Performance Analysis Methods

ESX Server 3

The wide deployment of VMware® Infrastructure 3 in today's enterprise environments has introduced a need for methods of optimizing the infrastructure performance. Key to this exercise is the ability to identify the root cause of performance problems on VMware ESX Server systems. This paper discusses the process for identifying performance bottlenecks on ESX Server systems and recommends actions to correct the problems you identify.

Because virtualization enables you to consolidate multiple physical servers onto a single machine, traditional operating system analysis tools that are unaware of virtualization either miss critical data or produce invalid results. VMware Infrastructure 3 provides two ways to monitor ESX Server performance.

- VirtualCenter provides capabilities for a simple, graphical first-pass analysis of host performance.
- The `esxtop` utility, the command line performance tool available on ESX Server, offers capabilities for more detailed monitoring of host performance. For information on using `esxtop`, see "Appendix B: Using the esxtop Utility" in the VMware Infrastructure 3 *Resource Management Guide* (see "References" on page 12 for a link).

This paper covers the following topics, with recommendations in each section to match problems you might identify.

CPU Analysis

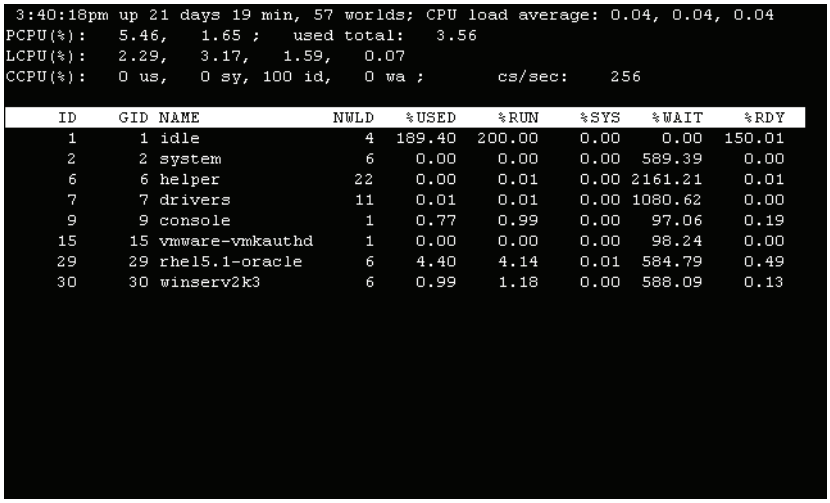
CPU load is generated by

- The guest operating system running inside the virtual machine
- The applications running in the virtual machine
- ESX Server, as it provides a virtual interface to the hardware

Although the work performed by ESX Server does cause some CPU load, applications in the virtual machines generate the great majority of processing on a system. A solid understanding of the workload profile of those applications, whether they are running in a virtual environment or directly on hardware, can help you analyze CPU usage.

Check CPU Utilization

Start `esxtop` by entering the command in the ESX Server host's service console. By default, `esxtop` shows CPU utilization. To ensure this data is displayed, press `c`. The following screen capture shows example data produced on a test system.

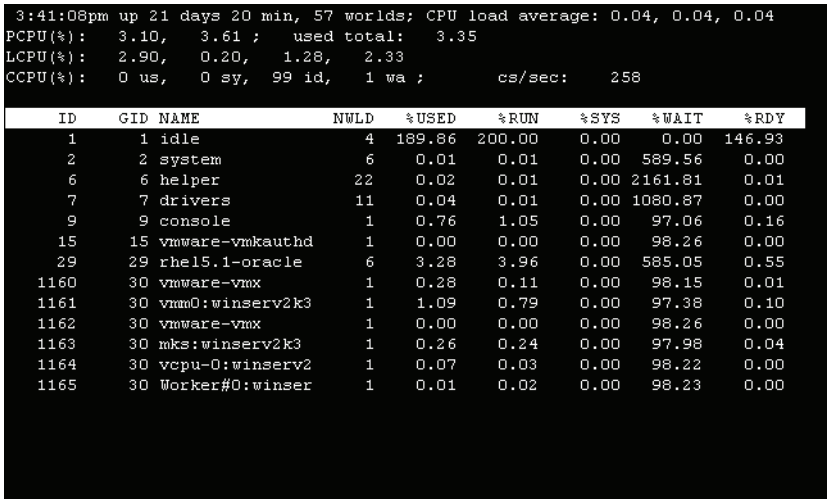


The output includes the following information:

