Lenovo

Using VMD Direct Assign in Linux KVM on Lenovo ThinkSystem Servers

Introduces the features of VMD Direct Assign

Explains which Lenovo servers support the feature

Provides instructions on how to configure VMD Direct Assign in Linux KVM

Describes how to verify operation of VMD Direct Assign in a virtual machine

Kelvin Shieh

Alpus Chen

Abstract

Intel VMD Direct Assign is a specific application of Intel Volume Management Device (Intel VMD) that aims to improve the storage experience in Hyperconverged Infrastructure (HCI) architecture, it takes the isolated storage subsystem and bypasses the hypervisor to assign Intel VMD control directly to a virtual machine running on the host.

This document provides a brief overview of the Intel VMD Direct Assign and describes how to configure and use VMD Direct Assign in Linux KVM on Lenovo® ThinkSystem™ servers. This document is intended for IT specialists and IT managers who are familiar with Linux KVM Virtualization.

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 3
How to use VMD Direct Assign in Linux KVM 5
Verify VMD Device in Guest OS 14
Resources 15
Authors 15
Notices 16
Trademarks 17

Introduction

Intel Volume Management Device (Intel VMD) is a feature introduced with the Intel Xeon Scalable processor family, which is used to help manage NVMe drives and provides features such as accidental hot plug, LED management, error isolation and bootable RAID for NVMe drives directly connected to the processor or PCH PCIe lanes. It allows direct control and management from the PCIe bus without additional hardware adaptors.

VMD maps the entire PCIe subtree to its own address space, enabling it to set up and manage the domain controlled by the VMD driver and enabled in UEFI. As Figure 1 shows, each root port consists of a set of x16 PCIe lanes that create a single Intel VMD Domain. The root port is like an integrated HBA, which provides a solid foundation for supporting the NVMe ecosystem.

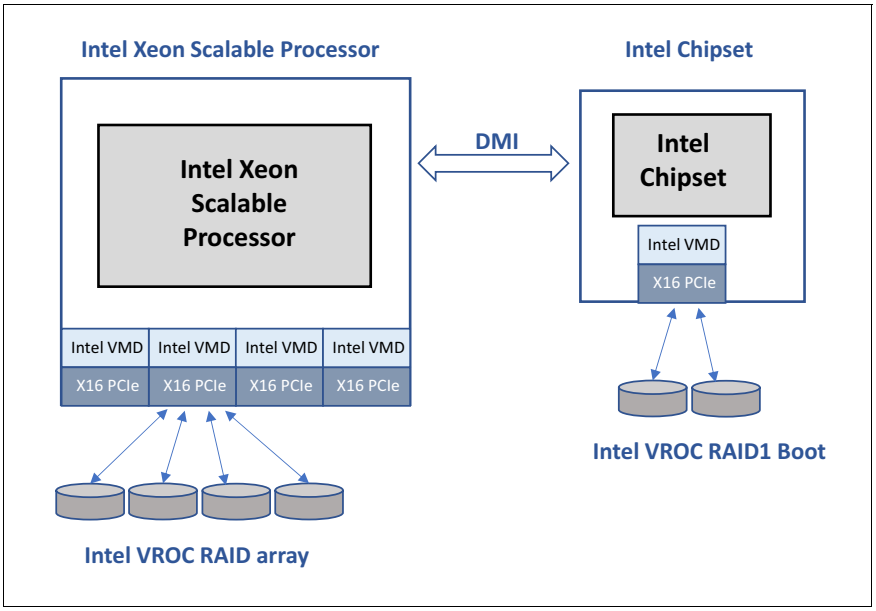

Figure 1 Intel VMD

Hypervisor technology has been developed to functionally divide bare metal server resources into virtual machines to improve IT flexibility and maximize hardware utilization. However, the resource management can cause increased latency and decreased performance of certain components.

Intel VMD Direct Assign is a specific application of Intel VMD that aims to improve the storage experience in HCI (Hyperconverged Infrastructure) architecture. As Figure 2 on page 4 shows, VMD Direct Assign takes the isolated storage subsystem and bypasses the hypervisor to assign a VMD controller directly to a virtual machine running on the host.

Each Intel VMD controller can be assigned up to one virtual machine, which will inherit the functions of Intel VMD and the performance of the NVMe drives without the storage bottleneck of the hypervisor.

