What is an unaligned I/O?

Applies to

- Clustered Data ONTAP 8
- SAN
- FlexPod
- Data ONTAP 7 and earlier

Answer

To explain an I/O alignment, it is first necessary to explain how LUNs are used as storage devices by the host operating system. LUNs and hard disks provide storage as blocks, where a block is an addressable unit of
storage measured in bytes. The traditional block size advertised by hard disks is 512 bytes. New hard disks such as advanced format drives and NetApp LUNs use a larger block size to store the data while continuing to advertise smaller 512 byte blocks to the host for compatibility purposes. The advertised blocks are referred to as logical blocks while the underlying storage block is referred to as a physical block. The NetApp LUN stores data in 4 KB physical blocks which yields eight 512 byte logical blocks per physical block.

The host operating system might begin a Read or Write (I/O) operation at any logical block. However, when the I/O begins at a logical block that is not at the start of a physical block, then the I/O is said to be misaligned.
I/Os are aligned only when they begin at a logical block; that is the first logical block in a physical block. That is to say, the I/O aligns to the physical block boundary.

Identifying I/O Alignment for LUNs

Data ONTAP contains a heuristic detection mechanism that automatically identifies misaligned I/O to LUNs. The heuristic checks the work by observing the position within a physical block where the I/O begins, as well as the I/O length. The heuristic's output will be one of the following: indeterminate, aligned, misaligned or partial writes. The indeterminate state is only applied when the LUN has not received enough I/O to decide on the alignment. For the LUN that is in constant use, this state will be short lived and the heuristic will identify one of the other states. The partial writes state identifies LUNs that receive the majority of I/Os smaller than 4kB. Typically, this state will apply to LUNs used by database transaction or redo logs. This state does not indicate an issue; it is used to separate this situation from the misaligned state. As workloads change, it might be necessary to reset the alignment heuristics by running the `lun stats -z` command.
The following example shows a LUN where the heuristics have identified the I/O as misaligned. The Read and Write alignment histograms record the percentage of I/O that begins at each of the eight positions within a physical block. This hypothetical example shows that 100% of read and write I/Os are starting in the eighth position, also known as a histogram bucket.

*> lun alignment show /vol/luns/my_lun

Multiprotocol type: linux

Alignment: misaligned

Write alignment histogram percentage: 0, 0, 0, 0, 0, 0, 0, 100
Read alignment histogram percentage: 0, 0, 0, 0, 0, 0, 0, 100

Partial writes percentage: 0
Partial reads percentage: 0

The next example shows a LUN where the heuristics have identified the I/Os as aligned. The Read and Write alignment histograms show 100% of the I/O's are beginning in the first position which is align to the physical block boundary.

*> lun alignment show /vol/luns/my_lun

Multiprotocol type: windows

Alignment: aligned

Write alignment histogram percentage: 100, 0, 0, 0, 0, 0, 0, 0
Read alignment histogram percentage: 100, 0, 0, 0, 0, 0, 0, 0

Partial writes percentage: 0
Partial reads percentage: 0

The final example shows the partial writes state on a LUN. Here you will see some amount of the percentage of total I/O being accounted for in the Partial writes and Partial reads fields. The exact amount of partial writes and reads are dependent on the application performing the I/O and the nature of the workload on that application. The percentages might not add up to 100, this is expected and does not indicate an issue with the heuristic calculation.

*> lun alignment show /vol/luns/my_lun
Multiprotocol type: linux

Alignment: partial-writes

Write alignment histogram percentage: 0, 0, 0, 0, 0, 0, 0, 0

Read alignment histogram percentage: 1, 6, 1, 3, 0, 4, 6, 6

Partial writes percentage: 0

Partial reads percentage: 68

For 7-mode nodes:

lun show -v all

/vol/Luns/Luns 2.0t (2198989880250) (r/w, online, mapped)
Comment: "Z: drive"
Serial#: xxxxxxxxxxxxx
Share: none
Space Reservation: enabled
Multiprotocol Type: windows_2008
Maps: xxxxx=0
Occupied Size: 1.8t (2000000000000)
Creation Time: Tue Oct 14 00:00:00 EST 2014
Alignment: misaligned
Cluster Shared Volume Information: 0x0
Space_alloc: disabled
report-physical-size: enabled

Contributing Factors to I/O Alignment for LUNs

The way a host operating system uses a LUN varies greatly by the OS type, partition scheme, file system and application. For the majority of situations where I/O is misaligned, the key contributing factor is the partition scheme employed by the host OS and will also be influenced by the file system or an application such as a database. As a rule of thumb, you should choose the Data ONTAP LUN OS type that most closely matches the host operating system to the OS type. The following table provides additional guidance. In some circumstances custom partition tables might be required. Some Data ONTAP LUN OS types use an offset known as a prefix to enable the default partition scheme used by the associated host operating system to be aligned. For LUNs with a prefix value greater than 0, a custom partition might create misaligned I/O.

