Implante o DNS VNF com a rede SRIOV no Openstack CVIM - Exemplo de configuração para o Prime Network Registrar (DNS)

Contents

Introdução

Este documento descreve a implantação passo a passo da CPNR no OpenStack Cisco Virtualized Infrastructure Manager (CVIM) usando SR-IOV e vinculação de backup ativo.

Pré-requisitos

Requisitos

A Cisco recomenda que você tenha conhecimento destes tópicos:

- Familiaridade com o OpenStack e conceitos de Single Root Input/Output Virtualization (SR-IOV)
- Conhecimento prático sobre o Cisco Virtual Interface Manager (VIM), o Cisco Elastic Services Controller e os comandos e a rede Linux

Componentes Utilizados

Este documento não se restringe a versões de software e hardware específicas.

As informações neste documento foram criadas a partir de dispositivos em um ambiente de laboratório específico. Todos os dispositivos utilizados neste documento foram iniciados com uma configuração (padrão) inicial. Se a rede estiver ativa, certifique-se de que você entenda o impacto potencial de qualquer comando

Informações de Apoio

No cenário de rede atual, as Funções de Rede Virtual (VNFs) desempenham um papel fundamental na habilitação de serviços de rede ágeis, escaláveis e eficientes. Para VNFs que exigem conectividade de rede de alto desempenho, o SR-IOV é uma tecnologia comumente usada. O SR-IOV permite que as VNFs ignorem o switch virtual do hipervisor e acessem diretamente os recursos da placa de rede (NIC) física, reduzindo assim a latência e aumentando o throughput.

Configuração

Antes de prosseguir com a implantação, verifique se esses pré-requisitos foram atendidos.

1. Requisitos de Hardware

- Placas de rede compatíveis com SR-IOV:
  - Pelo menos duas placas de rede físicas com capacidade de SR-IOV com SR-IOV habilitado no BIOS/Unified Extensible Firmware Interface (UEFI).
  - Exemplo:sriov0mapeado para o nó 0 de acesso à memória não uniforme (NUMA) e sriov1mapeado para o nó 1 de NUMA.

- Hosts compatíveis com NUMA:
  - Os nós de computação devem suportar a arquitetura NUMA.
  - O suporte a NUMA deve ser habilitado no BIOS/UEFI do host.

2. Identificação das Placas de Rede Intel

As placas de rede Intel XL710 e E810CQDA2 são comumente usadas para redes SR-IOV de alto

desempenho. Para verificar o modelo da placa de rede no host, consulte estas etapas:

Etapa 1. Usando o Comando lspci

Execute este comando para listar os dispositivos PCI (Peripheral Component Interconnect) relacionados às controladoras de rede:

```
lspci | grep -i ethernet
```

Exemplo de saída:

```
81:00.0 Ethernet controller: Intel Corporation Ethernet Controller XL710 for 40GbE QSFP+ (rev 02)
82:00.0 Ethernet controller: Intel Corporation Ethernet Controller E810-C for QSFP (rev 03)
```

Etapa 2. Verificando o XL710

Se a placa de rede for Intel XL710, você poderá ver Ethernet Controller XL710 na saída.

Etapa 3. Verificar o E810CQDA2

Se a placa de rede for Intel E810CQDA2, você poderá ver Ethernet Controller E810-Cna saída.

Etapa 4. Confirmando o Suporte ao Driver

Para verificar o driver da placa de rede em uso, execute:

```
ethtool -i
```

Exemplo de saída para XL710:

```
driver: i40e
version: 2.13.10
```

Exemplo de saída para E810CQDA2:

```
driver: ice
version: 1.7.12
```

Verifique se a versão do driver corresponde à matriz de compatibilidade para a distribuição do OpenStack e do Linux.

3. Configuração do BIOS/UEFI

- Habilitar SR-IOV:

Verifique se o SR-IOV está habilitado no BIOS/UEFI dos servidores.

