



# MYTHBUSTING GOES VIRTUAL

MATTIAS SUNDLING

ERIC SLOOF



# MYTHBUSTING GOES VIRTUAL



Eric Sloof  
VMware Certified Instructor  
NTPRO.NL  
@esloof

Mattias Sundling  
Evangelist  
Quest Software  
@msundling



## INTRODUCTION

- VMware vSphere evolves with every release
- Things that used to be true aren't true anymore
- Engage in virtualization communities and social media to get up to speed



## AGENDA/MYTHS

- 1) Defrag your Guest OS disks for best performance
- 2) E1000 is faster than VMXNET3
- 3) CBT causes significant overhead on your VMs
- 4) HA datastore heartbeats prevents host isolation
- 5) LSI SCSI is always better than Paravirtual SCSI



## MYTH 1

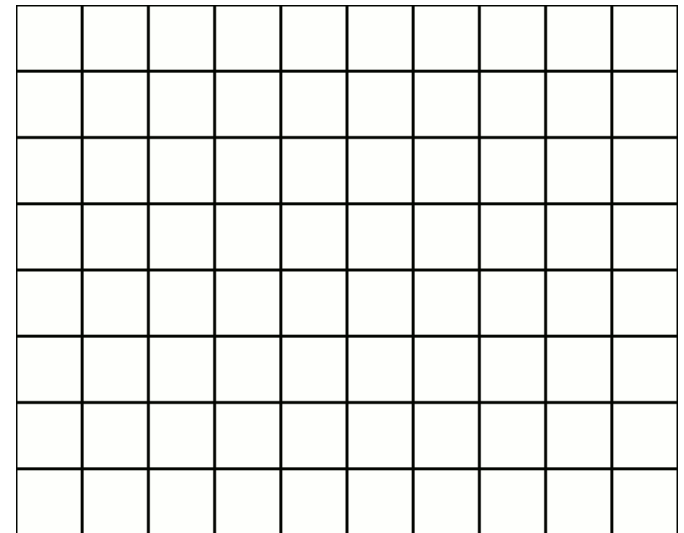
Defrag your Guest OS disks for best performance





## WHAT IS DISK DEFRAGMENTATION?

- Fragmentation occurs over time as you save, change or delete files
- Result in unnecessary reads/writes -> lower performance
- Disk Defrag tools can move data so files are sequentially on disk





## DEFRAGMENTATION: PHYSICAL VS VIRTUAL

- Physical Machine with single disk
  - Use defrag to improve performance
- Virtual Machine with SAN
  - No understanding of underlying disk geometry
  - Multiple VMs sharing a LUN = random disk I/O
  - Defrag process is heavy on disk I/O
  - Change Block Tracking (Backup & Replication)
  - Thin Provisioning
  - Snapshots
  - Linked Clones (vCloud Director & View)



## VM DEFRAG AND RECOMMENDATIONS

- Auto tiering
- SSD has no moving parts, not affected, shorten lifecycle

What does VMware and Storage Vendors recommend?

- Don't defrag VMs!

What does 3<sup>rd</sup> party defrag vendors say?

- Defrag your VMs!



## MYTH BUSTED

- Use defrag tools with caution
- Align your disks
- Paravirtual SCSI





## MYTH 2

E1000 is faster than VMXNET3





## WHAT'S AN E1000?

- The Intel 82545EM Gigabit Ethernet Controller
- VMware offers an emulated version of this controller
- Most operating systems are shipped with a 82545EM driver
- The 82545EM driver sucks! That's why Intel replaced it with e1000e aka 82574L



## WHAT'S AN E1000E?

- The Intel 82574L Gigabit Ethernet Controller
- In vSphere 5 (HW8), VMware offers an emulated version
- Windows 7 and Windows 2008 are shipped with drivers for the 82574L
- The 82574L is cool, but is it faster than an VMXNET3?



## WHAT'S VMXNET3?

- The VMXNET3 adapter is the next generation of Para virtualized NIC designed for performance
- The VMXNET3 network adapter is a 10Gb virtual NIC
- Drivers are shipped with the VMware tools and most OS are supported



# THE LABORATORY

The image displays a VMware vSphere interface with two overlapping windows. The background window is the PerformanceTest 7.0 application, showing benchmark results for a VM. The foreground window is the VM configuration page for a VM named 'clone from vm-left'.

