White paper

VIRTUALISATION AND FRAGMENTATION

Best practices for optimising virtual disks

Part 2

Written by Topix Media

Sponsored by PSS

Performance System Software S.A.
Rue de Genève 28
1225 Chêne-Bourg
Switzerland
+41 844 555771
info@pss.ch
www.pss.ch

White Paper

VIRTUALISATION AND FRAGMENTATION : Best practices for optimising virtual disks

Part 2

Contents

I. Purpose and Test Overview

The purpose of this white paper is to present strategies and best practices for achieving optimal disk I/O performance from virtual machine platforms. While it offers some broad practices it does not claim to be exhaustive.

This paper covers both current and near-future Windows-focused virtualisation technologies. Ideally this paper has the ambition to be used as a guide as well as a decision-making tool aid for those who will soon implement these technologies and as necessary, those seeking to optimize existing virtualised platforms with already degraded performance.

What is different?

There are already many publications on the subject, which stress the importance of defragmentation and the extraordinarily direct relationship between file fragmentation and virtual system performance. Where most reports have been anecdotal, this in-depth study will present empirical results.

Defining: Virtualisation

To “virtualise” something consists primarily to “give the impression” that something exists whereas physically it is not the case.

A well known use of this concept that almost everybody works with every day is the drive/volume. Our PCs and Macs, for instance, may show 4 logical volumes (C, D, E and F for the PC, HD1, HD2, HD3 and HD4 for Mac) while actually containing only one physical disk.

Partitioning technology allows us to split one physical object into 4 sections or “virtual” volumes. Subsequently we can use this same technology to pool multiple physical disks into one volume. Even our basic use of long term data storage has long been applying the basic concept behind virtualisation.

Defining: Virtualisation (Hardware/Machine/Platform)

Platform virtualisation consists of creating one or more virtual instances of an operating system (OS), typically known as a Guest OS or VM (virtual machine), that is under the false impression that it exclusively owns the hardware on which it believes it is directly installed.

Platform virtualisation “host” software is designed to allocate physical resources (hardware, CPU, memory, etc) statically or dynamically across virtual machines/guest operating systems.

While this ”host” software can be an application run on a full blown general purpose OS, more commonly in enterprise IT infrastructures a virtualisation host is a specialised software layer (a thin “hypervisor”), better streamlined for hosting multiple Guest OS’es on a shared platform. This direct installation is often, in the context of virtualisation terminology, referred to as a bare metal OS or Type-1 hypervisor. Common examples of ‘Type-1’ hypervisor software include Hyper-V™, XenServer™, ESX™ Server.

It is important to note that each hypervisor or guest carries out independently its own applications as if it were the only system on the "physical" hardware. This is perhaps the most critical concept to communicate, as addressing this is arguably the most difficult and most significant consideration for an IT professional looking to implement a virtualisation solution.

Defining: Virtual Hard Disks

There are numerous methods for storing these VMs on hypervisor managed disk subsystems. In most cases, each instance of a "guest" operating system stores its whole data & tasks, logs, etc in a single file (.vhd, .vmdk, etc…) commonly known as a virtual hard disk. Other options are supported on various hypervisor solutions such as "pass through" disks. The various virtual disk types all have their respective benefits and drawbacks and numerous publications already offer best practices for choosing the appropriate type based on the system's purpose.

The 'single' virtual disk file is at present the most common, as it allows for easier migrations and clustering support for the virtual hard disk, and depending on your choice of virtualisation platform, there are effectively two types. One type is a single virtual hard disk file of a fixed size. The benefits of such a file are relatively high and predictable performance at the cost of likely pre-allocating more physical storage space than necessary to accommodate growth. Another virtual disk type is a disk that allocates physical space only as needed, at the cost of some performance overhead. Again the choice of the disk type should be based on the business service level agreements.

The Disk is still the weak link

While virtualisation inexorably adds additional software overhead, the physical limitation of mechanical disk drives is still the bottleneck.

Processors and memory can carry out instructions of magnitude faster than mechanical disks. Therefore, the slower the disk (disk subsystem) is, the slower the entire system runs. This problem worsens unceasingly as the level of disk requests increases, with fragmentation exacerbating the amount of disk I/Os required to fulfill a request.

And you may have already guessed, this is particularly the case with virtualisation, where fragmentation continues to exist in Windows volumes encapsulated within virtual hard disks.

So, a combination of fragmentation and virtualisation tremendously increases traffic to the storage system.

