Advanced Format HDD Technology Overview

Last Update: July 2014

Describes the 512e and 4K drive formats

Explains the implications of using these drives

Lists supported operating systems

Helps with understanding the industry transition

Ilya Krutov

Abstract

Historically, a hard disk drive (HDD) media was physically formatted by using a sector size of 512 bytes. However, the need for higher-capacity storage and better data integrity has led to changes in how the data is structured and stored on the disk platter. These changes have resulted in the creation and adoption of the Advanced Format standard for HDDs, which introduced the sector size of 4 KB, by the industry.

However, many currently deployed systems and applications still assume 512-byte sector operations and might even optimize their storage I/O activity to align with the 512-byte sector boundaries.

This paper introduces the Advanced Format, describes its characteristics, and discusses implementation considerations for different IT environments to address potential compatibility and performance issues.

This paper is intended for IT professionals who want to learn more about the Advanced Format and plan to use the Advanced Format HDDs in their IT environments

At Lenovo® Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.

See a list of our most recent publications at the Lenovo Press web site:

http://lenovopress.com

Do you have the latest version? We update our papers from time to time, so check whether you have the latest version of this document by clicking the **Check for Updates** button on the front page of the PDF. Pressing this button will take you to a web page that will tell you if you are reading the latest version of the document and give you a link to the latest if needed. While you're there, you can also sign up to get notified via email whenever we make an update.

Contents

HDD industry transition

A typical hard disk drive is made up of multiple *platters* that are coated with a magnetic material to store data. The platters are divided into *tracks* and *sectors*, and the sector size represents a minimum amount of data that can be transferred to or from the disk drive during one I/O operation.

Fundamentally, the sector size was 512 bytes for many years, and systems and applications were designed and optimized to work with 512-byte sectors. In fact, many modern computer systems still assume 512-byte sector operations nowadays. Figure 1 shows typical 512-byte sector HDD structure.

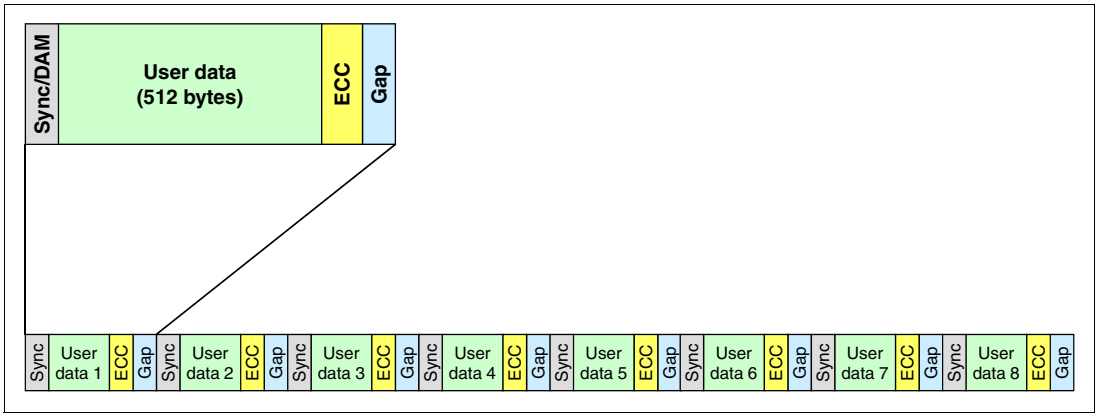

Figure 1 512-byte sector HDD structure

The structure of the 512-byte sector includes the following fields:

- Synchronization/Data Address Mark (Sync/DAM)

  The Sync/DAM field indicates the beginning of the sector and identifies the sector's number, location, and status.

- User data

  The User data field contains actual stored data.

- Error correcting code (ECC)

  The ECC field contains error-correcting code that is used to recover user data that might be damaged during the read or write operation.

- Gap

  The Gap field is used to separate sectors from each other.

Although the standard 512-byte sector was successfully used by the industry for many years, its size has become a limiting factor in achieving higher drive capacities and better error correction efficiency.

