SNIA

Education

Optimizing Your Storage for Server Virtualization

Chris Lionetti, NetApp

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

© 2010 Storage Networking Industry Association. All Rights Reserved.

Goal of the Session

- A frank discussion about advanced features within storage systems and how they can be used to optimize a virtual server infrastructure. These advanced features include snapshots, flash, SSD, de-dupe, and de-dupe aware cache and how they can solve problems such as boot storms, instant Virtual Machine cloning and deployments, and space optimization.
- Learning Objectives
  - Understand the Advantages/Disadvantages of SAN/NAS/DAS in a Virtual Server environment.
  - Understand how advanced storage features enable economies of scale and optimize assets.
  - Gain an understanding of the part that Virtual Servers play to gain hardware independence for highly availability services and

What this session ISN'T.

- This session is NOT a line by line comparison on different Hypervisors on the Market. Each Hypervisor has unique features and advantages.
  - Not all features discussed in this presentation are available to all hypervisors
- This is NOT a line by line comparison on different features of different storage vendors.
  - Not all features discussed in this presentation are available to all storage arrays
- When in doubt about the definition of a term, please refer to the SNIA dictionary.
  - All terms used in this presentation have been submitted.

© 2010 Storage Networking Industry Association. All Rights Reserved.

All computers IDLE at the same speed.

- Application requirements don't change as fast as the infrastructure does.
  - Most Applications need a single or pair of cores
  - Most Applications need less than 4GB of less of RAM.
- Common <$1500 servers today
  - Have 4 Cores (1 of 2 sockets used)
  - Have 4GB RAM (1 of 18 DIMM slots used)
- The represents a widening gap between application needs and optimized server configurations.

© 2010 Storage Networking Industry Association. All Rights Reserved.

That next floor tile will cost $20, the one after that will cost $100,000.

- Since (most) Application needs are usually low, obsolete servers are well suited to continue operation past the hardware proper lifecycle.
  - If it isn't broken don't touch it.
    - Corollary – Once it breaks you'll never be able to repair it.

- Common incentive to change is the massive retro-fit required when you hit a natural space barrier.
  - A Virtual Server deployment is really a power/space reclamation project for you datacenter.
    - I can already see it….”Trading Spaces – Datacenter Edition”.

© 2010 Storage Networking Industry Association. All Rights Reserved.

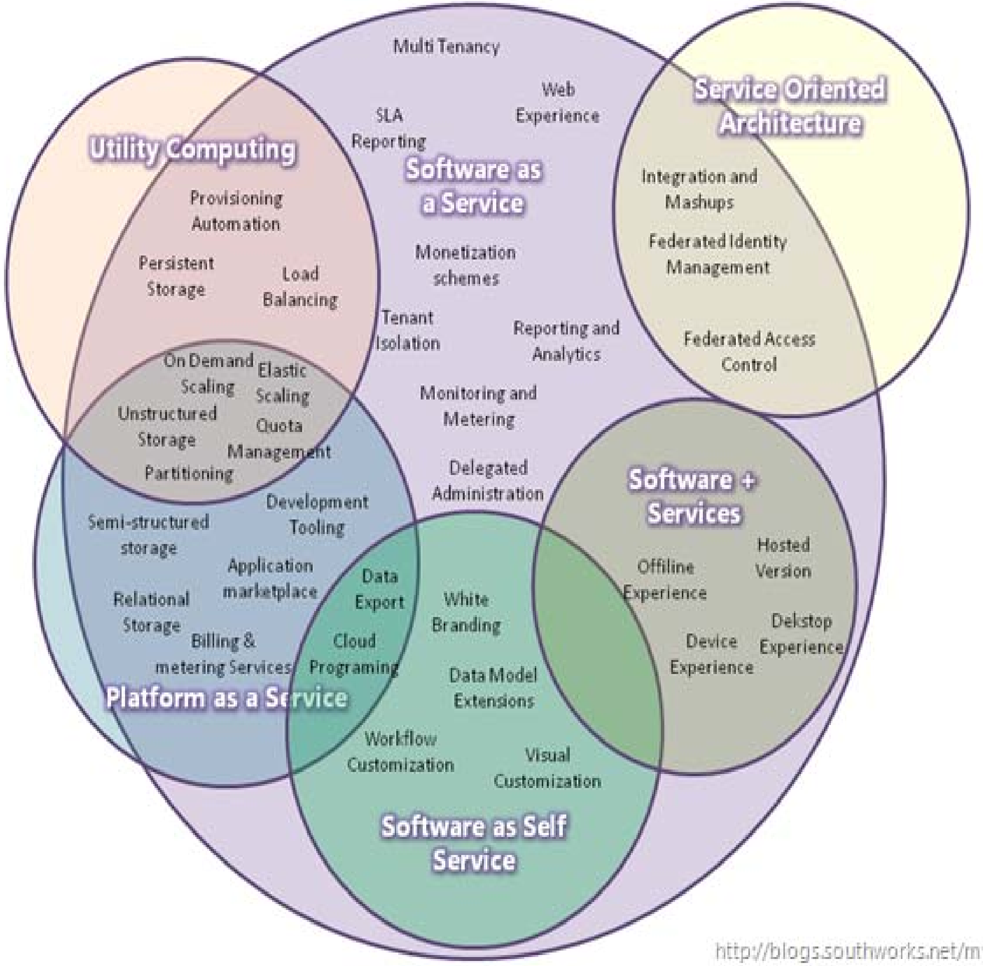
You can’t get there from here.

