

# Configurer DCPMM dans VMware ESXi pour le mode AppDirect

## Contenu

[Introduction](#)

[Conditions préalables](#)

[Conditions requises](#)

[Components Used](#)

[Informations générales](#)

[Configuration](#)

[Configurer le profil de service](#)

[Vérifier ESXi](#)

[Configurer NVDIMM de machine virtuelle](#)

[Configurer l'espace de noms dans la machine virtuelle](#)

[Dépannage](#)

[Informations connexes](#)

## Introduction

Ce document décrit le processus de configuration d'ESXi sur les serveurs de la gamme UCS B à l'aide de la mémoire persistante Intel® Optane™ (PMEM) en mode géré par l'hôte.

## Conditions préalables

### Conditions requises

Cisco vous recommande de prendre connaissance des rubriques suivantes :

- Gamme UCS B
- Concepts du module DCPMM (Data Center Persistent Memory Module) Intel® Optane™
- Administration VMware ESXi et vCenter Server

Assurez-vous que vous répondez à ces exigences avant d'essayer cette configuration :

- Reportez-vous aux directives PMEM du [guide de spécification](#) B200/B480 M5.
- Assurez-vous que le processeur est un processeur Intel® Xeon® évolutif de deuxième génération.
- Le rapport PMEM/DRAM (Dynamic Random Access Memory) répond aux exigences de la [Ko 67645](#).
- ESXi est 6,7 U2 + Express Patch 10 (ESXi670-201906002) ou ultérieur. Les versions antérieures à la version 6.7 ne sont pas prises en charge.
- UCS Manager et Server sont dans une version 4.0(4) ou supérieure. Pour la dernière version recommandée, veuillez consulter le site [www.software.cisco.com/](http://www.software.cisco.com/).

## Components Used

Les informations contenues dans ce document sont basées sur les versions de matériel et de logiciel suivantes :

- 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. Si votre réseau est en ligne, assurez-vous de bien comprendre l'incidence possible des commandes.

## Informations générales

Dans les serveurs UCS configurés pour le mode App Direct, les machines virtuelles VMware ESXi accèdent aux modules de mémoire NVDIMM (Non-Volatile Dual In-Line Memory Modules) DCPMM Optane.

Intel Optane DCPMM peut être configuré par l'intermédiaire de l'utilitaire de gestion IPMCTL via le shell Unified Extensible Firmware Interface (UEFI) ou via les utilitaires du système d'exploitation. Cet outil est conçu pour effectuer certaines des actions suivantes :

- Découverte et gestion des modules
- Mise à jour et configuration du micrologiciel du module
- Intégrité du moniteur
- Provisionner et configurer les espaces de noms, de région et d'objectif
- Déboguer et dépanner PMEM

UCS peut être configuré à l'aide d'une stratégie de mémoire persistante associée au profil de service pour faciliter l'utilisation.

L'utilitaire NDCTL (Non-Volatile Device Control) open source est utilisé pour gérer le sous-système de noyau Linux LIBNVDIMM. L'utilitaire NDCTL permet à un système de provisionner et d'effectuer des configurations en tant que régions et espaces de noms pour l'utilisation du système d'exploitation.

La mémoire persistante ajoutée à un hôte ESXi est détectée par l'hôte, formatée et montée en tant que data store PMem local. Afin d'utiliser le PMEM, ESXi utilise le format de système de fichiers VMFS (Virtual Machine Flying System)-L et un seul data store PMEM local par hôte est pris en charge.

Contrairement aux autres data stores, le data store PMEM ne prend pas en charge les tâches en tant que data stores traditionnels. Le répertoire d'accueil de la VM contenant les fichiers vmx et vmware.log ne peut pas être placé sur le data store PMEM.

PMEM peut être présenté à une machine virtuelle selon deux modes différents : Mode d'accès direct et mode disque virtuel.

- Mode d'accès direct  
Les machines virtuelles peuvent être configurées pour ce mode en présentant la région PMEM sous la forme d'une NVDIMM. Le système d'exploitation de la VM doit être sensible à

PMem pour utiliser ce mode. Les données stockées sur les modules NVDIMM peuvent persister pendant les cycles d'alimentation, car la mémoire NVDIMM agit comme une mémoire adressable par octet. Les barrettes NVDIMM sont automatiquement stockées dans le datastore PMem créé par l'ESXi lors de la formation du PMEM.

- Mode disque virtuel

Destiné aux systèmes d'exploitation traditionnels et hérités résidant sur la machine virtuelle afin de prendre en charge toutes les versions matérielles. Le système d'exploitation de la VM n'est pas nécessaire pour prendre en charge PMEM. Dans ce mode, un disque virtuel SCSI (Small Computer System Interface) traditionnel peut être créé et utilisé par le système d'exploitation de la machine virtuelle.

