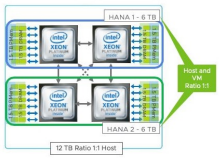
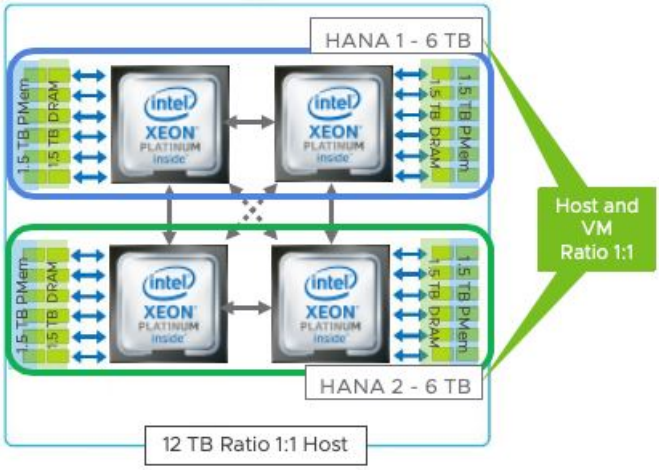


intel Optane Persistent Memory and SAP HANA Platform Configuration on VMware ESXi User Guide

Contents

intel Optane Persistent Memory and SAP HANA Platform Configuration on VMware ESXi



overview

Technology overview and deployment guidelines for using Intel Optane persistent memory with the SAP HANA platform on VMware ESXi.

This document aims to provide an update to the existing Intel and SAP co-publication, “Configuration Guide: Intel® Optane™ Persistent Memory and SAP HANA® Platform Configuration,” available online at intel.com/content/www/us/en/big-data/partners/sap/sap-hana-and-intel-optane-configuration-guide.html. This update will discuss the additional procedures needed to configure SAP HANA with Intel Optane persistent memory (PMem) running on a VMware ESXi virtual machine (VM).

In the existing guide, the operating system (OS)—either SUSE Linux Enterprise Server (SLES) or Red Hat Enterprise Linux (RHEL)—runs directly on bare metal or as the host OS in a non-virtualized setup. The steps to deploy SAP HANA with Intel Optane PMem in this non-virtualized server (which start on page 7 of the existing guide) are outlined as follows:

General steps

General steps: Configure Intel Optane PMem for SAP HANA

1. Install the management utilities.
2. Create App Direct regions (goal)—use interleaving.
3. Reboot the server—required to enable new configuration.
4. Create App Direct namespaces.
5. Create a file system on the namespace device.
6. Configure SAP HANA to use the persistent memory file system.
7. Restart SAP HANA to activate and start using Intel Optane PMem.

For deployment in a virtualized environment, this guide groups the steps for configuration of each component as follows:

Host:

1. Configure the server host for Intel Optane PMem using BIOS (vendor-specific).
2. Create App Direct interleaved regions, and verify that they are configured for VMware ESXi use.

VM:

3. Create a VM with hardware version 19 (VMware vSphere 7.0 U2) with NVDIMMs, and allow failover to another host while doing this.
4. Edit the VMX VM configuration file and make the NVDIMMs non-uniform memory access (NUMA)-aware.

OS:

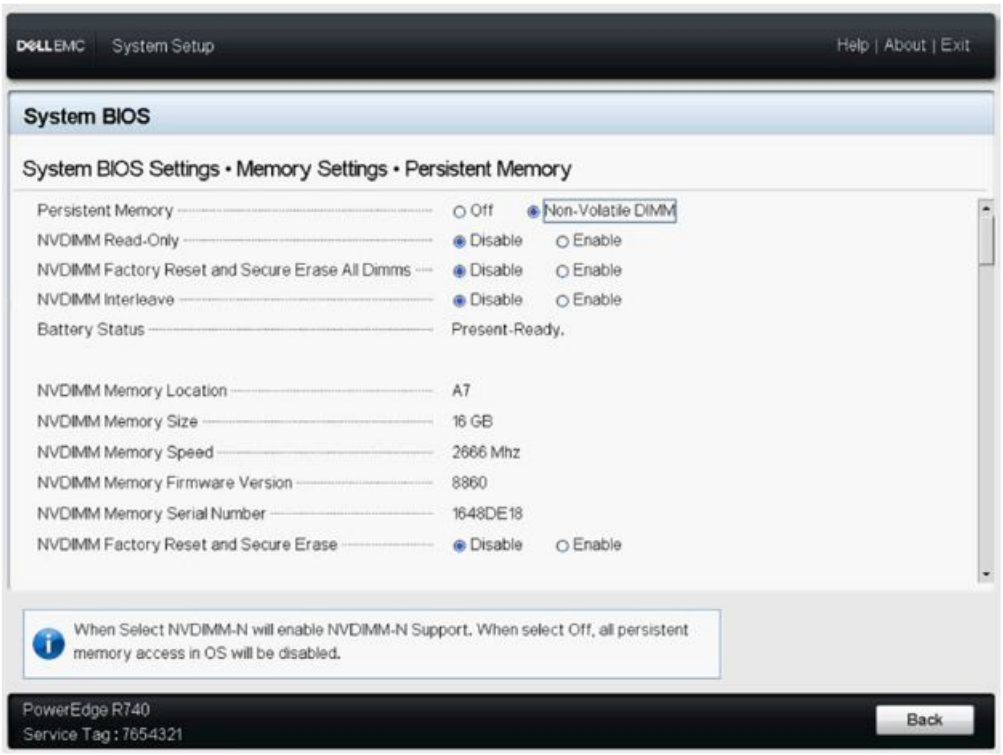
5. Create a file system on the namespace (DAX) devices in the OS.
6. Configure SAP HANA to use the persistent memory file system.
7. Restart SAP HANA to activate and start using Intel Optane PMem.

