

Technical white paper

NVMe Hot-Plug on Dell EMC PowerEdge servers running VMware vSphere or vSAN

Abstract

This white paper describes the support for Non-Volatile Memory Express (NVMe) Hot-Plug on Dell EMC PowerEdge servers running VMware vSphere or vSAN.

June 2020

Revisions

Date	Description
June 2020	Initial release

Acknowledgements

Authors: Syama Poluri, Himanshu Shrivastava, Thiru Navukkarasu, Krishnaprasad K

Support: Austin Bolen, Gurupreet Kaushik, Sherry Keller, IDD team

The information in this publication is provided "as is." Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © <06/23/2020> Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.

Table of contents

Re	vision	IS	2
Ac	knowl	edgements	2
		contents	
Ex	ecutiv	re summary	4
1		oduction	
	1.1	Audience and Scope	5
	1.2	Terminology	5
	1.3	NVMe Command-Line Interface	5
2	Use	cases for NVMe and hot-plug support on VMware ESXi	6
	2.1	Support for NVMe hot-plug operations	7
3	Sup	port for NVMe hot-plug on VMware vSphere 6.x and VMware vSAN 6.x	8
	3.1	Perform orderly removal of an NVMe device configured as a VMFS Datastore	8
	3.2	Perform orderly removal of an NVMe device configured as an RDM disk	9
	3.3	Perform orderly removal of an NVMe device on a VSAN cluster	10
	3.4	Perform orderly removal of an NVMe device configured as a host cache device	10
	3.5	Perform orderly removal of an NVMe device through Dell EMC iDRAC Service Module (iSM)	11
4	LED	management of NVMe devices in VMware ESXi	12
5	Sum	nmary	13
6	Refe	erences	14

Executive summary

This white paper describes the use cases for NVMe devices on Dell EMC PowerEdge servers running VMware vSphere or VMware vSAN and supported hot-plug operations, the best practices for performing hot-plug operations on NVMe devices through each use case.

4

1 Introduction

Non-Volatile Memory Express (NVMe) is an architecture that is optimized for solid-state devices (SSD) connected directly to a PCIe bus. NVMe devices consist of a PCIe host interface, SSD controller and a non-volatile memory. Unlike SAS or SATA interfaces, NVMe does not make use of host bus adapters (HBA).

1.1 Audience and Scope

The intended audience for this white paper includes IT Administrators, Channel Partners and users using hot-pluggable NVMe devices on Dell EMC PowerEdge servers running VMware vSphere or VMware vSAN. This white paper includes instructions for users to explore the hot-plug capabilities of NVMe devices.

1.2 Terminology

Hot insertion: Connecting an NVMe device when VMware ESXi operating system is booted up.

Surprise removal: Removing a device from the system without notifying the system beforehand.

Orderly removal: Removing an NVMe device after completing the pre-requisites, such as suspending all processes accessing the device.

Hot swap: Replacement of an NVMe SSD with a new SSD (from the same or different vendor) while the host is up and running. Hot swap is a surprise removal followed by a hot insertion operation with a different NVMe device.

1.3 NVMe Command-Line Interface

This white paper showcases several hot-plug operations enabled on NVMe devices via the command-line interface. VMware provides an NVMe namespace in esxcli. For more information, see <u>vSphere Command-Line Interface Reference</u>.

5

2 Use cases for NVMe and hot-plug support on VMware ESXi

Listed below are the use cases for NVMe in VMware ESXi environment.

Virtual Machine File System (VMFS) datastore: Virtual Machine File System is used to store virtual
machine disk files and associated files. Dell EMC and VMware support the usage of NVMe devices
for hosting virtual machines with the creation of VMFS. For more information about VMFS, see
Understanding VMFS Datastores.

Note: Surprise removal of NVMe device when it is in use by the VMFS is supported by Dell EMC and VMware from VMware ESXi 7.0 onwards. However, surprise removal is not supported by VMware ESXi 6.x versions when the device is in use.

• Raw Device Mapping (RDM): RDM provides virtual machines with direct access to a LUN. Dell EMC and VMware support the usage of NVMe devices for RDM use case.

Note: Surprise removal of NVMe device when it is configured as an RDM is not supported by Dell EMC and VMWare.

VMware vSAN: VMware vSAN supports NVMe devices as a cache and also as a capacity tier. For
information about the Dell EMC vSAN certified ReadyNodes supported, see the vSAN HCL from the
VMware Compatibility Guide.

