

Configurazione di DCPMM in VMware ESXi per la modalità AppDirect

Sommario

[Introduzione](#)

[Prerequisiti](#)

[Requisiti](#)

[Componenti usati](#)

[Premesse](#)

[Configurazione](#)

[Configurare il profilo del servizio](#)

[Verifica di ESXi](#)

[Configura NVDIMM macchina virtuale](#)

[Configura spazio dei nomi nella macchina virtuale](#)

[Risoluzione dei problemi](#)

[Informazioni correlate](#)

Introduzione

In questo documento viene descritto il processo di configurazione di ESXi su server UCS serie B utilizzando Intel® Optane™ Persistent Memory (PMEM) in modalità di gestione host.

Prerequisiti

Requisiti

Cisco raccomanda la conoscenza dei seguenti argomenti:

- UCS serie B
- Concetti relativi al modulo di memoria persistente (DCPMM) per il data center Intel® Optane™
- Amministrazione di VMware ESXi e vCenter Server

Prima di provare questa configurazione, accertarsi di soddisfare i seguenti requisiti:

- Consultare le linee guida PMEM sulla [guida alle specifiche](#) B200/B480 M5.
- Assicuratevi che la CPU sia costituita da processori Intel® Xeon® scalabili di seconda generazione.
- Il rapporto PMEM/DRAM (Dynamic Random Access Memory) soddisfa i requisiti di [KB 67645](#).
- ESXi è 6,7 U2 + Express Patch 10 (ESXi670-201906002) o successiva. Le release precedenti della versione 6.7 non sono supportate.
- UCS Manager e Server sono in versione 4.0(4) o successiva. Per l'ultima versione consigliata, visitare il sito www.software.cisco.com/.

Componenti usati

Le informazioni fornite in questo documento si basano sulle seguenti versioni software e hardware:

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

Le informazioni discusse in questo documento fanno riferimento a dispositivi usati in uno specifico ambiente di emulazione. Su tutti i dispositivi menzionati nel documento la configurazione è stata ripristinata ai valori predefiniti. Se la rete è operativa, valutare attentamente eventuali conseguenze derivanti dall'uso dei comandi.

Premesse

Nei server UCS configurati per la modalità App Direct, le macchine virtuali VMware ESXi accedono alle memorie persistenti DCPMM in modalità Optane Moduli di memoria in linea doppi non volatili (NVDIMM).

Intel Optane DCPMM può essere configurato tramite l'utility di gestione IPMCTL tramite la shell UEFI (Unified Extensible Firmware Interface) o tramite le utility del sistema operativo. Questo strumento è progettato per eseguire alcune delle azioni successive:

- Individua e gestisci moduli
- Aggiornare e configurare il firmware del modulo
- Monitoraggio stato
- Provisioning e configurazione di obiettivi, aree geografiche e spazi dei nomi
- Debug e risoluzione dei problemi relativi a PMEM

È possibile configurare UCS utilizzando un criterio di memoria persistente collegato al profilo del servizio per semplificarne l'utilizzo.

L'utilità NDCTL (Non-Volatile Device Control) open-source viene utilizzata per gestire il sottosistema del kernel Linux LIBNVDIMM. L'utilità NDCTL consente a un sistema di effettuare il provisioning ed eseguire configurazioni come aree e spazi dei nomi per l'utilizzo del sistema operativo.

La memoria persistente aggiunta a un host ESXi viene rilevata dall'host, formattata e montata come datastore PMem locale. Per utilizzare PMEM, ESXi utilizza il formato di file system VMFS (Virtual Machine Flying System)-L ed è supportato un solo datastore PMEM locale per host.

A differenza di altri archivi dati, l'archivio dati PMEM non supporta le attività come archivi dati tradizionali. La directory home VM con i file vmx e vmware.log non può essere inserita nell'archivio dati PMEM.

PMEM può essere presentato a una VM in due diverse modalità: Modalità di accesso diretto e modalità disco virtuale.

- Modalità accesso diretto

Le VM possono essere configurate per questa modalità presentando PMEMregion sotto forma di NVDIMM. Per utilizzare questa modalità, è necessario che il sistema operativo della

macchina virtuale riconosca PMem. I dati archiviati nei moduli NVDIMM possono essere mantenuti anche durante i cicli di alimentazione poiché la NVDIMM funge da memoria indirizzabile ai byte. I NVDIMM vengono automaticamente memorizzati nell'archivio dati PMem creato da ESXi durante la formattazione del modulo PMEM.