Notice that steps 5–7 for the OS configuration are identical to the existing guide, except that they are now applied to a guest OS deployment. This guide will therefore focus on steps 1–4 and the differences from a bare-metal installation.

Configure server host for Intel Optane PMem using BIOS

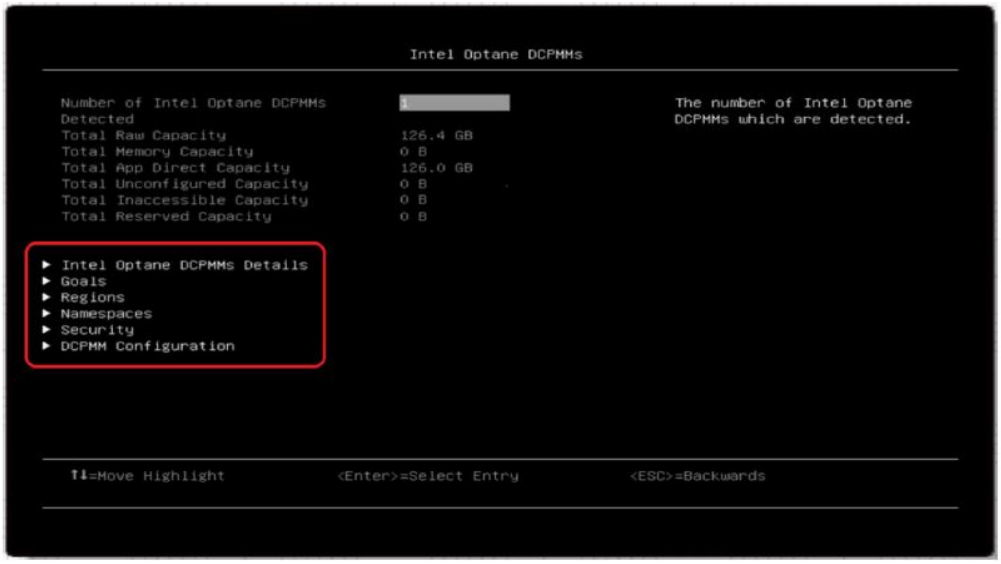
At the time of publication of the existing guide, the prescribed management utilities, ipmctl and ndctl, were mainly command-line interface (CLI)-based. Since then, newer systems produced by various OEM vendors have more widely adopted a graphical menu-driven user interface (UI) built-in to their Unified Extensible Firmware Interface (UEFI) or BIOS services. Each OEM has freely designed its UI to conform to its own style and framework of built-in utilities and controls.

As a result, the exact steps needed to configure Intel Optane PMem for every system will vary. Some examples of the Intel Optane PMem configuration screens from various OEM vendors are shown here to provide an idea of what these screens might look like and to illustrate the potential variety of UI styles that might be encountered.


DELL EMC System Setup
Help | About | Exit
System BIOS
System BIOS Settings • Memory Settings • Persistent Memory
Persistent Memory
Off
Non-Volatile DIMM
NVDIMM Read-Only
Disable
Enable
NVDIMM Factory Reset and Secure Erase All Dimms
Disable
Enable
NVDIMM Interleave
Disable
Enable
Battery Status
Present-Ready.
NVDIMM Memory Location
A7
NVDIMM Memory Size
16 GB
NVDIMM Memory Speed
2666 Mhz
NVDIMM Memory Firmware Version
8860
NVDIMM Memory Serial Number
1648DE18
NVDIMM Factory Reset and Secure Erase
Disable
Enable
When Select NVDIMM-N will enable NVDIMM-N Support. When select Off, all persistent memory access in OS will be disabled.
PowerEdge R740
Service Tag : 7654321
Back


System Configuration and Boot Management
Select Language
[English]
Launch Graphical System Setup
System Information
System Settings
Date and Time
Start Options
Boot Manager
BMC Settings
System Event Logs
User Security
Save Settings
Discard Settings
Load Default Settings
Exit Setup Utility
Display or modify system settings. Changes may not take effect immediately. Save any changed settings and reboot the system.
↑↓=Move Highlight
<Enter>=Select Entry
<ESC>=Exit Setup Utility


System Settings
<F1> Start Control
[Text Setup]
Devices and I/O Ports
Driver Health
Intel Optane DCPMMs
Legacy BIOS
Memory
Network
Operating Modes
Power
Processors
Recovery and RAS
Security
Storage
View and configure Intel Optane DCPMMs.
↑↓=Move Highlight
<Enter>=Select Entry
<ESC>=Backwards



