# Technical report:

# Storage Block Alignment with VMware Virtual Infrastructure and IBM System Storage N series

*Correctly aligning LUNs and virtual disk files*

*Document NS3593-0*

January 22, 2008

## Table of contents

# Abstract

*This technical report discusses how to correctly align LUNs and VMware .vmdk virtual disk files in VMware ESX Server to IBM System Storage N series storage blocks to achieve improved virtualized performance.*

# Introduction

In order to get the best performance from a storage system, the storage blocks on the logical unit number (LUN) must be aligned to the file system on the operating system (OS). This is typically done by setting the OS type when the LUN is created on the storage system. This document discusses LUN block alignment and the special considerations needed in a virtualized environment.

## LUN alignment basics

There are many cases where, by default, a file system block will not be aligned to the storage array. In Figure 1 below, you can see that the blocks do not align. This would mean that for each random read or write, two blocks would need to be read or written. This can negatively impact the performance of the storage array. Sequential writes will also be affected, although to a lesser extent.

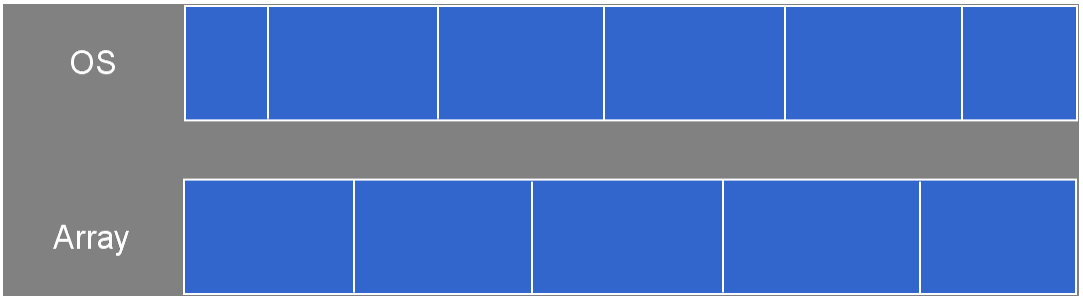


*Figure 1) Blocks not aligned.*

The blocks of this storage system are not aligned to the OS's file system.

In a nonvirtualized environment, the block alignment is done by selecting the appropriate LUN protocol type when the LUN is created. This will align the storage to the file system, as seen in Figure 2.

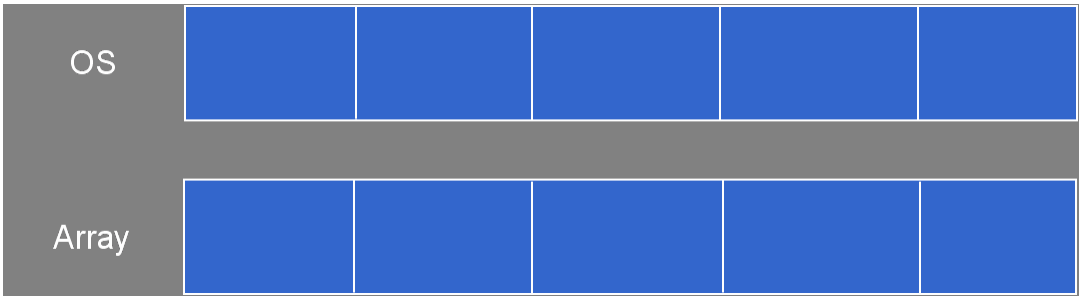


*Figure 2) Blocks aligned.*

The blocks of this storage system are correctly aligned to the OS's file system.

## LUN/virtual disk alignment with VMware Infrastructure 3 or ESX 2.5

Virtualization products such as VMware add another layer of complexity to alignment, as seen in Figure 3. Not only must the virtual machine file system (VMFS) datastore be correctly aligned to the storage blocks, but the guest OS file system must also be aligned.

