Configuring VMware VMDirectPath I/O Passthrough with NVMe SSDs on ThinkSystem Servers

Describes how to make NVMe drives available to ESXi virtual machines

Provides steps to implement VMDirectPath I/O on ESXi 7.0 and a RHEL 7.8 guest OS

Lists which NVMe Switch Adapters support the passthrough function

Shows how to confirm that the NVMe SSDs are working normally in VMs

Boyong Li

Abstract

VMDirectPath I/O (PCI passthrough) enables direct assignment of hardware PCI functions such as NVMe solid-state drives to a virtual machine. This gives the VM access to the PCI functions with minimal intervention from the ESXi host, potentially improving performance.

In this paper we describe how to configure NVMe SSDs passthrough as PCI devices to VMs on Lenovo® ThinkSystem™ servers. We provide step-by-step instructions using ESXi 7.0 U1.

This paper is intended for IT specialists and IT managers who want to learn more about NVMe SSDs passthrough.

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

NVMe passthrough enables direct assignment of hardware NVMe devices to VMs. This gives the VM access to the NVMe SSDs with minimal intervention from the ESXi host, potentially improving performance. It is suitable for performance-critical workloads such as storage acceleration for VMs, and other high-speed storage solutions such as NetApp storage solutions.

While VMDirectPath I/O can improve the performance of a VM, enabling it makes several important features of vSphere unavailable to the VM, such as Suspend and Resume, Snapshots, Fault Tolerance, and vMotion.

Figure 1 shows a workflow of DirectPath I/O.

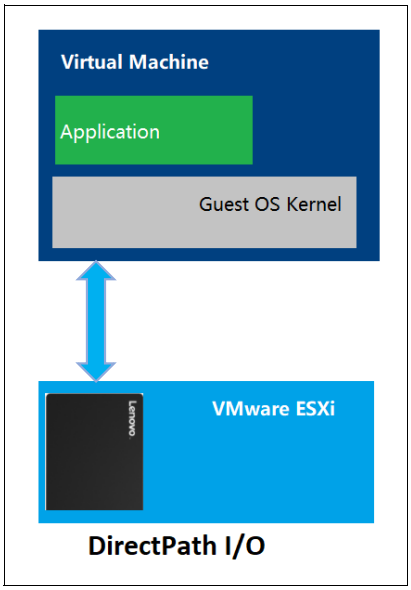


Figure 1 Passthrough workflow

The prerequisites for VMDirectPath I/O are as follows:

- Verify that your NVMe SSD devices and PCIe NVMe switchs are supported in the specific ThinkSystem server you are using. Refer to Table 1 on page 4.
- Verify that the host has Intel Virtualization Technology for Directed I/O (VT-d) or AMD I/O Virtualization Technology (IOMMU) is enabled in BIOS of the server.

Notes:

1. Not all ESX servers have full support for VMDirectPath. Please refer to:

   https://kb.vmware.com/s/article/2142307

2. A maximum of 16 passthrough devices is supported per VM on ESXi 6.x and 7.x. For the supported maximum devices number of each version of VMware vSphere, refer to:

   https://configmax.vmware.com/

Onboard NVMe ports and the ThinkSystem 1611-8P NVMe Switch Adapter support VMDirectPath I/O Passthrough as indicated in Table 1.

Table 1 Support for VMDirectPath I/O Passthrough

NVMe connection	Support for VMDirectPath I/O Passthrough
Onboard NVMe ports	
Intel Xeon Scalable processors (Gen 1, 2, 3)	Supported[a]
AMD EYPC processors (Gen 2, 3)	Supported
NVMe Switch Adapters	
810-4P NVMe Switch Adapter	No support
1610-4P NVMe Switch Adapter	No support
1610-8P NVMe Switch Adapter	No support
1611-8P NVMe Switch Adapter	Supported

a. For support, Intel VMD must be disabled.

For the latest server support of the NVMe Switch Adapters, refer to the Lenovo ThinkSystem RAID Adapter and HBA Reference:

https://lenovopress.com/lp1288-thinksystem-raid-adapter-and-hba-reference#term=nvme%2520switch

You can verify the PCIe NVMe SSDs compatibility of Lenovo ThinkSystem servers using the VMware Compatibility Guide, Figure 2, available at:

https://www.vmware.com/resources/compatibility/search.php?deviceCategory=io

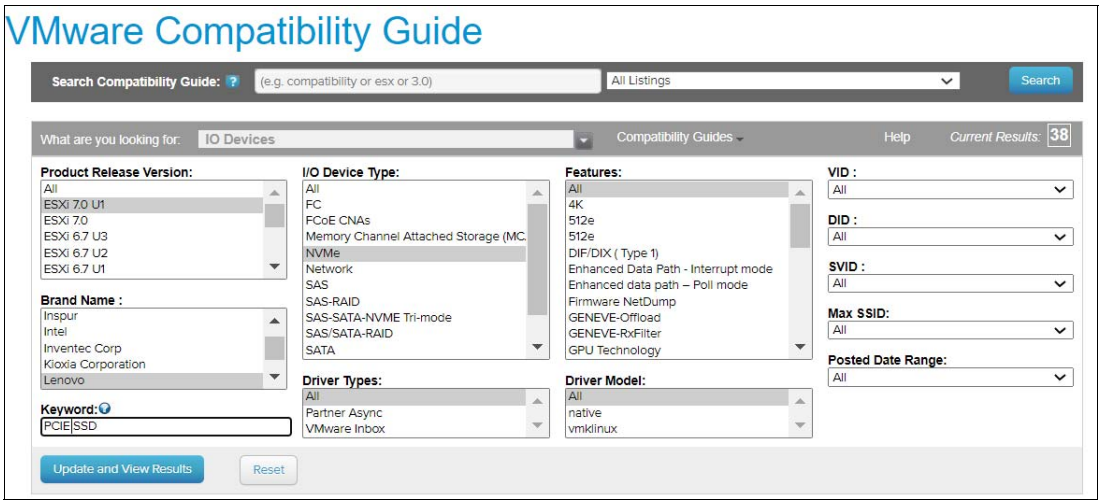


Figure 2 VMware Compatiblity Guide

In this paper, we are using the ThinkSystem SR650 V2 server as our test server, since it supports both onboard NVMe ports and the 1611-8P NVMe Switch Adapter.

