Lenovo

Analyzing Internal Storage Performance: Simple File Copy vs. Industry Standard Tools

Examines performance results reported by operating system file copy methods and industry standard evaluation tools

Evaluates internal storage bandwidth performance using two different storage form factors

Validates performance findings across operating systems

Explains the need for standardized tools when analyzing internal storage devices

Tristian "Truth" Brown

LENOVO PRESS

Abstract

Too often we search for answers where it is easiest to look, not actually where they can be found. This observational bias is known as the *streetlight* effect. A common example of this bias happens when internal storage devices are evaluated using simple operating system file copy commands instead of industry standard tools.

This paper discusses I/O bandwidth performance results when using operating system file copy methods versus industry standard tools. The goal is to highlight the need for trusted tools when evaluating the maximum performance capabilities of internal storage devices.

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

Introduction

Using operating system (OS)-based file copy tools to evaluate the performance of internal storage device is an unreliable methodology for measuring devices such as hard disk drives (HDDs), solid state drives (SSDs), and flash storage adapters. OS operations such as Linux's dd command and Windows drag & drop are examples of file copy tools that can distort the perceived performance of an internal storage device. This occurs because OS file copy tools do not provide users with the control necessary to accurately stress these devices.

This brief will investigate the achievable performance when an internal storage device workload is driven using an OS file copy command versus industry-standard I/O evaluation tools.

Hardware and methodology

All evaluations were executed on a 4-socket Lenovo System x3850 X6 with the following configuration:

- Four Intel E7-8867 v3 processors with a base frequency of 2.5 GHz
- 128 GB of memory per processor (512 GB total memory) operating at 1600 MHz
- Memory configured to operate in Independent Mode
- One io3 1.6 TB Enterprise Mainstream Flash Adapter, 00YA803
- One 600 GB 15K 12Gbps SAS 2.5" G3HS HDD, 00WG665

The operating systems evaluated were:

- Red Hat Enterprise Linux 6.5
- Windows Server 2012

The x3860 X6 server was configured for high performance so that neither memory nor processor speed would adversely impact either internal storage device's capabilities. The storage device on which both operating systems were installed was isolated from the internal storage devices under test. The performance of the storage device used for the operating systems exceeded that of the evaluated internal storage devices. Similar performance results are expected on other platforms, as long as no memory or processing bottlenecks are introduced.

All evaluations present a direct A-to-B comparison between industry-standard tools and simple OS-based file copy methods. Due to the nature of OS-based file copy tools, sequential read and sequential write I/O traffic was chosen as the equivalent Flexile I/O (fio) tester and Iometer workloads. Each device was then evaluated independently using uniform I/O traffic comprising 512 KB blocksize at a queue depth of 16. I/O activity at the device level was then monitored in one-second intervals utilizing iostat for Linux and Performance Monitor for Windows.

Linux evaluations

For Linux evaluations, fio and the Linux dd command were used to generate an I/O workload to the device under test. During each test, the Linux iostat tool probed I/O traffic at the device to independently measure performance at the operating system level. Throughout the evaluations each device was configured and accessed as raw JBOD storage.

HDD tests

The HDD manufacturer’s bandwidth specification is 250 MB/s for both sequential read and sequential write workloads. Therefore it was expected that an evaluation runtime of 40 seconds would result in 10 GB of transferred data. The fio evaluation tool results listed in Figure 1 display a bandwidth of 252 MB/s for Sequential Reads and a 256 MB/s for Sequential Writes and approximately 10GB of transferred data.

```

========== HDD fio Seq Read Evaluations =============

Seq Read, Block Size = 512K,524288, Queue Depth = 16, Runtime = 40s
/dev/sdb: (g=0): rw=read, bs=512K-512K/512K-512K/512K-512K, ioengine=libaio, iodepth=16
fio-2.1

Run status group 0 (all jobs):
   READ: io=10116MB, aggrb=252.7MB/s, minb=258833KB/s, maxb=258833KB/s, mint=40019msec, maxt=40019msec

Disk stats (read/write):
  sdb: ios=20176/0, merge=0/0, ticks=637486/0, in_queue=637588, util=99.79%


========== HDD fio Seq Write Evaluations =============

Seq Write, Block Size = 512K,524288, Queue Depth = 16, Runtime = 40s
/dev/sdb: (g=0): rw=write, bs=512K-512K/512K-512K/512K-512K, ioengine=libaio, iodepth=16
fio-2.1

Run status group 0 (all jobs):
  WRITE: io=10250MB, aggrb=256.0MB/s, minb=262177KB/s, maxb=262177KB/s, mint=40032msec, maxt=40032msec

Disk stats (read/write):
  sdb: ios=64/20435, merge=0/0, ticks=5/637597, in_queue=637846, util=99.79%

```

Figure 1 Hard Drive Sequential Read and Write performance results reported by fio tool

The fio tool evaluation results were then compared against the performance observed by the iostat monitoring tool. Figure 2 confirmed that the hard drive’s overall bandwidth matched the fio result and consistently performed at the expected bandwidth level for the duration of the evaluation.

