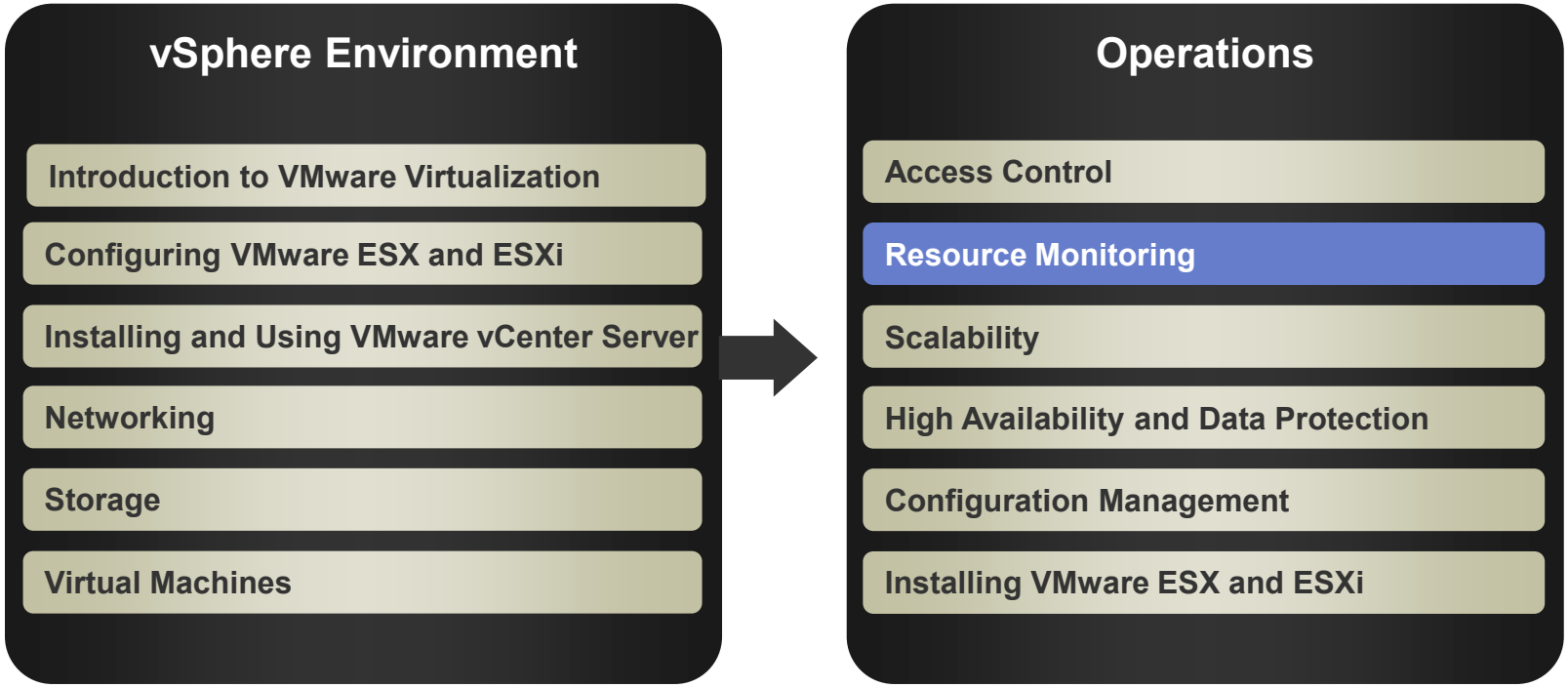
Resource Monitoring

Module 9

You Are Here



Importance

- Although the VMkernel works proactively to avoid resource contention, maximizing performance requires both analysis and ongoing monitoring.

Module Lessons

Lesson 1: Virtual CPU and Memory Concepts

Lesson 2: Monitoring Resource Usage

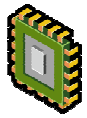
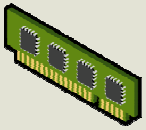
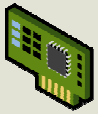
Lesson 3: Using Alarms

Lesson 1: Virtual CPU and Memory Concepts

Lesson Objectives

- Understand the different methods used by the VMkernel for optimizing CPU and memory usage

Systems for Optimizing VM Resource Use

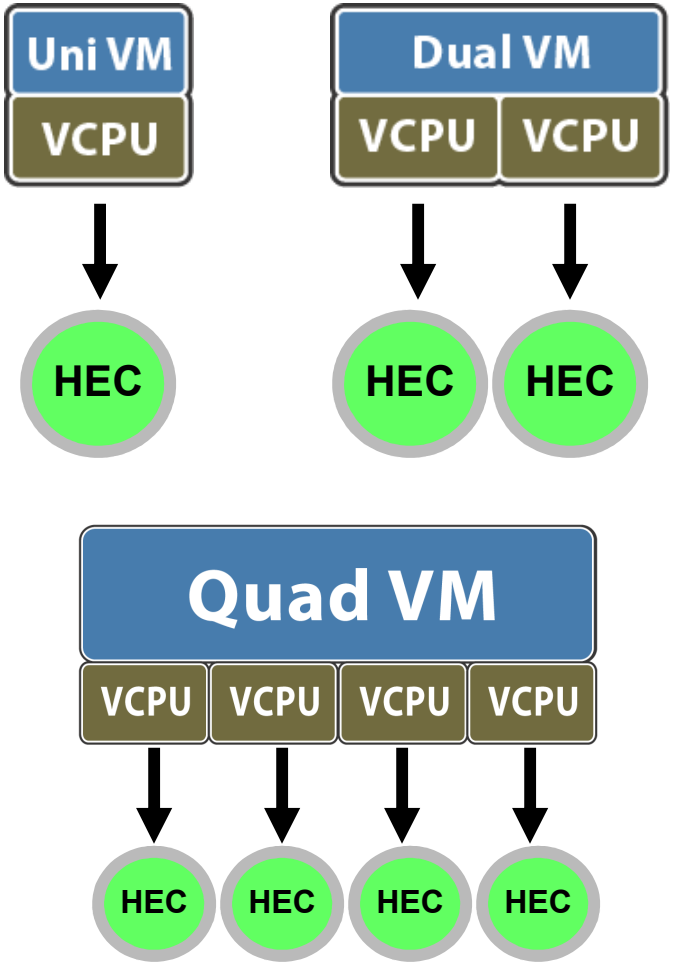
	Automatically managed by VMkernel	*Configured by virtual machine owner*	*Adjustable by administrator*
CPU cycles 	• Hyperthreading • Load balancing	• Virtual SMP	• Limit • Reservation • Share allocation
RAM 	• Transparent page sharing • `vmmemctl` • Use of VMkernel swap files for VMs	• Available memory	• Limit • Reservation • Share allocation
Disk bandwidth 		• Virtual machine file location	• Multipathing
Network bandwidth 		• Virtual switch with teamed NICs	• Traffic shaping

Virtual CPUs

A virtual machine can have up to eight virtual CPUs (VCPUs).

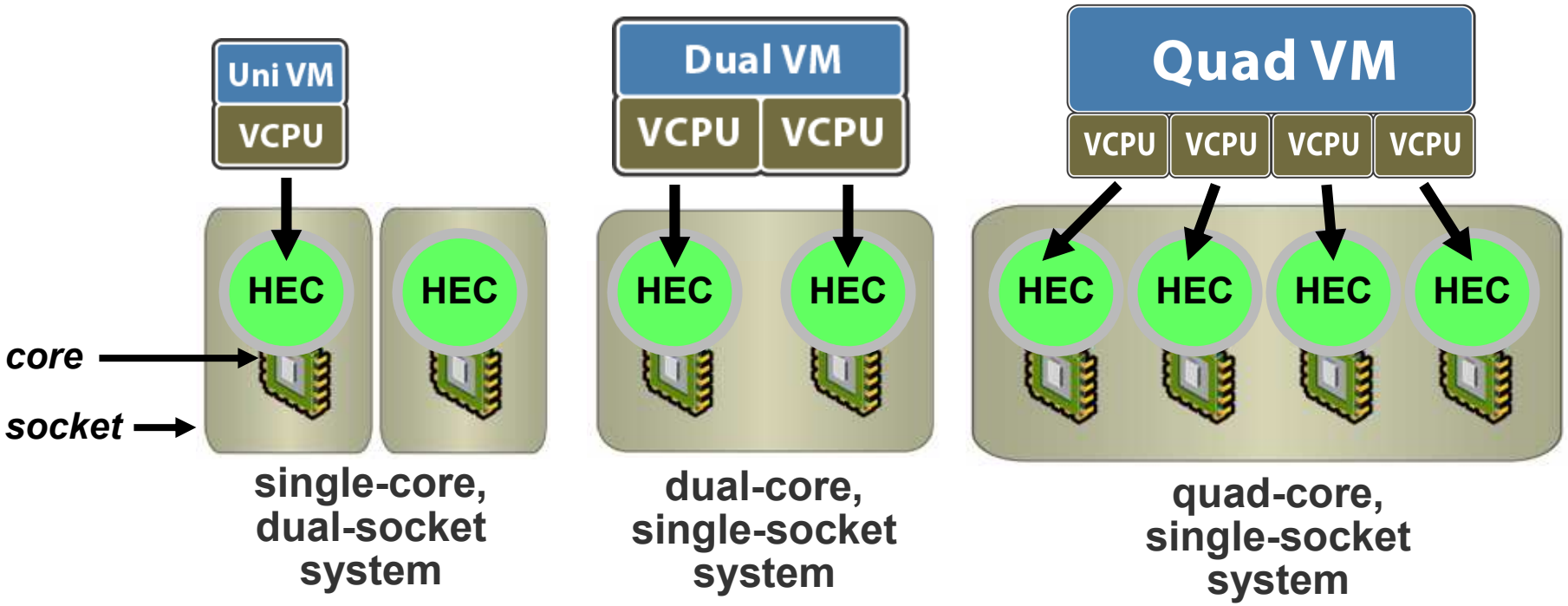
When a VCPU must be scheduled, the VMkernel maps a VCPU to a hardware execution context (HEC).

A hardware execution context is a processor's capability to schedule one thread of execution.



Hardware Execution Contexts

Different systems provide different numbers of hardware execution contexts.