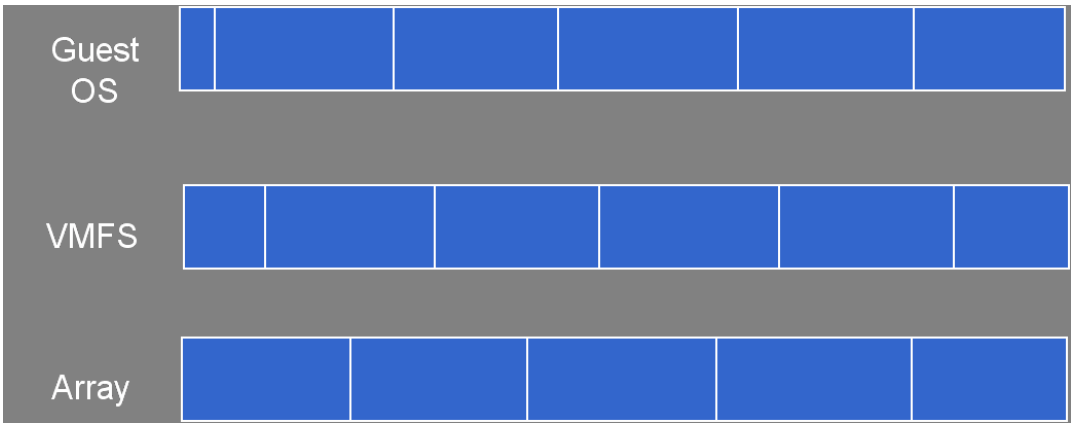


*Figure 3) Virtualized blocks not aligned.*

The blocks of this storage system are not correctly aligned to the virtualized file system or the guest OS.

## Correct LUN alignment

Selecting the correct LUN protocol type will align the VMFS file system to the storage array blocks, but the guest OS still needs to be aligned (see Figure 4). This process is discussed further later in the paper.

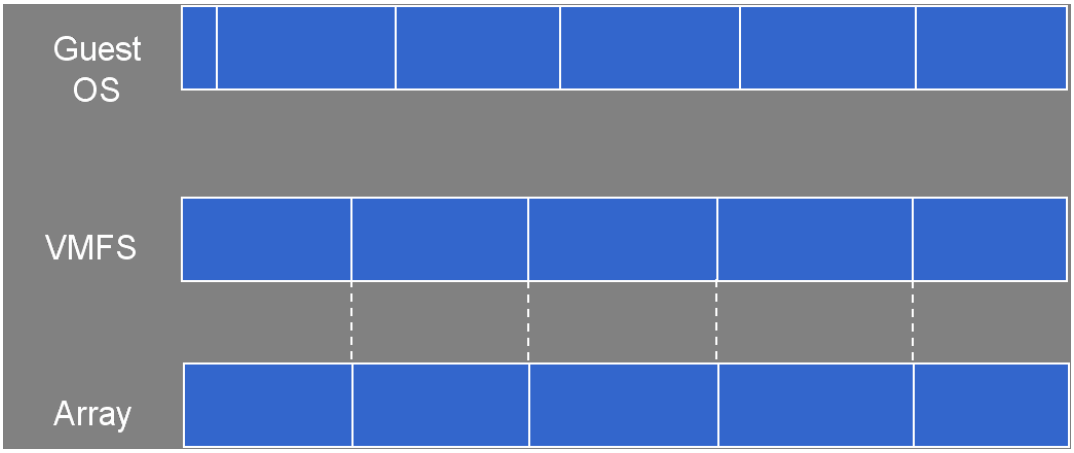


*Figure 4) Virtualized host correctly aligned to storage blocks.*

The blocks of this storage system are correctly aligned to the virtualized file system, but the guest OS is still unaligned.

## Counters that indicate improper alignment

There are various ways of determining if you do not have proper alignment. Using perfstat counters, under the wafl_susp section, “wp.partial_writes“, “pw.over_limit“, and “pw.async_read,“ are indicators of improper alignment. The “wp.partial write“ is the block counter of unaligned I/O. If more than a small number of partial writes happen, then IBM® System Storage™ N series with WAFL® (write anywhere file layout) will launch a background read. These are counted in “pw.async_read“; “pw.over_limit“ is the block counter of the writes waiting on disk reads.

Using IBM System Storage N series with Data ONTAP® 7.2.1 or newer, there are some per LUN counters to track I/O alignment:

lun:read_align_histo: 8bin histogram for reads that tracks how many 512b sectors off the beginning of a WAFL block an I/O was; reported as a % of reads.
lun:write_align_histo: same for writes.
lun:read_partial_blocks: % reads that are not a multiple of 4k.
lun:write_partial_blocks: same for writes.

