Lenovo

Analyzing the Performance of Intel Optane DC Persistent Memory in App Direct Mode in Lenovo ThinkSystem Servers

Introduces DCPMM App Direct Mode

Explores performance capabilities of persistent memory

Establishes performance expectations for given workloads

Discusses configurations for optimal DCPMM performance

Tristian “Truth” Brown

Travis Liao

Jamie Chou

LENOVO PRESS

Abstract

Intel Optane DC Persistent Memory is the latest memory technology for Lenovo ThinkSystem servers. This technology deviates from contemporary flash storage offerings and utilizes the ground-breaking 3D XPoint non-volatile memory technology to deliver a new level of versatile performance in a compact memory module form factor.

This paper focuses on the low-level hardware performance capabilities of Intel Optane DC Persistent Memory configured in App Direct Mode operation. This paper provides the reader with an understanding of workloads to produce the highest level of performance with this innovative technology.

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

There is a large performance gap between DRAM memory technology and the highest performing block storage devices currently available in the form of solid-state drives. Capitalizing on this opportunity, Lenovo® partnered with Intel, a key technology vendor, to provide the end customer with a novel memory module solution called Intel Optane DC Persistent Memory.

Intel Optane DC Persistent Memory provides unique levels of performance and versatility because it is backed by Intel 3D XPoint non-volatile memory technology instead of traditional NAND based flash. This technology has various implementations, however, this paper will focus solely on the performance of Intel Optane DC Persistent Memory when run in App Direct Mode operation.

Intel Optane DC Persistent Memory

Intel Optane DC Persistent Memory and it's implementation, the DC Persistent Memory module (DCPMM) is a byte addressable cache coherent memory module device that exists on the DDR4 memory bus and permits Load/Store accesses without page caching.

DCPMM creates a new memory tier between DDR4 DRAM memory modules and traditional block storage devices. This permits DCPMM devices to offer memory bus levels of performance, and allows application vendors to remove the need for paging, context switching, interrupts and background kernel code running.

DCPMMs can operate in three different configurations, Memory Mode, App Direct Mode, and Storage over App Direct Mode. This paper will focus on DCPMM devices operating in App Direct Mode and will analyze the performance of a system with DCPMMs running in this mode.

Figure 1 on page 3 shows the visual differences between a DCPMM and a DDR4 RDIMM. DCPMM devices physically resemble DRAM modules because both are designed to operate on the DDR4 memory bus. The uniquely identifying characteristic of a DCPMMs is the heat spreader that covers the additional chipset.

Figure 1 DCPMM (top) and a DDR4 RDIMM (bottom)

DCPMM modules can operate up to a maximum DDR4 bus speed of 2666MHz and are offered in capacities of 128GB, 256GB, and 512GB. The 128GB DCPMM devices can operate up to a maximum power rating of 15W whereas the 256GB and 512GB DCPMM devices can operate up to a maximum power rating of 18W.

Due to the calculation method and needed overhead for DCPMM device operation the actual usable capacity is slightly less than the advertised device capacity. Table 1 lists the expected DCPMM capacity differences as seen by the operating system.

Table 1 DCPMM advertised capacity relative to usable capacity in operating systems

Advertised DCPMM Capacity	Available DCPMM Capacity
128 GB	125 GB
256 GB	250 GB
512 GB	501 GB

DCPMM App Direct Mode configuration rules

The basic rules for installing DCPMM into a system are as follows:

- A maximum of 1x DCPMM device is allowed per memory channel
- DCPMM devices of varying capacity cannot be mixed with a system
- For each memory channel, DCPMM devices should be installed in the memory slot physically closest to the CPU unless it is the only DIMM in the memory channel

Figure 2 on page 5 shows a close-up of the SR950 system board, showing one second-generation Intel Xeon Scalable Processor with six DCPMMs and six DIMMs installed into the memory slots connected to the processor. The processor has two memory controllers, each providing three memory channels and each memory channel containing two DIMM slots.

As shown, the twelve modules installed are comprised of six RDIMMs and six DCPMM devices, with each DCPMM located in the memory slot electrically (and physically) closer to the processor for each memory channel.