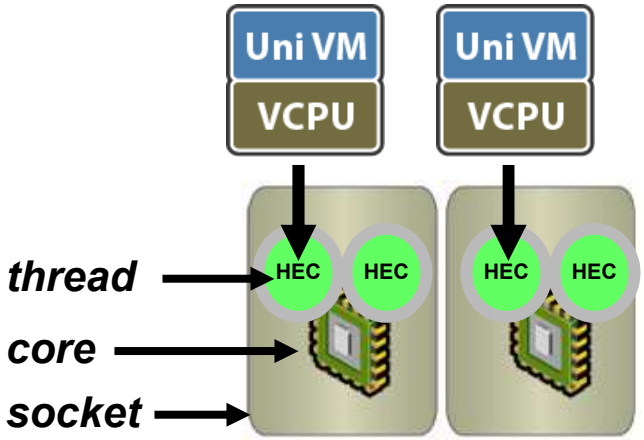


(Hyperthreading Not Enabled)

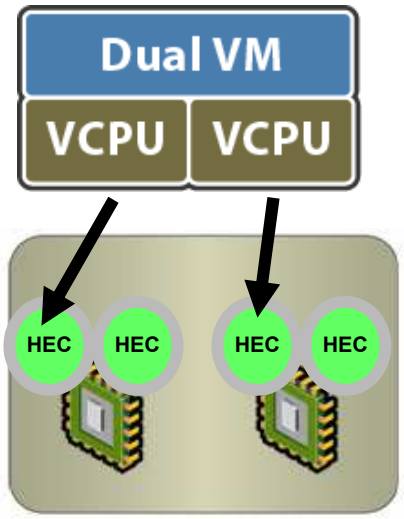
Hyperthreading

Enables a core to execute two threads, or sets of instructions, at the same time

Provides more hardware execution contexts for VCPUs to be scheduled



single-core, dual-socket system



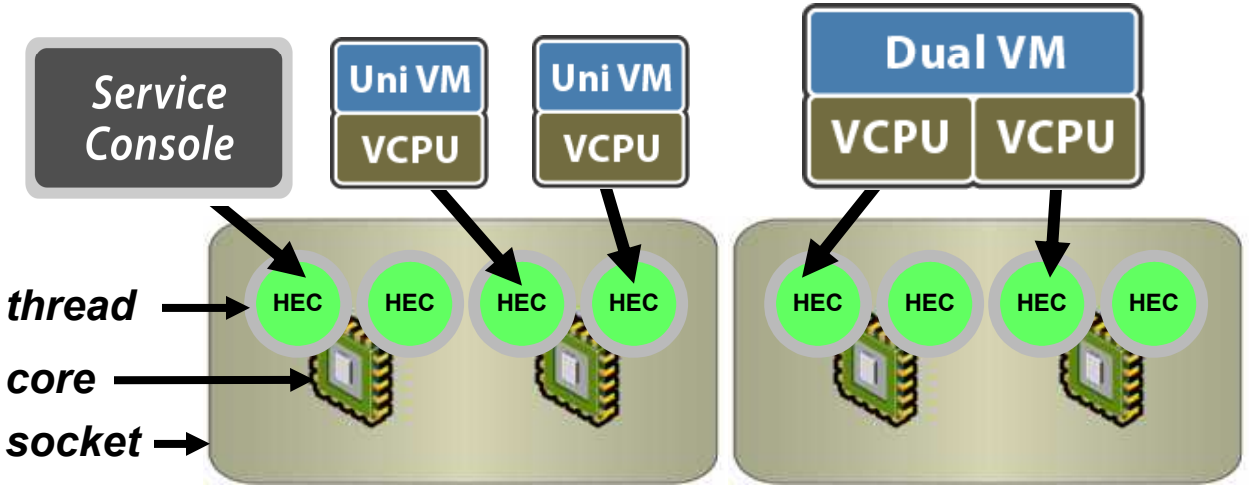
dual-core, single-socket system

VMkernel CPU Load Balancing

The VMkernel dynamically schedules virtual machines and the service console.

(VMware ESX™ only) The service console always runs on the first hardware execution context.

The VMkernel avoids scheduling multiple VCPUs on HECs in the same core.



hyperthreaded, dual-core, dual-socket system

Memory Virtualization Overview

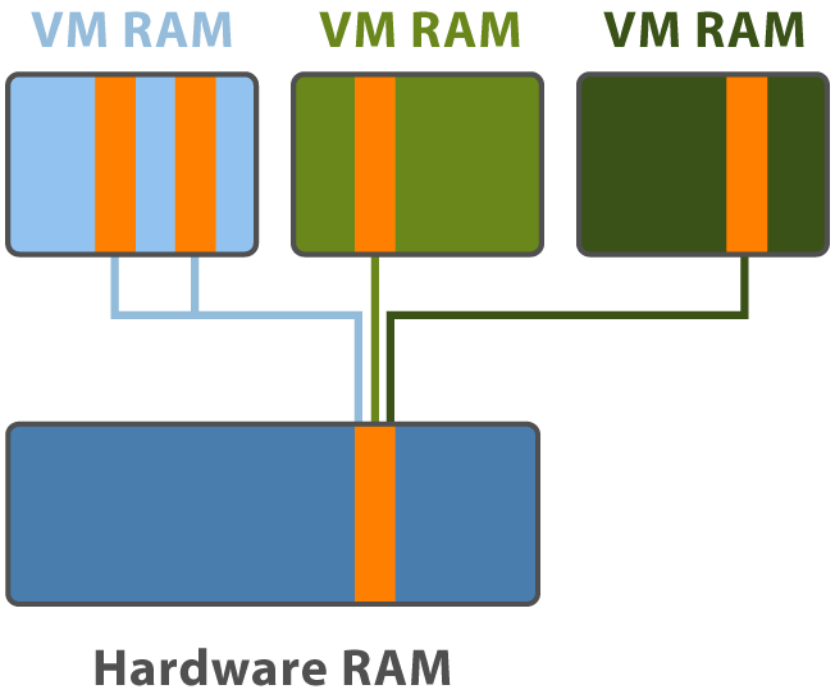
The VMkernel manages a machine's entire memory.

- On ESX and ESXi hosts, part of this memory is for the VMkernel.
- On ESX hosts only, part of this memory is for the service console.
- The rest is available for use by virtual machines.
  - Each virtual machine can reserve some amount of physical memory.
  - Each virtual machine also incurs some amount of overhead.

Virtual machines can use more memory than the physical machine has available.

- This is called memory overcommitment.

Transparent Memory Page Sharing



The VMkernel detects identical pages in virtual machines' memory and maps them to the same underlying physical page.

- No changes to guest operating system required

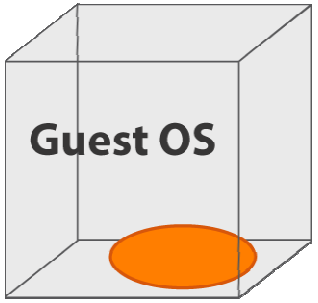
The VMkernel treats the shared pages as copy-on-write.

- Read-only when shared
- Private copies after write

Page sharing is always active, unless administratively disabled.

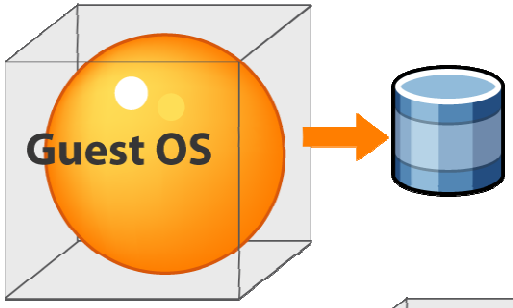
vmmemctl: Balloon-Driver Mechanism

Deallocate memory from selected virtual machines when RAM is scarce.



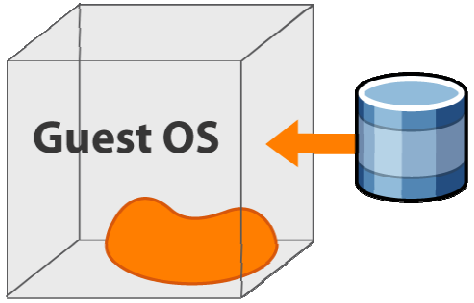
Ample memory. Balloon remains uninflated.

Inflate balloon.
(Driver demands memory from guest OS.)



Guest is forced to page out to its own paging area. The VMkernel reclaims memory.

Deflate balloon.
(Driver relinquishes memory.)



Guest can page in. ESX grants memory.

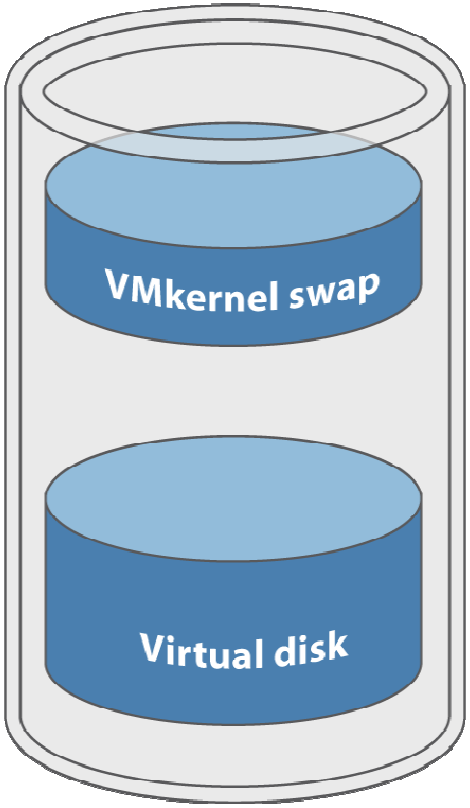
VMkernel Swap

Each powered-on virtual machine needs its own VMkernel swap file.

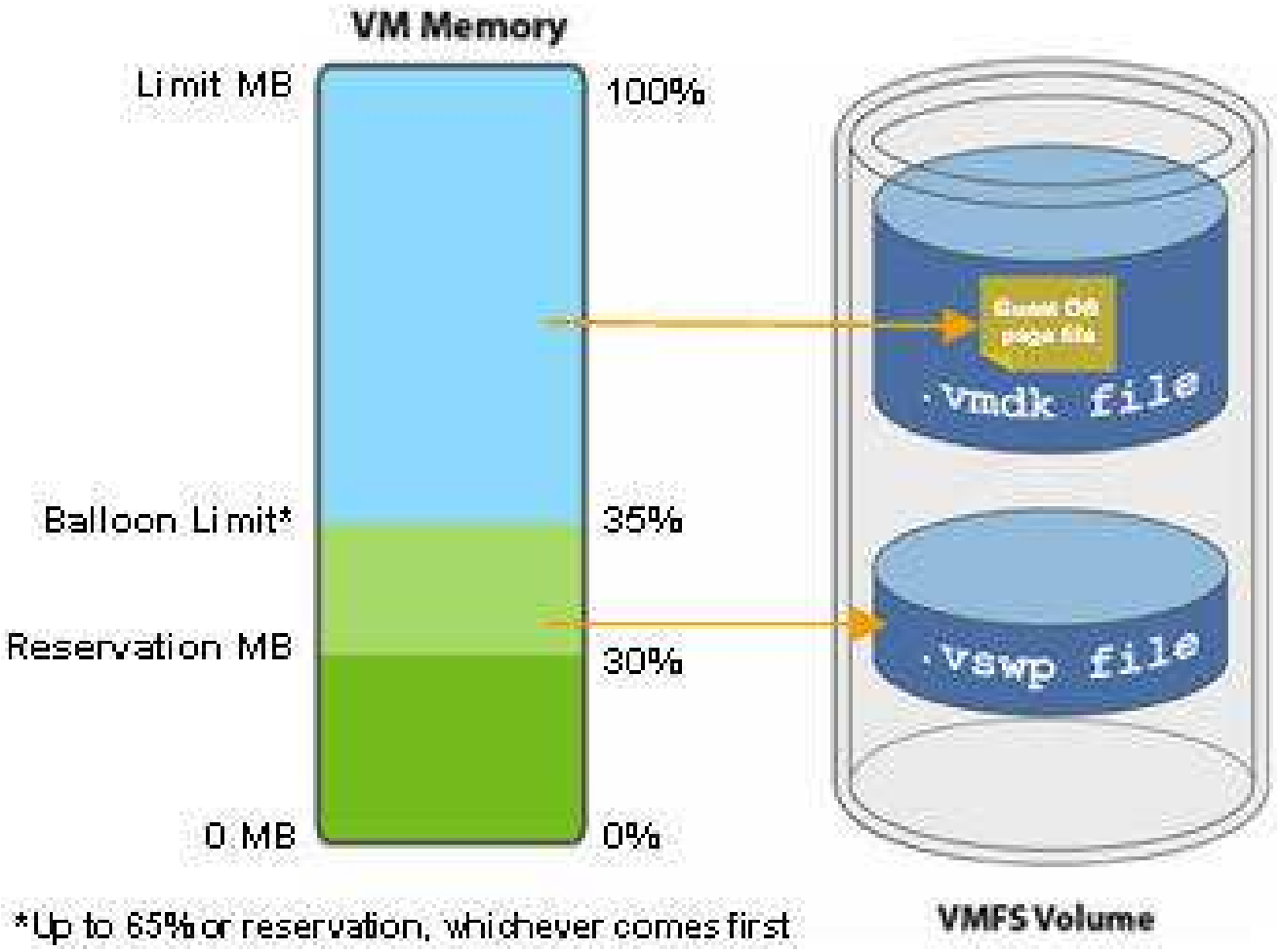
- Created when the VM is powered on, deleted when the VM is powered off
- Default location: same VMware vStorage VMFS volume as virtual machine's boot disk
- Size equal to the difference between the memory guaranteed to it, if any, and the maximum it can use
- Allows the VMkernel to swap out the virtual machine entirely if memory is scarce

Use of VMkernel swap is a last resort.

