

# Configurar o DCPMM no VMware ESXi para o modo AppDirect

## Contents

[Introduction](#)

[Prerequisites](#)

[Requirements](#)

[Componentes Utilizados](#)

[Informações de Apoio](#)

[Configurar](#)

[Configurar o perfil de serviço](#)

[Verificar ESXi](#)

[Configurar NVDIMM de máquina virtual](#)

[Configurar namespace na máquina virtual](#)

[Troubleshoot](#)

[Informações Relacionadas](#)

## Introduction

Este documento descreve o processo para configurar o ESXi nos servidores da série UCS (Unified Computing System) B usando o Intel® Optane™ Persistent Memory (PMEM) no modo gerenciado por host.

## Prerequisites

### Requirements

A Cisco recomenda que você tenha conhecimento destes tópicos:

- UCS B Series
- Conceitos do Módulo de Memória Persistente (DCPMM) do Data Center Intel® Optane™
- Administração do VMware ESXi e vCenter Server

Certifique-se de atender a estes requisitos antes de tentar esta configuração:

- Consulte as diretrizes de PMEM no [guia de especificação](#) B200/B480 M5.
- Verifique se a CPU é a segunda geração de processadores escaláveis Intel® Xeon®.
- A razão PMEM/Dynamic Random Access Memory (DRAM) atende aos requisitos conforme [KB 67645](#).
- O ESXi é 6,7 U2 + Express Patch 10 (ESXi670-201906002) ou posterior. Versões anteriores da 6.7 não são suportadas.
- O UCS Manager e o Server estão em uma versão 4.0(4) ou superior. Para obter a versão recomendada mais recente, visite [www.software.cisco.com/](http://www.software.cisco.com/).

## Componentes Utilizados

As informações neste documento são baseadas nestas versões de software e hardware:

- UCS B480 M5
- UCS Manager 4.1(2b)

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. Se a rede estiver ativa, certifique-se de que você entenda o impacto potencial de qualquer comando.

## Informações de Apoio

Nos servidores UCS configurados para o modo App Direct, as máquinas virtuais VMware ESXi acessam memórias persistentes Optane DCPMM Módulos de memória dual em linha (NVDIMMs) não voláteis.

O Intel Optane DCPMM pode ser configurado através do utilitário de gerenciamento IPMCTL através do shell da UEFI (Unified Extensible Firmware Interface) ou através dos Utilitários do SO. Esta ferramenta foi projetada para executar algumas das próximas ações:

- Descubra e gerencie módulos
- Atualizar e configurar o firmware do módulo
- Monitorar integridade
- Provisionar e configurar Meta, Região e Namespaces
- Depurar e solucionar problemas do PMEM

O UCS pode ser configurado usando uma política de memória persistente anexada ao perfil de serviço para facilitar o uso.

O utilitário NDCTL (Non-Volatile Device Control) de código aberto é usado para gerenciar o subsistema LIBNVDIMM Linux Kernel. O utilitário NDCTL permite que um sistema provisione e execute configurações como regiões e namespaces para uso do SO.

A memória persistente adicionada a um host ESXi é detectada pelo host, formatada e montada como um armazenamento de dados PMem local. Para usar o PMEM, o ESXi usa o formato de sistema de arquivos VMFS (Virtual Machine Flying System)-L e somente um datastore PMEM local por host é suportado.

Diferente de outros datastores, o datastore PMEM não suporta tarefas como datastores tradicionais. O diretório home da VM com os arquivos vmx e vmware.log não pode ser colocado no armazenamento de dados PMEM.

O PMEM pode ser apresentado a uma VM em dois modos diferentes: Modo de acesso direto e modo de disco virtual.

- Modo de acesso direto

As VMs podem ser configuradas para esse modo apresentando a região PMEM na forma de um NVDIMM. O sistema operacional VM deve estar preparado para PMem para usar este modo. Os dados armazenados nos módulos NVDIMM podem persistir em todos os ciclos de

energia, pois o NVDIMM atua como memória endereçável por bytes. Os NVDIMMs são automaticamente armazenados no armazenamento de dados PMem criado pelo ESXi ao formatar o PMEM.

