



A Clear Vision of Technology Solutions

# ESX 3 Performance Tuning

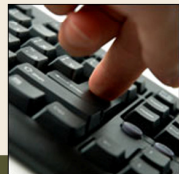
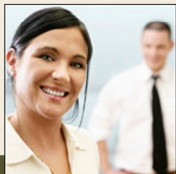
Eric Eiseman (VCP/VAC, RHCE)  
Long View Systems

August 15, 2007



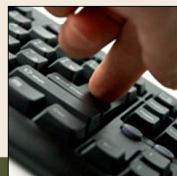
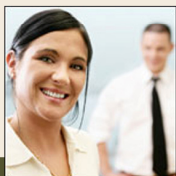
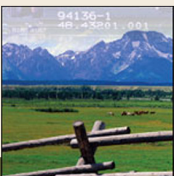
# Overview

- Perspective
  - ESX Hosts
  - Virtual Machines
- 4 Classic Parameters
  - CPU
  - RAM
  - Disk
  - Network



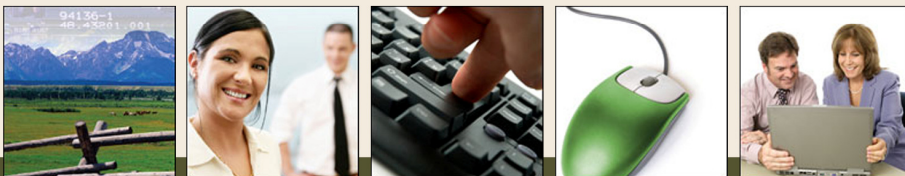
# ESX Host - CPU

- CPU's are CPU's
  - You can't make them go any faster
- Can choose the number of CPU's/Cores per ESX host
- Consider vSMP scheduling needs
  - vSMP machines must schedule all cores *together*
  - Ensure you have more CPU cores than the largest vSMP VM
  - i.e., 5+ cores for a 4 vSMP machine
  - Service console must schedule on pCPU0



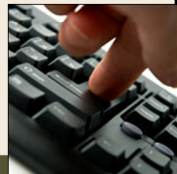
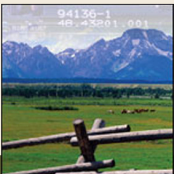
# ESX Host - CPU

- Keep all ESX hosts in a cluster vMotion compatible
  - Not just AMD vs. Intel, but age as well
  - [http://www.vmware.com/support/kb/enduser/std\\_adp.php?p\\_faqid=1991](http://www.vmware.com/support/kb/enduser/std_adp.php?p_faqid=1991)
  - [http://www.vmware.com/support/kb/enduser/std\\_adp.php?p\\_faqid=1992](http://www.vmware.com/support/kb/enduser/std_adp.php?p_faqid=1992)
- 32 cores max per ESX host
- More cores is not necessarily better
  - E.g., 32 CPU host may max RAM before running out of scheduling locations
- Service Console applications can greatly affect CPU of the whole host
  - Management tools, iSCSI Software Initiator, NFS, VMotion, etc.



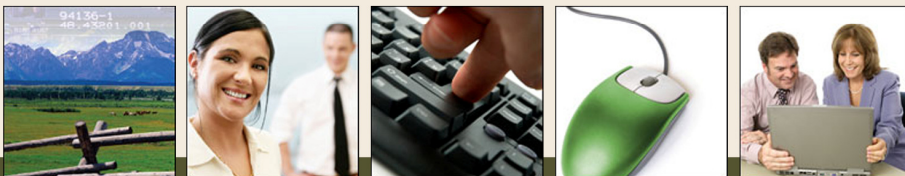
# ESX Host - RAM

- More RAM is not necessarily better
  - A 64GB host may max CPU's before addressing all of its RAM
- Consider HA requirements
  - A 1 host failure HA cluster will reserve  $\sim 1/N$  of the RAM on each host (20% for 5 hosts)
- Have more RAM than the largest VM
  - More than 16GB for a 16GB VM
  - Service Console uses 800MB (272MB by default)
- Over-allocation is fine to a point (1.25x – 1.33x)
  - RAM actually in use should still be under 1.0x
  - Don't swap if you can ever avoid it



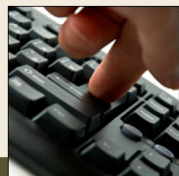
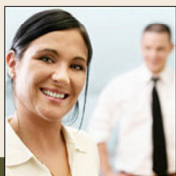
# ESX Host - Disk