- The `PCPU(%)` line in the header shows utilization for the physical processors on the host system by core and the total physical CPU usage. The comma-delimited data first displayed shows core utilization followed by `used total`, which averages utilization of all cores.
- The `LCPU(%)` line shows the percentage of CPU utilization per logical CPU. The percentages for the logical CPUs mapping to a single physical core add up to 100 percent. This line appears only if hyperthreading is present and enabled.
- The `CCPU(%)` line shows the percentages of total CPU time as reported by the ESX Server service console. If you run any third-party software, such as management agents and backup agents, inside the service console, you might see a high `CCPU(%)` number.
- An idle "world" is running. In ESX Server, a world is a managed execution entity similar to the operating system concept of a process. The `%USED` entry of that idle world displays the percentage of CPU cycles that remain unused. If `esxtop` reports less than 100 percent utilization for the idle world, that means only a fraction of one physical core remains available for additional work. The maximum value for this number can be many hundreds of percent (up to 100 percent for each core) small numbers here represent heavily loaded systems.
- Check the utilization (`%USED`) of the virtual machines you want to analyze. The virtual machines are reported here with the names specified at the time they were created. As with the idle world's row,

utilization for each virtual machine can exceed 100 percent. A virtual machine that uses two virtual CPUs, for example, can show up to 200 percent CPU utilization.

- You can expand the group data for a virtual machine you want to examine in more detail.To do so, press `e`, then enter the group ID number (shown in the `GID` column) for the virtual machine. The screen capture below contains a CPU-expanded information display for GID 30 from the previous screen capture. When you expand the display, `esxtop` expands rows and provides counter data for every world in the group. This data includes:
  - `vmm`*X*—For each virtual CPU provided to the virtual machine, `esxtop` displays a virtual machine monitor (VMM) world. This world performs the majority of the work required to execute and virtualize the virtual machine's code (operating system, application, and hypervisor).
  - `vcpu-`*X*—ESX Server creates a `vcpu-`*X* world to assist the VMM world for each virtual CPU. The primary work of this world is virtualization of I/O devices.
  - `mks`—This line reports data associated with servicing interrupts for mouse, keyboard, and screen.
  - `vmware-vmx`—The VMX worlds assist in maintenance and communications with other worlds and generally do not represent a material portion of the group utilization.



Evaluate the CPU Data and Correct the System

In general, to evaluate the CPU data that `esxtop` provides, consider the system's load. Is the system overloaded with too many virtual machines? Is the guest operating system using all of its virtual CPUs and does it require more or faster processors? Are all guest operating systems waiting for I/O? For example:

- Check the `PCPU(%)` line to see if utilization for all cores is near 100%. If so, the system is saturated. If multiple virtual machines are competing for the CPUs, try to reduce the number of virtual machines on the system or find other means of decreasing the load on the system. See "CPU Saturation of the Host" on page 4 for more details.
- See if the `PCPU(%)` line shows an unequal load across processor cores with some at saturation and some remaining near idle. If so, applications within the virtual machine are utilizing all of the cores provided to them. Increase the virtual machine's virtual CPU count, if possible, and verify that the guest operating system is making use of the additional cores. If the application supports horizontal scalability, you can run multiple virtual machines to use the additional cores. See "CPU Saturation of a Virtual Machine" on page 5 for more details.
- If all CPUs remain underutilized, either the application in the virtual machine is misconfigured or the virtual machine is waiting for I/O operations to complete. See "Low CPU Utilization" on page 5 for more details.

CPU Saturation of the Host

You can use both the PCPU(%) and %USED counters to identify systems that are using all physical CPUs. It is possible, however, for the virtual machines on the system to utilize nearly all of the processor cycles without actually requesting the additional cycles that are available. This near-saturation case is the sign of a heavily loaded system.

A better sign of overutilization on a host is ready time (%RDY). When any world's ready time starts to climb, that world is spending the reported percentage of its time waiting for some CPU to become available for work. Ready time above 10 percent is worth investigation and may be a sign of an overutilized host. For a more detailed discussion of ready time, see the VMware document "Ready Time Observations" (see "References" on page 12 for a link).

Host saturation is a clear sign that too much work is assigned to a single server. This is usually a result of overly aggressive consolidation ratios. Overcommiting CPU resources in this way degrades performance. Consider the following remedies:

- Verify that VMware Tools is installed in every virtual machine on the system. In addition to many other benefits, VMware Tools provides a network driver (vmxnet) that is necessary for efficient virtual machine networking.
- If you are using VMware Distributed Resource Scheduler (DRS), verify that the all systems in the DRS cluster are carrying load when the server you are interested in is overloaded. If they are not, increase the aggressiveness of the DRS algorithm and check virtual machine reservations against other hosts in the cluster to ensure virtual machines can migrate. Increase the number of servers in the DRS cluster so virtual machines from the server you are evaluating can migrate to servers with available resources.
- Increase the CPU resources available to the virtual machines by increasing the number or improving the performance of CPUs or cores on some of the systems in the DRS cluster.
- Set CPU reservations for the virtual machines that most need the processing power to guarantee that they get the CPU cycles they need.
- Ensure you are using the newest version of ESX Server. Newer versions of ESX Server provide better efficiency and CPU-saving features such as TCP segmentation offload, large memory pages, and jumbo frames.
- Reduce the CPU resource footprint of running virtual machines. For example, you can take the following measures:
  - Decrease disk or network activity or both for applications that cache data. You can do this by increasing the amount of memory provided to the virtual machine. Doing so can lower I/O and reduce the need for ESX Server to virtualize the hardware.
  - Take some of the load off the CPU by replacing software I/O with dedicated hardware (such as iSCSI HBAs or TCP segmentation offload NICs).
  - Reduce the virtual CPU count for guests to the minimum number required to execute the workload. For instance, a single-threaded application in a four-way guest takes advantage of only a single virtual CPU. But the hypervisor must maintain the three idle virtual CPUs, wasting CPU cycles that could be used for other work.