There are two scenarios corresponding to the different connection modes of NVMe SSDs:

1. Directly connected to the PCIe NVMe ports on the motherboard
2. Connected to a PCIe NVMe Switch Adapter which is plugged into the PCIe riser on the motherboard.

Scenario 1: Connecting the NVMe SSDs to onboard NVMe ports

This scenario is divided into three steps:

- “Step 1: Configure the UEFI Options”
- “Step 2: Install vSphere and VM, then configure NVMe SSDs passthrough” on page 7
- “Step 3: Check the NVMe SSDs on VM” on page 11

Step 1: Configure the UEFI Options

The PCIe NVMe ports (from PCIe port1 to port6) on the Lenovo SR650 V2 motherboard are multi-purpose ports. If we enable Intel VMD (Volume Management Device) in the UEFI setup, the PCIe port will be controlled by the VMD function. We need to configure these ports as the PCIe port and disable the VMD function in UEFI setup option.

AMD processor servers: AMD processor-based servers do not need a UEFI setting enabled, so this step can be skipped for these servers.

1. Configure the UEFI setup: Enter the UEFI setup by pressing the F1 at server booting, choose **System Settings** → **Devices and I/O Ports** → **Intel VMD technology** → **Enable**/**Disable Intel VMD** and then change the option to **Disabled** as shown in Figure 3.

Figure 3 Intel VMD disabled

2. Return to the main menu, save settings and reboot.
3. Reenter UEFI setup by pressing the F1 at server boot, choose **System Settings** → **Storage**, and check that the NVMe SSDs have been identified in the list as shown in Figure 4.

