

Solucionar problemas de HA do CNDP Cluster Manager e recuperar

Contents

[Introduction](#)

[Prerequisites](#)

[Requirements](#)

[Componentes Utilizados](#)

[Informações de Apoio](#)

[O que é SMI?](#)

[O que é SMI-BM ou CNDP?](#)

[O que é um gerenciador de cluster SMI?](#)

[O que é DRBD?](#)

[Problema](#)

[Procedimento de manutenção](#)

Introduction

Este documento descreve o procedimento para solucionar problemas de Alta Disponibilidade (HA) no gerenciador de cluster da Plataforma de Implantação Nativa na Nuvem (CNDP).

Prerequisites

Requirements

A Cisco recomenda que você tenha conhecimento destes tópicos:

- Infraestrutura Cisco Subscriber Microservices (SMI)
- Arquitetura 5G CNDP ou SMI-Bare-metal (BM)
- DRBD (Distributed Replicated Block Device, dispositivo de bloco replicado distribuído)

Componentes Utilizados

As informações neste documento são baseadas nestas versões de software e hardware:

- SMI 2020.2.2.35
- Kubernetes v1.21.0
- DRBD 8 9 10

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. Se a rede estiver ativa, certifique-se de que você entenda o impacto potencial de qualquer comando.

Informações de Apoio

O que é SMI?

O Cisco SMI é uma pilha em camadas de tecnologias e padrões de nuvem que permitem aplicativos baseados em microsserviços das unidades de negócios Cisco Mobility, Cable e Broadband Network Gateway (BNG) - todas com funções semelhantes de gerenciamento de assinantes e requisitos semelhantes de armazenamento de dados.

Atributos:

- Pilha de nuvem em camadas (tecnologias e padrões) para fornecer implantações de cima para baixo e também acomodar a infraestrutura atual de nuvem do usuário.
- Ambiente de execução comum compartilhado por todos os aplicativos para funções que não sejam de aplicativos (armazenamento de dados, implantação, configuração, telemetria, alarme). Isso fornece interação e experiência consistentes para todos os pontos de contato e pontos de integração do usuário.
- Aplicativos e Common Execution Environments são implantados em contêineres de microsserviço e conectados com uma Intelligent Service Mesh.
- API exposta para implantação, configuração e gerenciamento, a fim de permitir a automação.

O que é SMI-BM ou CNDP?

O Cisco SMI-BM ou CNDP é uma plataforma bare-metal com curadoria que fornece a infraestrutura para implantar Funções de Rede Virtual (VNF - Virtual Network Functions) e Funções Nativas de Nuvem (CNFs - Cloud-Native Functions), que possibilita a mobilidade, o cabo e as unidades de negócios BNG da Cisco.

Atributos:

- BM que elimina a sobrecarga relacionada ao Virtualized Infrastructure Manager (VIM).
- Melhor desempenho: Mais núcleos para aplicativos Execução mais rápida de aplicativos
- Fluxo de trabalho de implantação automatizado, integrado com o Pacote de Função Central (CFP - Core Function Packet) do Network Services Orchestrator (NSO)
- Pilha com curadoria para implantar funções de rede (NFs) Cisco 5G
- Guia simplificado de pedidos e implantação

O que é um gerenciador de cluster SMI?

Um gerenciador de cluster é um `keepalived` cluster usado como o ponto inicial para implantação de cluster de plano de controle e plano de usuário. Ele executa um cluster Kubernetes de nó único e um conjunto de PODs que são responsáveis por toda a configuração do cluster. Somente o gerenciador de cluster primário está ativo e o secundário assume somente em caso de falha ou é desativado manualmente para manutenção.

O que é DRBD?

O DRBD é usado para aumentar a disponibilidade dos dados. É um componente de software de código aberto baseado em Linux que facilita a substituição de sistemas de armazenamento compartilhado por um espelho de rede. Resumindo, você pode dizer que este é um "espelho RAID 1 baseado em rede para os dados".

Problema

O gerenciador de cluster é hospedado em um cluster de 2 nós com DRBD e `keepalived`. O HA pode se quebrar e também pode entrar em um estado `split-brain`. Este procedimento ajuda a recuperar o cluster quebrado. O estado desejado de HA do gerenciador de cluster é que `cluster manager1 (CM1)` é principal e `cluster manager2 (CM2)` é secundário. Aqui, o CM1 é a vítima do `split-brain`.

Procedimento de manutenção

Faça login no gerenciador de cluster e verifique o status do cluster DRBD.

```
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%
```

Neste cenário, o CM2 é primário e o cluster está no modo autônomo. CM1 é atualmente secundário e está em `wait for connection` estado.

Este é o estado correto do cluster:

```
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 -
```

Mova o CM VIP de CM-2 para CM-1 e torne CM-1 primário (o VIP do gerenciador de cluster é `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 vlanlxx ----> here is the server IP
inet 10.x.x.65/26 scope global secondary vlanlxx. ----> here is the VIP
```

Identificar o recurso DRBD (compartilhado pela rede):

```
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
```

Verifique o arquivo de configuração DRBD para obter os detalhes do recurso:

```
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;
```

Agora, execute a recuperação DRBD:

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
```

Iniciar **keepalived** processo em ambos os CMs. VRRP com a ajuda de **keepalived** seleciona CM-1 como principal, com base no recurso principal conectado **/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
```

Verifique o status DRBD em CM-1 e CM-2. Ele deve ser transformado para o estado de cluster correto agora.

```
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 é montado somente no nó primário.

```
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
```

Sobre esta tradução

A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês ([link fornecido](#)) seja sempre consultado.