- Performance will be noticeably slow.



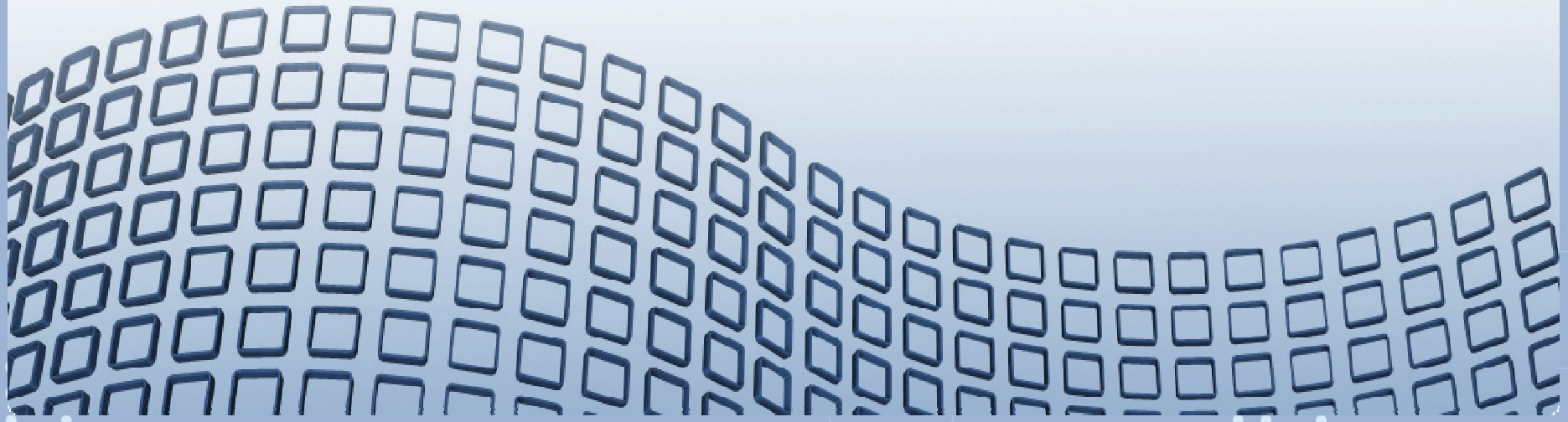
Ballooning vs. VMkernel Swapping



Lesson Summary

- The VMkernel uses hyperthreading and load balancing to manage CPU allocated across virtual machines.
- The VMkernel uses transparent page sharing, the balloon driver mechanism, and VMkernel swap files to manage memory allocation across virtual machines.

Lesson 2:
Monitoring Resource Usage



Lesson Objectives

- Monitor a virtual machine's resource usage
  - CPU
  - Memory
  - Disk
  - Network bandwidth

Performance-Tuning Methodology

Assess performance.

- Use appropriate monitoring tools.
- Record a numerical benchmark before changes.

Identify the limiting resource.

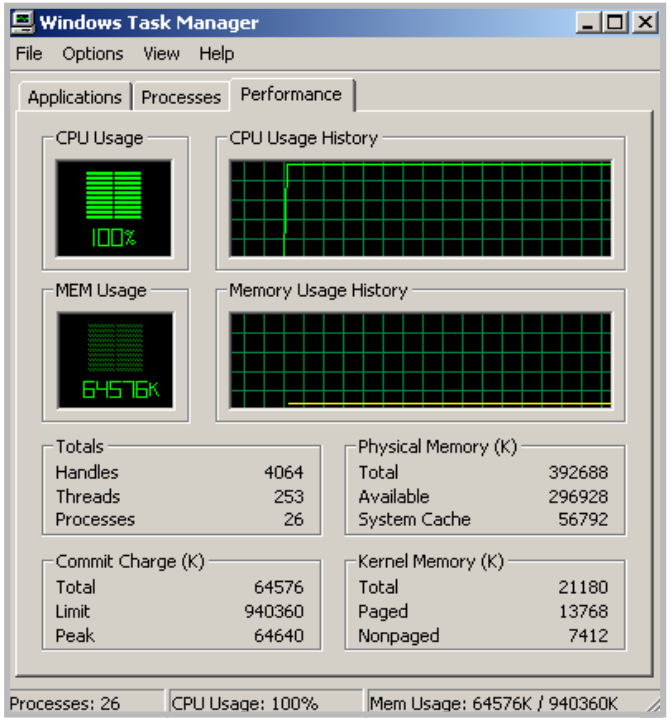
Make more resource available.

- Allocate more.
- Reduce competition.
- Log your changes!

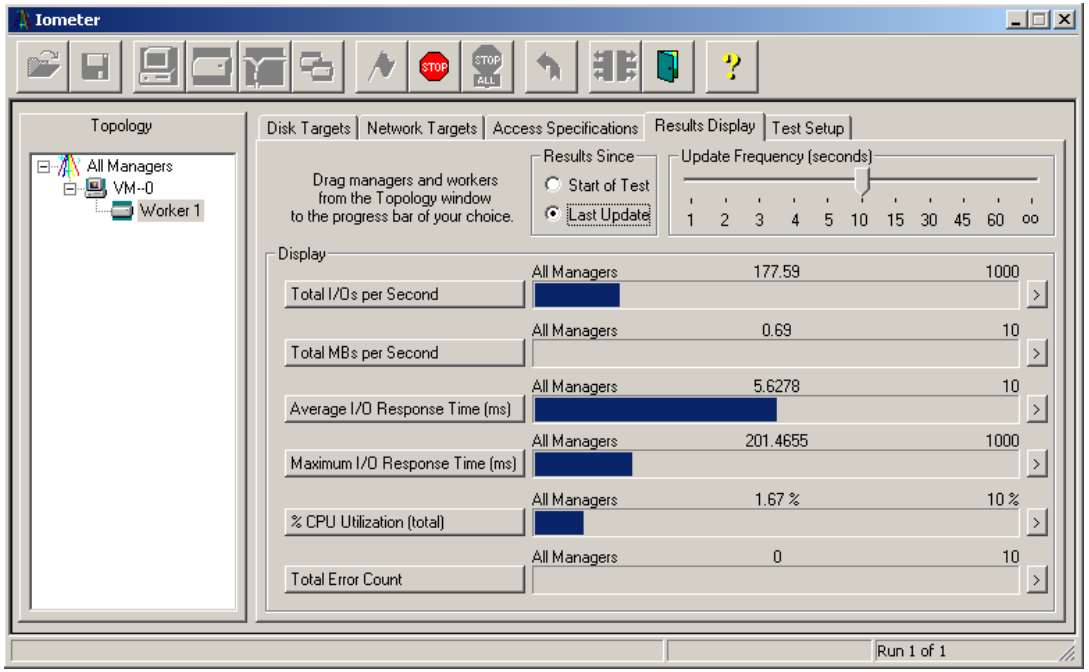
Benchmark again.