Figure 4 SDD listing

4. Select an entry in the list to view the details of the drive, as shown in Figure 5.

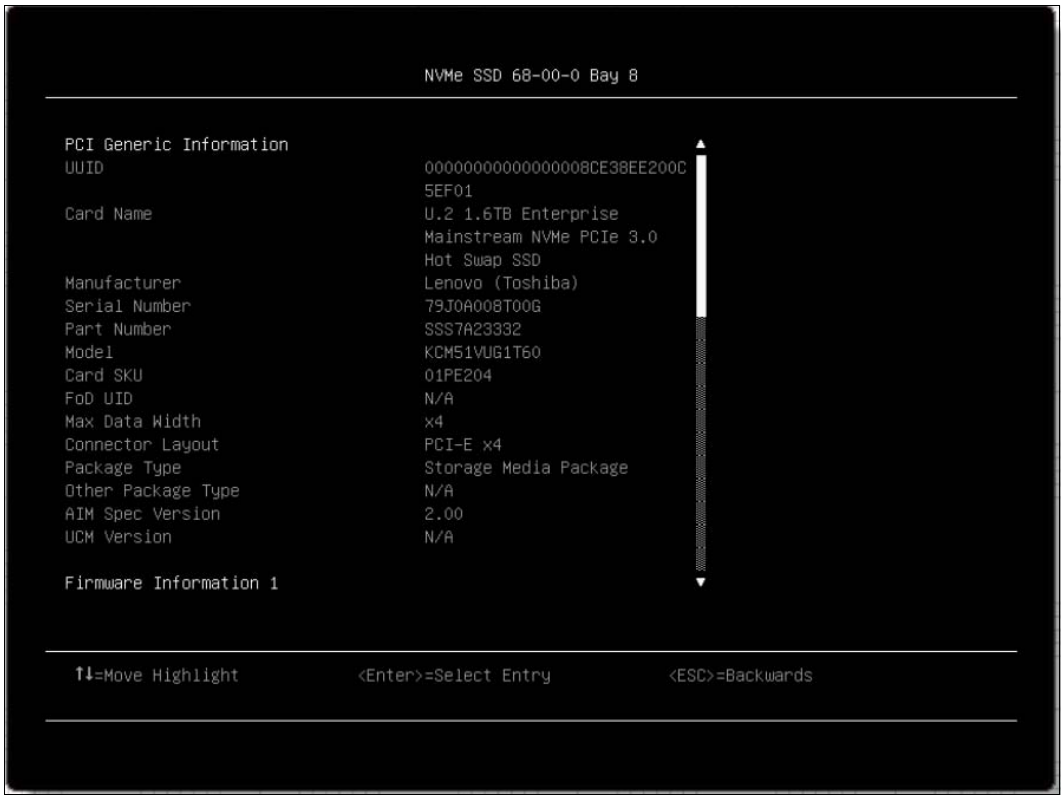


Figure 5 Details of the selected SSD

5. Return to the main menu, save settings and reboot.

Step 2: Install vSphere and VM, then configure NVMe SSDs passthrough

1. Install the OS. In our testing, we installed ESXi 7.0 U1 on the ThinkSystem SR650 V2 server, Figure 6.

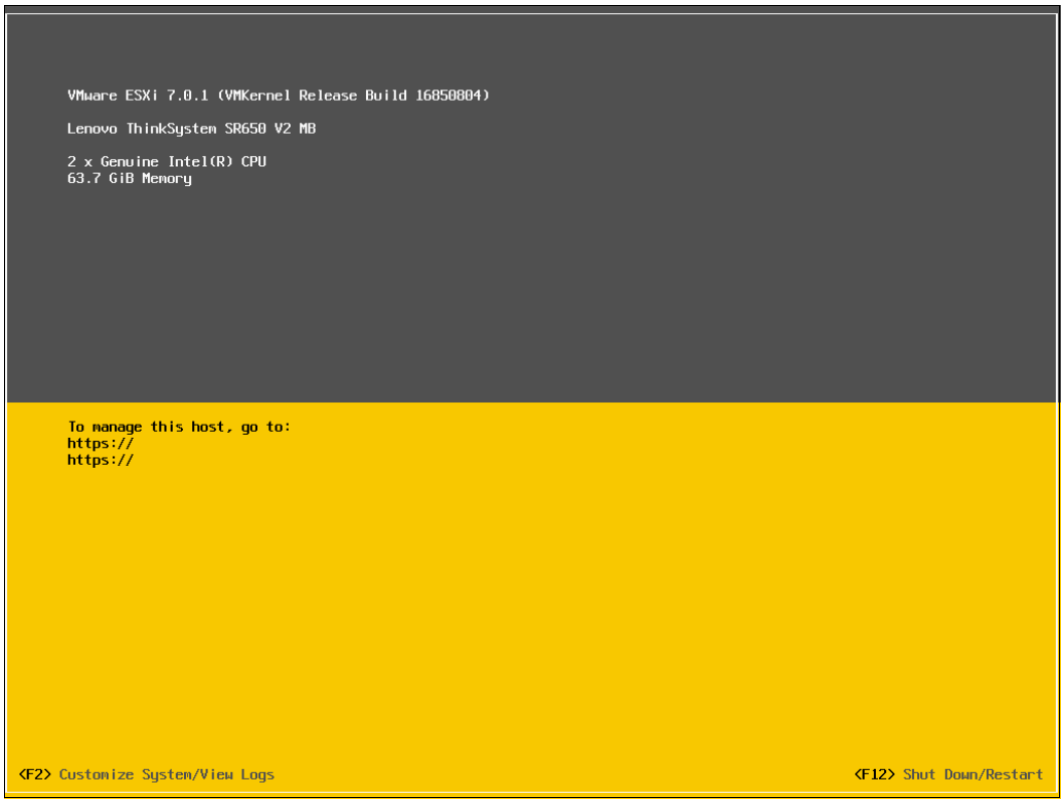


Figure 6 Installing ESXi 7.0 U1

2. Enter the Host client, Figure 7

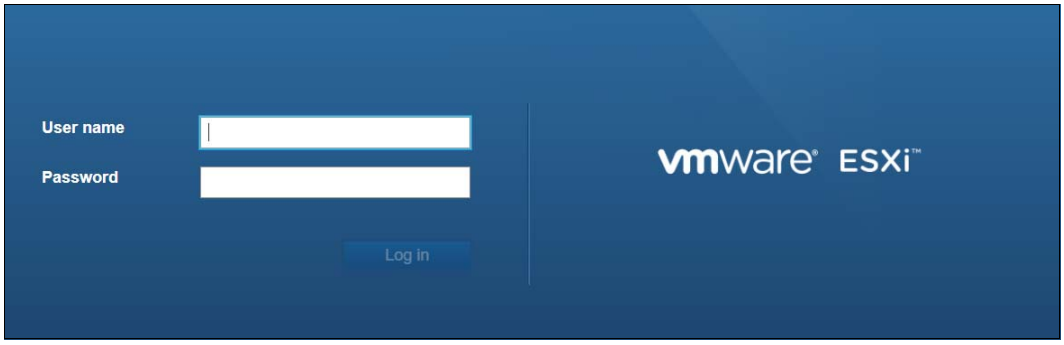


Figure 7 Login to the host client

3. Create a new VM and install OS on the VM as shown in Figure 8. We install REHL 8.3 as an example.