With constant improvement in areal densities, the 512-byte sector occupies less and less space on the drive's platter. If, for example, the platter has a media defect within the sector, more data bits stored in this sector might be corrupted with increasing areal densities (see Figure 2 on page 4). This limitation requires more robust and advanced error correction be implemented within the HDD to be able to recover from higher number of bits lost. In addition, today's HDDs with leading areal densities are already reaching limits on the number of bits that can be corrected within 512-byte sectors.

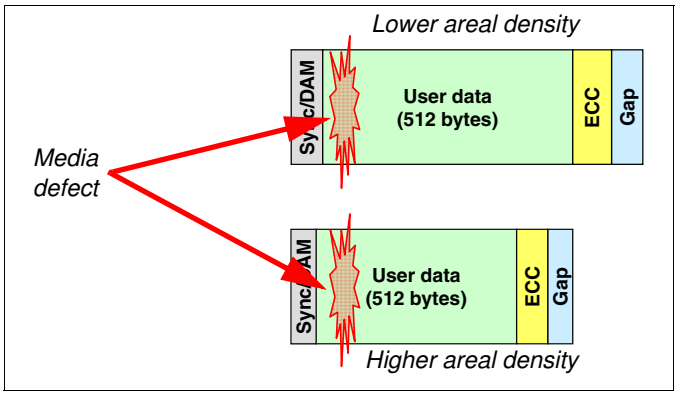

Figure 2 Media defect: Lower versus higher areal density

Another point to consider is the sector granularity, which is the ratio of sector size to overall data storage size. For example, very fine granularity is good when the application manages very small and discrete amounts of data. However, if the application manages large amounts of data or manages data in large blocks, fine granularity becomes less efficient. Many of today's applications can manage gigabytes of data, and can generate storage I/O requests ranging from 4 KB to 1 MB, depending on the application.

As a result, changing to a larger sector size within the industry became a fundamental need to improve error correction and format efficiency. As a response to these needs, the HDD industry introduced the Advanced Format.

Introduction to the Advanced Format

The storage industry has been working on the transition to larger sector hard disk drives since the early 2000s. In 2009, through the coordinated effort within the International Disk Drive Equipment and Materials Association (IDEMA), the Advanced Format standard was formalized and approved as the name for the new standard 4-KB sectors. HDD manufacturers are committed to this new standard and agreed to change to the Advanced Format as a long-term strategy.

Advanced Format introduces the sector size of 4,096 bytes (4 KB) and longer error-correcting code (ECC), and keeps other control fields (Sync/DAM and Gap), as shown in Figure 3.

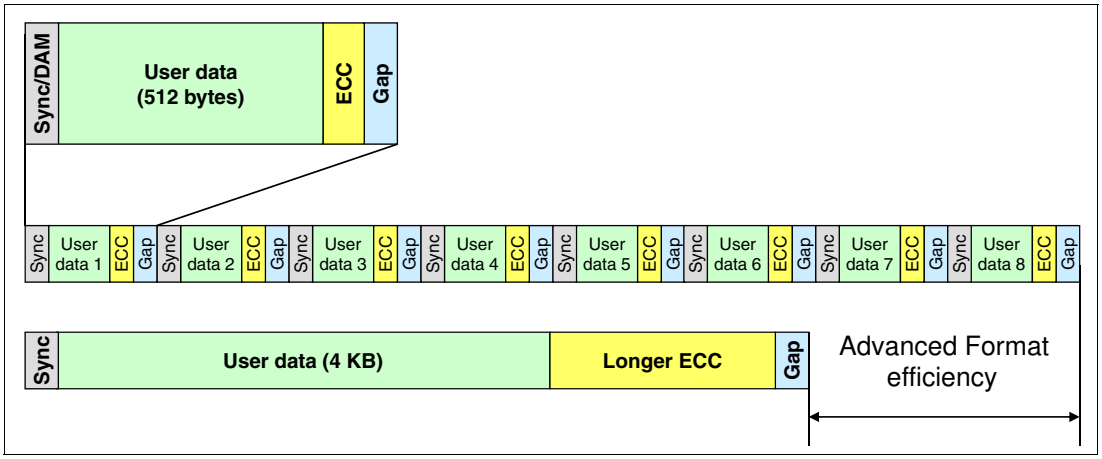

Figure 3 Advanced Format

From the data storage point of view, eight existing data sectors are now stored in a single 4-KB sector. This configuration helps to eliminate multiple control fields, such as *Sync/DAM* and *Gap*, and therefore improves storage efficiency. In addition, larger and more powerful ECC algorithms can be used to improve data integrity with higher areal densities.

