```
openssl x509 -text -in
```

SNIA NSF | NETWORKING STORAGE

Storage Networking for Virtualization Best Practice Considerations

Live Webcast
January 17, 2019

Today's Presenters

SNIA NSF | NETWORKING STORAGE

J Metz
Cisco

Jason Massae
VMware

Cody Hosterman
Pure Storage

SNIA-At-A-Glance

SNIA-At-A-Glance

170
industry leading organizations

3,500
active contributing members

50,000
IT end users & storage pros worldwide

Learn more: **snia.org/technical** **@SNIA**

SNIA Legal Notice

SNIA NSF | NETWORKING STORAGE

- The material contained in this presentation is copyrighted by the SNIA unless otherwise noted.
- 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.
- 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.

Agenda

- Topics
  - Virtualization Review
  - Multipathing
  - iSCSI
  - Fibre Channel
  - NFS
  - Queuing

What is our **goal** today?

Not to recommend specific settings. Not to talk about vendor X or Y.

But to help you ask the right questions.

RELIABILITY IS KEY

What are the Most Common Storage Issues Support Hears?

Hello This is Support, is This a New or Existing Case?

iSCSI

- Configuration
- Pathing
- Connectivity

FC

- Configuration
- Pathing
- Queuing

- Latency
- Connectivity
- Reliability

Queuing

- Setting
- Mis-Match
- Consistency

NAS

- Configuration
- Version
- Connectivity

Virtualization & Storage

Virtualization Vendors

- VMware ESXi
- Microsoft Hyper-V
- KVM
- Citrix XenServer
- Red Hat Enterprise Virtualization

A variety of offerings—but fundamentally similar

Virtualization

- Virtualization is meant to provide abstraction of physical underlying infrastructure:
  - Network
  - CPU
  - Memory
  - Storage

- Provides flexibility of application/operating system deployment, mobility, efficiency, etc. etc. etc.

Virtualization & Storage

- Common question:

“Hey, you virtualized my database and now my performance sucks”

Does virtualization *really* cause performance issues?

Is virtualization fundamentally *overhead*?

Short answer: No, not really.

Virtualization & Storage

Virtualization is about **SHARED** resources—make what you have used more efficiently.

But because of this, virtualization is designed for **FAIRNESS** by default.

No one should be able to use everything by default.

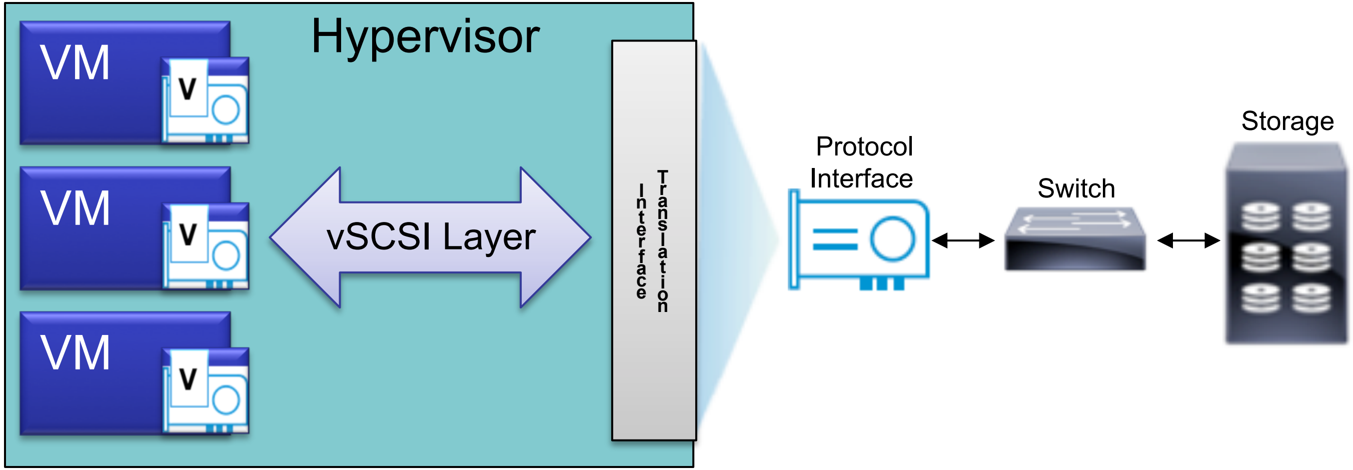
Virtualization & Storage

- This is easier for CPUs, Memory.
- Give VM B 2 CPUs, VM B 4 CPUs etc.

- You can overprovision, but it is easier to tell when that host is 85% full from CPU usage. Or Memory.