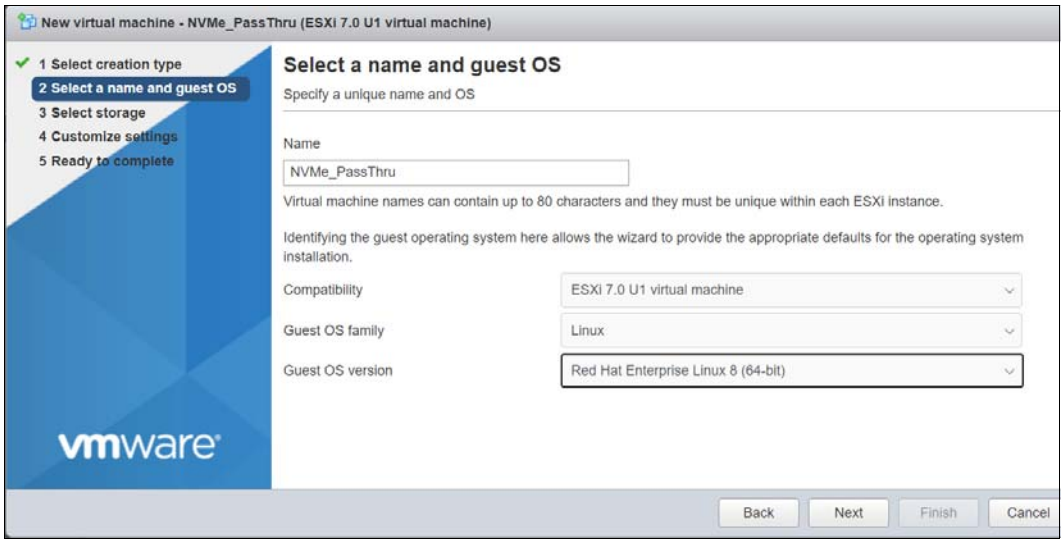


Figure 8 Creating the guest VM

4. Check the VM that has been installed, Figure 9.

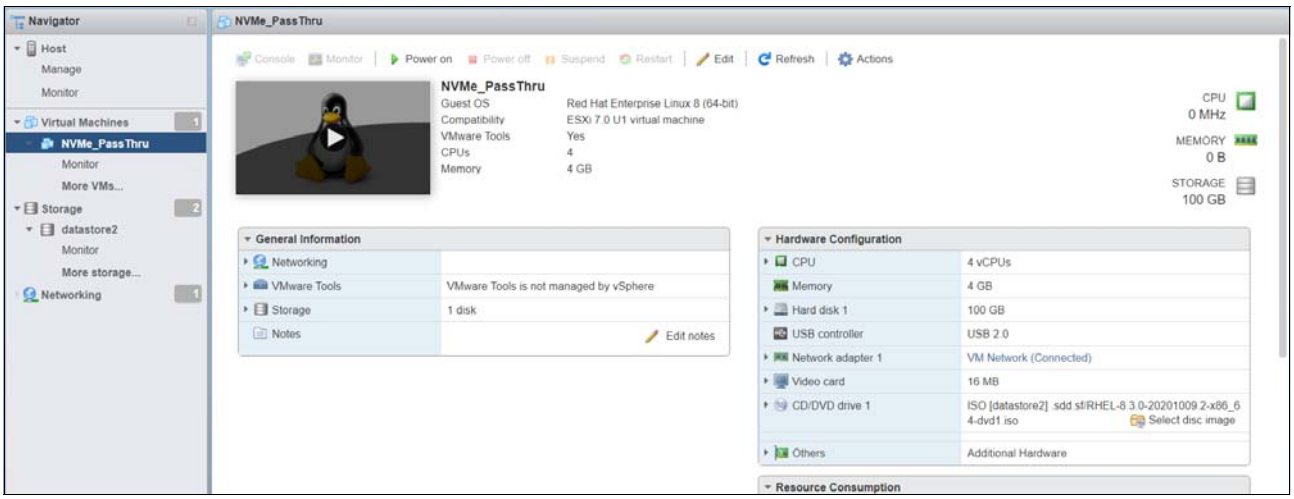


Figure 9 VM configuration

5. From the left-hand navigation menu, click **Manage** and select **Hardware** → **PCI Devices** and find the two NVMe SSD devices. Select them click **Toggle Passthrough**, Figure 10

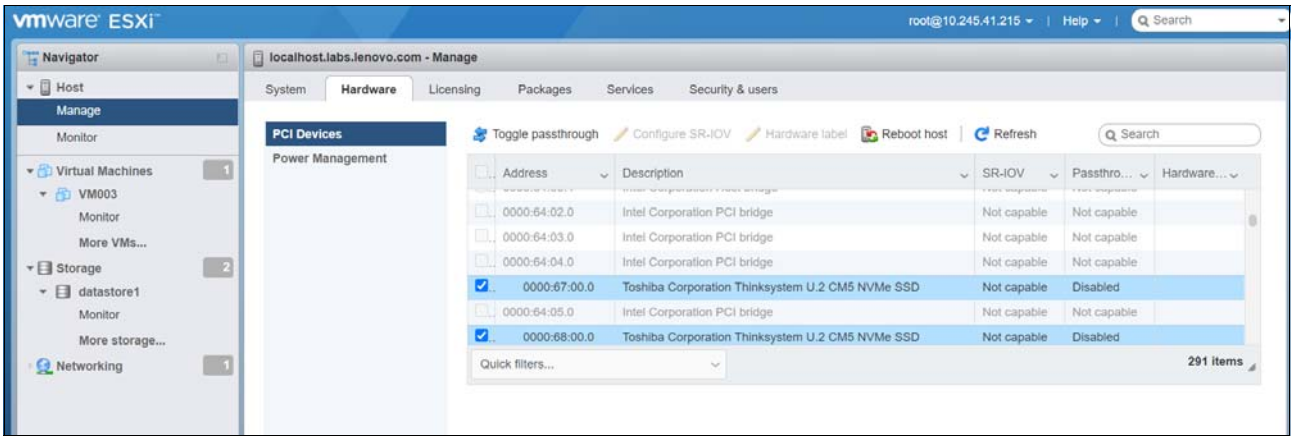


Figure 10 Enabling passthrough on the SSDs

6. Some versions of VMware vSphere may prompt that system needs to reboot ESXi host to take effect. Reboot the server if prompted to do so at this point.
7. The status of passthrough for the devices should now change to “Active” as shown in Figure 11.

