

# 排除CNDP群集管理器HA故障並恢復

## 目錄

[簡介](#)

[必要條件](#)

[需求](#)

[採用元件](#)

[背景資訊](#)

[什麼是SMI?](#)

[什麼是SMI-BM或CNDP?](#)

[什麼是SMI群集管理器？](#)

[什麼是DRBD?](#)

[問題](#)

[維護程式](#)

## 簡介

本檔案介紹對雲原生部署平台(CNDP)叢集管理器中的高可用性(HA)問題進行疑難排解的程式。

## 必要條件

### 需求

思科建議您瞭解以下主題：

- 思科使用者微服務基礎架構(SMI)
- 5G CNDP或SMI裸機(BM)架構
- 分散式複製區塊裝置(DRBD)

### 採用元件

本文中的資訊係根據以下軟體和硬體版本：

- SMI 2020.02.2.35
- Kubernetes v1.21.0
- DRBD 8.9.10

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

### 什麼是SMI?

Cisco SMI是雲技術和標準的分層堆疊，支援來自思科移動性、電纜和寬頻網路網關(BNG)業務部門的基於微服務的應用 — 所有這些業務部門都具有相似的使用者管理功能和相似的資料儲存要求。

屬性：

- 層雲堆疊（技術和標準），提供自上而下的部署，並適應當前的使用者雲基礎設施。
- 所有應用程式共用通用執行環境，用於非應用程式功能（資料儲存、部署、配置、遙測、警報）。這樣可為所有使用者觸控點和整合點提供一致的互動和體驗。
- 應用和常見執行環境部署於微服務容器中，並與智慧服務網狀網連線。
- 公開的API用於部署、配置和管理，以實現自動化。

## 什麼是SMI-BM或CNDP?

思科SMI-BM或CNDP是一個託管裸機平台，提供部署虛擬網路功能(VNF)和雲原生功能(CNF)的基礎設施，支援思科移動性、電纜和BNG業務部門。

屬性：

- BM可消除與Virtualized Infrastructure Manager(VIM)相關的開銷。
- 效能提高：更多應用核心更快的應用程式執行
- 自動化部署工作流程，與Network Services Orchestrator(NSO)Core Function Packet(CFP)整合
- 用於部署思科5G網路功能(NF)的託管堆疊
- 簡化訂單和部署指南

## 什麼是SMI群集管理器？

群集管理器是2節點 `keepalived` 集群用作控制平面和使用者平面集群部署的初始點。它運行一個單節點Kubernetes群集和一組POD，它們負責整個群集設定。只有主群集管理器處於活動狀態，輔助群集管理器僅在出現故障或因維護而手動關閉時進行接管。

## 什麼是DRBD?

DRBD用於提高資料的可用性。它是一個基於Linux的開放原始碼軟體元件，便於網路映象替換共用儲存系統。簡而言之，您可以說這是「基於網路的資料的Raid 1映象」。

## 問題

群集管理器託管在具有分散式複製塊裝置(DRBD)的2節點群集中，`keepalived`. HA可以折斷，並且也可以進入大腦分裂狀態。此過程有助於恢復損壞的群集。群集管理器HA的所需狀態是 `cluster manager1` (CM1) 是主要和 `cluster manager2` (CM2) 是次要的。這裡CM1是腦分裂的受害者。

## 維護程式

登入到群集管理器並檢查DRBD群集的狀態。

```
cloud-user@pod-name-cm-1:~$ drbd-overview
0:data/0  WFConnection Secondary/Unknown UpToDate/DUnknown
```

```
cloud-user@pod-name-cm-2:~$ drbd-overview
```

```
0:data/0 StandAlone Primary/Unknown UpToDate/DUnknown /mnt/stateful_partition ext4 568G 147G 392G 28%
```

在此方案中，CM2為主群集處於獨立模式。CM1當前處於輔助狀態並處於 **wait for connection** 狀態。

以下是群集的正确狀態：

```
cloud-user@pod-name-deployer-cm-1:~$ drbd-overview
```

```
0:data/0 Connected Primary/Secondary UpToDate/UpToDate /mnt/stateful_partition ext4 568G 364G 176G 68%
```

```
cloud-user@pod-name-deployer-cm-2:~$ drbd-overview
```

```
0:data/0 Connected Secondary/Primary UpToDate/UpToDate Move the CM VIP to CM-1 from CM-2 and make CM-1 as primary -
```

將CM VIP從CM-2移動到CM-1，並將CM-1作為主節點(集群管理器VIP為 10.x.x.65)

On CM-2 issue below command

```
cloud-user@pod-name-cm-2:~$sudo systemctl restart keepalived
```

On CM-1 issue below command (Make sure that the VIP is now switched over to CM-1)

```
cloud-user@pod-name-cm-1:~$ip a s | grep 10.x.x
```

```
inet 10.x.x.70/26 brd 10.x.x.127 scope global vlan1xx ----> here is the server IP
```

```
inet 10.x.x.65/26 scope global secondary vlan1xx. ----> here is the VIP
```