**PerformanceTest 7.0 Results:**

- Video Adapters: Description: VMware SVGA 3D, Mfg: VMware, Inc., Memory: 0MB, Driver (Date): 7.14.1.1050 (2-7-2011)
- Disk Information: Drive Letter (Number): C:\ ( 0 ), Model Number: (N/A), Disk Size: 31.9 GB, Free Space: 14.6 GB
- Hardware components: HARDDISK, CD DRIVE, MOTHERBOARD, MEMORY, VIDEO CARD

**VM Configuration (clone from vm-left):**

Property	Value
Guest OS:	Microsoft Windows 7 (64-bit)
VM Version:	8
CPU:	1 vCPU
Memory:	2048 MB
Memory Overhead:	98,87 MB
VMware Tools:	Running (Current)
IP Addresses:	192.168.178.111
DNS Name:	CloneVMLeft
EVC Mode:	N/A
State:	Powered On
Host:	esx4-l.ntpro.local
Active Tasks:	
vSphere HA Protection:	N/A

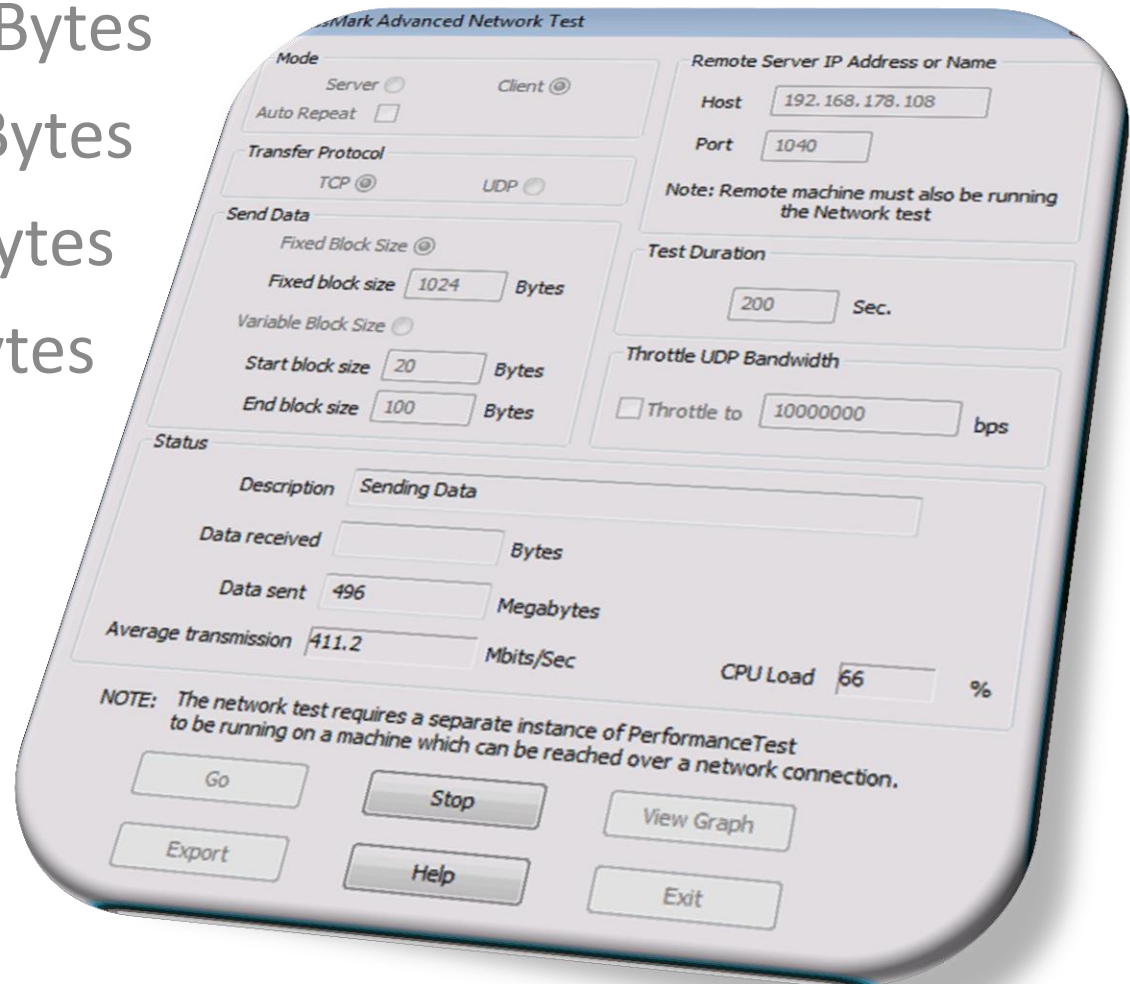
**VM Inventory:**

- MythBusters
  - esx4-l.ntpro.local
    - clone from vm-left
    - vm-left
    - vMotion
    - Windows 8
    - Windows XP
    - Windows8
  - esx4-r.ntpro.local
    - VMworld 2011
      - 192.168.178.128
      - 192.168.178.177
      - 192.168.2.102
      - CentOS
      - DC.NTPRO.LOCAL
      - dsl
      - DSL-2
      - EFI
      - vEOS
      - View 4.5