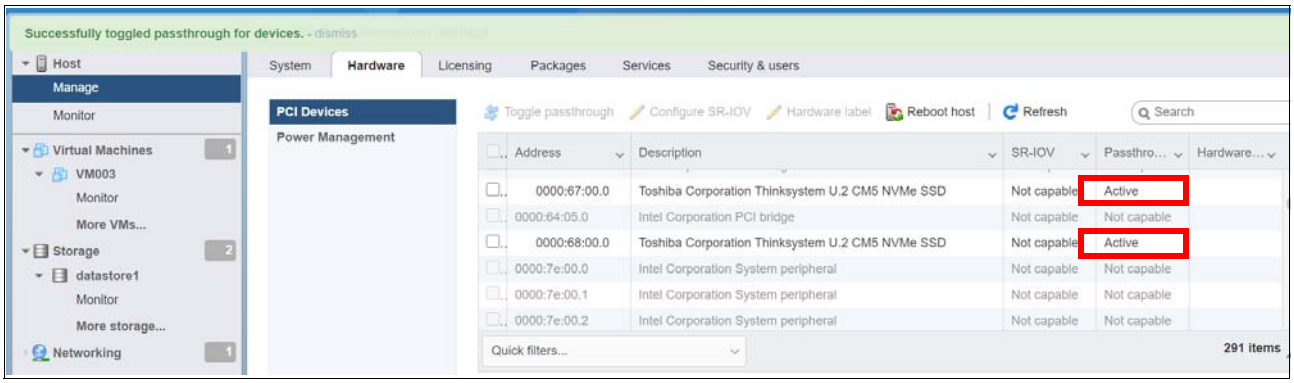


Figure 11 Drives are active

8. Go to the VM, Click the **Edit** button to edit settings of this VM, Figure 12.

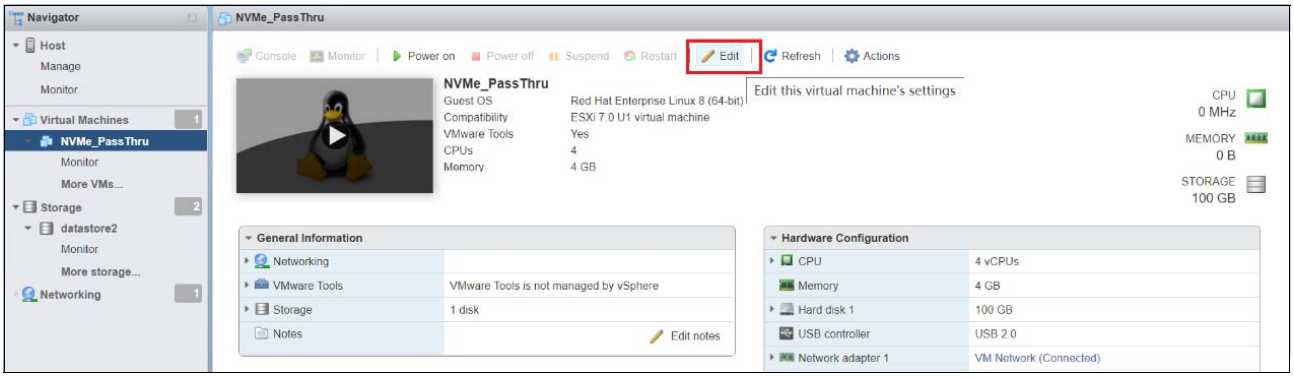


Figure 12 Edit the VM

9. Click the **Add other device** → **PCI device**, Figure 13.

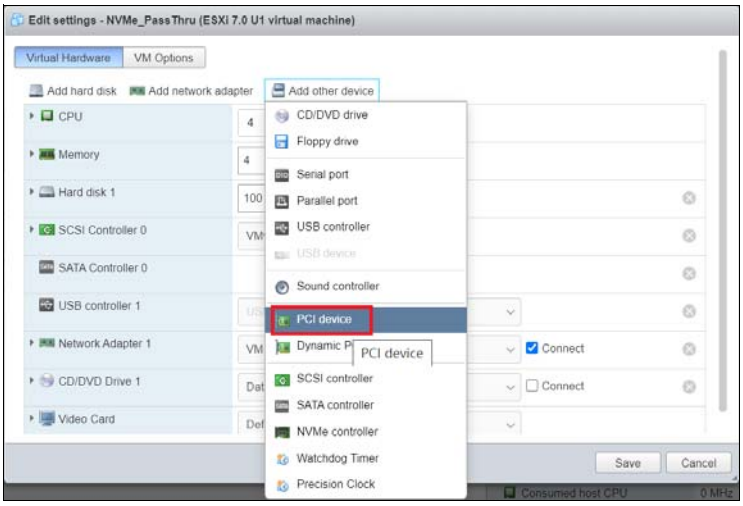


Figure 13 Add PCI device to the VM

10. Add the two NVMe SSDs as new PCI device and click Save to save the configuration, Figure 14.

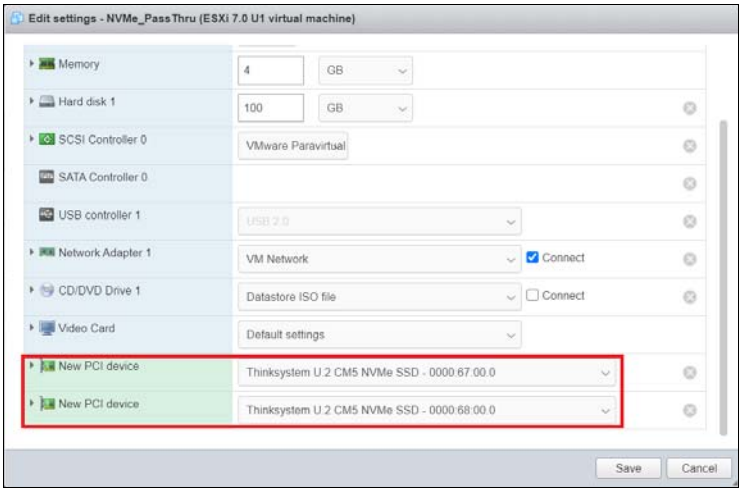


Figure 14 Add the SSDs to the VM

11. When the device is assigned, the VM must have a memory reservation for the full configured memory size. This means that we need to configure the memory reservation for NVMe devices and requires setting the reservation size equal to the memory capacity size. In Figure 15, you can either choose **Reserve all guest memory (All locked)** or fill the same size in the Reservation option, and save the configuration.

