在VMware ESXi for AppDirect模式下配置DCPMM

目录

简介

本文档介绍在主机托管模式下使用英特尔® Optane™持久内存(PMEM)在统一计算系统(UCS)B系列服务器上配置ESXi的过程。

先决条件

要求

Cisco 建议您了解以下主题：

- UCS B系列
- 英特尔® Optane™数据中心永久内存模块(DCPMM)概念
- VMware ESXi和vCenter Server管理

尝试进行此配置之前，请确保满足以下要求：

- 请参阅B200/B480 M5规范指南中的PMEM指南。
- 确保CPU是第二代英特尔®至强®可扩展的处理器。
- PMEM/动态随机访问内存(DRAM)比率符合每KB 67645的要求。

- ESXi是6.7 U2 + Express补丁10(ESXi670-201906002)或更高版本。不支持6.7版本之前的版本。
- UCS Manager和服务器的版本为4.0(4)或更高版本。有关最新推荐版本，请访问www.software.cisco.com/。

使用的组件

本文档中的信息基于以下软件和硬件版本：

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

本文档中的信息都是基于特定实验室环境中的设备编写的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您的网络处于活动状态，请确保您了解所有命令的潜在影响。

背景信息

在为应用直接模式配置的UCS服务器中，VMware ESXi虚拟机访问Optane DCPMM永久内存非易失性双列直插内存模块(NVDIMM)。

英特尔Optane DCPMM可通过IPMCTL管理实用程序通过统一可扩展固件接口(UEFI)外壳或操作系统实用程序进行配置。此工具旨在执行下一些操作：

- 发现和管理模块
- 更新和配置模块固件
- 监控运行状况
- 调配和配置目标、区域和命名空间
- 调试PMEM并排除故障

UCS可以使用附加到服务配置文件的永久内存策略进行配置，以便于使用。

开源非易失性设备控制(NDCTL)实用程序用于管理LIBNVDIMM Linux内核子系统。NDCTL实用程序允许系统将配置调配和执行为用于操作系统的区域和命名空间。

添加到ESXi主机的永久内存由主机检测、格式化并作为本地PMem Datastore装载。为了使用PMEM，ESXi使用虚拟机飞行系统(VMFS)-L文件系统格式，并且每台主机仅支持一个本地PMEM数据存储。

与其他Datastore不同，PMEM Datastore不支持传统Datastore的任务。包含vmx和vmware.log文件的VM主目录无法放置在PMEM数据存储上。

PMEM可以以两种不同模式呈现给虚拟机：直接访问模式和虚拟磁盘模式。

- 直接访问模式
  通过以NVDIMM形式显示PMEM区域，可以为此模式配置VM。VM操作系统必须具有PMem感知才能使用此模式。NVDIMM模块上存储的数据可以在电源周期中持续，因为NVDIMM充当字节可寻址存储器。NVDIMM在创建PMEM时自动存储在由ESXi创建的PMem数据存储上。

- 虚拟磁盘模式
  适用于驻留在VM上的传统和传统操作系统，以支持任何硬件版本。VM操作系统不需要具有PMEM感知。在此模式下，VM操作系统可以创建和使用传统的小型计算机系统接口(SCSI)虚拟磁盘。