The Advanced Format establishes a path to higher capacities, better data reliability, and lower cost per gigabyte (GB) ratio. At the same time, systems and applications (including system's hardware, firmware, UEFI, drivers, operating systems, middleware, and applications) need to be designed to support 4-KB sector HDDs.

Making the transition to the Advanced Format as smooth as possible to achieve the long-term benefits with minimal side effects is the key focus of the HDD storage industry.

For more information about the Advanced Format standard, see the following website:

- Advanced Format (AF) Technology - IDEMA

  http://www.idema.org/?page_id=98

Advanced Format hard disk drive types

With traditional 512-byte sector HDDs, the minimum amount of data that can be addressed and transferred to or from the media in a single I/O operation is 512 bytes. Many systems and applications were developed and optimized for 512-byte sector formats, and transition to 4-KB sector HDDs might cause unexpected compatibility issues for the existing hardware and software. Advanced Format HDDs address potential compatibility issues by introducing two types of sectors to separate physical media transfer blocks from addressable blocks:

- Physical sector

  Physical sector is the minimum amount of data that the *HDD* can read from or write to the *physical media* in a single I/O operation. For Advanced Format HDDs, the physical sector size is 4 KB.

- Logical sector

  Logical sector is the *addressable logical block*, which is the minimum amount of data that the HDD can address. This amount is also the minimum amount of data that the *host system* can deliver to or request from the *HDD* in a single I/O operation. Advanced Format HDDs support 512-bytes and 4-KB logical sector sizes.

This separation allows applications that query the drive's sector sizes to detect drive format and properly align their storage I/O operations to sector boundaries. For applications that expect 512-byte sector HDD formats and do not query sector sizes, this separation establishes a path to 512-byte emulation.

There are two types of Advanced Format HDDs:

- 4-KB native (4Kn) HDDs

  The 4Kn HDD directly maps 4-KB logical sectors (or blocks) to the 4-KB physical sectors.

- 512-byte emulation (512e) HDDs

  The 512e HDD transparently translates 512-byte logical block I/O requests into 4-KB physical sector operations. Each physical sector contains eight logical blocks.

Advanced Format 4Kn HDDs

The Advanced Format 4Kn HDDs transfer data to and from host by using native 4-KB blocks. The system must support 4Kn HDDs at all levels: architecture, disk partition structures, UEFI, firmware, adapters, drivers, operating system and software.

Consideration: If not all system components support Advanced Format 4Kn HDDs, consider using Advanced Format 512e HDDs or traditional 512-byte sector format HDDs. Using 4Kn HDDs in the configurations with system components that do not support native 4-KB transfers might lead to unexpected results and is not supported.

For additional information, see “Disk controller considerations” on page 7 and “Operating systems and applications considerations” on page 8.

Advanced Format 512e HDDs

Many existing hardware and software components are still designed around 512-byte sectors and expect the data be addressed, sent, and received by using the 512-byte I/O blocks. Advanced format 512e HDDs maintain compatibility with existing applications.

The 512e HDDs transparently map logical 512-byte blocks to the 4-KB physical sectors, where each physical sector contains eight logical blocks. This approach is shown in Figure 4.