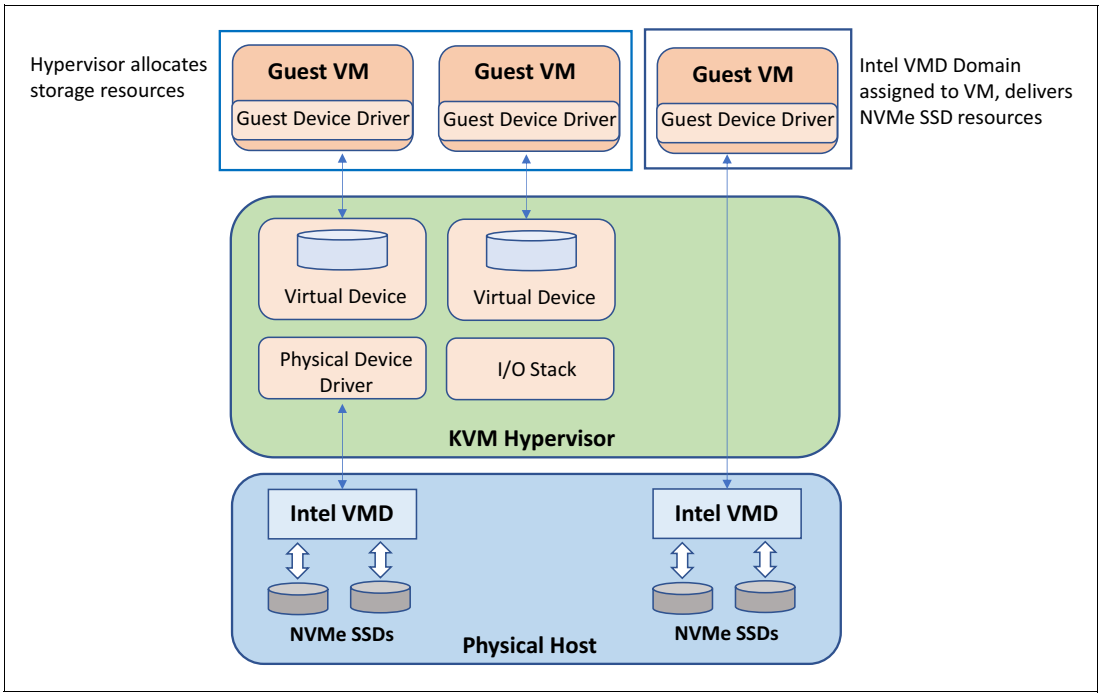

Figure 2 VMD direct assign

For more details about the Intel VMD and VMD Direct Assign, visit:

https://www.intel.com/content/www/us/en/architecture-and-technology/intel-volume-management-device-overview.html

The Lenovo paper discussing Intel VMD Direct Assign with VMware ESXi is available from the Lenovo Press web site:

https://lenovopress.com/lp1536-using-intel-vmd-direct-assign-with-vmware-esxi

Table 1 lists Lenovo ThinkSystem servers with the third-generation Intel Xeon Scalable processors (codename: Ice Lake) and the minimum version of UEFI firmware that supports the Intel VMD Direct Assign function.

Table 1 UEFI supported versions for Intel VMD Direct Assign

Lenovo ThinkSystem servers with 3rd Gen Intel Xeon Scalable processors (Ice Lake)	VMD Direct Assign function with Linux OS	Supported UEFI version
Lenovo ST650 V2 Server	Supported	1.02 or later
Lenovo SR630 V2 Server	Supported	1.02 or later
Lenovo SR650 V2 Server	Supported	1.02 or later
Lenovo SR670 V2 Server	Supported	1.02 or later
Lenovo SD630 V2 Server	Supported	1.02 or later
Lenovo SD650 V2 Server	Supported	1.02 or later
Lenovo SD650-N V2 Server	Supported	1.02 or later
Lenovo SN550 V2 Server	Supported	1.02 or later

For more information about how to update firmware in Lenovo ThinkSystem server, refer to the support page:

https://datacentersupport.lenovo.com/us/en/solutions/ht511325

In this paper, we will demonstrate Intel VMD Direct Assign function on Lenovo SD630 V2 server. As shown in Figure 3, the SD630 V2 supports either 2x 7mm 2.5-inch NVMe drives or 1x 15mm 2.5-inch NVMe drive directly connected to the PCIe ports on the system board.

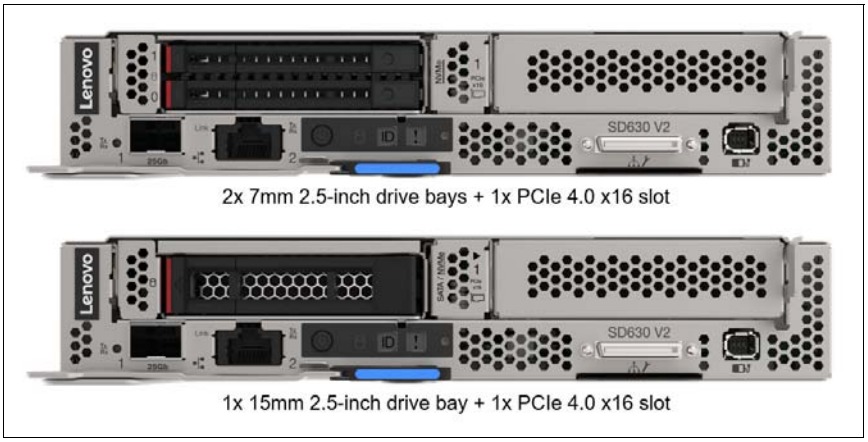

Figure 3 Front panel view of Lenovo ThinkSystem Server SD630 V2 server

Table 2 lists the operating system versions that support Intel VMD Direct Assign

Table 2 OS support for Intel VMD Direct Assign

Supported Host Linux OS	Supported Guest Linux OS
RHEL 8.2 and newer	RHEL 8.0 and newer
SLES 15 SP2 and newer	SLES 15 SP2 and newer

How to use VMD Direct Assign in Linux KVM

This section describes the steps we have taken to configure and enable Intel VMD Direct Assign for Linux OS with KVM. The configuration of our SD630 V2 test server is listed in Table 3 below. The host OS is installed on the M.2 SATA SSD.

OS boot drive: The NVMe drive cannot be installed as a host OS device because it is under the VMD domain that will be assigned to the guest OS.