- Modo de Disco Virtual

Destinado ao SO tradicional e herdado que reside na VM para suportar qualquer versão de hardware. O SO da VM não precisa ser compatível com PMEM. Nesse modo, um disco virtual Small Computer System Interface (SCSI) pode ser criado e usado pelo SO da VM.

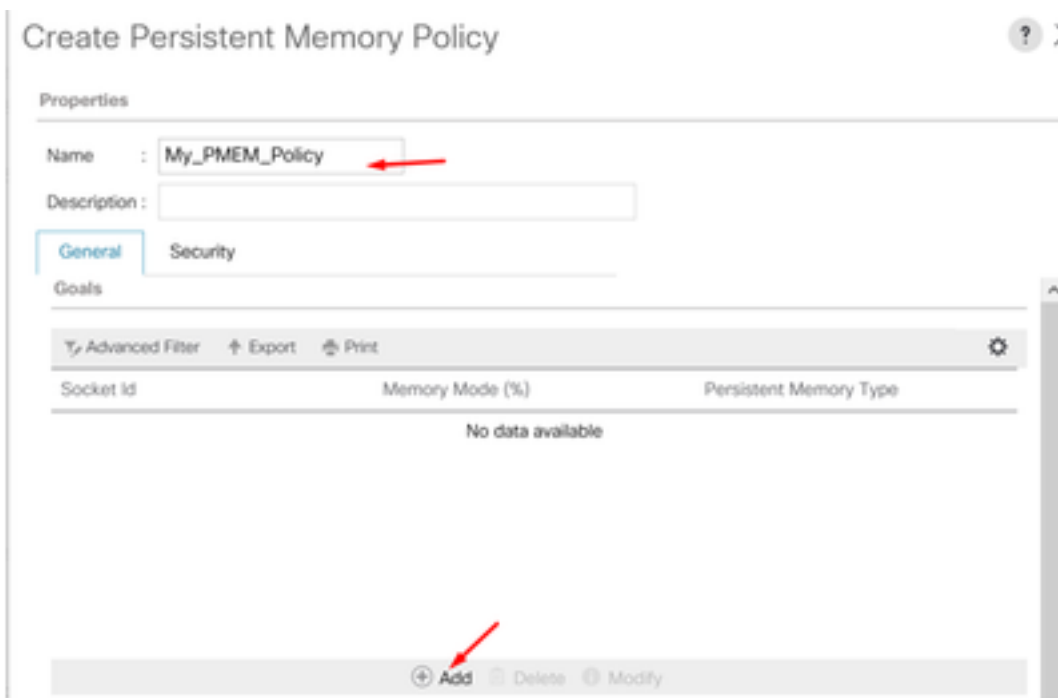
Este documento descreve a configuração para usar uma máquina virtual no modo de acesso direto.

## Configurar

Este procedimento descreve como configurar o ESXi em servidores UCS Blade Series usando o Intel Optane DCPMM.

### Configurar o perfil de serviço

1. Na GUI do UCS Manager, navegue até **Servers > Persistent Memory Policy** e clique em **Add** como mostrado na imagem.



2. Criar **meta**, verifique se o **modo de memória** é 0%, como mostrado na imagem.

## Create Goal



### Properties

Socket ID :  All Sockets

Memory Mode (%) :

Persistent Memory Type :  App Direct  App Direct Non Interleaved

OK

Cancel

3. Adicione a política PMEM ao perfil de serviço desejado.

Navegue até **Service Profile > Policies > Persistent Memory Policy** e anexe a política criada.