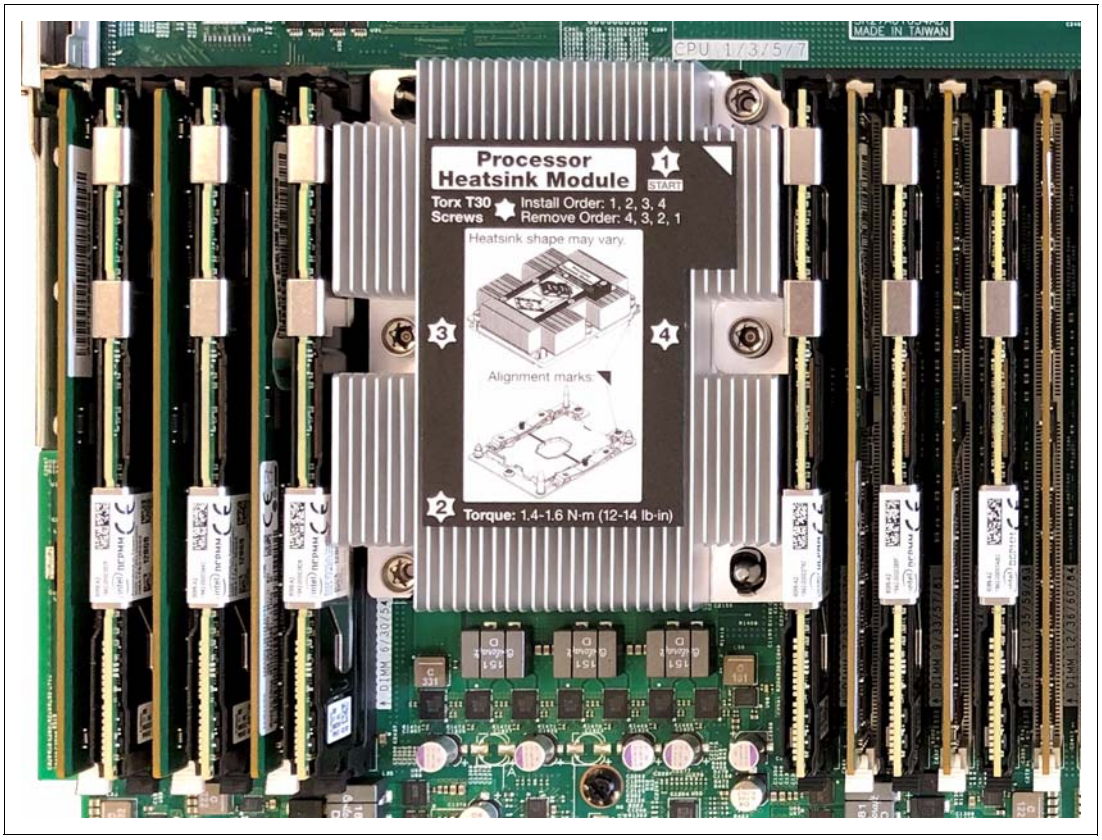

Figure 2 Intel Xeon Scalable Processor with 6 DCPMMs and 6 RDIMMs (SR950)

App Direct Mode

App Direct Mode operation is when a DCPMM is configured as a direct accessible storage module residing on the memory bus. From a software perspective, DCPMM exists between ultra-fast DRAM system memory and conventional block storage devices such as solid-state drives (SSDs). This allows for applications to directly access DCPMM devices at fairly high bandwidth levels and much lower latencies than can be seen with the highest performing enterprise SSDs. App Direct Mode also permits application data to remain available on the memory bus even after a system reboot or a power cycle (persistence).

The key DCPMM App Direct Mode requirement is that application vendors modify their software to fully function and utilize the inherent performance capabilities of App Direct Mode. When properly optimized, DCPMM devices can provide up to multiple tens of GB/s of throughput at nanosecond latencies versus single digit GB/s of throughput at microsecond latencies seen with high performing NAND-based flash storage devices.

If an application hasn't been modified to support App Direct Mode, it can utilize DCPMM in Storage over App Direct Mode operation which is a more conventional setup using a supported DAX model in the operating system (OS). The performance associated with this configuration is discussed in the Lenovo Press paper, *Analyzing the Performance of Intel Optane DC Persistent Memory in Storage over App Direct Mode in Lenovo ThinkSystem™ Servers*, available at https://lenovopress.com/LP1085.

