512e and 4Kn Disk Formats

This whitepaper provides the context for 512e and 4Kn disk format migration, as well as pointing out the long-term benefits to customers and potential pitfalls to avoid when moving from 512-byte to 4K sector formats.

Dell Engineering
May 2015

Author: Chetan Kumar, Dell Enterprise Disk Engineering

A Dell Technical White Paper

Contents

1 Overview

A change is coming in the hard drive industry. As storage densities dramatically increase, one of the most elemental aspects of hard drive design — the logical block format size known as a sector has remained constant. The storage industry has transitioned over to a new type of format for media, known as Advanced Format, which has a 4 KB physical sector size. This change brings two new types of media to the enterprise market:

- 4 KB native: This media has no emulation layer and directly exposes 4 KB as its logical and physical sector size. The overall issue with this new type of media is that the majority of current and legacy applications and operating systems do not query for and align I/Os to the physical sector size, which can result in unexpected failed I/Os.
- 512-byte emulation (512e): This media has an emulation function and exposes 512 bytes as its logical sector size (similar to a regular disk today), but makes its physical sector size information (4 KB) available. The overall issue with this new type of media is that the majority of applications and operating systems do not understand the existence of the physical sector size, which can result in a number of issues.

Table 1 Format types

Format type	Bytes per sector value	Bytes per physical sector value
512n	512	512
512e	512	4,096
4Kn	4,096	4,096

Beginning in late 2009, accelerating in 2010, and hitting mainstream in 2011 for client-based HDDs, hard drive companies began migrating away from the legacy sector size of 512 bytes to a larger, more efficient sector size of 4,096 bytes, generally referred to as 4K sectors, and now referred to as Advanced Format by The International Disk Drive Equipment and Materials Association (IDEMA). Enterprise HDDs are also moving to this format, but are slower in adoption. The first Advanced Format enterprise HDD became available in 2012, with a limited set in 2013 and a more general distribution in 2014.

This provides the context for this migration, as well as the long-term benefits and potential pitfalls to avoid when moving from 512 bytes to 4K sectors.

2 Background

The legacy sector format contains a Gap section, a Sync section, an Address Mark section, a Data section and Error Correction Code (ECC) section as shown in Figure 1.

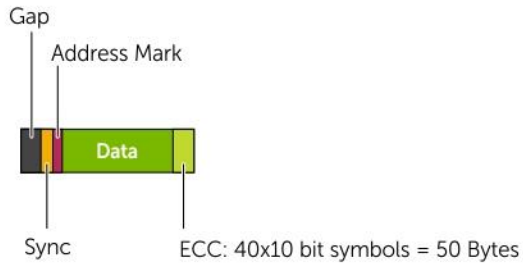

Figure 1 Legacy sector format

The sector layout is structured as follows:

- Gap: Separates sectors.
- Sync: Sync marks the beginning of the sector and also provides timing alignment.
- Address Mark: Not only stores information to identify the sector's number and location but also provides sector status.
- Data: Is designated to store user data.
- ECC: Error correction codes that are used to repair and recover gap, sync, address mark and data fields which may be corrupted during reads or writes.

The 512-byte sector has non-data-related overhead for the Gap, Sync, and Address Mark sections for every 512-byte. Reducing the amount of space used for error correction code improves format efficiency. As the demand for hard drive capacity growth has increased, format efficiency with 512-byte sectors has become a limiting design element. To keep up with the capacity growth demand it was imperative for the hard drive industry had to innovate new methods to improve the error correction efficiency.

512-byte sectors were optimal when managing small, discrete amounts of data. However, applications common in modern computing systems manage data in large blocks, much larger in fact than the legacy 512-byte sector size. The migration to larger sectors has become a fundamental need relative to gaining improvements in error correction and achieving format efficiencies.

3 Long-term benefits of 4K sectors

All hard drive manufacturers have decided to transition to Advanced Format, the industry must adapt to and embrace this change to minimize potential negative side effects. Short-term benefits to end users are not dramatic in terms of immediate capacity increases, however, the migration to 4K-sized sectors will most definitely provide quicker paths to higher areal densities and hard drive capacities, as well as more robust error correction.