Disk fragmentation in a virtual environment

Fragmentation appears as soon as file segments are distributed into noncontiguous logical space, which typically translates quite directly to physical space. In the case of virtual systems (see image), fragmentation occurs within each virtual hard disk by that Guest OS, and on the virtual platform (e.g. Hyper-V – which uses NTFS) itself.

[1] http://msdn.microsoft.com/en-us/library/cc768526.aspx

Splattering the various parts of files rather than writing a file contiguously, results in a slow hard drive now having to move mechanical parts to locate all the fragments for a given file.

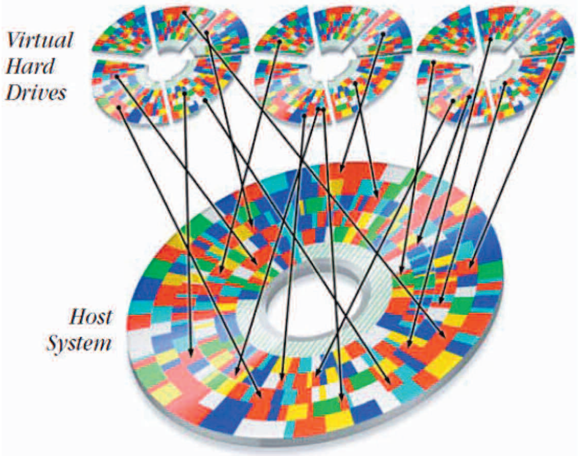

Image 6 (hierarchical disk fragmentation)

Obviously, accessing the various fragments of files, disseminated in several places, considerably slows down access speed.

The more virtual machines there are the more aggregate fragmentation. In fact, the greater the current degree of fragmentation (as file fragmentation increases the likelihood of free space fragmentation) the greater the probability each new file created will fragment.

Shortly, this accumulating fragmentation results in a deceleration of virtual platform performance. It also causes perverse effects: a greater monopolisation of the resources and of I/O.

In extreme cases they are likely to compromise the reliability of the architecture with consequences which go from slow performance annoyance to disastrous system and application "freezes" or hangs.

Virtualisation considerations

There is a world of difference between casual use of Parallel Desktop™ on a Mac (to emulate a PC) and a Hyper-V platform set up to concurrently run tens or even hundreds of virtual machines at, or close to, physical system performance levels.

Without a doubt, hypervisor-based solutions are a must for enterprise needs. The hypervisor's thin framework is far superior in minimising the overhead of that added layer of software. At present, implementing a hypervisor-based virtualisation infrastructure is still relatively complex, requiring powerful hardware, sophisticated management tools, and a very high level of IT competency in order to achieve a satisfactory result.

Maintaining a high service level agreement (SLA) requires deep knowledge of production requirements that core business applications such as DBMS will generate. Provisioning hardware that is capable of supporting requirements is typically the initial step in designing a virtual infrastructure. However, Virtualisation Administrators will need to combine both hardware and software symbiotically to achieve their ultimate goal of consolidating physical systems and increasing efficiencies while averting bottlenecks.

As we described earlier, solving fragmentation is a paramount software consideration that correlates closely to hardware performance. In fact, adding more disk spindles to increase disk I/O performance is the most common misapplied solution for file fragmentation (on both virtual and physical systems).

In all cases, from consumer to enterprise application of the technology, file fragmentation undoubtedly remains a major governing factor for the performance ceiling of virtual machines. The remainder of this paper will explore the performance benefits of solving fragmentation and the specific technological needs that an enterprise virtual infrastructure requires from such a solution.

It will be necessary, on each host and each guest, under Windows, to implement a defragmentation solution; ideally one with the ability to communicate between the host and its guests.

II. Methodology

Hardware setup:

The host was a dual ML350G5 Xeon with 3.0 GHz equipped with 2 GB of RAM and, in RAID1, three SAS 146 GB 15k RPM internal disks to which were connected three external disks: one of 500 GB 7,200 RPM in USB 2.0, a second disc of 1 TB with 7,200 RPM in Firewire 400 and the third one “concatenated” or spanned disk of 2 TB in eSATA.

A number of various operating systems were created as virtual machines including:

- Windows Server 2008
- Windows Vista Ultimate x64
- Windows Vista Professional x64
- Windows XP SP3 x86
- Linux Ubuntu 9.0.4

Software setup (V-locity):

The V-locity host component was installed on the Hyper-V test server, and the V-locity Guest component was installed on the Windows guest virtual machines.

Installing the V-locity components is transparent and can be deployed across a network to the target Host and Guest operating system just the same as any other software application.