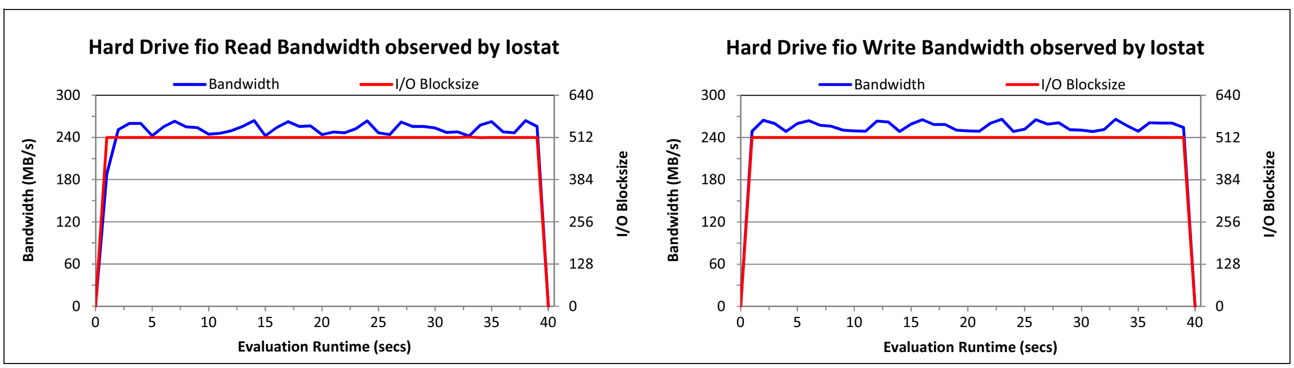

Figure 2 Hard Drive Read (left) and Write (right) performance observed at device by iostat tool

The Linux dd file copy command is not dictated by a runtime; therefore, a 10 GB test file was utilized to determine the device’s bandwidth performance. Using the dd command to match the performance observed with the fio evaluation tool proved to be difficult. The first major issue was that the dd command did not adhere to the specified I/O blocksize of 512 KB. The second issue was that a 250 MB/s transfer rate was achieved but not sustained for the duration of the evaluation.

The dd file copy command results listed in Figure 3 on page 5 displays a bandwidth of 145 MB/s for sequential reads and 210 MB/s for sequential writes. The iostat results in Figure 4 on page 5 show that there was “dead” time in each file transfer. The greater-than-40-second runtime, combined with inconsistent performance throughout the duration of each test, yielded the underperforming bandwidth results.

```
========== HDD dd Seq Read Evaluations =============

File Read from Device Using dd --> Blocksize=512K, Filesize Approx=10GB

[root@localhost Linux_device_logs]# dd if=/dev/sdb of=/home/test_file_10GB.tmp bs=524288 count=19073
19073+0 records in
19073+0 records out
9999745024 bytes (10 GB) copied, 68.858 s, 145 MB/s


========== HDD dd Seq Write Evaluations =============

File Write to Device Using dd --> Blocksize=512K, Filesize Approx=10GB

[root@localhost Linux_device_logs]# dd if=/home/test_file_10GB.tmp of=/dev/sdb bs=524288 count=19073
19073+0 records in
19073+0 records out
9999745024 bytes (10 GB) copied, 47.625 s, 210 MB/s
```

Figure 3 Hard Drive Sequential Read and Write performance results reported by Linux dd Command

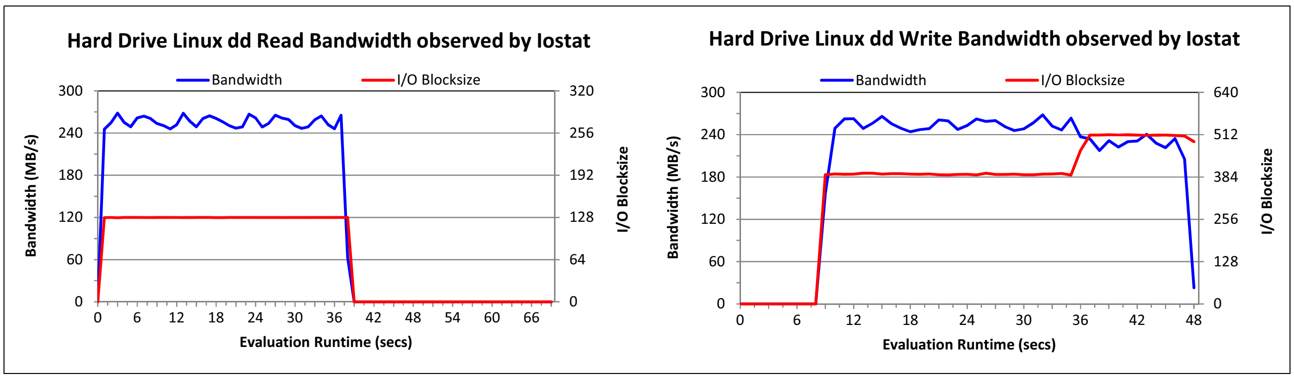

Figure 4 Hard Drive Read (left) and Write (right) performance observed at device by theiostat tool

Flash storage adapter tests

The PCIe Flash storage adapter manufacturer’s bandwidth specification for sequential reads is 2800 MB/s and for sequential writes is 1700 MB/s. Due to the performance capabilities of the device, a 65GB test file was utilized for the dd file copy evaluations. For the fio tests the device’s non-uniform Read/Write bandwidth specification dictated an evaluation runtime of 23 seconds for the sequential read workload and a runtime of 38 seconds for the sequential write workloads. These evaluation runtimes result in approximately 65GB of transferred data.

The fio evaluation tool results listed in Figure 5 on page 6 display a bandwidth of 2841 MB/s for Sequential Reads and 1728 MB/s for Sequential Writes, which results in approximately 65GB of transferred data. Again, the performance observed by iostat displayed in Figure 6 on page 6 align with the fio evaluation tool’s output and confirm that performance remained consistent throughout the evaluation.

```
========== PCIe Flash fio Seq Read Evaluations =============

Seq Read, Block Size = 512K,524288, Queue Depth = 16, Runtime = 23s
/dev/fiob: (g=0): rw=read, bs=512K-512K/512K-512K/512K-512K, ioengine=libaio, iodepth=16
fio-2.1

Run status group 0 (all jobs):
   READ: io=65393MB, aggrb=2841.7MB/s, minb=727315KB/s, maxb=727649KB/s, mint=23011msec, maxt=23012msec

Disk stats (read/write):
  fiob: ios=130153/0, merge=0/0, ticks=1461734/0, in_queue=1460334, util=99.65%


========== PCIe Flash fio Seq Write Evaluations =============

Seq Write, Block Size = 512K,524288, Queue Depth = 16, Runtime = 38s
/dev/fiob: (g=0): rw=write, bs=512K-512K/512K-512K/512K-512K, ioengine=libaio, iodepth=16

Run status group 0 (all jobs):
  WRITE: io=65724MB, aggrb=1728.8MB/s, minb=442437KB/s, maxb=442778KB/s, mint=38016msec, maxt=38018msec

Disk stats (read/write):
  fiob: ios=85/131034, merge=0/0, ticks=6/2420314, in_queue=2415601, util=99.83%