- Modalità disco virtuale

Progettato per sistemi operativi tradizionali e legacy che risiedono su VM per supportare qualsiasi versione hardware. Non è necessario che il sistema operativo VM riconosca PMEM. In questa modalità, è possibile creare e utilizzare un disco virtuale SCSI (Small Computer System Interface) tradizionale dal sistema operativo VM.

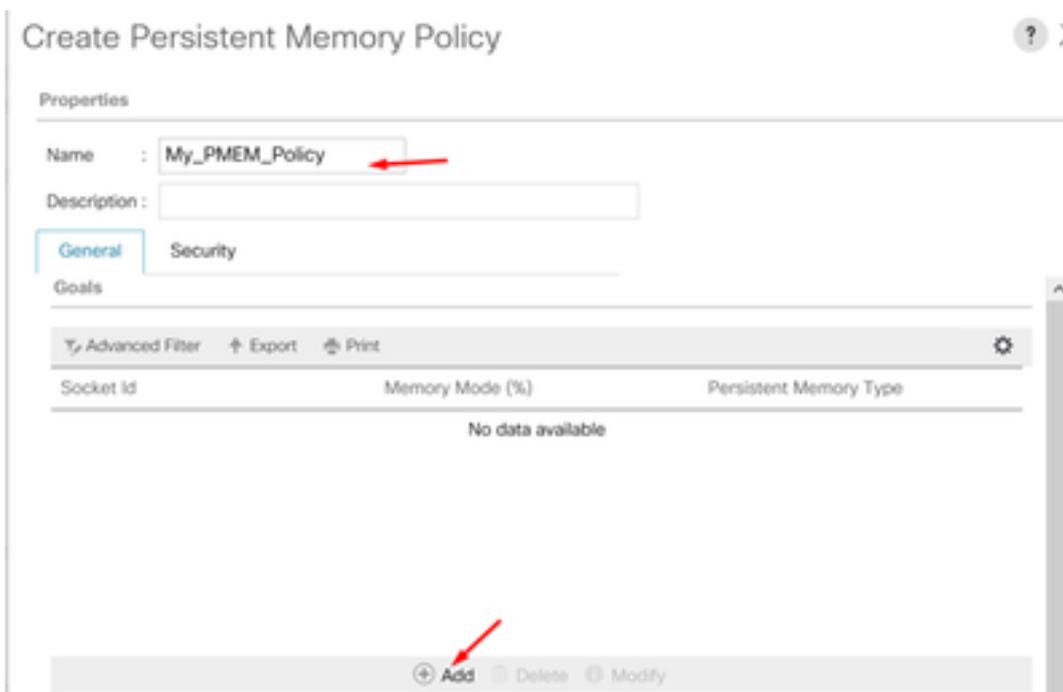
In questo documento viene descritta la configurazione per l'utilizzo di una macchina virtuale in modalità di accesso diretto.

Configurazione

In questa procedura viene descritto come configurare ESXi sui server blade UCS utilizzando il DCPMM Intel Optane.

Configurare il profilo del servizio

1. Nell'interfaccia utente di UCS Manager, selezionare **Server > Criterio memoria persistente** e fare clic su **Add** (Aggiungi), come mostrato nell'immagine.



2. Create **Goal**, assicuratevi che la **modalità Memoria** sia 0%, come mostrato nell'immagine.

Create Goal



Properties

Socket ID : All Sockets

Memory Mode (%) :

Persistent Memory Type : App Direct App Direct Non Interleaved

OK

Cancel

3. Aggiungere i criteri PMEM al profilo di servizio desiderato.

Selezionare **Profilo servizio > Criteri > Criterio memoria persistente** e allegare il criterio creato.

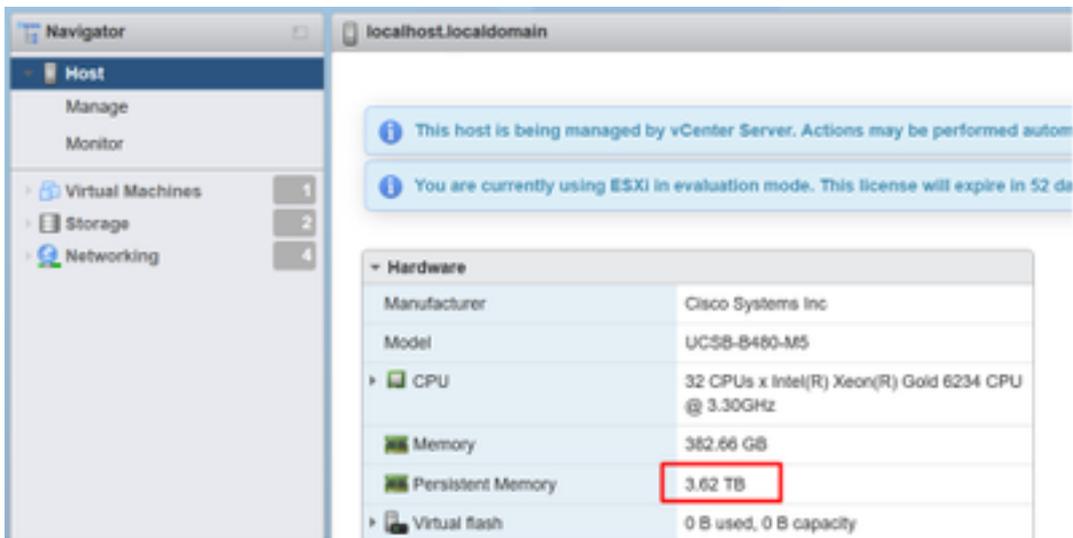
4. Verificare lo stato di salute della regione.