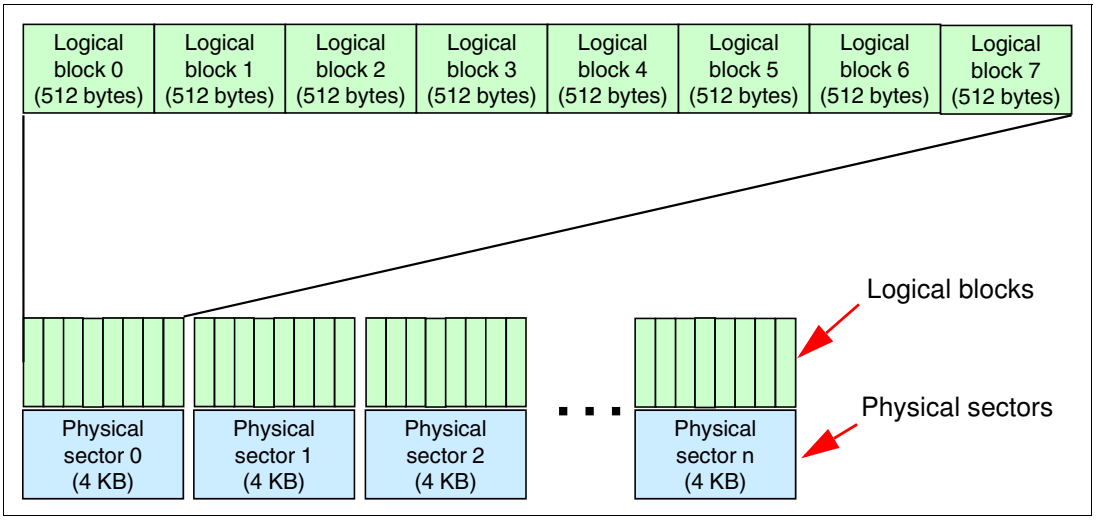

Figure 4 512-byte logical block mapping

With the Advanced Format HDDs, when the application issues a 512-byte READ operation, the HDD reads the entire 4 KB physical sector that contains this requested logical block. The HDD then passes this 512-byte block to the application. When a 512-byte write operation is requested, a *read-modify-write* sequence is initiated. It includes these steps:

1. The entire 4-KB physical sector that contains the addressed 512-byte logical block is read from the HDD platter to the HDD buffer.
2. The HDD locates the 512-byte logical block that needs to be overwritten within the 4-KB block and overwrites it.
3. The entire 4-KB sector is written back to the HDD platter.

Because of this behavior of the 512e mode, certain performance considerations must be taken into account:

- Ideally, the operating system and applications interact with an HDD by using the 4-KB blocks.
- It is optional but desirable for each 4-KB I/O operation to be *aligned* with the 4-KB physical sector. Therefore, it is better for each 4-KB read or write request to involve reading from or writing to only one physical sector.

Misalignment, where 4-KB operations are split across two 4-KB sectors, can have a significantly negative performance impact because each 4-KB I/O request forces the drive to manipulate with data from two physical sectors. Proper alignment is especially important for write operations, where each operation performs two I/O transfers (read and write). The alignment concept is illustrated in Figure 5.