```

Figure 5   PCIe Flash Adapter Sequential Read and Write performance results reported by fio tool

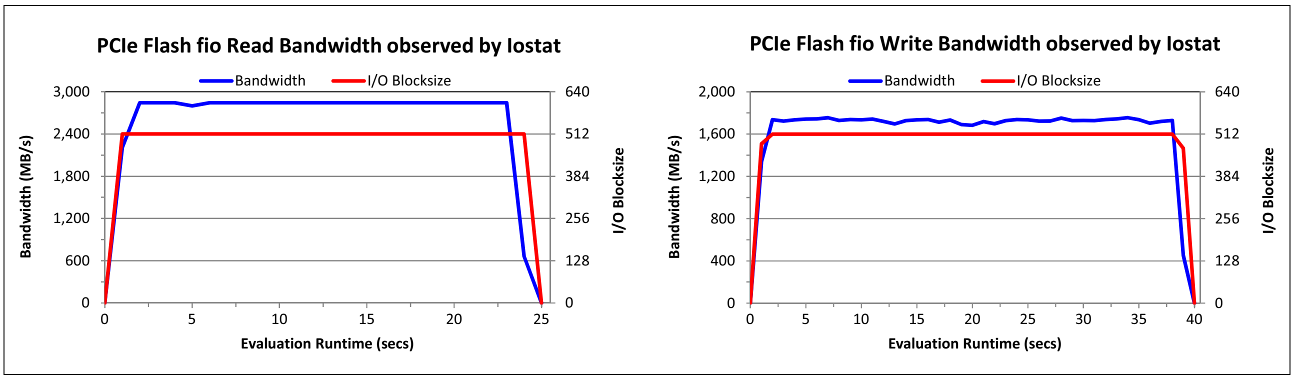

Figure 6   PCIe Flash Adapter Read (left) and Write (right) performance observed at device by iostat tool

As with the HDD, the dd command was not able to match the bandwidth performance level achieved with the fio evaluation tool. The traffic varied at the device and a major deviation from the HDD evaluation was that maximum sequential *read* bandwidth performance was never achieved. This can be caused by many factors but is mostly due to the 128KB read block size chosen by the dd command. It is clear that the dd command did not provide enough traffic to saturate the device and demonstrate maximum sequential read performance.

Conversely, the dd command was able to drive enough sequential *write* traffic to saturate the device for a portion of the evaluation but not consistently for the duration. The 838 MB/s was a result of no data being written to the PCIe Flash device for approximately 40 seconds of the 77 seconds of evaluation time. A deep analysis of how Linux executes the dd file copy command is needed to understand this occurrence but that is beyond the scope of this brief.

Lenovo insight: As demonstrated in Figure 7 and Figure 8 below, without monitoring I/O traffic at the device level it is very easy to assume inaccurate internal storage device performance based on the results provided by the Linux dd command.

```

======= PCIe Flash dd Seq Read Evaluations =============

File Read from Device Using dd --> Blocksize=512K, Filesize Approx=65GB

[root@localhost Linux_device_logs]# dd if=/dev/fiob of=/home/test_file_65GB.tmp bs=524288 count=123977
123977+0 records in
123977+0 records out
64999653376 bytes (65 GB) copied, 126.245 s, 515 MB/s


======= PCIe Flash dd Seq Write Evaluations =============

File Write to Device Using dd --> Blocksize=512K, Filesize Approx=65GB

[root@localhost Linux_device_logs]# dd if=/home/test_file_65GB.tmp of=/dev/fiob bs=524288 count=123977
123977+0 records in
123977+0 records out
64999653376 bytes (65 GB) copied, 77.5684 s, 838 MB/s

```

Figure 7 Flash Adapter Sequential Read and Write performance results reported by the dd command

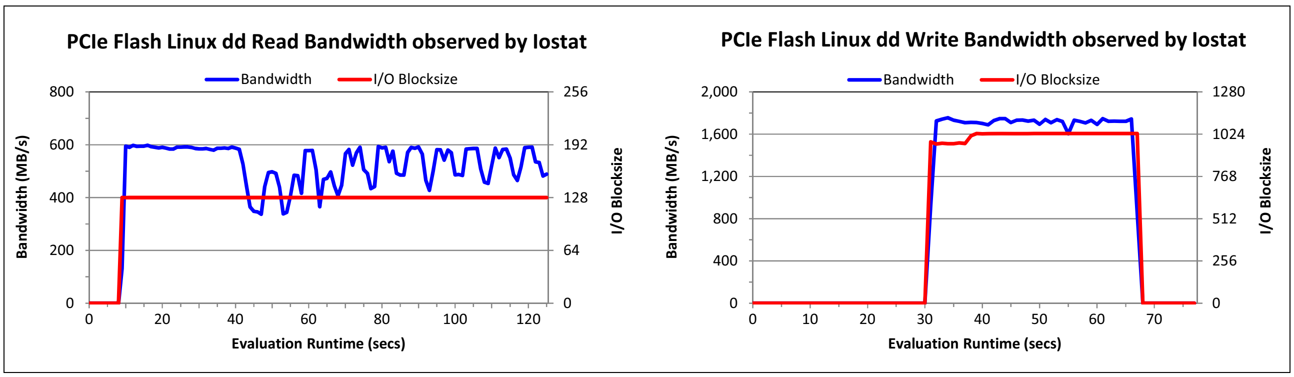

Figure 8 PCIe Flash Adapter Read (left) and Write (right) performance observed at device by iostat tool

Windows evaluations

For Windows evaluations, the evaluation tool Iometer was used to generate an I/O workload while the Performance Monitor tool probed I/O traffic at the device. The exact workload specifications were carried over from the Linux evaluations.