確定DRBD資源 ( 通過網路共用 )：

```
cloud-user@pod-name-deployer-cm-1:/$ cat /etc/fstab
```

```
#/mnt/stateful_partition/data /data none defaults,bind 0 0 ----> /data is the resource
```

```
#/mnt/stateful_partition/home /home none defaults,bind 0 0
```

```
cloud-user@pod-name-deployer-cm-1:/$ df -h
```

```
Filesystem Size Used Avail Use% Mounted on
```

```
udev 189G 0 189G 0% /dev
```

```
tmpfs 38G 22M 38G 1% /run
```

```
/dev/sda1 9.8G 3.5G 5.9G 37% /
```

```
tmpfs 189G 0 189G 0% /dev/shm
```

```
tmpfs 5.0M 0 5.0M 0% /run/lock
```

```
tmpfs 189G 0 189G 0% /sys/fs/cgroup
```

```
/dev/sda4 71G 1.5G 66G 3% /tmp
```

```
/dev/sda3 71G 11G 57G 16% /var/log
```

```
/dev/drbd0 568G 365G 175G 68% /mnt/stateful_partition -->/dev/drbd0 is the device name
```

```
tmpfs 38G 0 38G 0% /run/user/1000
```

```
cloud-user@pod-name-deployer-cm-1:/$ lsblk
```

```
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT
```

```
sda 8:0 0 744.1G 0 disk
```

```
sda1 8:1 0 10G 0 part /
```

```
sda2 8:2 0 10G 0 part
```

```
sda3 8:3 0 72.2G 0 part /var/log
```

```
sda4 8:4 0 72.2G 0 part /tmp
```

```
sda5 8:5 0 577.6G 0 part
```

```
drbd0 147:0 0 577.5G 0 disk /mnt/stateful_partition ----> /dev/sda5 is used to create drbd0
```

檢查DRBD配置檔案以獲取資源詳細資訊：

```
cloud-user@pod-name-deployer-cm-1:/$ cat /etc/drbd.d/data.res
```

```
resource data {
```

protocol C; --->**Synchronous replication protocol. Local write operations on the primary node are considered completed only after both the local and the remote disk write have been confirmed. As a result, loss of a single node is guaranteed not to lead to any data loss**

....  
....

```
device /dev/drbd0;  
disk /dev/disk/by-partlabel/smi-state; --> This translates to /dev/sda5  
meta-disk internal;  
floating 10.192.1.2:7789;  
floating 10.192.1.3:7789;
```

現在，執行DRBD恢復：

On CM-2

```
cloud-user@pod-name-cm-2:~$ sudo systemctl stop keepalived ---> stop to avoid VRRP VIP switchover  
cloud-user@pod-name-cm-2:~$ sudo drbdadm disconnect data ---> data is the cluster resource  
cloud-user@pod-name-cm-2:~$ sudo drbdadm secondary data ---> Make it secondary manually  
cloud-user@pod-name-cm-2:~$ sudo drbdadm connect --discard-my-data data ---> Force discard of all modifications on the split brain victim  
cloud-user@pod-name-cm-2:~$ drbd-overview status
```

On CM-1:

```
cloud-user@pod-name-cm-1:~$ sudo systemctl stop keepalived ---> stop to avoid VRRP VIP switchover  
cloud-user@pod-name-cm-1:~$ sudo drbdadm connect data ---> Data will be connected as primary  
cloud-user@pod-name-cm-1:~$ drbd-overview status
```

開始 **keepalived** 兩個控制器上的進程。借助於的**VRRP keepalived** 根據連線的主資源選擇CM-1作為主資源 **/data**:

```
cloud-user@pod-name-cm-1:~$ sudo systemctl start keepalived  
cloud-user@pod-name-cm-1:~$ sudo systemctl status keepalived
```

```
cloud-user@pod-name-cm-2:~$ sudo systemctl start keepalived  
cloud-user@pod-name-cm-2:~$ sudo systemctl status keepalived
```

檢查CM-1和CM-2上的DRBD狀態。現在必須將其轉換為正確的群集狀態。

```
cloud-user@pod-name-deployer-cm-1:~$ drbd-overview  
0:data/0 Connected Primary/Secondary UpToDate/UpToDate /mnt/stateful_partition ext4 568G 364G  
176G 68%
```

```
cloud-user@pod-name-deployer-cm-2:~$ drbd-overview  
0:data/0 Connected Secondary/Primary UpToDate/UpToDate Move the CM VIP to CM-1 from CM-2 and  
make CM-1 as primary
```

**/data** 僅在主節點上裝載。

```
cloud-user@pod-name-deployer-cm-1:~$ df -h  
Filesystem Size Used Avail Use% Mounted on  
udev 189G 0 189G 0% /dev  
tmpfs 38G 22M 38G 1% /run  
/dev/sda1 9.8G 3.5G 5.9G 37% /  
tmpfs 189G 0 189G 0% /dev/shm  
tmpfs 5.0M 0 5.0M 0% /run/lock  
tmpfs 189G 0 189G 0% /sys/fs/cgroup
```

```
/dev/sda4 71G 1.5G 66G 3% /tmp
/dev/sda3 71G 11G 57G 16% /var/log
/dev/drbd0 568G 364G 175G 68% /mnt/stateful_partition
tmpfs 38G 0 38G 0% /run/user/1000
```

```
cloud-user@pod-name-deployer-cm-secondary:~$ df -h
```

```
Filesystem Size Used Avail Use% Mounted on
```

```
udev 189G 0 189G 0% /dev
tmpfs 38G 2.3M 38G 1% /run
/dev/sda1 9.8G 2.0G 7.3G 22% /
tmpfs 189G 0 189G 0% /dev/shm
tmpfs 5.0M 0 5.0M 0% /run/lock
tmpfs 189G 0 189G 0% /sys/fs/cgroup
/dev/sda3 71G 9.3G 58G 14% /var/log
/dev/sda4 71G 53M 67G 1% /tmp
tmpfs 38G 0 38G 0% /run/user/1000
```

## 關於此翻譯

思科已使用電腦和人工技術翻譯本文件，讓全世界的使用者能夠以自己的語言理解支援內容。請注意，即使是最佳機器翻譯，也不如專業譯者翻譯的內容準確。Cisco Systems, Inc. 對這些翻譯的準確度概不負責，並建議一律查看原始英文文件（提供連結）。