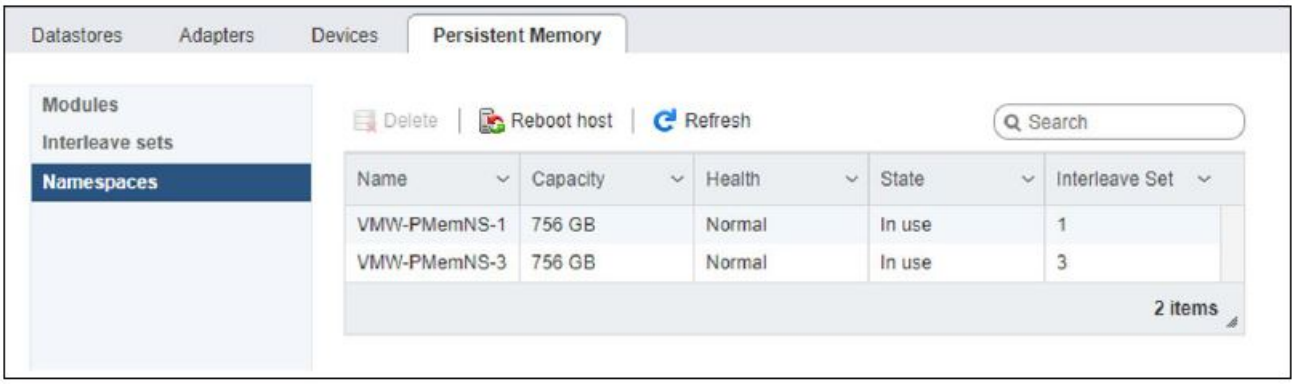
Regardless of the UI style differences, the goal of provisioning Intel Optane PMem to create App Direct mode regions remains the same for both bare-metal and virtualized use cases like VMware ESXi. Previous steps that were performed using a CLI are simply replaced by a menu-driven or form-style UI procedure to get the same end result. That is, to create interleaved App Direct regions across all the sockets that have Intel Optane PMem installed.

To help navigate through this process more easily, the following table provides links to the latest documentation and guides published by some of the top-tier OEM vendors for SAP HANA. Follow the steps from these guides to create interleaved App Direct regions for each socket, and then complete the process with a reboot of the system to enable the new configuration. Consult your OEM technical team or Intel support with any questions.

OEM vendor	**Intel Optane PMem configuration guide/document**	**Online link**
Cisco	“Cisco UCS: Configuring and Managing Intel® Optane™ Data Center Persistent Memory Modules”	cisco.com/c/en/us/td/docs/unified_computing/ucs/persistent- memory/b_Configuring_Managing_DC-Persistent-Memory- Modules.pdf
Dell Technologies	“Dell EMC NVDIMM-N Persistent Memory User Guide” (Intel Optane PMem 100 series)	https://dl.dell.com/topicspdf/nvdimm_n_user_guide_en-us.pdf
Dell Technologies	“Dell EMC PMem 200 Series User Guide”	https://dl.dell.com/topicspdf/pmem_15g_en-us.pdf
Fujitsu	“DCPMM (Data Center Persistent Memory) Command-Line Interface”	https://ssl.syncsearch.jp/click?url=https%3A%2F%2Fsupport.ts.fujitsu.com%2FSearch%2FSWP1235322.asp&query=dcpmm&site=7215VAWV
Fujitsu	“Configure DCPMM (Data Center Persistent Memory) in UEFI Setup”	https://ssl.syncsearch.jp/click?url=https%3A%2F%2Fsupport.ts.fujitsu.com%2FSearch%2FSWP1235339.asp&query=dcpmm&site=7215VAWV
Fujitsu	“Configure DCPMM (Data Center Persistent Memory) on Linux”	https://ssl.syncsearch.jp/click?url=https%3A%2F%2Fsupport.ts.fujitsu.com%2FSearch%2FSWP1235054.asp&query=dcpmm&site=7215VAWV

OEM vendor	Intel Optane PMem configuration guide/document	Online link
HPE	HPE Persistent Memory User Guide for HPE ProLiant Gen10 servers and HPE Synergy”	http://itdoc.hitachi.co.jp/manuals/ha8000v/hard/Gen10/ DCPMM/P16877-002_en.pdf
HPE	“Intel Optane persistent memory 100 series for HPE User Guide”	https://support.hpe.com/hpesc/public/ docDisplay?docId=a00074717en_us
Lenovo	“How to change the Intel® Optane™ DC Persistent Memory Module operating modes through UEFI”	https://datacentersupport.lenovo.com/us/en/products/ servers/thinksystem/sr570/7y02/solutions/ht508257- how-to-change-the-intel-optane-dc-persistent-memory- module-operating-modes-through-uefi
Lenovo	“Enabling Intel Optane DC Persistent Memory on Lenovo ThinkSystem Servers”	https://lenovopress.com/lp1167.pdf
Lenovo	“Implementing Intel Optane DC Persistent Memory with VMware vSphere”	https://lenovopress.com/lp1225.pdf
Supermicro	“Intel 1st Gen DCPMM Memory Configuration for the Intel Purley Platform”	https://www.supermicro.com/support/resources/memory/ DCPMM_1stGen_memory_config_purley.pdf
Supermicro	“Intel® Optane™ Persistent Memory 200 Series Configuration for Supermicro X12SPx/X12Dxx/ X12Qxx Motherboards”	https://www.supermicro.com/support/resources/memory/ Optane_PMem_200_Series_Config_X12QP_DP_UP.pdf

Create App Direct interleaved regions and verify their configuration for VMware ESXi use
The OEM UEFI or BIOS menus typically provide UI screens to confirm that the App Direct regions have been created for each socket. With VMware, you can also use the web client or the esxcli command to verify this. From the web client, go to Storage, and then select the Persistent Memory tab.