- How Many of you have seen the Green Datacenter Talk?
  - Neat, huh. Too bad you can’t do most of it.
- You need abilities that non-virtual servers just DON’T have.
  - Evacuate a server without interrupting service.
    - Let some servers go into deep sleep, while other servers are fully utilized.
  - Replace obsolete servers with a more efficient equipment when the business/economic needs justify it.
    - Today these decisions are trumped by inability to bring a service down. We run obsolete because the pain of moving is too great.
- Doing Nothing has a cost and risk associated with it as well.

© 2010 Storage Networking Industry Association. All Rights Reserved.

Wait…..Isn't Cloud the Future?

- Depends what you mean by cloud
  - Cloud is many things to many people
- Will likely require application rewrite.
- Read the fine print (SLA)



http://blogs.southworks.net/mwoloski

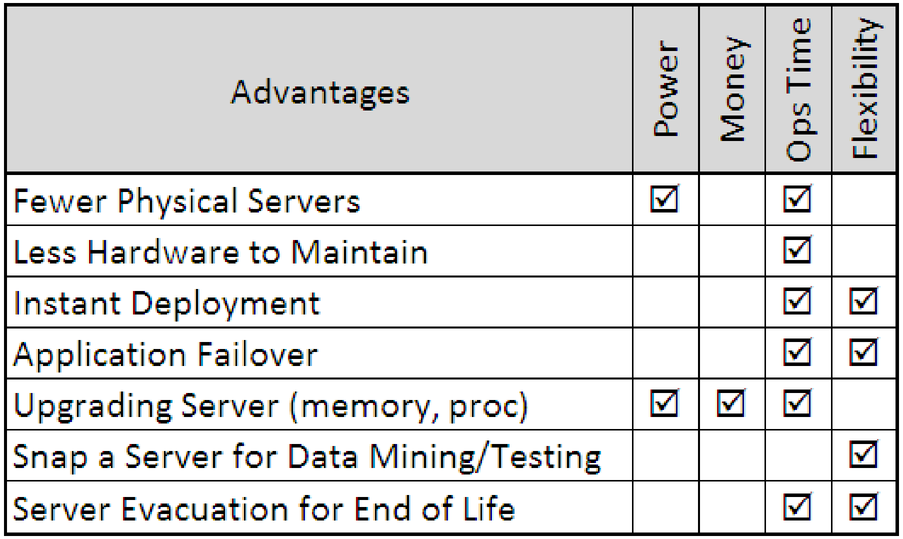
© 2010 Storage Networking Industry Association. All Rights Reserved.

What Everybody Knows.