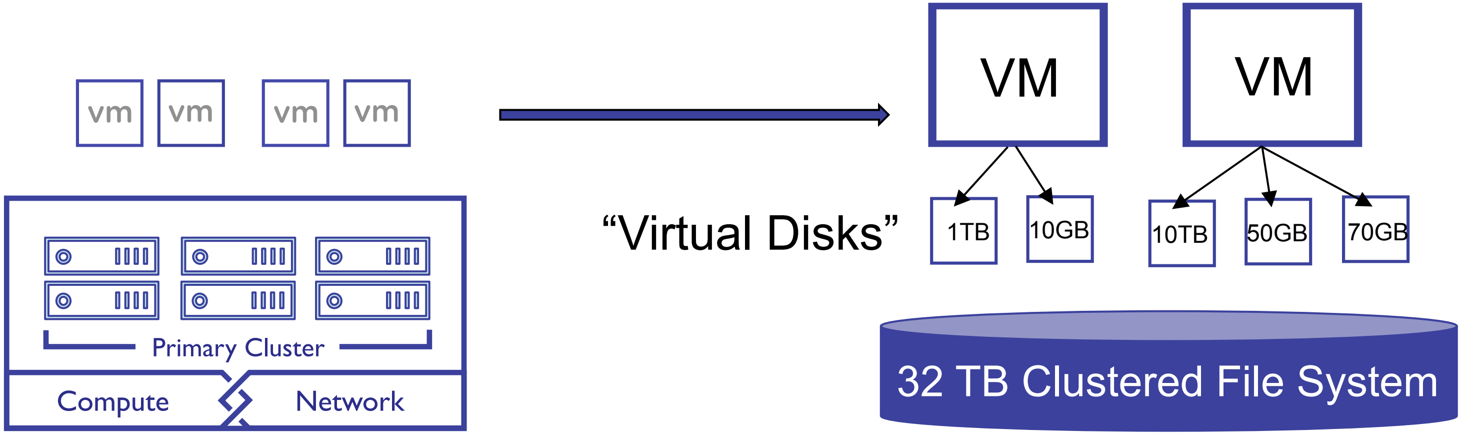
- What about Storage? It is not so straight forward.

Hypervisor Storage Overview



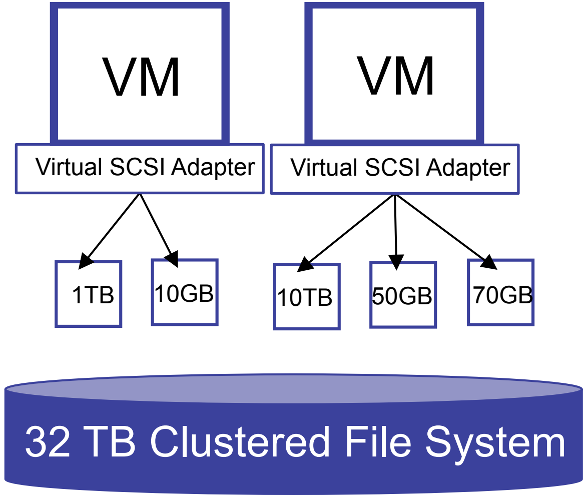
vSCSI Layer

Virtualization & Storage



NFS, or host-formatted block device (FCoE, FC, iSCSI, etc.)

What is a Virtual Disk?



- Key points:
  - A virtual disk is a **file** on a **file system** that looks like local storage to the VM. Hypervisor assigns its own VPD information in that virtual disk
  - To the OS, the virtual disk **looks the same**. Whether that underlying storage is FC, iSCSI, NFS, or something else.
  - Storage array → Network → HBA → File System → Virtual Disk → Virtual HBA → OS

Multipathing

Multipathing

- For best performance, resiliency, use more than one path to your storage.
- More than one HBA, more than one switch, more than one target port.

The basic tenet of redundancy:

No single failure should cause an outage.

Multipathing

- Having additional paths to failover to is **critical**, but using them all at once is **optimal**

There are a variety of multipathing algorithms:

- Fixed Path/Most Recently Used
- Round Robin
- Least Queue Depth
- Latency Optimized

Multipathing

- MRU/Fixed uses one path. Avoid when possible.
- Round Robin is nice, but “dumb”.
- LQD or Latency is generally the best

- Key Points:
  - Multipathing is configured in the hypervisor, not the guest
  - Defaults may not be optimal
  - Talk to your STORAGE vendor for best practices
  - Look into ways to set best practices by default

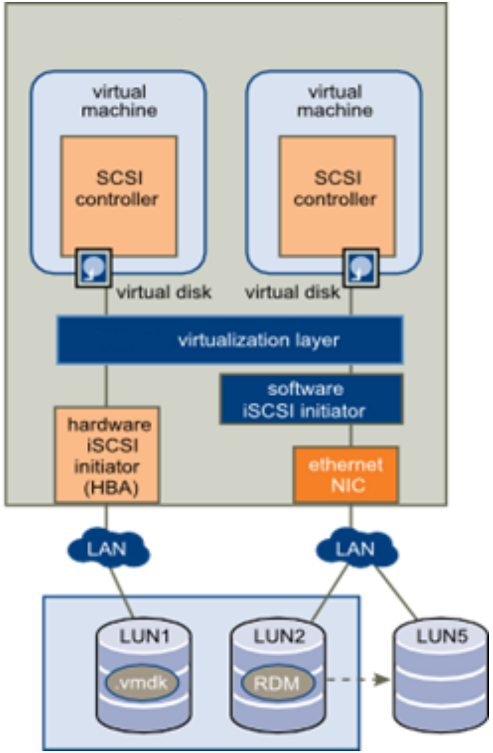
Multipathing

- Consistency (horizontally (e.g. multipathing across hosts), vertically (e.g. MTU))
- Multipathing means something different between block and file at network layer

iSCSI

Avoid iSCSI Issue, Follow the Guidelines.