The new Advanced Format standard of a 4K-byte sector essentially combines eight legacy 512-byte sectors into a single 4K-byte sector. The Advanced Format standard uses the same number of bytes for Gap, Sync and Address Mark, but increases the ECC. This yields a format efficiency of 97 percent, almost a 10 percent improvement. Together, the benefits of improved format efficiency and more robust error

correction make the transition to 4K sectors well worth the effort. Managing this transition properly to capture the long-term benefits with minimal side effects is a key focus for the hard drive industry.

4 Understanding the impacts of the 4K transition

As noted earlier, there are many aspects of modern computing systems that continue to assume that sectors are always 512 bytes. To transition the entire industry over to the new 4K standard and expect all of these legacy assumptions to suddenly change is not realistic. Over time, the implementation of native 4K sectors, where both the host and hard drive exchange data in 4K blocks, will take place. Until then, Dell and other companies will also implement the 4K sector transition in conjunction with a technique called 512-byte sector emulation.

The most critical aspect of a smooth and successful transition to 4K sectors used in Advanced Format is performance. Whether you are a system builder, OEM, integrator, IT professional, or end user building or configuring a computer, to ensure you have the performance you need for a successful transition, use the operating system to align partitions on a 4Kn drive boundary.

For operating system versions that are not 4K-sector aware, use third-party software or utilities to create hard drive partitions. To ensure the software or utilities you are using are 4K-sector aware, check with your Dell support team.

4.1 512-byte sector emulation

The 512-byte emulation is acceptable because it does not force complex changes in legacy computing systems. However, it carries the potential for lower performance, particularly when writing data that does not neatly correspond to eight translated legacy sectors, as evident by the 512-byte emulation writing process.

4.1.1 Emulated read and write processes

The 512e HDD read and write processes are illustrated below.

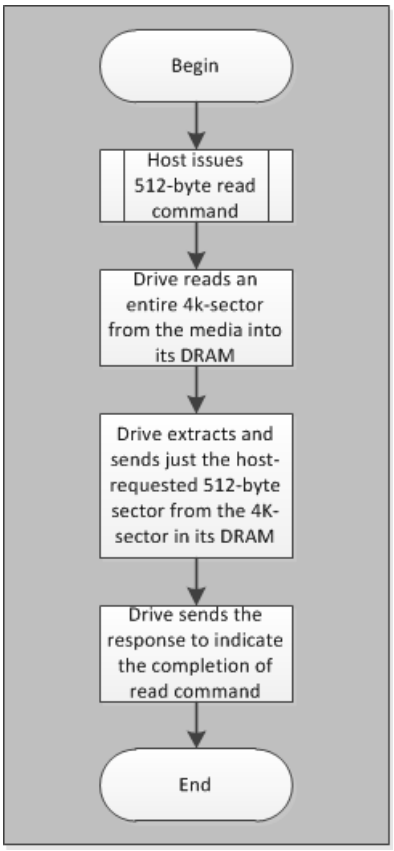

Figure 2 Potential read sequence for 512-byte emulation

The process of reading the 4K block of data and reformatting to the specific 512-byte emulated sector requested by the host computer is performed in the drive's DRAM memory and does not measurably impact performance.

A write process will be more complicated, particularly when data the host computer attempts to write is a subset of a physical 4K sector. In these cases, the hard drive must first read the entire 4K sector containing the targeted location of the host write request, merge the existing data with the new data, and then rewrite the entire 4K sector as shown in Figure 3. The fundamental reason for this is that the drive cannot write just a portion of the 4K sector, but can only write the entire sector at once.