Due to the nature of Window's file copy tools, both devices were configured using default NTFS partitions for all evaluations.

Command line-based file copy commands, such as copy and xcopy, don't display performance metrics; therefore, they were excluded from consideration. Experimentation displayed an insignificant performance difference between Windows' built-in drag & drop, copy & paste, and cut & paste file options. Therefore Windows cut & paste instruction was chosen as the method for file transfer.

To investigate the sequential read bandwidth, a test file was transferred from the device under test to the Windows Desktop. For the sequential write bandwidth, a test file was transferred to the device under test from the Windows Desktop.

HDD tests

For Iometer Read and Write evaluations, a runtime of 40 seconds was used to simulate approximately 10 GB of data being transferred to and from the device. The HDD Iometer evaluation results listed in Figure 9 display sequential read bandwidth of 243 MB/s and Figure 10 displays a sequential write bandwidth of 242 MB/s. The small delta in performance with respect to the fio Linux results was expected due to a combination of factors such as the NTFS partition overhead, Windows operating system, or test evaluation tolerance. Despite the lower performance, Iometer was still able to achieve within 3% of the manufacture's specified performance capabilities of the device.

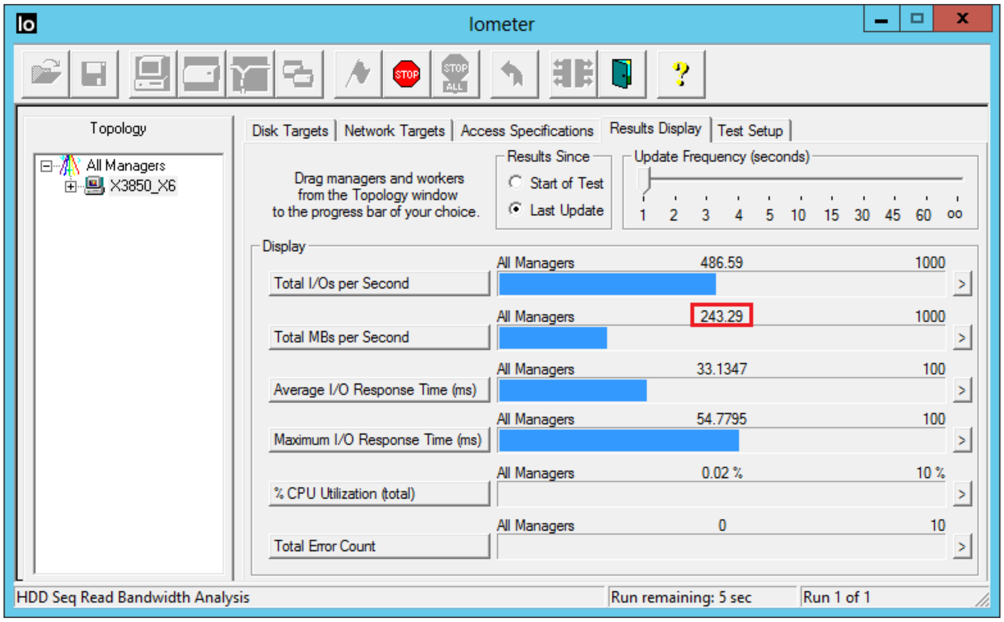

Figure 9 Hard Drive Sequential Read performance results measured by Iometer tool

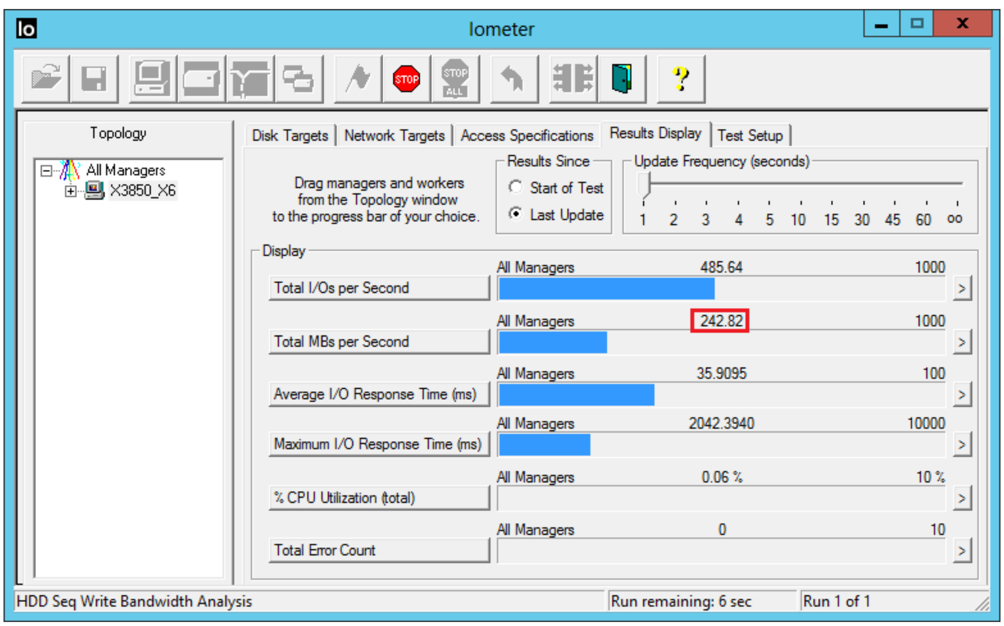

Figure 10 Hard Drive Sequential Write performance results measured by Iometer tool

Figure 11 on page 9 displays the performance monitor tool statistics, which confirm the Iometer results and display a consistent level of performance for the duration of the evaluation.