Do not *make casual changes to production systems!*

Guest Operating System Monitoring Tools

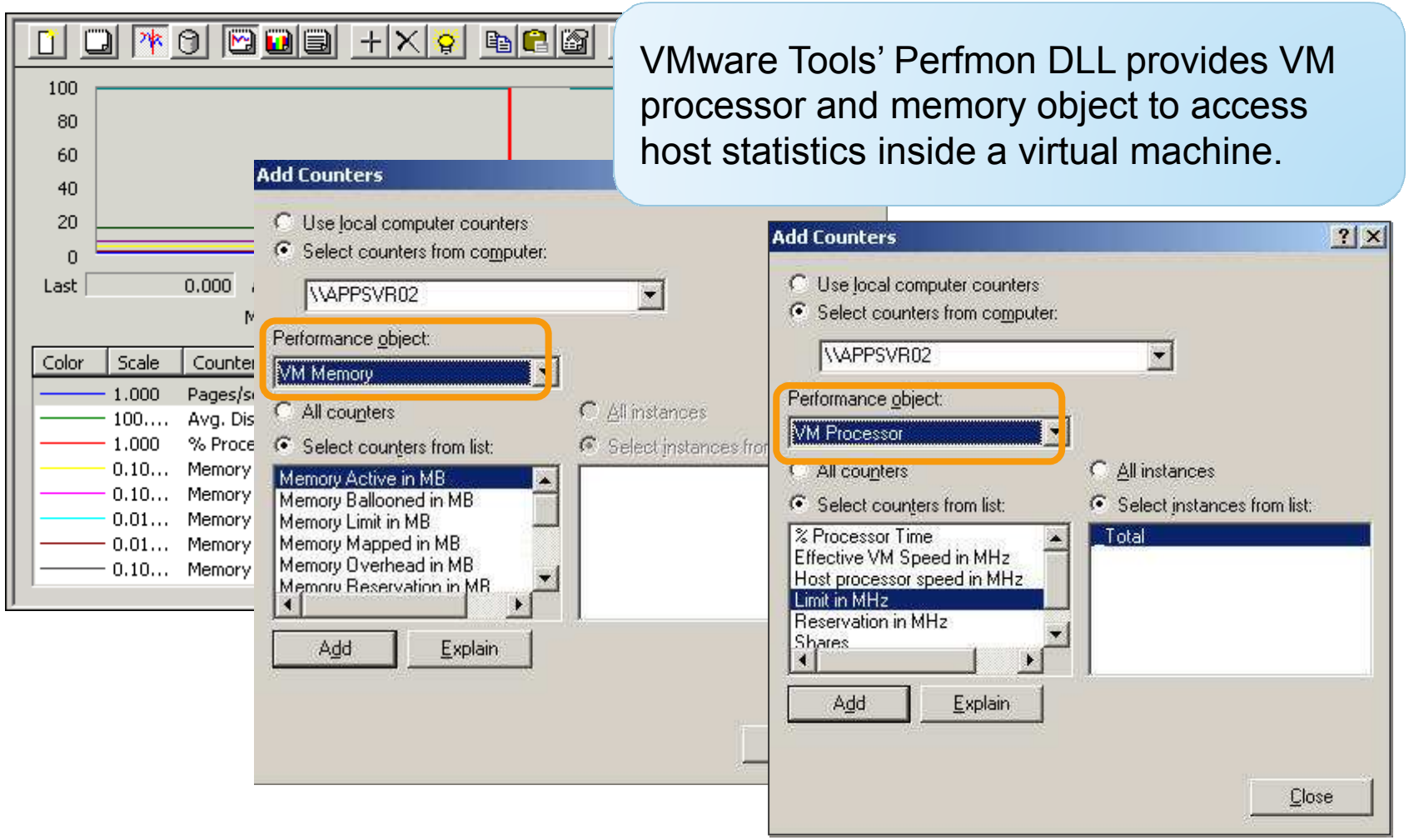


Task Manager



Iometer

Using Perfmon to Monitor VM Resources

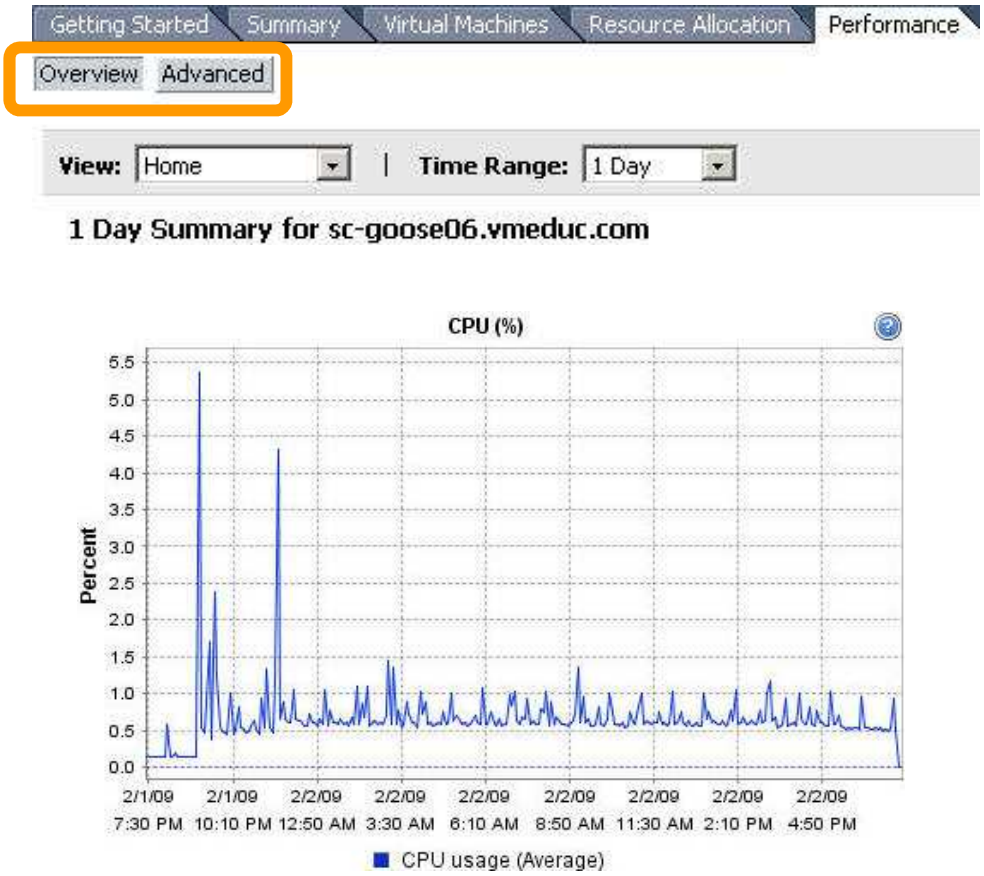


VMware Tools' Perfmon DLL provides VM processor and memory object to access host statistics inside a virtual machine.

vCenter Server Performance Charts

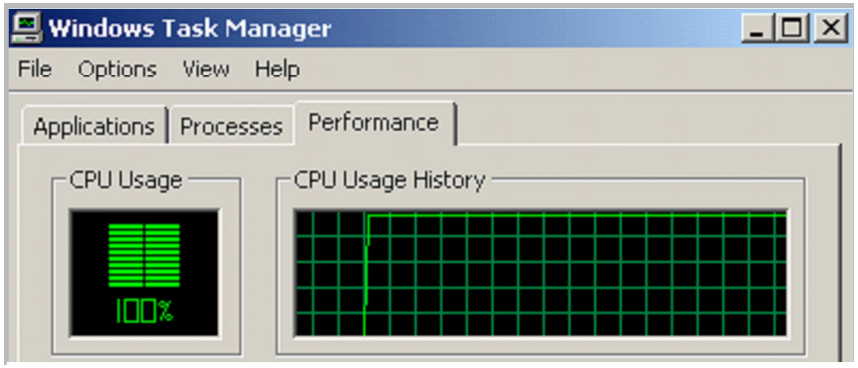
The Performance tab displays two kinds of charts for hosts and virtual machines:

- Overview charts
  - Display the most common metrics for an object
- Advanced charts
  - Display data counters not shown in the Overview charts

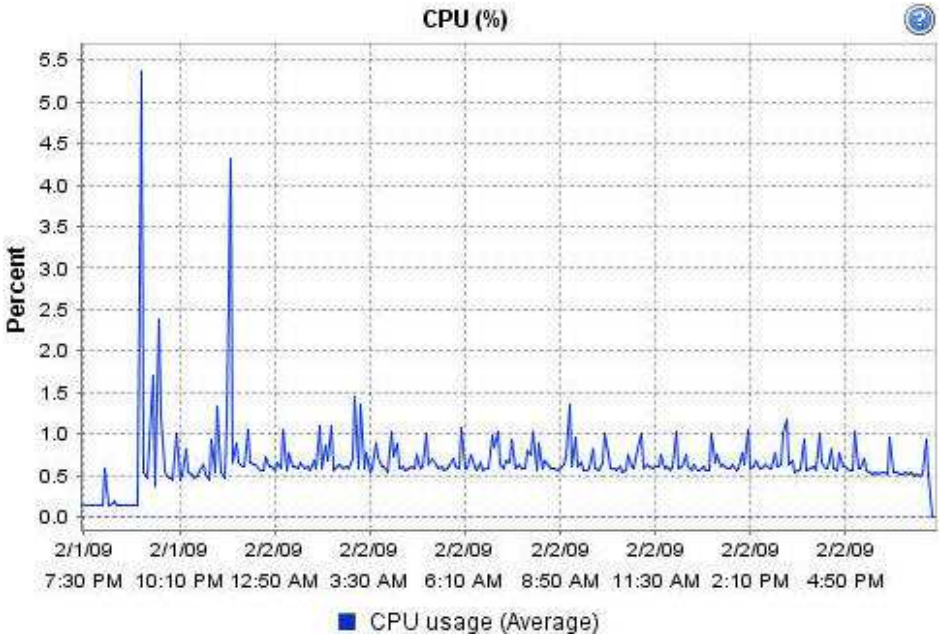


Interpreting Data from the Tools

VMware vCenter™ Server monitoring tools and guest operating system monitoring tools provide different points of view.

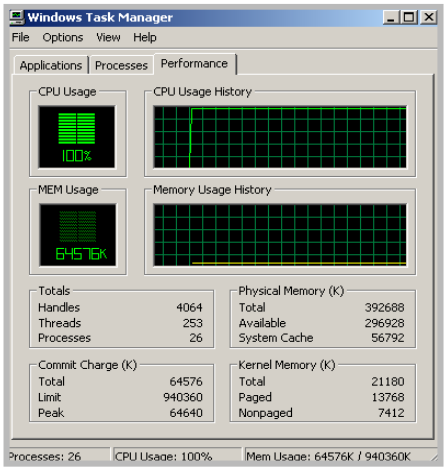


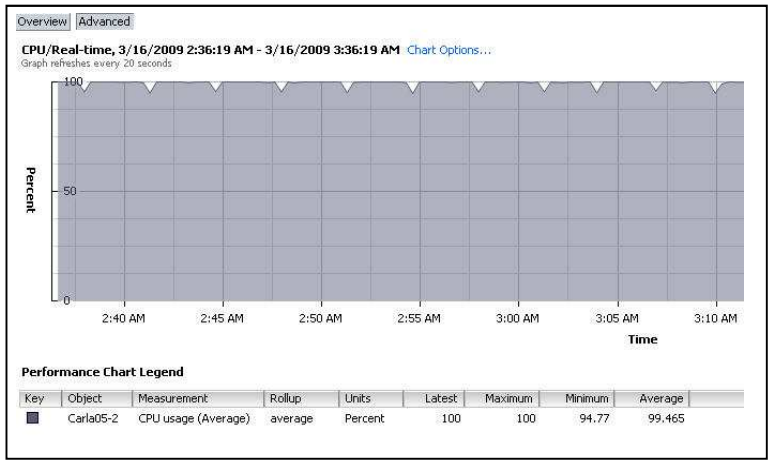
Task Manager in guest operating system



CPU Usage chart for host

Is the Virtual Machine CPU-Constrained?

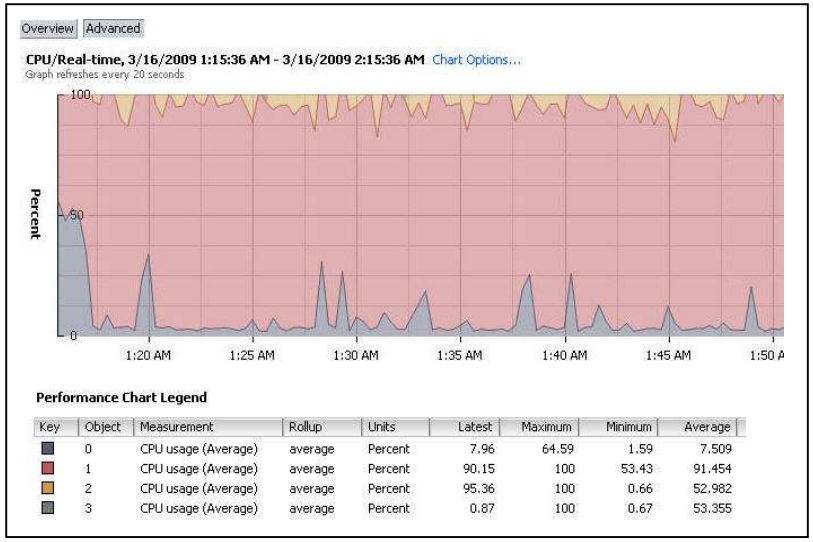




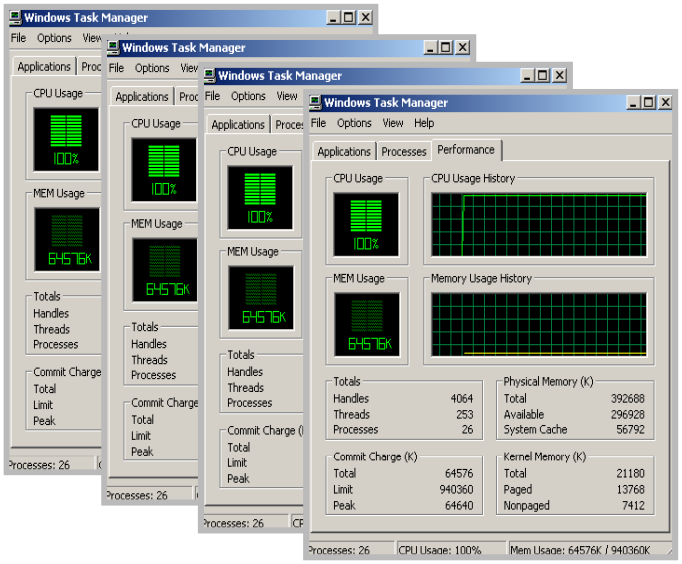
Check the virtual machine's CPU utilization.

If CPU utilization is continuously high, then the virtual machine is constrained by CPU.

However, the host might have enough CPU for other virtual machines to run.



Are Virtual Machines CPU-Constrained?