- There exist a number of commonly known advantages to running virtual servers.
  - We’ve been oversubscribing Switch ports for years.
  - We’ve been oversubscribing SAN based storage for years.
    - When do you oversubscribe the Servers Processors and Memory
    - When do you oversubscribe the Servers HBAs and NICs

- Warning regarding VMs

  Most VMs run NICs and HBAs significantly hotter than physical servers generally do.



Advantages	Power	Money	Ops Time	Flexibility
Fewer Physical Servers	☑		☑	
Less Hardware to Maintain			☑	
Instant Deployment			☑	☑
Application Failover			☑	☑
Upgrading Server (memory, proc)	☑	☑	☑	
Snap a Server for Data Mining/Testing				☑
Server Evacuation for End of Life			☑	☑

© 2010 Storage Networking Industry Association. All Rights Reserved.

"I do not think that word means what you think it means" -Vizzini

- Hypervisor
  - The software layer which allows multiple VMs to exist on a single physical asset. This layer allocates the assets according to the VM config files listed below.
- Bare Metal Hypervisor versus Full Hypervisor
  - Most Hypervisor choices offer a Bare Metal version that is extremely thin and requires fewer patches. Requires more expertise to deploy, but also offers a lower attack vector.
- Virtual Machine Config, Virtual Machine Memory File
  - VM Config – Stores machine state type information. May include such things as the type of processor being emulated, the enumerated list of virtual devices present in the virtual machine, etc.
  - VM Memory File – Contains the stored contents of system memory for a virtual machine, might be used to allow a VM to be paused without restarting, or to be moved from physical server to physical server.

© 2010 Storage Networking Industry Association. All Rights Reserved.

"I do not think that word means what you think it means" -Vizzini

Emulated / Synthetic / Mapped Devices

- You can map a pNIC directly to a VM, or you can create a vSwitch from a pNIC, then map a vNIC to that vSwitch. A synthetic device is one that is optimized for a VM and offers enhanced performance over emulate drivers. A directly mapped device is another alternative.

Virtual Hard Drive vs Direct Access Device vs NAS

- You can allow the hypervisor to virtualized a device and create a VHD file and present that virtual device to the VM.
- Alternately you can map the VM directly to the device (VMDK, Pass-Through).
- On some hypervisors the VHD can be accessible to the hypervisor via a NAS share instead of via direct block access.
  - A last method exists to map the device directly to the VM inside the VM instead of having the hypervisor do the mapping.

© 2010 Storage Networking Industry Association. All Rights Reserved.

“I do not think that word means what you think it means”-Vizzini

Parent/Host OS versus Guest/Child OS

- The Host OS is where the Virtualization Hypervisor management exists. It could be called the Management OS, or the Parent Partition, or the Host OS. This Host OS can then orchestrate the many Guest VMs that will share the physical system/resources.

Management versus Storage versus VM Network

- The Guest OS will generally live on a VM Network that hosts customer traffic. The Management LAN is generally reserved for Management traffic.
- A vSwitch may be configured to support many customer VLANs, and each host could use VLAN tagging. Each VM could be isolated to a dedicated customer VLAN outside of the control of the VM.
  - Buzzword Bingo = Secure Multi-Tenancy

© 2010 Storage Networking Industry Association. All Rights Reserved.

Protocols, Tigers and Bears, oh my…

- Hypervisors can connect many ways
  - iSCSI via Software Initiator, iSCSI via TOE, iSCSI on CEE
  - FC, FCoE on CEE
  - CIFS/NFS
  - Beware the details/Fine Print.
    - Default behavior inside the hypervisor changes depending on how you connect.
      - i.e. Hypervisor Z supports application aware snapshots, but only if using VHDs.
      - i.e. Hypervisor X supports Thin Provisioned by not zeroing disks, but only via NFS
      - i.e. Hypervisor W supports virtual FC HBAs inside a VM, but Hypervisor Y doesn't
- Don't get suckered in, Hypervisors are usually Free.
  - it's the management platforms that cost you money.

© 2010 Storage Networking Industry Association. All Rights Reserved.