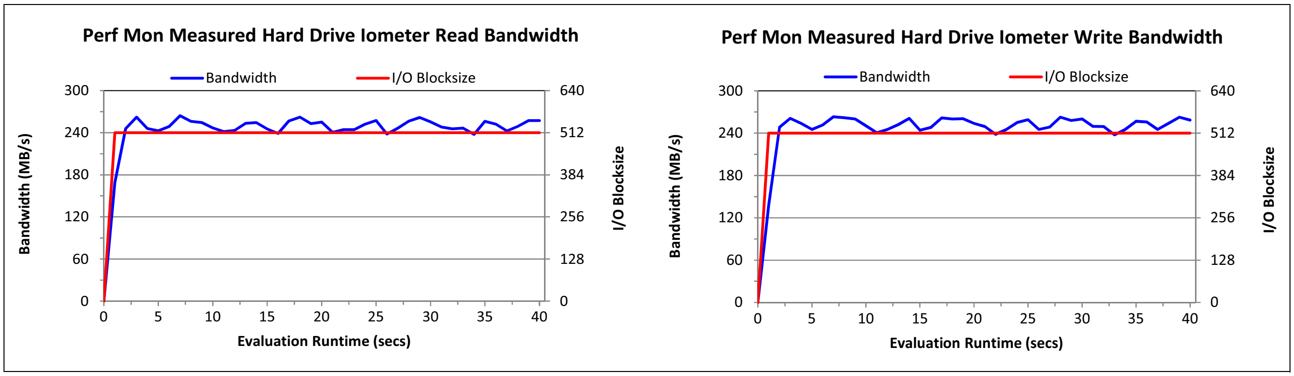

Figure 11 Hard Drive Read (left) and Write (right) performance observed at device by Performance Monitor Tool

Originally, a test file of 10 GB was used to investigate drive bandwidth performance. But due to the short transfer duration and the throughput reported at the file copy progress window, it was determined that the 65 GB test file was better suited for bandwidth analysis.

The file copy progress window consistently displayed a 5.6x increase in bandwidth performance resulting in an approximately 1.4 GB/s throughput rate for read transfers (Figure 12) as well as write transfers (Figure 13). This level of increased performance is not feasible given the hardware limitations and device throughput specification of 250 MB/s. Therefore the performance reported at the progress window was determined to be inaccurate.

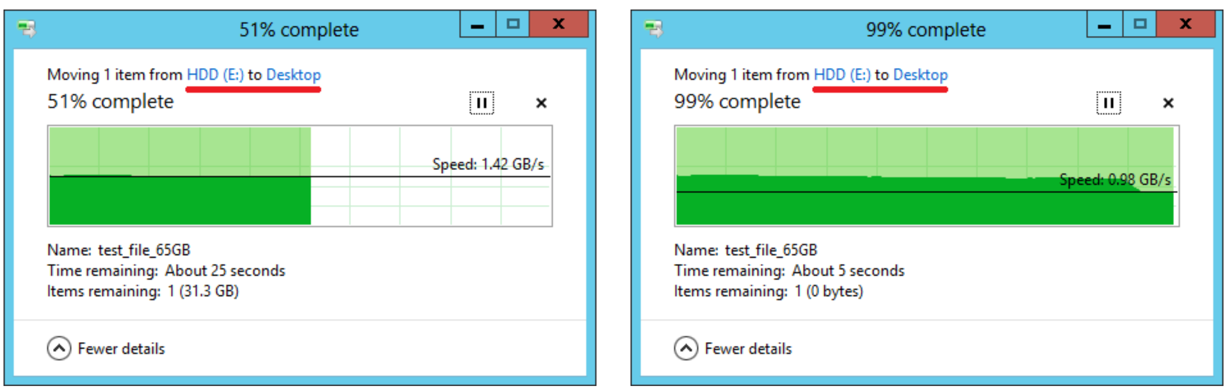

Figure 12 Hard Drive Sequential Read bandwidth performance displayed by progress window

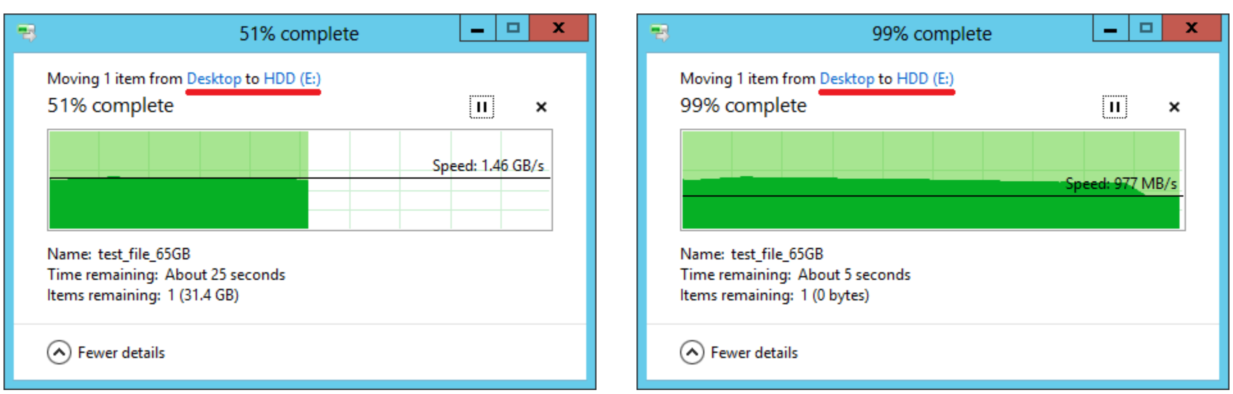

Figure 13 Hard Drive Sequential Write bandwidth performance displayed by progress window

Lenovo insight: The performance displayed by in the proceeding progress windows (Figure 12 and Figure 13) vastly exceeds not only the manufacture's bandwidth specification but also the current capabilities of hard disk drive technology. These numbers are clearly inaccurate!