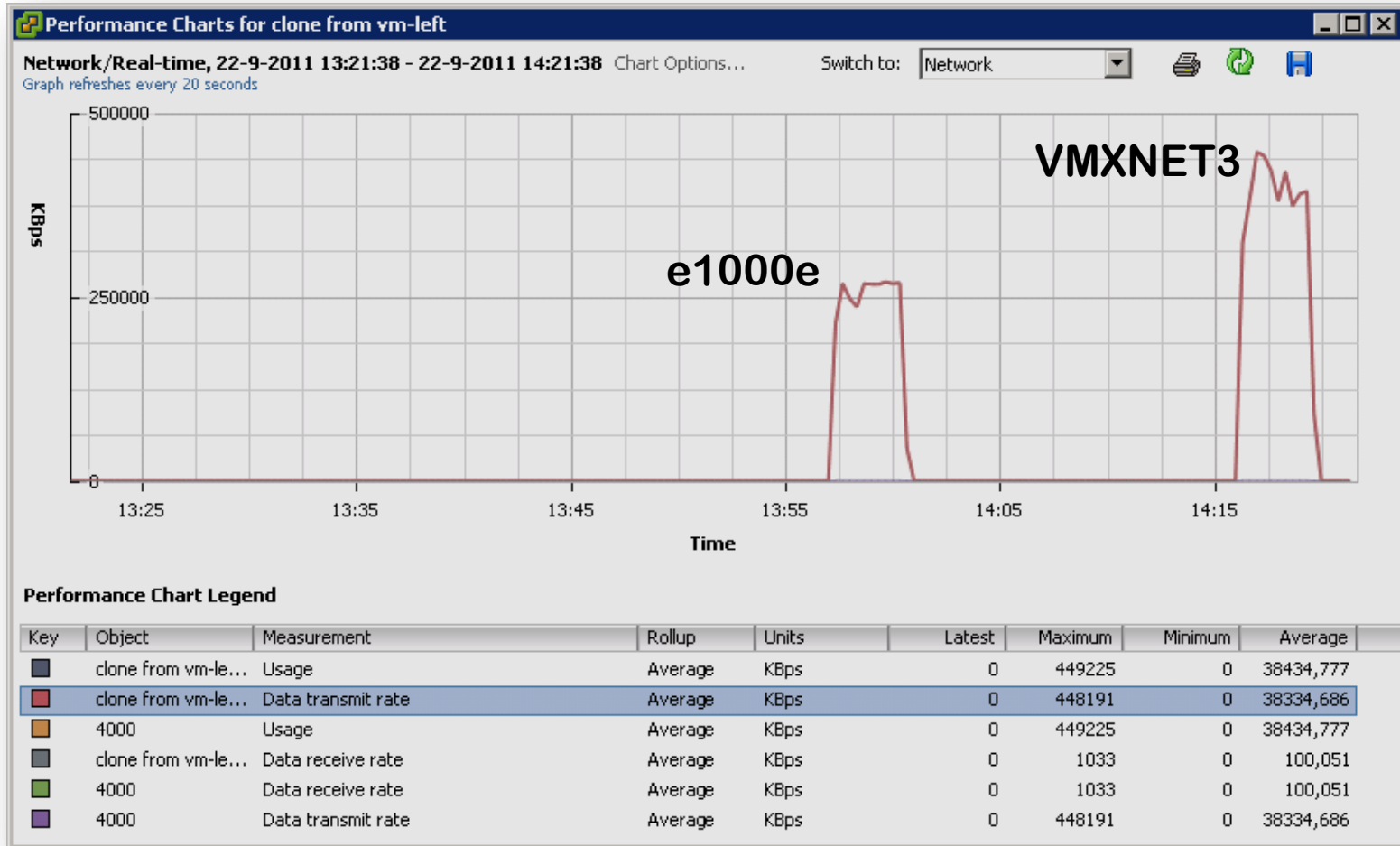
## THE TOOLS

- UDP 1024 Bytes
- TCP 1024 Bytes
- UDP 16 Kbytes
- TCP 16 Kbytes





# TCP 16K - WHAT'S FASTER?





# UDP 1024 BYTES - WHAT'S FASTER?

## VMXNET3

e1000e

esx4-lntpro.local - PuTTY  
12:27:53pm up 4 days 23:01, 278 worlds, 2 VMs, 2 vCPUs; CPU load average: 0.25, 0.08, 0.03