As you will see, a default namespace is automatically created per region. (This example is for a two-socket system.) For esxcli, you can use the following command:

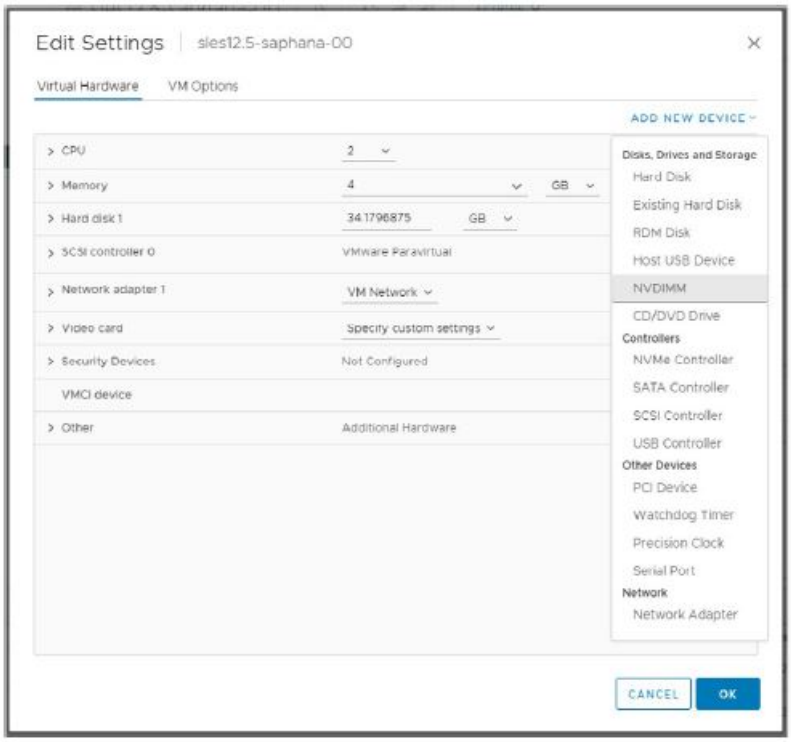
```
[root@localhost:~] localcli --plugin-dir /usr/lib/vmware/esxcli/int hardwareinternal nvd namespace details list

Namespace VMW-PMemNS-1:
   UUID of Namespace: 45591b24-2132-5427-be9e-46a04a52aff3
   Name of Namespace: VMW-PMemNS-1
   Namespace Size: 811748818944
   Namespace State: InUse
   Health summary: OK
   Interleave set ID: 0x1

Namespace VMW-PMemNS-3:
   UUID of Namespace: 74510d89-5920-5a64-8d77-e67bfb49ed5c
   Name of Namespace: VMW-PMemNS-3
   Namespace Size: 811748818944
   Namespace State: InUse
   Health summary: OK
   Interleave set ID: 0x3
```

Create a VM with hardware version 19 (VMware vSphere 7.0 U2) with NVDIMMs, and allow failover to another host
Deploy a VM with a supported guest OS (SLES or RHEL for SAP HANA) and SAP HANA 2.0 SPS 04 or greater installed
There are multiple ways to provision and deploy vSphere VMs. These techniques are best described and covered by VMware's online document library at "VMware vSphere—Deploying Virtual Machines" (**https://docs.vmware.com/en/VMware-vSphere/7.0/com.vmware.vsphere.vm_admin.doc/GUID-39D19B2B-A11C-42AE-AC80-DDA8682AB42C.html**).

To choose the best method for your environment, you will need to create a VM with the appropriate supported OS and install SAP HANA on it like you would on a physical (bare-metal) server.
Create App Direct namespaces on the deployed VM by adding Intel Optane PMem (NVDIMM) devices

Once the VM is deployed, the Intel Optane PMem devices should be added. Before you are able to add NVDIMMs to the VM, check if the Intel Optane PMem regions and namespaces were created correctly in the BIOS. Ensure that you have selected all Intel Optane PMem (100%). Also ensure that Persistent memory type is set to App Direct Interleaved. Memory Mode should be set to 0%.

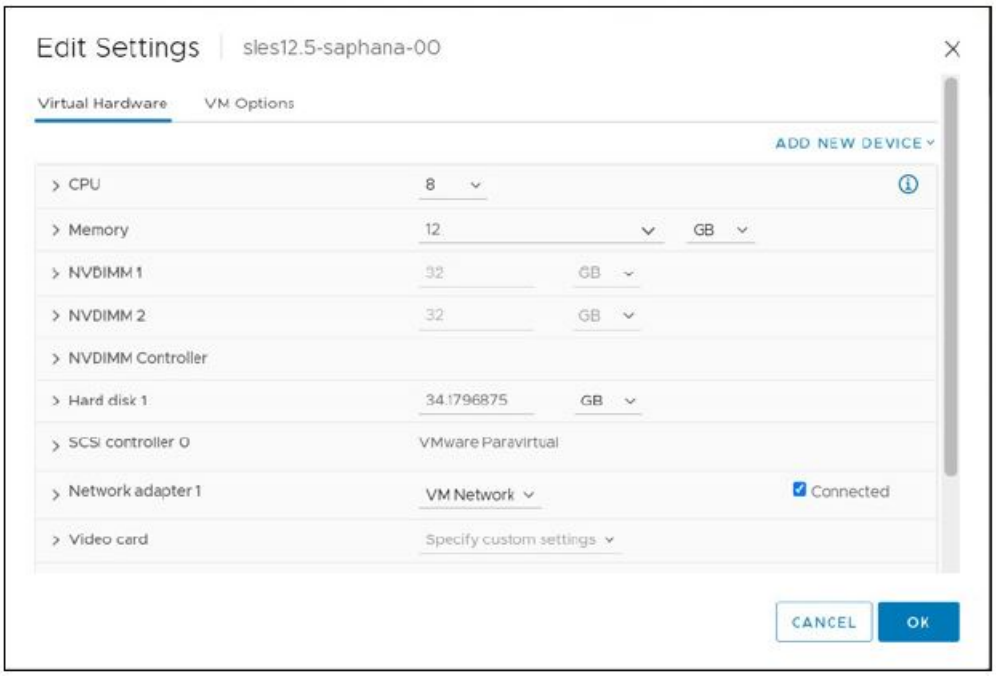
Power off the VM, and then edit the VM settings by using the Add new device option and selecting NVDIMM. The standard practice is to create one NVDIMM device per host CPU socket. Refer to a best practices guide from your OEM if available.
This step will also automatically create the namespaces.