For the sequential read evaluation (left chart in Figure 14), the Performance Monitor tool recorded unusual behavior at the device level. Despite many attempts, Performance Monitor was unable to observe any meaningful read statistics when transferring the test file from the HDD to the desktop. Although Figure 14 shows a negligible amount of read traffic, the test file was verified at the completion of the transfer operation. An in-depth investigation beyond the scope of this brief is needed to fully characterize how the file copy transfer command is executed in Windows.

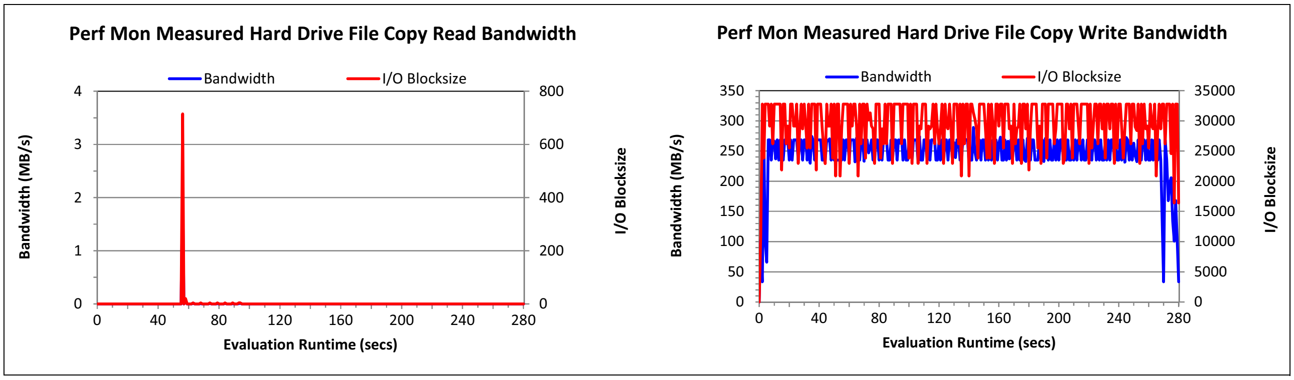

Figure 14 Hard Drive Read (left) and Write (right) performance observed at device by Performance Monitor Tool

For the sequential write evaluation (right chart in Figure 14), the Performance Monitor tool was able to record transfer statistics. The duration of the file transfer operation did not align with the reported 1.4GB/s throughput displayed at the progress window. Given the specification transfer rate of 250 MB/s, the 65 GB test file should have completed within roughly 260 seconds, but the Performance Monitor tool registered write traffic for a much longer duration. These captured statistics confirm that the file copy operation performance displayed at the progress window does not directly correlate with the data being transferred at the device level.

Flash Storage Adapter tests

The PCIe Flash adapter Iometer evaluations achieved a slightly lower Sequential Read throughput than what was reported in Linux. This was expected due to the overhead associated with configuring the NTFS partition, along with the change in OS to Windows. Factors such as increased workers, a queue depth greater than 16, or larger I/O blocksize could yield a higher maximum bandwidth throughput. None of these techniques were implemented so fio's Linux evaluations and Iometer's Windows evaluations would remain uniform.

For the Iometer PCIe Flash Adapter evaluations a runtime of 23 seconds was utilized for the sequential read workload and a runtime of 38 seconds for the sequential write workload. The PCIe Flash device Iometer evaluation results listed in Figure 15 on page 11 displays a sequential read bandwidth of 2357 MB/s and Figure 16 on page 11 displays a sequential write bandwidth of 1728 MB/s.

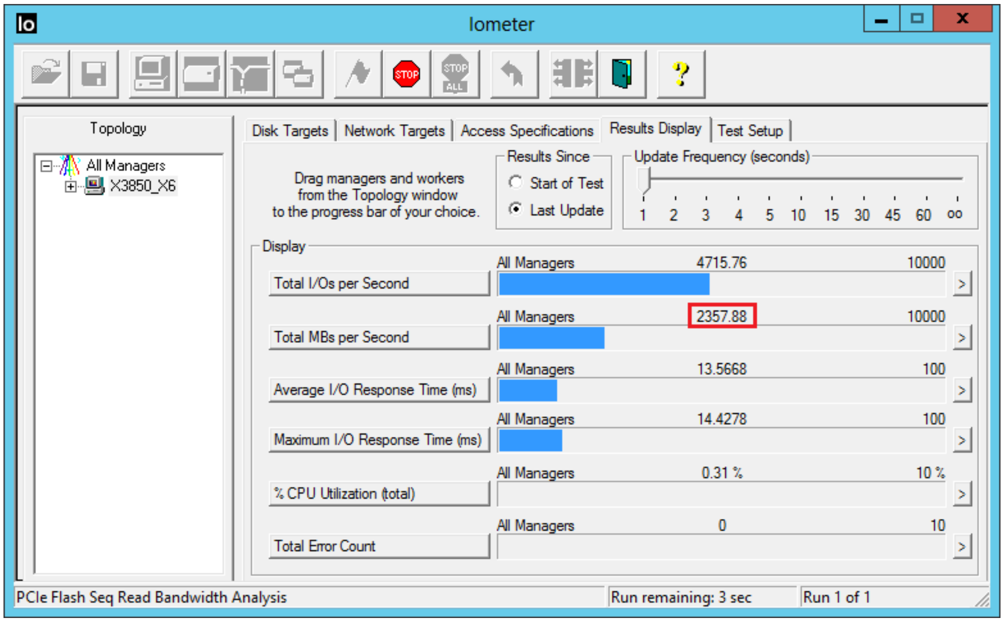

Figure 15 PCIe Flash Adapter Sequential Read performance results measured by Iometer tool