    Few applications fully utilize two or more virtual CPUs, and virtual machines are often committed to a special purpose with a single application on each virtual machine. The guest operating system and the hypervisor must expend CPU cycles managing multiple virtual CPUs. If the applications are not using those virtual CPUs, you can improve system efficiency as a whole by reducing the virtual CPU count for the virtual machines.
  - For virtual machines created from physical machines using VMware Converter, analyze the virtual machine resources as well as the applications running inside the virtual machine. Stop any unnecessary services running inside the virtual machine. Also, reduce the number of virtual CPUs and the amount of memory count to the minimum required to execute the workload.

In general, the easiest way to address CPU bottlenecks when the virtual machines are correctly configured is to increase processing power at the cluster level. If VirtualCenter reports fully utilized CPUs for all hosts in a cluster, you need to increase cluster resources or decrease the number of virtual machines.

CPU Saturation of a Virtual Machine

As with host CPU saturation, you can see virtual machine CPU saturation when the value of `%USED` for a virtual machine is high. In contrast to what you see with host CPU saturation, the idle world might report a large amount of free computational resources and the virtual machine's ready time (`%RDY`) might remain low. You can see this behavior when a single virtual machine utilizes all of the processors allocated to it but additional CPUs remain unused on the host. You can confirm the virtual machine's utilization of all of its virtual CPUs by expanding the virtual machine's world on the `esxtop` CPU screen. If you confirm that the virtual CPUs are saturated, you have the following options:

- Verify that VMware Tools is installed in every virtual machine on the system. In addition to many other benefits, VMware Tools provides a network driver (vmxnet) that is necessary for efficient virtual machine networking.
- If possible, increase the number of virtual CPUs provided to the virtual machine. Because the application in the virtual machine is successfully using all of its virtual CPUs, it may continue to scale as you increase the virtual CPU count. Pay attention to the `vmm`*`X`* world for each virtual CPU after you increase the virtual CPU count to verify that the virtual machine is making use of its newly provided resources. The addition of virtual CPUs imposes additional overhead on the host even when the virtual CPUs are not being used. Thus you should carefully assess the virtual machine's needs to avoid increasing the virtual CPU count unnecessarily.
- If possible, power on multiple virtual machines running the same application. The value of this option depends on how well the application supports horizontally scalable configuration. An application might perform better when running in multiple virtual machines, each with a single virtual CPU, than it does in a single SMP virtual machine.
- Utilize faster processors. Because processor performance is continually increasing, the option of upgrading processors or migrating the virtual machine to a system with newer processors can provide more total throughput to the virtual machine.
- Set CPU reservations for the virtual machines that most need the processing power to guarantee that they get the CPU cycles they need.
- Reduce the CPU resource footprint of running virtual machines. For example, you can take the following measures:
  - Decrease disk or network activity or both for applications that cache data. You can do this by increasing the amount of memory provided to the virtual machine. Doing so can lower I/O and reduce the need for ESX Server to virtualize the hardware.
  - Take some of the load off the CPU by replacing software I/O with dedicated hardware (such as iSCSI HBAs or TCP segmentation offload NICs).

Low CPU Utilization

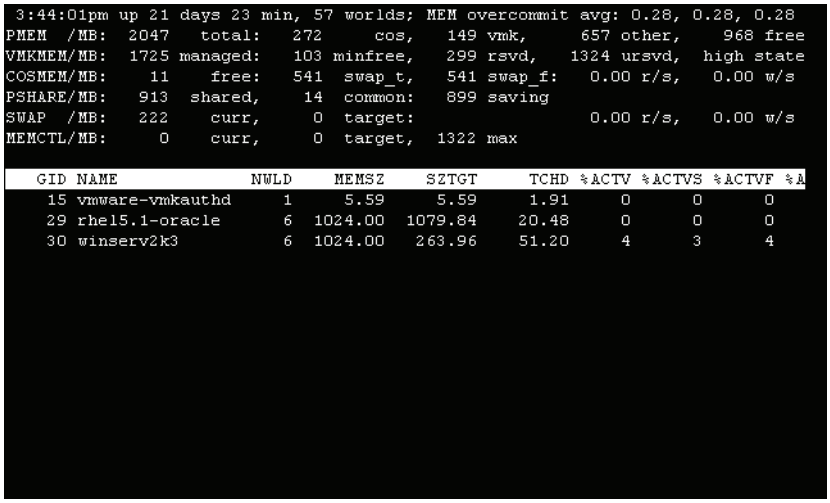
If you have confirmed performance problems, low CPU utilization is usually a sign of inefficiently designed datacenter architecture. The design might be flawed in the configuration of an individual virtual machine or in the connectivity between various components. The following sections discuss methods for investigating system-level components such as memory and then systemwide components such as networking and storage.