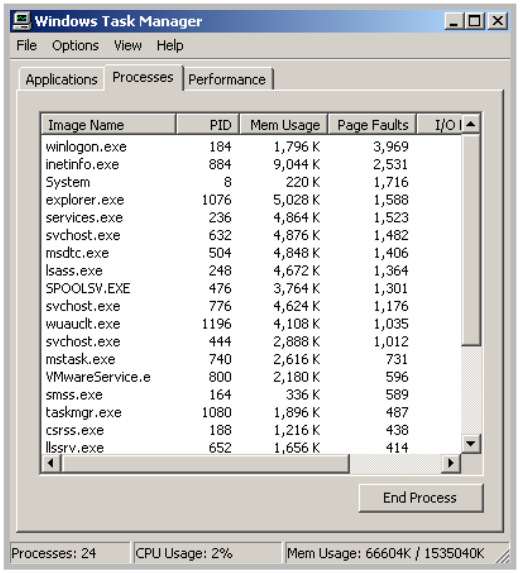
Task Manager of several OSes

CPU Ready graph of several VMs

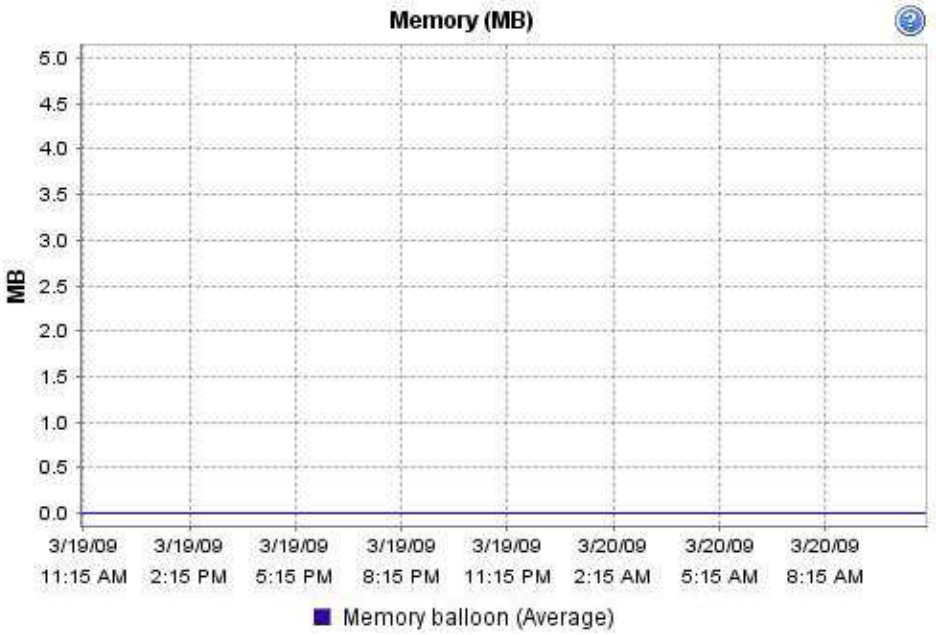
Multiple virtual machines are constrained by CPU if:

- There is high CPU utilization in the guest operating system.
- There are relatively high CPU ready values for the virtual machines.

Is a Virtual Machine Memory-Constrained?



Task Manager inside VM

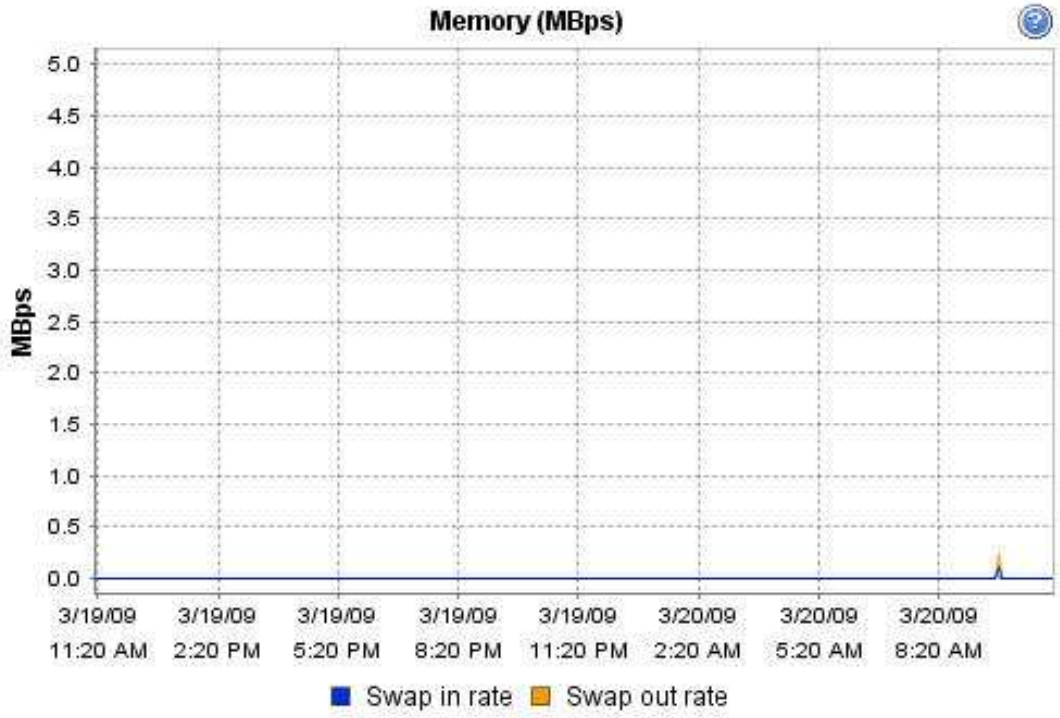


VM ballooning activity

Check the virtual machine's ballooning activity:

- If ballooning activity is high, this might not be a problem if all virtual machines have sufficient memory.
- If ballooning activity is high and the guest operating system is swapping, then the virtual machine is constrained for memory.

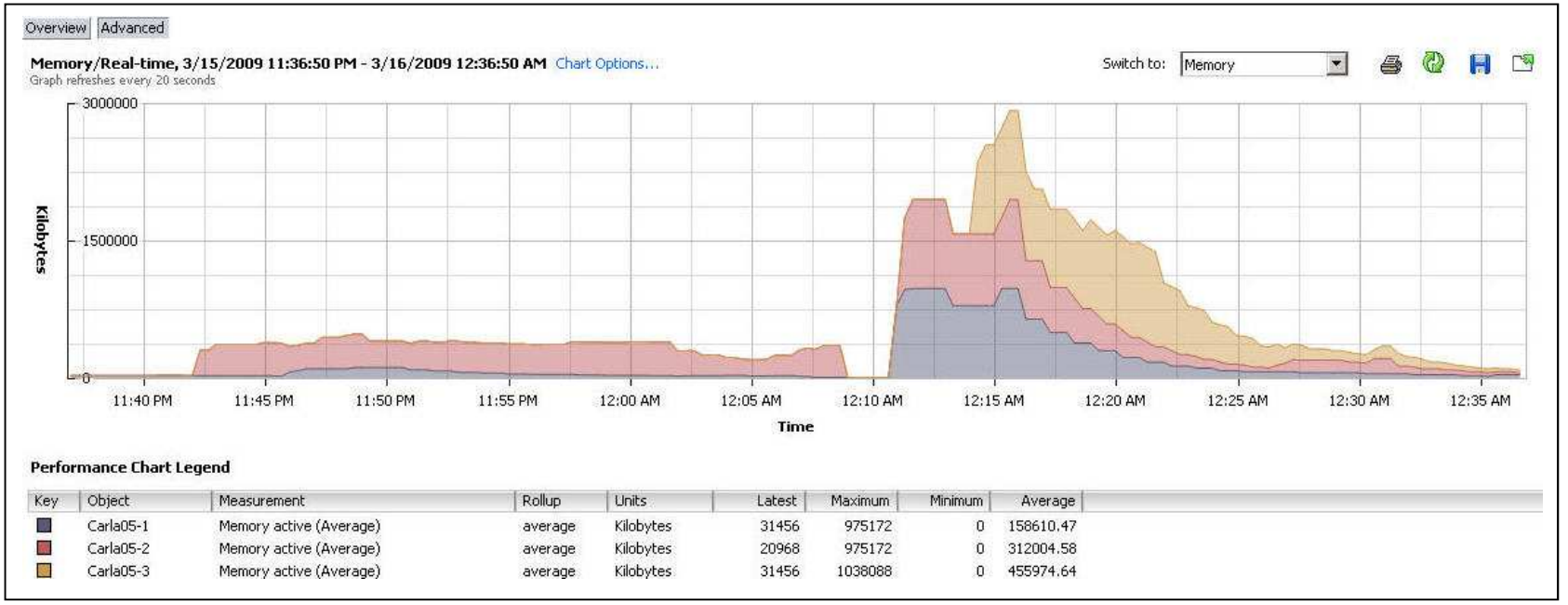
Is the Host Memory-Constrained?



Host swap-ins/swap-outs

If the virtual machines are being swapped in and out, then the host memory is probably overcommitted.

Monitoring Active Memory of a Virtual Machine



Key	Object	Measurement	Rollup	Units	Latest	Maximum	Minimum	Average
	Carla05-1	Memory active (Average)	average	Kilobytes	31456	975172	0	158610.47
	Carla05-2	Memory active (Average)	average	Kilobytes	20968	975172	0	312004.58
	Carla05-3	Memory active (Average)	average	Kilobytes	31456	1038088	0	455974.64

Monitor for increases in active memory on the host:

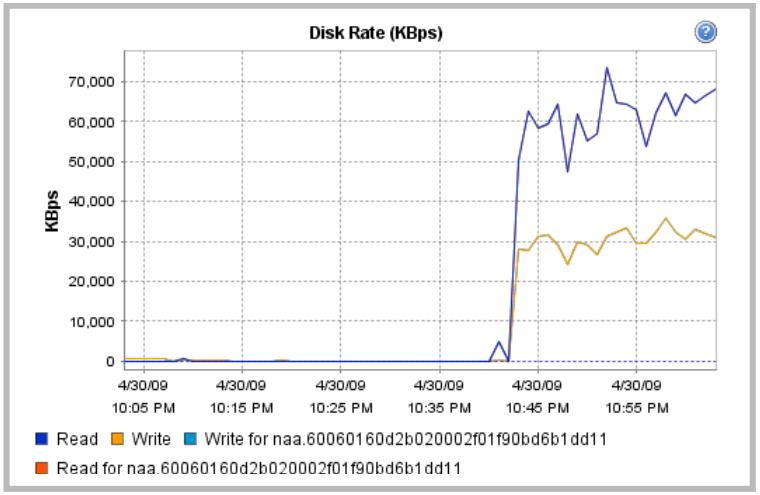
- Host active memory refers to active physical memory used by virtual machines and the VMkernel.
- If amount of active memory is high, this could lead to virtual machines that are memory-constrained.

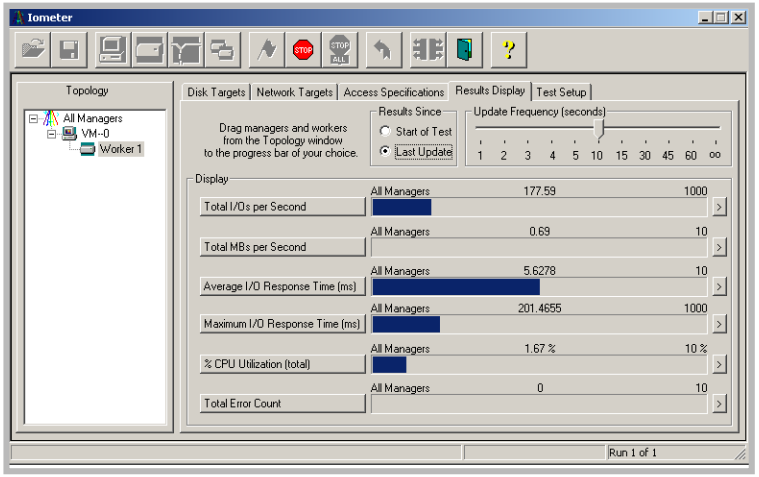
Are Virtual Machines Disk-Constrained?

Disk-intensive applications can saturate the storage or the path.

If you suspect that a virtual machine is constrained by disk access:

- Measure the effective bandwidth between virtual machine and the storage.
- Measure the resource consumption using performance graphs.