App Direct Mode Performance Analysis

This section describes the results of our analysis of performance of App Direct Mode.

Hardware configuration evaluation environment

Intel Memory Latency Checker (MLC) was used to quantify this innovative technology because it is a well-established industry performance evaluation tool. MLC is fully compatible with DCPMM App Direct Mode operation and has the ability to stress and measure performance and latency at the memory bus level.

The App Direct Mode operation performance data established in this paper is based on a single-socket system configured with the details provided in Table 2. DCPMM support rules require mirrored hardware configurations across sockets therefore the results shown in the following charts are representative of each individual socket of a multi-socket sever.

Table 2 System configurations for single-socket DCPMM App Direct Mode performance evaluations

Configurations	**2-2-2**	**2-2-1**	**2-1-1**
Processor	1x Intel Xeon Platinum 8280L	1x Intel Xeon Platinum 8280L	1x Intel Xeon Platinum 8280L
Operating system	SLES12 SP4	SLES12 SP4	SLES12 SP4
Memory	6x 32GB RDIMM, 2933 MHz	6x 32GB RDIMM, 2933 MHz	6x 32GB RDIMM, 2933 MHz
DCPMM	6x DCPMM, 2666 MHz ► 128GB ► 256GB ► 512GB	4x DCPMM, 2666 MHz ► 128GB ► 256GB ► 512GB	2x DCPMM, 2666 MHz ► 128GB ► 256GB ► 512GB

App Direct scaling analysis and maximum bandwidth performance

The DCPMM App Direct Mode configurations listed in Table 2 dictate that each CPU socket will contain 6x DD4 DRAM memory modules for 2x, 4x, or 6x DCPMM device populations.

Figure 3 on page 7 shows the expected maximum bandwidth available when an application is designed to utilize DCPMMs in App Direct Mode. DCPMMs are optimized for warm data sets therefore performance benefits are greater for reads operations versus write operations. Write (NT) refers to non-temporal writes which are write operations that aren't cached within the processor.

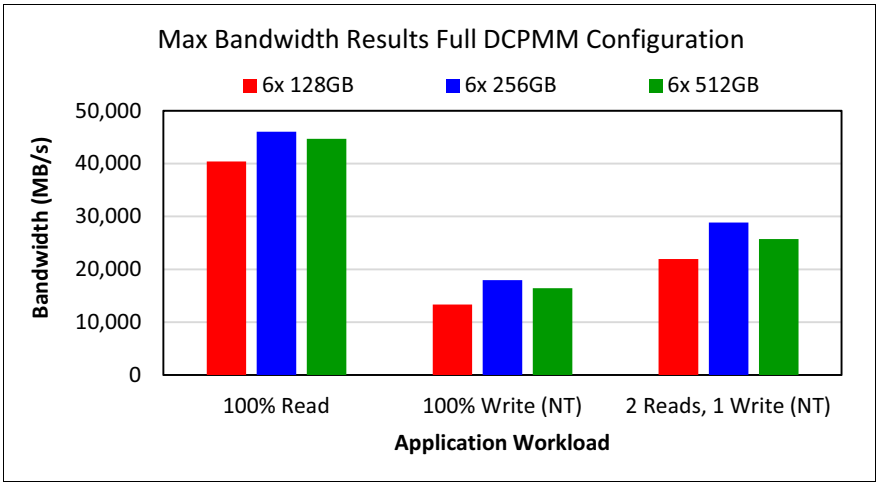

Figure 3 Maximum Socket Bandwidth for 6x DCPMM configuration

As shown in Figure 3, a read-intensive application workload could see up to over 40 GB/s per socket where as a write intensive workload could see roughly 15 GB/s per socket for a fully populated DCPMM socket. Under a more real-world 2 reads / 1 write (NT) workload, performance can vary but typically can fall within the 20~30 GB/s per socket throughput range. To put in this context, the maximum capable bandwidth a PCIe 3.0 x8 device is just under 8 GB/s.

Figure 4 demonstrates the near linear bandwidth performance when scaling from 2x DCPMMs per socket up to 6x DCPMMs per socket for 256GB DCPMM capacity. Figure 3 demonstrates the relatively similar maximum throughput performance between the DCPMM capacity offerings.