- Habilite a tecnologia de virtualização para E/S direcionada (VT-d)/AMD-Vi:

O Intel VT-d ou AMD-Vi deve ser habilitado para a funcionalidade de passagem de PCI e SR-IOV.

4. Configuração do OpenStack

- Principais serviços do OpenStack:

Verifique se os serviços do OpenStack, como Nova, Neutron, Glance e Keystone, estão instalados e configurados.

- Configuração do neutrão:

A Neutron deve suportar Openvswitch (OVS) para redes de orquestração/gerenciamento e SR-IOV para redes de aplicativos/serviços.

- Configuração de SR-IOV:

Os nós de computação devem ser configurados para suportar SR-IOV, com Funções Virtuais (VFs) criadas nas placas de rede.

5. Imagem VNF do Cisco Prime Network Registrar (CPNR)

- Compatibilidade de imagem VNF:

A imagem CPNR VNF deve suportar interfaces SR-IOV e incluir os drivers necessários.

- Carregar para Visão Geral:

Verifique se a imagem CPNR VNF está disponível no OpenStack Glance.

6. Acesso administrativo

- CLI do OpenStack:

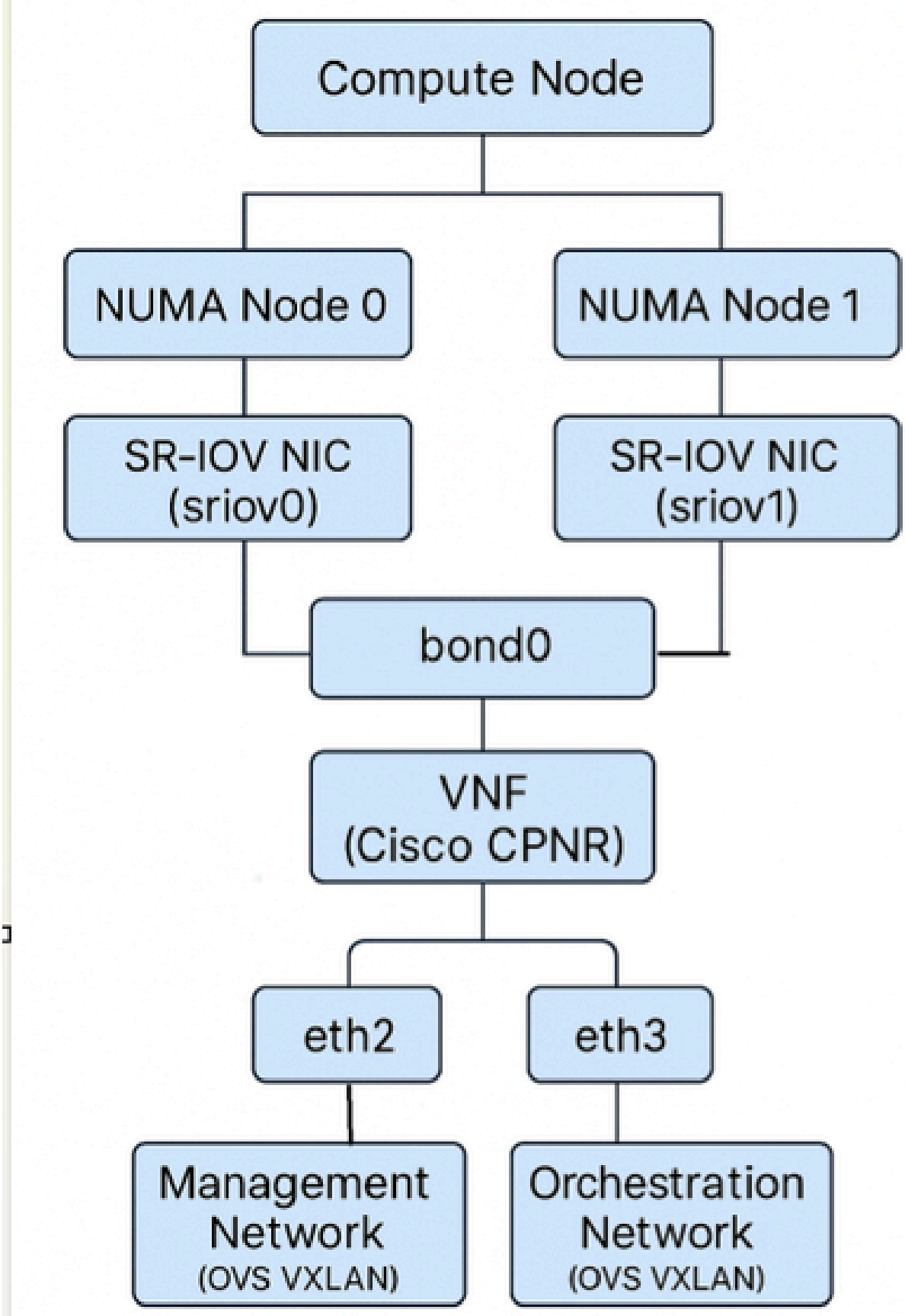
Garanta acesso à CLI do OpenStack para criar redes, variações e iniciar o VNF.