Monitoring Disk Latency

To determine disk performance problems, monitor two disk latency data counters:

- Kernel disk command latency
  - The average time spent in the VMkernel per SCSI command
  - High numbers (greater than 2–3ms) represent either an overworked array or an overworked host.
- Physical device command latency
  - The average time the physical device takes to complete a SCSI command
  - High numbers (greater than 15–20ms) represent a slow or overworked array.

Are Virtual Machines Network-Constrained?

Network-intensive applications often bottleneck on path segments outside the ESX host.

- Example: WAN links between server and client

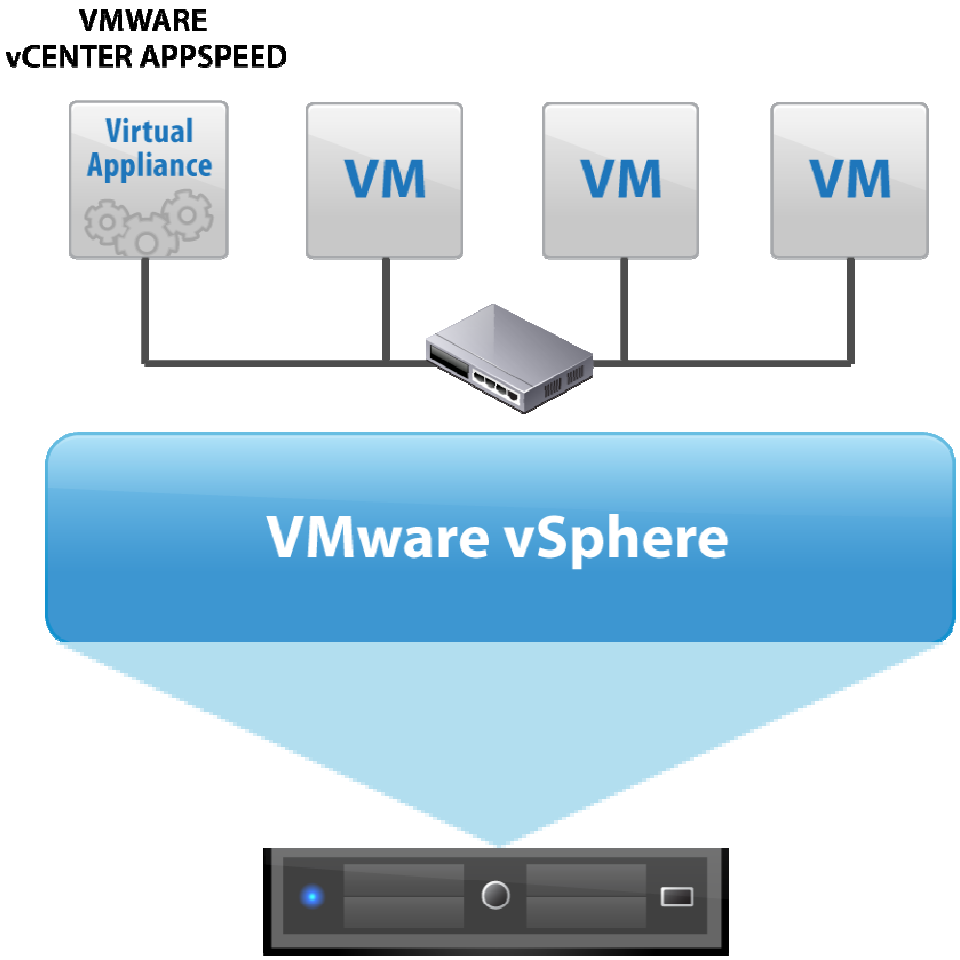
If you suspect that a virtual machine is constrained by the network:

- Confirm that VMware Tools is installed.
  - Enhanced network drivers are available.
- Measure the effective bandwidth between the virtual machine and its peer system.

Application Performance Management: AppSpeed

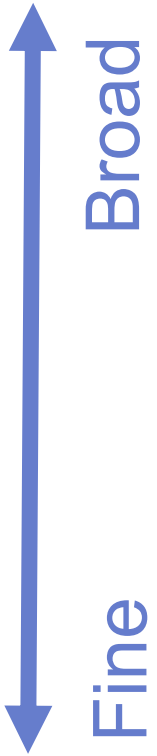
A virtual appliance for proactive application performance management

Provides visibility into the performance of multitier applications running in virtual machines



Improving Virtual Machine Performance

Methods for improving virtual machine performance:



Use a VMware Distributed Resource Scheduler cluster.

Use storage multipathing.

Use NIC teaming.

Modify resource pool's CPU and memory limits and reservations.

Modify virtual machine's CPU and memory and reservations.

Use network traffic shaping.

Lab 15

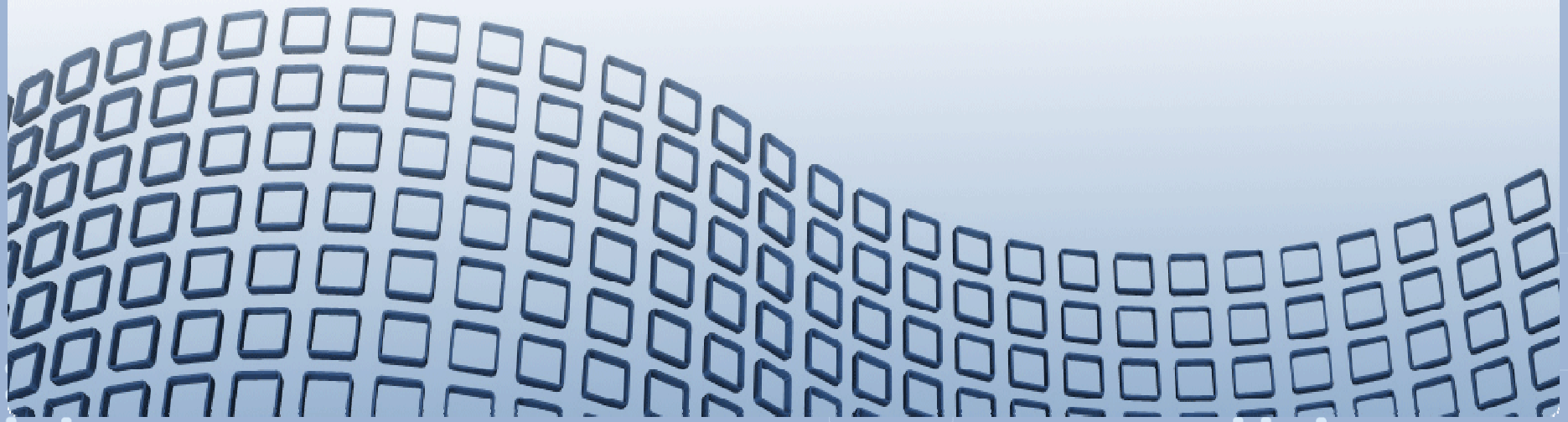
In this lab, you will see how CPU workload is reflected by system-monitoring tools.

1. Monitor CPU utilization using vCenter Server.
2. Run a CPU-intensive application.
3. Undo changes made to your virtual machines for this lab.

Lesson Summary

- VMware Tool's Perfmon DLL provides a way to get CPU and memory statistics inside a Windows virtual machine.
- Use the Performance tab for real-time graphs showing a virtual machine's resource usage.
- Virtual machine performance can be improved by using shares and reservations, balancing the load with DRS, and storage and network multipathing.

Lesson 3: Using Alarms



Lesson Objectives

- Create alarms with condition-based triggers
- Create alarms with event-based triggers
- View and acknowledge triggered alarms

What Is an Alarm?

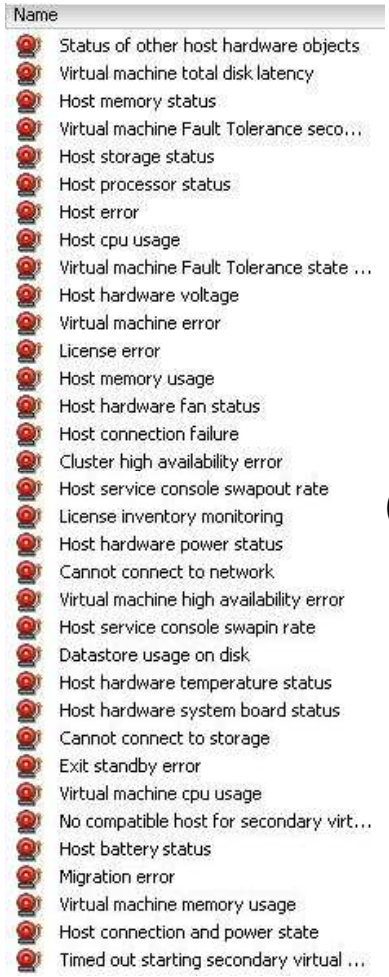
An alarm is a notification that occurs in response to selected events or conditions that occur with an object in the inventory.

Default alarms exist for various inventory objects.

- Many default alarms for hosts and virtual machines

You can create custom alarms for a wide range of inventory objects.

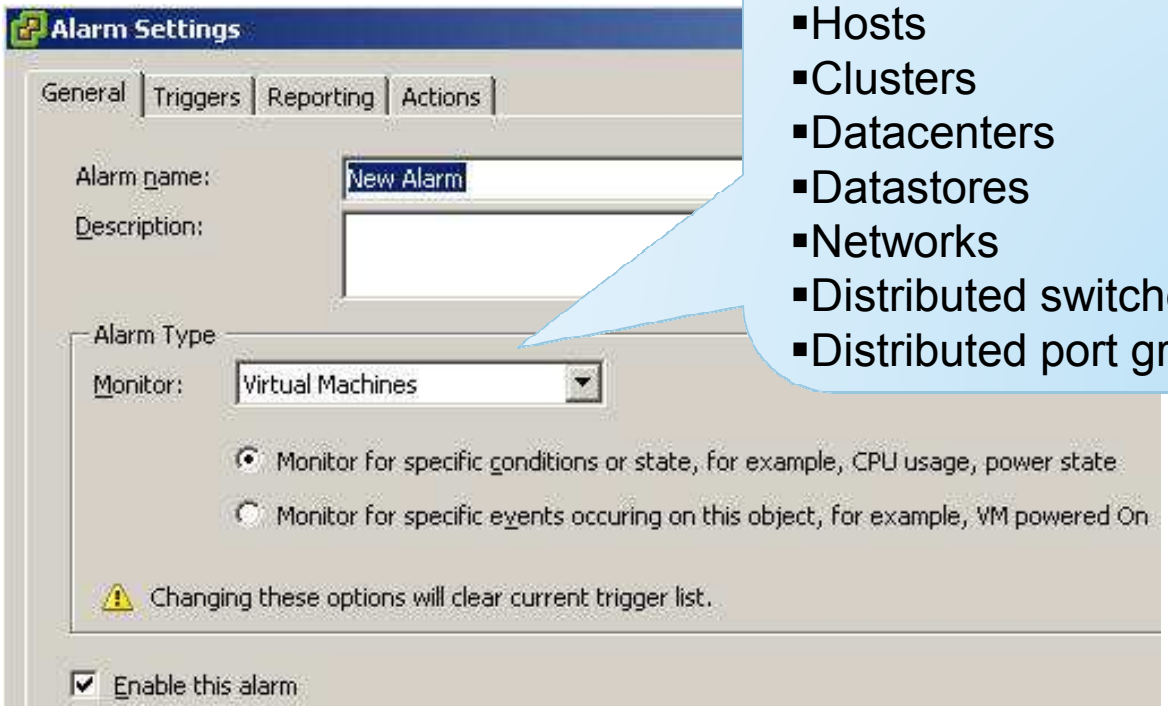
- Virtual machines, hosts, clusters, datacenters, datastores, networks, distributed switches, and distributed port groups



Default Alarms (partial list)

Creating an Alarm

Right-click inventory object, then choose Alarm **>** Add Alarm**.**



Alarm support for:

- Virtual machines
- Hosts
- Clusters
- Datacenters
- Datastores
- Networks
- Distributed switches
- Distributed port groups

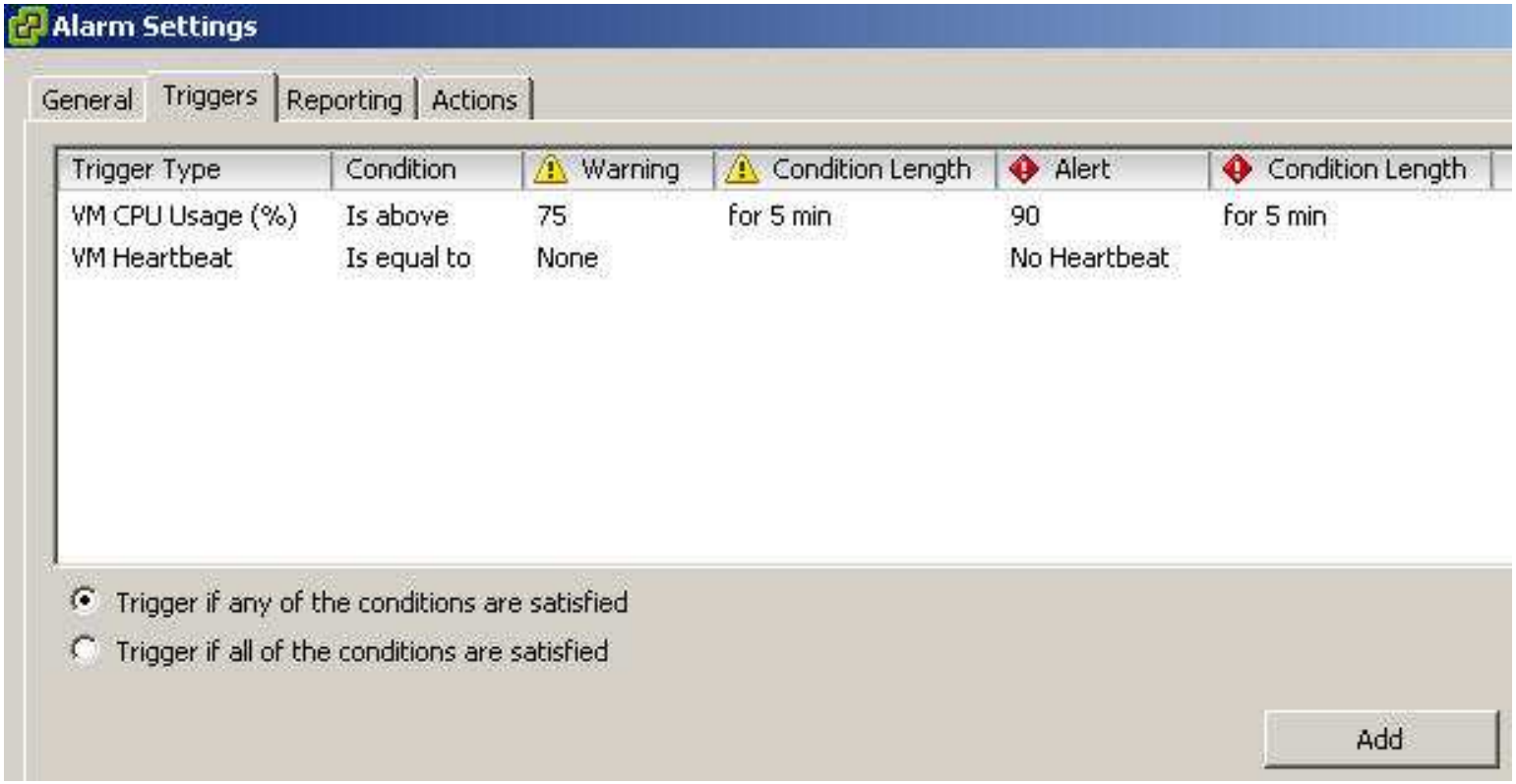
Alarm Triggers

An alarm is comprised of a trigger. There are two types:

- Condition, or state, trigger – Monitors the current condition or state; for example:
  - A virtual machine's current snapshot is above 2GB in size.
  - A host is using 90 percent of its total memory.
  - A datastore has been disconnected from all hosts.
- Event – Monitors events; for example:
  - The health of a host's hardware has changed.
  - There are insufficient licenses in the datacenter.
  - A distributed virtual port group has been reconfigured.

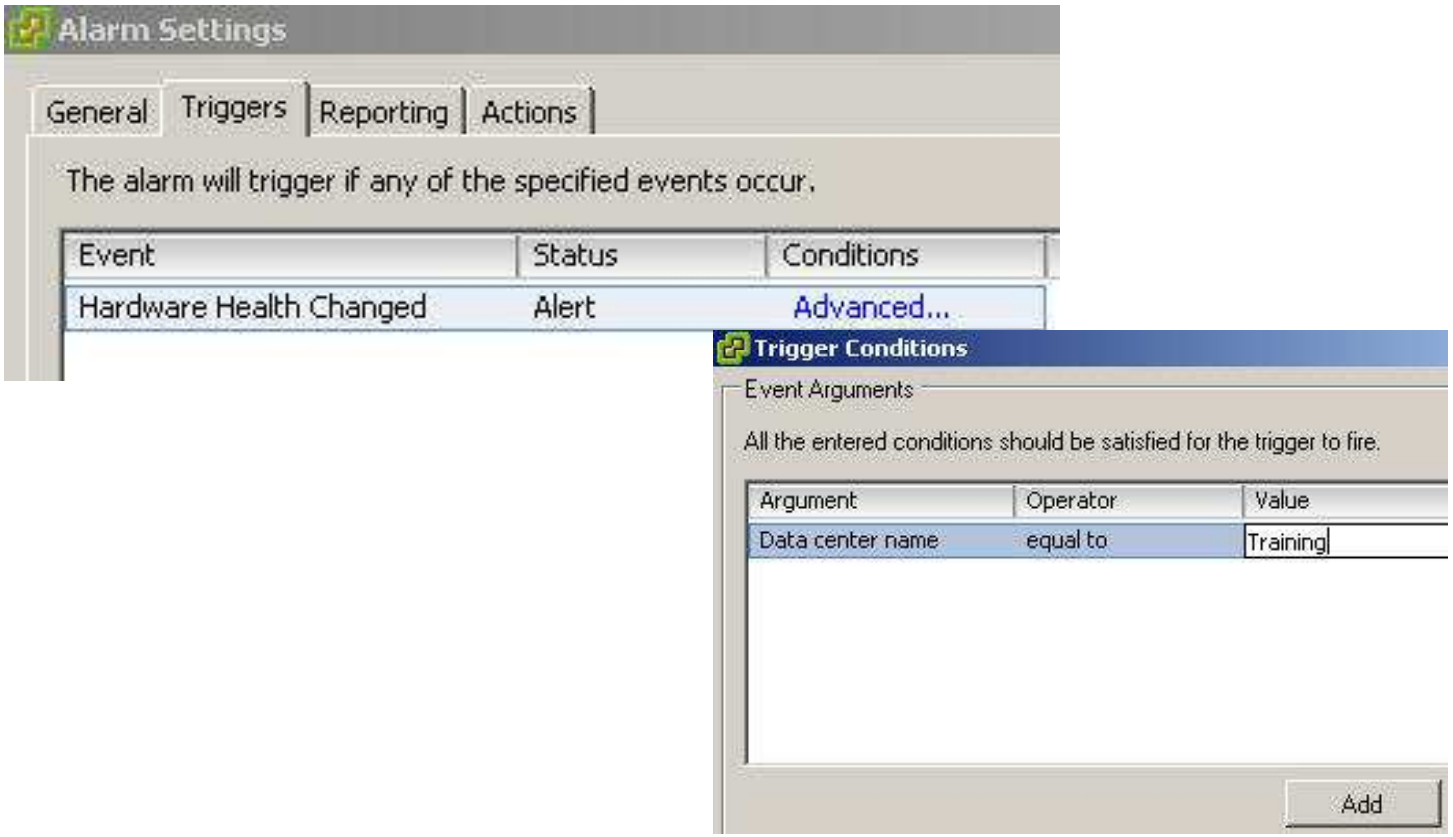
Configuring Condition Triggers

Condition triggers for a virtual machine



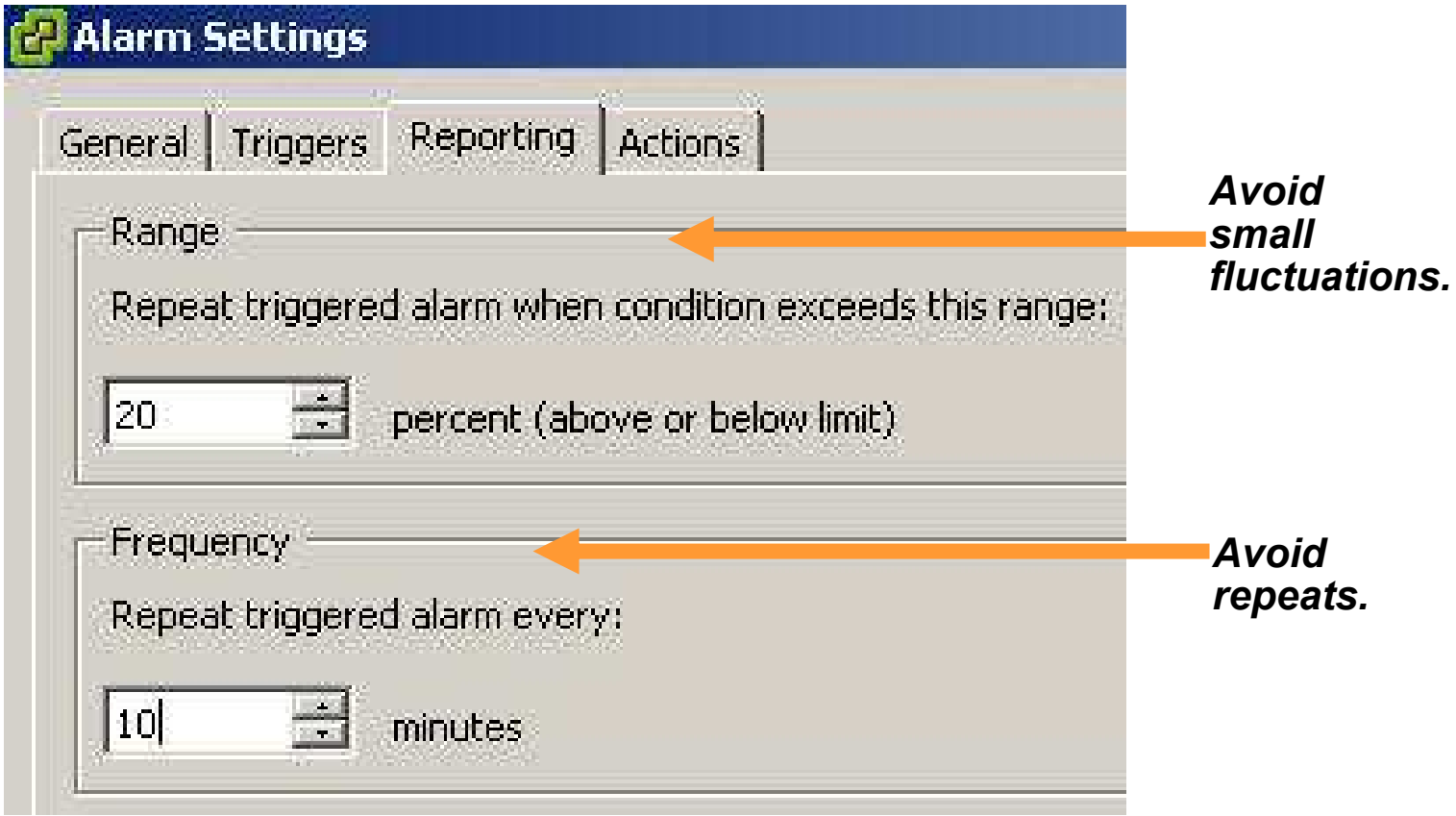
Configuring Event Triggers

Event trigger for a host



Configuring Reporting Options

Use the Reporting pane to avoid needless re-alarms.