Especially network configurations!



Network

- Isolated/dedicated traffic (VLANs)
- Port Binding is best practice

Load Balance

- Distribute paths to LUNs among Service Providers

Redundancy

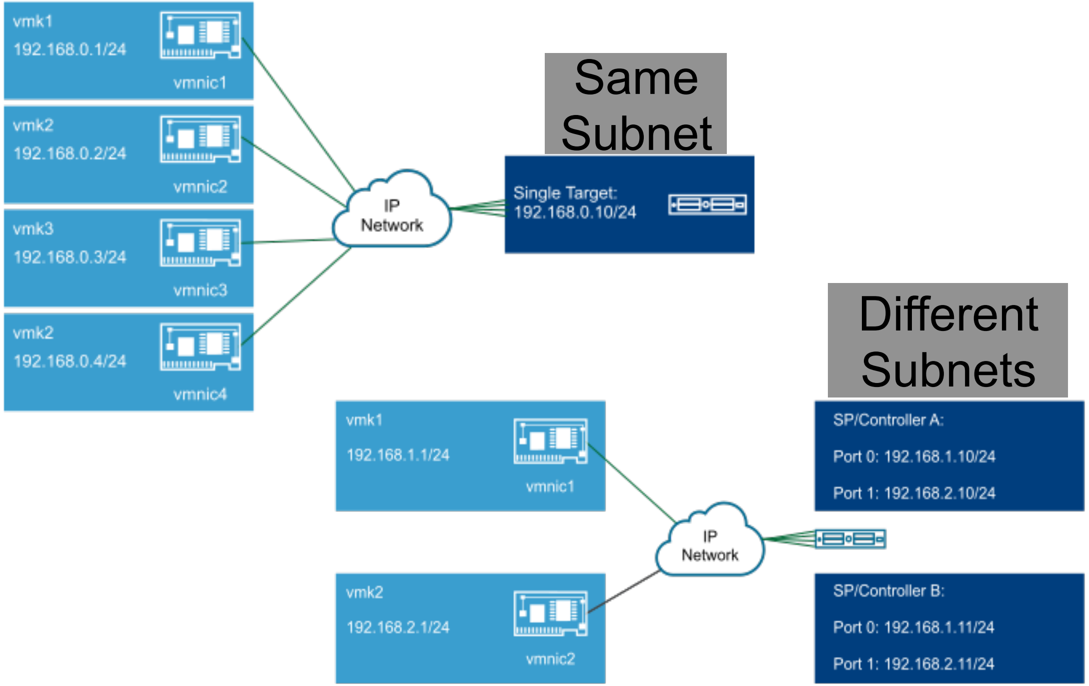
- Make sure initiator is connected to all network adapters used for iSCSI
- Redundant, dedicated interfaces and connections

Storage

- Don’t share LUNs outside virtual environment
- Place each LUN on a RAID group capable of necessary performance

Should You Use Teaming or Port Binding?

Best Practice is to use Port Binding for iSCSI



Same Subnet

Different Subnets

Port Binding

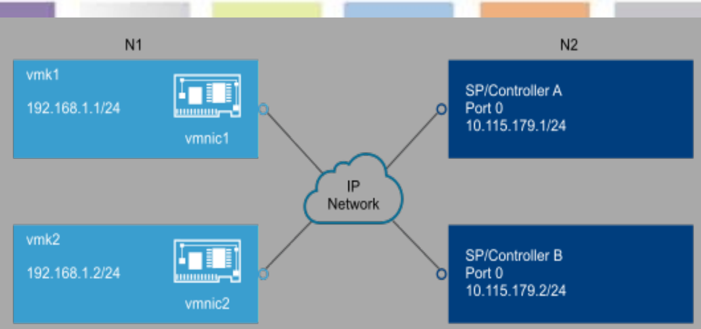
- Fails over I/O to other paths
- Load balancing over multiple paths
- iSCSI initiator creates sessions from all bound ports to all configured target portals

NIC Teaming

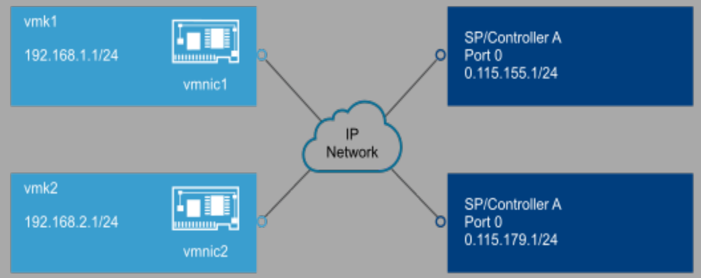
- Array Targets are in a different broadcast domain and subnet
- Only provides fault tolerance at NIC/port
- Use if routing is required
- Able to use separate gateway per vmkernel

Using Separate Gateways per VMkernel Port.