- Use Shared Storage
  - Fibre Channel SAN
  - iSCSI
  - NFS
- Choice depends on I/O rate of VM workloads and your wallet



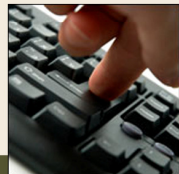
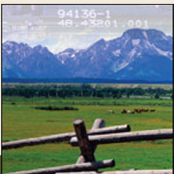
# Fibre Channel SAN

- Pro's
  - Ol' reliable
  - Highest potential performance (4GB fabric) if ESX can generate enough traffic
- Con's
  - Most complex to manage (if you don't already have a FC deployment)
  - Most expensive to acquire
  - Separate SAN fabric to manage



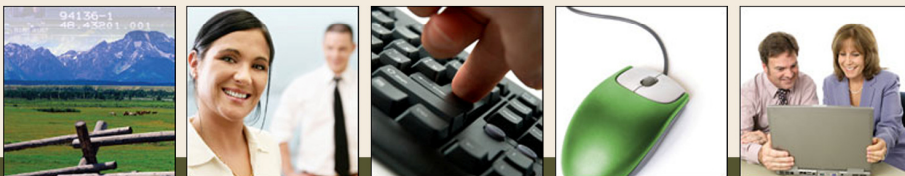
# iSCSI

- Pro's
  - Cheaper than Fibre Channel
  - Reuse existing networking infrastructure
  - Still block level access formatted as vmfs3
- Con's
  - Newer protocol (RFC 3720 from 2002)
  - 1GB max
  - HBA's are new
  - Software initiator can't use jumbo frames
  - CPU overhead if not using HBA's



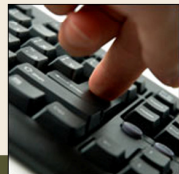
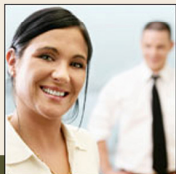
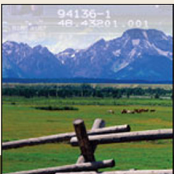
# NFS

- Pro's
  - Oldest protocol (RFC 1094 in 1989)
  - Reuse existing networking infrastructure
  - Least on-disk usage (thin disk vmdk's)
- Con's
  - 1GB max
  - File level protocol controlled by NFS host (locking)
  - ESX cannot see the file system
  - No RDM's
  - Need NFS over TCP (UDP is default for older/cheaper devices)
  - CPU overhead



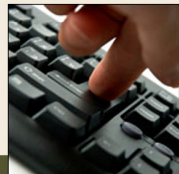
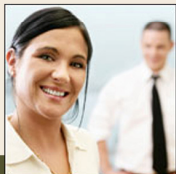
# Disk Recommendations

- RAID Types
  - Use multi-RAID set technologies
    - EMC – RAID5 MetaLUN's
    - HP – RAID5 vRAID's
    - NetApp – RAID-DP Aggregates
  - Best balance of performance vs. cost
- ESX will auto-align on EMC disks
- Use iSCSI HBA's on heavily used systems
  - Offload CPU time to hardware
  - Allows jumbo frames (9000 is a good size)
- Use dedicated NFS/iSCSI hosting hardware
  - NetApp, Celerra, etc.
- Tune the “fabric” (Fibre Channel or Ethernet)



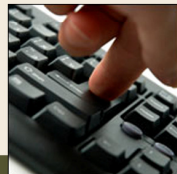
# VMFS Recommendations

- Limit vmfs datastores to a reasonable size
  - Metadata locking issues
  - 10 to 12 VM's per LUN
- Do not use extents
  - Create a new LUN if possible
- Use RDM's for disks over a certain size
  - No reason for vmdk/vmfs overhead for a single 2TB file
  - Pick a line in the sand (50GB, 100GB, etc.)
  - Easier to grow RDM's (SAN operation only)



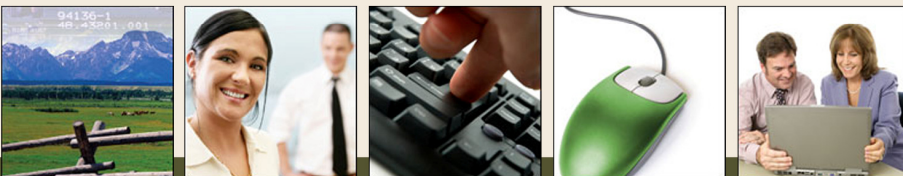
# ESX Host - Network

- Use 1000/full everywhere
  - auto/auto in most instances
- Turn off as many negotiations as possible
  - Use “spanning-tree portfast “ or “spanning-tree portfast trunk”
  - LACP/PAgP off
- At least 2 pNIC’s per vSwitch
  - Including Service Console/vMotion/vmkernel
- Etherchannel (802.3ad) is usually not worth the effort
  - Will you have more than 1 GB of traffic to a single VM?
  - All links must go to the same switch