Note: Data ONTAP might report I/O misalignments on properly aligned LUNs. In general, these misalignment
warnings can be disregarded as long as you are confident that your LUN is correctly provisioned and your partitioning table is correct.

Example: Boot LUNs are targets for various logs and other smaller writes. The small writes done will look like partial writes, which in turn look like misaligned IO. They can be safely ignored. Most boot LUNs will have some number of partial writes and therefore might look misaligned even though the LUN geometry and LUN type are correct. If you know the LUN geometry is correct then ignore the misalignment.

**Database Log LUNs**

LUNs that are used as the log location for a Database application, such as Oracle (redo log) or Microsoft's SQL Server (transaction log), can see a very random distribution of unaligned write I/O across the alignment histogram. This is normal and expected behavior for this type of I/O. That said, in cases where we know the LUN is being used as a DB log location, and we've confirmed that correct lun type, and offset are being used, no further corrective actions are possible or necessary.

<table>
<thead>
<tr>
<th>LUN ostype</th>
<th>Prefix (bytes)</th>
<th>Prefix (sectors)</th>
<th>Operating System</th>
</tr>
</thead>
<tbody>
<tr>
<td>windows</td>
<td>32,256</td>
<td>63</td>
<td>Windows 2000, 2003 (MBR format)</td>
</tr>
<tr>
<td>windows_gpt</td>
<td>17,408</td>
<td>34</td>
<td>Windows 2003 (GPT format)</td>
</tr>
<tr>
<td>windows_2008</td>
<td>0</td>
<td>0</td>
<td>Windows 2008 and later</td>
</tr>
<tr>
<td>hyper_v</td>
<td>0</td>
<td>0</td>
<td>Windows 2008 Hyper-V abd later</td>
</tr>
<tr>
<td>linux</td>
<td>0</td>
<td>0</td>
<td>All Linux distributions *</td>
</tr>
<tr>
<td>xen</td>
<td>0</td>
<td>0</td>
<td>Citrix XenServer</td>
</tr>
<tr>
<td>vmware</td>
<td>0</td>
<td>0</td>
<td>VMware ESX*</td>
</tr>
<tr>
<td>solaris</td>
<td>1MB</td>
<td>2048</td>
<td>Solaris *</td>
</tr>
<tr>
<td>solaris_efi</td>
<td>17,408</td>
<td>34</td>
<td>Solaris *</td>
</tr>
</tbody>
</table>
* Indicates additional consideration might be required, see the information below.

**Additional Considerations for Linux**

Linux distributions offer a wide variety of ways to use a LUN including as raw devices for databases, volume managers and file systems. It is not necessary to create partitions on a LUN when used as a raw device or when used as a physical volume in a volume manager. If the LUN will be used without a volume manager, then a good practice to follow is to partition the LUN such that it has one partition that begins at an aligned offset, that is a sector that is an even multiple of eight logical blocks. For details on creating aligned partitions of Linux, refer to [How to create aligned partitions in Linux for use with NetApp LUNs, VMDKs, VHDs and other virtual disk containers](#).

**Additional Considerations for VMware ESX/ESXi**

Typically, the unaligned I/O seen on a type VMware LUN will be due to VMDK alignment issues introduced by the guest OS partitioning of its disk.

**Additional Considerations for Solaris**

Solaris offers a wide variety of ways to use a LUN including many different file systems and volume managers. No special partitioning is necessary for Solaris; however, it is important to know when to use either the solaris or solaris_efi LUN ostype. You should use the solaris_efi LUN ostype only for LUNs larger than 990GB that will be formatted with UFS. The solaris_efi LUN ostype is used to offset the default starting location of the first partition after the EFI disk label, which is sector 34. You should use the solaris LUN ostype for all other configurations include SVM, VxVM, ZFS and UFS smaller than 990GB. On Solaris versions 10u8 and newer, where ZFS is being used, a LUN ostype of solaris should be used along with adding physical-block-size:4096 to the host's sd.conf file (ssd.conf for Solaris x86).

**I/O Alignment for Files**

I/O alignment for files works exactly like it does for LUNs. Typically misaligned I/O for files occurs with those files used as virtual disks or disk images accessed via a NAS protocol. It is therefore important that virtual disks have their partitions aligned to file block boundaries for optimal I/O throughput.
Additional Information

N/A