Memory Analysis

Host memory utilization includes all memory used by virtual machines on the host and all memory used by ESX Server itself. The monitoring capabilities in ESX Server do not help you detect improper usage or configuration of memory within a virtual machine. You must use traditional monitoring tools in the guest operating system to identify memory-hungry applications or shortages that lead to swapping inside the virtual machine.

Check Memory Utilization

Start `esxtop` by entering the command in the ESX Server host's service console. Press `m` to display the memory counters.



The output includes the following information:

- The header shows host data that affects all virtual machines running on the host. The physical memory row (`PMEM`) shows the total RAM installed on the system, the amount used by the service console (`cos`), the memory used by the VMkernel (`vmk`), and other statistics.
- The next few rows show host-level memory statistics for various ESX Server subsystems:
  - `VMKMEM`—shows memory statistics for the ESX Server VMkernel.
  - `COSMEM`—displays the memory statistics reported by the ESX Server service console.
  - `PSHARE`—displays ESX Server page-sharing statistics.
  - `SWAP`—displays ESX Server swap usage statistics.
  - `MEMCTL`—displays statistics for the memory balloon driver.
- Data for each virtual machine on the host appears as a row in the table at the bottom of the display. The following counters are of particular interest when evaluating virtual machine memory usage:
  - The total memory allocated to the virtual machine appears in the `MEMSZ` column.
  - The memory that is actively in use by the guest operating system and its applications is reported in the touched and active counters (`TCHD` for touched memory and `%ACTV`, `%ACTVS`, and `%ACTVF` for active memory). `%ACTVS` and `%ACTVF` provide slow and fast averages of the `%ACTV` counter. Either the `%ACTV` or `TCHD` counters can serve as good predictors of memory usage. When the actively used memory of one or more virtual machines exceeds the amount of memory on the host, the server starts to swap and performance degrades significantly.
  - Knowing the activity caused by to the balloon driver can be useful. When the balloon driver is active in the guest operating system, the virtual machine's `MCTL?` counter is set to `Y`. The amount of memory the balloon driver is using in a specific guest operating system is reported under `MCTLSZ`. ESX Server uses the balloon driver to recover memory from less memory-intensive virtual machines so it can be used by those with larger active sets of memory. ESX Server takes this memory management step before it resorts to swapping memory to disk.
  - The rate at which ESX Server is swapping memory to and from disk is displayed by the swap write (`SWW/s`) and swap read (`SWR/s`) counters. These counters should remain near zero for high-performing steady-state operations. A sustained rate of a significant number of MB/s is a certain sign that the host does not have enough memory.

- For NUMA systems, the NUMA counters displaying migration count (NMIG) and remote and local memory (NRMEM and NLMEM, respectively) can offer information indicating whether the virtual machine is utilizing NUMA memory inefficiently. Memory access across NUMA nodes is inefficient and migrations of memory across nodes slow down execution.
- The memory overhead required to maintain each virtual machine is displayed by the OVHD counter. Knowing this additional usage can help you plan virtual machine and host configurations.

Evaluate the Memory Data

Analysis of memory utilization on an ESX Server host requires not just investigation of server-side statistics but also a solid understanding of the applications that are running in virtual machines on the host. When memory is short on the host, ballooning and swapping might be visible in esxtop. Swapping has a significant impact on performance. When memory is short within a virtual machine, the guest operating system swaps memory to disk.

To evaluate the data provided by esxtop, consider the following factors:

- Is memory short on the host? Swapping (SWW/s and SWR/s) is a certain sign of this problem. Heavy use of the balloon driver might also suggest a shortage of memory on the host, but ballooning has only a very slight impact on guest performance.
- Can you address memory deficiencies by resizing virtual machines? Check the memory usage of critical applications running in the virtual machines to help you decide whether you can decrease the amount of RAM provided to those virtual machines. Some operating systems expand to utilize all available memory even though this approach provides little or no benefit to the application. Reducing the memory space and correcting oversized caches frees memory for other virtual machines.
- Is the total active memory for all virtual machines (TCHD or %ACTV) consistently exceeding the total available memory? If so, you must either add more memory to the host or migrate virtual machines to another DRS cluster.
- Are the guest operating systems swapping memory to disk? If a virtual machine has too little memory, the guest operating system swaps inside the virtual machine. This swapping appears the same to ESX Server as any other disk activity, but you should investigate it and solve the problem using traditional operating system analysis tools.
- Can you see NUMA migrations (NMIG) on the system? The NMIG column reports total migrations since the virtual machine was powered on. If this number continues to climb, the virtual machine is being migrated from node to node. The repeated migrations certainly degrade performance.
- Does the amount of memory located on a remote NUMA node (NRMEM) remain at a non-zero number? This may be a sign that the memory assigned to a virtual machine exceeds the memory of a single NUMA node. If the virtual machine is using more memory than is available on a single node, some of its memory is certain to be located on a remote node. Remote memory access is quite slow compared to local memory access.