Table 3 ThinkSystem SD630 V2 server configuration

Component	Configuration
Server	ThinkSystem SD630 V2 Server
CPU	2x Intel Xeon Scalable Processors 3rd Generation Platinum 8352V CPU @2.10 GHz
Memory	16x DDR4 3200 MHz 16GB RDIMM
M.2 SATA	1x M.2 240GB SATA 6Gbps SSD

Component	Configuration
NVMe drive	1x Intel Optane SSD P4800X 1.5 TB 2.5-inch 15mm NVMe
Host hypervisor OS	RHEL 8.2 with KVM package
Guest VM OS	RHEL 8.2

The overall steps to configure and enable Intel VMD Direct Assign are covered in the following sections:

- “Step 1: Determine VMD controller PCI address in the system” below
- “Step 2: Use OneCLI tool to enable UEFI VMD Direct Assign option” on page 7
- “Step 3: Install Host Linux OS RHEL 8.2 with Virtualization Packages” on page 8
- “Step 4: Create RHEL 8.2 VM guest OS” on page 10
- “Step 5. Install Intel replacement VMD driver in guest OS” on page 12
- “Step 6. Add VMD PCI Device Controller to the Guest OS” on page 13

Step 1: Determine VMD controller PCI address in the system

First, we must enable the VMD function in the UEFI setup menu and find the PCI address of the VMD controller in the system.

1. Power on the system and press F1 to enter UEFI setup menu.
2. Select **System Settings** → **Devices and I/O Ports** → **Intel VT for Directed I/O (VT-d)** and make sure the option is set to **Enabled** as below Figure 4 shows.

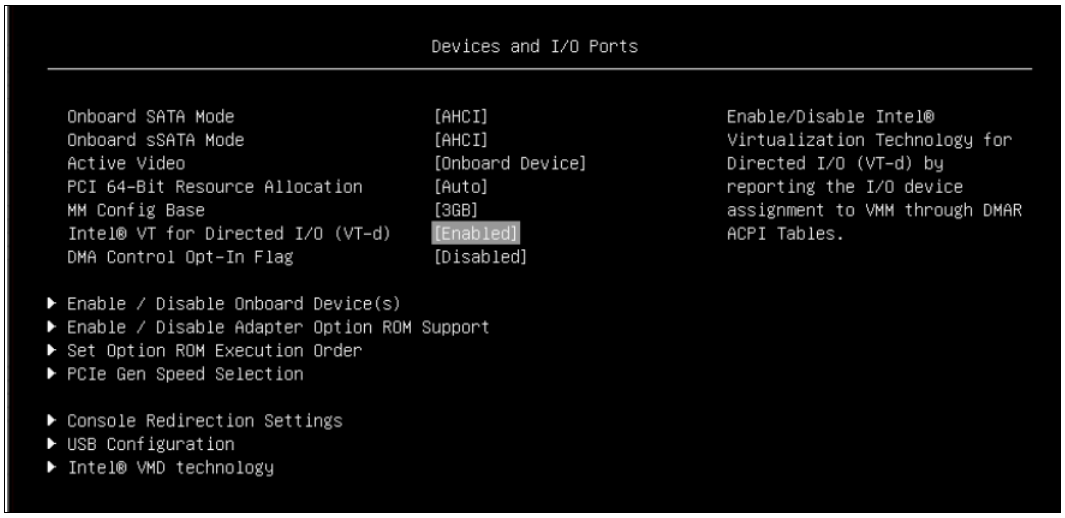

Figure 4 Intel VT-d option set to Enabled

3. Select **Intel VMD Technology** setting and set the option **Enable/Disable Intel VMD** to Enabled, as shown in Figure 5.

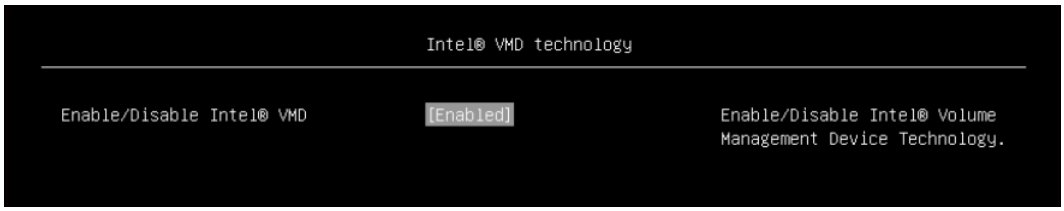

Figure 5 Intel VMD option set to Enabled

4. Save the settings and reboot the system to take effect, and then press F1 to renter the UEFI setup menu.
5. Go to **System Settings** → **Storage** → **Intel Virtual RAID on CPU** → **All Intel VMD Controllers**, select the **Non-RAID Physical Disk** and the PHYSICAL DISK INFO window will display that the VMD controller PCI address is `64:00.5`, as shown in Figure 6.

