Jimmy May's Blog

SQL Server Performance, Best Practices, & Productivity

Disk Partition Alignment (Sector Alignment): Make the Case: Save Hundreds of Thousands of Dollars

Jimmy May **8 May 2009 2:15 PM** | **11**

I recently collaborated with Microsoft PFE Daniel Janik to create a template to make the case for disk partition alignment. Perhaps your customers or stakeholders within your organization can benefit. This work was recently broadcast throughout PFE DLs as well as the April 2009 *SQLRAP Newsletter*. Thanks also to Cindy Gross & Ward Pond for their keen eyes.

This information is available elsewhere on my blog & the Internets. It's provided here as public service for the first time for convenient access.

Make the Case for Disk Partition Alignment (Sector Alignment)

Introduction
Disk partition alignment is a best practice and must be applied during disk setup. Partition alignment provides a significant increase in system performance. Microsoft engineers have shown again-&-again that disk alignment can improves performance by not merely the nominal 10 – 15% in RAID systems claimed in official documentation, but commonly 20%, 30%, or more.

How to Save Hundreds of Thousands of Dollars
If your SAN cost $1,000,000, & partition alignment provides 10%, 20%, 30%, or more throughput & IOPs at better latency, then alignment *arguably provides a direct savings of $100,000, $200,000, $300,000 or more*.

Likewise, failure to do so is like throwing that money away.

Look at it another way. Say you have 24 disks on three shelves dedicated to SQL Server. Wouldn't it be great to have the equivalent of an additional shelf of disks? Say you have 200 disks. How different would your life be if you were able to get the performance of the equivalent of 250 disks or more? Imagine the perf gain for hundreds or thousands of disks. As my friend Danielle Nguyen said, we can "make huge perf gains and save huge $$ on additional hardware…especially when a data center has 100s of servers".

Typical Results
Here're the results of an experiment which have been shown to be typical.

Data was collected for disk latency, duration, and other relevant metrics. The Avg. Disk Transfers/sec counters of the PhysicalDisk and LogicalDisk performance objects were used to measure disk latency. Disk latency is a fundamental measure of disk performance.
The experiment was simple, yet convincing. The results were consistent and significant.

Analysis resulted in the following conclusions:

• Disk alignment produced significant improvement compared to nonaligned disks. The measurements document enhancements in excess of 30% for disk latency and duration.
• The performance of six aligned disks was comparable to or better than eight nonaligned disks.

Disk Alignment Experiment
73GB 15K SAS DAS RAID 10
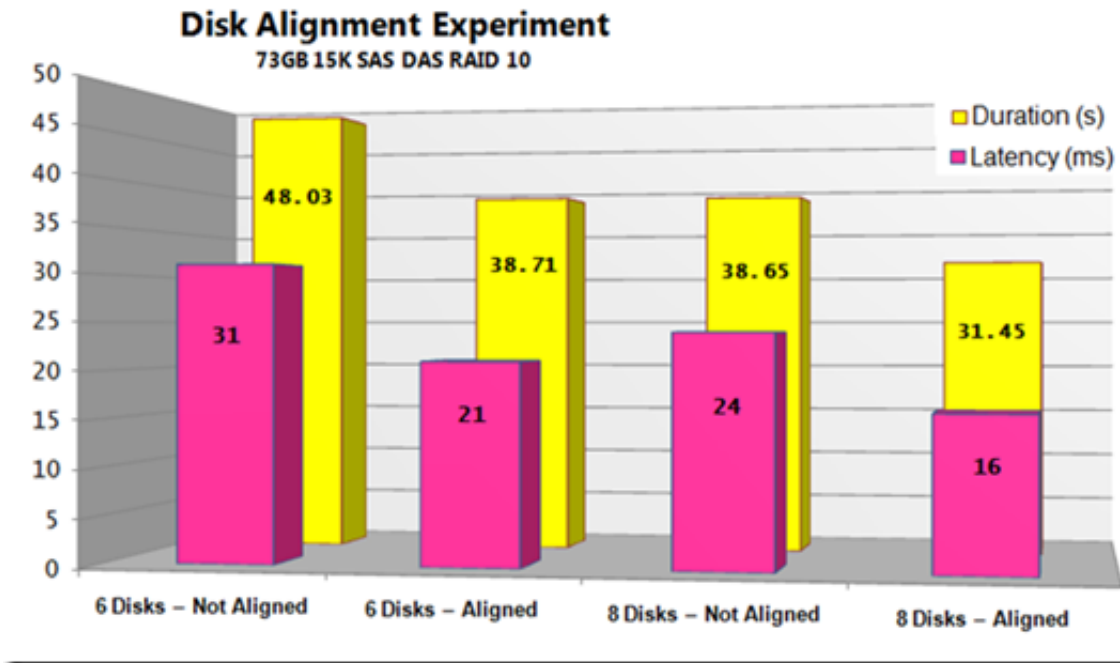
The Fundamental Physics
When the file allocation unit (cluster) size is defined at 64KB per best practice yet partition alignment is not performed, multiple IOPs are required to satisfy single requests.