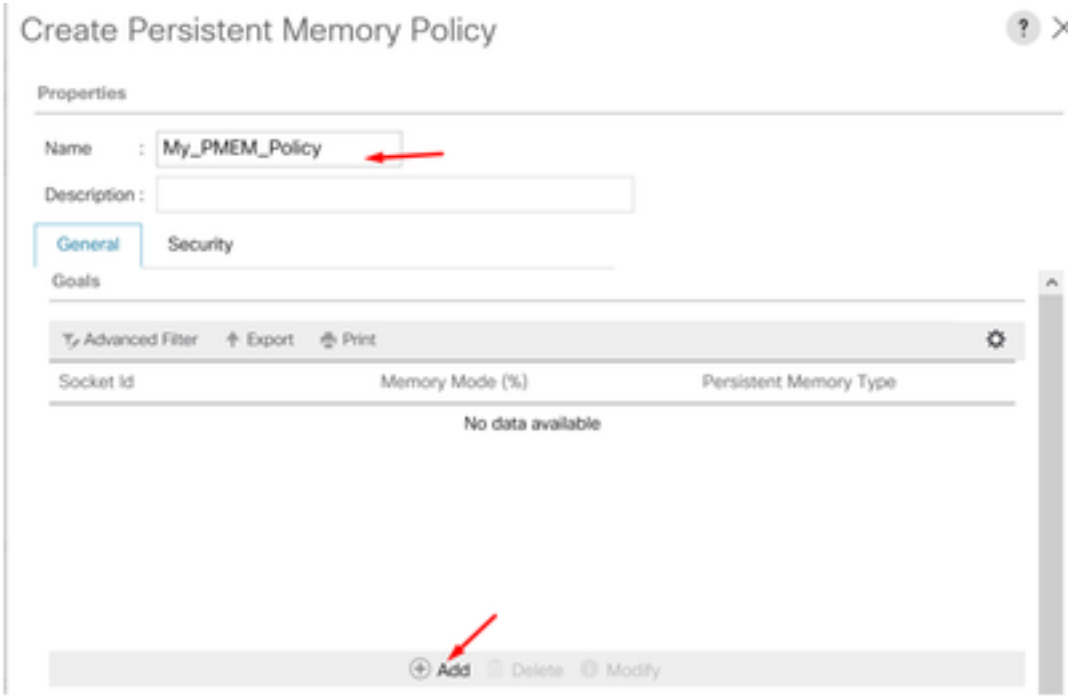
本文档介绍在直接访问模式下使用虚拟机的配置。

配置

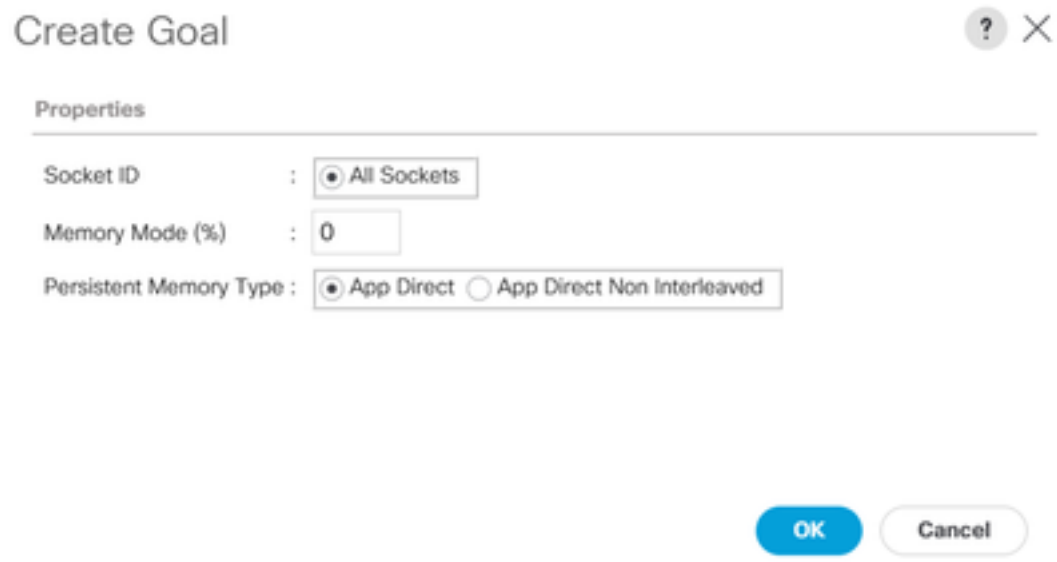
此程序介绍如何使用Intel Optane DCPMM在UCS刀片服务器上配置ESXi。