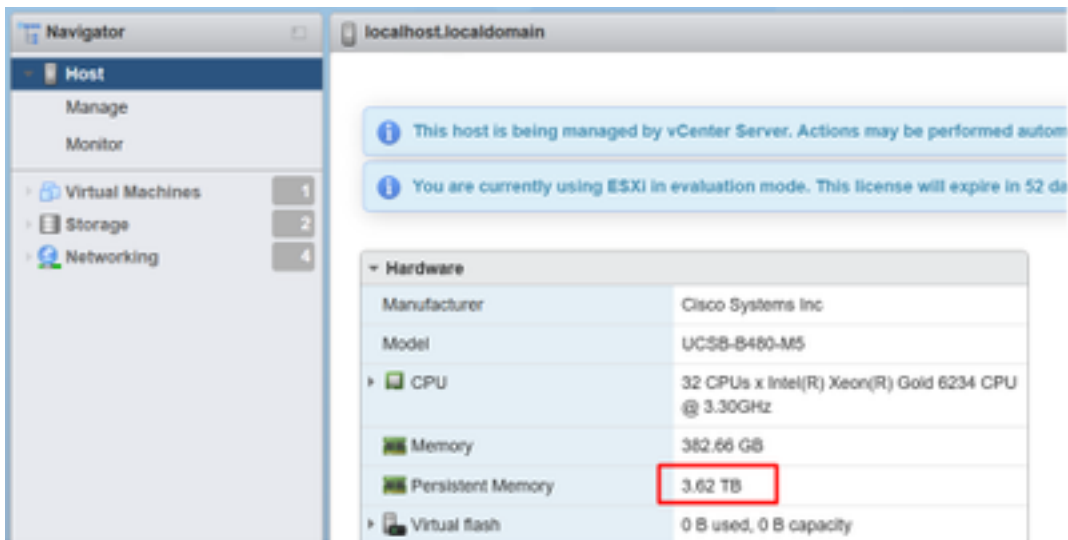
4. Verifique a saúde da região.

Navegue até o **Servidor** selecionado > **Inventário > Memória persistente > Regiões**. O tipo AppDirect está visível. Esse método cria uma região por soquete de CPU.

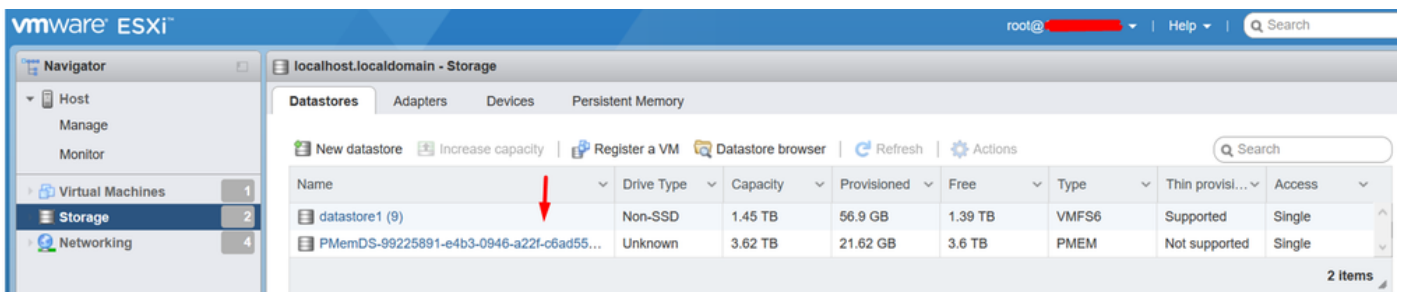
Id	Socket Id	Local DIMM Slot	DIMM Locator Ids	Type	Total Capacity (..)	Free Capacity (..)	Health Status:
1	Socket 1	Not Applicable	DIMM_A2.DIMM..	AppDirect	928	928	Healthy
2	Socket 2	Not Applicable	DIMM_G2.DIMM..	AppDirect	928	928	Healthy
3	Socket 3	Not Applicable	DIMM_N2.DIMM..	AppDirect	928	928	Healthy
4	Socket 4	Not Applicable	DIMM_U2.DIMM..	AppDirect	928	928	Healthy

## Verificar ESXi

1. No console da Web, o host exibe o PMEM total disponível.

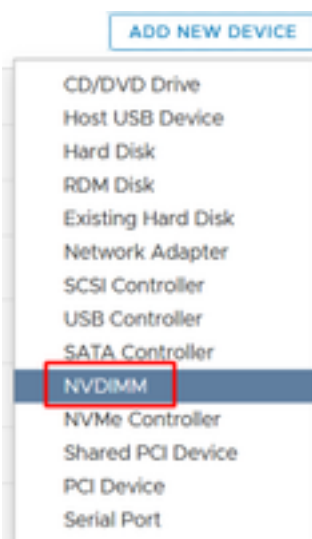


2. O ESXi exibe um armazenamento de dados especial composto pela quantidade total de PMEM, como mostrado na imagem.



## Configurar NVDIMM de máquina virtual

1. No ESXi, as máquinas virtuais acessam o Optane DCPMM PMEM como NVDIMMs. Para atribuir um NVMDIMM a uma máquina virtual, acesse a máquina virtual pelo vCenter e navegue para **Actions > Edit Settings (Ações > Editar configurações)**, clique em **ADD NEW DEVICE (ADICIONAR NOVO dispositivo)** e selecione **NVDIMM** conforme mostrado na imagem.