All default settings for V-locity were left selected. The user interface and configuration options are very similar to the current Diskeeper 2009 product, making use of V-locity a seamless transition for those already familiar with Diskeeper.

Shortly after installation is was noticed that a “dialog” was automatically established between the V-locity guests and the host components, indicating they had discovered each other and were synchronising resource usage. Specifically this communication channel, based on this new “enlightened” iteration of Diskeeper Corporation’s proprietary InvisiTasking™ technology, allows each V-locity component, within each OS, to execute its defragmentation processes without any impact to any other virtual or host system.

Due to the fact that the built-in Windows Disk Defragmenter has no such solution to coordinate its actions on a Hyper-V platform, Microsoft recommends disabling their native utility in the Hyper-V section of their white paper “Performance Tuning

Guidelines for Windows Server 2008".

Test environment :

Two Hyper-V platforms were established for the series of tests:

1. The first was a fresh installation of Windows Server 2008 Enterprise Edition x64 running the Hyper-V role to this machine being integrated in a workgroup of about fifteen workstations.

2. A second "Server Core" installation running Windows Server 2008R2 Release Candidate (version available June 1st 2009).

The starting environment with respect to fragmentation for this set of tests recreated starting environments similar to those in part 1 of this white paper. Four separate test cases with unique and different conditions were used.

To assist in replicating a production SQL environment we used a simple script that created "words" from 32 up to 64 characters long, which were then written, erased and renamed approximately 60,000 times.

Additionally, we also copied a 4.7 GB DVD iso image, BLOBs, as well as a non-compressed SQL file for a total size close to 5 GB of data.

Last but not least, we installed programs on each virtual machine: office automation suites, image processing programs & utilities, multimedia applications, messaging and instant messaging, browsers, and multimedia players.

Application performance and general data access, including through the network, for the virtual systems all have data transfer ratios equivalent to those observed from earlier samples of non-virtualised systems (as performed in part 1 of this paper).

VHD provisioning considerations:

It was observed, by monitoring the size and the behavior of the .vhd files, that V-locity's defragmentation algorithms are designed to minimise growth of dynamic virtual hard disks. Unless a defragmentation tool is specifically designed in this respect it can artificially and incorrectly grow a dynamic virtual hard disk.

On a related note, V-locity will be automatically disabled on Differencing Disks for this very reason.

This unique feature will assist a system administrator to target appropriate currently over-provisioned dynamic virtual hard disks to return space to their storage pools.

III. Test Results & Analysis

V-Locity was installed on all the Windows systems noted (hosts and guests) and employed to defragment files and consolidate free space. Performance results were similar to those of the non-virtualised systems (displayed below). It was also noted that V-locity balanced its resource usage on each virtual machine so as to eliminate

[2] http://www.microsoft.com/whdc/system/sysperf/Perf_tun_srv.mspx

any potential resource contention with production requirements of any system. Tests were performed on four unique starting environments.

Activity	Enviroment	Fragmented (seconds)	Defragmented (seconds)	Performance Gain %
File Copy (5GB)	Test Case 1	38	31	**18.4%**
	Test Case 2	80	62	**22.5%**
	Test Case 3	104	57	**45.0%**
	Test Case 4	306	61	**80.0%**
Opening larger Office files (100 pages Word and or sheet Excel 2007 size about 5MB)	Test Case 1	13	9	**30.5%**
	Test Case 2	14	11	**21.5%**
	Test Case 3	15	11	**26.5%**
	Test Case 4	25	13	**32.0%**
Disk to Disk Backup (using VSS method to the SATA attached disks)	Test Case 1	1395	996	**28.5%**
	Test Case 2	2950	2287	**22.5%**
	Test Case 3	7200	6600	**8.5%**
	Test Case 4	11250	11030	**2.0%**
Anti-Virus Scan	Test Case 1	58	39	**32.0%**
	Test Case 2	84	59	**29.5%**
	Test Case 3	105	62	**41.0%**
	Test Case 4	186	118	**36.5%**
Bidirectional flows with the Exchange server, server side	Test Case 1	8	6	**25.0%**
	Test Case 2	14	9	**35.5%**
	Test Case 3	19	12	**37.0%**
	Test Case 4	26	16	**38.5%**
Bidirectional flows with the Exchange server, client side	Test Case 1	7	6	**14.0%**
	Test Case 2	14	9	**35.5%**
	Test Case 3	25	13	**32.0%**
	Test Case 4	36	19	**47.0%**
"Bulk" insert into a SQL database of 100,000 random values line	Test Case 1	27	22	**18.5%**
	Test Case 2	34	29	**14.5%**
	Test Case 3	55	33	**40.0%**
	Test Case 4	96	49	**49.0%**
Creation in a SQL database of a series of tables with unique key	Test Case 1	17	16	**5.5%**
	Test Case 2	24	22	**8.5%**
	Test Case 3	49	43	**12.0%**
	Test Case 4	76	62	**18.5%**
Within a complex SQL request	Test Case 1	27	26	**3.5%**
	Test Case 2	29	27	**7.0%**
	Test Case 3	48	40	**16.5%**
	Test Case 4	81	64	**21.0%**
Within a complex SQL request	Test Case 1	37	35	**5.5%**
	Test Case 2	49	37	**24.5%**
	Test Case 3	68	52	**23.5%**
	Test Case 4	96	61	**36.5%**