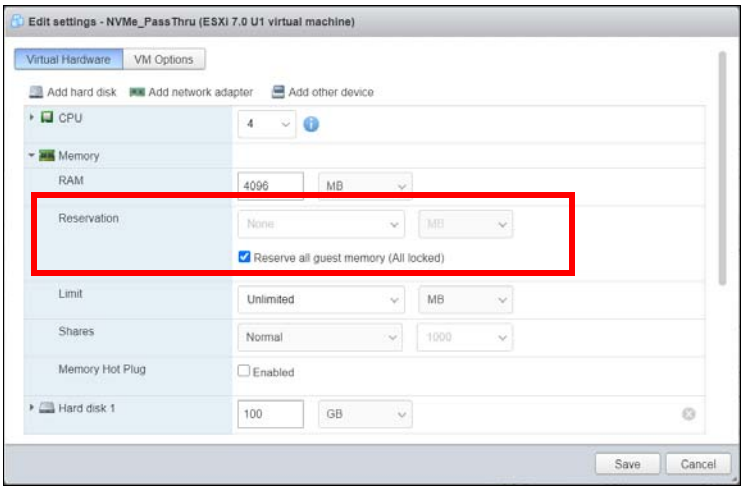


Figure 15 Memory reservation

Step 3: Check the NVMe SSDs on VM

1. Power on the VM. Our test environment uses a RHEL 8.3 guest OS.
2. Log in as an administrator and open a terminal.
3. Enter the command to find the NVMe SSDs which we configured as available via passthrough to the VM:

```
lspci | grep -i nvme
```

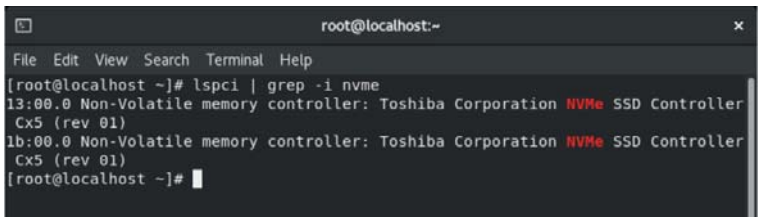


Figure 16 lspci command output

4. Check the dmesg to make sure there are no errors or warnings in the log:

```
dmesg | grep - i nvme
```

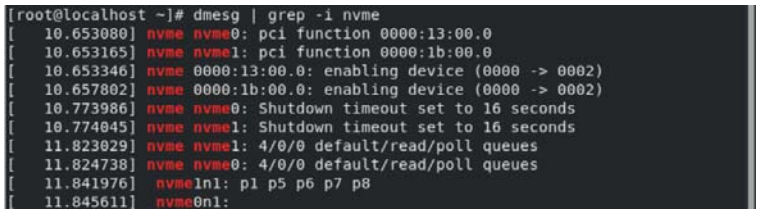
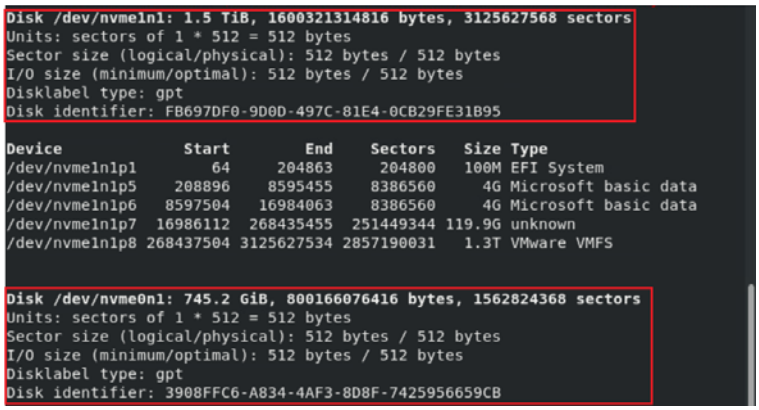


Figure 17 dmesg command output

5. Use fdisk to check if the hard disk has been mounted successfully.

```
fdisk -l
```



```
Disk /dev/nvme1n1: 1.5 TiB, 1600321314816 bytes, 3125627568 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: gpt
Disk identifier: FB697DF0-9D0D-497C-81E4-0CB29FE31B95

Device             Start        End    Sectors   Size Type
/dev/nvme1n1p1        64     204863     204800   100M EFI System
/dev/nvme1n1p5    208896    8595455    8386560     4G Microsoft basic data
/dev/nvme1n1p6   8597504   16984063    8386560     4G Microsoft basic data
/dev/nvme1n1p7  16986112  268435455  251449344 119.9G unknown
/dev/nvme1n1p8 268437504 3125627534 2857190031   1.3T VMware VMFS

Disk /dev/nvme0n1: 745.2 GiB, 800166076416 bytes, 1562824368 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: gpt
Disk identifier: 3908FFC6-A834-4AF3-8D8F-7425956659CB
```

Figure 18 fdisk command output

6. Now, the NVMe devices have been configured passthrough to the VM successfully, and you can access and use it as a real drive in the virtual OS.

Scenario 2: Connecting the NVMe SSDs to an NVMe Switch

In this scenario, the SSDs you wish to enable passthrough on are connected via an NVMe Switch Adapter.

Note: First, we need to confirm that the NVMe switch supports PCI passthrough function. At present, there is a ThinkSystem 1611-8P NVMe Switch Adapter that can support this function on the corresponding Lenovo serves.

Step 1: Configure the UEFI Options

1. Enter the UEFI setup by pressing the F1 at server booting, choose **System Settings** → **Storage**, and check that the NVMe switch and NVMe SSDs have been identified in the list, Figure 19.