The following diagrams document the phenomena for cluster sizes of 4KB & 64KB. *For the latter, each-&-ever I/O request requires two IOPs!*
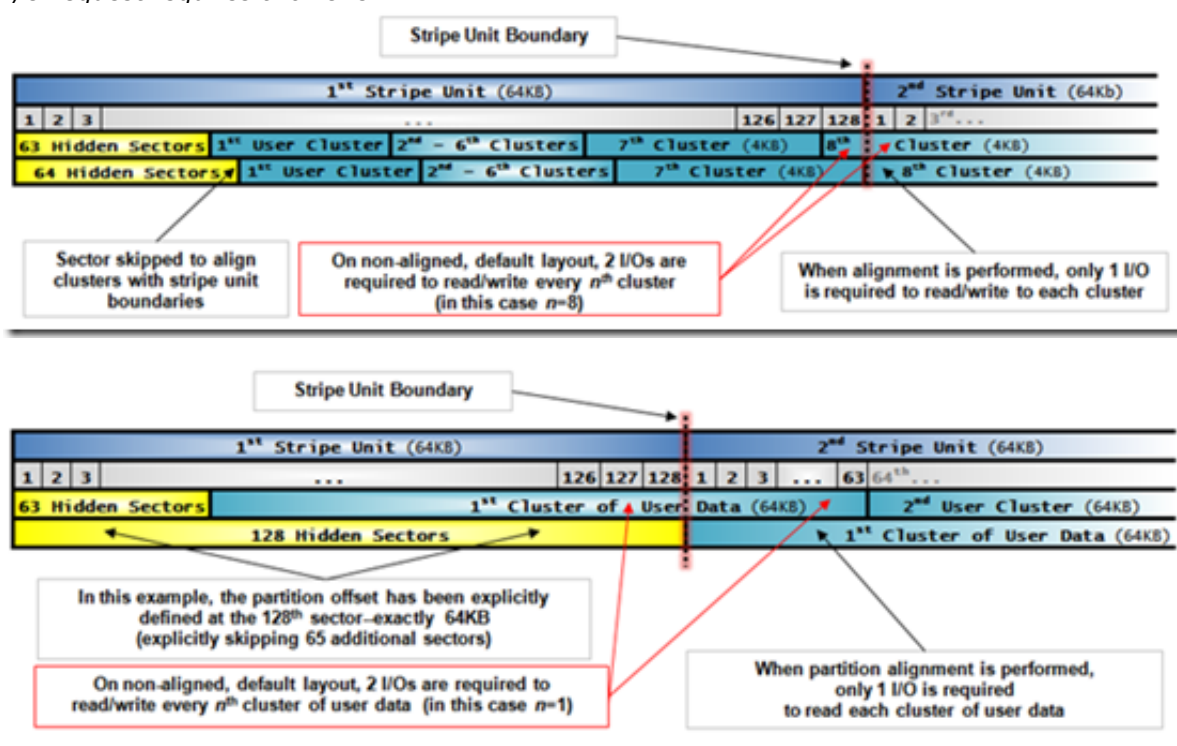
The Remedy
The remedy is simple but has a big gotcha. The good news is that partition alignment is simple to perform; the bad news is that partition alignment must be done at partition creation time, prior to partitions being formatted. This is great if you have a new SAN, but it might be painful to convert large amounts of existing data on misaligned partitions.