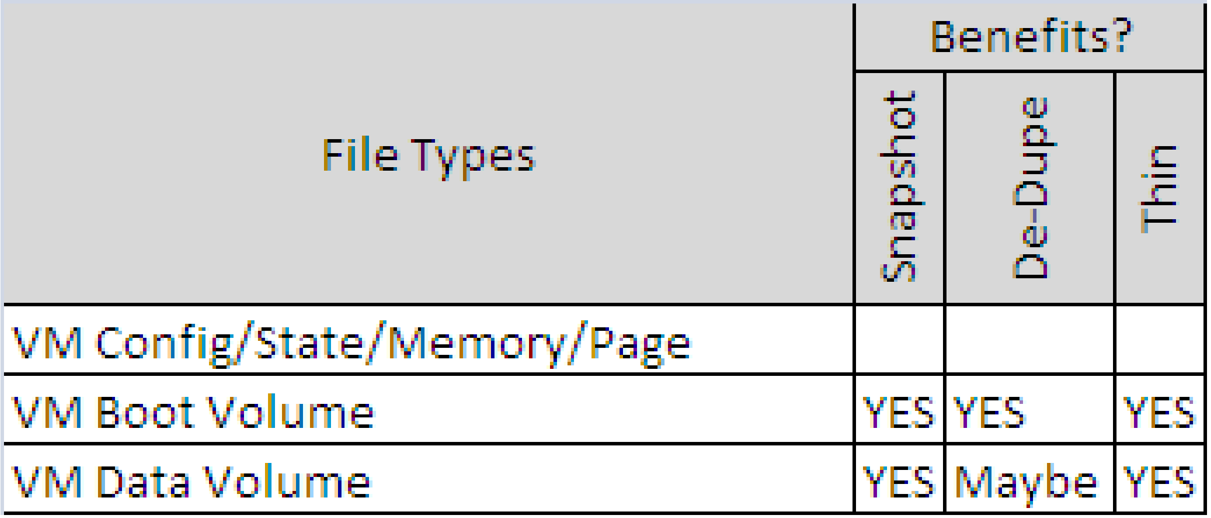
Choose Wisely…..

- Consider carefully which Hypervisor you choose.
  - Some choices about how to store your VMs may lock you into a vendor. i.e. VHDs are an industry standard, as are RAW devices mapped to a VM.
- Don’t assume that choice you make today will be where you are 5 years from now.
  - Consider carefully not only how you get into a hypervisor but how you escape that hypervisor with your VMs intact.
  - Many Hypervisors have system management software that enable you to convert the competitors VMs over.
  - Understand the licensing aspects of the hypervisor of choice. VMs can save significant fees if considered up front.

© 2010 Storage Networking Industry Association. All Rights Reserved.

“Ya Gotta Keep’em Seperated” -Offspring

- Consider the different datatypes and load.
  - Separation of the data types allows a Storage Administrator realize benefits otherwise not obtainable.



File Types	Benefits?		
	Snapshot	De-Dupe	Thin
VM Config/State/Memory/Page			
VM Boot Volume	YES	YES	YES
VM Data Volume	YES	Maybe	YES

Send in the Clones….

- Your Target device should support snapshots.
  - These snapshots will be the basis for your VM Templates.
  - A Snapshot and remap command to the target array will allow you to deploy a new fully functional VM in seconds.
  - The more Snapshots you can make from a template, the less physical storage your VMs will require.
  - The more snapshots you can make from a template the more likely a set of blocks that is needed for a VM to boot will be in your arrays cache memory.

© 2010 Storage Networking Industry Association. All Rights Reserved.

Data Storage on a Diet

- Consider deploying your VMs using Thin provisioning.
  - Boot drives are rarely ever filled to capacity, but easier to ask to for too much and not need it.
    - So create the VM Boot drives as Thin Volumes, and don't zero them on creation.
    - Verify that your array if it supports Thin and Snap that is supports them together.
  - If your array doesn't support Thin Provisioning, some Hypervisors.
    - Research : Dynamic VHD, Eager Zeroed Thick,
    - Warning ; Never Defrag a Thin Provisioned Volume or a Snapshot as it will greatly expand the allocated space.
    - Operationally 90% of servers need ½ of their storage over their life. 5% of the servers will have been right sized, while 5% will need 8x their original storage.