- Privilégios de raiz ou de administrador:

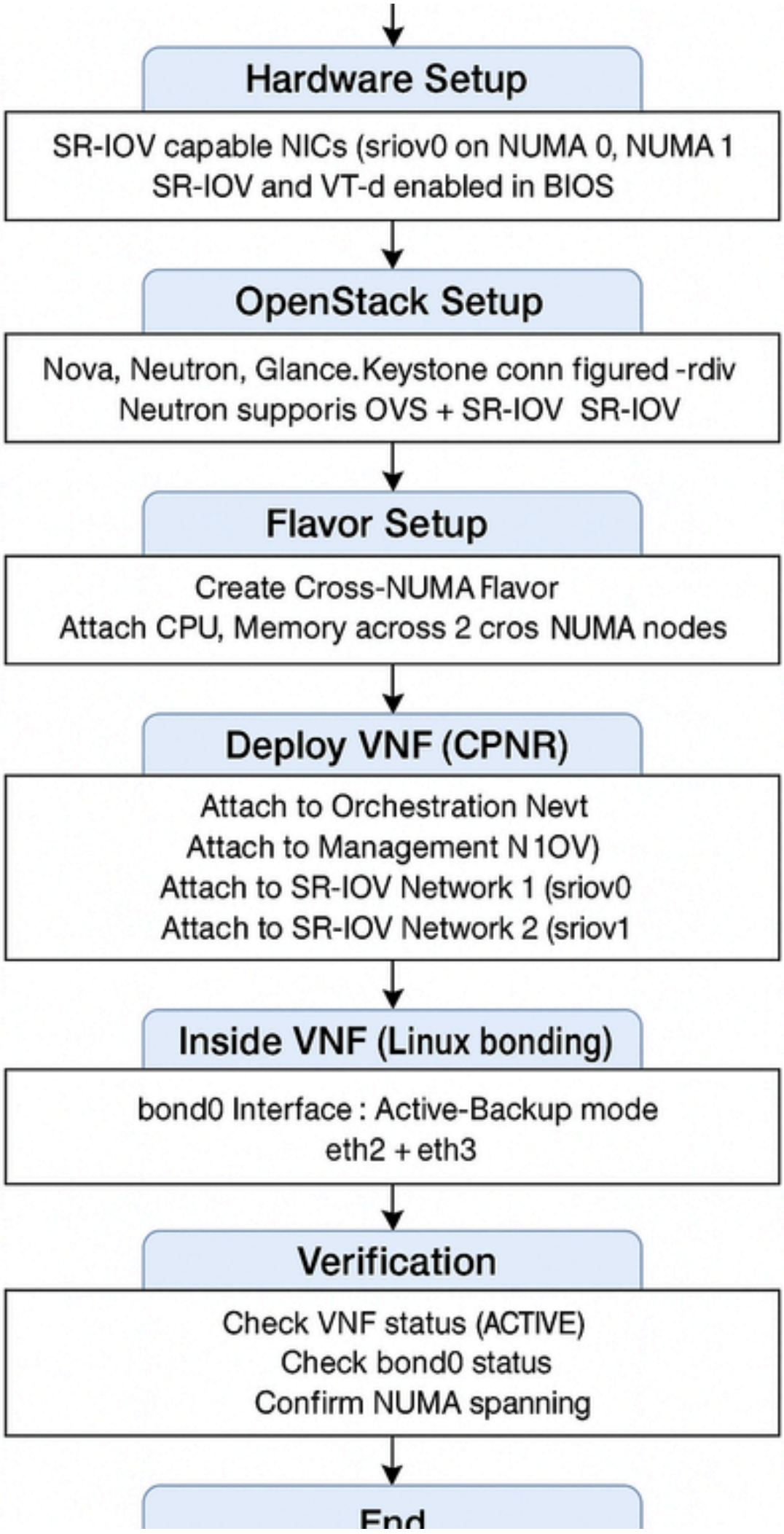
Acesso de raiz ou administrativo para configurar a rede no host Linux e dentro do VNF.