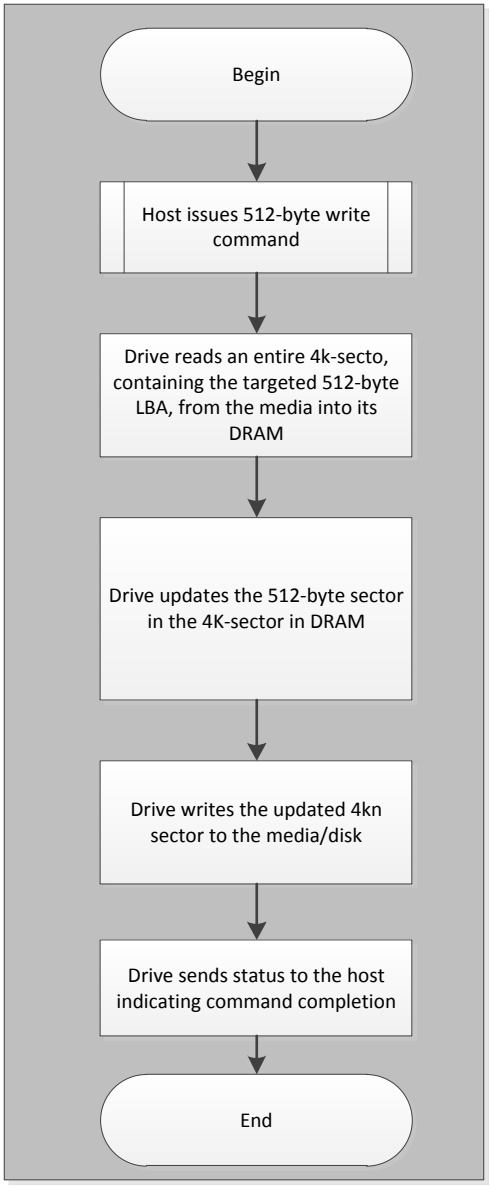

Figure 3 Potential write sequence for 512-byte emulation

In this instance, the hard drive must perform extra mechanical steps in the form of reading a 4K sector, modifying the contents and then writing the data. This process is called a read-modify-write (RMW) cycle, which is undesirable because it has a negative impact on hard drive performance. Minimizing the probability and frequency of read-modify-write instances is the most important aspect of making the transition to 4K sectors smooth and painless.

The causes of read-modify-write reduced performance include:

- Write requests that are misaligned because of logical to physical partition misalignment
- Write requests smaller than 4K in size
- Write requests that are not multiples of 4K in size

4.1.2 Aligned versus unaligned hard drive partitions

Both 512e/4Kn HDDs use 4KB sectors but the read-write operations depends on the transfer size request and alignment with Logical Block Address (LBA).

Each 512-byte sector on the drive is assigned a unique LBA, from zero (0) to the number required based on the drive capacity. The host requests a specific block of data using the assigned LBA. When the host requests to write data, an LBA address is returned at the end of the write request telling the host where the data is located. This becomes important in the transition to 4K sectors since for every 4K sector there are eight different possibilities for where the host LBA starts.

When LBA 0 is aligned to the first virtual 512-byte block in the 4K physical sector, the logical to physical alignment condition for 512-byte emulation is termed Alignment 0. Another possible alignment is when LBA 0 is aligned to the second virtual 512-byte block in the 4K physical sector. This situation is termed Alignment 1 and is shown in comparison to the Alignment 0 condition in Figure 4. There are six additional possibilities for unaligned partitions that can result in read-modify-write events similar to the Alignment 1 condition.

	Sector n								Sector n+1							
Drive LBA	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Host Aligned Command									Host sends 4K bytes write command							
Drive Operations									Drive writes 4k bytes aligned, no RMWs							
									Drive sends write completion status							

Figure 4 Aligned write scenario

Alignment 0 conditions work very well with the new 4K sectors in the Advanced Format standard. This is because a hard drive can easily map eight contiguous 512-byte sectors into a single 4K sector by storing 512-byte write requests in the hard drive's cache (DRAM) until enough contiguous 512-byte blocks are received to form a 4K sector. Since modern computing applications deal with chunks of data that are typically larger than 4K, runts (transfers smaller than 4K) are extremely rare.

	Sector n								Sector n+1							
Drive LBA	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Host Unaligned Command					Host sends 4K bytes write command											
Drive Operations	Because host transfer is unaligned, drive reads 2x 4k byte sectors into its DRAM															
	Updates both 4k sectors, n and n+1, with the host 4k bytes of data															
	Sectors 4-11 will be updated in DRAM															
	Drive writes the updated sectors n and n+1 to the media															
	Drive sends write completion status															

Figure 5 Unaligned write scenario

When hard drive partitions are created that result in an unaligned condition as shown in figure 5, read-modify-write (RMW) cycles occur that can slow hard drive performance. See section 5.5 for information on avoiding these cycles when implementing Advanced Format hard drives.