If any of these except for read/write_align_histo[0] is nonzero, you had some misaligned I/O.

# LUN and .vmdk alignment steps

## Aligning virtual disks (.vmdks) to an IBM N series VMFS datastore

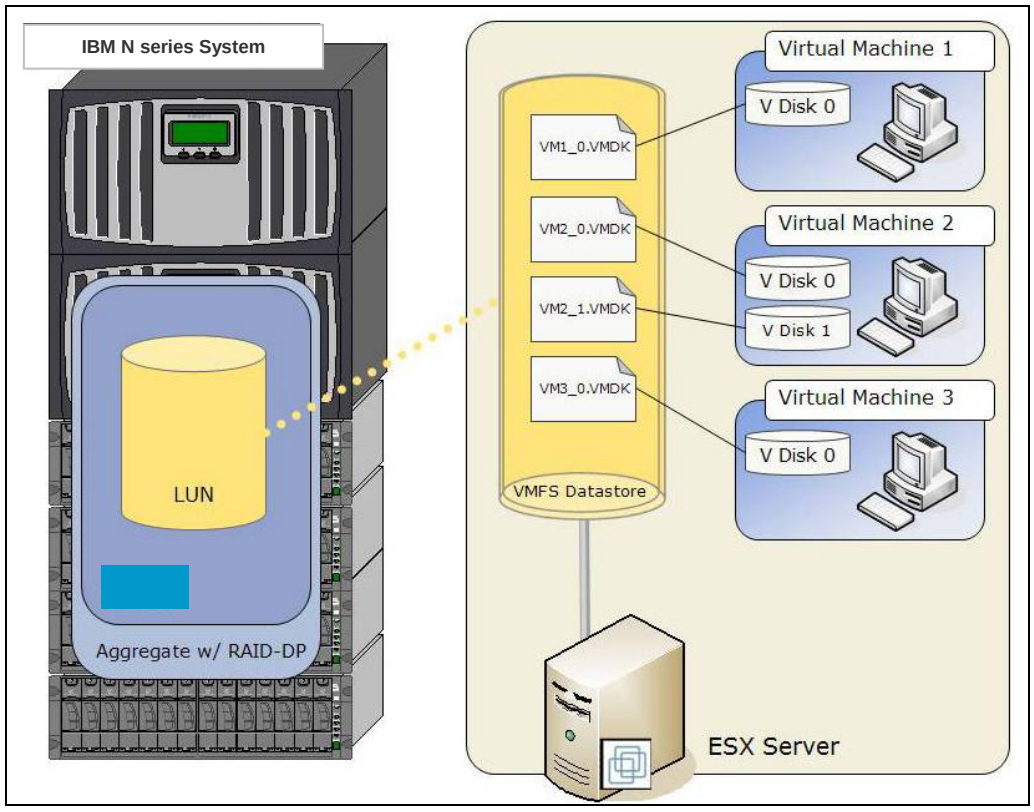


*Figure 5) VMFS datastore on IBM N series storage.*

In VMware Infrastructure 3 (VI3), the IBM N series storage system can be aligned to the VMware VMFS file system by selecting VMware from the LUN Protocol Type when the LUN is created. This creates the block alignment as seen in Figure 4 above. To correctly align the Microsoft® Windows® guest OS to the storage blocks, use the following steps.

Create a Virtual Machine using Virtual Infrastructure Client.

After the Virtual Machine is created, we need to add a second hard drive. Add a second hard drive using the following commands:

Right-click the Virtual Machine and select “Edit Settings…”
Click “Add…”
Select “Hard Drive” and click “Next>.”
Select “Create a new virtual disk” and click “Next>.”
Use the disk capacity you would like your secondary drive to be, then click “Next>.”
Click “Next>”, then click “Finish,” and then click “OK.”

This creates a second hard drive named “{Virtual_Machine_Name}_1-flat.vmdk” This vmdk (virtual machine disk) may be removed from this Virtual Machine and presented to another virtual machine to perform the next section. Or Windows can be installed on the first drive of this Virtual Machine with the next section performed on the second drive just created.