配置服务配置文件

1.在UCS Manager GUI中，导航至Servers **> Persistent Memory Policy**，然后单**击Add**，如图所示。



2.创建**目标**，确保**内存**模式为0%，如图所示。

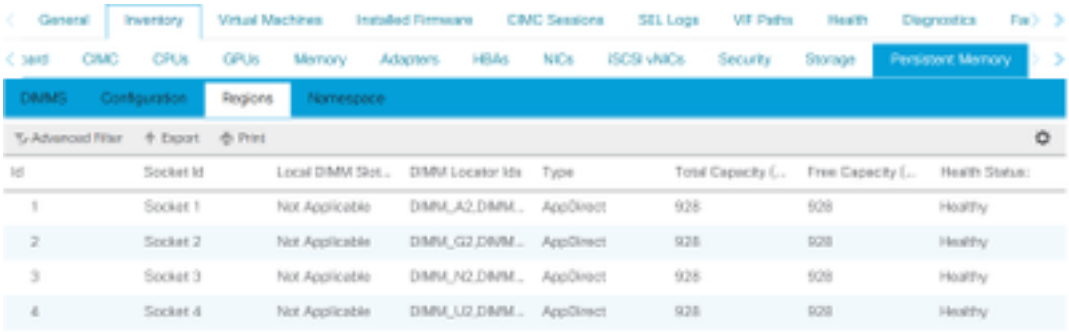


3.将PMEM策略添加到所需的服务配置文件。

导航至**Service Profile > Policies > Persistent Memory Policy**，然后附加创建的策略。

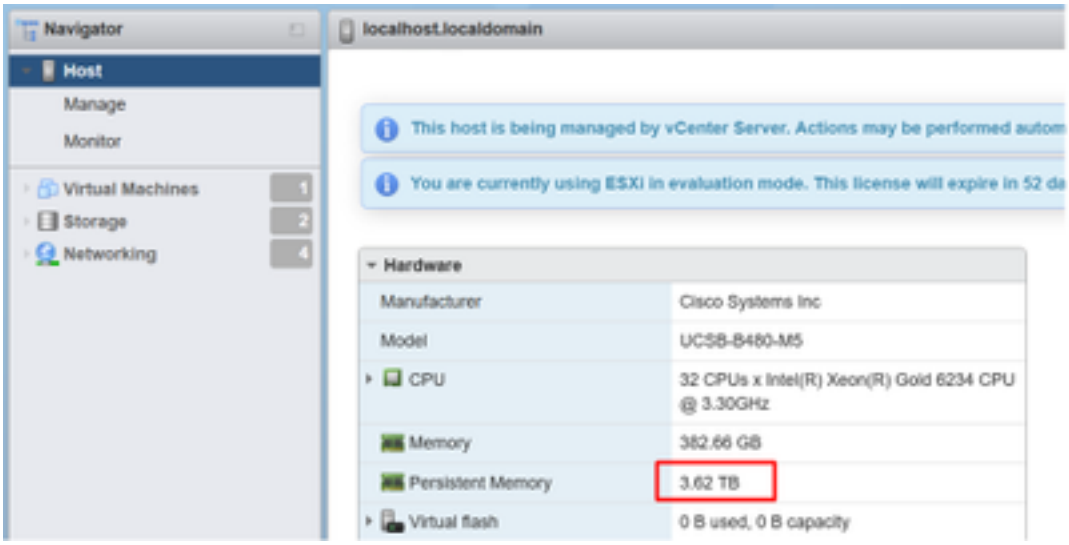
4. 检验区域的健康状况。

导航至选定的服**务器>资产>持久内存>区域**。AppDirect类型可见。此方法为每个CPU插槽创建一个区域。

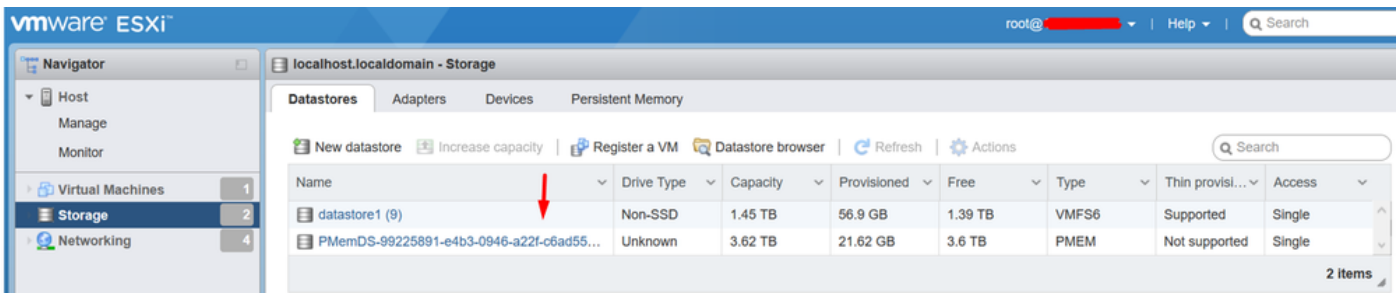


验证ESXi

1.在Web控制台中，主机显示可用PMEM总数。



2. ESXi显示由PMEM总量组成的特殊数据存储，如图所示。



配置虚拟机NVDIMM

1.在ESXi中，虚拟机以NVDIMM形式访问Optane DCPMM PMEM。要将NVMDIMM分配给虚拟机，请通过vCenter访问虚拟机并导航至**操作>编辑设置**，单击**添加新设备**并选择**NVDIMM**，如图所示。