**Nota:** Ao criar uma máquina virtual, verifique se a compatibilidade do SO atende à versão mínima exigida que suporta a Memória persistente do Intel® Optane™; caso contrário, a opção **NVDIMM** não aparece nos itens selecionáveis.

2. Defina o tamanho do NVDIMM como mostrado na imagem.



## Configurar namespace na máquina virtual

1. O utilitário **NDCTL** é usado para gerenciar e configurar o PMEM ou NVDIMM.

No exemplo, o Red Hat 8 é usado para configuração. A Microsoft tem cmdlets PowerShell para gerenciamento de namespace de memória persistente.

Faça o download do utilitário **NDCTL** usando a ferramenta disponível de acordo com a distribuição do Linux

Por exemplo:

```
# yum install ndctl # zypper install ndctl # apt-get install ndctl
```

2. Verifique a região NVDIMM e o namespace criados por padrão pelo ESXi, quando o NVDIMM estiver atribuído à máquina virtual, verifique se o espaço corresponde à configuração. Certifique-se de que o modo do namespace esteja definido como **raw**, isso significa que o ESXi criou o namespace. Para verificar, use o comando:

```
# ndctl list -RuN
```

```
admin@localhost:/etc
[admin@localhost etc]$ ndctl list -RuN
{
  "regions": [
    {
      "dev": "region0",
      "size": "20.00 GiB (21.47 GB)",
      "available_size": 0,
      "max_available_extent": 0,
      "type": "pmem",
      "persistence_domain": "unknown",
      "namespaces": [
        {
          "dev": "namespace0.0",
          "mode": "raw",
          "size": "20.00 GiB (21.47 GB)",
          "blockdev": "pmem0"
        }
      ]
    }
  ]
}
```

3. (Opcional) Se o namespace ainda não tiver sido criado, um namespace pode ser criado com o comando:

```
# ndctl create-namespace
```

O comando **ndctl create-namespace** cria um novo namespace no modo **fsdax** por padrão e cria um novo **/dev/pmem([x].[y])** dispositivo. Se um namespace já tiver sido criado, essa etapa poderá ser ignorada.

4. Selecione o modo de acesso PMEM, os modos disponíveis para configuração são:

- Modo do setor:

Apresenta o armazenamento como um dispositivo de bloqueio rápido, o que é útil para aplicativos antigos que ainda não podem usar memória persistente.

- Modo Fsdax:

Permite que os dispositivos de memória persistente suportem acesso direto ao NVDIMM. O acesso direto ao sistema de arquivos requer o uso do modo **fsdax**, para permitir o uso do modelo de programação de acesso direto. Esse modo permite a criação de um sistema de arquivos sobre o NVDIMM.

- Modo Devdax:

Fornece acesso bruto à memória persistente usando um dispositivo de caracteres DAX. Os sistemas de arquivos não podem ser criados em dispositivos usando o modo **devdax**.

- Modo Bruto:

Esse modo tem várias limitações e não é recomendado para usar a memória persistente.