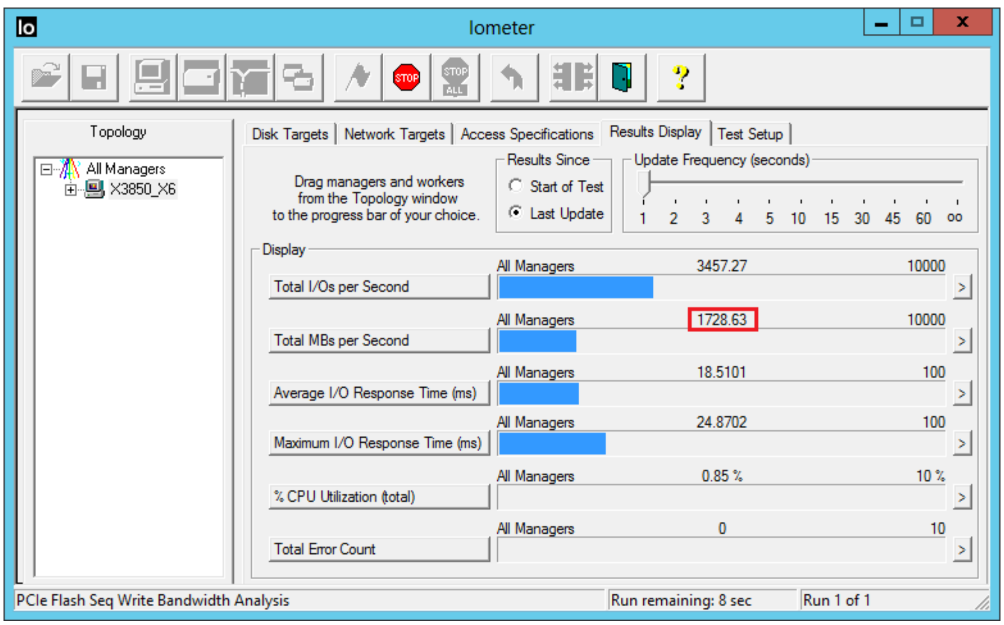

Figure 16 PCIe Flash Adapter Sequential Write performance results measured by Iometer tool

Figure 17 on page 12 confirms that the PCIe flash device’s overall bandwidth matched the Iometer result and consistently performs at the expected bandwidth level throughout the duration of the evaluation.

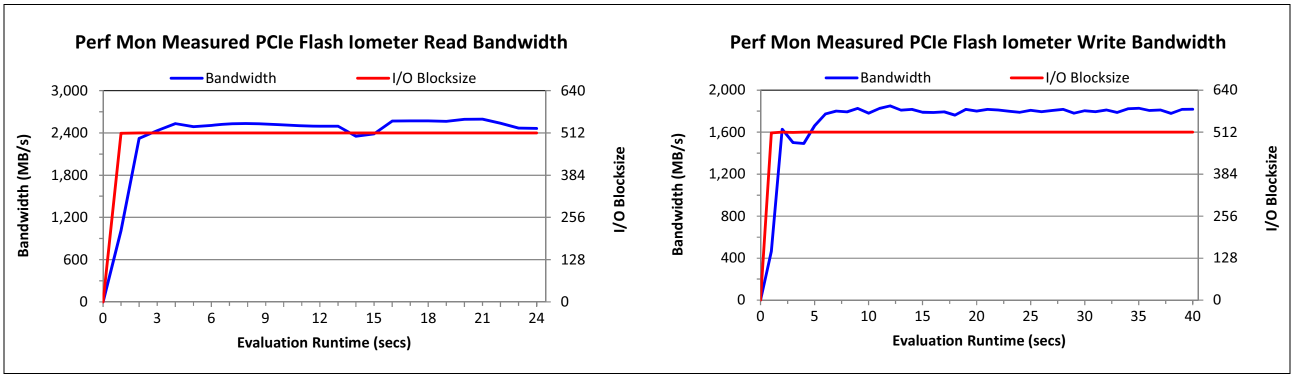

Figure 17 PCIe Flash Adapter Read (left) and Write (right) performance observed at device by Performance Monitor Tool

To measure the PCIe Flash device's file copy bandwidth performance, a 65 GB test file was used. For sequential read evaluation, the progress window displayed a throughput performance of roughly 1.4 GB/s and for the sequential write evaluation the progress displayed roughly 1.5 GB/s.

Unlike the HDD, these results fall within the capabilities of the device; however, they do not meet the performance expectation. The sequential read performance degraded by roughly 40% and the sequential write performance by more than 10%. The performance reported at the progress windows did not align with the devices' known capabilities.

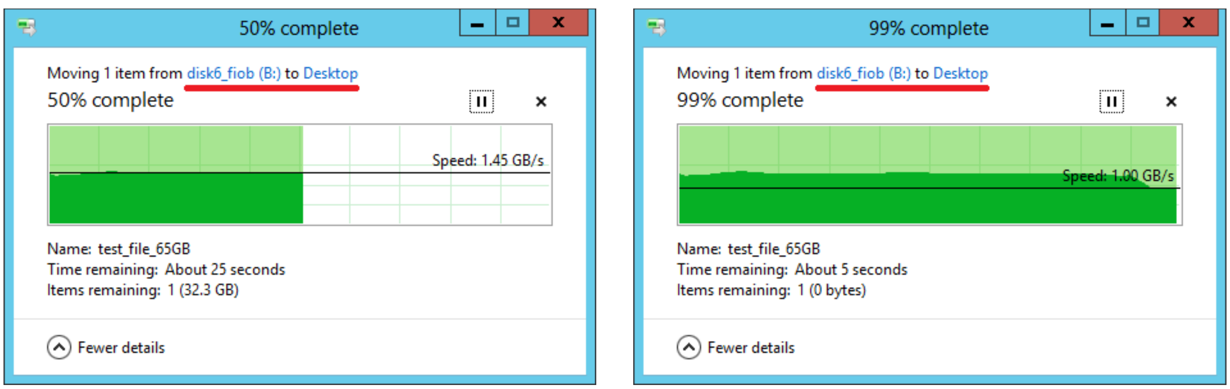

Figure 18 PCIe Flash Adapter Sequential Read bandwidth performance displayed by progress window