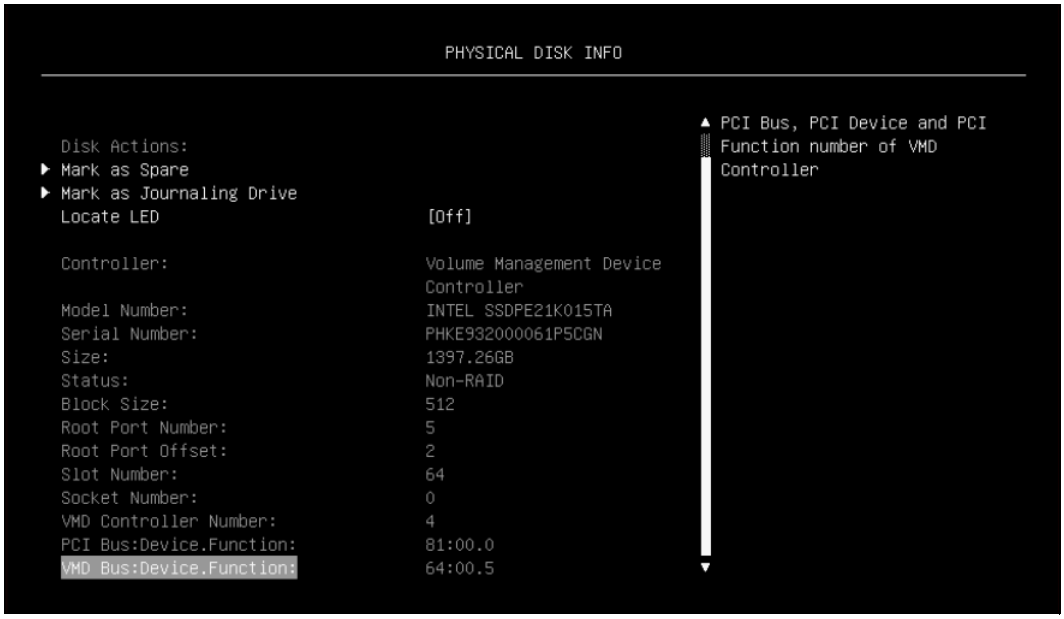

Figure 6 VMD PCI address from UEFI menu

Step 2: Use OneCLI tool to enable UEFI VMD Direct Assign option

The UEFI option VMD for Direct Assign can be enabled by using the Lenovo XClarity™ Essentials OneCLI tool. Download the tool from the Lenovo Support site:

https://datacentersupport.lenovo.com/us/en/solutions/ht116433

Use the following steps to enable UEFI VMD Direct Assign option from the operating system remotely:

1. Run the following OneCLI command to enable VMD Direct Assign:

```
C:\onecli>OneCli config set DevicesandIOPorts.VMDforDirectAssign Enabled --imm
<USERID>:<PASSWORD>@<IP>
```

Figure 7 shows how to enable VMD Direct Assign option from operating system remotely.

```
C:\kel\kel_tools\onecli>OneCli config set DevicesandIOPorts.VMDforDirectAssign Enabled --imm USERID:PASSWORD@10.241.67.1
80

Lenovo XClarity Essentials OneCLI lxce_onecli01u-3.4.0
(C) Lenovo 2013-2022 All Rights Reserved

OneCLI License Agreement and OneCLI Legal Information can be found at the following location:
    "C:\kel\kel_tools\onecli\Lic"

[1s]Certificate check finished [100%][=============================================================================>]

Start to connect BMC at 10.241.67.180 to apply config set
Invoking SET command ...
Connected to BMC at IP address 10.241.67.180 by IPMI
DevicesandIOPorts.VMDforDirectAssign=Enabled
Changes completed successfully, but these changes will not take effect until next restart the system.
Succeed.

C:\kel\kel_tools\onecli>
```

Figure 7 Enable VMD Direct Assign option via OneCLI command

2. Reboot the system.
3. Run the following OneCLI command to make sure that VMD Direct Assign option is enabled in the system.

   ```
   C:\onecli>OneCli config show DevicesandIOPorts.VMDforDirectAssign --imm
   <USERID>:<PASSWORD>@<IP>
   ```

   Figure 8 shows VMD Direct Assign option is enabled in the test system.

```
C:\kel\kel_tools\onecli>OneCli config show DevicesandIOPorts.VMDforDirectAssign --imm USERID:PASSWORD@10.241.67.180

Lenovo XClarity Essentials OneCLI lxce_onecli01u-3.4.0
(C) Lenovo 2013-2022 All Rights Reserved

OneCLI License Agreement and OneCLI Legal Information can be found at the following location:
    "C:\kel\kel_tools\onecli\Lic"

[1s]Certificate check finished [100%][=============================================================================>]

Start to connect BMC at 10.241.67.180 to apply config show
Invoking SHOW command ...
Connected to BMC at IP address 10.241.67.180 by IPMI
DevicesandIOPorts.VMDforDirectAssign=Enabled
Succeed.
```

Figure 8 Check VMD Direct Assign option via OneCLI command

Step 3: Install Host Linux OS RHEL 8.2 with Virtualization Packages

In this setup, we will install the Host Linux OS RHEL 8.2 on the M.2 SATA drive.

Note: The Host Linux OS needs to be installed on a separate drive, rather than the one that we will do the VMD Direct Assign to a guest OS.

1. Follow the standard process to install RHEL 8.2 as the Host OS. Ensure that the server OS has a GUI, and Virtualization Packages are selected and installed, as Figure 9 shows.