Correct Memory Configuration on the System

The steps you must take to resolve memory shortages are simple: use less memory or add more to the system. The following recommendations are variations on this theme:

- Verify that VMware Tools is installed in every virtual machine on the system and that the memory balloon driver has not been disabled. (The balloon driver is always on by default and can be disabled manually using text-based advanced configuration tools. It should be disabled only in extremely rare cases.) When it can use the balloon driver to reclaim memory within the guest operating systems, ESX Server is able to take memory from virtual machines that are not using it and make it available to those that do need it.
- Provide more memory to the DRS cluster. As total resources go up, VirtualCenter balances virtual machines across the cluster in a way that provides virtual machines the memory they need.

- Set memory reservations to provide the minimal amount of memory required for the operating system and critical applications. This approach allows for sustained, fast access for critical code and provides hints to VirtualCenter for optimal positioning of virtual machines across the DRS cluster.
- Make sure the amount of memory used by the VMkernel to maintain the virtual machines is acceptable. This value, reported for each virtual machine by the overhead counter (`OVHD`), is dependent on the memory size of the virtual machine, the number of virtual CPUs provided to the virtual machine, and whether the virtual machine is running a 64-bit operating system. Fewer virtual machines on the host, fewer aggregate virtual CPUs, and lower-precision operating systems (32-bit as opposed to 64-bit) lower this number. You can free resources for all virtual machines in the cluster by reducing any of these factors in the cluster.
- Size virtual machines on NUMA systems to guarantee that each virtual machine's memory fits on a single node. If there is a mismatch, you must either decrease the memory allocated to a virtual machine or increase the node's memory size.
- Size guests appropriately according to their needs. For example:
  - Depending on the access pattern for the data, databases might not benefit from the last doubling of cache size. Experiment with smaller cache sizes and see if performance drops. If the smaller cache size does not degrade performance, decrease the virtual machine's available memory so other virtual machines can use that memory.
  - Check the guest operating system's statistics for swapping inside the virtual machine. Provide memory as needed and pay attention to `esxtop` statistics to see if providing the additional memory generates a new bottleneck on the host.

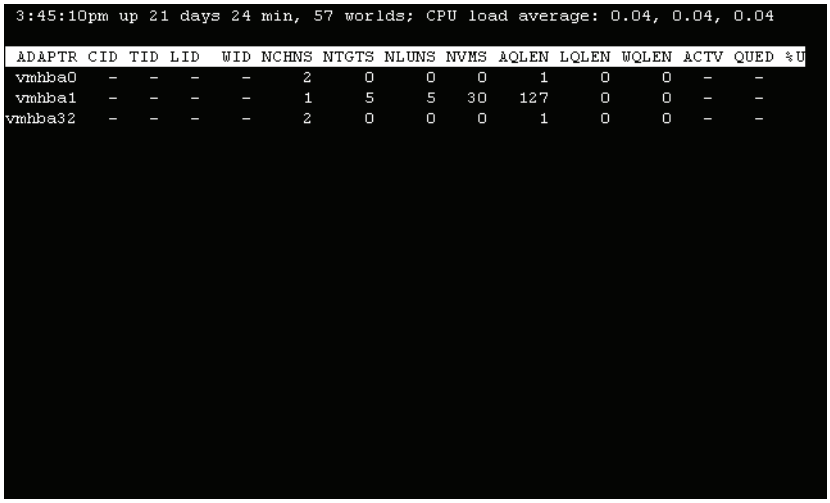
Storage Analysis

Storage often limits the performance of enterprise workloads. Traditional means of analysis are sound for evaluating storage performance in virtual deployments. This section introduces tools for identifying heavily used resources and virtual machines that place high demands on their storage systems. You can then apply traditional methods to correct the problems.

This section does not cover iSCSI storage using software initiators. When virtual machines access iSCSI storage through the hypervisor's iSCSI initiator or a software initiator inside the guest operating system, the storage traffic appears on the VMkernel network or the virtual machine's network stack. See "Network Analysis" on page 11 for more information.

Check Storage Utilization

Start `esxtop` by entering the command in the ESX Server host's service console. Press `d` to display the disk adapter information.