Para alterar o modo para o modo **fsdax**, use o comando:

```
ndctl create-namespace -f -e
```

Se um **dev** já tiver sido criado, o namespace dev será usado para formatar e modificar o modo para **fsdax**.

```
admin@localhost:/etc
File Edit View Search Terminal Help
    "size": "20.00 GiB (21.47 GB)",
    "blockdev": "pmem0"
  }
}
}
}
}
[admin@localhost etc]$ ndctl create-namespace -f -e namespace0.0 --mode fsdax
failed to reconfigure namespace: Permission denied
[admin@localhost etc]$ sudo ndctl create-namespace -f -e namespace0.0 --mode fsdax
[sudo] password for admin:
{
  "dev": "namespace0.0",
  "mode": "fsdax",
  "map": "dev",
  "size": "19.69 GiB (21.14 GB)",
  "uuid": "09658ac7-16ea-4c3d-8fbe-e9dae854ddf0",
  "sector_size": 512,
  "blockdev": "pmem0",
  "numa_node": 0
}
[admin@localhost etc]$
```

**Observação:** esses comandos exigem que a conta tenha privilégios de raiz, talvez seja necessário **sudo**.

5. Crie um diretório e um sistema de arquivos.

O Direct Access ou DAX é um mecanismo que permite que os aplicativos acessem diretamente a mídia persistente da CPU (por meio de cargas e lojas), ignorando a pilha de E/S tradicional. Os sistemas de arquivos de memória persistente habilitados para DAX incluem ext4, XFS e Windows NTFS.

Exemplo de sistema de arquivos XFS criado e montado:

`sudo mkdir < directory route (e.g./mnt/pmem) > sudo mkfs.xfs < /dev/devicename (e.g. pmem0) >`

```
admin@localhost:/etc
File Edit View Search Terminal Help
}
[admin@localhost etc]$ mkdir /mnt/pmem
mkdir: cannot create directory '/mnt/pmem': Permission denied
[admin@localhost etc]$ sudo mkdir /mnt/pmem
[admin@localhost etc]$ sudo mkfs.xfs /dev/pmem0
meta-data=/dev/pmem0      isize=512    agcount=4, agsize=1290112 blks
=                sectsz=4096  attr=2, projid32bit=1
=                crc=1      finobt=1, sparse=1, rmapbt=0
=                reflink=1
data        =                bsize=4096  blocks=5160448, imaxpct=25
=                sunit=0    swidth=0 blks
naming      =version 2      bsize=4096  ascii-ci=0, ftype=1
log         =internal log  bsize=4096  blocks=2560, version=2
=                sectsz=4096  sunit=1 blks, lazy-count=1
realtime    =none        extsz=4096  blocks=0, rtextents=0
[admin@localhost etc]$
```

6. Monte o sistema de arquivos e verifique se ele foi bem-sucedido.

`sudo mount`



```
admin@localhost:/etc
File Edit View Search Terminal Help
[admin@localhost etc]$ sudo mount /dev/pmem0 /mnt/pmem/
[admin@localhost etc]$ // verify the mount was successful
bash: //: Is a directory
[admin@localhost etc]$ df -h /mnt/pmem/
Filesystem      Size  Used Avail Use% Mounted on
/dev/pmem0      20G  173M   20G   1% /mnt/pmem
[admin@localhost etc]$
```

A VM está pronta para usar o PMEM.

## Troubleshoot

Geralmente, é recomendado montar este sistema de arquivos habilitado para DAX usando a opção `-o dax` mount, se um erro for encontrado.

```
[admin@localhost etc]$ sudo mount -o dax /dev/pmem0 /mnt/pmem/
mount: /mnt/pmem: wrong fs type, bad option, bad superblock on /dev/pmem0, missing codepage or helper program, or other error.
```

O reparo do sistema de arquivos é executado para garantir a integridade.

```
[admin@localhost etc]$ sudo xfs_repair /dev/pmem0
[sudo] password for admin:
Phase 1 - find and verify superblock...
Phase 2 - using internal log
- zero log...
- scan filesystem freespace and inode maps...
- found root inode chunk
Phase 3 - for each AG...
- scan and clear agi unlinked lists...
- process known inodes and perform inode discovery...
- agno = 0
- agno = 1
- agno = 2
- agno = 3
- process newly discovered inodes...
Phase 4 - check for duplicate blocks...
- setting up duplicate extent list...
- check for inodes claiming duplicate blocks...
- agno = 0
- agno = 1
- agno = 2
- agno = 3
Phase 5 - rebuild AG headers and trees...
- reset superblock..
Phase 6 - check inode connectivity...
- resetting contents of realtime bitmap and summary inodes
- traversing filesystem ...
- traversal finished ...
- moving disconnected inodes to lost+found ...
Phase 7 - verify and correct link counts...
done
[admin@localhost etc]$
```

Como solução alternativa, a montagem pode ser montada sem a opção `-o dax`.