注意： 创建虚拟机时，请确保操作系统兼容性符合支持英特尔® Optane™持久内存的最低要求版本，否则**NVDIMM**选项不会出现在可选项中。

2.设置NVDIMM大小，如图所示。

在虚拟机中配置命名空间

1. NDCTL实**用程**序用于管理和配置PMEM或NVDIMM。

在本例中，Red Hat 8用于配置。Microsoft具有用于持久内存命名空间管理的PowerShell cmdlet。

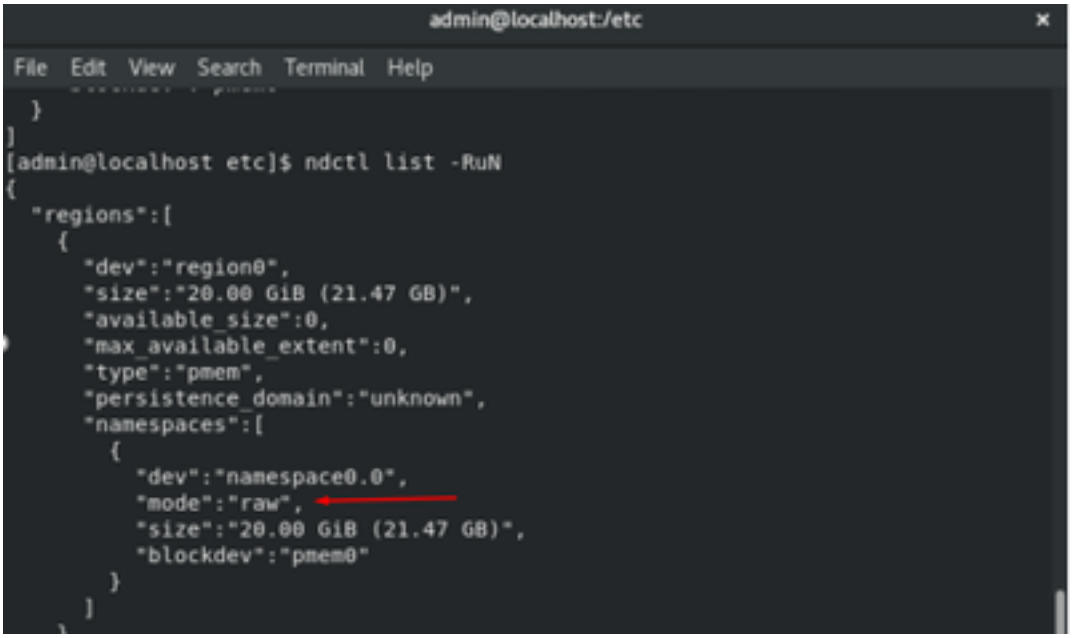
根据Linux发行版，使用**可用工具**下载NDCTL实用程序

例如：

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

2. 验证ESXi默认创建的NVDIMM区域和命名空间，当NVDIMM分配到虚拟机时，验证空间与配置的匹配。确保命名空间的模式设置为原始，**这**意味着ESXi已创建命名空间。要进行验证，请使用以下命令：

```
# ndctl list -RuN
```



3.（可选）如果尚未创建命名空间，则可以使用以下命令创建命名空间：

```
# ndctl create-namespace
```

默认情况下，ndctl create-namespace命令在fsdax模式下创建新命名空间，并创建新的/dev/pmem([x].[y])设备。如果已创建命名空间，则可跳过此步骤。

