

排除CNDP群集管理器HA问题并恢复

目录

[简介](#)

[先决条件](#)

[要求](#)

[使用的组件](#)

[背景信息](#)

[什么是SMI?](#)

[什么是SMI-BM或CNDP?](#)

[什么是SMI Cluster Manager?](#)

[什么是DRBD?](#)

[问题](#)

[维护程序](#)

简介

本文档介绍对云本地部署平台(CNDP)集群管理器中的高可用性(HA)问题进行故障排除的过程。

先决条件

要求

Cisco 建议您了解以下主题：

- 思科用户微服务基础设施(SMI)
- 5G CNDP或SMI裸机(BM)架构
- 分布式复制块设备(DRBD)

使用的组件

本文档中的信息基于以下软件和硬件版本：

- SMI 2020.02.2.35
- Kubernetes v1.21.0
- DRBD 8.9.10

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

什么是SMI?

Cisco SMI是云技术和标准的分层堆栈，支持来自思科移动性、电缆和宽带网络网关(BNG)业务部门的基于微服务的应用 — 所有这些业务部门都具有相似的用户管理功能和相似的数据存储要求。

Attributes:

- 层云堆栈（技术和标准），提供自上而下的部署，并适应当前用户云基础设施。
- 所有应用程序共享通用执行环境，用于非应用程序功能（数据存储、部署、配置、遥测、警报）。这样可为所有用户接触点和集成点提供一致的交互和体验。
- 应用和常见执行环境部署在微服务容器中，并与智能服务网状网连接。
- 公开的API，用于部署、配置和管理，以实现自动化。

什么是SMI-BM或CNDP?

思科SMI-BM或CNDP是一个托管裸机平台，提供部署虚拟网络功能(VNF)和云原生功能(CNF)的基础设施，支持思科移动性、电缆和BNG业务部门。

Attributes:

- 消除与虚拟基础设施管理器(VIM)相关开销的BM。
- 性能改善: 更多应用核心更快的应用程序执行
- 自动化部署工作流程，与网络服务协调器(NSO)核心功能数据包(CFP)集成
- 用于部署思科5G网络功能(NF)的托管堆栈
- 简化的订单和部署指南

什么是SMI Cluster Manager?

集群管理器是2节点 `keepalived` 集群用作控制平面和用户平面集群部署的初始点。它运行一个单节点 Kubernetes集群和一组POD，负责整个集群设置。只有主群集管理器处于活动状态，辅助群集管理器仅在出现故障时进行接管，或手动关闭以进行维护。

什么是DRBD?

DRBD用于提高数据的可用性。它是一个基于Linux的开源软件组件，便于网络镜像替换共享存储系统。简而言之，您可以说这是“基于网络的Raid 1数据镜像”。

问题

集群管理器托管在带有分布式复制块设备(DRBD)的双节点集群中，`keepalived`. HA可以破碎，也可以进入大脑分裂状态。此过程有助于恢复中断的群集。集群管理器HA的所需状态是 `cluster manager1` (CM1) 是主要和 `cluster manager2` (CM2) 是次要的。这里，CM1是脑分裂的受害者。

维护程序

登录集群管理器并检查DRBD集群的状态。

```
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%
```

在此场景中，CM2为主用而集群处于独立模式。CM1当前处于辅助状态并处于 **wait for connection** 状态。

集群的正确状态如下：

```
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 -
```

将CM VIP从CM-2移至CM-1，并将CM-1作为主设备(集群管理器VIP为 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
```

确定DRBD资源 (通过网络共享)：

```
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
```

检查DRBD配置文件以获取资源详细信息：

```
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;
```

现在，执行DRBD恢复：

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
```

开始 **keepalived** 两个配置模块上的进程。借助于，**VRRP keepalived** 根据连接的主资源选择CM-1作为主资源 **/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
```

检查CM-1和CM-2上的DRBD状态。现在必须将其转换为正确的群集状态。

```
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 仅在主节点上装载。

```
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
```

关于此翻译

思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。