Some hypervisors may allow separate gateways per VMkernel port to be configured.



All bound vmkernel ports in one subnet (N1) and configure all target portals in another subnet (N2)



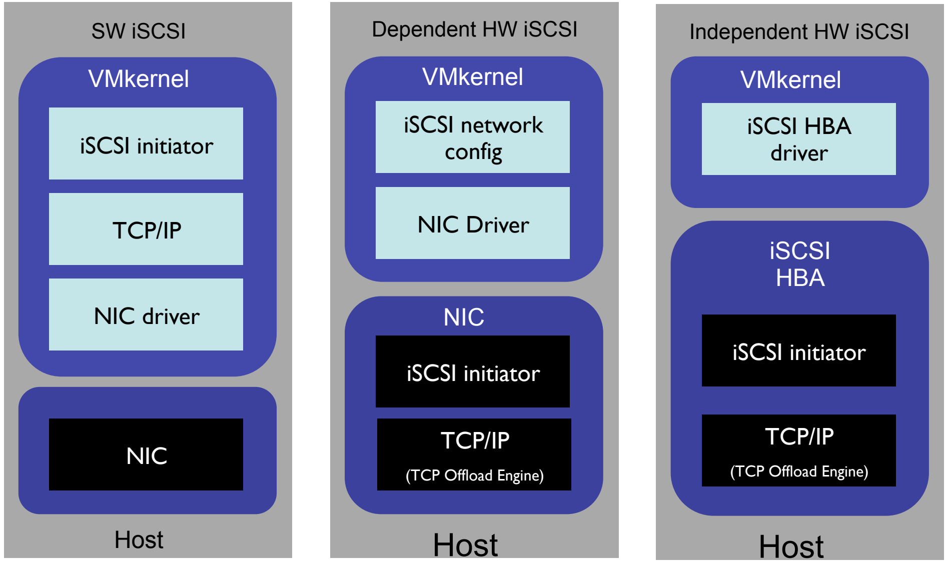
Use static routing when using different subnets

With separate gateways per VMkernel, you use port binding to reach targets in different subnet

- With these configurations you can use port binding to reach targets in different subnets
- You may also configure static routes when initiators and targets are in different subnets

What iSCSI Adapters Can You Use?

There are 3 types of iSCSI adapters: Software, Dependent and Independent.



Standard NIC adapter

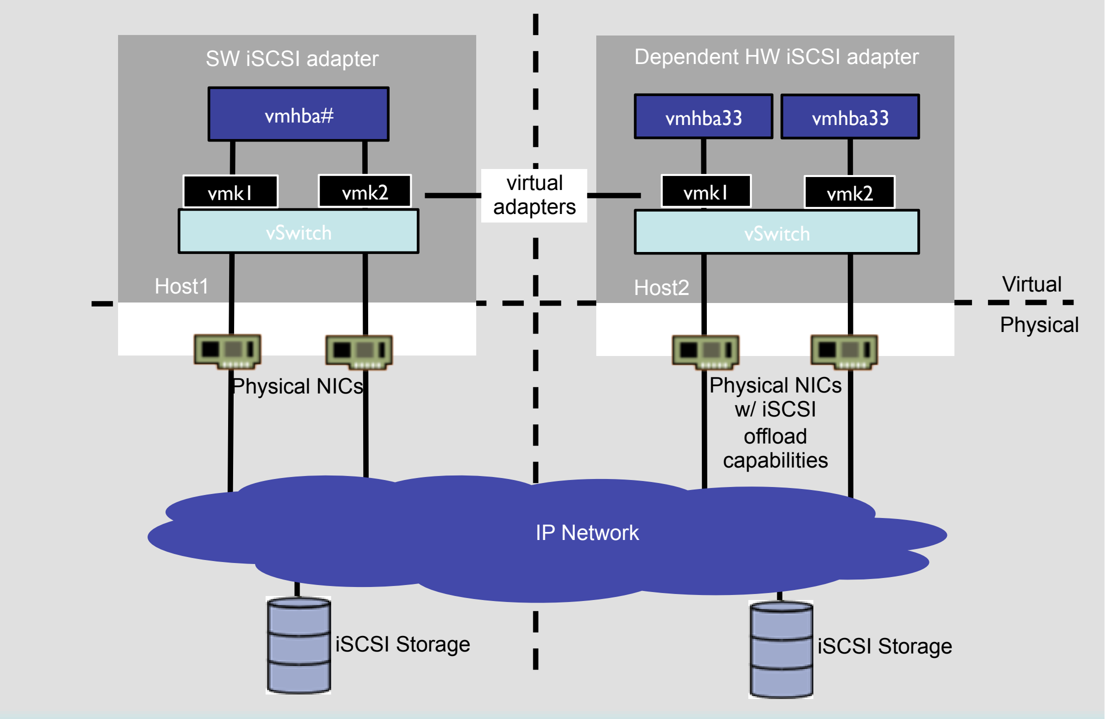
Third party adapter depends on VMware networking

Third party adapter offloads iSCSI, network processing, and management from host

- Mix of SW & HW adapters is not supported
- IPv4 & IPv6 is supported on all 3
- SW iSCSI provides near line rate
- HW iSCSI provide lower CPU Utilization
- Recommend Jumbo Frames (MTU 9000)

How many iSCSI Adapters Should You Use?