On ESX Server 3.5, you can press v to display the storage system information per virtual machine or press u to display the information per storage device. The same counters are displayed in each listing. The output includes the following information:

Each of the three storage views displays information in a particular order. You can expand groups in the disk displays to view more detailed information:

Adapter View

In the adapter view (d), each physical HBA appears on a row of its own, identified by the appropriate adapter name. You can check this short name against the more descriptive data provided through the Virtual Infrastructure Client to identify the hardware type.

Press e to expand the HBAs listed in the adapter view. The expanded display shows worlds that are using the HBAs. Locate a virtual machine's world ID in the WID column to find the data for activity related to that virtual machine.

Virtual Machine Disk View

In the virtual machine disk view (v), each row represents a group of worlds on the ESX Server host. Each virtual machine appears on a row of its own, and esxtop displays rows for the service console, system, and other worlds that are less important when you are analyzing storage. The groups IDs (GID) match those shown on the CPU screen, and you can expand the listing by pressing e.

Press e to expand the worlds data for each virtual machine in the virtual machine disk view.

Disk Device View

In the disk device view (u), each device appears on its own row.

Press e in the disk device view to show usage by each world on the host.

Counters to Check

The output includes the following information:

- The ACTV counter provides a snapshot of current activity. The QUED counter lists the number of queued commands that the host will process after ACTV commands have finished. A sustained number of ACTV commands is healthy and indicates continuing disk activities. A sustained number of queued commands indicates a heavily loaded system.
- The LOAD counter provides an estimate of the utilization of a single HBA. It represents the ratio of the number of commands that are active or queued to the total number of commands that can be active or queued at one time. Thus a LOAD value of 1.0 means that both the active buffer and queue are full. At this point, the server begins failing to execute commands.
- The %USD counter provides the percentage of the queue depth used by VMkernel active commands. Very high values indicate the likelihood that commands are queued, and you may need to adjust the queue depths for system's HBAs.
- You can view the total latency seen from a virtual machine to the array in the virtual machine latency counter (GAVG/cmd), which shows the sum of the latencies caused by the hardware (DAVG/cmd) and those caused by the VMkernel (KAVG/cmd).
- Aborted commands as displayed by the aborted commands per second counter (ABRTS/s) represent commands issued by the guest operating system after it determines that a storage request cannot be fulfilled. Aborts are a sign that the storage system cannot meet the demands of the guest operating system.

Evaluate the Storage Data

It is important to have a solid understanding of the storage architecture and equipment in your environment before attempting to analyze performance data. Consider the following questions:

- Is the host or any of the guest operating systems swapping memory to disk? You can check the guest operating system's swap activity using traditional operating system tools. You can see data on host swap activity in two `esxtop` counters—`SWR/s` and `SWW/s`— as described in "Check Memory Utilization" on page 6.
- Are commands being aborted? This is a certain sign that the storage subsystem is unable to handle requests as the guest operating systems expect. Corrective action might include hardware upgrades, storage redesign, or virtual machine reconfiguration.
- Is the queue large? Although less dangerous than aborted commands, queued commands are a sign that you need to upgrade hardware or redesign the storage system.
- Is the storage array responding at expected rates? Storage vendors provide latency statistics for their hardware that you can check against the latency statistics in `esxtop`. When the latency numbers are high, the hardware could be overworked by too many servers. For example, 2–5 ms latencies are usually a sign of a healthy storage system reading data on the array cache, 5–12 ms latencies reflect a healthy storage architecture in which data is being read randomly across the disk, and 15 ms latencies or greater possibly represent an overutilized or misbehaving array.

Correct Storage Configuration on the System

If you confirm that you have storage system problems, consider the following possible methods of correcting the problems:

- Reduce the virtual machines' and host's need for storage.
  - Some applications, such as databases, can use system memory to cache data and avoid disk access. Check the virtual machines to see if they can benefit from increased caches and provide more memory to the virtual machines if appropriate and if resources permit. The additional memory may reduce the burden on the storage system.
  - Eliminate as much swapping as possible to reduce the burden on the storage system. First, verify that the virtual machines have the memory they need by checking swap statistics in the guest operating system. Provide memory if resources permit. Next, as described in "Correct Memory Configuration on the System" on page 7, eliminate host swapping.
- Configure the HBAs and RAID controllers for optimal use.
  - Increase the number of outstanding disk requests for the virtual machine by adjusting the `Disk.SchedNumReqOutstanding` parameter. For detailed instructions, see the "Equalizing Disk Access Between Virtual Machines" section in the VMware Infrastructure 3 *Fibre Channel SAN Configuration Guide*. See "References" on page 12 for a link to this document.
  - Increase the queue depths for the HBAs. Check the section "Setting Maximum Queue Depth for HBAs" in the detailed instructions, see the "Equalizing Disk Access Between Virtual Machines" section in the VMware Infrastructure 3 *Fibre Channel SAN Configuration Guide* for detailed instructions. See "References" on page 12 for a link to this document.
  - Make sure the appropriate caching is enabled for the disk controllers. Use the tools provided by the controller vendor to verify this setting.
- If latencies are high, inspect array performance using the vendor's array tools. When too many servers simultaneously access common elements on an array, the disks might have trouble keeping up. Consider array-side improvements to increase throughput.