Passare al **Server** selezionato > **Inventario > Memoria persistente > Aree**. Il tipo AppDirect è visibile. Questo metodo crea una regione per socket CPU.

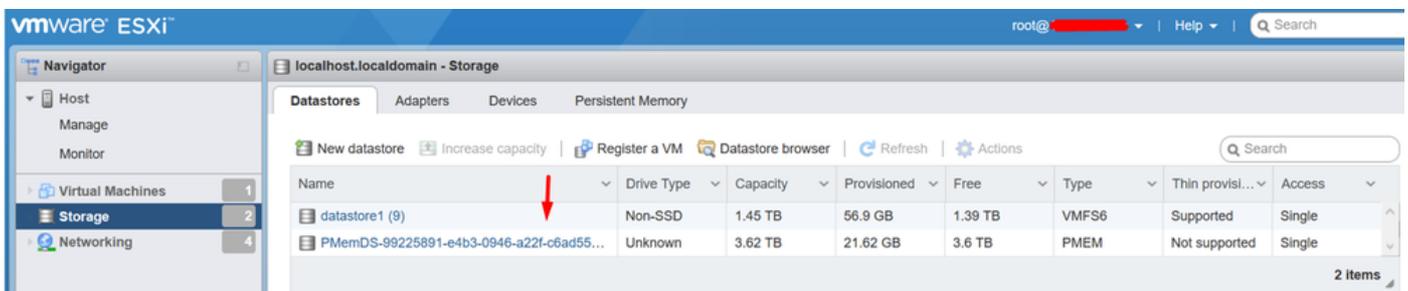
Id	Socket Id	Local DIMM Slot	DIMM Locator Ids	Type	Total Capacity (..)	Free Capacity (..)	Health Status
1	Socket 1	Not Applicable	DIMM_A2.DIMM..	AppDirect	928	928	Healthy
2	Socket 2	Not Applicable	DIMM_G2.DIMM..	AppDirect	928	928	Healthy
3	Socket 3	Not Applicable	DIMM_N2.DIMM..	AppDirect	928	928	Healthy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

Verifica di ESXi

1. Nella console Web, l'host visualizza il totale PMEM disponibile.



2. ESXi visualizza un datastore speciale composto dalla quantità totale di PMEM, come mostrato nell'immagine.



Configura NVDIMM macchina virtuale

1. In ESXi, le macchine virtuali accedono al modulo PMMM DCPMM in modalità Optane come NVDIMM. Per assegnare un NVMDIMM a una macchina virtuale, accedere alla macchina virtuale tramite vCenter e selezionare **Azioni > Modifica impostazioni**, fare clic su **AGGIUNGI NUOVO DISPOSITIVO** e selezionare **NVDIMM** come mostrato nell'immagine.



Nota: Quando si crea una macchina virtuale, verificare che la compatibilità del sistema operativo soddisfi la versione minima richiesta che supporta la memoria persistente Intel® Optane™. In caso contrario, l'opzione **NVDIMM** non viene visualizzata negli elementi selezionabili.

2. Impostare le dimensioni della NVDIMM come illustrato nell'immagine.



Configura spazio dei nomi nella macchina virtuale

1. L'utilità **NDCTL** viene utilizzata per gestire e configurare PMEM o NVDIMM.

Nell'esempio, Red Hat 8 viene utilizzato per la configurazione. Microsoft dispone di cmdlet PowerShell per la gestione persistente dello spazio dei nomi della memoria.