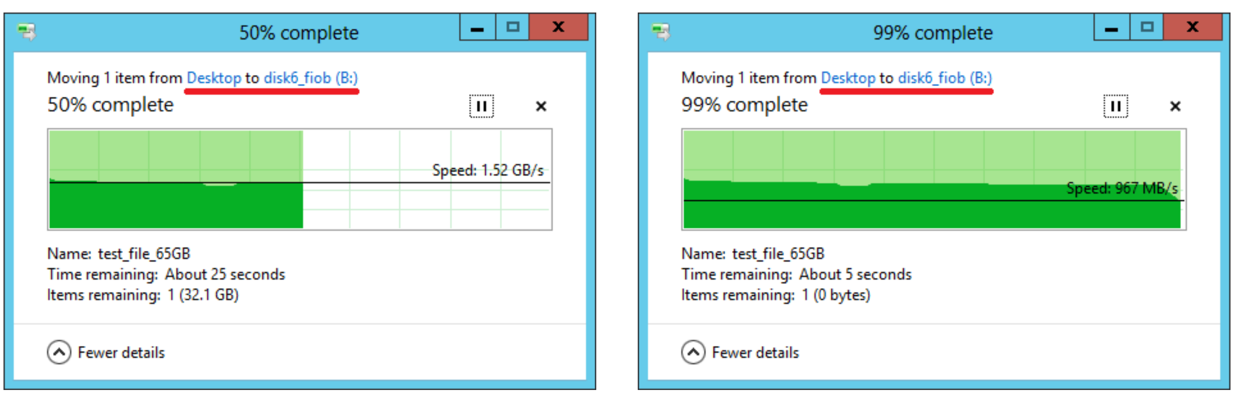

Figure 19 PCIe Flash Adapter Sequential Write bandwidth performance displayed by progress window

Lenovo insight: The manufactures specification states asymmetrical Sequential Read/Write performance, yet the progress windows display roughly a 5% difference between the Read and Write file transfers.

Again the Performance Monitor tool did not observe any meaningful read statistics when transferring the test file from the PCIe Flash device to the desktop, as shown in the left chart in Figure 20. The recorded data shows that the file copy operation at the device level did not mirror the actual I/O transfer at the progress window.

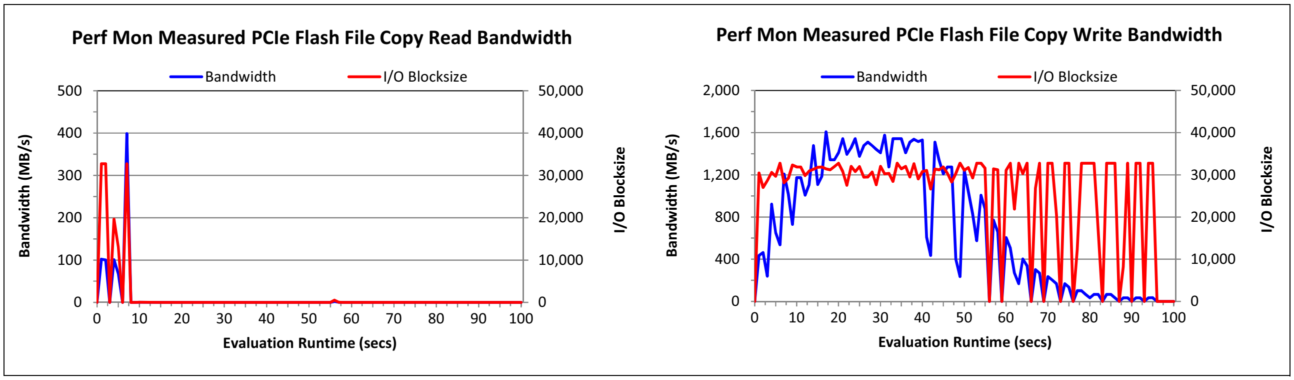

Figure 20 PCIe Flash Adapter Read (left) and Write (right) performance observed at device by Performance Monitor Tool

For the sequential write evaluation (right side of Figure 20), the Performance Monitor tool was able to observe the file transfer statistics. As identified in the Hard Drive Sequential write evaluations, the duration of the file transfer operation did not align with the reported 1.5GB/s throughput displayed at the progress window.

Conclusion

When measuring the performance of an internal storage device, a trusted evaluation tool is a must. In this brief, Iometer and fio were chosen because they are both open source and widely utilized in the industry. While it might seem more convenient to use file copy commands for quick performance evaluations, both Iostat and Performance Monitor confirmed that the performance reported by these methods should not be trusted.

If deviating from known industry tools is unavoidable, the end user should first confirm that the internal storage device is in fact receiving the correct level of traffic. Specifically, you want to make sure the I/O workload, blocksize, and queue depth is consistent over the duration of the evaluation.

Finally, any tool used for performance investigations must be able to drive enough work to the device such that a saturation point can be achieved for the duration of an evaluation.

Author

Tristian "Truth" Brown is a Hardware Performance Engineer on the Lenovo Server Performance Team in Raleigh, NC. He is responsible for the hardware analysis of high-performance, flash-based storage solutions for System x servers. Truth earned a bachelor's degree in Electrical Engineer from Tennessee State University and a master's degree in Electrical Engineering from North Carolina State University. His focus areas were in Computer Architecture and System-on-Chip (SoC) microprocessor design and validation.

Thanks to the following people for their contributions to this project:

- David Watts, Lenovo Press

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.

This document was created or updated on June 14, 2016.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp0523

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)®	System x®

The following terms are trademarks of other companies:

Intel, and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

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

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.