Maximizing VMware ESX Performance Through Defragmentation of Guest Systems

Presented by

RAXCO SOFTWARE

July, 2010

Table of Contents

Executive Overview

It has long been recognized that file and free space fragmentation are detrimental to Windows® system performance. There are numerous articles by third party authors and in Microsoft's *Technet* knowledgebase about the negative impact of file and free space fragmentation on read/write performance, and the need to perform regular defragmentation. In a nutshell, file fragmentation increases the time it takes to read a file; while free space fragmentation increases the time it takes to write a file. In both instances, the extra workload increases the demand for CPU and memory while issuing excess IO to the disk. This excess IO is often the cause of resource bottlenecks which on physical servers, manifests itself in slow server response times, slow application launches, unacceptable file backup/imaging time, wasted disk space and application errors.

This paper details the results of testing performed to determine if there was any measurable performance benefit to be derived from defragmenting virtual servers. With server virtualization, an organization can deploy several virtual instances of Windows Server on a single physical machine. The combined workload of these machines must share the resources of their physical host. Given the well-documented track record of the Windows file system to fragment files and free space; this testing set out to determine if fragmentation inside Windows guests had any impact on VMware ESX performance.

The testing, conducted on a VMware ESX cluster, indicated that file defragmentation and free space consolidation of Windows server guests has a positive effect on the performance of both the Windows guest and the ESX host. Based on the metrics detailed in this report we conclude:

- File defragmentation and free space consolidation combine to significantly reduce the total number of IO that need to traverse the ESX virtual storage stack
- Free space consolidation significantly improves the number of large writes to the disk, improving throughput
- File defragmentation and free space consolidation dramatically reduce disk latency, the time it takes an IO to complete
- File defragmentation and free space consolidation increase sequential IO
- File defragmentation and free space consolidation of Windows guests in a VMware environment reduce the total overhead on the hypervisor
- Free space consolidation improves write throughput for the Windows guests

Test Equipment and Methods

The following system configuration was used in the testing.

ESX server configuration

ESX Version:	4.0
Motherboard:	Intel S5000PSL
CPU Type:	Intel(R) Xeon(R) CPU E5345 @ 2.33GHz
Number of CPUs:	2
Cores per CPU:	4
Logical Processors:	8
Memory:	32 GB

Storage Configuration

RAID controller:	Adaptec RAID 3805
Number of Drives:	4
Drive Type:	WD1001FALS 1TB 7200 RPM 32MB Cache
Total Capacity:	4.0 TB
Number of LUNS:	2
LUN 1 RAID level:	5
LUN 1 Capacity:	2.00 TB
LUN 1 Partitions:	1
LUN 1 Name:	IOTesting
LUN 2 RAID level:	5
LUN 2 Capacity:	744.75 GB
LUN 2 Partitions:	1

VM Configuration

Number of VMs:	5
Operating System:	Windows Server 2008 R2 (64-bit)
Memory:	2GB
Number of CPUs:	2
SCSI Controller:	LSI Logic (no SCSI bus sharing)
Number of Disks:	1
Size of Disk:	50 GB
Provisioning Type:	Thick
Backing Datastore:	IOTesting
Virtual Memory:	None (pagefile disabled)
Network:	Enabled

Testing Overview

Fragmentation in a Virtual Environment

File fragmentation is a function of how the file system allocates space to a file. To create a file, the NTFS file system looks at its $Bitmap metadata file to determine where space is available. The $Bitmap file identifies which logical clusters are in use and which ones are free. If the file system cannot allocate space for the entire file in a contiguous string of logical clusters, the file will be fragmented. It is important to note that this means a file is fragmented as the file system sees it prior to being written to the disk. When a read/write request is received for that file, the Master File Table is accessed and it provides the starting logical cluster number (LCN) and the run length for each fragment needed to satisfy the requested read/write range. The more fragments there are, the longer it takes to access the file. If a file request spans 50 fragments, the file system needs to report the 50 starting LCNs and run lengths to the disk controller.