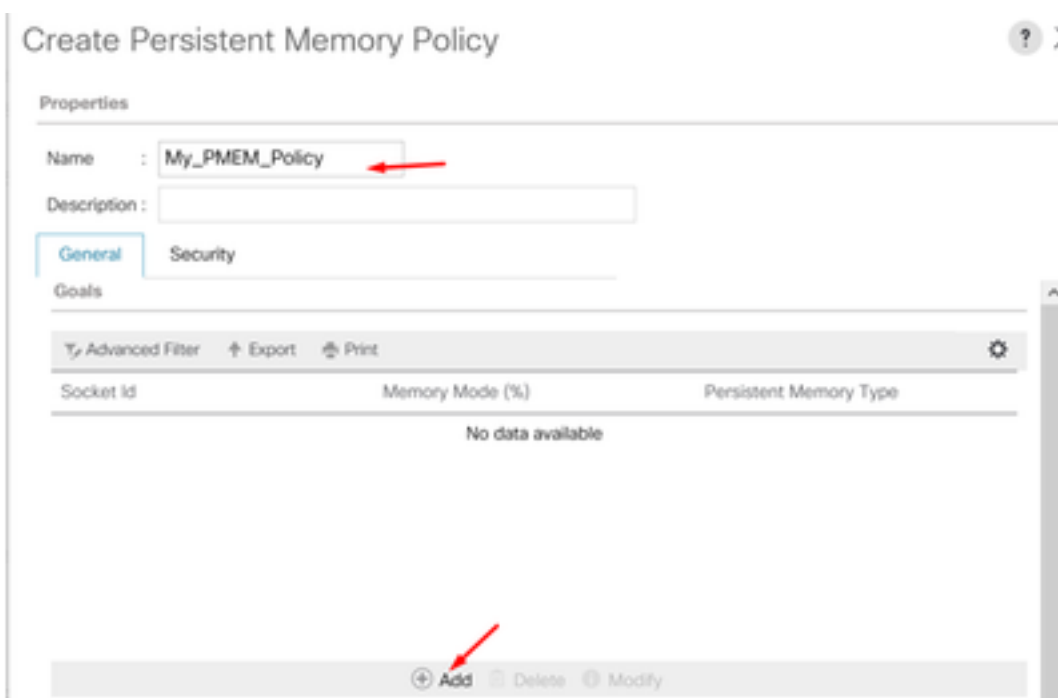
Ce document décrit la configuration pour utiliser une machine virtuelle en mode Accès direct.

## Configuration

Cette procédure décrit comment configurer ESXi sur des serveurs lames UCS utilisant Intel Optane DCPMM.

### Configurer le profil de service

1. Dans l'interface utilisateur graphique d'UCS Manager, accédez à **Serveurs > Stratégie de mémoire persistante** et cliquez sur **Ajouter** comme indiqué dans l'image.



2. Créez **Objectif**, assurez-vous que le **mode Mémoire** est égal à 0 % comme indiqué dans l'image.

## Create Goal



### Properties

Socket ID :  All Sockets

Memory Mode (%) :

Persistent Memory Type :  App Direct  App Direct Non Interleaved

OK

Cancel

3. Ajoutez la stratégie PMEM au profil de service souhaité.

Accédez à **Service Profile > Politiques > Persistent Memory Policy** et joignez la stratégie créée.

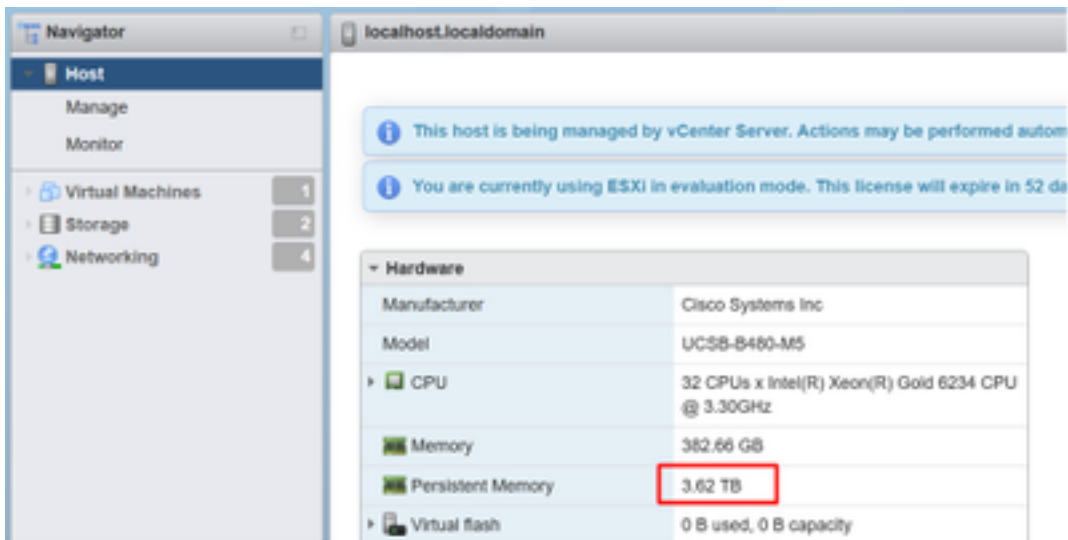
4. Vérifiez l'état de santé de la région.