Visão geral da arquitetura

Diagrama de Conectividade da Interface de Rede do VNF


Compute Node
NUMA Node 0
NUMA Node 1
SR-IOV NIC
(sriov0)
SR-IOV NIC
(sriov1)
bond0
VNF
(Cisco CPNR)
eth2
eth3
Management
Network
(OVS VXLAN)
Orchestration
Network
(OVS VXLAN)

Fluxograma


Hardware Setup
SR-IOV capable NICs (sriov0 on NUMA 0, NUMA 1
SR-IOV and VT-d enabled in BIOS
OpenStack Setup
Nova, Neutron, Glance.Keystone conn figured -rdiv
Neutron supporis OVS + SR-IOV SR-IOV
Flavor Setup
Create Cross-NUMA Flavor
Attach CPU, Memory across 2 cros NUMA nodes
Deploy VNF (CPNR)
Attach to Orchestration Nevt
Attach to Management N1OV)
Attach to SR-IOV Network 1 (sriov0
Attach to SR-IOV Network 2 (sriov1
Inside VNF (Linux bonding)
bond0 Interface : Active-Backup mode
eth2 + eth3
Verification
Check VNF status (ACTIVE)
Check bond0 status
Confirm NUMA spanning
End

2. Vinculação de backup ativo:

    - Uma interface bond0 é criada usando placas de rede SR-IOV (eth2 de sriov0 e eth3 de sriov1).
    - O modo de backup ativo garante redundância e tolerância a falhas sem exigir configurações no switch.

3. Rede OpenStack:

    - Redes de orquestração e gerenciamento: Baseado em Openvswitch para tráfego administrativo e de controle.
    - Redes de aplicativos/serviços: Baseado em SR-IOV para tráfego de alto desempenho.

Por que o modo Cross-NUMA é necessário

1. Rede sensível a NUMA no OpenStack

NUMA é uma arquitetura de memória onde cada CPU (e sua memória local e dispositivos) é agrupada em um nó NUMA. No OpenStack, o posicionamento com reconhecimento de NUMA garante que os VNFs sejam atribuídos de forma ideal aos recursos no mesmo nó NUMA para minimizar a latência e maximizar o desempenho.