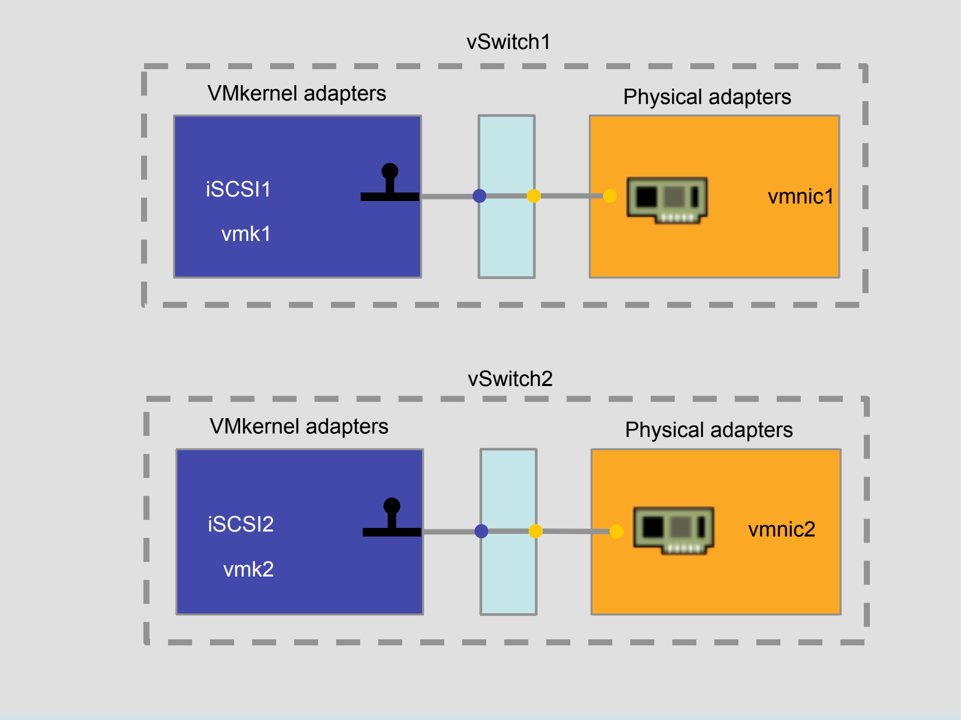
Depends on the iSCSI adapter type.



- SW iSCSI adapter:
  - Only one adapter is needed
- Dependent adapter:
  - Each vmkernel is paired with a single adapter
  - 2 or more for redundancy
- Independent adapter:
  - Do not require Vmkernel
  - 2 or more for redundancy

How do You Configure Multiple Adapters for iSCSI or iSER?

Adapter Mapping on Separate Virtual Switches.

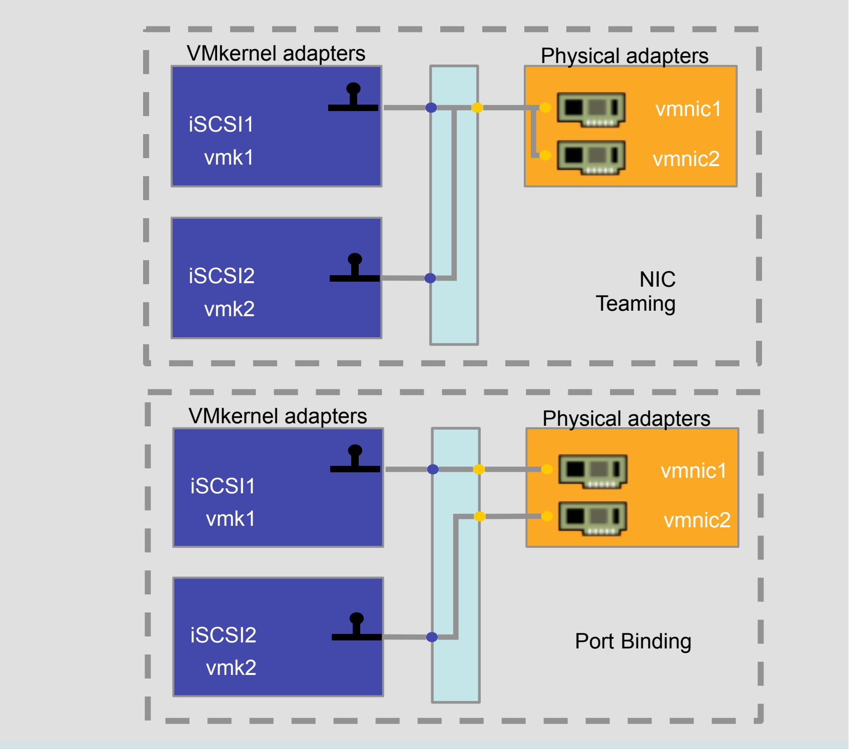


Multiple Switch Config

- Designate a separate switch for each virtual-to-physical adapter pair
- Physical network adapters must be on the same subnet as the storage

How do You Configure Multiple Adapters for iSCSI or iSER?