4.2 Small or runt writes

In modern computing applications, data such as documents, images and video streams are much larger than 512 bytes. Therefore, hard drives can store these write requests in cache until there are enough sequential 512-byte blocks to build a 4K sector. As long as hard drive partitions are aligned, the hard drive can easily map 512-byte sectors into 4K sectors without any performance penalties. There are, however, certain low-level processes, such as meta data writes for example, that can force a hard drive to handle runt writes that are not associated with unaligned partitions. When I/O size is either smaller than 4K or not a multiple of 4K-sector size, the drive must handle the RMW and small writes in the same way. Dell recommends eliminating the sub-4K or non-4K multiple transfers to maximize performance.

4.3 Mixing drives

If you decide to use a mix of drive types, but do not resolve these issues, the overall storage performance may be lower than expected. To mix drives, you must have a good understanding of the operating system, applications, and the configuration such as RAID, volumes, and so on. To ensure compatible drive mix, check with your Dell support team.

5 Preparing for and managing the 4K transition

Now that you understand the benefits of migrating to 4K sectors, as well as the potential impacts to performance, let's look at ways to manage this transition through the context of applications and operating systems.

5.1 Managing 4K sectors in the Windows environment

The most important aspect of managing the transition to 4K sectors is related to the 512-byte emulation drive alignment issues described in section 4. Advanced Format drives work well in an Alignment 0 condition, where the physical-to-logical starting position are equal. Alignment conditions are created when the hard drive partition(s) is created.

5.2 Enterprise Windows support for 4K sector media

Table 2 lists the Microsoft Windows support policy for various media and their resulting reported sector sizes.

- Logical sector: The unit that is used for logical blocks addressing for the media. We can also think of it as the smallest unit of write that the drive can accept.
- Physical sector: The unit for which read and write operations to the device are completed in a single operation. This is the unit of atomic write.

Table 2 Windows support

Drive formats	Reported logical sector size	Reported physical sector size	Supported versions
512-byte Native, 512n	512 bytes	512 bytes	All Windows versions
Advanced Format, 512e, AF, 512-byte Emulation	512 bytes	4096 bytes	Windows Server 2012 Windows Server 2008 R2 with MS KB 982018 Windows Server 2008 R2 SP1

			Windows Server 2008 with MS KB 2553708
Advanced Format, AF, 4K Native	4096 bytes	4096 bytes	Windows Server 2012 (4K data disks are supported and as boot disks in UEFI mode)

Note that Windows Server 2003 and Windows Server 2003 R2 do not support 512e or 4Kn media. While the system may boot up and operate minimally, there may be functionality issues, data loss, or sub-optimal performance. Dell does not recommend using 512e media with Windows Server 2003.

There are a number of software utilities (such as Diskpart and Paragon) that are widely used by system builders, OEMs, value-added resellers, and IT managers for aligning partitions when the operating system doesn't offer or support partition alignment out of the box. Systems today typically consist of multiple hard drive partitions. This means that each partition on the hard drive must be created with 4K-aware partitioning software to make sure proper alignment and performance is ensured.

5.3 Managing 4K sectors in the Linux environment

The key strategies in managing the transition to 4K sectors in a Windows environment also apply to Linux.

Table 3 Linux support

Drive formats	Reported logical sector size	Reported physical sector size	Supported versions
512-byte Native, 512n	512 bytes	512 bytes	All Linux versions
Advanced Format, 512e, AF, 512-byte Emulation	512 bytes	4096 bytes	RHEL 6.1* SLES 11 SP2** Ubuntu 13.10 Ubuntu 12.04.4
Advanced Format, AF, 4K Native, 4Kn	4096 bytes	4096 bytes	RHEL 6.1* SLES 11 SP2** Ubuntu 13.10 Ubuntu 12.04.4

*Red Hat Enterprise Linux 6 supports 4K-sector devices as data disks. 4K-sector boot disks are supported in UEFI mode only.
**SUSE Linux Enterprise fully supports 4K sector drives in all conditions and architectures with one exception. The 4KB/sector hard disk drives are not supported as a boot drive on x86_64 systems booting with a legacy BIOS.