- Use the command line utility DiskPart to implement alignment when creating new partitions.

Two Essential Correlations; Three Variables: Partition Offset, File Allocation Unit Size, & Stripe Unit Size

There are two correlations which when satisfied are a fundamental precondition for optimal disk I/O performance. **The results of the following calculations must result in an integer value:**

```
Partition_Offset ÷ Stripe_Unit_Size
Stripe_Unit_Size ÷ File_Allocation_Unit_Size
```

Of the two, the first is by far the most important for optimal performance.

Starting Partition Offset
Use this formula to obtain the starting partition offsets for existing partitions:

```
wmic partition get BlockSize, StartingOffset, Name, Index
```

File Allocation Unit Size
Run this command for each drive to see the file allocation unit size reported in bytes per cluster.

```
fsutil fsinfo ntfsinfo c:
fsutil fsinfo ntfsinfo d:
etc...
```

Stripe Unit Size
The value for stripe unit size must be obtained from your SAN man (or woman).

Dynamic Volumes
Note that the dynamic volumes complicate matters a bit. See this post for more information.

A Common Misalignment Example
The following demonstrates a common misalignment scenario: Given a starting partition offset for 32,256 bytes (31.5 KB) and stripe unit size of 65,536 bytes (64 KB), the result is 0.4921875. This is not an integer; therefore the offset & strip unit size are not correlated. This is consistent with misalignment.

Disk I/O Subsystem Configuration
Configuring optimal disk performance is often viewed as much art as science. Yet an understanding of best practices can result in significant improvements in performance. Some of the many factors which affect disk I/O performance include the number, size, & speed of disks; file allocation unit size; configuration of HBAs & fabric switches; network bandwidth; cache on disk, controllers, & SAN; whether disks are dedicated, shared, or virtualized; RAID level; bus speed; number of paths from disk I/O subsystem to server; driver versions for all components, stripe size, stripe unit size, & workload. Disk partition alignment is the foundation for optimal disk performance. Failure to do so is incompatible with performance & scalability.

References

¡New!
Disk Partition Alignment (Sector Alignment) Best Practices: Characterization, Analysis, and Configuration for Optimal Performance of Windows Disks—Technical Note Series
http://sqlcat.com/whitepapers/archive/2009/05/11/disk-partition-alignment-best-practices-for-sql-server.aspx

Disk Partition Alignment (Sector Alignment) for SQL Server: Part 1: Slide Deck
http://blogs.msdn.com/jimmymay/archive/2008/10/14/disk-partition-alignment-for-sql-server-slide-deck.aspx

Disk Partition Alignment (Sector Alignment) for SQL Server: Part 4: Essentials (Cheat Sheet)
http://blogs.msdn.com/jimmymay/archive/2008/12/04/disk-partition-alignment-sector-alignment-for-sql-server-part-4-essentials-cheat-sheet.aspx

An updated version of the Disk Partition tool for Windows Server 2003 is available
http://support.microsoft.com/default.aspx?scid=kb;en-us;923076&sd=rss&spid=3198

Pre-deployment I/O Best Practices (*Volume Alignment and NTFS Allocation Unit Size*)
http://www.microsoft.com/technet/prodtechnol/sql/bestpractice/pdpliobp.mspx

Disk Subsystem Performance Analysis for Windows
http://www.microsoft.com/whdc/archive/subsys_perf.mspx

Administrivia

> *<Note: Edited for clarity on 20090510>*
> *<Note: Link to whitepaper added & "Disk I/O Subsystem Configuration" section added>*

Jimmy May**,** MCDBA, MCSE, MCITP: DBA + DB Dev
Senior Performance Consultant: SQL Server
A.C.E.: Assessment, Consulting, & Engineering Services

http://blogs.msdn.com/jimmymay

This post was written with the PracticeThis.com plugin for Windows Live Writer

If it is fast and ugly, they will use it and curse you; if it is slow, they will not use it.
    —Computer science professor, billionaire, & entrepreneur David Cheriton

Comments

Disk Partition Alignment (Sector Alignment): Make the Case with this Template | ASP NET Hosting
8 May 2009 2:30 PM

PingBack from http://asp-net-hosting.simplynetdev.com/disk-partition-alignment-sector-alignment-make-the-case-with-this-template/

MSFT Bob
19 Dec 2009 9:34 PM

Jimmy,

If a disk contains multiple partitions and the first is not aligned, will this result in inefficient I/O on the ensuing partitions as well?