Note: Surprise removal of NVMe device when it is in use by the vSAN cache or the capacity tier is supported by Dell EMC and VMWare.

Host swap cache: The host cache is a shared swap area created for virtual machines. It is generally
created on a low latency device such as solid-state drives and is used as a write-back cache for
virtual machine swap files. Dell EMC supports usage of NVMe devices for this use case.

Note: Surprise removal of NVMe device when it is in use for host cache is supported by Dell EMC and VMWare. However, users must note that applications or guest operating systems are prone to failure depending on their usage.

PCle passthrough: Dell EMC does not support NVMe device as a passthrough device for generic
use cases. However, there are appliance use cases from Dell EMC that support this use case
explicitly.

Note: Surprise removal of NVMe device when it is configured as PCIe passthrough is not supported by Dell EMC and VMWare.

Note: Dell EMC does not support booting VMware ESXi from an NVMe device. For more information, see <u>Dell Express Flash NVMe PCle SSD User's Guide</u>.

Dell EMC does not support using an NVMe device for PCIe passthrough or VMDirectPath I/O or for Dynamic Direct Path I/O use cases. For more information, see VMware Knowledge Base article 2142307.

2.1 Support for NVMe hot-plug operations

The following table describes the respective version of VMware ESXi and Dell EMC Generations of PowerEdge servers that support hot-plug operations on NVMe devices.

Table 1 Supported NVMe hot-plug operations

Dell EMC PowerEdge	Dell EMC customized VMware ESXi 6.x		Dell EMC customized VMware ESXi 7.0 GA (A00 Build# 15843807) *		Dell EMC customized VMware ESXi 7.0 (A02 Build# 16324942)		
generation	Supported	Unsupported	Supported	Unsupported	Supported		
Intel 14G	Hot insertion Orderly removal	Surprise removal	Hot insertion Orderly removal	Surprise removal	 Hot insertion Surprise removal Orderly removal		
AMD 14G Naples	Hot insertionOrderly removal	Surprise removal	Hot insertion Orderly removal	Surprise removal	Hot insertion Surprise removal Orderly removal		
AMD 15G Rome	Hot insertion** Orderly removal**	Surprise removal**	Hot insertion Orderly removal	Surprise removal	Hot insertion Surprise removal Orderly removal		

^{*} Dell EMC does not support NVMe surprise removal for VMware ESXi 7.0 GA. For more information see, VMware Knowledge Base article 78390 and VMware vSphere 7.x on Dell EMC PowerEdge Servers Release Notes. However, VMware have announced support for hot removal in VMware ESXi 7.0 GA. For more information, see VMware Knowledge Base article 78297.

^{**} Dell EMC PowerEdge servers with AMD Rome processors are required to have both interrupt remapping and DMA Remapping disabled to support NVMe hot-plug on VMware vSphere and vSAN 6.x versions explicitly. This restriction is lifted from VMware vSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards. For more information, see VMware VMware VSphere 7.0 onwards.

3 Support for NVMe hot-plug on VMware vSphere 6.x and VMware vSAN 6.x

VMware introduced support for surprise removal on NVMe devices from VMware ESXi 7.0 onwards. Surprise removal of NVMe is not supported in the previous versions of VMware ESXi when the NVMe device is in use. For more infomration, see VMware Knowledge Base article 2151404. Surprise removal of NVMe devices when they are in use is not supported by Dell EMC and VMware. The user should instead perform orderly removal. The following section describes the directions for performing an orderly removal of an NVMe device in the various VMware use cases.

3.1 Perform orderly removal of an NVMe device configured as a VMFS Datastore

Follow the steps:

- 1. Identify the virtual machines or processes which are accessing the NVMe device. Use the command: esxcli storage core device world list to see all processes and use filter to find the specific NVMe device.
- Use the following command to set the autoclaim rescan to false to prevent devices from being redetected:

esxcli storage core claiming autoclaim -enabled=false

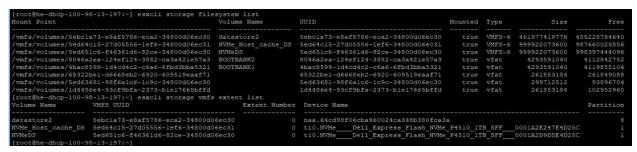


Figure 1 Checking the status of the filesystem mount

- Check the status of the filesystem mount using the command: esxcli storage filesystem list
- 4. Power off any virtual machine using the VMFS Datastore. Orderly removal will fail if any virtual machine is using the VMFS Datastore.