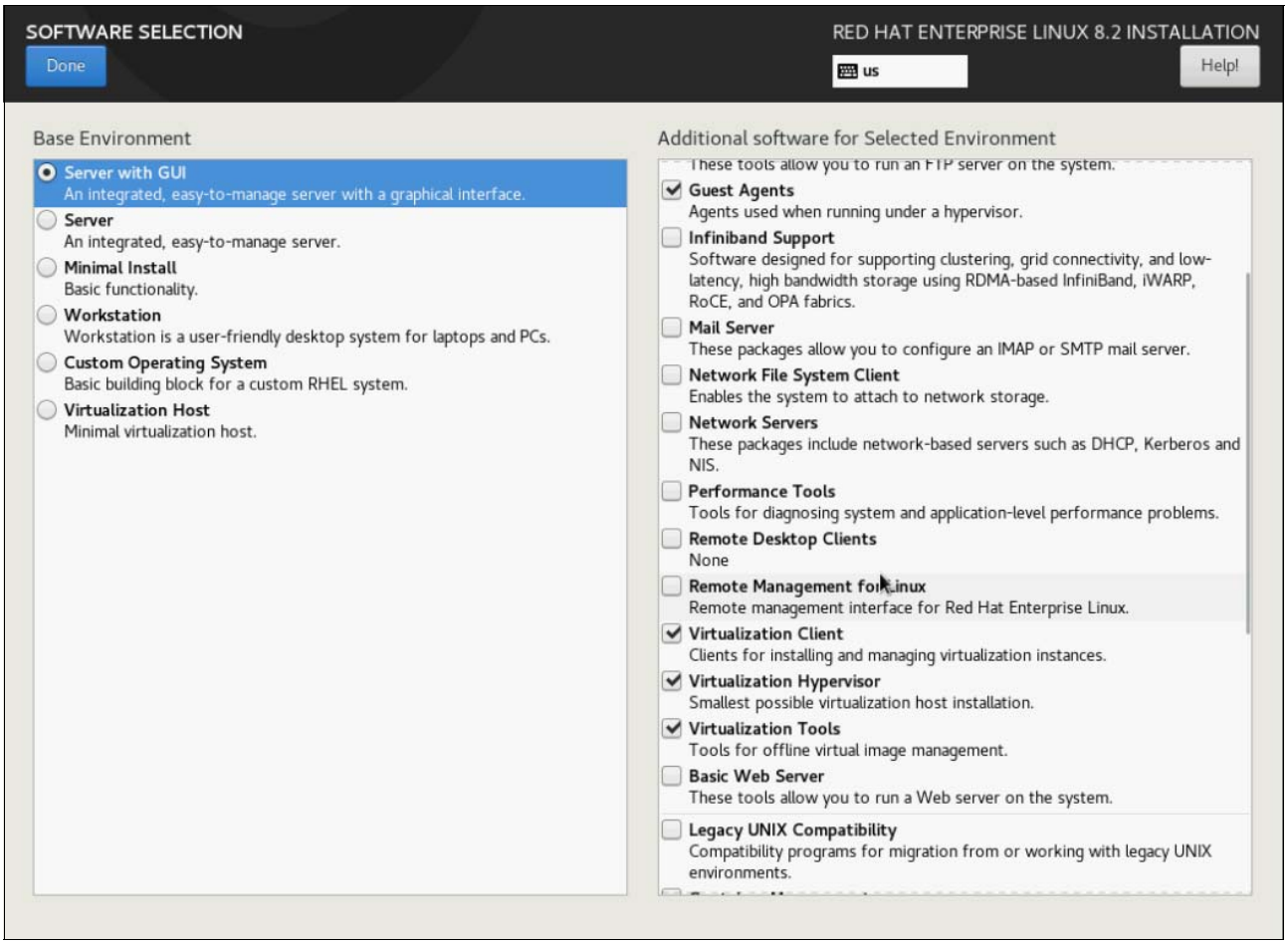

Figure 9 Server with GUI and Virtualization Packages are selected

2. Confirm that the virtualization related services are working.

   After installing the Host OS, run below commands to make sure that the Virtualization packages and all related services are installed and can run successfully:

   ~# **libvirtd**

   If the above command shows error, try to restart the service:

   ~# **systemctl restart libvirtd**

   And then check the virtualization status by:

   ~# **systemctl status libvirtd**

   Check if kvm is visible in lsmod output:

   ~# **lsmod | grep kvm**

   You will see something like below:

```
kvm_intel      294912      0
kvm            786432      1 kvm_intel
irqbypass      16384       1 kvm
```

3. Stop and disable Firewall using the following commands:

```
~# systemctl stop firewalld
~# systemctl disable firewalld
```

4. Check whether the virtual network is active.

```
~# virsh net-list
```

You will see something like below:

```
Name     State    Autostart   Persistent
----------------------------------------------------------
default  active   yes         yes
```

If virtual network is not active, do:

```
~# virsh net-start default
```

5. Enable the IOMMU passthrough mode in the Host OS.

Edit the grub file (/etc/default/grub), add the parameter `intel_iommu=on` to the end of GRUB_CMDLINE_LINUX line, and save this file.

Then rebuild grub, using the following command:

```
~# grub2-mkconfig > /boot/efi/EFI/redhat/grub.cfg
```

6. Reboot the host OS to make the change effective.

Step 4: Create RHEL 8.2 VM guest OS

Using scp or another FTP tool to transfer the RHEL 8.2 ISO file to the Host OS /root/ directory. We will access this ISO to create a guest OS.

1. Launch Virtualization Manager using the following command:

```
~# virt-manager
```

Select **File** → **New Virtual Machine**.