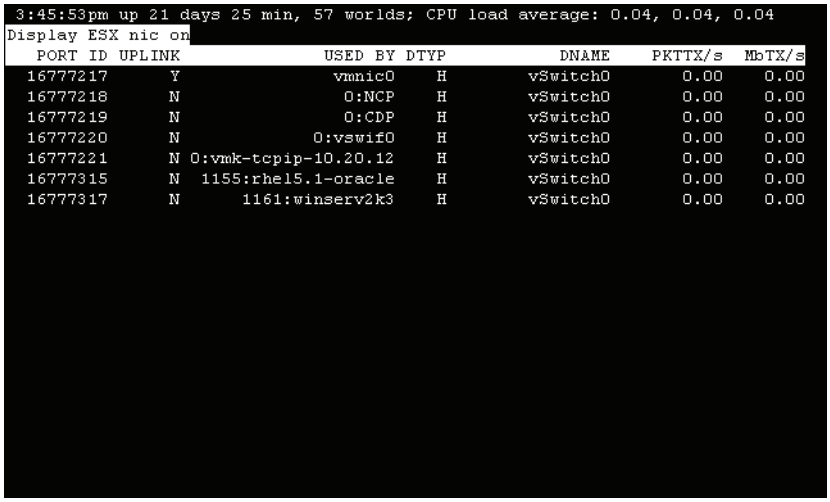
- Balance load across the available physical resources.
  - Spread heavily used storage across LUNs being accessed by different adapters. The presence of separate queues for each adapter can yield some efficiency improvements.
  - Use multipathing or multiple links if the combined disk I/O is higher than the capacity of a single HBA.
  - Using VMotion, migrate I/O-intensive virtual machines to different ESX Server hosts, if possible.
- Upgrade hardware, if possible. Storage system performance often bottlenecks storage-intensive applications, but for the very highest storage workloads (many tens of thousands of I/Os per second), CPU upgrades on the ESX Server host increase the host's ability to handle I/O.

Network Analysis

Network analysis is usually a straightforward process for which typical native techniques are valid. Checking for load relative to link throughput and looking for dropped packets can identify all but the most subtle of problems.

Check Network Utilization

Start `esxtop` by entering the command in the ESX Server host's service console. Press `n` to display the network information.



The following properties of this display are worth particular attention:

- Each row presents data for one network-related item on the host, for example: a physical NIC (`vmnic`*X*), a virtual switch interface (`vswif`*X*), a virtual machine (contains the virtual machine name), the VMkernel network stack (`vmk-tcpip-`*A.B.C.D*).
- The network items are organized by the virtual switch to which they are attached. The virtual switch name is listed in the `DNAME` column.
- Network traffic on the hypervisor's iSCSI initiator appears on the VMkernel network row, which contains the name `vmk-tcpip-`*A.B.C.D*, where *A.B.C.D* is the VMkernel IP address.
- Network traffic on iSCSI initiators configured in the guest operating system appear on the line for the virtual NIC displayed using the virtual machine's name as shown on the `esxtop` network screen.
- You can calculate total throughput for each item by summing the total transmitted data (`MbTX/s`) and received data (`MbRX/s`) for that item. As the physical hardware becomes saturated, the system begins to drop transmitted and received packets (`%DRPTX` and `%DRPRX`, respectively). Depending on the protocol, the system may retransmit the dropped packets at a later time.

Evaluate the Network Data

To evaluate the network data, consider the following questions:

- Do the physical NIC's reported speed and duplex setting match the expectation of the hardware? Hardware connectivity issues might cause a NIC to autonegotiate a lower speed or half-duplex mode.
- Do appropriate network items show a significant load? For instance, is a network-intensive load in a virtual machine actually generating the expected network activity on its virtual NIC? Are storage-intensive loads generating traffic on the virtual NIC or the VMkernel NIC (`vmkNIC`) when software initiators are used on the hypervisor or in the guest operating system?
- Is the network traffic flowing on appropriate NICs? A typical ESX Server host might have network traffic generated by virtual machines, network traffic from the iSCSI protocol, VMotion-related network traffic, and network activity associated with the service console. You should have separate NICs to handle these different kinds of network packets.
- During periods of saturation, does the total throughput (`MbTX/s` summed with `MbRX/s`) match expectations? Either the guest operating system or the other end of the communication link might be throttling the performance.
- Are packets being dropped? When overworked, the hardware refuses packets. those packets are reported as dropped transmitted (`%DRPTX`) and dropped received (`%DRPRX`) packets.