PORT-ID	USED-BY	TEAM-PNIC	DNAME	PKTTX/s	MbTX/s	PKTRX/s	MbRX/s	%DRPTX	%
16777217	Management	n/a	vSwitch0	0.00	0.00	0.00	0.00	0.00	0.00
16777218	vmnic0	-	vSwitch0	4.16	0.02	4.36	0.00	0.00	0.00
16777220	vmk0	vmnic0	vSwitch0	3.96	0.02	2.58	0.00	0.00	0.00
16777253	310219:clone from vm	vmnic0	vSwitch0	203597.24	1655.84	0.99	0.00	0.00	0.00
16777254	310244:vm-left	vmnic0	vSwitch0	0.20	0.00	203594.66	1655.82	0.00	0.00
33554435	Management	n/a	DvsPortset-2	0.00	0.00	0.00	0.00	0.00	0.00
33554436	vmk2	vmnic2	DvsPortset-2	0.00	0.00	0.00	0.00	0.00	0.00
33554438	vmnic2	-	DvsPortset-2	0.00	0.00	0.59	0.00	0.00	0.00
33554439	vmnic3	-	DvsPortset-2	0.00	0.00	0.59	0.00	0.00	0.00
33554442	vmk1	vmnic3	DvsPortset-2	0.00	0.00	0.00	0.00	0.00	0.00

esx4-lntpro.local - PuTTY  
12:37:47pm up 4 days 23:11, 278 worlds, 2 VMs, 2 vCPUs; CPU load average: 0.25, 0.08, 0.03

PORT-ID	USED-BY	TEAM-PNIC	DNAME	PKTTX/s	MbTX/s	PKTRX/s	MbRX/s	%DRPTX	%
16777217	Management	n/a	vSwitch0	0.00	0.00	0.00	0.00	0.00	0.00
16777218	vmnic0	-	vSwitch0	5.55	0.03	5.15	0.00	0.00	0.00
16777220	vmk0	vmnic0	vSwitch0	4.95	0.03	3.96	0.00	0.00	0.00
16777256	311074:vm-left	vmnic0	vSwitch0	0.40	0.00	129648.59	1058.38	0.00	0.00
16777257	309044:clone from vm	vmnic0	vSwitch0	129648.78	1054.42	0.79	0.00	0.00	0.00
33554435	Management	n/a	DvsPortset-2	0.00	0.00	0.00	0.00	0.00	0.00
33554436	vmk2	vmnic2	DvsPortset-2	0.00	0.00	0.40	0.00	0.00	0.00
33554438	vmnic2	-	DvsPortset-2	0.00	0.00	1.59	0.00	0.00	0.00
33554439	vmnic3	-	DvsPortset-2	0.00	0.00	1.19	0.00	0.00	0.00
33554442	vmk1	vmnic3	DvsPortset-2	0.00	0.00	0.40	0.00	0.00	0.00
50331649	Management	n/a	DvsPortset-1	0.00	0.00	0.00	0.00	0.00	0.00
67108865	Management	n/a	vSwitch1	0.00	0.00	0.00	0.00	0.00	0.00



## MYTH BUSTED

- VMXNET3 is much faster than e1000 or e1000e
- VMXNET3 has less CPU overhead compared to e1000 or e1000e
- VMXNET3 is more stable than e1000 or e1000e