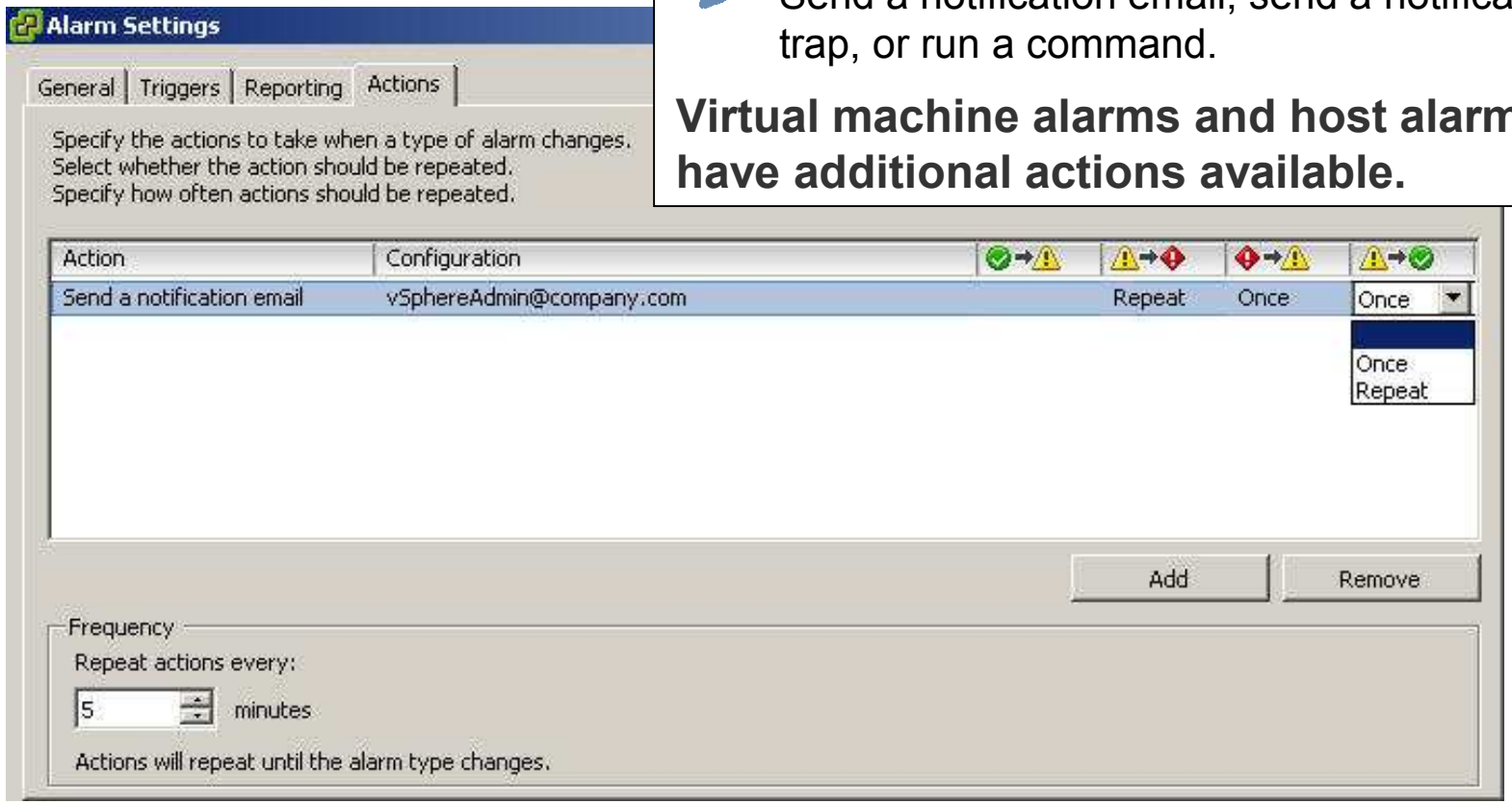


Configuring Actions

Every alarm type has the following actions:

- Send a notification email, send a notification trap, or run a command.

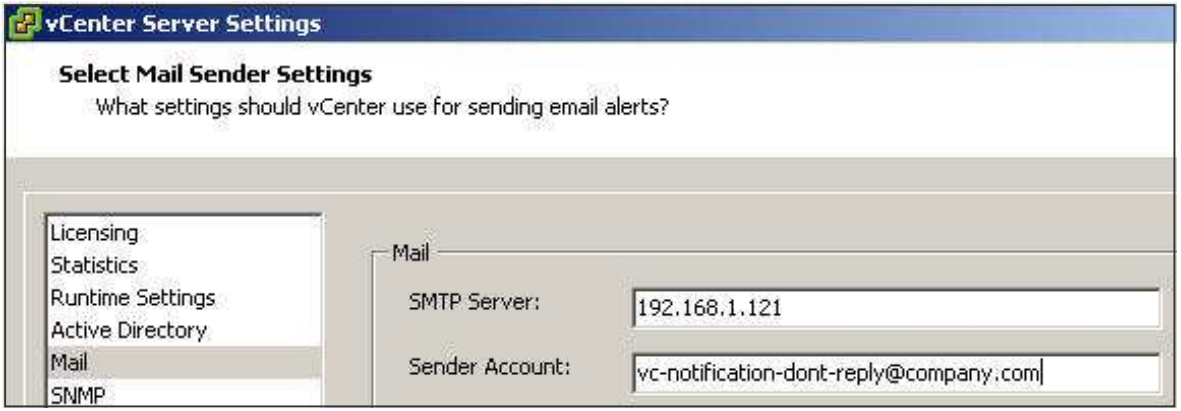
Virtual machine alarms and host alarms have additional actions available.



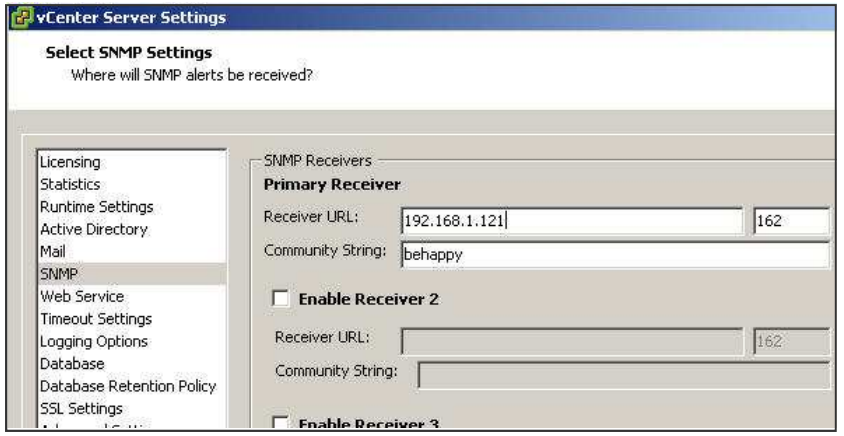
Configuring vCenter Server Notifications

In the menu bar, choose Administration **>** vCenter Server Settings**.**

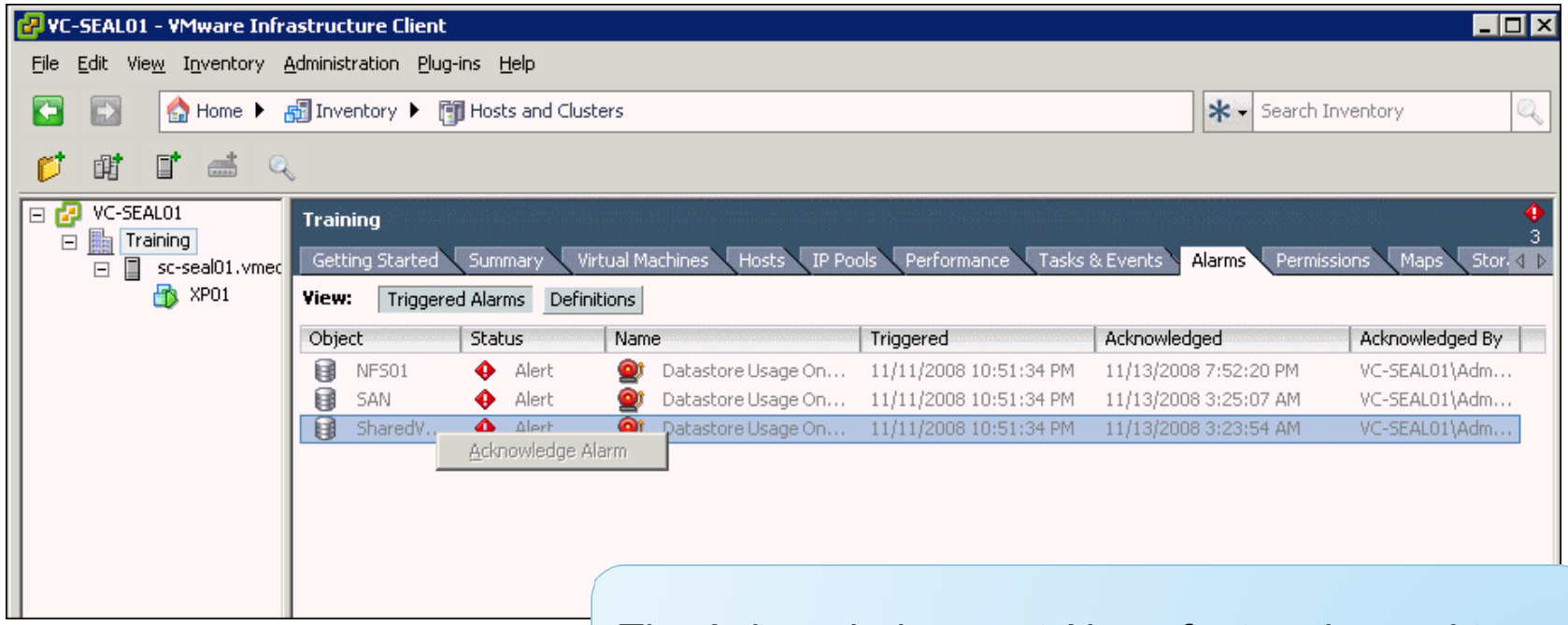
Click Mail **to set SMTP parameters.**



Click SNMP **to specify trap destinations.**



Viewing and Acknowledging Triggered Alarms



The Acknowledgement Alarm feature is used to track when triggered alarms are addressed.

Lab 16

In this lab, you will demonstrate the vCenter Server alarm feature.

1. Create a virtual machine alarm.
2. Trigger the virtual machine alarm, then acknowledge it.
3. Disable the alarms.

Lesson Summary

- Condition-based alarms monitor the current condition or state of virtual machines, hosts, and datastores.
- Event-based alarms monitor events that occur in response to operations occurring with an inventory object.
- The Acknowledgement Alarm feature is used to track when triggered alarms have been addressed.

Key Points

- The VMkernel has built-in mechanisms (such as CPU load balancing and transparent page sharing) for managing the CPU and memory allocation on an ESX/ESXi host.
- The Performance tab allows you to monitor a host or virtual machine's performance in real time or over a period of time.
- Monitor your vCenter Server inventory using alarms, which notify you when selected events or conditions have occurred.