This same behavior occurs inside each virtualized Windows server. File and free space fragmentation occur within each VMDK and impose a performance penalty on the system. File and free space fragmentation are relevant in a virtual environment because the finite resources of the host must be shared with other virtual machines. If file and free space fragmentation creates a resource bottleneck on one virtual machine, the remaining VMs are going to be deprived access to those host resources.

Products Used in Tests

To test the effect of file defragmentation and free space consolidation, we used Raxco Software's PerfectDisk. PerfectDisk is unique in that it is the only disk defragmentation solution that defragments files and consolidates the free space on the disk into the largest possible contiguous chunk. Defragmenting files improves read access time, but consolidating the free space improves write access and slows the re-fragmentation of the disk.

Testing Procedure

To conduct the tests, we needed to be able to collect performance metrics from the VMware environment. We used the VMware ***vscsiStats*** utility, which intercepts IO between the VMware kernel and the Monitor levels. The vscsiStats utility sorts every IO coming through the storage stack into various categories for performance analysis. A good article on using vscsiStats for can be found here http://communities.vmware.com/docs/DOC-10095. The vscsiStats data provided the quantitative data we used to determine the benefits of file defragmentation and free space consolidation.

A disk was formatted with NTFS and populated with ISO images and other random length files. A custom tool was used to fragment the resulting collection of files. The fragmented disk was imaged so it could be restored to provide an identical starting point in subsequent tests. Finally, five (5) virtual machines were created using identical copies of the fragmented disk. These machines were designated VM1 through VM5 and they used the fragmented disk in the first (the before) set of tests.

The next step was to defragment the imaged disk with PerfectDisk. Five copies of the defragmented disk were also made so we had identical but separate test disks. These disks would be used by the VMs for the second comparative (the after) set of tests.

All extraneous activity on the ESX cluster was shut down to ensure the vscsiStats counters were only counting IO related to the test activities. The vscsiStats were enabled on the five VMs and Microsoft Office was installed on each machine. Upon completion of the installation, the vscsiStats collection on all machines was stopped and the data collected. Next, all five disks were defragmented using PerfectDisk. The vscsiStats counters were restarted and MS SQL was installed. The following tables represent the average condition of the five disks which were designated: Fragmented Disk and PerfectDisk Disk.

	Pre- Software Installation	Post-Software Installation
File System	NTFS	NTFS
Bytes/Cluster	4096	4096
File Fragmentation	29.9%	28.7%
Directory Fragmentation	0.2%	0.6%
Free Space Fragmentation	100%	100%
Metadata Fragmentation	10.7%	12.2%
Excess File Fragments	378723	428174
Excess Directory Fragments	87	525
Excess Metadata Fragments	65	79
Total Number of Files	122322	135676
Total Number of Directories	33099	35783

Table 1 - Fragmented Disk Pre/Post Software Installation Details

	Pre- Software Installation	Post-Software Installation
File System	NTFS	NTFS
Bytes/Cluster	4096	4096
File Fragmentation	0%	0.2%
Directory Fragmentation	0%	0.3%
Free Space Fragmentation	0.5%	0.1%
Metadata Fragmentation	0%	2%
Excess File Fragments	3	1869
Excess Directory Fragments	0	307
Excess Metadata Fragments	0	13
Total Number of Files	122239	135592
Total Number of Directories	33099	36096

Table 2- PerfectDisk Pre/Post Software Installation Details

Test Procedure Anomalies

The intent of the testing was to install both MS Office and MS SQL to a fragmented and a defragmented disk on multiple VMs and to collect the vscsiStats data for both tests. We were able to successfully install MS Office on all five VMs. When installing MS SQL on the VMs, we encountered installation errors just prior to the completion of the tests on two machines. We restored the disks and repeated the process several times only to have the installation stop short of completion.

The vscsiStats used in this report are averages from the three machines that did complete the entire software installation process for MS Office and MS SQL. It is important to note that while the vscsiStats information presented is for three machines, the two machines that did not complete the tests were running at the same time in the same cluster, but their vscsiStats data were not used in the averages since their installations aborted.

MS Office Installation Statistics

MS Office was installed on the five VMs created for this test. For ease of presentation we took the five sets of statistics from these identical disks and compiled averages for the entire set. The vscsiStats data presented here are the average compilations.