Edit NVDIMMs size as needed, and then select Allow failover on another host for all NVDIMM devices.

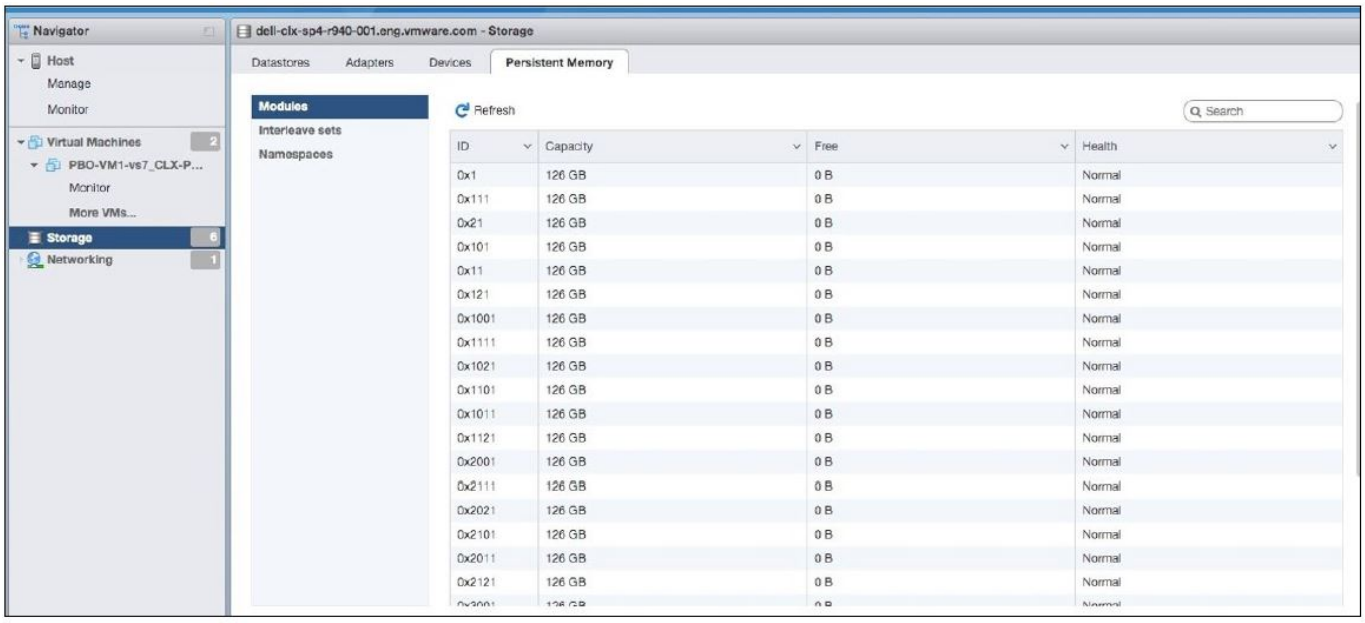
If there is no NVDIMM device listed, try to upgrade VM compatibility. Select the VM, select Actions > Compatibility > Upgrade VM Compatibility, and ensure that the VM is compatible with ESXI 7.0 U2 and later.

After adding the NVDIMM devices successfully, your VM configuration settings should look like this:

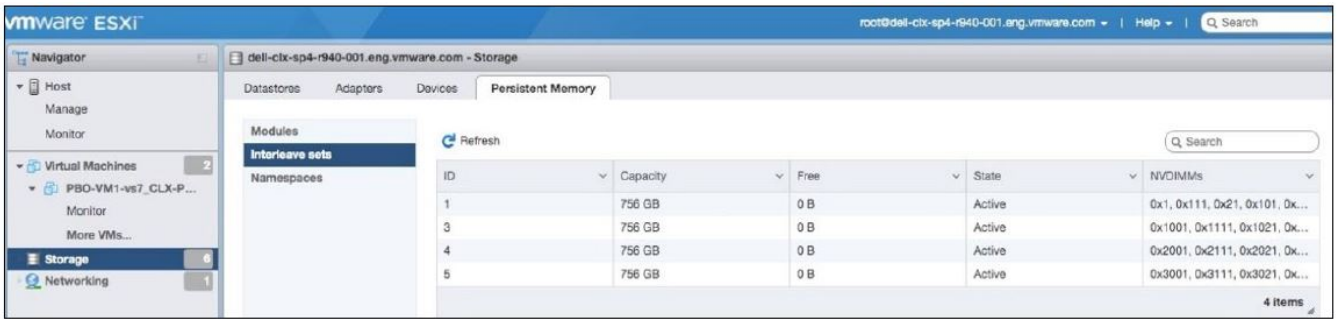


If the configurations were done correctly, the VMware ESXi Intel Optane PMem storage views should look like the following figures.