- As placas de rede SR-IOV são NUMA-Local:
    - Cada NIC física está ligada a um nó NUMA específico. Por exemplo:
        - sriov0 está conectado ao nó NUMA 0.
        - sriov1 está conectado ao nó NUMA 1.
- Limitação de Modo NUMA Único:
    - Quando um VNF é iniciado no modo NUMA único, o OpenStack permite apenas que o VNF se conecte a NICs que são locais ao nó NUMA onde o VNF é iniciado. Isto significa:
        - Se o VNF for iniciado em NUMA 0, ele só poderá se conectar a NICs em sriov0.
        - Se o VNF for iniciado em NUMA 1, ele só poderá se conectar a NICs em sriov1.

2. Por que o modo Cross-NUMA é necessário

O CPNR VNF requer acesso a:

- Rede de orquestração (Openvswitch, não depende de NUMA)
- Rede de gerenciamento(Openvswitch, independente de NUMA)
- Rede 1 SR-IOV: Connected tosriov0(nó NUMA 0)
- Rede SR-IOV 2: Connected tosriov1(nó NUMA 1).

Nesta instalação, o CPNR VNF requer acesso a NICs SR-IOV de NUMA 0 (sriov0) e NUMA 1 (sriov1) para fornecer redundância e alta disponibilidade. Para isso:

- O VNF deve ser iniciado no modo cross-NUMA, que permite que o OpenStack aloque CPU, memória e NICs de vários nós NUMA.
- Isso garante que o VNF possa se conectar a NICs em sriov0 e sriov1, permitindo o uso de ambas as portas SR-IOV em uma configuração de ligação de backup ativo.

Limitação de tamanho de contrarcontrole para portas OVS

O que é o Conntrack?

Conntrack é um recurso do kernel do Linux usado para rastrear conexões de rede, particularmente para NAT (Network Address Translation Conversão de Endereço de Rede) e regras de firewall. Para portas baseadas em OVS no OpenStack, o contrtrack é usado para gerenciar o estado da conexão e aplicar regras de grupos de segurança.

Como o Conntrack afeta as portas OVS

1. Tabela Contrack:

    - Cada conexão ativa consome uma entrada na tabela de controle.
    - O tamanho da tabela conntrack é limitado pelo parâmetro thenf_conntrack_max.

2. Limite padrão:

    - Por padrão, o tamanho da tabela de controle de contexto é de entradas 65536. Para cargas de trabalho com altas taxas de conexão (por exemplo, VNFs com muitos fluxos simultâneos), esse limite pode ser rapidamente esgotado, resultando em pacotes descartados.

3. Impacto nas portas OVS:

    - Se a tabela de controle estiver cheia, novas conexões serão descartadas, o que pode afetar seriamente o desempenho da VNF.
    - Isso é especialmente relevante para as redes deOrquestraçãoeGerenciamento, que usam portas OVS.

Como reduzir as limitações do Contrack

1. Aumentar Tamanho da Tabela de Controle de Conversão:

    - Exibir o limite atual:

```
sysctl net.netfilter.nf_conntrack_max
```

- Aumentar o limite:

```
sysctl -w net.netfilter.nf_conntrack_max=262144
```

- Torne a alteração persistente:

```
echo "net.netfilter.nf_conntrack_max=262144" >> /etc/sysctl.conf
```

2. Monitorar Uso do Controle de Controle:

   Verifique as estatísticas do controle de contexto:

```
cat /proc/sys/net/netfilter/nf_conntrack_count
```

3. Otimizar Regras do Grupo de Segurança:

   Reduza o número de regras aplicadas às portas OVS para minimizar a sobrecarga do controle de contorno.

Como o SR-IOV resolve problemas de contrastação

1. Elimina a Dependência Contracelular

As portas SR-IOV ignoram o caminho de dados OVS e os recursos do kernel do Linux como conntrack. Isso remove completamente a sobrecarga de rastreamento de conexão.

2. Maior escalabilidade

Ao contrário das portas OVS, que são limitadas pelo tamanho da tabela de controle (nf_conntrack_max), as portas SR-IOV podem lidar com um número praticamente ilimitado de conexões.

3. Latência reduzida

Ao descarregar o processamento de pacotes para o hardware da placa de rede, as portas SR-IOV eliminam a latência introduzida pelo processamento de controle baseado em software.

