

在VMware ESXi中為AppDirect模式配置DCPMM

目錄

[簡介](#)

[必要條件](#)

[需求](#)

[採用元件](#)

[背景資訊](#)

[設定](#)

[配置服務配置檔案](#)

[驗證ESXi](#)

[配置虛擬機器NVDIMM](#)

[在虛擬機器中配置名稱空間](#)

[疑難排解](#)

[相關資訊](#)

簡介

本文檔介紹在主機託管模式下使用Intel® Optane™ Persistent Memory(PMEM)在統一計算系統(UCS)B系列伺服器上配置ESXi的過程。

必要條件

需求

思科建議您瞭解以下主題：

- UCS B系列
- 英特爾® Optane™ 資料中心永久記憶體模組(DCPMM)概念
- VMware ESXi和vCenter Server管理

嘗試此組態之前，請確保符合以下要求：

- 請參閱B200/B480 M5規格指南上的PMEM[指南](#)。
- 確保CPU是第二代英特爾®至強®可擴展處理器。
- 根據[KB 67645](#),PMEM/動態隨機存取記憶體(DRAM)比率符合[要求](#)。

- ESXi為6.7 U2 + Express Patch 10(ESXi670-201906002)或更高版本。不支援早期的6.7版本。
- UCS Manager和伺服器為4.0(4)版本或更高版本。有關最新推薦的版本，請訪問www.software.cisco.com/。

採用元件

本文中的資訊係根據以下軟體和硬體版本：

- UCS B480 M5
- UCS管理器4.1(2b)

本文中的資訊是根據特定實驗室環境內的裝置所建立。文中使用到的所有裝置皆從已清除（預設）的組態來啟動。如果您的網路運作中，請確保您瞭解任何指令可能造成的影響。

背景資訊

在配置為App Direct模式的UCS伺服器中，VMware ESXi虛擬機器可以訪問Optane DCPMM永久記憶體非易失性雙列直插式記憶體模組(NVDIMM)。

英特爾Optane DCPMM可通過IPMCTL管理實用程式通過統一可擴展韌體介面(UEFI)外殼或通過作業系統實用程式進行配置。此工具旨在執行以下某些操作：

- 發現和管理模組
- 更新和配置模組韌體
- 監控運行狀況
- 設定和配置目標、區域和名稱空間
- 對PMEM進行調試和故障排除

可以使用連線到服務配置檔案的永久記憶體策略配置UCS，以便於使用。

開源非易失性裝置控制(NDCTL)實用程式用於管理LIBNVDIMM Linux核心子系統。NDCTL實用程式允許系統調配和執行配置作為供作業系統使用的區域和名稱空間。

新增到ESXi主機的永久記憶體由主機檢測、格式化並裝載為本地PMem資料儲存區。為了使用PMEM，ESXi使用虛擬機器飛行系統(VMFS)-L檔案系統格式，並且每個主機僅支援一個本地PMEM資料儲存區。

與其他資料儲存不同，PMEM資料儲存不支援傳統資料儲存任務。具有vmx和vmware.log檔案的VM主目錄不能放在PMEM資料儲存上。

PMEM可以在兩種不同模式下呈現給VM:直接訪問模式和虛擬磁碟模式。

- 直接訪問模式
可以通過以NVDIMM的形式顯示PMEMregion來為此模式配置VM。VM作業系統必須具有PMem感知才能使用此模式。由於NVDIMM充當位元組可定址記憶體，因此儲存在NVDIMM模組上的資料在電源週期中可持續。在形成PMEM時，NVDIMM自動儲存在ESXi建立的PMem資料儲存中。
- 虛擬磁碟模式
適用於駐留在VM上的傳統和舊作業系統，以便支援任何硬體版本。VM作業系統不需要具有PMEM感知能力。在此模式下，VM操作系統可以建立和使用傳統的小型電腦系統介面(SCSI)虛擬磁碟。