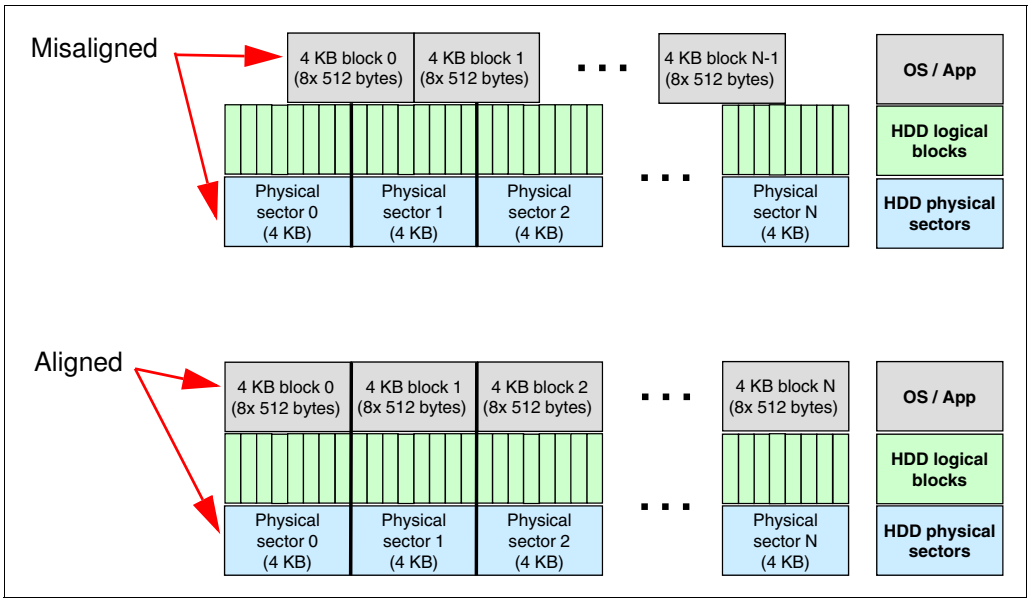

Figure 5 Physical sector and I/O block alignment

Depending on environment, aligned disk I/O can be achieved at the HDD partitioning level, at the operating system and application storage I/O level, or both.

Consideration: Misaligned storage I/O might have negative impact on system performance (sometimes to a significant extent). Aligned storage I/O can help achieve better performance of the storage subsystem.

Disk controller considerations

Non-RAID host bus adapters (HBAs) directly pass the information about the drive's sector sizes to the host. Therefore, the Advanced Format compatibility issues are unlikely.

RAID controllers virtualize physical storage space, and they present logical volumes that are seen as physical storage devices to the host. RAID controllers must be able to recognize the Advanced Format drives to avoid any potential interoperability issues. In addition, if the operating system or application that runs on the host performs storage I/O optimization by

detecting the sector size and aligning storage I/O accordingly, the RAID controller is also able to report storage drive's physical and logical sector sizes to the host.

Support for the Advanced Format HDDs for RAID controllers is typically released as a controller firmware upgrade.

Tip: See the release notes for the most recent firmware upgrades available for IBM ServeRAID™ controllers to confirm Advanced Format support.

Generally, do not mix standard 512-byte and advanced 4-KB format drives in the same RAID array because it might lead to performance issues. For example, if the host assumes 512-byte sector operations, the RAID array might experience performance issues due to misaligned I/O on the 4-KB drives. For more information, see "Advanced Format 512e HDDs" on page 6. However, if the RAID controller is able to report the 4-KB sector size for the logical drive that consists of both 512-byte and 4-KB HDDs to the host for alignment purposes, the host's storage I/O operations are aligned and performance degradation is unlikely.

Operating systems and applications considerations

Operating systems and applications must recognize Advanced Format HDDs to avoid potential interoperability issues that are related to the 4-KB sector format. Advanced Format aware software can properly run the required storage I/O alignment tasks, and this process ensures efficient storage I/O operations. If the software is not able to natively support 4-KB sector alignments, serious performance issues can occur. The effect of these issues can be minimized if proper configuration is made for the operating system and applications.

HDD partition alignment

HDD partitions must first be aligned with 4-KB sector sizes. Logical partitions created by the operating system must start on the 4-KB physical sector boundary to achieve proper alignment. Many modern operating systems recognize the Advanced Format, and they automatically align their partitions to start on the physical sector boundary during installation. If the operating system does not support automatic partition alignment during installation, third-party disk partitioning and alignment tools can be used to align the partitions.

File system logical block size

Consider configuring the *cluster size* for the file system of 4 KB (or multiples of 4 KB). This configuration helps ensure that the file system I/O requests are aligned with the physical sector boundaries. Cluster sizes of less than 4 KB on an Advanced Format HDD can lead to multiple read-modify-write cycles with write operations on a single sector. This misalignment can cause not only cause performance degradation, but also data loss.

Unbuffered writes

Advanced-Format-unaware applications designed and optimized for 512-byte sector operations or applications that run small storage I/O transfers might also cause multiple read-modify-write operations on a single sector due to buffered or unbuffered writes that are not aligned with 4-KB sector boundaries. In addition, certain internal file structures for the applications might be designed for 512-byte sectors, which also causes unaligned I/O.