IO Counts

In a VMware environment, it is well understood that more IO operations (IOPS) are bad, and fewer IOPS are good. The first vscsiStats metric we looked at was total IO and its breakdown into read IO and write IO after the MS Office installation. In Table 3 we see the total average IO was reduced 21.3% when the disk was defragmented. As one would expect with software installation, almost all of the IO reduction benefit came in the form of writes. Since identical software was installed on identical disks, this indicates defragmentation and free space consolidation produced substantial reductions in total IO. When the file system can find enough contiguous free space to create large writes, the total IO count will be reduced.

	Fragmented Disk	PerfectDisk Disk	% Improvement
Total IO Count	37191	29238	21.3
Read IO Count	3066	2799	8.7
Write IO Count	34125	26439	22.5

Table 3 -Total IO Counts-MS Office Test

IO Size

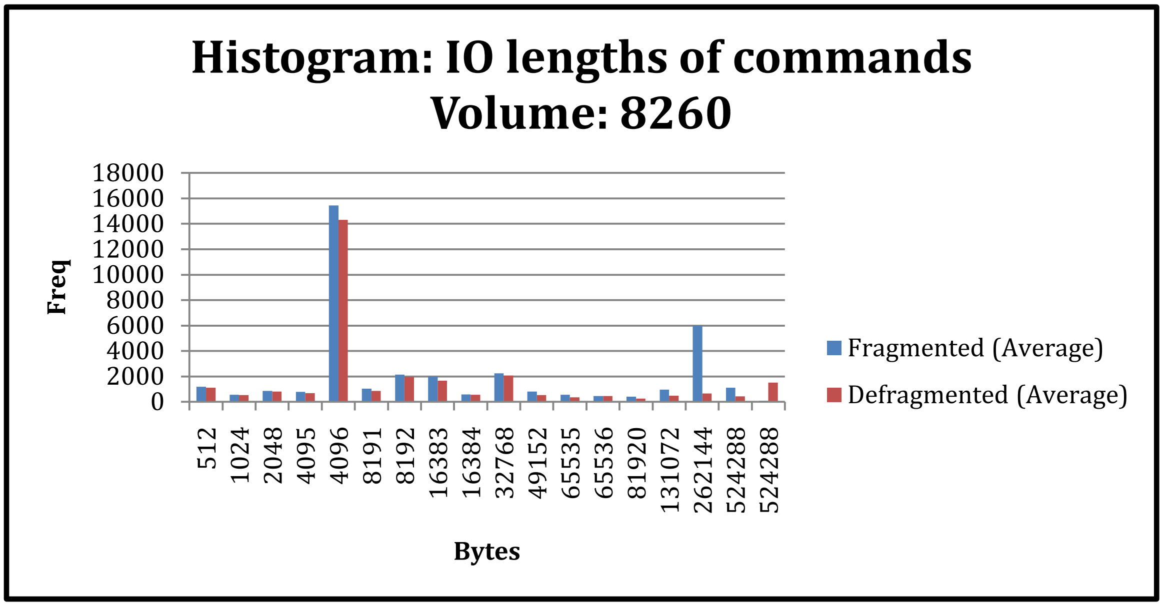
The vscsiStats utility sorts each IO, based on its size, into one of 18 buckets ranging in size from 512 bytes to >524K bytes. The two largest buckets are 524K and >524K and the larger the IO the better since this improves throughput.

The fragmented disk had all of its files and 21GB of free space scattered all over the disk, while the PerfectDisk disk had contiguous files and the free space consolidated into one large chunk. Contiguous free space is the key to creating larger IO. Table 4 shows the total number of IO to each of the two largest buckets.

	Fragmented Disk	PerfectDisk Disk
Total IO Equal to 524K	1103	422
Total IO > 524K	81	1514
Read IO Equal to 524K	0	0
Read IO >524K	23	32
Write IO Equal to 524K	1103	422
Write IO >524K	58	1482

Table 4 - IO Distribution by Size –MS Office Test

The results show the PerfectDisk disk was able to perform **18.7 times** as many total IO greater than 524K than the same disk with fragmented free space. The vscsiStats histogram below shows substantially more IO in the rightmost column.