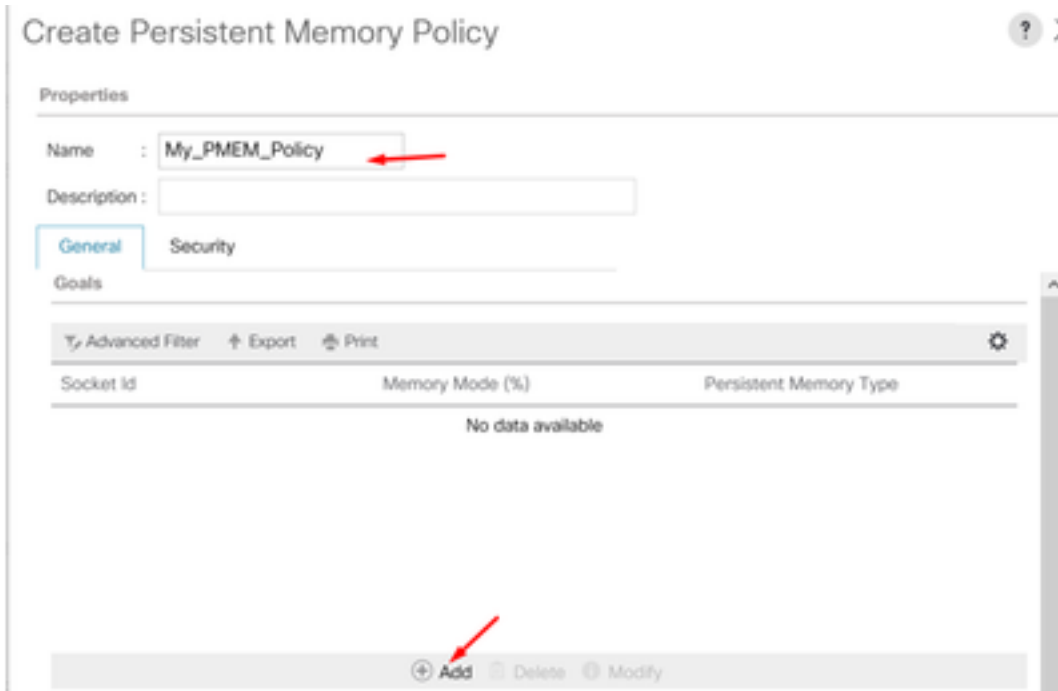
本文檔介紹在直接訪問模式下使用虛擬機器的配置。

設定

此過程介紹如何使用Intel Optane DCPMM在UCS刀片系列伺服器上配置ESXi。

配置服務配置檔案

1. 在UCS Manager GUI中，導航至**Servers > Persistent Memory Policy**，然後點選Add，如下圖所示。



Create Persistent Memory Policy

Properties

Name : My_PMEM_Policy

Description :

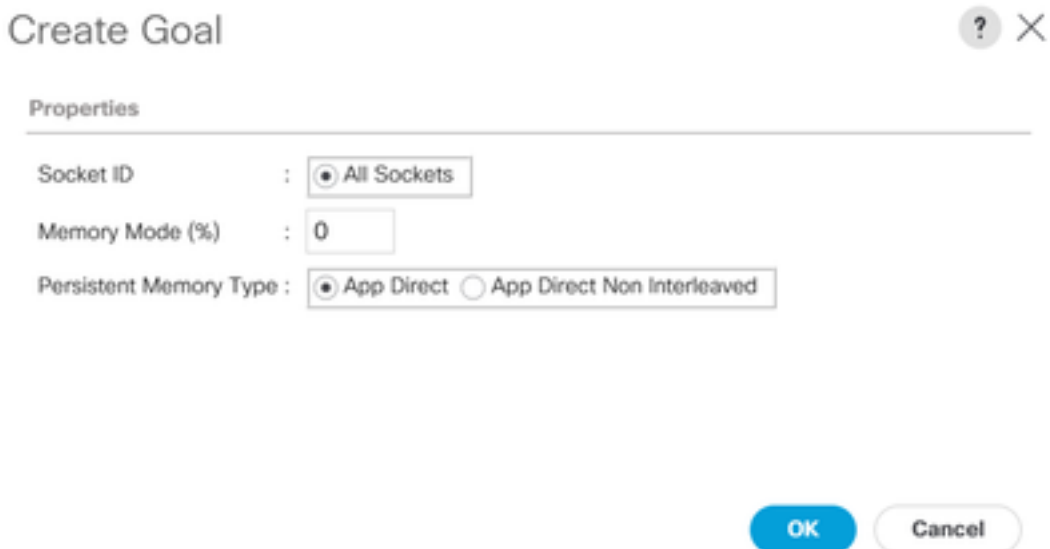
General Security

Goals

| Socket Id | Memory Mode (%) | Persistent Memory Type |
|-------------------|-----------------|------------------------|
| No data available | | |