Adapter Mapping on a Single Virtual Switch

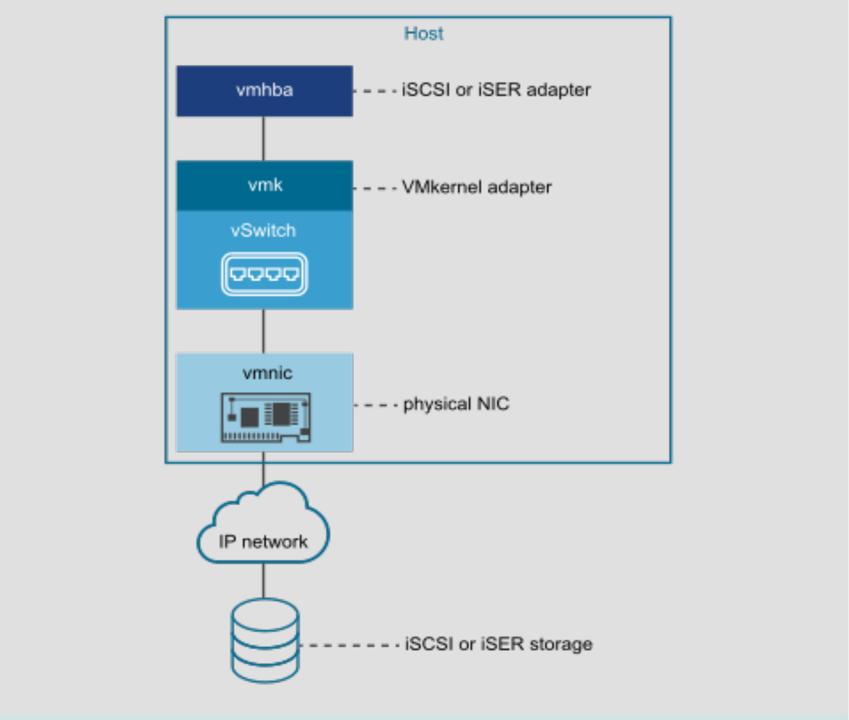


Single Switch Config

- Add all NICs and VMkernel adapters to single virtual switch
- Not suitable for iSER because iSER does not support NIC teaming*
- *You can use single switch, for iSCSI, if each iSCSI VMkernel is bound to a single NIC

How Should You Configure the Network for iSCSI and iSER?

Follow these rules when configuring the port binding

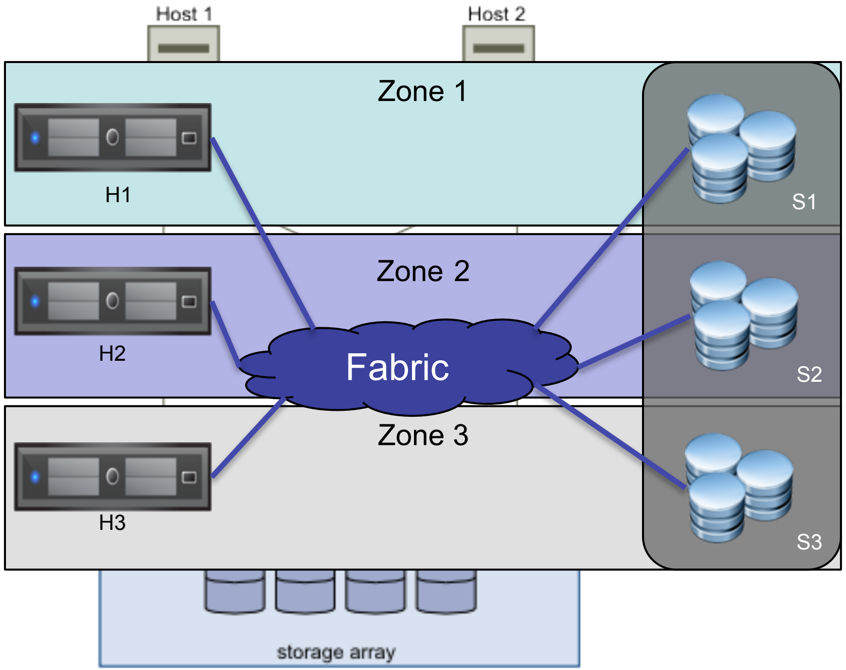


- You can connect the SW iSCSI adapter with any physical NIC available
- Dependent iSCSI adapters must be connected only to their own physical NICs
- You must connect the iSER adapter only to the RDMA-capable network adapter (RNIC)

Fibre Channel

What are Critical FC Details for Virtual Environments?

Be consistent and redundant in your FC connectivity.



- Ensures fault tolerance
- Make sure zoning is correct
- Ensure HW FW is current, supported and consistent
- Allocate adequate resources
- Do not change pathing policies

NFS

NFS Best Practices

Follow Array Vendor’s Recommendations When Possible

- Do not mix NFS client versions for same datastore
- Use separate VMKernel for NFS traffic
- Avoid routing and hops
- NFS v3 and v4.1 ESXi support AUTH_SYS security
  - NFS v4.1 supports two Kerberos security models, krb5 and krb5i
- Use NFS Multipathing with NFS v4.1

NFS Best Practices

Allocate sufficient resource, ensure reliable connectivity.