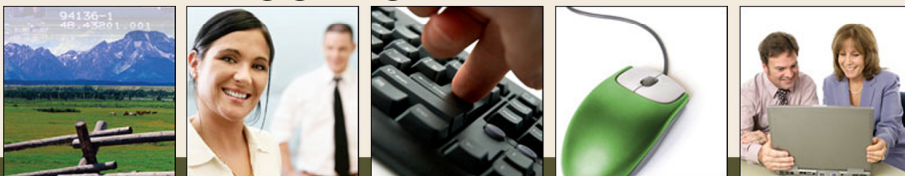
# ESX Host - Misc

- Remember:
  - 1Gb network = 1GHz of CPU time
- Check for hotfixes monthly
  - Many memory leaks and race conditions have been fixed in the past year
- Let DRS do your work for you
  - Teach it business rules with affinity and anti-affinity
- Use 802.1q trunking when you need a large number of networks
  - Easier than managing 12+ pNIC's



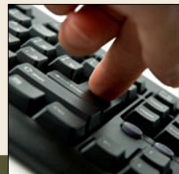
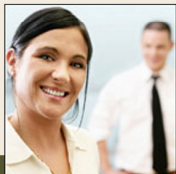
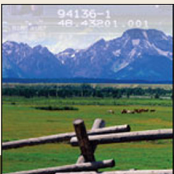
# Virtual Machines - CPU

- Use the least number of vCPU's possible
  - Single threaded applications won't benefit from having a 2<sup>nd</sup> CPU
  - SQL/Oracle servers are good vCPU candidates
  - Prove the additional CPU's will improve performance with external testing
- Use Shares instead of reservations and limits.
  - Shares dynamically change as DRS vMotions machines around
  - Static limits and reservations may tune you into a corner



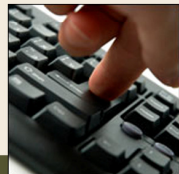
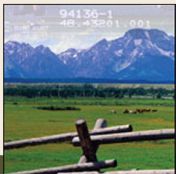
# Virtual Machines - RAM

- Allocate what a VM actually needs to perform
  - Windows touches all RAM on boot (sawtooth graphs)
  - SQL will cache to all available RAM whether it uses it or not (measure cache hit/miss rates and adjust)
  - Runaway processes will be limited
- Smaller “puzzle piece” sizes will give DRS more flexibility on where to place VM’s



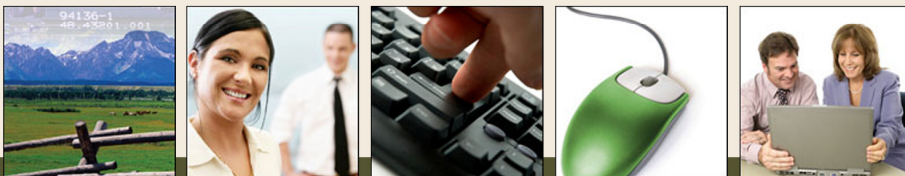
# Virtual Machines - Disk

- Pick a reasonable default disk size
  - 10 to 20 GB
  - Increasing vmdk's is easier than shrinking
  - Idle space on your average server
- Use RDM's for very large disk or high I/O needs (decrease layers of abstraction)
- Use LSI Logic *everywhere*
  - Newer SCSI implementation than Buslogic



# Virtual Machines – Network

- Place VM's that talk to each other heavily on the same
  - ESX host
  - vSwitch
  - Portgroup
- Network transfers are now at RAM copy speed

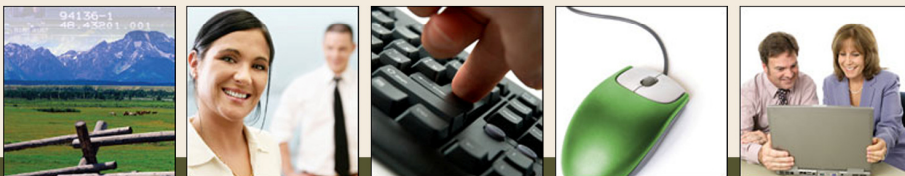


A Clear Vision of Technology Solutions



# Virtual Machines - Misc

- *Always* load the VMware Tools
  - Updated drivers
  - Allows vmkernel to see inside the guest kernel (balloon driver, etc.)
- Disable all peripherals you won't use
  - COM, LPT, Floppy, etc.
  - Device Manager/Kudzu will scan them