V-locity Return on Investment (ROI)

While the performance numbers do speak for themselves, it is always important to qualify technical data from a business perspective; namely, "how much money will this make/save". It comes down to dollars and cents and to the return on the investment made. The great news is that V-locity, the world's only virtual platform disk optimiser for Windows® Server 2008 Hyper-V™, offers not only outstanding performance, but also very substantial cost savings. Frankly, V-locity is an obvious investment to make in these times of economic difficulties, as it enables a company to reduce capital expenditures and benefit from measurable savings in hardware and other hard dollar costs, as well as sizeable time savings and employee productivity gains. V-locity is a must-do investment if you care about the performance and reliability of your systems and about the financial health of your company.

Our simulation of the economic impact of V-locity on a 1,000 user installation resulted in the following compelling economic benefits: Recovery of all your investment in V-locity due to hardware and other hard dollar savings in only 10 months.

Once we also factored in the savings of time and improvements in productivity (according to a very conservative estimate), the payback time was reduced to only 6 ½months!

The ROI on V-locity is over 100%, you get paid back in the first year!

Here are the assumptions behind the numbers in the text based on 1,000 workstations:

1. V-locity cost of $149.99 for each of 250 CPU cores, total $37,497.50, rounded to $37,500

2. Hardware savings per annum from:
    a. HDD AFR rate reduction from 10.5% to 4% : 65 HDD x $199.95 each = $12,996.75
    b. HDD useful life increased from 3 to 4 years
        i. 1,000 units x $199.95 cost spread over 3 years is $66,650
        ii. 1,000 units x $199.95 cost spread over 4 years is $49,987
        iii. Savings is $66,650 – 49,987 = $16,662

3. User support / Help desk – estimated 240 hours, (average 15 minutes saved per workstation per year), of an IT professional @ $56/hour = $13,440 per annum

4. Time Savings and Productivity Gains of a savings of: 10 secs/file x 10 files/day x 220 days x 1,000 users x $42/hour / 3,600 seconds/hour = $256,667 per annum of potential savings and, considered only 10% of that as "real economic recovery", to be conservative about it = $25,667 per annum

Total estimated annual savings is:
$12,997+16,662+13,440+25,667 = $68,766 per annum. Compared to an investment of $37,500 this gives a payback in only 6.5 months with an ROI well over 100%.

[3] http://www.cs.cmu.edu/~bianca/fast07.pdf

[4] Price quotes are available from software resellers or Diskeeper Corporation Europe.

White paper sponsored by Diskeeper Corporation

Bio of Topix Technologies SAS

Founded in 1984, Topix Technologies Ltd was deeply involved, in its first years, in industrial engineering and relative computing.

A lot of well known industrialists, quoted on French, English & American Stock Exchanges used Topix Technologies systems & services which were, at the beginning, developed on VAX and PDP computers from Digital Equipment Corp. and all these solutions were further transferred on upcoming standards ten years later.

Bernard Marx, an engineer of the famous Ecole Centrale de Paris, PhD, founded Topix Technologies, a company which diversifies its activities by leverage buy-outs as well as by important investment in the bearish « Media & Entertainment » sector, a sector for which real time and performing solutions, relying on heavy multimedia files cannot afford being fragmented.

In 2009, Topix Group (Topix Technologies, Topix Medias & Topix Press) continues its strategy of developing content management solutions, streaming technologies and publishing specialised printed magazines, newspapers and newsletters. The target audience mainly consists of local authorities & administrations throughout Europe.

Contact : Bernard Marx – Phone. +33 (0)1 55 74 62 00 - Fax. +33 (0)1 55 74 62 10
Postal adress : 5 rue François Ponsard – 75116 Paris – France
Mail : bernard.marx@topix-technologies.fr

Ecole Centrale de Paris PhD
Carried out by Topix Technologies SAS