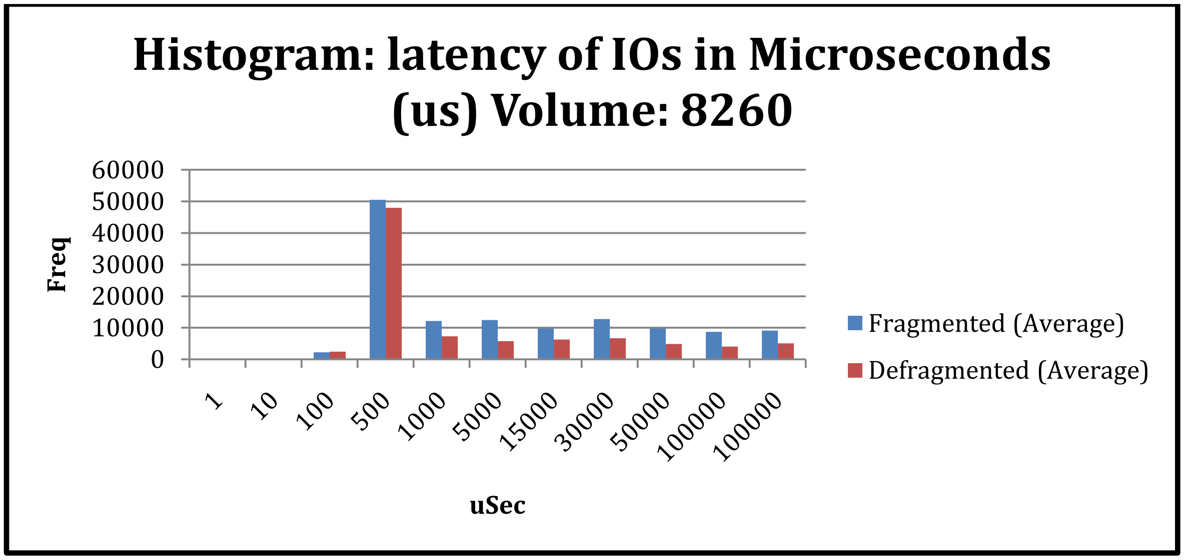
IO Latency

The third metric evaluated was disk latency. Latency is the amount of time it takes the system to complete an IO. Again, vscsiStats sorted each IO into one of 11 buckets ranging from 1 microsecond (1μsec.), to greater than 100,000 microseconds (100ms). A fast IO would be anything faster than 15000 microseconds (15ms), which is about the rated IO access time for a hard drive. A slow IO is anything taking longer than 30ms to complete. Table 5 shows the total IO distribution for all of the vscsiStats buckets greater than 15ms.

	30ms	50ms	100ms	>100ms
Total IO - Fragmented	1129	931	1512	8085
Total IO- PerfectDisk	944	741	1069	3618