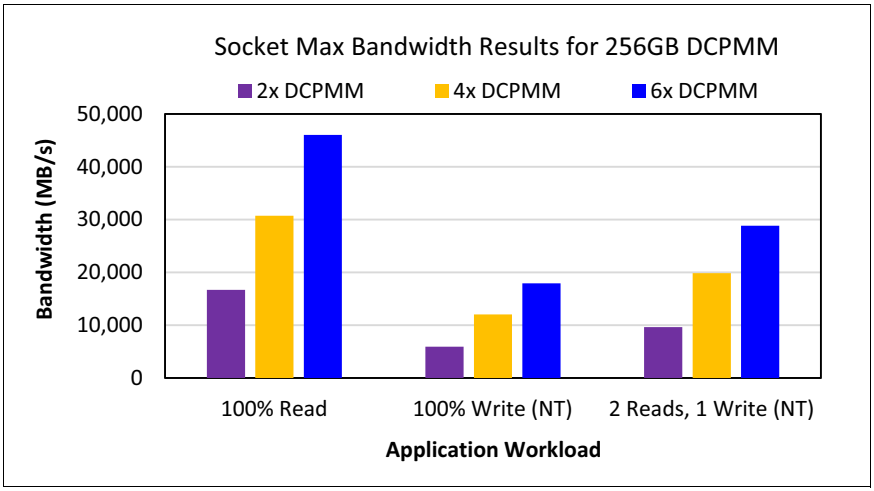

Figure 4 Maximum Scaling Socket Bandwidth for 256GB DCPMM Configuration

DCPMM bandwidth and loaded latency analysis

Figure 5 displays DCPMM measured bandwidth at varying loaded latency values for three hardware configurations, 2-2-2, 2-2-1, and 2-1-1, as listed in Table 2 on page 6. The most optimal configuration is a fully populated socket with 6x DCPMM devices for any given capacity (2-2-2).

Applications that favor sequential workloads will see the highest bandwidth coupled with the lowest latency prior to device saturation. Applications that exhibit very random workloads will experience lower bandwidth at slightly higher latencies.

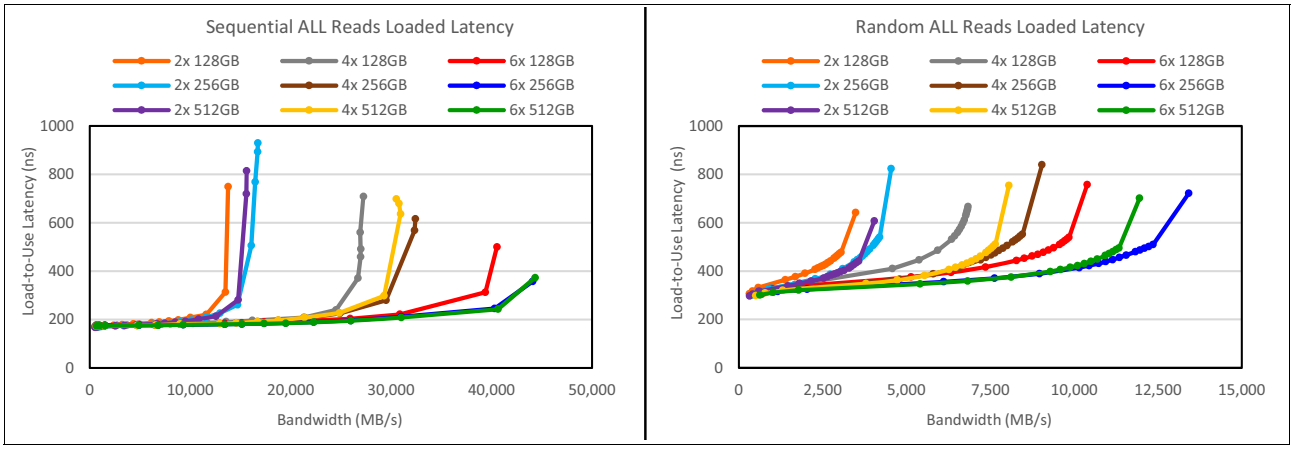

Figure 5 DCPMM Sequential (left) and Random (right) All Reads Loaded Latency Performance

In Figure 6, the MLC evaluation was modified to simulate 2 reads / 1 write (NT) to better align with real world scenarios. DCPMM devices provide higher levels of throughput at lower latencies for sequential operations versus random operations. Overall, workloads that are more read intensive and sequential in nature will see the best performance.