2. Choose the ISO file by browsing the /root as shown in Figure 10.

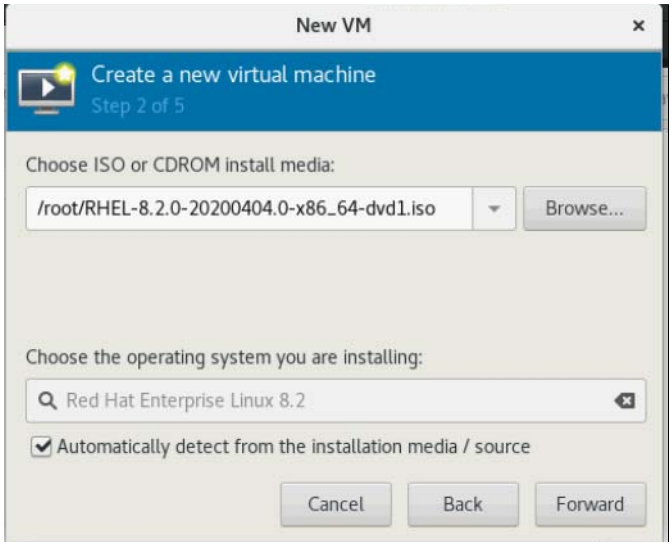

Figure 10 Choose the ISO file for guest OS creation

3. Choose Memory and CPU settings, Figure 11.

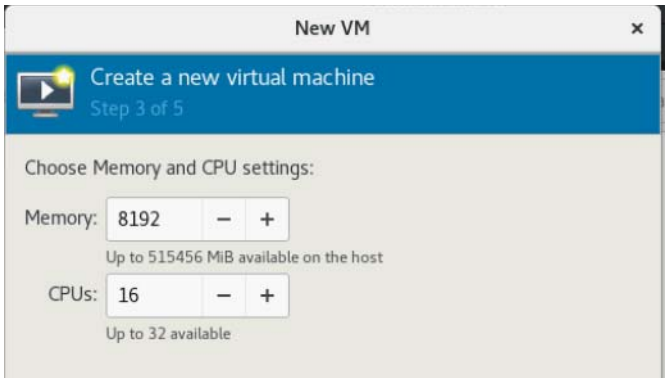

Figure 11 Choose Memory and CPU settings

4. Create a disk image for the virtual machine, Figure 12.

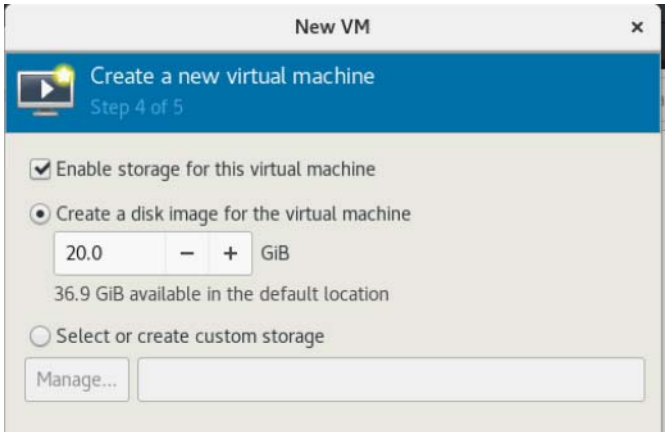

Figure 12 Create a disk for the guest OS

5. Follow the standard process to install RHEL 8.2 as VM guest OS, Figure 13.

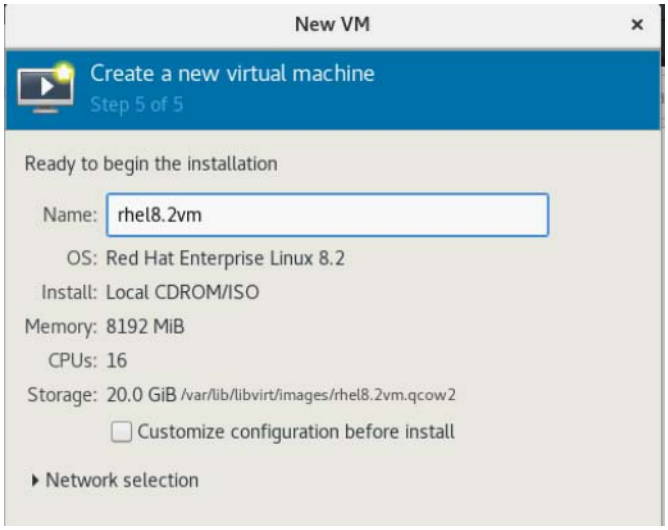

Figure 13 Ready for guest OS creation

Step 5. Install Intel replacement VMD driver in guest OS

After guest OS installation is completed, install the replacement VMD driver from Intel to enhance security and optimize performance.

1. Go to Intel website and download Intel replacement VMD driver of VMD Direct Assign setup for RHEL 8.2 VM. Store it in a USB storage device for use.

   https://www.intel.com/content/www/us/en/search.html?ws=text#q=vmd&t=All

2. Use the Virtual Machine USB Redirection function to load the USB storage device that has already stored the Intel replacement VMD driver, Figure 14.

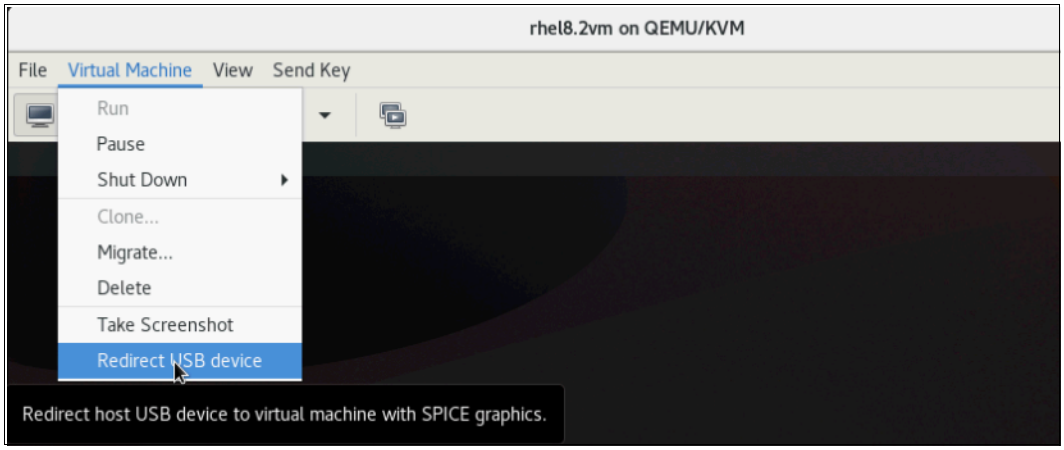

Figure 14 Enable the Virtual Machine USB Redirection function

3. In the guest OS, create a sub-directory, such as vm-driver, to store the Intel replacement VMD driver. Copy the Intel replacement VMD driver from the USB storage device to the VM- driver sub-directory.
4. Install the replacement driver in the guest OS.