From within Windows in the Virtual Machine, if you run “System Information” (msinfo32.exe in Start-> Run) and select Components → Storage → Disks, scroll to the bottom and you will see the Partition Starting Offset information. This number needs to be perfectly divisible by 4096. The default .vmdk, you will see the Partition Starting Offset set to 32,256 as seen in Figure 6. 32,256 / 4096 = 7.875, and thus this file system is not correctly aligned.

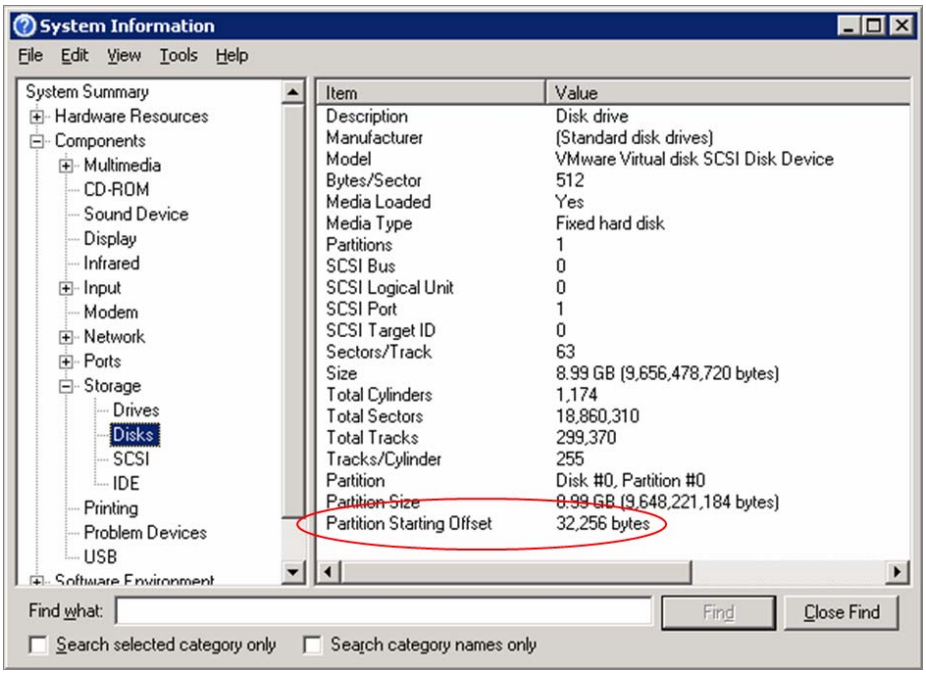


*Figure 6) Partition Starting Offset of an unaligned file system on a .vmdk.*

The disk is not aligned to the storage blocks as this number needs to be perfectly divisible by 4096.

To correctly set the Partition Starting Offset, use the following steps:

Start a Command Prompt.

C:\>diskpart

DISKPART> list disk

Two disks should be listed.

DISKPART> select disk 1

This selects the second disk drive.

DISKPART> list partitions

This step should result in a message stating “There are no partitions on this disk to show.” This message confirms the disk is blank.

DISKPART> create partition primary align=64

Viewing the second disk using “System Information,” the Partition Starting Offset is now 65,536, as seen in Figure 7. 65,536 / 4096 = 16, and thus this file system is properly aligned to the storage blocks.