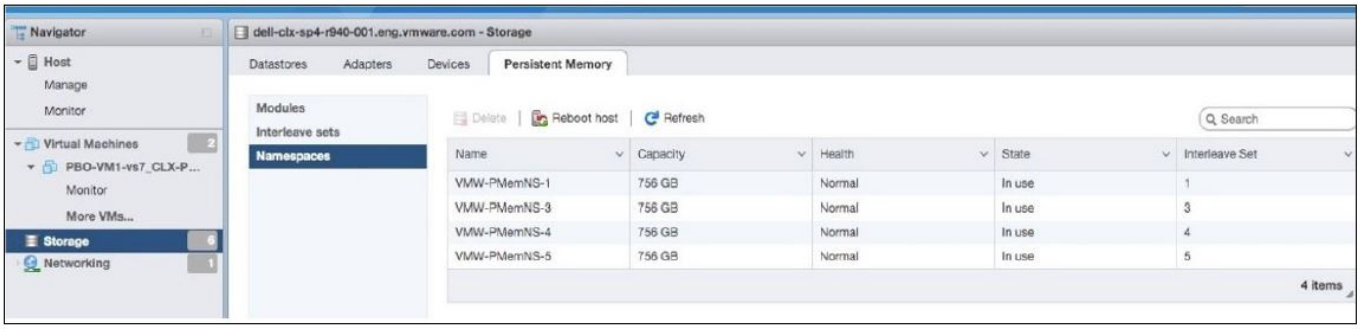
VMware ESXi Intel Optane PMem storage view—modules



VMware ESXi Intel Optane PMem storage view—interleave sets

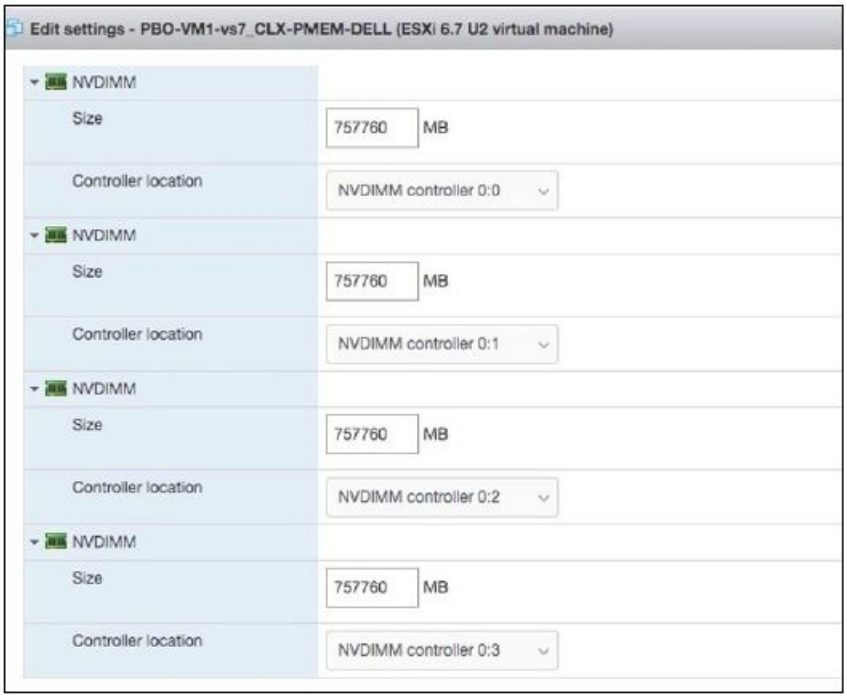


VMware ESXi PMem storage view—namespaces



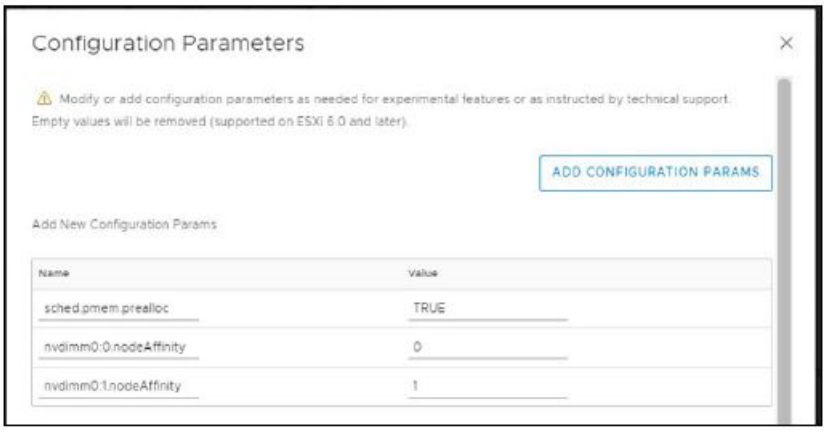
Note: Shown interleave set numbers depend on the hardware configuration and might be different for your system.
Next, you can add NVDIMMs and NVDIMM controllers to your SAP HANA VM. To use all available memory in your system, select the maximum size possible per NVDIMM.

NVDIMM creation via VMware vCenter graphical user interface

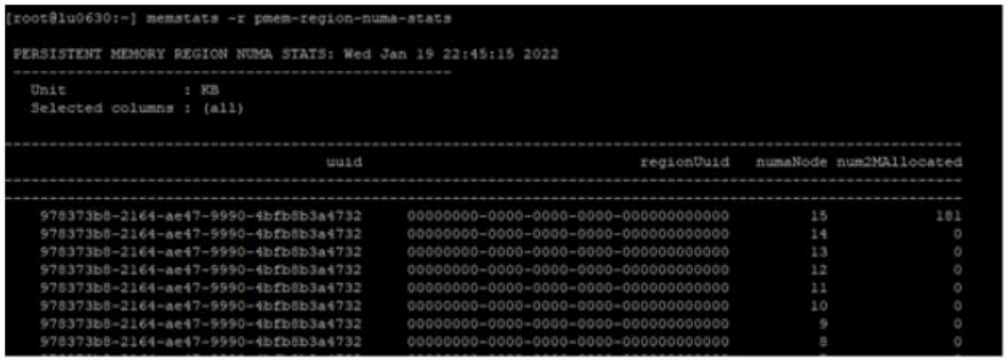


Edit the VMX VM configuration file and make the NVDIMMs NUMA-aware
By default, Intel Optane PMem allocation in VMkernel for VM NVDIMMs does not consider NUMA. This can result in the VM and the allocated Intel Optane PMem running in different NUMA nodes, which will cause NVDIMMs' access in the VM to be remote, resulting in poor performance. To avoid this, you must add the following settings to a VM configuration using VMware vCenter
(more details about this step can be found in VMware KB 78094).
In the Edit settings window, select the VM Options tab, and then click Advanced.
In the Configuration Parameters section, click Edit configuration, select the Add Configuration Params option, and enter the following values:

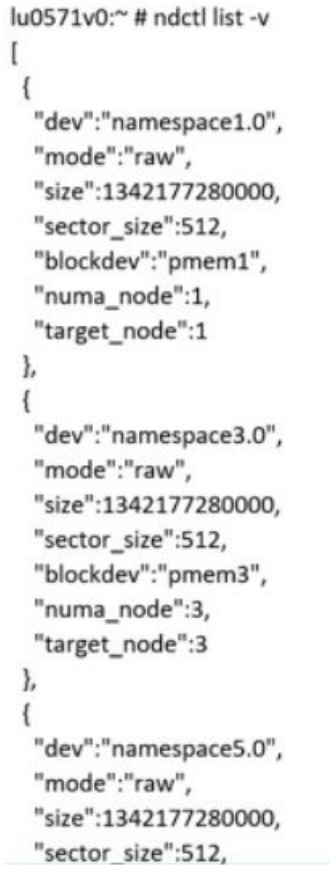
```
Key Value
-----------------------------------------
nvdimm.mode = “independent-persistent”
sched.pmem.prealloc TRUE
nvdimm0:0.nodeAffinity 0
nvdimm0:1.nodeAffinity 1
nvdimm0:2.nodeAffinity 2
nvdimm0:3.nodeAffinity 3
```



To verify that the Intel Optane PMem region allocation is distributed across NUMA nodes, use the following VMware ESXi command:
memstats -r pmem-region-numa-stats



Create a file system on the namespace (DAX) devices in the OS
To complete the configuration process, proceed to steps 5–7 of the bare-metal configuration guide, starting on page 13. These steps describe how to complete the OS configuration.
Just as in the case of a bare-metal server configuration, restarting the VM after the last step, Set the SAP HANA Base Path, will activate Intel Optane PMem for SAP HANA use.
You can check whether NVDIMMs devices are mounted properly using the following ndctl command:



Set namespaces to “fsdax” mode
You might have noticed at this point that the namespaces created were in “raw” mode. In order to be properly used by SAP HANA, they need to be converted to “fsdax” mode. Use the following command to perform this:
ndctl create-namespace -f -e <namespace> –mode=fsdax
Remounting the App Direct namespaces and file systems after the VM reboots
VMware enabled high-availability (HA) functionality in vSphere 7.0 U2 for Intel Optane PMem–enabled SAP HANA VMs.1 However, to ensure complete data transfer, additional steps are needed to prepare Intel Optane PMem for SAP HANA use so that it can automatically reload the data from shared (conventional) storage after the failover.

The same steps can be applied to remount the App Direct namespaces and file systems every time a VM reboots or is migrated. Refer to “Implementing High Availability in VMware vSphere 7.0 U2 for SAP HANA with Intel® Optane™ Persistent Memory” (**intel.in/content/www/in/en/architecture-and-technology/vmware-vsphere-ha-sap-hana-optane-pmem.html**) for more details.

solutions

Why deploy SAP HANA on VMware solutions?
VMware has had SAP HANA production support since 2014 and non-production support since 2012.

Superior scalability for x86 on-premises hypervisors for SAP HANA

- Host support for up to 768 logical CPUs and 16 TB RAM
- SAP HANA scale-up capabilities support up to eight socket-wide VMs with 448 vCPUs and 12 TB RAM
- SAP HANA scale-out capabilities support up to 32 TB
- Virtual SAP HANA and SAP NetWeaver® performance deviation of a single VM to bare-metal systems certified to pass SAP standards
- Full SAP HANA workload-based sizing support

- On roadmap: 18 TB Intel Optane PMem SAP HANA systems

Broadest Intel x86 hardware and vendor support for SAP HANA

- Support for all major Intel CPUs:
  - Intel Xeon processor v3 family (Haswell)
  - Intel Xeon processor v4 family (Broadwell)
  - 1st Generation Intel Xeon Scalable processors (Skylake)
  - 2nd Generation Intel Xeon Scalable processors (Cascade Lake)
  - 3rd Generation Intel Xeon Scalable processors (Cooper Lake)
  - 3rd Generation Intel Xeon Scalable processors (Ice Lake, in progress)
  - 4th Generation Intel Xeon Scalable processors (Sapphire Rapids, in progress)
- Support for 2-, 4-, and 8-socket server systems
- Full Intel Optane PMem support
- Support for vSphere from all major SAP hardware partners, both for on-premises implementations and in the cloud

Appendix

Optional step: Enable ipmctl in the UEFI shell
In the absence of a BIOS menu system to configure Intel Optane PMem, the UEFI CLI can still be used to configure a system for the use of SAP HANA running on VMware ESXi. To execute the equivalent of step 1 above, a UEFI shell can be enabled at boot time to run the ipmctl management utility from the CLI:

1. Create a bootable UEFI shell USB flash drive with the FAT32 file system.
   Note: Some system vendors provide a boot option to enter the UEFI shell from their start-up menu, in which case you have the option to not have to make the USB flash drive bootable or to use another storage device accessible from the UEFI shell. Consult your specific documentation or support resource for details.
2. Copy the UEFI executable file ipmctl.efi from the Intel Optane PMem firmware package to the flash drive (or other storage device chosen). Once again, your system vendor will provide the Intel Optane PMem firmware package for your system.
3. Boot your system to enter the UEFI shell.
   For a bootable USB flash drive, the typical steps would be:
   - Plug in the USB flash drive into an open USB port on the host and turn it on.
   - Enter the Boot menu to display all bootable sources.
   - Select the bootable UEFI shell USB flash drive.
4. Select the file system of your drive and navigate to the path where the impctl.efi file was copied.
   For bootable USB flash drives, often the file system is FS0, but it can vary, so try FS0, FS1, FS2, and so on.

```
Shell> FS3:
FS3:\> ls
Directory of: FS3:\
01/19/2019  00:28           1,543,360  ipmctl.efi
          1 File(s)   1,543,360 bytes
          0 Dir(s)
FS3:\> _
```

5. Execute ipmctl.efi help to list all available commands. For additional information, refer to “IPMCTL User Guide.”
   Create App Direct regions

Use the Create Goal command to create an interleaved region configured for App Direct Mode:

ipmctl.efi create -goal PersistentMemoryType=AppDirect

```
FS3:\> ipmctl.efi create -goal PersistentMemoryType=AppDirect
The following configuration will be applied:
 SocketID | DimmID | MemorySize | AppDirect1Size | AppDirect2Size
==========================================================================
 0x0000   | 0x0011 | 0 B        | 126.0 GiB      | 0 B
 0x0000   | 0x0021 | 0 B        | 126.0 GiB      | 0 B
 0x0000   | 0x0001 | 0 B        | 126.0 GiB      | 0 B
 0x0000   | 0x0111 | 0 B        | 126.0 GiB      | 0 B
 0x0000   | 0x0121 | 0 B        | 126.0 GiB      | 0 B
 0x0000   | 0x0101 | 0 B        | 126.0 GiB      | 0 B
 0x0001   | 0x1011 | 0 B        | 126.0 GiB      | 0 B
 0x0001   | 0x1021 | 0 B        | 126.0 GiB      | 0 B
 0x0001   | 0x1001 | 0 B        | 126.0 GiB      | 0 B
 0x0001   | 0x1111 | 0 B        | 126.0 GiB      | 0 B
 0x0001   | 0x1121 | 0 B        | 126.0 GiB      | 0 B
 0x0001   | 0x1101 | 0 B        | 126.0 GiB      | 0 B
Do you want to continue? [y/n] y_
```

Complete the memory provisioning (create goal) process by rebooting the server to enable the new settings. After the reboot, the newly created DIMM-interleave-sets are represented as persistent memory “regions” of App Direct Mode capacity. To view the region setup, use the List Regions command:

ipmctl show -region

This command returns output similar to the following:

```
FS3:\> ipmctl show -region
 RegionID| SocketID | PersistentMemoryType | Capacity  | FreeCapacity | HealthState
==========================================================================================
 0x0001  | 0x0000   | AppDirect            | 756.0 GiB | 0 B          | Healthy
 0x0002  | 0x0001   | AppDirect            | 756.0 GiB | 0 B          | Healthy
```

```
default:~ # lsblk
NAME    MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
sda       8:0    0 34.2G  0 disk
├─sda1    8:1    0    2G  0 part [SWAP]
└─sda2    8:2    0 32.2G  0 part /
pmem0   259:0    0 31.5G  0 disk
pmem1   259:1    0 31.5G  0 disk
```

```
default:~ # mkfs.xfs -m reflink=0 -f /dev/pmem0
meta-data=/dev/pmem0             isize=512    agcount=4, agsize=2064256 blks
         =                       sectsz=4096  attr=2, projid32bit=1
         =                       crc=1        finobt=0, sparse=0, rmapbt=0, reflink=0
data     =                       bsize=4096   blocks=8257024, imaxpct=25
         =                       sunit=0      swidth=0 blks
naming   =version 2              bsize=4096   ascii-ci=0 ftype=1
log      =internal log           bsize=4096   blocks=4031, version=2
         =                       sectsz=4096  sunit=1 blks, lazy-count=1
realtime =none                   extsz=4096   blocks=0, rtextents=0
default:~ # mkfs.xfs -m reflink=0 -f /dev/pmem1
meta-data=/dev/pmem1             isize=512    agcount=4, agsize=2064256 blks
         =                       sectsz=4096  attr=2, projid32bit=1
         =                       crc=1        finobt=0, sparse=0, rmapbt=0, reflink=0
data     =                       bsize=4096   blocks=8257024, imaxpct=25
         =                       sunit=0      swidth=0 blks
naming   =version 2              bsize=4096   ascii-ci=0 ftype=1
log      =internal log           bsize=4096   blocks=4031, version=2
         =                       sectsz=4096  sunit=1 blks, lazy-count=1
realtime =none                   extsz=4096   blocks=0, rtextents=0
```

Documents / Resources

	intel Optane Persistent Memory and SAP HANA Platform Configuration on VMware ESXi [pdf] User Guide Optane Persistent Memory and SAP HANA Platform Configuration on VMware ESXi, SAP HANA Platform Configuration on VMware ESXi, Platform Configuration on VMware ESXi, Configuration on VMware ESXi, VMware ESXi

References

- Data Center Partners, Technologies, and Solutions
- Fujitsu EMEIA : Fujitsu Global
- VMware vSphere 7.0 U2 and vSphere HA for SAP HANA with DRAM and Intel® Optane™ PMem in App-Direct Mode - Virtualize Applications
- Page Not Found - Lenovo Support US
- ssl.syncsearch.jp/click?url=https%3A%2F%2Fsupport
- Super Micro Computer, Inc. | Support | Resources