+ Add - Delete Modify

2. 建立目標，確保Memory Mode為0%，如下圖所示。



Create Goal

Properties

Socket ID : All Sockets

Memory Mode (%) : 0

Persistent Memory Type : App Direct App Direct Non Interleaved

OK Cancel

3. 將PMEM策略新增到所需的服務配置檔案。

導航到**Service Profile > Policies > Persistent Memory Policy**並附加建立的策略。

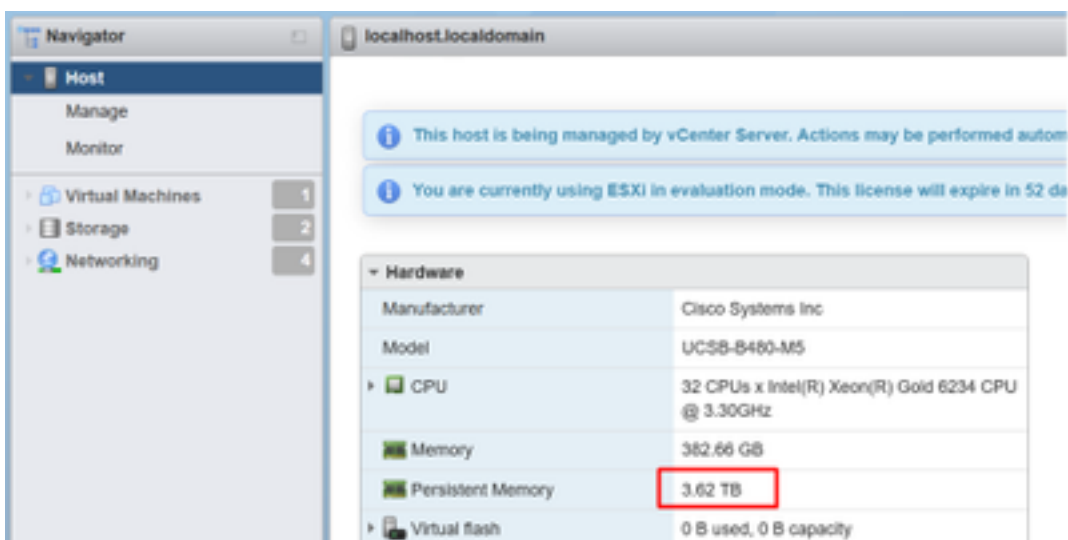
4. 驗證區域的健康狀況。

導航到選定的**Server > Inventory > Persistent Memory > Regions**。AppDirect型別可見。此方法每個CPU插槽建立一個區域。

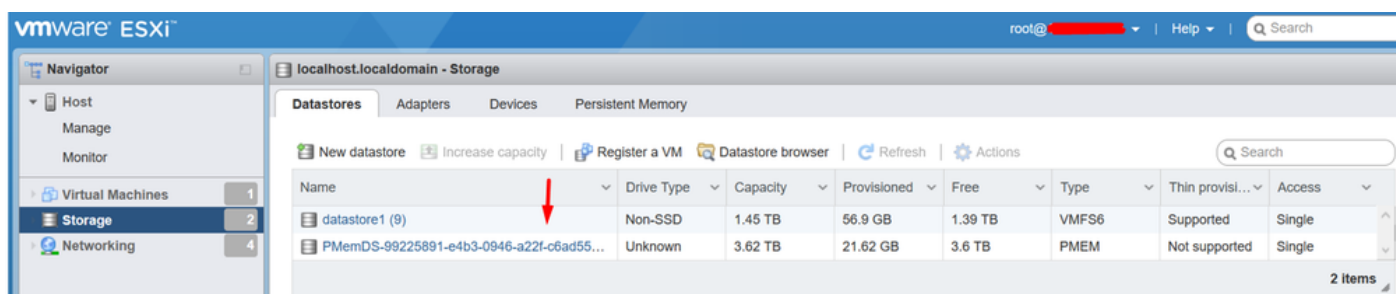
| General Inventory Virtual Machines Installed Firmware CIM Sessions SEL Logs VIF Paths Health Diagnostics File > | | | | | | | |
|---|-----------|-----------------|-----------------|-----------|---------------------|--------------------|---------------|
| 340 CIM CPUs GPUs Memory Adapters HBAs NICs iSCSI vNICs Security Storage Persistent Memory > | | | | | | | |
| DIMM Configuration Regions Namespace | | | | | | | |
| TV-Advanced Filter Export Print | | | | | | | |
| ID | Socket ID | Local DIMM Slot | DIMM Locator ID | 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 |