4.选择PMEM访问模式，可用于配置的模式为：

- 扇区模式：

  将存储显示为快速块设备，这对于仍无法使用持久内存的旧式应用程序非常有用。

- Fsdax模式：

  允许永久内存设备支持直接访问NVDIMM。文件系统直接访问需要使用fsdax模式，以便启用直接访问编程模型。此模式允许在NVDIMM顶部创建文件系统。

- Devdax模式：

  使用DAX字符设备提供对永久内存的原始访问。 无法使用devdax模式在设备上创建文件系统。

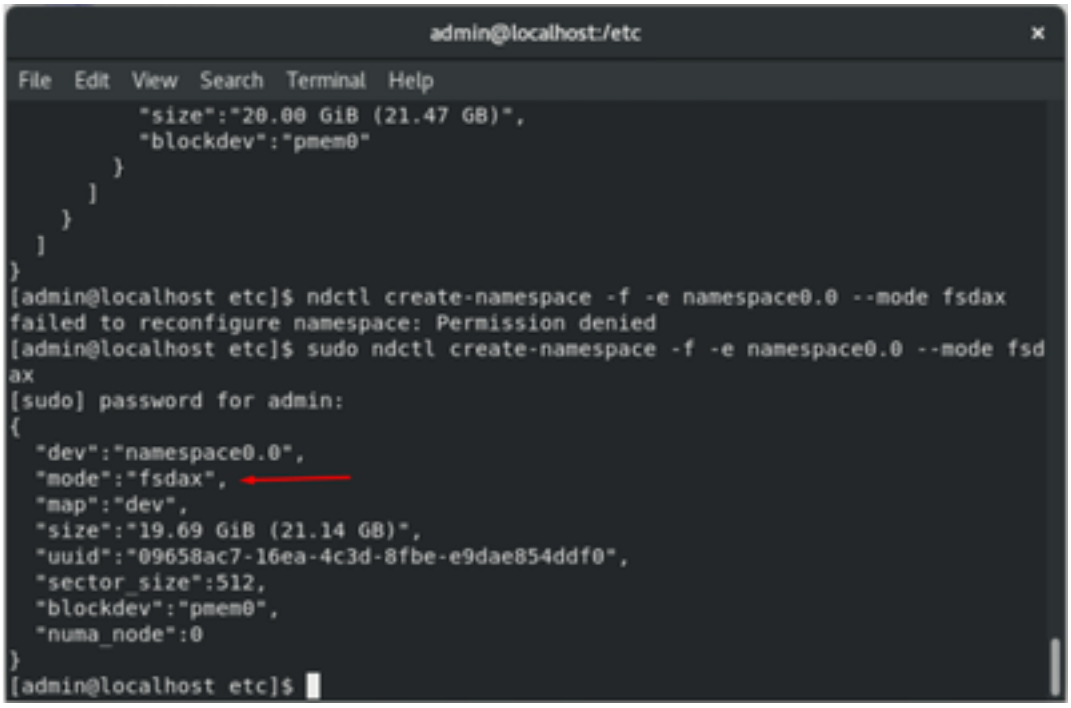
- 原始模式:

  此模式有几个限制，不建议使用持久内存。

要将模式更改为fsdax模式，请使用以下命令：

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

如果已创建dev，则使用dev命名空间将模式格式化并修改为fsdax。



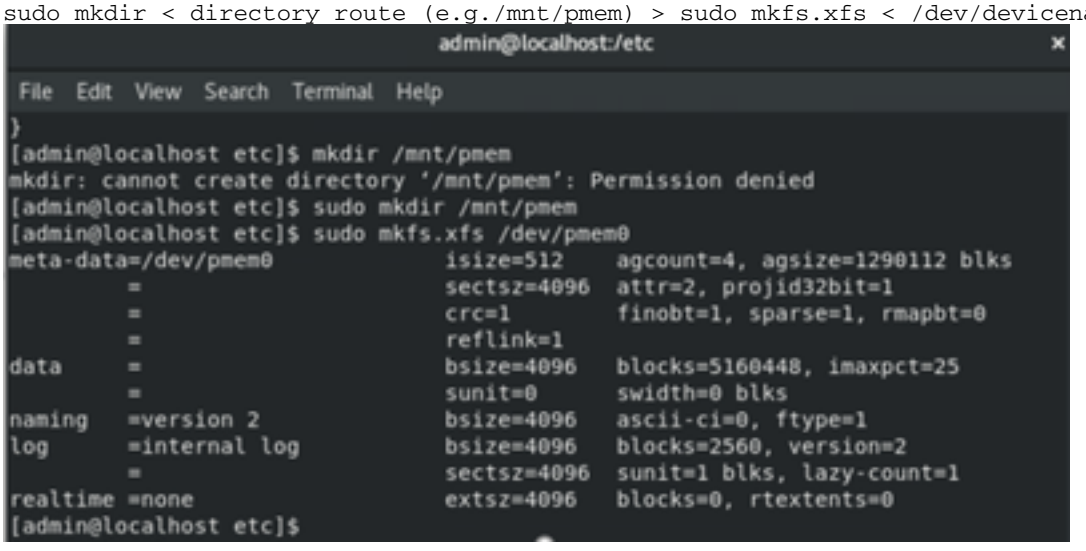
注意：这些命令要求帐户具有根权限，**可**能需要sudo命令。

5.创建目录和文件系统。

直接访问或DAX是一种机制，允许应用程序通过CPU（通过加载和存储）直接访问持久介质，而绕过传统I/O堆栈。支持DAX的永久内存文件系统包括ext4、XFS和Windows NTFS。

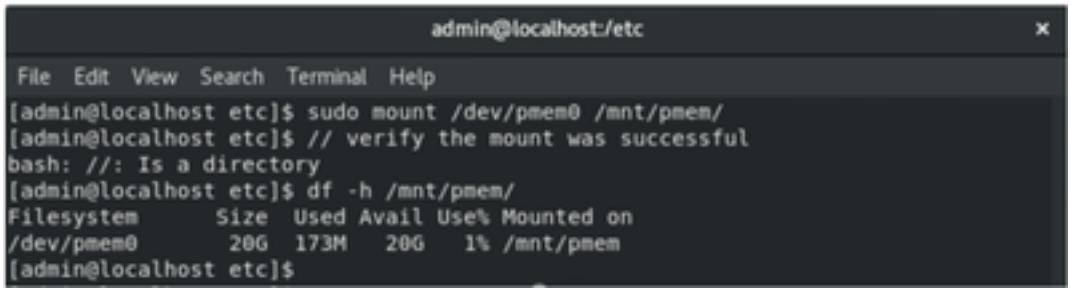
创建和装载的XFS文件系统示例：

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



6.安装文件系统并验证是否成功。

```
sudo mount
```



VM已准备好使用PMEM。

故障排除

如果发现错误，通常建议使用 — o dax装载选**项装**载此启用DAX的文件系统。

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

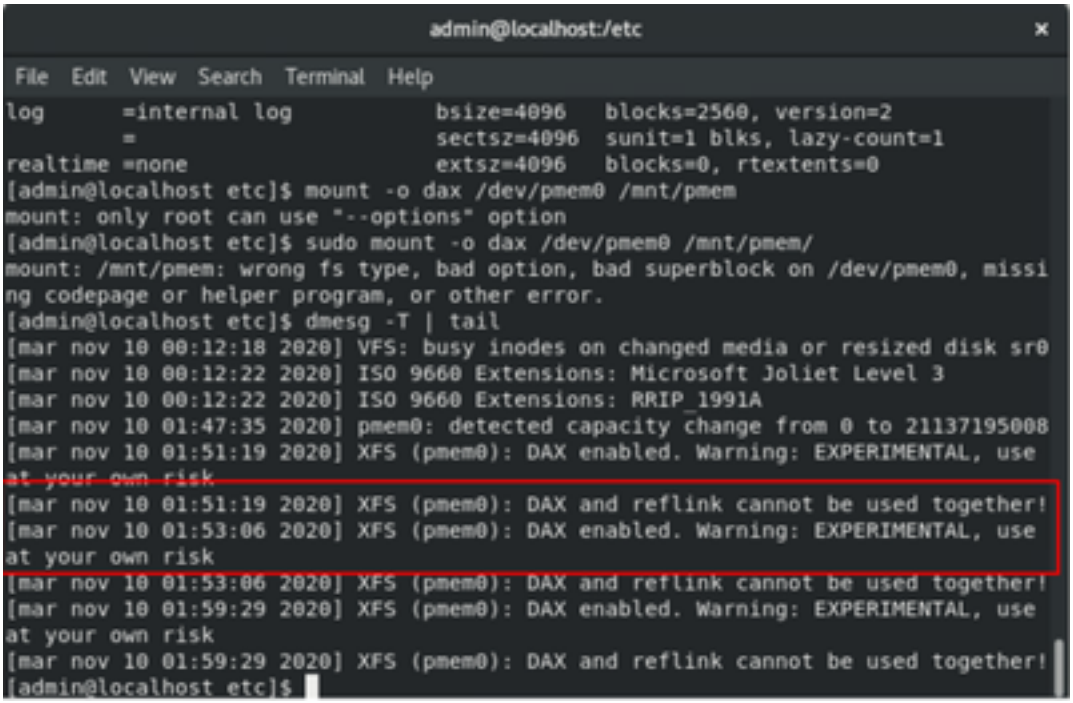
执行文件系统修复以确保完整性。

```
[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]$
```