- 10Gb recommended
- More Throughput = More Performance
- Minimize latency
- Ensure redundant pathing
- Any advanced configs must be identical on all hosts!

Queuing

What is a queue depth limit?

- A queue is a line, and a queue depth limit is how “wide” that line is. Essentially, how many “things” can be allowed through at once.

Example:

- One grocery clerk can help one customer at a time (queue depth limit of 1). So, if there are two customers, one must wait for the first to finish (added latency).
- If there are two clerks (queue depth limit of 2), two customers can be helped at a time and neither has to wait (no added latency)

What is a queue depth limit?

- In terms of storage, a queue depth limit has many names:
  - Outstanding I/Os
  - Concurrent threads
  - In-flight I/Os
- If queue depth limit is 32, 32 I/Os can be processed at once. The 33rd must wait and the 33rd then has added latency because it has to wait.

Queue limits. Queue limits Everywhere.

- Storage Array Queue Depth Limit
- Device Queue Depth Limit
- Virtual SCSI Adapter Queue Depth Limit
- Virtual Disk Queue Depth Limit

Queuing Key Points

- Hypervisors are designed by default to provide some level of fairness.
- A change in one place might just move the bottleneck. Know where those bottlenecks might be.
- OS or application owners NO LONGER have full control of the storage stack. They don't control multipathing, queues, etc.

Storage Array Queue Limits

- This is really the first consideration.
- If a volume or a target on an array has a low limit—there is no point to increase anything above it.

- For storage arrays with per-volume or per-target limits, volume/target parallelization is the key.

Not host tuning.

HBA Device Queue Limit

- This is a HBA setting that controls how many outstanding I/Os can be queued on a particular device
- No different than a physical server in concept.

Main difference?

This is at the hypervisor level, NOT the OS level (same change as multipathing).

Type	Common Default Value
QLogic	64
Brocade	32
Emulex	32
Cisco UCS	32
Software iSCSI	128

Virtual Machine Queues

- Every virtual machine has two main queues:
  - Per-virtual SCSI adapter
  - Per-virtual disk
- Different types of virtual SCSI adapters have different defaults and maximums.
- If you are experiencing latency in the application, but the hypervisor is configured to the max, the latency might be introduced here.
- Hypervisor does NOT know about these queues or latency introduced because of them

Virtual Machine Queues

- Simply upgrading virtual adapters or increasing their internal is **unlikely to improve** performance (or really at all)
- Otherwise, you are just moving the bottleneck from the guest to the hypervisor:

Little's Law: The long-term average number of customers in a stable system L is equal to the long-term average effective arrival rate, λ, multiplied by the average time a customer spends in the system, W.

$$L = \lambda W$$

Little’s Law in Action

- Let’s use our grocery store analogy again:
- If one customer takes 1 minute to check out (aka latency) and there is one clerk, a store can serve 60 customers in an hour (aka IOPS).
- If there are two clerks, the store can serve 120 customers in an hour.

Some quick math…

- Let’s suppose the latency should be .5 ms:
  - 1 second = 1,000 ms
  - 1,000 ms / .5 ms per IO = 2,000 IOPS
  - 2,000 IOPS * 96 outstanding I/Os = 192,000 IOPS

With this latency, we would expect 192,000 IOPS

Do I need to change this stuff?

YES, *IF:* you see host-introduced latency and/or you need more available throughput or IOPS, increase the queue depth limits.

NO, *BECAUSE:* Most workloads are distributed across VMs, hosts and/or volumes (parallel queues)

NO, *BECAUSE:* Low-latency arrays are less likely to need changes—they empty out the queue (i.e. service the I/Os) very fast

Again: Hypervisors are designed by default to provide some level of fairness.