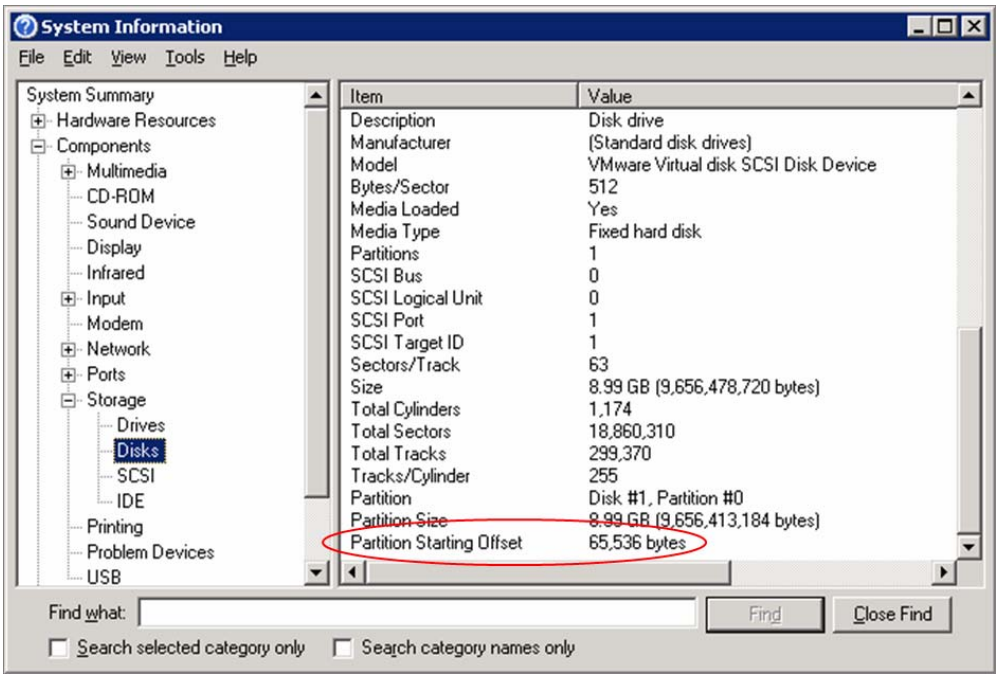


*Figure 7) Partition Starting Offset of a correctly aligned file system on a .vmdk.*

The disk is correctly aligned to the storage blocks as this number is perfectly divisible by 4096.

This second drive can now be removed from this Virtual Machine and presented to another Virtual Machine as Drive 0 for Windows to be installed. It is suggested that a copy of this .vmdk be saved as a “Gold Copy.” This Gold Copy can then be manually copied to a virtual machine directory on VMFS for proper alignment. Another method for this would be to install Windows on this correctly aligned .vmdk, then select the “Clone to Template” or “Convert to Template” option in the Virtual Infrastructure Client. The Virtual Machines deployed from this correctly aligned .vmdk template will also be correctly aligned.