Accédez à **Serveur sélectionné > Inventaire > Mémoire persistante > Régions**. Le type AppDirect est visible. Cette méthode crée une région par socket de processeur.

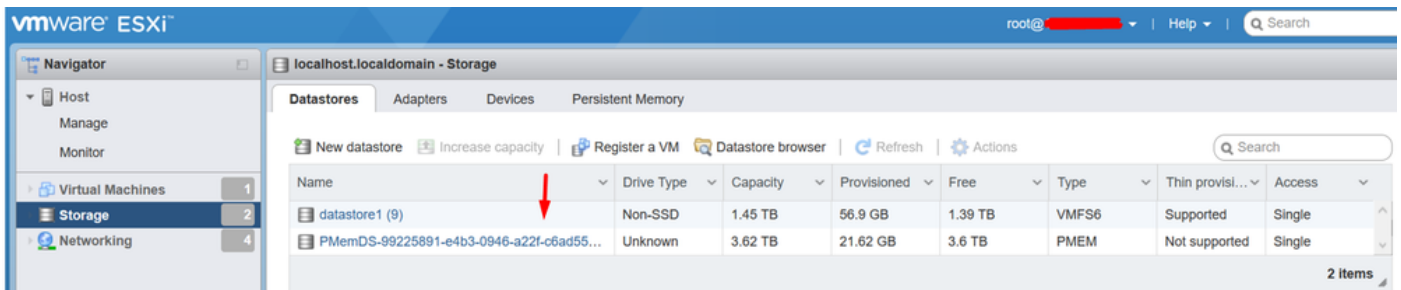
Id	Socket Id	Local DIMM Slot	DIMM Locator Ids	Type	Total Capacity (..)	Free Capacity (..)	Health Status:
1	Socket 1	Not Applicable	DIMM_A2.DIMM..	AppDirect	928	928	Healthy
2	Socket 2	Not Applicable	DIMM_G2.DIMM..	AppDirect	928	928	Healthy
3	Socket 3	Not Applicable	DIMM_N2.DIMM..	AppDirect	928	928	Healthy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

## Vérifier ESXi

1. Dans la console Web, l'hôte affiche le nombre total de modules PMEM disponibles.

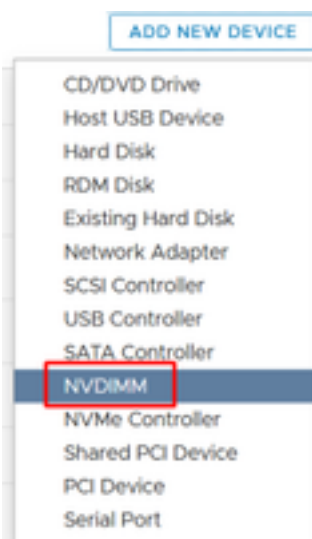


2. ESXi affiche un data store spécial composé de la quantité totale de PMEM, comme l'illustre l'image.



## Configurer NVDIMM de machine virtuelle

1. Dans ESXi, les machines virtuelles accèdent à Optane DCPMM PMEM en tant que NVDIMM. Afin d'affecter une NVMDIMM à une machine virtuelle, accédez à la machine virtuelle via vCenter et accédez à **Actions > Edit Settings**, cliquez sur **ADD NEW DEVICE** et sélectionnez **NVDIMM** comme indiqué dans l'image.



**Remarque** : lorsque vous créez une machine virtuelle, assurez-vous que la compatibilité du système d'exploitation correspond à la version minimale requise qui prend en charge la mémoire persistante Intel® Optane™, sinon l'option **NVDIMM** n'apparaît pas dans les éléments sélectionnables.