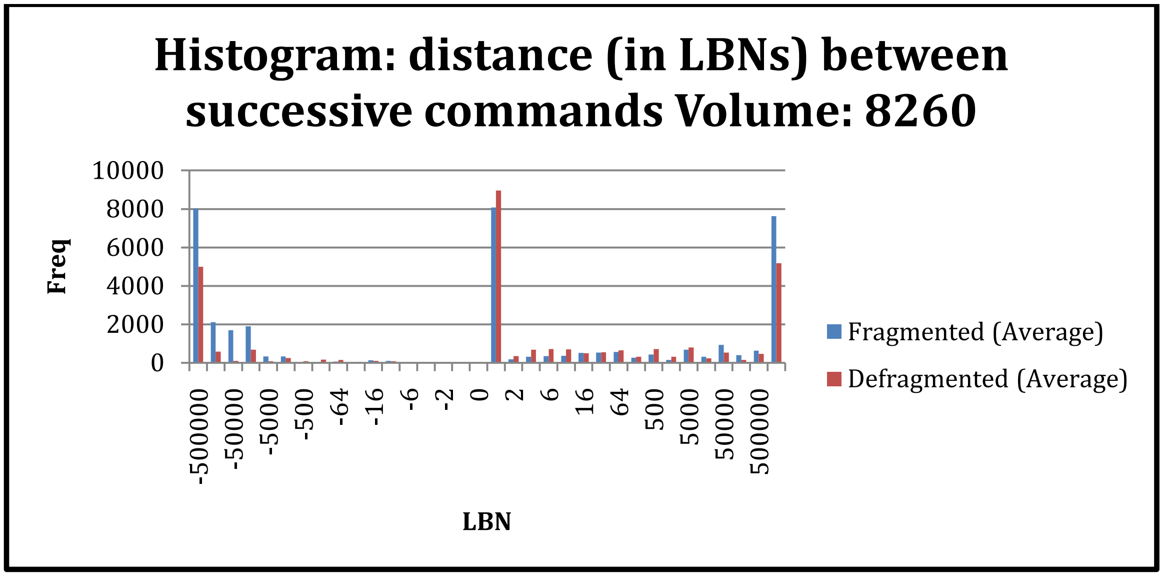
Table 5 - IO Distribution by Latency-MS Office Test

These results show the PerfectDisk disk significantly reduced the total number of slower IO. In fact, the PerfectDisk disk reduced IO taking 100ms or longer by 51.2%.   This data is complementary to Table 4 which shows the same disk doing a greater number of large IO. Latency should decrease as the average IO size increases. The vscsiStats histogram illustrates the decrease in IO latency.



Sequential IO

The vscsiStats utility measures the distance in logical block numbers (LBN) between seeks. The vscsiStats histogram shows that for the MS Office installation, the defragmented disk increased the sequential IO, the number of IO that was only one logical block away (center columns) and reduced the numbers of IO that were at the far right and left of the scales. The fragmented disk had only one LBN between successive commands 23% of the time, while the PerfectDisk disk increased the number of successive commands to 33%.



MS SQL Installation Statistics

Following the completion of the MS Office installation, we defragmented the disk again and installed MS SQL on all the VMs. The MS SQL installation is a larger package with more files and larger files than the MS Office installation, which means more work for the file system. As noted above in Testing Anomalies, we repeatedly encountered a MS SQL installation error on two of the machines which precluded our capturing their statistics. The data presented here are the averages for the three machines where the MS SQL installation succeeded, but where the other two VMs were still running and competing for cluster resources.

<u>IO Counts</u>

The availability of consolidated free space made a considerable difference in the number of total IO performed. The total IO count on the PerfectDisk disk was 37,177 less than the fragmented disk. Reducing the number of IO that need to traverse the VMware virtual stack reduces the hypervisor overhead and frees up CPU and memory resources.

	Fragmented Disk	PerfectDisk Disk	% Improvement
Total IO Count	127703	90526	36.5
Read IO Count	33467	29000	14.3
Write IO Count	94236	61526	44.6

Table 6 -Total IO Count-MS SQL Test

IO Size

As we saw in the MS Office installation, the availability of contiguous free space meant larger disk IO. The MS SQL installation produced a similar result, with the PerfectDisk disk producing **11.9 times** as many total IO >524K. The large IO size means fewer total IO, as we saw in the previous table. The net result is a reduction in VMware overhead and increased host resources.

	Fragmented Disk	PerfectDisk Disk
Total IO Equal to 524K	2512	848
Total IO > 524K	247	2959
Read IO Equal to 524K	33	7
Read IO >524K	125	65
Write IO Equal to 524K	2480	841
Write IO >524K	122	2894