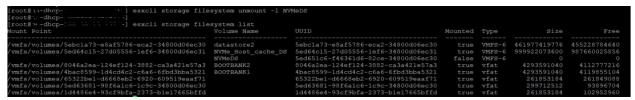


Figure 2 Unmount the VMFS Datastores on the NVMe device

- 5. Unmount the VMFS Datastores on the NVMe device using the command: esxcli storage filesystem unmount -1 NVMeDS
- 6. Detach the NVMe device using the command:

 esxcli storage core device set --state=off -d <nvme device name>

7. Remove the NVMe device physically.

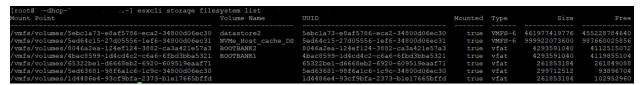


Figure 3 Check VMFS Datastore is not connected to the ESXi

8. Make sure the VMFS Datastore is not connected to the ESXi using the command: esxcli storage filesystem list

The NVMe device has been safely removed from VMware ESXi and the server. It is now safe to replace the NVMe device with a different device or hot insert the same NVMe device if required. When performing hot insert to an NVMe device, follow the attached operations and set autoclaim to true. For more information, see VMware Knowledge Base article 2151404.

3.2 Perform orderly removal of an NVMe device configured as an RDM disk

Follow the steps:

- Check the NVMe device usage using the command: esxcli storage core device world list.
- 2. The output must show the world or process name for the specific device name.

Figure 4 Check the world or process name for the NVMe device name

- 3. Perform orderly removal of the NVMe device by first unmapping the NVMe device or power off the virtual machine itself.
- 4. Perform hot remove on the NVMe device and check if the hot-plug slot is disabled. Use the lspci command to check if the NVMe device exists in ESXi.

If there are no references to the specific NVMe device, then the NVMe device has been successfully removed.

Note: If a different NVMe device is inserted back into the system after the hot removal, then the RDM link configured to the virtual machine will not be available automatically. The user needs to re-create RDM and point to the VM manually.

3.3 Perform orderly removal of an NVMe device on a VSAN cluster

VMware vSAN is a software-based storage solution for VMware hyperconverged solutions. Embedded in the hypervisor and leveraging industry-standard servers, vSAN delivers flash-optimized, high-performance computation and storage for hyperconverged infrastructures. vSAN can make use of NVMe devices in both cache and capacity tier.

Dell EMC has vSAN readynodes that support NVMe devices as both a cache tier and a capacity tier providing every available NVMe solution. For more information on the vSAN readynodes supported by Dell EMC, see VMware Compatibility Guide

Listed below are recommendations to perform orderly removal of an NVMe device from a vSAN cluster:

- If the NVMe device that is to be removed by hot removal is in the cache tier, the overall process depends on the use case. Ensure to evacuate the data before removing the disk as the diskgroup will be earased when the NVMe device is hot removed.
- If the NVMe device that is to be removed by hot removal is in the capacity tier, ensure that the device
 is accessed carefully as it depends upon FTT (Failure to Tolerate) and other parameters. For more
 information on general recommendations for removing a disk from vSAN host, see <u>VMware</u>
 <u>Knowledge Base article 2150593</u>.

3.4 Perform orderly removal of an NVMe device configured as a host cache device

Follow the steps:

 Check the NVMe device usage. Go to vCenter Server > Host > Configure. Under the Storage tab, select Host Cache Configuration. The NVMe device utilized as a host cache is displayed.

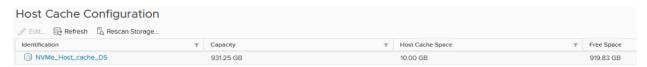


Figure 5 Identifying the NVMe device usage

Edit the host cache configuration for the NVMe device to set the host cache space to zero bytes. This ensures no space from NVMe datastore is used as host cache space.

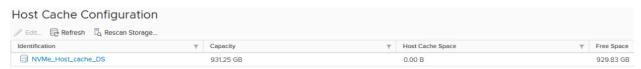


Figure 6 Host Cache Space set to zero.