Scaricare l'utility **NDCTL** utilizzando lo strumento disponibile in Linux Distribution

Ad esempio:

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

2. Verificare la regione NVDIMM e lo spazio dei nomi creati per impostazione predefinita da ESXi, quando NVDIMM viene assegnato al sistema virtuale, verificare che lo spazio corrisponda alla configurazione. Assicurarsi che la modalità dello spazio dei nomi sia impostata su **raw**, ovvero che ESXi abbia creato lo spazio dei nomi. Per procedere alla verifica, utilizzare il comando:

```
# 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. (Facoltativo) Se lo spazio dei nomi non è già stato creato, è possibile crearlo con il comando:

```
# ndctl create-namespace
```

Per impostazione predefinita, il comando **ndctl create-namespace** crea un nuovo spazio dei nomi in modalità **fsdax** e crea un nuovo dispositivo **/dev/pmem([x].[y])**. Se è già stato creato uno spazio dei nomi, è possibile ignorare questo passaggio.

4. Selezionare la modalità di accesso PMEM. Le modalità disponibili per la configurazione sono:

- Modalità settore:

Presenta lo storage come un dispositivo di blocco rapido, utile per le applicazioni legacy che non sono ancora in grado di utilizzare la memoria persistente.

- Modalità Fsdax:

Consente ai dispositivi di memoria persistente di supportare l'accesso diretto alla NVDIMM. L'accesso diretto al file system richiede l'utilizzo della modalità **fsdax** per consentire l'utilizzo del modello di programmazione ad accesso diretto. Questa modalità consente la creazione di un file system sulla NVDIMM.

- Modalità Devdax:

Fornisce l'accesso raw alla memoria persistente utilizzando un dispositivo carattere DAX. I file system non possono essere creati su dispositivi che utilizzano la modalità **devdax**.

- Modalità raw:

Questa modalità presenta diverse limitazioni e non è consigliata per l'utilizzo della memoria persistente.

Per modificare la modalità in **modalità fsdax**, utilizzare il comando:

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

Se è già stato creato un **dev**, lo spazio dei nomi dev viene utilizzato per formattare e modificare la modalità in **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]$

```

Nota: per questi comandi è necessario che l'account disponga di privilegi root, potrebbe essere richiesto il comando **sudo**.

5. Creare una directory e un file system.

Direct Access o DAX è un meccanismo che consente alle applicazioni di accedere direttamente ai supporti persistenti dalla CPU (attraverso carichi e archivi), ignorando il tradizionale stack di I/O. I file system di memoria persistente abilitati per DAX includono ext4, XFS e Windows NTFS.

Esempio di file system XFS creato e montato:

`sudo mkdir < directory route (e.g. /mnt/pmем) > sudo mkfs.xfs < /dev/devicename (e.g. pmем0) >`

```

admin@localhost:/etc
File Edit View Search Terminal Help
}
[admin@localhost etc]$ mkdir /mnt/pmем
mkdir: cannot create directory '/mnt/pmем': Permission denied
[admin@localhost etc]$ sudo mkdir /mnt/pmем
[admin@localhost etc]$ sudo mkfs.xfs /dev/pmем0
meta-data=/dev/pmем0      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
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. Installare il file system e verificare che sia stato eseguito correttamente.

`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 VM è pronta per l'utilizzo di PMEM.

Risoluzione dei problemi

In genere, si consiglia di montare questo file system abilitato per DAX utilizzando l'opzione **-o dax** mount, se viene rilevato un errore.

```
[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.
```

Il ripristino del file system viene eseguito per garantire l'integrità.

```
[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]$
```

Per ovviare al problema, è possibile montare il dispositivo di montaggio senza l'opzione **-o dax**.

Nota: In **xfsprogs** versione 5.1, per impostazione predefinita i file system XFS vengono creati con l'opzione **reflink** attivata. In precedenza era disattivato per impostazione predefinita. Le opzioni **reflink** e **dax** si escludono a vicenda, il che provoca un errore di montaggio.

"Impossibile utilizzare contemporaneamente DAX e reflink!" l'errore può essere visualizzato in **dmesg** quando il comando mount non riesce:

```
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]$
```

Per risolvere il problema, rimuovere l'opzione **-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]$
```

Installazione con Next4 FS.

Il file system EXT4 può essere utilizzato come alternativa in quanto non implementa la funzione di collegamento rapido ma supporta 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]$
```

Informazioni correlate

- [Guida introduttiva: Provisioning della memoria persistente CC Intel® Optane™](#)
- [Configurazione della memoria persistente](#)
- [Utility di gestione ipmctl e ndctl per la memoria persistente Intel® Optane™](#)
- [Documentazione e supporto tecnico – Cisco Systems](#)