12.3.3 Performance Monitoring Facts

A *bottleneck* occurs when a component is unable to keep up with demand and subsequently slows down other processes or functions. When a system seems to respond slowly, it's important to be able to accurately identify the component(s) that are causing the problem so you can take the proper actions to improve performance.

Common components to examine in order to improve performance are:

- Processor (CPU)
- Hard disk
- Memory
- Network

One way to identify components that are causing a bottleneck is to monitor system performance statistics. These statistics give you a measure of the activity of a certain aspect of the system. By recognizing abnormal statistics, you can identify the component that is overloaded or not responding appropriately. Windows identifies system performance statistics using the following terms:

- A *counter* is a specific statistic you can monitor (such as the amount of free memory or the number of bytes sent on a network card).
- An *object* is a statistic group, often corresponding to a specific type of hardware device or software process (such as the processor or memory).

Use Task Manager, System Monitor, Resource Monitor, and Performance Monitor to track statistics. You should be familiar with the meaning and use of the following counters:

Counter	Description

% Processor Time (Processor Utilization)	*Processor utilization* is the amount (percentage) of time the processor spends doing non-idle tasks.
	▪ Processor utilization should be relatively low, up to 40% on average.
	▪ Processor utilization will spike (85 - 90% or higher) when a major task is launched or a significant task is performed.
	▪ Utilization is reported for each processor in a multi-processor or multi-core system. A CPU that supports Hyper-Threading will show two utilization graphs for each processor.
	▪ If the processor utilization is consistently high (over 90%), then the CPU is likely the bottleneck.
	  ▪ Check the running processes to see the CPU use of each process. If possible, delay or pause non-critical processes or run them during off hours.
	  ▪ A process that has hung could show 100% CPU use. If the process does not complete after a period of time and does not respond, end the process to return CPU use to normal.
	  ▪ A computer with a virus might show an unknown process consuming most of the processor time. Use the internet to identify the function of unknown processes.
	  ▪ Configure the processor *affinity* to specify that a specific process use a certain processor in a multi-processor system.
	  ▪ Upgrade to a faster CPU or add more cores to the system.

% Disk Time (Highest Active Time)	The % Disk Time statistic identifies the percentage of time that the disk subsystem is busy reading from and writing to disk. If this value is consistently over 90%, check the following other statistics to identify the source of the high disk activity: ▪ Average Disk Queue Length ▪ Memory statistics
Average Disk Queue Length	The disk *queue* holds read and write requests that are waiting to be processed by the disk controller. The average disk queue length tells you the number of read and write requests that are typically waiting to be processed. ▪ A high number indicates that the system has requested data from the hard disk, or has tried to save data to the hard disk, but that request could not be fulfilled immediately (i.e. it has to wait). ▪ This number should be below 2 times the number of disk spindles. Most physical hard disks have a single spindle (although some newer drives have 2 or 3). RAID arrays will have at least one spindle per physical disk. If this statistic shows consistently waiting read/write requests, you might need to upgrade your disks. ▪ Choose a faster disk (higher RPM and faster access time). ▪ Use a RAID-0 configuration to improve disk access.
Available, Used, and Free Physical Memory	You can use Task Manager to quickly identify the use of physical memory in your system. ▪ The total installed memory value reflects the amount of memory available to the operating system. On a 32-bit system, this value will be less than 4 GB, even if you have 4 GB of memory installed. This value could also be slightly less than the amount of installed physical RAM if the video adapter shares the system memory. The amount of memory used for this purpose is displayed under hardware reserved. ▪ The cached value identifies memory that is being used for a disk cache to improve read/write operations from the hard disk. ▪ The available value identifies how much memory is unassigned.

	If the amount of memory in use is close to the amount of RAM installed, you might need to add RAM or quit some running programs to free up memory.
Memory Committed Bytes (Commit Charge)	When a process runs, the operating system assigns memory to the process. The amount of committed memory identifies how much memory has been assigned to running processes. Be aware of the following conditions indicated by this statistic: ▪ If the value exceeds the amount of physical RAM, then the page file is being used instead of physical RAM. At some point, this will start to cause a bottleneck. ▪ To temporarily make more memory available, quit running programs or increase the page file size. However, the only permanent solution is to add more physical memory.
Page File Usage	The page file usage identifies the amount or percentage of the page file that is being used. ▪ A common recommendation is for the page file to be 1.5 to 2 times larger than the physical memory. In most cases, you will let the system manage the page file size. ▪ It is normal for the page file to show some use, even when the system has sufficient physical memory. ▪ When the page file use percentage is near 100%, you can increase the page file size as a temporary measure. Adding more memory is the best permanent solution.
Memory Pages per Second	The operating system allocates memory to processes in 4,096 KB blocks called *pages*. ▪ Instead of assigning physical memory addresses, the operating system assigns *virtual memory* addresses to shield the process from the details of the physical memory storage system. ▪ The *paging supervisor* is a process that maintains a table that correlates virtual memory addresses with the actual physical memory locations. When physical memory is low, data in RAM that is currently not being used by the CPU can be moved to the hard disk in order to free up memory for other processes. ▪ The area on the hard disk used for storing the contents of RAM is called the *page file*. ▪ When the CPU needs to access data in RAM, a page fault (also called a hard fault) occurs when that data does not exist in RAM but is instead in the page file. ▪ *Paging* is the process of moving data from RAM to disk and back. Before the CPU can work with data required by a process, that data must be placed into RAM. The memory pages per second statistic identifies the number of hard faults that occur each second. A high number for this statistic accompanied by high disk activity (% Disk Time or the disk activity light constantly flashing) could indicate a condition known as *thrashing*. ▪ With thrashing, the demand for memory and the low amount of physical RAM means that the system must be constantly moving data from RAM, to disk, and back. ▪ The negative effects associated with paging increases as the amount of memory increases past the amount of physical RAM. While some paging is normal, as the demands on memory increase, the amount of paging will at some point reach a point where thrashing occurs and the effect on performance is noticeable--even to the point of making the system unusable. ▪ As a temporary solution, you can quit some running programs in order to decrease the demand for RAM. The only long-term solution is to add more physical RAM. ▪ Increasing the page file size will have no effect unless you are also experiencing out of memory errors. The problem is not that there isn't sufficient combined memory, but that the amount of physical memory is insufficient.
Network Utilization	Network utilization identifies the amount of traffic sent and received by a network connection. ▪ Utilization is listed as a percentage of the total available theoretical bandwidth (such as 100 Mbps for a Fast Ethernet connection). ▪ Poor performance that has low CPU, disk, and memory statistics but high network utilization could indicate a bottleneck at the network adapter.