## MYTH 3

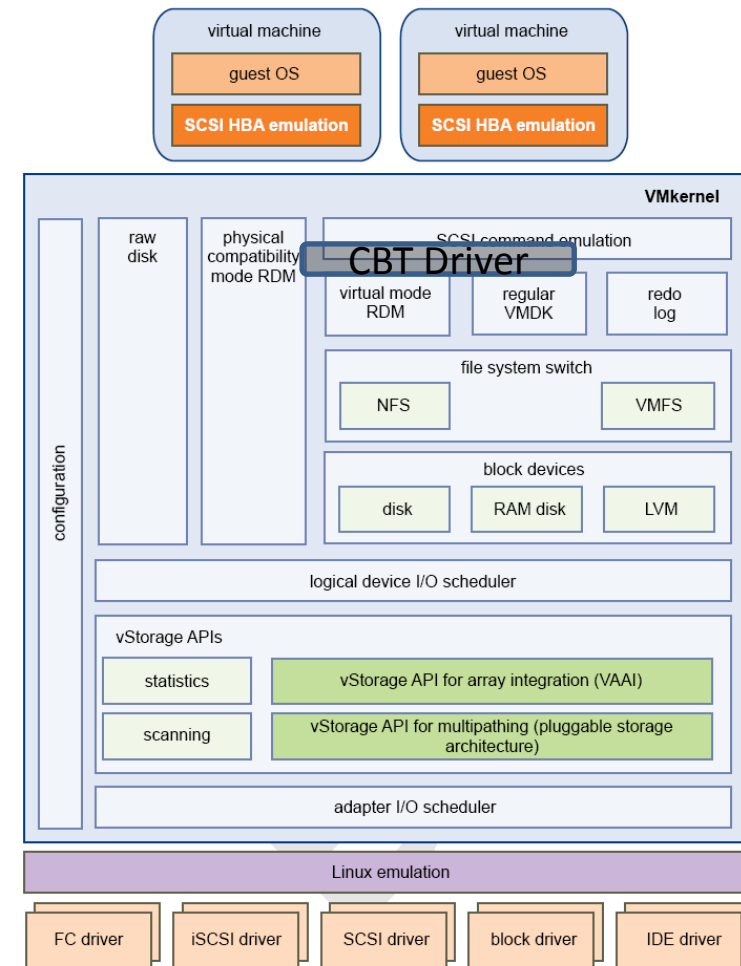
CBT CAUSES SIGNIFICANT OVERHEAD ON YOUR VMs





## WHAT IS CBT?

- Driver inside VMkernel
- Identifies change blocks within virtual disks
- Block size based on VMDK size
- Backup window significantly reduced
- Requirements: vSphere 4+ and Virtual HW v7+
- Limitations: pRDM, iSCSI within VM
- Enable through vCenter or backup application





## CBT OVERHEAD

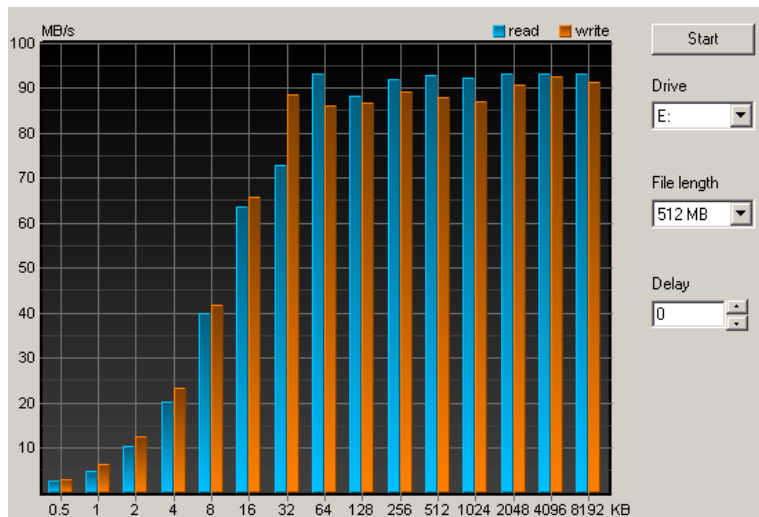
- Memory
  - Max 256 KB / Disk (2TB)
  - 1.25 KB / 10 GB VMDK
- CPU
  - Turning on a bit in bitmap when an I/O request completes
- Storage
  - Space
    - .ctk file 0,5 MB / 10 GB VMDK
  - I/O
    - Every time disk gets closed, change tracking info written to disk