Table 7 - IO Distribution by Size –MS SQL Test

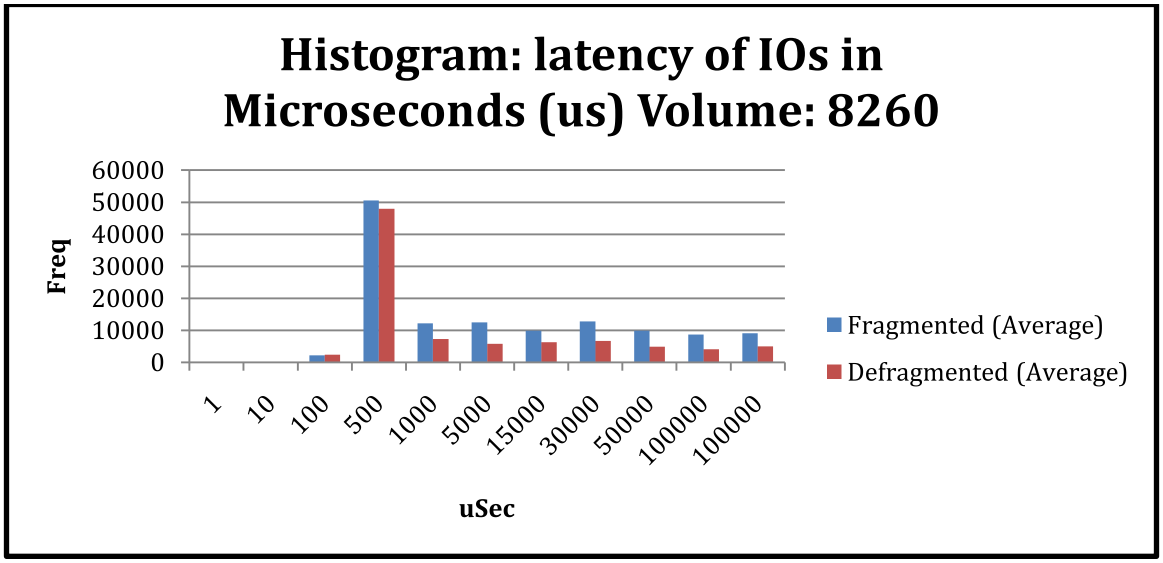
IO Latency

The IO latency distribution for the MS SQL tests illustrates that the PerfectDisk disk reduced the number of slow IO. Since consolidated free space allows larger IO to be written, there are fewer total IO. Latency is reduced when the system can complete a read/write in a single IO. The PerfectDisk disk reduced by 48.7% the total number of IO taking more than 15ms. This reduction was relatively constant across each of the buckets sizes.

	30ms	50ms	100ms	>100ms
Total IO-Fragmented	12749	9877	8700	9116
Total IO-PerfectDisk	6707	4923	4081	5053

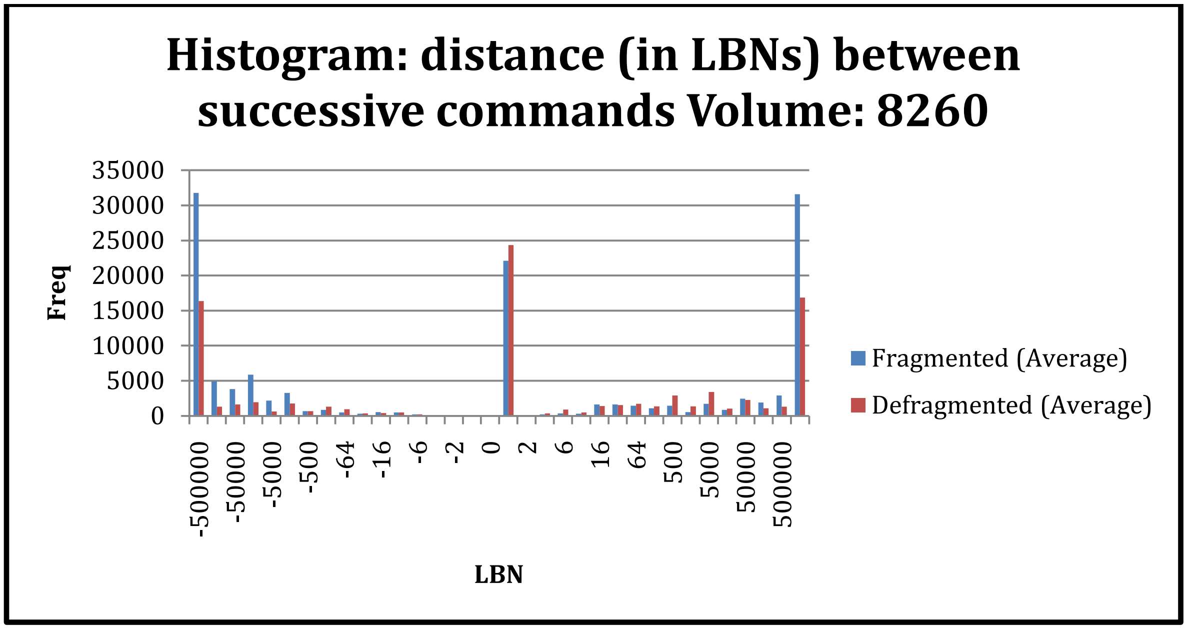
Table 8 - IO Distribution by Latency- MS SQL Test