## Aligning virtual disks to an IBM N series NFS datastore

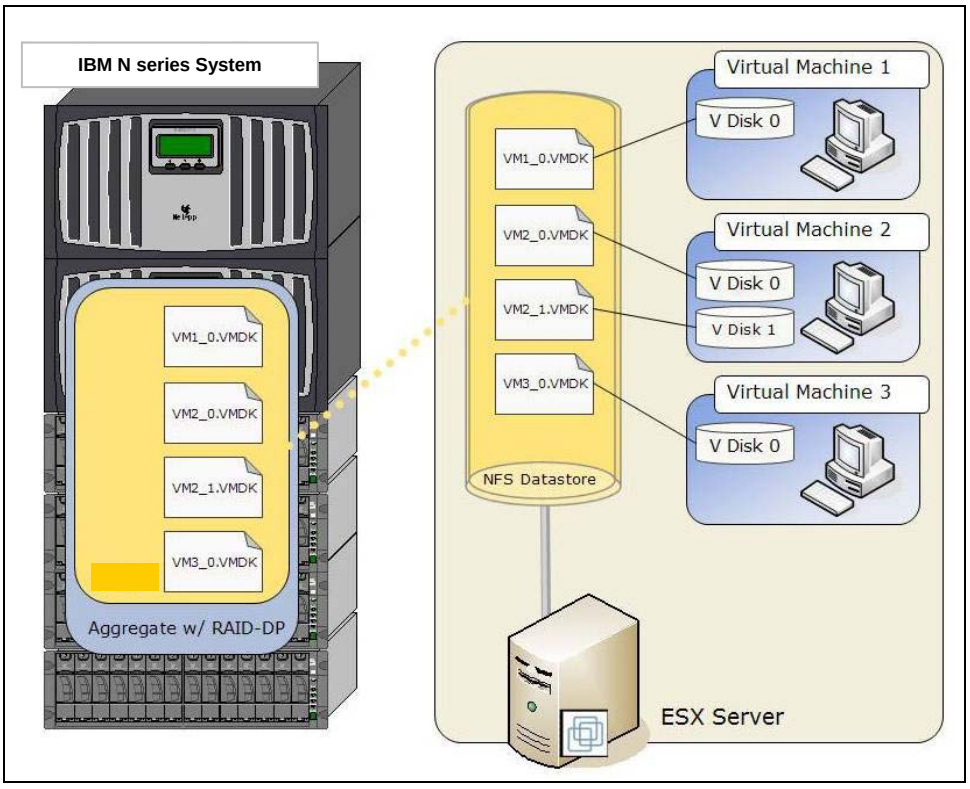


*Figure 8) NFS datastore on IBM N series storage.*

Since a .vmdk is a virtual block device, a .vmdk still needs to be correctly aligned to a network file system (NFS) datastore mounted from an IBM N series NFS share. To correctly align a .vmdk file on an NFS datastore, use the steps listed in “Aligning .vmdks to an IBM N series VMFS datastore.”

## Aligning RDMs to an IBM N series storage system

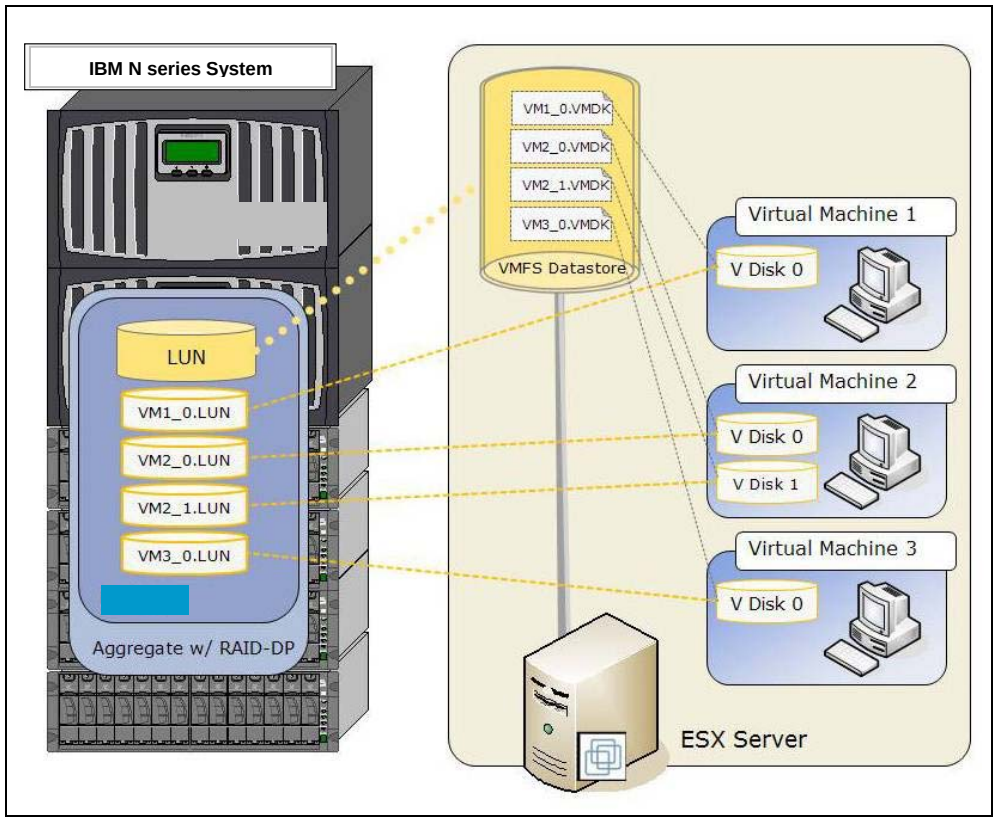


*Figure 9) RDMs on* IBM N series *storage.*

When using raw device mappings (RDMs), the file system will be correctly aligned if the protocol type of the IBM N series LUN is set to the guest OS type. If you are using Windows as a guest OS, set the LUN Protocol Type to Windows on the IBM N series storage controller, and so on. Do not use the VMware LUN Protocol Type for RDMs. No other steps are necessary for proper alignment.

## Trademarks and special notices



References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Network Appliance, the Network Appliance logo, Data ONTAP and WAFL are trademarks or registered trademarks of Network Appliance, Inc., in the U.S. and other countries.

Microsoft, Windows, Windows NT, 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.

Information is provided "AS IS" without warranty of any kind.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.