驗證ESXi

1. 在Web控制檯中，主機顯示可用的PMEM總數。



2. ESXi顯示一個由PMEM總量組成的特殊資料儲存，如下圖所示。



配置虛擬機器NVDIMM

1. 在ESXi中，虛擬機器作為NVDIMM訪問Optane DCPMM PMEM。若要將NVMDIMM分配給虛擬機器，請通過vCenter訪問虛擬機器並導航到操作>編輯設定，按一下新增新裝置，然後選擇NVDIMM，如下圖所示。



附註： 建立虛擬機器時，請確保作業系統相容性滿足支援英特爾®光纖™永久記憶體的最低要求版本，否則可選專案中不顯示NVDIMM選項。

2. 如圖所示設定NVDIMM的大小。



在虛擬機器中配置名稱空間

1. NDCTL實用程式用於管理和配置PMEM或NVDIMM。

在示例中，Red Hat 8用於配置。Microsoft具有用於永久記憶體名稱空間管理的PowerShell cmdlet。

根據Linux發行版使用可用工具下載NDCTL實用程式

例如：

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

2. 驗證預設情況下由ESXi建立的NVDIMM區域和名稱空間，當將NVDIMM分配給虛擬機器時，驗證空間與配置相匹配。請確保將名稱空間的模式設定為**raw**，這表示ESXi已建立名稱空間。若要驗證，請使用命令：

```
# 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. (可選) 如果尚未建立名稱空間，可以使用以下命令建立名稱空間：

```
# ndctl create-namespace
```

預設情況下，`ndctl create-namespace`命令在`fsdax`模式下建立一個新名稱空間，並建立一個`/dev/pmem([x].[y])`裝置。如果已建立名稱空間，則可以跳過此步驟。

4. 選擇PMEM訪問模式，可用於配置的模式包括：

- Sector Mode:

將儲存呈現為快速塊裝置，這對於仍不能使用永久記憶體的傳統應用程式非常有用。

- Fsdax模式：

允許永久記憶體裝置支援直接訪問NVDIMM。檔案系統直接訪問要求使用`fsdax`模式，以便啟用直接訪問程式設計模型。此模式允許在NVDIMM上建立檔案系統。

- Devdax模式：

使用DAX字元裝置提供對永久記憶體的原始訪問。無法在使用`devdax`模式的裝置上建立檔案系統。

- 原始模式：

此模式有多種限制，建議不要使用永續性記憶體。

若要將模式變更為`fsdax`模式，請使用命令：

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

如果已經建立了`dev`，則`dev`名稱空間用於格式化模式並將其修改為`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]$
```

注意：這些命令要求帳戶具有root許可權，可能需要sudo命令。

5. 建立目錄和檔案系統。

直接訪問或DAX是一種機制，允許應用程式直接從CPU訪問永續性介質（通過載入和儲存），繞過傳統的I/O堆疊。支援DAX的永久記憶體檔案系統包括ext4、XFS和Windows NTFS。

建立和裝載的XFS檔案系統示例：

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   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. 裝載檔案系統並驗證是否成功。

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]$
```

VM已準備好使用PMEM。

疑難排解

如果發現錯誤，通常建議使用-o dax掛載選項掛載此啟用DAX的檔案系統。

```
[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.
```

執行檔案系統修復以確保完整性。

```
[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]$
```

作為解決方法，可以不使用 -o dax選項來裝載裝載。

附註：在xfsprogs版本5.1中，預設設定是在啟用reflink選項的情況下建立XFS檔案系統。以前預設情況下禁用此功能。reflink和dax選項互相排斥，從而導致安裝失敗。

「DAX和reflink不能一起使用！」 mount命令失敗時，可在dmesg中看到錯誤：


```
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]$
```

解決方法是，刪除-o dax選項。

```
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]$
```

使用ext4 FS裝載。

EXT4檔案系統可用作替代方案，因為它不實施重新連結功能，但支援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]$
```

相關資訊

- [快速入門手冊：配置英特爾® Optane™ DC永久記憶體](#)
- [永久記憶體配置](#)

- [用於英特爾® Optane™ 永久記憶體的管理實用程式ipmctl和ndctl](#)
- [技術支援與文件 - Cisco Systems](#)