Por que o modo de backup ativo é escolhido para portas SR-IOV

em CPNR VM

O modo de vinculação de backup ativo é particularmente adequado para essa implantação devido à sua simplicidade, tolerância a falhas e compatibilidade com interfaces SR-IOV. Veja por quê:

1. Redundância sem complexidade

- Modo de backup ativo: Apenas uma interface (a interface ativa) transmite e recebe tráfego em um determinado momento. A(s) outra(s) interface(s) permanece(m) em modo de espera.
- Se a interface ativa falhar (por exemplo, devido a uma falha de link ou a um problema de hardware), o vínculo alternará automaticamente para uma interface em standby. Isso garante conectividade de rede contínua sem exigir intervenção manual.

2. Não é necessário um Grupo de Agregação de Enlaces (LAG)

- Ao contrário de outros modos de vinculação (por exemplo, 802.3adorbalance-alb), o modo de backup ativo não requer configurações de LACP (Link Aggregation Control Protocol) ou do lado do switch.
- Isso é especialmente importante para portas SR-IOV, pois os VFs SR-IOV geralmente não suportam configurações LACP ou LAG.

3. Failover contínuo

- O failover é quase instantâneo, com interrupção mínima do tráfego.
- Quando a interface ativa falha, a ligação promove imediatamente uma interface em standby para o status ativo.

4. Independência do hardware

O modo de backup ativo funciona independentemente dos switches físicos subjacentes ou do hardware. A lógica de failover reside inteiramente no kernel do Linux, tornando-o altamente portátil e versátil.

5. Otimizado para SR-IOV

Os VFs de SR-IOV estão vinculados a NICs físicas e nós NUMA específicos. Usando o modo de backup ativo, você pode combinar VFs de diferentes nós NUMA em uma única interface de ligação lógica (bond0). Isso garante alta disponibilidade, ao mesmo tempo em que faz uso eficiente dos recursos NUMA.

O modo de backup ativo é um dos modos mais simples e mais amplamente usados na vinculação com o Linux. Ele foi projetado para fornecer alta disponibilidade, garantindo que o tráfego continue a fluir sem interrupções, mesmo se uma das interfaces vinculadas falhar. Esta é uma explicação detalhada de como o modo Ative-Backup funciona, suas características principais e suas vantagens.

O que é uma interface Linux Bond?

Uma interface de ligação no Linux combina duas ou mais interfaces de rede em uma única interface lógica. Essa interface lógica, conhecida como ligação (por exemplo, ligação0), é usada para fornecer:

- Redundância: Garantindo alta disponibilidade de conectividade de rede.
- Melhoria de desempenho: Em outros modos (por exemplo,balance-error802.3ad), ele também pode agregar largura de banda.

Como funciona o modo de backup ativo?

No modo Ative-Backup, apenas uma interface (chamada de interface ativa) é usada em um determinado momento para transmitir e receber tráfego. A(s) outra(s) interface(s) permanece(m) em modo standby. Se a interface ativa falhar, uma das interfaces em standby será promovida para o status ativo e o tráfego será automaticamente redirecionado para a nova interface ativa.

Principais recursos do modo de backup ativo

1. Interface única ativa:

    - A qualquer momento, apenas uma interface física na ligação está ativa para transmitir e receber tráfego.
    - As interfaces em standby são completamente passivas, a menos que ocorra um failover.

2. Failover automático:

    - Se a interface ativa falhar (por exemplo, devido a um problema de hardware, desconexão de cabo ou falha de link), o vínculo automaticamente muda para uma interface em standby.
    - O failover é perfeito e não requer intervenção manual.

3. Suporte a failback:

    Quando a interface com falha for restaurada, ela poderá se tornar ativa automaticamente novamente (se configurada para fazer isso) ou permanecer no modo de espera, dependendo da configuração de vinculação.