   Use below command from OS command line to install VMD replacement driver:

   ~# **rpm -ivh kmod-iavmd-0.13.1.821-rhel_82.x86_64.rpm**

   The output is shown in Figure 15.

```
[root@localhost RHEL-8-2-0-]# cd kel
[root@localhost kel]# ls
 iavmd-0.13.1.821-rhel_82.src.rpm
 kmod-iavmd-0.13.1.821-rhel_82.x86_64.rpm
'Set-up RHEL VM in ESXi_RHEL_8_2_Signed.pdf'
[root@localhost kel]# cp * /root/vm-driver/
[root@localhost kel]# cd /root/vm-driver/
[root@localhost vm-driver]# ls
 iavmd-0.13.1.821-rhel_82.src.rpm
 kmod-iavmd-0.13.1.821-rhel_82.x86_64.rpm
'Set-up RHEL VM in ESXi_RHEL_8_2_Signed.pdf'
[root@localhost vm-driver]# rpm -ivh kmod-iavmd-0.13.1.821-rhel_82.x86_64.rpm
warning: kmod-iavmd-0.13.1.821-rhel_82.x86_64.rpm: Header V4 RSA/SHA256 Signatur
e, key ID c343c1b0: NOKEY
Verifying...                          ################################# [100%]
Preparing...                          ################################# [100%]
Updating / installing...
   1:kmod-iavmd-0.13.1.821-rhel_82    ################################# [100%]
[root@localhost vm-driver]#
```

Figure 15 Install Intel replacement VMD driver in guest OS

5. Rebuild the initramfs using the following command:

   ~# **dracut -f**

6. Enable the Direct Assign mode in the Guest OS.

Edit the grub file (`/etc/default/grub`), add the Direct Assign parameter `iavmd.direct_assign=1` to the end of `GRUB_CMDLINE_LINUX` line, and save this file.