The vscsiStats histogram illustrates the latency reduction.



Sequential IO

The vscsiStats utility measures the distance in logical block numbers (LBN) between seeks. The vscsiStats histogram shows that for the MS SQL installation, the defragmented disk increased the sequential IO, the number of IO that was only one logical block away (center columns) and reduced the numbers of IO that were at the far right and left of the scales. The fragmented disk had only one LBN between successive commands 19.6% of the time. The PerfectDisk disk has one LBN between successive commands 31.6% of the time.



Applicability to Microsoft's Hyper-V

Raxco chose to perform these tests on the ESX platform because VMware provides the vscsiStats utility that was able to collect the appropriate metrics. Given what we have seen in our work with Hyper-V, we have no reason to believe the results would be much different with Windows guests on that platform.

The Hyper-V platform differs from ESX in one significant way. In addition to having file and free space fragmentation in the Windows guests, Hyper-V also has file and free space fragmentation on the Windows host which compounds the fragmentation issue. As a result, with Hyper-V it is necessary to defragment both the guest and the host in order to eliminate virtual machine resource contention.

Summary

Performance improvements from file defragmentation and especially free space consolidation are commonplace on physical Windows servers. Under the NTFS file system, normal file activity such as creation, extension, truncation and deletion, contribute to both file and free space fragmentation. The net effect of both kinds of fragmentation is increased IO which in turn increases the demand for CPU and memory.

In a virtual environment, multiple virtual machines share the resources of their physical host. A virtualized Windows server guest behaves just like a physical server. All of the file system workings are exactly the same. Files and free space inside the VMDK fragment and create an extra IO load that increases the demand for CPU and memory resources on the ESX host. The purpose of this testing was to determine if the effects of fragmentation had measureable effect on resources in a virtual environment.

Based on the test results we conclude the following:

- Free space consolidation improves disk write performance. The installation of MS Office and MS SQL Server showed disk write improvements of 22.5% and 44.6% respectively when comparing the fragmented disk to the defragmented disk.
- File defragmentation and free space consolidation improve overall IO performance. The total IO counts improved 21.3% and 36.5% respectively on the benchmarked disks.
- Free space consolidation improves total throughput. During the software installations, the defragmented disk was able to perform more IO greater than 524K in size. For the MS Office and MS SQL installations, the defragmented disk generated **25 times** and **23 times more** IO >524K respectively.
- File defragmentation and free space consolidation combine to improve disk latency. Using 30ms or slower as the definition of a slow I0, we saw the fragmented disk on the MS Office install performed 11,657 slow IO, while the defragmented disk only did 6272, an 46% improvement. On the MS SQL install, the improvement on the defragmented disk was 51.3%.

- File defragmentation and free space consolidation improve productivity. While the vscsiStats do not time the work done, we did note installation time for MS Office was about 20 minutes on the fragmented disk and 15 minutes on the defragmented disk. The MS SQL installation was approximately 75 minutes on the fragmented disk and 50 minutes on the PerfectDisk disk.

The test results indicate the defragmentation of files and the consolidation of free space on Windows guests combines to reduce the resource overhead on VMware ESX. The benefit to the end user is better guest and host performance, better guest disk space utilization and potentially enough resource conservation to support one or more additional guests on a host.