© 2010 Storage Networking Industry Association. All Rights Reserved.

Storage nibbles itself larger

- Snapshots grow over time.
  - Imagine a pack of VMs all receiving a needed patch.
  - You need a method to reconcile those snap changes back together
    - You can either update the template and recreate those new VMs (hard, ugly)
  - Different storage vendors have different approaches.
    - KNOW the implications of the method they ask you to follow.
- Thinly Provisioned LUNs grow over time
  - Does your storage vendor have a method to support LUN Shrink?
  - Does your Hypervisor allow for LUN Shrinking a Volume?
  - Does the Guest OS allow for a LUN shrink? Can it be done live?

Pumping out the Ocean

- De-Duplication on the array allows for a much smaller working data set.
  - This new smaller data set can allow for a different class of drives
    - Where SATA were allocated before, higher performance SAS or FC drives may now make more sense.
    - Where SAS or FC drives were allocated before, higher performance SSD type drives may make more sense.
- A much higher percentage of a VM boot drive may be able to live in the Array cache.
  - If the Array cache is Snap/Thin/De-Dupe aware, those blocks will have an very high hit ratio.
  - A high hit ratio on re-used blocks means those blocks will stay in cache longer. The cache is satisfying hot blocks more efficiently.

© 2010 Storage Networking Industry Association. All Rights Reserved.

The Grey cloud over the silver lining

- Boot Storms are Scary.
  - A Boot storm is where a large set of VMs are all trying to simultaneously boot from the SAN.
    - Only three ways exist to prevent boot storms
      - Boot to local DAS type volumes – and forgo most of the benefits of VM on SAN
      - Back the Array with a MASSIVE number of spindles to handle any boot storm
      - Ensure that the majority of the blocks needed for the Boot Volumes are services from read cache instead of disks.
    - By utilizing Snapshots, Thin Provisioning, and De-Dupe, you can shrink the number of blocks required to solve the boot storm.
    - Increasing System Read cache is another common method to solve a boot storm.
  - While this occurs much more often in a Virtual Desktop Infrastructure (VDI) it can also exist in a server farm.

© 2010 Storage Networking Industry Association. All Rights Reserved.

Just a Flash in the SAN

- Alternate methods to reduce the effects of Boot Storms.
  - Moving very hot or commonly used blocks to SSD assist in servicing many more VMs with fewer physical spindles.
  - Using Flash or large DRAM as a massive read cache extension can also help with satisfying these VM needs.

© 2010 Storage Networking Industry Association. All Rights Reserved.

200% Disk Utilization, but that's Unpossible

- The accepted DAS utilization metric is that storage on DAS is on average 50% utilized.

- The accepted SAN utilization metric is that storage on SAN is 75 to 80% utilized.

- Using Snapshots, Thin Provisioning, De-Dupe as well as template based deployments, you should see numbers well beyond 150% utilization.

    - Common misconception is that 100% utilized is the maximum that you can achieve.

© 2010 Storage Networking Industry Association. All Rights Reserved.

Homework. Aww Man….

- Make sure you read the literature from a Storage vendor on how they integrate with your Hypervisor of choice
- Make sure you consider the process of converting your Physical Servers to Virtual Servers. This feature is embedded in each Management Platform from each Hypervisor vendor.
- Consider a Proof of concept of the features, and make sure you test limits of support.
- Talk to reference accounts and find out if the IT staff, IT budget, and amount of Deployed Storage increased or decreased as a result of the move to Virtual.

© 2010 Storage Networking Industry Association. All Rights Reserved.

Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: trackvirtualizationapplications@snia.org

Many thanks to the following individuals for their contributions to this tutorial.

- SNIA Education Committee

Nancy Clay
Jonathan Goldnick
Rob Peglar

Matias Woloski
Alex Dunn

© 2010 Storage Networking Industry Association. All Rights Reserved.