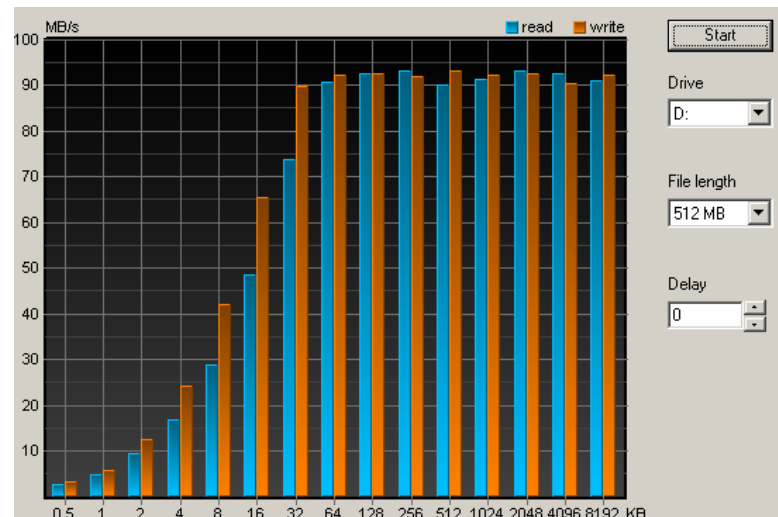
# TEST/RESULT

- Very little overhead CPU, Mem, Storage
- Could not measure it, except .ctk file
- No negative impact on disk I/O

## CBT Disabled



## CBT Enabled





## MYTH BUSTED

- CBT overhead is very small
- Backup window and host resources reduced significant
- Use CBT if your backup solution supports it





## MYTH 4

HA datastore heartbeats prevent host isolation





## DATSTORE HEARTBEATS ? HOST-X-HB

- host-X-hb (where X is the host's MOID) – Located on each heartbeat datastore, this file is used to check for slave liveness through the heartbeat datastore.
- This file is checked by the master host if the master loses network heartbeats from the slave.
- For VMFS datastores, the vSphere HA agent locks this file with an exclusive lock and relies on the VMkernel heartbeat to indicate liveness.
- For NFS datastores, vSphere HA periodically updates the time stamp to this file to indicate liveness.



## DATASTORE HEARTBEATS ? HOST-X-POWERON

- host-X-poweron (where X is the host's MOID) – A per-host file that contains the list of all virtual machines that are powered on.
- This file is used as a communication channel if a management network outage occurs.
- Isolated slaves use this file to tell the master that it is isolated as well as to tell the master which virtual machines it has powered off.



## THE SLAVE DOES NOT RESPOND

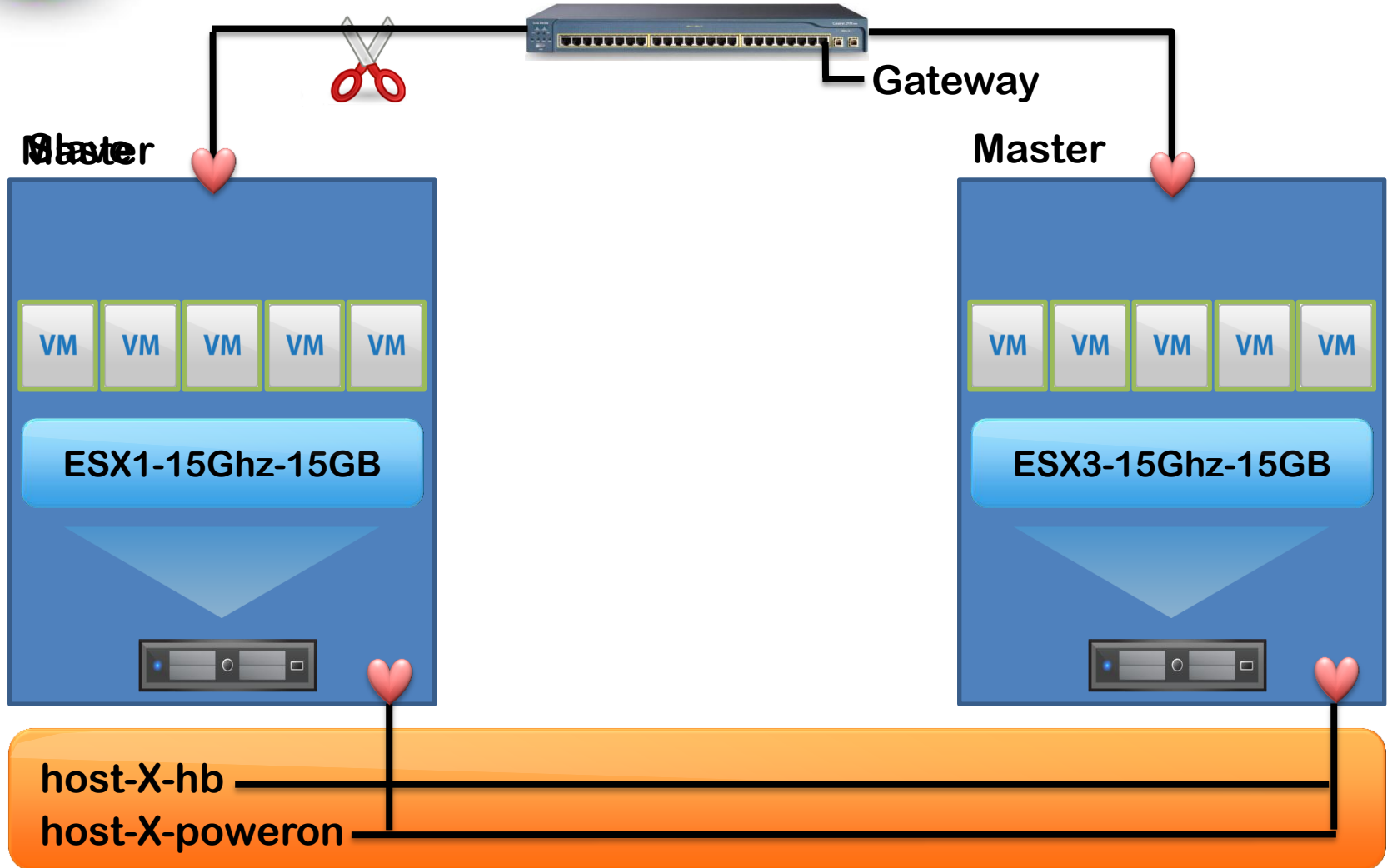
The master host must determine whether the slave host:

- Actually crashed
- Is not responding because of a network failure
- The HA agent is in an unreachable state

The absence of both a network and datastore heartbeat indicates full host failure.



# THE LABORATORY





## MYTH BUSTED

- Datastores are used as a backup communication channel to detect virtual machine and host heartbeats.
- Datastore heartbeats are used to make the distinction between a failed, an isolated or a partitioned host.





## MYTH 5

LSI LOGIC SCSI IS ALWAYS BETTER THAN PVSCSI





## WHAT IS PVSCSI?

- Paravirtual SCSI
  - Introduced in vSphere 4, improved in vSphere 4.1
  - Designed for high performance (+12%)
  - Requires less resources on vSphere Host (-18%)
  - Supports only Win 2003+, RHEL5+, SUSE Linux Enterprise 11 SP1+, Ubuntu 10.04 +, Linux 2.6.33+
  - Virtual HW 7+

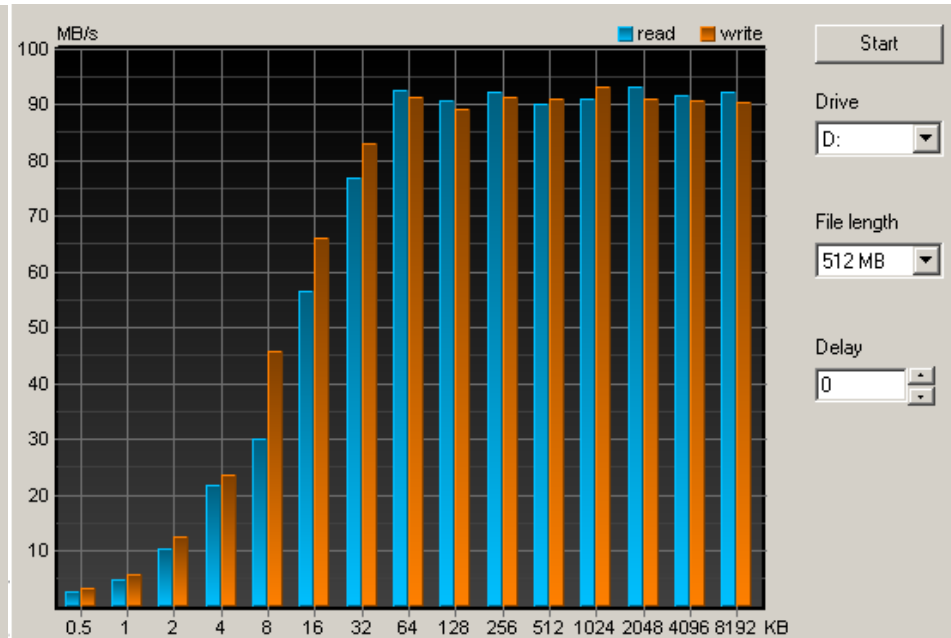
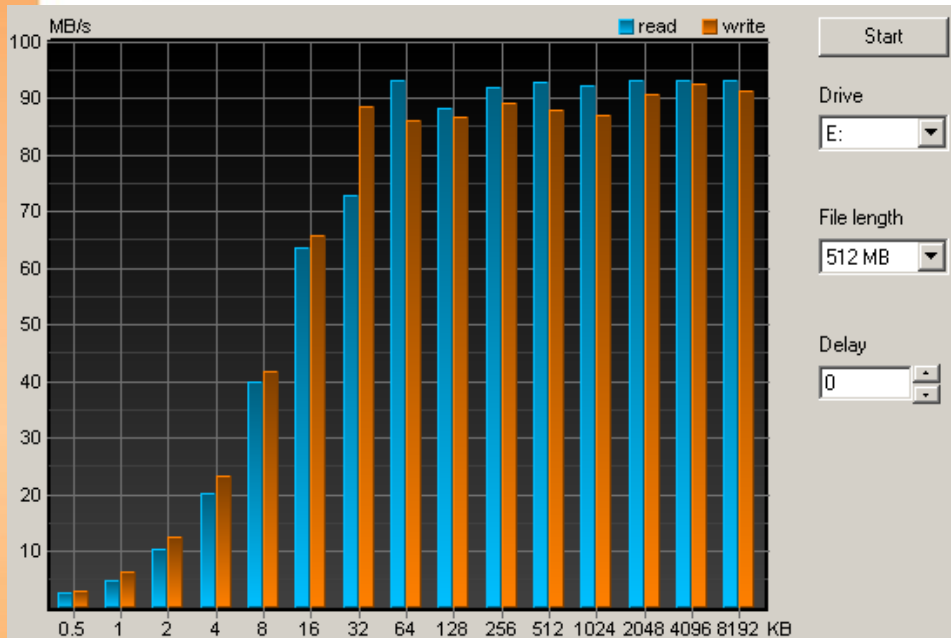
Hardware	Summary
Memory	1024 MB
CPUs	1
Video card	Video card
VMCI device	Restricted
SCSI controller 0	LSI Logic Parallel
Hard disk 1	Virtual Disk
CD/DVD drive 1	Client Device
Network adapter 1	VM Network
SCSI controller 1	Paravirtual
