside of this approach is that support for this functionality is limited to a small number of disk arrays.

Another area where storage managers must define new processes and policies is when a machine is EOL (end of life). Physical machines are often EOL, and there is a well-defined process in place that managers follow. In virtual environments, the EOL of a virtual machine is not the same as that of a physical machine. There needs to be a clear definition of how and when a virtual machine is at EOL. Additionally, as in physical environments, physical machines need to be replaced, which effects virtual machine's environment. A physical server for the virtualized environment and virtual machines must have a life cycle that can be easily repeated and that aligns with the business.

Virtual Machine Migration

To facilitate migration of a virtual machine across servers in a server pool or cluster, all physical servers in the pool or cluster must have access to the virtual machine and its data. Shared access can be achieved in a number of ways. If the storage used is network-attached, servers can access data via shared protocols such as NFS. If the storage is FC- or iSCSI-attached, then the approach used

by VMware and XenSource are different. VMware leverages a global file system, VMFS, where each virtual machine is identified as a VMDK file. When vmotion or migration occurs, servers in a pool or cluster access the necessary data through this file system and file designation. In XenSource, the virtual machine that needs to be migrated must reside on a Logical Unit Number (LUN) that is visible to other physical servers. The management component of the solution facilitates and manages access to the LUN.

Storage Provisioning and Allocation

Server virtualization addresses the need to bring applications to market at a faster rate, but what is not often addressed is how to bring storage provisioning and allocation in step with the deployment of a virtual machine. In many organizations, the processes of identifying an available resource and provisioning it may take days, resulting in delays. In virtualized environments, this is made even more difficult.

Defining RAID sets, carving out a LUN, and allocating the LUN to the physical machine and then to each virtual machine while delivering necessary performance and capacity becomes a disk layout challenge. In VMware, VMFS abstracts the raw partitions requiring server administrators to map the file system with VMDK representing virtual machines and their data. The same can be accomplished with network-attached storage (NAS) where the file system of the NAS abstracts raw partitions providing necessary capacity and performance via NFS or CIFS. In XenSource architecture, there is no native file system; server and XenSource administrators follow the same process to allocate storage as in a physical environment and are not limited by a file system.

Storage Tiers and Migration

The way data is accessed, the value of data to an organization, and the relevance of data may all change over time. To optimize storage resources, storage managers have been seeking a way to tier storage and seamlessly migrate data across the tiers to achieve greater efficiency and lower cost of acquisition and ownership. The challenge is that migration is not a simple task. Whether migrating volumes or whole storage arrays, it's important to plan ahead in order to avert any possible delays and disasters. This is a problem for both physical and virtual environments.

Visibility Into the Environment and Troubleshooting

A data packet travels through many components in its path to the disk drive. Imagine a data packet beginning its journey at the application level, being mapped by the file system to a location on a disk drive, exiting through a HBA to speed along through the storage-area network cloud to a storage controller that defines what LUN is part of the volume belonging to the application server, and finally landing on a sector on a disk drive.

The increasing complexity of datacenter environments is making it more difficult for administrators to maintain visibility into their environments, and to identify problems and their root causes. Server virtualization creates an additional level of abstraction, decreasing the visibility and the ability to effectively troubleshoot problems.

Optimizing Storage for Performance and Capacity

CPU are getting faster, memory is getting bigger, but the disk drives haven't changed much in their performance characteristics, making achieving comparable performance more challenging. Additionally, disk drives have increased in capacity from the standard of a 36 GB drive in 2002 to at least 146 GB in 2007.

The combination of both trends, limited performance gain and exponential growth in capacity, has created a challenge for storage managers. There is a need to leverage as many disk drives as available for I/O requirements, but often that is achieved at the expense of capacity utilization. This is further exacerbated by the use of SATA drives that range in size from 500 GB to 1 TB. This challenge is real for both physical and virtual server environments, though virtualization makes the problem stand out more.

Conclusion

Although virtualization has many benefits, including improved resource utilization, lower energy consumption, and faster time to market, it also exacerbates many of the storage management issues that managers have sought to resolve. Virtualization is a layer of abstraction, and with abstraction comes lack of visibility into the environment necessary to make sound decisions about storage.

Additionally, there are challenges specific to virtualized environments, and the solutions must be specific as well. When organizations begin designing virtualized environments, it's critical to consider storage. Storage, and how storage resources are managed, will define the success or failure of your virtualization implementation. The following are areas that should be discussed:

- How does our selected server virtualization platform handle virtual machine mappings to storage, and what additional tools might be needed to optimize the provisioning and allocation process?
- Based on what applications might be running on the virtual machines, what performance is required for each on the storage end and what might be layout requirements necessary to achieve this performance?
- Whether production or test and development, data must be protected, which poses a challenge for many virtualized environment; make consideration for how the data must be

protected in order to ensure timely restores and make adjustments that will facilitate desired result.

- Have all your requirements for storage in place before making a decision to which storage system, software, and services will deliver the desired functionality.

By planning upfront, managers will save time and money and will deliver an architecture and processes that are repeatable and contribute to business' overall success.

COPYRIGHT NOTICE

The analyst opinion, analysis, and research results presented in this IDC Executive Brief are drawn directly from the more detailed studies published in IDC Continuous Intelligence Services. Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from IDC. Contact IDC Go-to-Market Services at gms@idc.com or the GMS information line at 508-988-7610 to request permission to quote or source IDC or for more information on IDC Executive Briefs. Visit www.idc.com to learn more about IDC subscription and consulting services or www.idc.com/gms to learn more about IDC Go-to-Market Services.

Copyright 2008 IDC. Reproduction is forbidden unless authorized.