To help eliminate the issues related to unaligned writes, look at configuring buffering on unbuffered writes and configuring buffered writes in 4-KB blocks or multiples of 4-KB blocks to be aligned with 4-KB physical sector size.

Data mobility

Another area where interoperability issues might arise is data mobility, which is when data is moved from one data store to another one or when the software replication is established across two or more locations. If these data stores have different sector format, the different sector size might be an unexpected event for the application that is being moved from one location to another. For such configurations, consider using the same HDD formats to avoid interoperability issues.

Virtualized environments

For virtualized environments, there are these potential sources of misaligned storage I/O:

- A hypervisor's partition that stores virtual machines (VMs) such as a VMFS partition in a VMware environment
- An OS's partition within a VM disk file that stores an operating system (OS) or data such as an NTFS partition in a Windows Server 2008 VM
- An application that runs within a VM

These potential misalignments are illustrated in Figure 6. Even if only one file system block needs to be accessed, the process might incur numerous additional storage I/O operations due to misalignment. For example, to write to the OS block 1 in Figure 6 in the misaligned configuration, two VMFS blocks need to be written, and each block requires two read-modify-write cycles. In the aligned configuration, only one VMFS block is written, and no read-modify-write sequence is initiated for OS block 1 write operation.

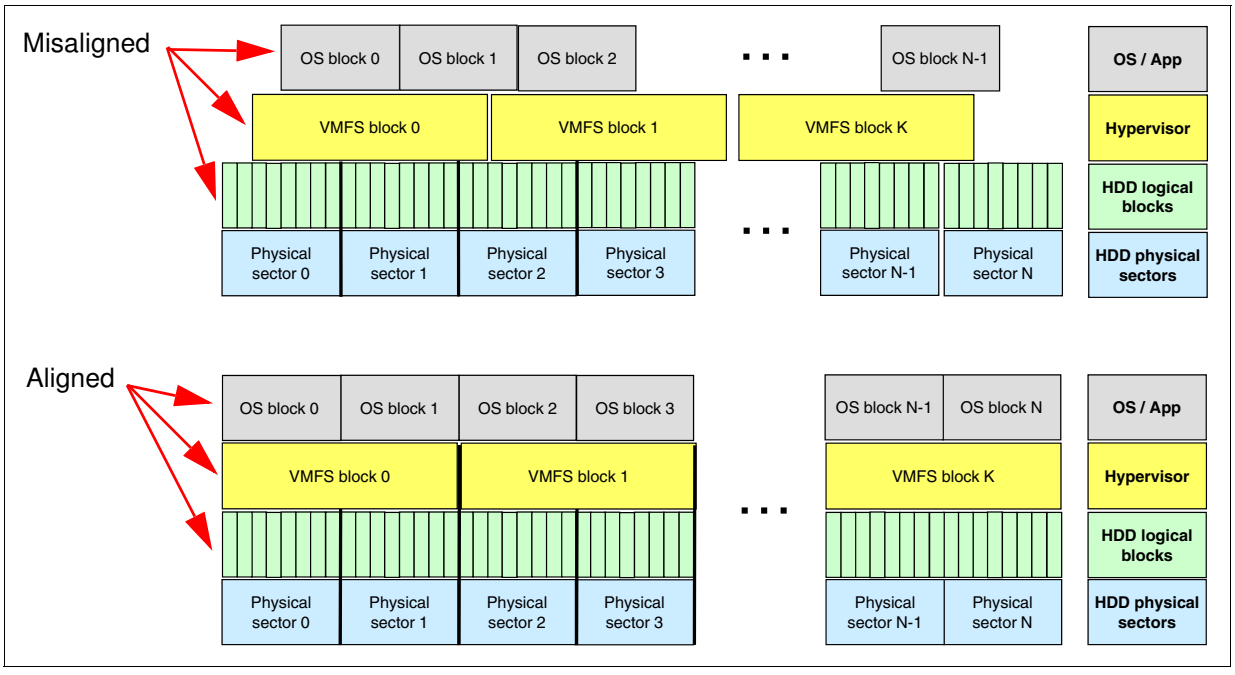

Figure 6 Unaligned versus aligned partitions: Virtualized environments

Therefore, it is critical to ensure that both VMFS and VM guest OS partitions are aligned with 4-KB sector boundaries. VMFS partition alignment is performed automatically or manually by the hypervisor management tools while OS partition alignment is done by the guest OS tools (also automatically or manually). In addition, the guest OS's file system logical block and application-specific considerations within the guest OS also apply.