Hard disk 2	Virtual Disk



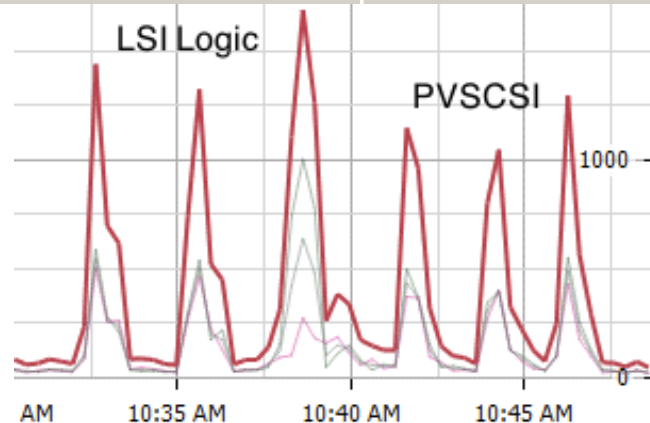
# TEST RESULT LSI LOGIC VS. PVSCSI

## LSI Logic

## PVSCSI



- Very similar disk throughput
- Lower Host CPU Utilization for PVSCSI



Host CPU utilization  
90MB/s – 60 000  
IOPS



## MYTH BUSTED

- PVSCSI is equal or faster than LSI SCSI
- PVSCSI requires less Host resources
- Used to have more limitations
- PVSCSI is better in all ways so why aren't we using it on all supported VMs?
- Takes time to change behaviour





## SUMMARY

- Use disk defragmentation tools with caution
- Use VMXNET3 wherever possible
- CBT has very little overhead
- HA datastore heartbeats are used as a communication channel after management network outage
- PVSCSI is equal or faster and requires less resources than LSI SCSI



## QUESTIONS



### **Eric Sloof**

VMware Certified Instructor  
NTPRO.NL

[esloof@ntpro.nl](mailto:esloof@ntpro.nl), [@esloof](https://twitter.com/esloof)

### **Mattias Sundling**

Evangelist  
Quest Software

[mattias.sundling@quest.com](mailto:mattias.sundling@quest.com), [@msundling](https://twitter.com/msundling)