Back to Little's Law

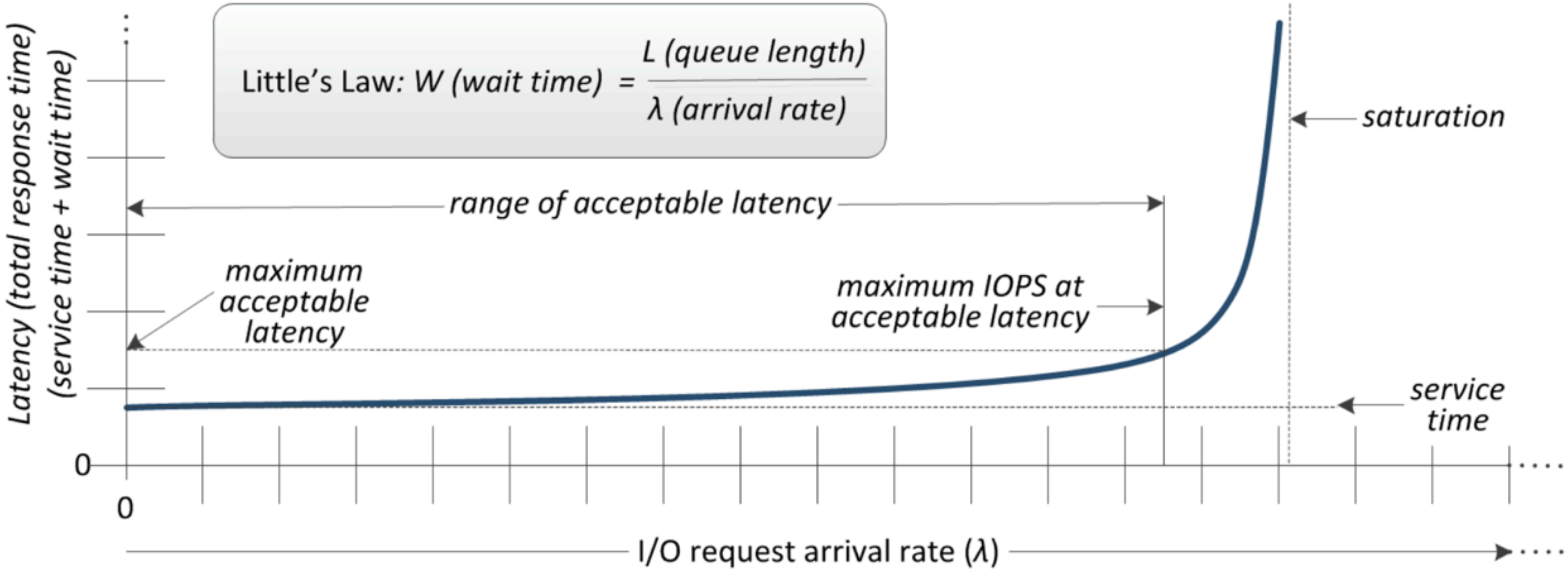


Figure 4: I/O Saturation

口子窖
出口
EXIT

Performance Management

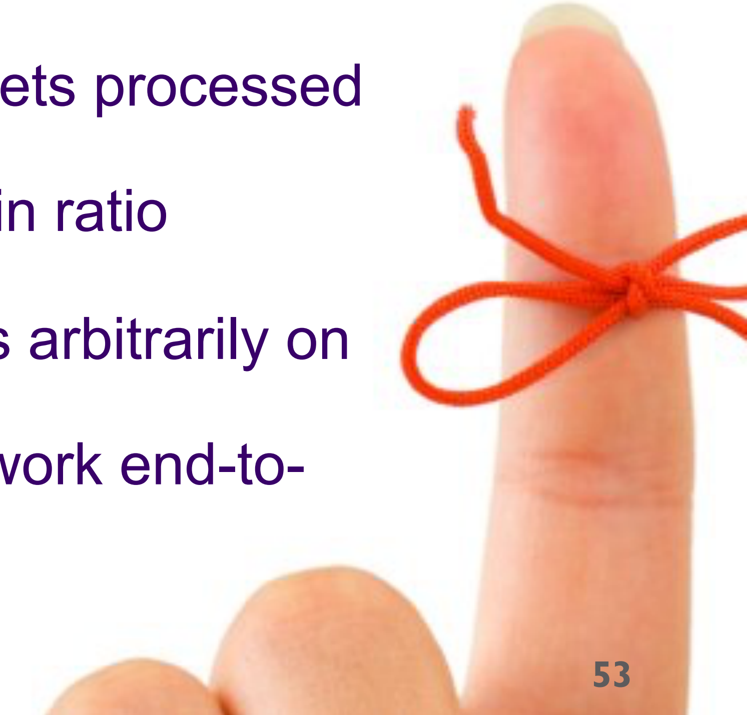
SNIA NSF | NETWORKING STORAGE

- Maybe change queue depths
- Hypervisor-level performance QoS
- Array-level QoS
- Performance Balancing tools

Basic decision: decide is something becomes abnormal, who should be punished?

Summary

- Applications are the primary litmus test for the proper storage networking protocol
- Be aware of the impact of where I/O gets processed – in the hypervisor or on the hardware
- Understand oversubscription and fan-in ratio requirements when necessary
- Don’t change storage network settings arbitrarily on a host
- Understand and know how protocols work end-to-end



- Session Resources
- https://bit.ly/2Lw0WaY

Upcoming NSF Webcasts

SNIA NSF | NETWORKING STORAGE

What NVMe™/TCP Means for Networked Storage
January 22, 2019
Register: https://www.brighttalk.com/webcast/663/344698

Networking Requirement for Hyperconvergence
February 5, 2019
Register: https://www.brighttalk.com/webcast/663/341209

The Scale-Out File System Architecture Overview
February 28, 2019
Register: https://www.brighttalk.com/webcast/663/346111

After This Webcast

- Please rate this webcast and provide us with feedback
- This webcast and a PDF of the slides will be posted to the SNIA Networking Storage Forum (NSF) website and available on-demand at www.snia.org/library
- A full Q&A from this webcast, including answers to questions we couldn't get to today, will be posted to the SNIA-NSF blog: sniansfblog.org
- Follow us on Twitter @SNIANSF

SNIA NSF | NETWORKING STORAGE

Thank You