Ask the Performance Team Blog

Thoughts from the EPS Windows Server Performance Team

Performance tuning Windows Server 2008 R2 Pt 2

Tim Newton - MSFT 5 Nov 2010 4:00 AM 4

In our previous post, we discussed basic performance tuning related to CPU, RAM and Pagefile. In this post, I would like to go into physical disks, the disk subsystem and power management for disks.

When it comes to physical disks, you have to make a decision between speed, storage and interface type. Consumer and many Enterprise drives are now SATA, which makes upgrades relatively cheap and easy. More exotic interfaces, such as Fibre Channel or SAS tend to be very fast, but also much more expensive. For any drive, a good rule of thumb is that rotational speed is directly related to transfer rate. Typical SATA drives run at 7200 RPM, but there are a few that run at 10,000 RPM. SAS and Fibre Channel drives are available at speeds of up to 15,000 RPM with a corresponding increase in speed. The downside is that larger capacity drives are typically not available at the higher speeds. Having more cache on the drive is also important, with many drives now available with 16 or 32 MB built in.

Another factor to take into account is the advent of relatively cheap SSD (solid-state) drives. These drives have phenomenal access times and random reads and writes, but generally suffer in sequential reads\write and capacity. The first couple of generations of SSD had serious performance issues under certain write conditions, with the exception of a handful of specific models with optimized controllers. The newer controllers however are said to largely mitigate this problem, which is making SSDs a more attractive proposition for many. The cost is also much higher than a comparably sized rotational hard disk, but coming down, albeit slowly.

The interface types available to you are often very limited depending on which model of hardware you buy. Workstations and lower-end servers are often only available in SATA, which is fine, but prevents you from using the fastest drives available. Fibre-Channel or SAS controllers are faster, but of course more expensive and only available on certain models unless you add an aftermarket controller card. Depending on the controller, you also have to make the decision whether or not you need to use a fault-tolerant or other RAID configuration. RAID 1 (Mirroring) allows continual protection of data, but you only get half the capacity out of the installed drives. RAID 0 (Striping) yields better performance, but if any single drive goes bad, you could end up losing your data or having to restore from backup. I personally use a pair of 10,000 RPM drives in a RAID 0 setup at home, and the performance is phenomenal, I just have to be aware that if anything goes wrong I will have to go through a slightly more complex restore process than if I had a regular single drive setup.

Peak and sustained bandwidth requirements are also important. The rate at which data can be accessed is determined by the controller type, the number of physical disks installed, the type of disks installed and use of RAID. Having more disks installed, running in RAID 0 or something else with striping, and on a SAS or Fibre Channel controller is going to be about as fast as you can get, but is not cheap. If you decide to use RAID, you also have to be educated on what stripe size you need to configure to maximize your throughput.

All of this has to be taken into account when building out your storage solution. Faster drives are often desirable for the OS install drive, since having a faster drive makes Windows itself run faster. However, for data drives, extremely high throughput might not be necessary and more storage space is the deciding factor. Many people are now installing a small SSD as their boot drive and traditional drives for storage so as to maximize speed versus cost versus capacity. To make an accurate determination of your needs, you have to take into account the read:write ratio, sequential versus random access needs, block sizes and concurrency needs. What is right for your needs is of course up to you to decide, you have to take into account cost, capacity, speed and the need for any sort of RAID or backup solution on the backend.

Regardless of how you choose to set up your storage system, the Performance Monitor counters you use to monitor it will be the same:

%Disk Read Time, %Disk Write Time, %Disk Time, %Idle Time: All of these can be important, but keep in mind that they are only reliable when dealing with single disks. Having disks in RAID or a SAN setup can make these numbers inaccurate.

Average Disk Queue Length: A good counter to monitor if requests are backed up on your disk. Any sustained number higher than 2 is considered problematic. However, just like the counters above, this one is not considered reliable except when dealing with single physical disk per volume configurations.

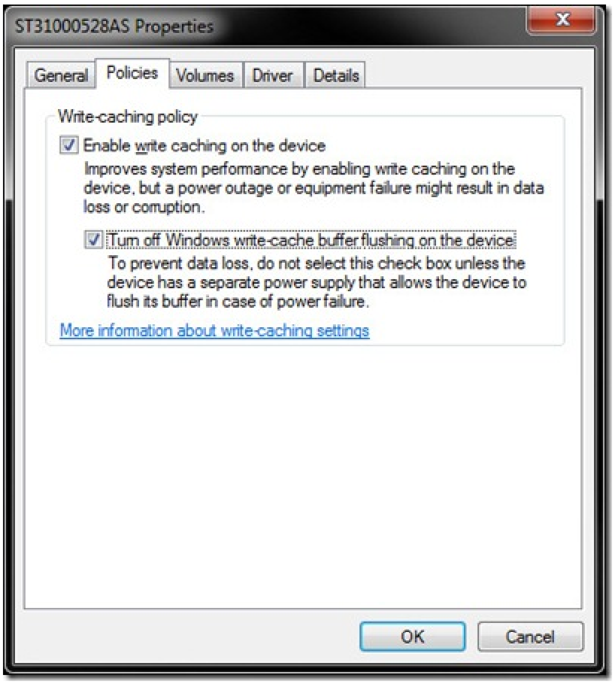
Average Disk Second/Read, Average Disk Second/Write, Average Disk Second/Transfer: These are probably my favorite disk counters. They tell you how long, in seconds, a given read or write request is taking. (Transfer is considered a Read\Write round trip, so is pretty much what you get if you add the other two.) These counters tell you real numbers that deal directly with disk performance. For instance, if you see that your Average Disk Seconds/Write is hitting .200 sustained, that tells you that your write requests are taking a full 200 ms to complete. Since modern disks are typically rated at well under 10 ms random access time, any number much higher than that is problematic. Short spikes are okay, but long rises on any of these numbers tell you that the disk or disk subsystem simply is not keeping up with load. The good thing is, these being real numbers, the number of disks or how they are configured is really not important. High numbers equate to bad performance regardless.