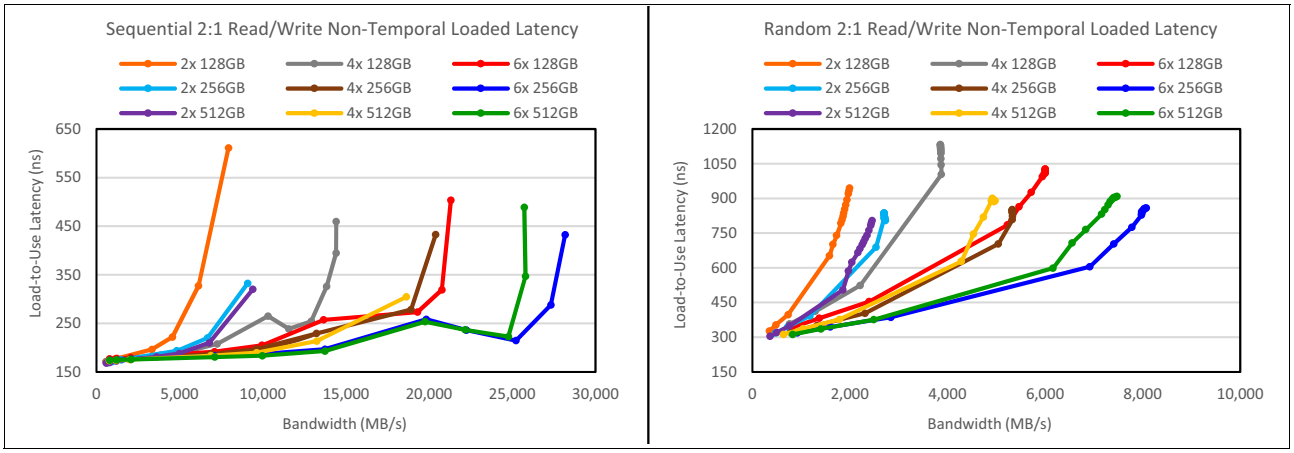

Figure 6 DCPMM Sequential (left) and Random (right) Read/Write (NT) Loaded Latency Performance

DCPMM performance relative to DDR4 system memory

This section establishes DCPMM device performance in reference to DDR4 DRAM system memory. We analyzed performance using 32GB DDR4 2933MHz RDIMM memory modules. The DRAM modules were clocked to 2666 MHz to align with the maximum operating bus speed seen when DCPMM devices are installed in a system.

The displayed DCPMM performance is representative of a 6x DCPMM per socket configuration and the DRAM performance is representative of 6x RDIMMs per socket (a 2-2-2 configuration).

Figure 7 and Figure 8 display per socket performance gap between DRAM system memory and DCPMM devices operating in App Direct Mode. This result clearly demonstrates that the DCPMM operation goal is for warm data that favors sequential workloads. DRAM system

memory is not as sensitive to workload characteristics; therefore, performance is relatively equal between the workloads for up to roughly 75 GB/s per socket bandwidth.