Figure 19 PCIe device listing - adapter and NVMe drives

2. Check the NVMe switch properties, and ensure the firmware is in recent versions, Figure 20.

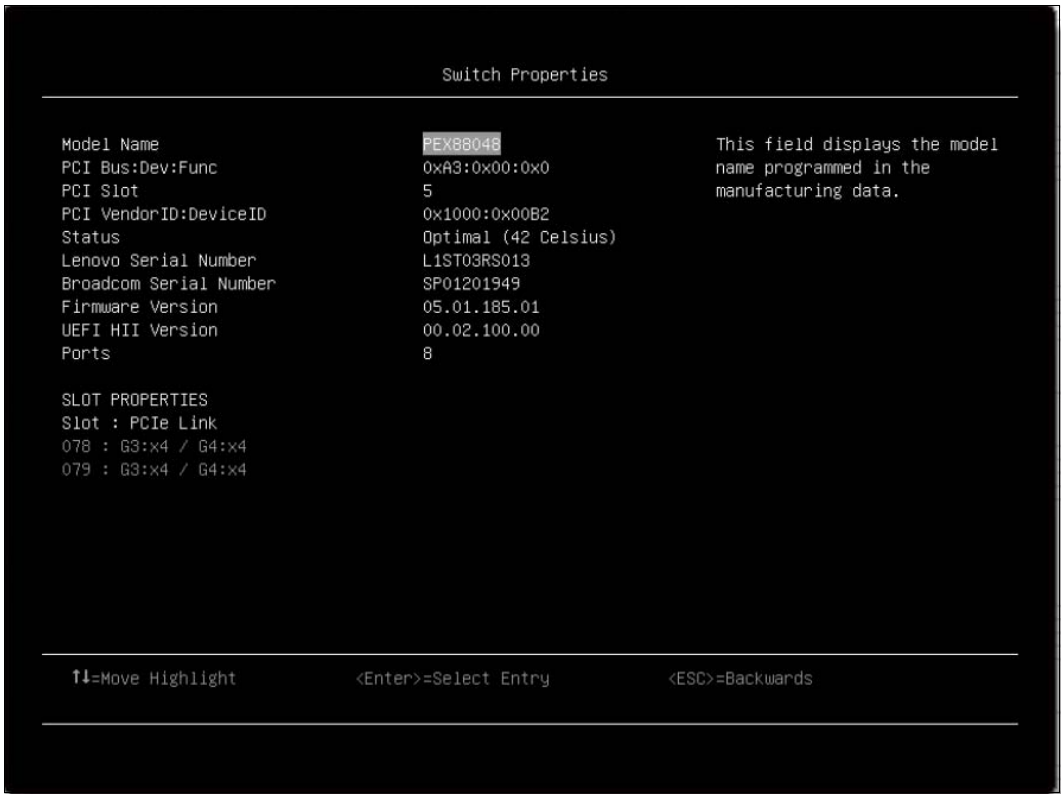


Figure 20 Switch Adapter properties

Step 2: Configure NVMe SSDs passthrough

We can repeat the steps in “Step 2: Install vSphere and VM, then configure NVMe SSDs passthrough” on page 7, then configure NVMe SSDs passthrough.