Current Disk Queue Length: Don’t bother with this counter. It measures instantaneous disk queue numbers, and as such is subject to extreme variance. You can get much more useful numbers by using the Average Disk Queue Length counters mentioned above.

Disk Bytes/Second, Disk Read Bytes/Second, Disk Write Bytes/Second: These counters tell you how much data is being read from or written to your disks being monitored. The number by itself doesn’t mean anything, but in combination with other counters can be telling. For instance, let’s say it is reporting that you are writing 100 MB/second to the disk. Is this too much? Not if overall performance is still good. If however, your disk access times are getting high, you can see if you can correlate this to the load reported by this counter and calculate exactly how much your disks can handle before becoming slow.

Disk Reads/Second, Disk Writes/Second, Disk Transfers/Second: These counters tell you how many operations are occurring per second, but not how much as being transferred. It is possible for instance to have a large amount of very small requests causing performance issues, but overall throughput may not be high enough to seem problematic. This counter can be used in the same way as the Disk Bytes/Second counters, but measure operations instead of throughput.

In relations to Power Management, there are two performance options you should be aware of. These are **'Enable write caching'** and **'Enable Advanced Performance',** or as it is now called, **'Turn off Windows write-cache buffer flushing on the device'**. These options are found by going into Device Manager and opening the properties of the drive:



ST31000528AS Properties

General | Policies | Volumes | Driver | Details

Write-caching policy

☑ Enable write caching on the device
Improves system performance by enabling write caching on the device, but a power outage or equipment failure might result in data loss or corruption.

☑ Turn off Windows write-cache buffer flushing on the device
To prevent data loss, do not select this check box unless the device has a separate power supply that allows the device to flush its buffer in case of power failure.

More information about write-caching settings

OK | Cancel

Write caching allows the OS to assume that a write has completed, even though the disk subsystem may not yet have actually completed the action of physically writing to disk. Some disk controllers include cache that allows requests to be stored there while the disk is busy doing other things. This allows the OS to continue on with its work without having to worry about the state of the data on its way to the disk. With write caching off, the OS will have to wait if it thinks the write is still in progress. Keep in mind that using write caching is only 100% safe if the computer or disk subsystem has some sort of power backup. Writes tend to happen very quickly, but if power was interrupted while data was still in the buffer, than that data would be lost.

The 'Advanced Performance' option requires write caching be enabled. This option strips all write-through flags from the disk requests and removes all flush-cache commands. If you have backup power for your disk subsystem, you should not need these flags or commands since any dirty data that resides in cache is protected and assumed to be 'in-order'. Even if power is lost to the the actual drive, the cache manager can retry the write operation from the cache once power is restored.

So, that is all for now. Until next time..

Tim Newton

Share this post :

Tweet 16 Like 4 Share 1 Save this on Delicious

Comments

6 Nov 2010 8:27 AM

Dmitry Mashkov

Tim, I am sorry to say that, but this kind of information floats between dozens of MS articles and does not change a bit since early "SQL Server 7.0 Performance Tuning" article which provided the consolidated information for tuning the base OS and SQL from all point of view. Unfortunately, no new articles referring to Windows 2003/2008/R2/RX add anything new (updating drive RPM numbers does not count - this does not change the approach itself), that has really changed in this releases. Republishing the same info is just a contamination of TechNet.

I'd love to see really new information, more detailed than the same list of basic counters with replaced OS name...

12 Nov 2010 8:41 AM

pjha

Excellent article!!!! Cleared a lot of my doubts about disks. Keep them coming.

15 Jun 2012 12:48 PM

Pankaj

I have a Question here..!! If I Checked My Virtual Disk Properties > Policies, I found that Write Caching is Enabled by Default (better Peformance)..!! Is it safe?

We not have a Good Power backup.

Please suggest. I should change it to Write Caching Disabled for better performance of System..

14 Feb 2014 7:07 PM

Bucket

"This car can go really fast, but the seatbelts are disabled when it does." That's what caching is.