Changes have been made to both the Linux kernel and utilities to support Advanced Format drives. These changes ensure that all partitions on Advanced Format drives are properly aligned on 4K sector boundaries. Kernel support for Advanced Format drives is available in kernel versions 2.6.31 and above. Support for portioning and formatting Advanced Format drives is available in the following Linux utilities:

- Fdisk: GNU Fdisk is a command line utility that partitions hard drives. Versions 1.2.3 and above support Advanced Format drives.
- Parted: GNU Parted is a graphical utility for partitioning hard drives. Versions 2.1 and above support Advanced Format drives.

5.4 VMware support

The key strategies in managing the transition to 4K sectors in a Windows environment also apply to VMware. VMware is yet to show concrete plans to support 512e and 4Kn drives.

Table 4 VMware support

Operating system	512e	4Kn
VMware ESXi	TBD	TBD

5.5 Dealing with unaligned conditions

Using a 4K-aware version of an operating system to create hard drive partitions is a simple, straightforward method for avoiding unaligned conditions. Some third-party firms offer utilities that examine existing hard drive partitions and realign them as needed. This alternative takes additional time and adds steps to the system building or upgrading process. Dell has been developing more sophisticated methods and design systems to manage unaligned conditions to mitigate negative performance impacts.

5.6 512e/4Kn application support

Not all applications are 4K physical sector aware. Table 5 summarizes the 4K application support. When an application is 512e/4Kn aware, the I/Os will be compliant to the file system partition, and the drives will run at expected performance level.

Table 5 Application support

Application	512e	4Kn	Comments
Oracle	Yes	Yes	
Microsoft Exchange	Yes	No	4Kn support ~2016*
SQL	Yes	Yes	
VDI	TBD	TBD	TBD

*It is more of supportability issue that Microsoft has not fully tested/validated Exchange with 4Kn drives. Since Exchange does its own replication, it is very sensitive to the disk types, in particular to disk sector sizes and does not recommend having different disk types as part of the same Database availability group. There are issues where the replication can fail; for example, you have a 512n disk hosting one DB copy and a 512e disk hosting another DB copy. See the following article for more details: http://blogs.technet.com/b/exchange/archive/2013/04/24/exchange-2010-database-availability-groups-and-disk-sector-sizes.aspx

6 Drive labels

The Advanced Format Logo Program was created by IDEMA and the Advanced Format Marketing Work Group to easily identify hard disk drives that employ 4K sectors. While usage of the logo is optional, AF

logos may be seen populating hard drive product labels, product pages, and various literatures to indicate the usage of AF technologies versus legacy sector size architectures that were used in earlier drives.

The AF emulation logo has one rounded corner and is used on any client or enterprise hard drive that is equipped with industry-standard emulation techniques. AF 512e is the current standard by which downward compatibility with legacy sector formats is achieved.

The AF native logo identifies the presence of AF native technologies, where data using long data sector format standards is both recorded on the drive and passed to the host in the AF format. Unlike AF emulation, there is no modification of the sector size by which the data is processed or communicated to the host.

Figure 6 Advanced Format emulation logos

Figure 7 Advanced Format native logos

7 Conclusion

The industry transition away from the legacy 512-byte sector is a certainty, and gives hard drive suppliers another tool for driving improved areal densities. Dell and our customers both benefit from this transition through higher capacity hard drives and continued reliability.

The key to a smooth transition is a well-educated storage community who can avoid potential performance pitfalls. The most critical aspect of a smooth and successful transition to 4K sectors used in Advanced Format is performance. Whether you are building or configuring a computer, do the following to ensure optimal performance:

- Use an operating system to align partitions on a 4Kn boundary
- Use third-party software or utilities to create hard drive partitions; verify with your Dell team to ensure the products you are using have been updated and confirmed to be 4K

Together with our industry colleagues and customers, we can make the transition to Advanced Format 4K sectors smooth and efficient, leveraging the long-term potential benefits for Dell and our customers.

8 Additional resources

http://www.seagate.com/tech-insights/advanced-format-4k-sector-hard-drives-master-ti/

http://msdn.microsoft.com/en-us/library/windows/desktop/hh848035(v=vs.85).aspx

http://blogs.technet.com/b/exchange/archive/2013/04/24/exchange-2010-database-availability-groups-and-disk-sector-sizes.aspx

http://www.idema.org/?page_id=2900