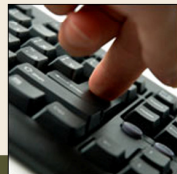


A Clear Vision of Technology Solutions



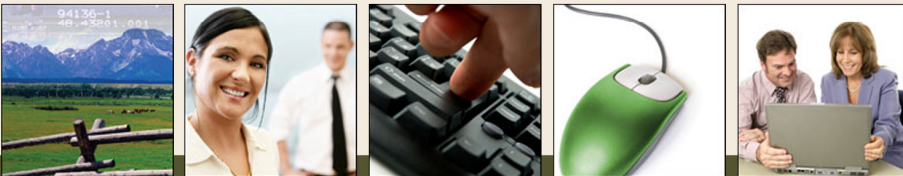
# Monitoring/Tuning

- Must coordinate stats from inside a VM to the behavior of the ESX host
- Inside the VM context is relative to the allocated shares
  - Watch out for Resource Pools – shares are relative to the pool that owns them
- External (client) testing is the gold standard
  - Slow terminal server sessions on a SQL server doesn't mean very much
  - What is the response time for real world queries?
  - Match the testing to the production workload



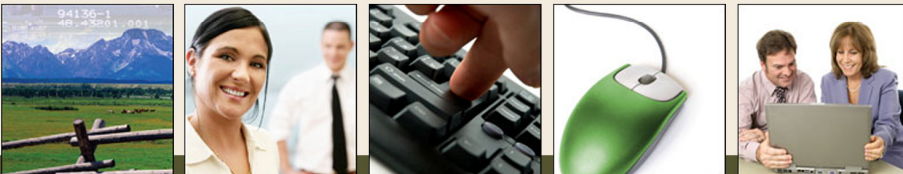
# Virtual Machine Monitoring

- Establish a baseline with Perfmon and Task Manager
- Understand what “normal” is before trouble starts
- What are your SLA’s?
- Dev environments tend to have a never ending need
- How fast is fast enough?
- Average CPU Utilization
- Peak CPU Utilization
- CPU Time
- Processor Queue Length
- Memory Usage
- Peak Memory Usage
- Page Faults
- I/O Reads vs. Write
- I/O Bytes (Read and Write)
- Disk Queue Lengths
- Network Bytes Received/Sent
- Output queue Length



# ESX Host Monitoring

- Virtual Center
  - Changing graph timescale changes sample size
- esxtop
  - Be sure to change modes (c - cpu, d - disk, m - memory)
- %RDY is a red flag
  - VM's are spending time in a queue waiting to be scheduled
- Swapping stats
  - Out of RAM?
- Consistent CPU times > 80%
  - Need another host?
- Disk queue lengths
  - Read vs. Write characteristics
  - What disk I/O loads are in competition on the SAN?

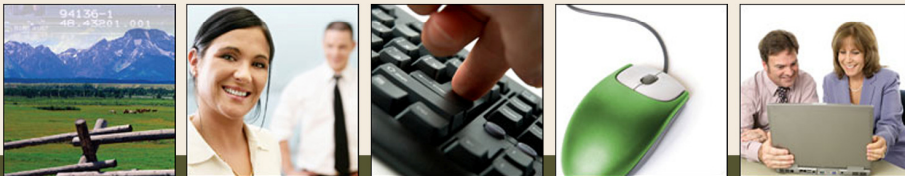


A Clear Vision of Technology Solutions



# Questions?

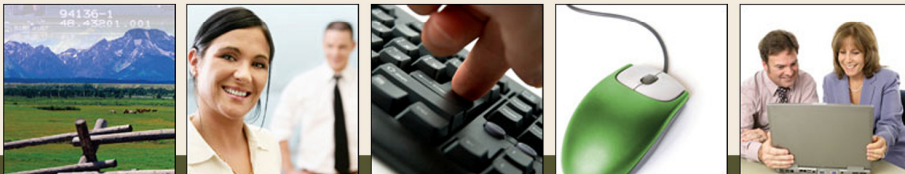
- ????



A Clear Vision of Technology Solutions

# Thank You

- My E-mail
  - [eric.eiseman@lvs1.com](mailto:eric.eiseman@lvs1.com)
- More Long View Systems Info
  - [denversales@lvs1.com](mailto:denversales@lvs1.com)



A Clear Vision of Technology Solutions