7. Rebuild grub, using the following command:

   ~# **grub2-mkconfig > /boot/efi/EFI/redhat/grub.cfg**

8. Reboot the guest OS to make the change effective.

Step 6. Add VMD PCI Device Controller to the Guest OS

1. Open virt-manager, shutdown virtual machine, and select **VM** → **Edit** → **Virtual Machine Details** . Figure 16

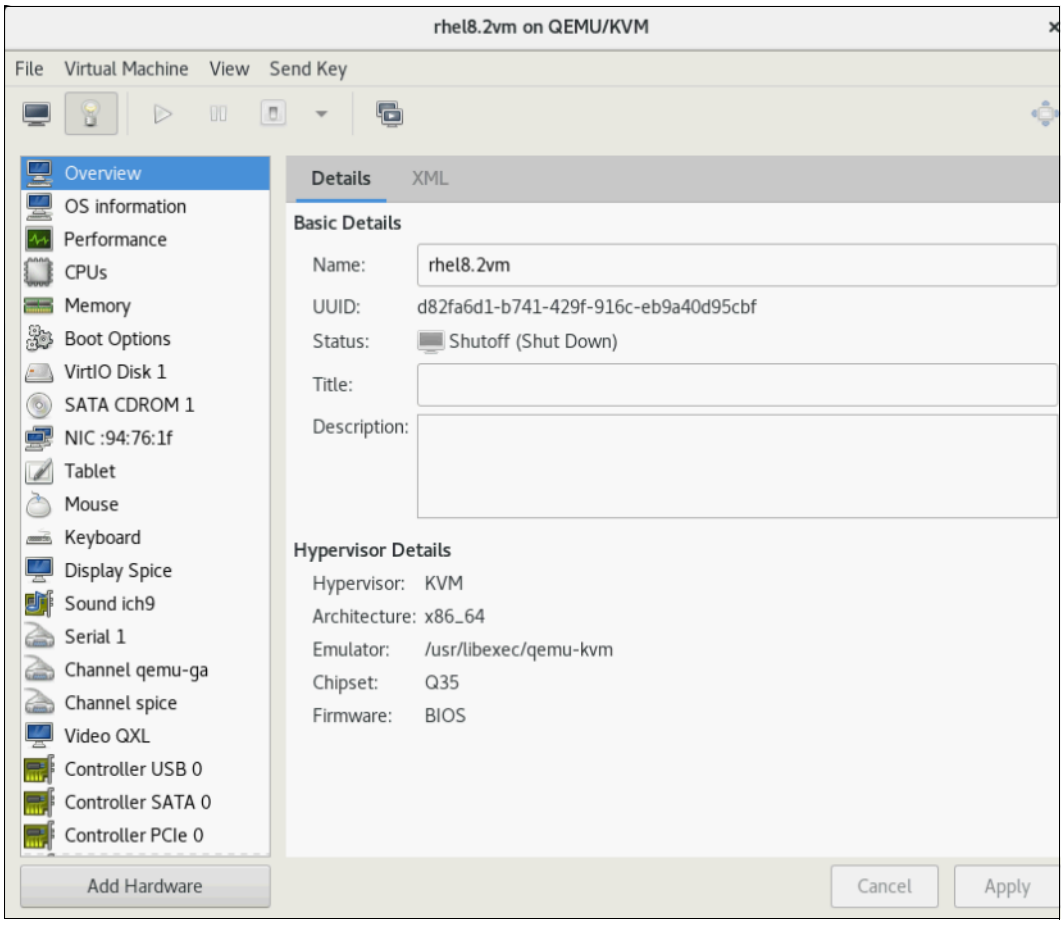

Figure 16 Add Hardware

2. Select **View** → **Details** (or click on Light Bulb icon) and click **Add Hardware**. Figure 17 appears.

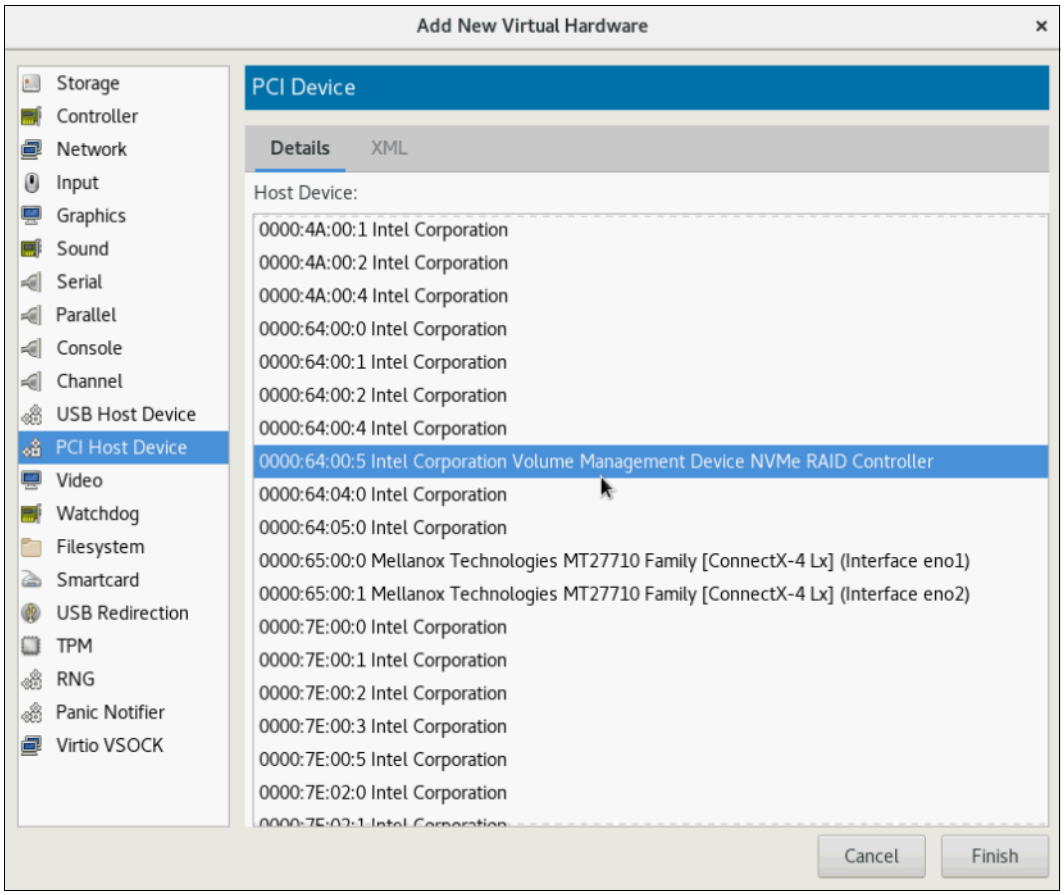

Figure 17 Add New Virtual Hardware window

3. In the Add New Virtual Hardware window, select **PCI Host Device** on the left, then scroll down and select **Intel Corporation Volume Management Device NVMe RAID Controller** as shown in Figure 17.

   The device ID shown in Figure 17 should be consistent with the BIOS VMD Controller detail ID, PCI 0000:64:00.5 in Figure 6 on page 7.

4. Press the Finish button to complete the configuration.

Verify VMD Device in Guest OS

Since all settings and configurations are completed in both Host OS and guest OS, do below to verify whether the VMD Direct Assign works normally.

1. Power on and boot the VM guest OS RHEL 8.2 with VMD replacement driver.

2. Log in to the guest OS as root user and open a terminal. Using the **lsblk** command to list all the block devices. The nvme0n1 device appears as a dedicated NVMe drive in the guest OS as shown in Figure 18.

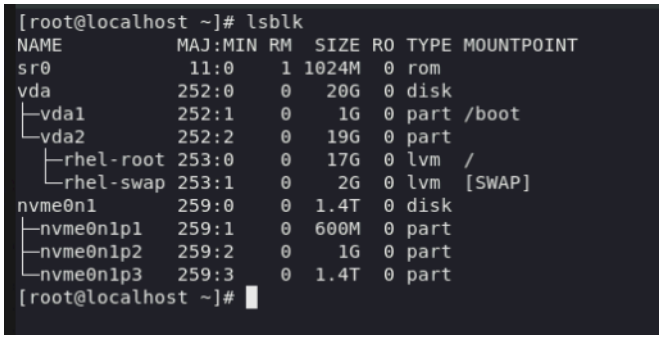

Figure 18 The nvme0n1 device appears as dedicated drive in the guest OS

Resources

- Intel Volume Management Device (Intel VMD)

  https://www.intel.com/content/www/us/en/architecture-and-technology/intel-volume-management-device-overview.html
- KVM-DirectAssign-RHEL82VM

  https://www.intel.com/content/www/us/en/search.html?ws=text#q=vmd&sort=relevancy
- Lenovo ThinkSystem SD630 V2 Server Product Guide

  https://lenovopress.com/lp1394-thinksystem-sd630-v2-server

Authors

Kelvin Shieh is the OS Development Technical Lead for the Lenovo Infrastructure Solutions Group, based in Taipei, Taiwan.

Alpus Chen is an OS Engineer at the Lenovo Infrastructure Solutions Group in Taipei, Taiwan. As a specialist in Linux and VMware technical support for several years, he is interested in operating system operation and recently focuses on VMware OS.

Thanks to the following specialists for their contributions and suggestions:

- Gary Cudak, Lenovo OS Architect
- 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 May 30, 2022.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp1576

Trademarks

Lenovo and the Lenovo logo 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 from https://www.lenovo.com/us/en/legal/copytrade/.

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

Lenovo®	Lenovo(logo)®
Lenovo XClarity™	ThinkSystem™

The following terms are trademarks of other companies:

Intel, Intel Optane, Xeon, 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.

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