Operating systems and hypervisors summary

Table 1 summarizes the information about whether the operating system natively supports Advanced Format HDDs (both 4Kn and 512e), as well as support for 4-KB (or multiples of 4 KB) file system logical blocks and automatic partition alignment.

Table 1 Advanced format aware and unaware server operating systems

Server operating system or hypervisor	4-KB[a] logical block	AF 4Kn	AF 512e	Automatic partition alignment
Microsoft Windows Server 2012 / 2012 R2	Yes	Yes	Yes	Yes
Microsoft Windows Server 2008 / 2008 R2	Yes	No	Yes	Yes
Microsoft Windows Server 2003 / 2003 R2	Yes	No	No	No
Red Hat Enterprise Linux 6	Yes	Yes	Yes	Yes
Red Hat Enterprise Linux 5	Yes	No	No	No[b]
SUSE Linux Enterprise Server 11	Yes	Yes	Yes	No[b]
SUSE Linux Enterprise Server 10	Yes	No	No	No[b]
VMware vSphere 4.x (ESXi)	Yes[c]	No	No	Yes
VMware vSphere 5.x (ESXi)	Yes[d]	No	No	Yes

a. Also supports multiples of 4 KB.
b. You can use Linux Partitioning Utility to create partitions.
c. VMware vSphere 4.x uses VMFS 3 with a configurable block size of 1 MB, 2 MB, 4 MB, or 8 MB.
d. VMware vSphere 5.x uses VMFS 5 with a fixed block size of 1 MB.

In Table 1, Microsoft supports Advanced Format 512e HDDs starting from Windows Server 2008 with extra updates installed, and native 4Kn mode is supported starting from Windows Server 2012.

Important: Microsoft does not support Windows Server 2003 / 2003 R2 with Advanced Format drives (both 4Kn and 512e).

For more information about Microsoft Windows support for the Advanced Format HDDs, see the following publication:

- Advanced format (4K) disk compatibility update

  http://msdn.microsoft.com/en-us/library/windows/desktop/hh848035.aspx

Linux support for the Advanced Format HDDs is available in Linux kernel version 2.6.31 or later. Disk partition utilities such as *fdisk* and *parted* were also modified to support Advanced Format drives.

VMware vSphere infrastructure is unaware of the Advanced Format HDDs. However, it automatically performs partition alignment for VMFS partitions that are created with vSphere Client or VMware vCenter. For VMFS3 (vSphere 3.x or 4.x), partitions are aligned with 64-KB boundaries. For VMFS5 (vSphere 5.x), partitions are aligned with 1-MB boundaries.

Authors

This paper was produced by the following team of specialists:

Ilya Krutov is a Project Leader at Lenovo Press. He manages and produces pre-sale and post-sale technical publications on various IT topics, including x86 rack and blade servers, server operating systems, virtualization and cloud, networking, storage, and systems management. Ilya has more than 15 years of experience in the IT industry, backed by professional certifications from Cisco Systems, IBM, and Microsoft. During his career, Ilya has held a variety of technical and leadership positions in education, consulting, services, technical sales, marketing, channel business, and programming. He has written more than 200 books, papers, and other technical documents. Ilya has a Specialist's degree with honors in Computer Engineering from the Moscow State Engineering and Physics Institute (Technical University).

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service.

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/redp5119

Trademarks

Lenovo, the Lenovo logo, and For Those Who Do are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. These and other Lenovo trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by Lenovo at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of Lenovo trademarks is available on the Web at http://www.lenovo.com/legal/copytrade.html.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Lenovo®
Lenovo(logo)®
ServeRAID™

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows Server, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.