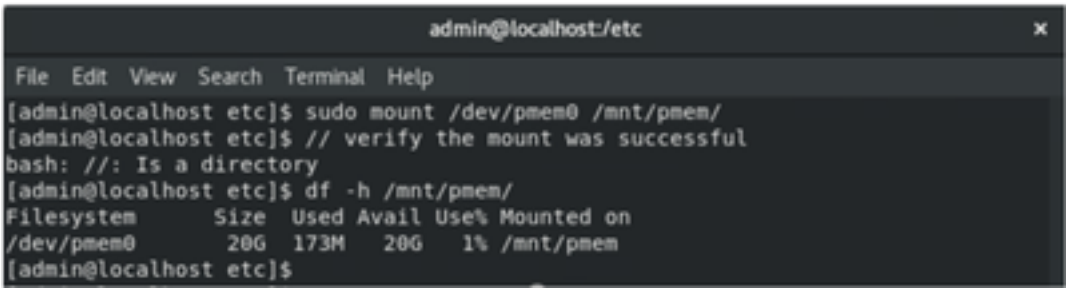
解决方法是，安装时无需使用**-o dax选**项。

> **注意**：在**xfsprogs** 5.1版中，默认为在启用reflink选项的情况下创建**XFS文**件系统。以前默认禁用了它。**repsink**和**dax**选项是互斥的，这会导致安装失败。

“DAX和反光灯不能一起使用！”当mount命令失败时，**dmesg**中会显示错误：

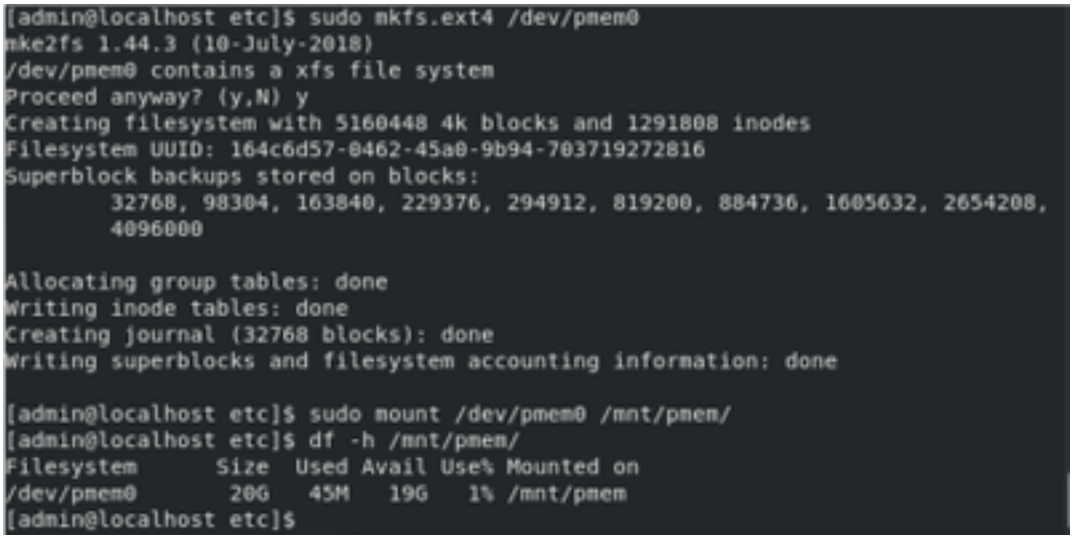


作为解决方法，请删除**-o dax选**项。



使用ext4 FS装载。

EXT4文件系统可用作替代系统，因为它不实施反射功能，但支持DAX。



相关信息

- [快速入门指南：调配英特尔® Optane™ DC永久内存]
- [持久内存配置]

- 用于英特尔® Optane™持久内存的管理实用程序ipmctl和ndctl
- 技术支持和文档 - Cisco Systems