3. Unmount the datastore. This can be done on the GUI or the CLI using the command: esxcli storage filesystem unmount -1 NVMe_Host_cache_DS

[root@he-dhcp-100-98-13-197:~] esxcli storage filesystem unmount -1 NVMe_Host_cache_DS [root@he-dhcp-100-98-13-197:~] esxcli storage filesystem list [root@he-dhcp-100-98-13-197:~] esxcli storage filesystem list									
Mount Point	Volume Name	UUID	Mounted	Type		Free			
/vmfs/volumes/5ebc1a73-e8af5786-eca2-34800d06ec30	datastore2	5ebc1a73-e8af5786-eca2-34800d06ec30	true	VMFS-6	461977419776	455228784640			
/vmfs/volumes/5ed651c6-f46361d6-82ce-34800d06ec30	NVMeDS	5ed651c6-f46361d6-82ce-34800d06ec30		VMFS-6	999922073600	998397444096			
	NVMe Host cache DS	5ed64c15-27d05556-1ef6-34800d06ec31	false	VMFS-6		0			
/vmfs/volumes/8046a2ea-124ef124-3882-ca3a421e57a3	BOOTBANK2	8046a2ea-124ef124-3882-ca3a421e57a3	true	vfat	4293591040	4112252928			
/vmfs/volumes/4bac8599-1d4cd4c2-c6a6-6fbd3bba5321	BOOTBANK1	4bac8599-1d4cd4c2-c6a6-6fbd3bba5321	true	vfat	4293591040	4119855104			
/vmfs/volumes/65322be1-d6668eb2-6920-609519eaaf71		65322be1-d6668eb2-6920-609519eaaf71		vfat	261853184	261849088			
/vmfs/volumes/5ed63681-98f6a1c6-1c9c-34800d06ec30		5ed63681-98f6a1c6-1c9c-34800d06ec30		vfat	299712512	93896704			
/vmfs/volumes/1d4486e4-93cf9bfa-2373-b1e17665bffd		1d4486e4-93cf9bfa-2373-b1e17665bffd		vfat	261853184	102952960			

Figure 7 Unmount the datastore

4. Perform hot removal of the NVMe device and check if the hot-plug slot is disabled. Use the lspci command to check if the NVMe device exists in ESXi.

If there are no references to the specific NVMe device, then the NVMe device has been successfully removed.

3.5 Perform orderly removal of an NVMe device through Dell EMC iDRAC Service Module (iSM)

Dell EMC iDRAC Service Module (iSM) can be used to remove Non-Volatile Memory Express (NVMe) Pheripheral Component Interconnect Express (PCIE) Solid State Device (SSD) without powering off or restarting the system.

When hot removal is performed, all the activities associated with the device must be stopped to prevent data loss which can be achieved using the **Prepare to Remove** option from the **Configuration** tab in the iDRAC. This allows the NVMe PCIe SSD to be safely removed.

Follow the steps:

- Check the NVMe device usage from the ESXi using the command: esxcli storage core device world list
- 2. List the VMFS extents available on the host using the command: esxcli storage vmfs extent list
- 3. Make sure you have iSM installed on VMware ESXi. Use the command: esxcli software vib list | grep dcism
- 4. Click on the **Prepare to Remove** option for the NVMe device to be removed. Go to **iDRAC** > **Configuration** > **Controller (CPU number)** > Select the NVMe device from the **drop-down** menu.

Once the iDRAC notifies that the NVMe device is successfully removed, the device will no longer be utilized by VMware ESXi. The device can be physically removed as it is not used by the filesystem.

4 LED management of NVMe devices in VMware ESXi

Starting from ESXi 7.0 patch build# 16324942, Dell EMC and VMware support LED management of NVMe devices in VMware ESXi. Dell EMC and VMware provide an esxcli namespace to monitor the LED indications of NVMe devices. For more information, see <u>Dell Knowledge Base article SLN321848</u>.

5 Summary

This white paper is an exclusive documentation on NVMe hot-plug support for Dell EMC Generations of PowerEdge servers in VMware vSphere and vSAN environments and describes use cases of NVMe device(s). This document will be updated progressively when there is more information on the support stances, known issues and best practices for hot plug insertion or hot plug removal techniques. Details of NVMe hot-plug support on Dell EMC 14th Generation of PowerEdge servers and above are discussed in this white paper.

6 References

- NVMe Hot-Plug Challenges and Industry Adoption
- Dell Express Flash NVMe PCle SSD User's Guide
- vSphere Support for hot-swap of local disks
- vSphere and vSAN support for Hot-plug of NVMe SSDs on AMD EPYC Processors
- Supported scenarios of NVME hot plug or removal in ESXi 7.0
- Dell EMC VMware ESXi 7.x Documentation