2. Définissez la taille de NVDIMM comme indiqué dans l'image.



## Configurer l'espace de noms dans la machine virtuelle

1. L'utilitaire **NDCTL** est utilisé pour gérer et configurer le module PMEM ou NVDIMM.

Dans l'exemple, Red Hat 8 est utilisé pour la configuration. Microsoft dispose d'applets de commande PowerShell pour la gestion de l'espace de noms de mémoire persistant.

Téléchargez l'utilitaire **NDCTL** à l'aide de l'outil disponible conformément à la distribution Linux

Exemple :

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

2. Vérifiez la région NVDIMM et l'espace de noms créés par défaut par ESXi, lorsque la NVDIMM est attribuée à la machine virtuelle, vérifiez que l'espace correspond à la configuration. Assurez-vous que le mode de l'espace de noms est défini sur **raw**, ce qui signifie qu'ESXi a créé l'espace de noms. Afin de vérifier, utilisez la commande :

```
# ndctl list -RuN
```

```
admin@localhost:/etc
File Edit View Search Terminal Help
[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. (Facultatif) Si l'espace de noms n'a pas encore été créé, vous pouvez créer un espace de noms à l'aide de la commande suivante :

```
# ndctl create-namespace
```

La commande **ndctl create-namespace** crée un nouvel espace de noms en mode **fsdax** par défaut et crée un nouveau périphérique **/dev/pmem([x].[y])**. Si un espace de noms a déjà été créé, cette étape peut être ignorée.

4. Sélectionnez le mode d'accès PMEM, les modes disponibles pour la configuration sont les suivants :

- Mode Secteur :

Présente le stockage comme un périphérique de blocage rapide, ce qui est utile pour les applications héritées qui ne sont toujours pas en mesure d'utiliser la mémoire persistante.

- Mode Fsdax :

Permet aux périphériques mémoire persistants de prendre en charge l'accès direct à la mémoire NVDIMM. L'accès direct au système de fichiers nécessite l'utilisation du mode **fsdax**, afin d'activer l'utilisation du modèle de programmation d'accès direct. Ce mode permet de créer un système de fichiers au-dessus de la NVDIMM.

- Mode Devdax :

Fournit un accès brut à la mémoire persistante à l'aide d'un périphérique de caractères DAX. Les systèmes de fichiers ne peuvent pas être créés sur des périphériques utilisant le mode **devdax**.

- Mode brut :

Ce mode présente plusieurs limitations et n'est pas recommandé pour l'utilisation de la mémoire persistante.

Afin de changer le mode en mode **fsdax**, utilisez la commande :

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

S'il y a un **dev** déjà créé, l'espace de noms dev est utilisé pour formater et modifier le mode en **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]$
```

**Remarque** : ces commandes exigent que le compte dispose de privilèges root, la commande `sudo` peut être requise.

### 5. Créez un répertoire et un système de fichiers.

Direct Access ou DAX est un mécanisme qui permet aux applications d'accéder directement aux supports persistants à partir du processeur (via les caches et les magasins), en contournant la pile d'E/S traditionnelle. Les systèmes de fichiers de mémoire persistante compatibles DAX incluent ext4, XFS et Windows NTFS.

Exemple de système de fichiers XFS créé et monté :

```
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. Montez le système de fichiers et vérifiez qu'il a réussi.

```
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]$
```

La machine virtuelle est prête à utiliser PMEM.

## Dépannage

Il est généralement recommandé de monter ce système de fichiers compatible DAX à l'aide de l'option de montage **-o dax**, si une erreur est détectée.

```
[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.
```

La réparation du système de fichiers est exécutée pour garantir l'intégrité.

```
[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]$
```

Comme solution de contournement, le montage peut être monté sans l'option **-o dax**.

**Note:** Dans **xfsprogs** version 5.1, la valeur par défaut est de créer des systèmes de fichiers XFS avec l'option de **réflink** activée. Auparavant, elle était désactivée par défaut. Les options de **réflecteur** et de **dax** s'excluent mutuellement, ce qui provoque l'échec du montage.

«DAX et reflink ne peuvent pas être utilisés ensemble ! » l'erreur peut être vue dans **dmesg** lorsque la commande mount échoue :

```
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]$
```

Pour contourner le problème, supprimez l'option **-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]$
```

Montage avec ext4 FS.

Le système de fichiers EXT4 peut être utilisé comme alternative car il n'implémente pas la fonctionnalité de reflink mais prend en charge 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]$
```

## Informations connexes

- [Guide de démarrage rapide : Provisionner la mémoire permanente Intel® Optane™ DC](#)
- [Configuration de la mémoire permanente](#)
- [Utilitaires de gestion ipmctl et ndctl pour mémoire permanente Intel® Optane™](#)
- [Support et documentation techniques - Cisco Systems](#)