1. Activate the passthrough function of the NVMe SSDs, Figure 21.

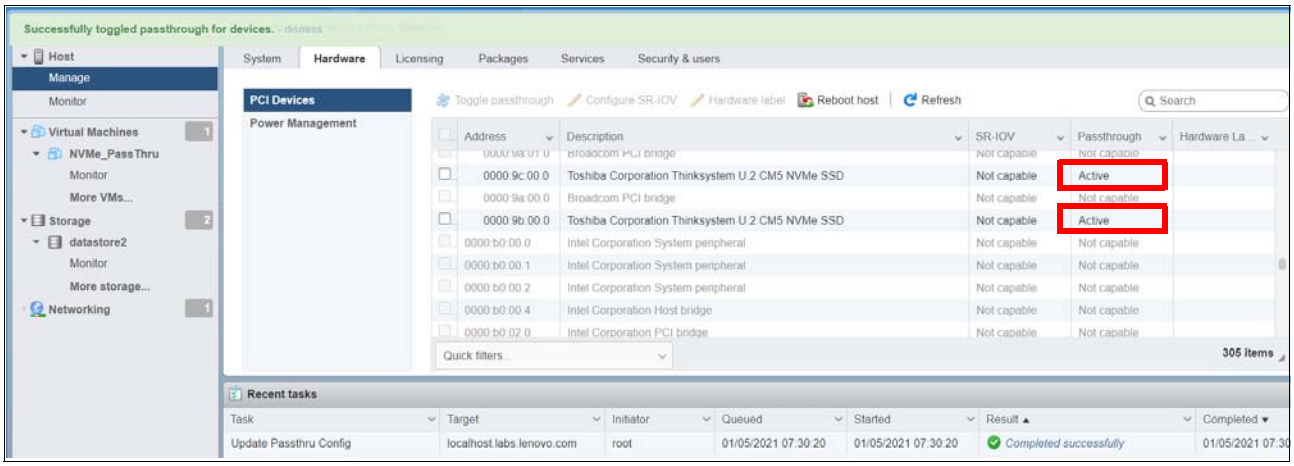


Figure 21 Activate passthrough on the NVMe SSDs

2. Add the NVMe SSDs to the VM, Figure 22.

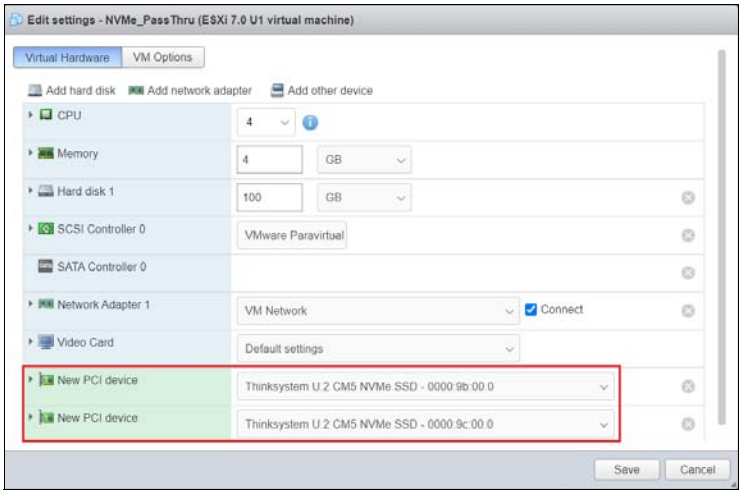


Figure 22 Add the SSDs to the VM

3. Keep the reservation size equal to the memory capacity size, Figure 23.

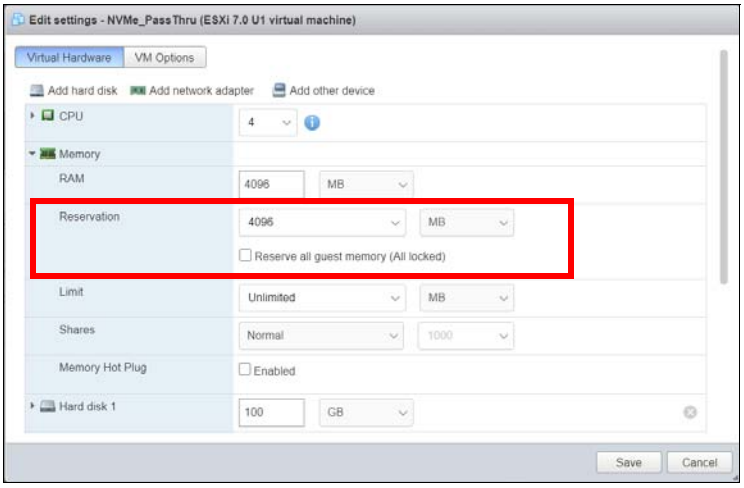


Figure 23 Memory reservation

Step 3: Check the NVMe SSDs on VM

At the end, we can find these NVMe SSDs in the VM and verify there are no errors or warnings of NVMe SSDs in the log. Use the same commands as described in “Step 3: Check the NVMe SSDs on VM” on page 11. The output is shown in Figure 24.

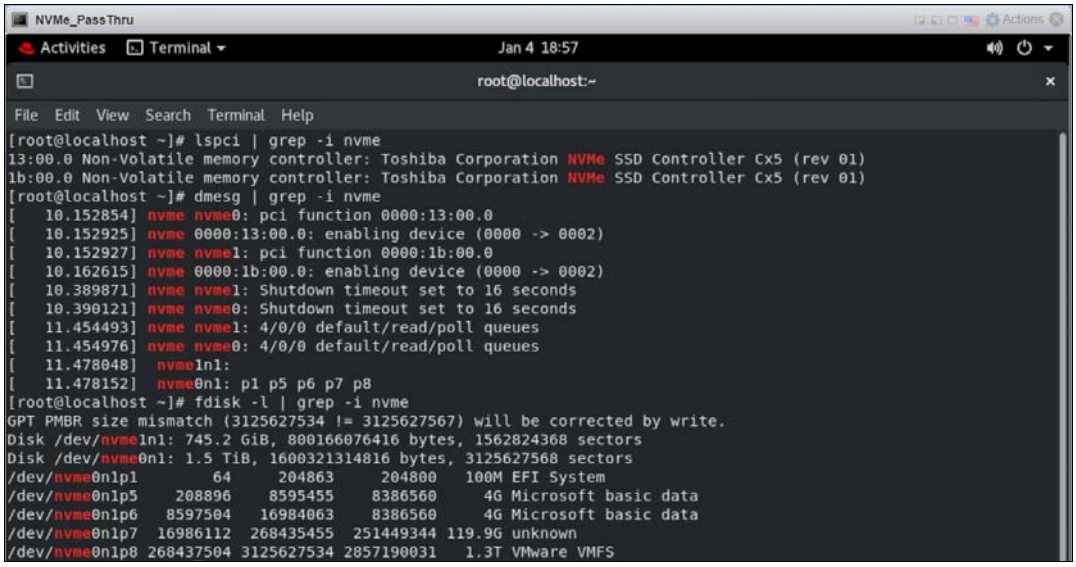


Figure 24 output from lspci, dmesg and fdisk commands

Now, the NVMe devices have been configured passthrough to the VM successfully, and you can access and use it as a real drives in the virtual OS.

References

- vSphere VMDirectPath I/O and Dynamic DirectPath I/O: Requirements for Platforms and Devices

  https://kb.vmware.com/s/article/2142307

- VMware Configuration Maximums

  https://configmax.vmware.com/

- vSphere VCG

  https://www.vmware.com/resources/compatibility/search.php

- Storage Options for ThinkSystem Servers

  https://lenovopress.com/lp0761-storage-options-for-thinksystem-servers

- NVMe-Rich Configurations of the ThinkSystem SR650

  https://lenovopress.com/lp0904-nvme-rich-configurations-of-the-sr650

Author

Boyong Li is the OS Technical Leader of the Lenovo Infrastructure Solutions Group in Beijing, China. He is an experienced software architecture, BIOS and OS engineer and is responsible for technical support for both Windows and VMware.

Thanks to the following specialists for their contributions and suggestions:

- Chengcheng Peng, Lenovo VMware Engineer
- Gary Cudak, Lenovo OS Architect and WW Technical Lead
- 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 April 14, 2021.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp1464

Trademarks