Correct Network Configuration on the System

If you confirm that you have networking problems, consider the following possible methods of correcting the problems:

- Make sure that the hardware is configured to run at its maximum capacity. Verify that 1 Gb NICs are not autonegotiating down to 100 Mb/s because they connected to an older switch. Similarly, ensure that NICs are running in full-duplex mode.
- When network throughput seems lower than expected, apply traditional network diagnosis techniques to investigate every link in the connection. Low throughput at the ESX Server host is not necessarily caused by server configuration.
- Verify that VMware Tools is installed in all guest operating systems and that TSO, jumbo frames, and 10 Gb Ethernet are enabled, where possible.
- Bond multiple physical NICs to virtual switches that show high utilization.
- Provide separate virtual switches their own physical NICs and separate network-intensive virtual machines on their own virtual switches.
- If virtual machines running on the same ESX Server host communicate with each other, connect them to a dedicated virtual switch so all network transfers occur in memory and not as packets are shipped over the physical network.

References

- "Ready Time Observations"
  http://www.vmware.com/pdf/esx3_ready_time.pdf
- VMware Infrastructure 3 *Fibre Channel SAN Configuration Guide*
  http://www.vmware.com/pdf/vi3_35/esx_3/r35/vi3_35_25_san_cfg.pdf
- VMware Infrastructure 3 *Resource Management Guide*
  http://www.vmware.com/pdf/vi3_esx_resource_mgmt.pdf

Appendix: Counters

The following tables include descriptive information on counters mentioned in this document:

Table 1. CPU counters

Counter	Description
GID	Group ID.
%USED	The percentage of CPU that is used by a world or group.
NWLD	The number of worlds in a group. When this number is greater than one, the row can be expanded to display information on each world.
%RDY	The percentage of time that a world or group is waiting for a processor to be available to execute its workload.

Table 2. Memory counters

Counter	Description
MEMSZ	The amount of memory (in MB) allocated to a virtual machine at the time of its creation.
TCHD	The amount of memory (in MB) that has been touched (recently used) by a virtual machine. In this case "recently" means within the past minute or two.
%ACTV	Instantaneous view of the percentage of memory pages that have been used by a virtual machine in the previous seconds. Unlike TCHD, which counts pages by following working sets, %ACTV is updated more frequently and is based on a sample of the entire memory pool.
%ACTVS	Slow moving average of the %ACTV counter.
%ACTVF	Fast moving average of the %ACTV counter.
NHN	The NUMA home node. This is the node on which a virtual machine is booted. Migrations that have occurred since the virtual machine started running would result in the virtual machine running on another node or nodes.
NMIG	The number of NUMA node migrations since a virtual machine was booted. The ESX Server scheduler should avoid NUMA migrations, so if this number continues to climb during normal operations, some tuning of the virtual machines may be required.
NRMEM	The amount of memory that exists on a remote NUMA node.
NLMEM	The amount of memory that exists on the local NUMA node.
N%L	The percentage of the virtual machine's memory that exists on the local NUMA node. N%L = NLMEM / (NRMEM+NLMEM)
MCTL?	Set to Y when the balloon driver is active in the guest and N when not.
MCTLSZ	This counter reports the amount of memory that the balloon driver is currently reclaiming for use by other virtual machines.
OVHD	The amount of memory used by the VMkernel to maintain and execute a virtual machine.

Table 3. Storage counters

Counter	Description
ACTV	The number of I/O operations that are currently active. This number represents operations the host is processing and can serve as a snapshot view of storage activity. When this number hovers near zero, the storage system is not being used. If this number is consistently something other than zero, the system is constantly interacting with the storage.
QUED	The number of I/O operations that require processing but have not yet been addressed. Commands are queued and awaiting management by the kernel when the driver's active buffer is full (see ACTV). Occasionally a queue forms and, as a result, this counter displays a small, non-zero QUED number, but any significant (double-digit) average of queued commands means the storage hardware is unable to keep up with the host's needs.
DAVG/cmd	The average amount of time it takes a device (HBA, array, and everything in between) to service a single request.

Table 3. Storage counters

Counter	Description
KAVG/cmd	The average amount of time it takes the VMkernel to service a disk operation. Because this number represents time spent by the CPU to manage I/O and processors are orders of magnitude faster than disks, it should be much, much less than DAVG.
GAVG/cmd	The total latency seen from the virtual machine when performing an I/O operation. GAVG = DAVG+KAVG.
ABRTS/s	The rate at which disk operations are being aborted. Abort commands are issued by the guest operating system when the storage system has not responded within an acceptable amount of time (as defined by the guest operating system or application.)

Table 4. Network counters

Counter	Description
MbTX/s	The number of megabits per second that are transmitted from the network item.
MbRX/s	The number of megabits per second that are received at the network item.
%DRPTX	The percentage of packets for which transmission was attempted but unsuccessful. The packets were dropped.
%DRPRX	The percentage of packets that should have been received but were not. The packets were dropped.

VMware, Inc. 3401 Hillview Ave., Palo Alto, CA 94304 www.vmware.com

Revision 20080311 Item: TN-056-PRD-01-01