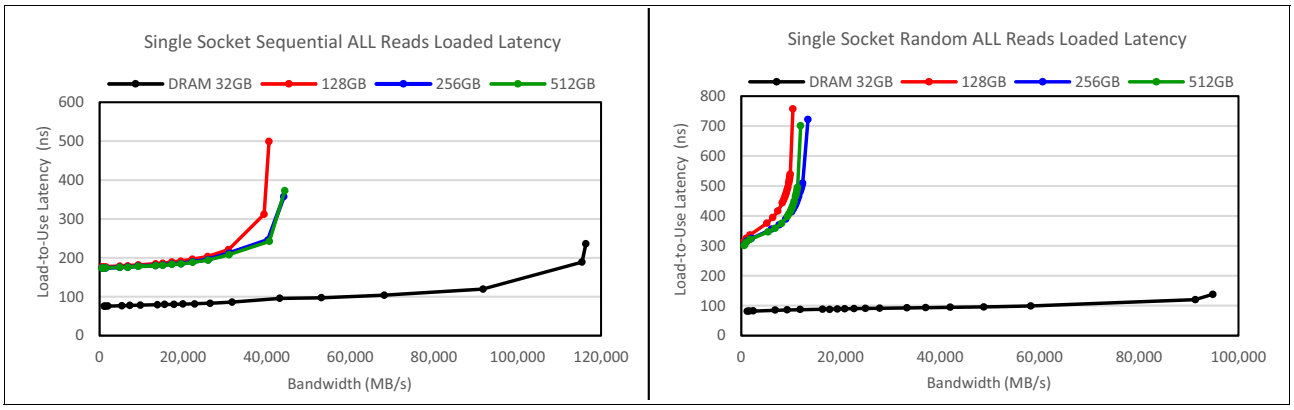

Figure 7 Single Socket DRAM vs. DCPMM Sequential (left) and Random (right) All Reads Loaded Latency Performance

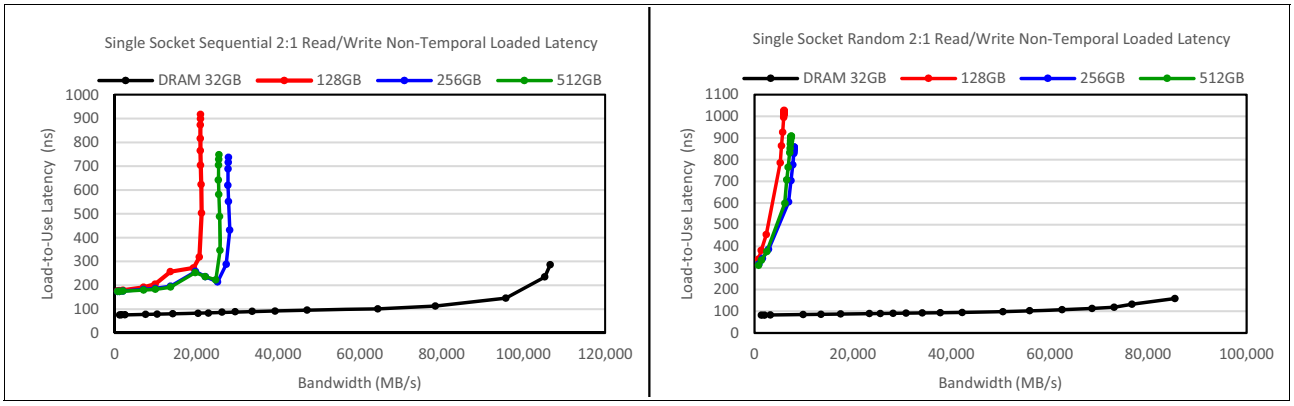

Figure 8 Single Socket DRAM vs. DCPMM Sequential (left) and Random (right) Read/Write (NT) Loaded Latency Performance

Conclusion

Intel Optane DC Persistent Memory offerings in Lenovo ThinkSystem servers are focused on providing high performing supplemental technology to ultra-fast DRAM system memory. The only requirement is that application vendors modify their software to fully function and take advantage of DCPMM device in App Direct Mode operation.

In short, when an application utilizes DCPMM devices under optimal performance conditions the end result is a new level of performance historically not available with persistent technology.

About the authors

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 Data Center Group. 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 are in Computer Architecture and System-on-Chip (SoC) microprocessor design and validation.

Travis Liao is a Hardware Performance Engineer in the Lenovo Data Center Group Performance Laboratory based in Taipei. His focus is modelling and validating performance of server storage subsystem including RAID controllers, SSDs and software RAID. Travis holds a Master's Degree in Electronic Engineering from National Taiwan University in Taiwan.

Jamie Chou is an Advisory Engineer in the Lenovo Data Center Group Performance Laboratory in Taipei Taiwan. Jamie joined Lenovo in November 2014. Prior to working on server performance, he worked on system software development, automation, and Android system performance. Jamie received a Master’s Degree and a PhD from the department of Computer Science and Information Engineering, Tamkang University, Taiwan.

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

- David Watts, Lenovo Press
- Lenovo RDC Performance Team

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 April 2, 2019.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp1083

Trademarks