I'm thinking specifically of the scenario where the vendor has created a small OEM partition at the beginning of the disk that is not aligned.

Thanks,

Bob L

Jimmy May
20 Dec 2009 7:26 AM

Hello, Bob, you've asked an insightful qustion. I indirectly mention this in the white paper. It's a real concern even in Windows Server 2008. This is why you must *always* validate your partitions. Trust, but verify. Make sure the correlations cited here, in other posts, & the white paper. Let me know if you have any other questions.

Bob Leithiser
20 Dec 2009 8:07 AM

Thanks after reading the article more thoroughly it makes sense that all partitions need to be aligned and not just the starting one. It's important to consider that, because many Win7 users may be assuming their partitions are aligned, when they are not due to the OEM partition. For road warriors who rely on getting every once of throughput out of their laptops, they may want to wipe and reload.

MSFT Bob
20 Dec 2009 8:16 AM

"Virtual drives and the host drives on which they reside must be aligned for optimal performance. The guidelines described here apply to the respective guest & host operating systems"

Does this mean in the scenario where you are using a file-based VHD rather than dedicated disk for a VM, that if you create a partition from within a guest VM, that the guest needs to align the partition? I'm not quite clear on this one, since isn't the physical I/O for the VM handled by the host? Would be interesting to see in general the impact on partition aligning on VMs and best practices. Thank you.

Jimmy May
20 Dec 2009 1:20 PM

Bob, yes, both host & guest must be aligned. I am hoping to do get funding to do experiments to document the impact & provide guidance.

However, be mindful, because of the complexity of HDD internals, single disk & 2 disk mirrors (such as you'd find in a desktop or notebook) are not impacted by DPA. See the "black box" sections of my slide deck.

As we've documented repeately, the impact is substantial for many enterprise disk I/O subsystems with, for example, RAID 5 or RAID 10.

Michael
9 Feb 2010 11:26 AM

If (this is hypothetical) a server had a single disk (no RAID of any kind), does the partition still need to be aligned?

Jimmy May
9 Feb 2010 5:09 PM

Great question. It turns out owing to modern disk internals, DPA for single disks is irrelevant. However, it does indeed apply--& our work shows repeatedly that DPA makes a significant impact--for RAID-0, RAID-10, RAID-5, & RAID-6 arrays. See the white paper or my deck for additional info.

Disk Partition Alignment Best Practices for SQL Server whitepaper

http://msdn.microsoft.com/en-us/library/dd758814.aspx

Disk Partition Alignment deck

http://blogs.msdn.com/jimmymay/archive/2008/10/14/disk-partition-alignment-for-sql-server-slide-deck.aspx

Michael
11 Feb 2010 7:52 AM

Excellent, thank you. So to follow up on my initial question, if a VM is located on NFS storage, why would I need to align the GOS? The VMDK shows as a single disk w/o RAID. And the NFS abstracts the VMDK from the underlying block storage.

Peter
15 Feb 2010 7:57 PM

Michael with newer hard drives, such as the Western Digital EARS Caviar Green series that have an internal sector size of 4096b but report 512b to the operating system may indeed need to be aligned to avoid a performance hit.

http://www.osnews.com/story/22872/Linux_Not_Fully_Prepared_for_4096-Byte_Sector_Hard_Drives

has one users experience with misaligned single drives. SSD drives also run in to issues like this with there large erase blocks.

Jimmy May
12 Mar 2010 12:07 PM

Thanks, Peter. Michael, also be mindful that disk partition alignment is a physical phenomenon. If virtual partitions are configured in such a way that their virtual file allocation units (that is, their clusters) don't align on the same physical boundaries as those of the host, then you're likely to get the very same

performance hit that our work has seen many times in all manner of environments. VMWare has a body of work related to alignment which might be useful to you.