**Note:** No `xfsprogs` versão 5.1, o padrão é criar sistemas de arquivos XFS com a opção `reflink` habilitada. Anteriormente, estava desativado por padrão. As opções `reflink` e `dax` são mutuamente exclusivas, fazendo com que a montagem falhe.

"O DAX e o reflink não podem ser usados juntos!" o erro pode ser visto no **dmesg** quando o comando mount falha:

```
admin@localhost:/etc
File Edit View Search Terminal Help
log      =internal log          bsize=4096   blocks=2560, version=2
         =                   sectsz=4096  sunit=1 blks, lazy-count=1
realtime =none              extsz=4096   blocks=0, rtextents=0
[admin@localhost etc]$ mount -o dax /dev/pmem0 /mnt/pmem
mount: only root can use "--options" option
[admin@localhost etc]$ sudo mount -o dax /dev/pmem0 /mnt/pmem/
mount: /mnt/pmem: wrong fs type, bad option, bad superblock on /dev/pmem0, missing
codepage or helper program, or other error.
[admin@localhost etc]$ dmesg -T | tail
[mar nov 10 00:12:18 2020] VFS: busy inodes on changed media or resized disk sr0
[mar nov 10 00:12:22 2020] ISO 9660 Extensions: Microsoft Joliet Level 3
[mar nov 10 00:12:22 2020] ISO 9660 Extensions: RRIP_1991A
[mar nov 10 01:47:35 2020] pmem0: detected capacity change from 0 to 21137195008
[mar nov 10 01:51:19 2020] XFS (pmem0): DAX enabled. Warning: EXPERIMENTAL, use
at your own risk
[mar nov 10 01:51:19 2020] XFS (pmem0): DAX and reflink cannot be used together!
[mar nov 10 01:53:06 2020] XFS (pmem0): DAX enabled. Warning: EXPERIMENTAL, use
at your own risk
[mar nov 10 01:53:06 2020] XFS (pmem0): DAX and reflink cannot be used together!
[mar nov 10 01:59:29 2020] XFS (pmem0): DAX enabled. Warning: EXPERIMENTAL, use
at your own risk
[mar nov 10 01:59:29 2020] XFS (pmem0): DAX and reflink cannot be used together!
[admin@localhost etc]$
```

Como solução alternativa, remova a opção **-o dax**.

```
admin@localhost:/etc
File Edit View Search Terminal Help
[admin@localhost etc]$ sudo mount /dev/pmem0 /mnt/pmem/
[admin@localhost etc]$ // verify the mount was successful
bash: //: Is a directory
[admin@localhost etc]$ df -h /mnt/pmem/
Filesystem      Size  Used Avail Use% Mounted on
/dev/pmem0      20G  173M   20G   1% /mnt/pmem
[admin@localhost etc]$
```

Montagem com ext4 FS.

O sistema de arquivos EXT4 pode ser usado como uma alternativa porque não implementa o recurso de relink, mas suporta DAX.

```
[admin@localhost etc]$ sudo mkfs.ext4 /dev/pmem0
mke2fs 1.44.3 (10-July-2018)
/dev/pmem0 contains a xfs file system
Proceed anyway? (y,N) y
Creating filesystem with 5160448 4k blocks and 1291808 inodes
Filesystem UUID: 164c6d57-0462-45a0-9b94-703719272816
Superblock backups stored on blocks:
    32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632, 2654208,
    4096000

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done

[admin@localhost etc]$ sudo mount /dev/pmem0 /mnt/pmem/
[admin@localhost etc]$ df -h /mnt/pmem/
Filesystem      Size  Used Avail Use% Mounted on
/dev/pmem0      20G   45M   19G   1% /mnt/pmem
[admin@localhost etc]$
```

## Informações Relacionadas

- [Guia de início rápido: Provisione memória persistente Intel® Optane™ DC](#)
- [Configuração de memória persistente](#)
- [Utilitários de gerenciamento ipmctl e ndctl para memória persistente Intel® Optane™](#)
- [Suporte Técnico e Documentação - Cisco Systems](#)