4. Sem requisitos de switch:

    - Ao contrário de outros modos de vinculação (por exemplo, 802.3adorbalance-rr), o modo de backup ativo não requer nenhuma configuração especial nos switches físicos (por exemplo, LAG ou LACP).

- Isso o torna ideal para cenários em que a configuração do lado do switch não é possível ou ao ligar funções virtuais SR-IOV, que não suportam LAG.

5. Monitoramento:

   - A ligação monitora continuamente a integridade de todas as interfaces do membro usando o parâmetro `imon`(Media Independent Interface Monitor).
   - Se uma falha de link for detectada, o vínculo imediatamente alterna para uma interface em standby íntegra.

Como o tráfego flui no modo de backup ativo

Operação normal

1. Interface ativa:

   - O tráfego flui exclusivamente através da interface ativa (por exemplo,eth2em uma ligação ofeth2e th3).
   - A interface em standby (eth3) permanece ociosa e não transmite nem recebe tráfego.

2. Monitoramento:

   - A ligação monitora periodicamente o status de todas as interfaces do membro. Isso é feito com:
     - mímon: Verifica o status do link de cada interface em um intervalo configurável (por exemplo, a cada 100 ms).
     - Monitoramento do Address Resolution Protocol (ARP) (opcional): Envia solicitações ARP para garantir que a interface ativa esteja acessível.

Cenário de failover

1. Falha de link na interface ativa:

   Se a interface ativa (eth2) falhar (por exemplo, cabo desconectado, falha de hardware da placa de rede ou link inativo), a conexão detectará imediatamente a falha usando monitor miniimonor ARP.

2. Failover automático:

   - Os switches de ligação à interface em standby (eth3), que se torna a nova interface ativa.
   - O tráfego é roteado novamente através da nova interface ativa sem exigir intervenção manual.

3. Tempestividade de failover:

O processo de failover é quase instantâneo (normalmente em alguns milissegundos, dependendo do intervalo de meia hora).

Cenário de failback

1. Restauração da interface com falha:
    - Quando a interface que falhou anteriormente (eth2) é restaurada, ela pode:
        - Recuperar automaticamente a função ativa (se configurado para fazer isso).
        - Permanecer no modo de espera (comportamento padrão).
2. Continuidade de tráfego:

    O failback é perfeito, garantindo que não haja interrupção nos fluxos de tráfego em andamento.

Caso de uso: Vinculação de backup ativo com portas SR-IOV

O modo de backup ativo é particularmente adequado para interfaces SR-IOV porque:

- Os VFs SR-IOV geralmente não suportam protocolos de agregação de links como o LACP.
- O vínculo no modo de backup ativo pode fornecer redundância sem qualquer configuração no lado do switch.

Por exemplo:

- eth2é mapeado para um onsriov0 (nó 0 NUMA) de SR-IOV VF.
- eth3é mapeado para um onsriov1 SR-IOV VF(nó NUMA 1).
- A ligação (bond0) combina essas interfaces, fornecendo failover perfeito entre VFs de SR-IOV.

Etapa 1. Rede OpenStack

A CPNR VNF requer estas quatro redes:

1. Rede de orquestração: Para tráfego de controle e orquestração (baseado em Openvswitch).
2. Rede de gerenciamento: Para acesso administrativo (baseado em Openvswitch).
3. Rede 1 SR-IOV: Tráfego de aplicativo/serviço em sriov0.
4. Rede SR-IOV 2: Tráfego de aplicativo/serviço em sriov1.

Implantação passo a passo:

Etapa 1.1. Criar Redes Openvswitch

- Rede de orquestração:

```
openstack network create --provider-network-type vxlan orchestration-network
```

- Rede de gerenciamento:

```
openstack network create --provider-network-type vxlan management-network
```

Etapa 1.2. Criar sub-redes para redes Openvswitch

- Sub-rede de orquestração:

```
openstack subnet create --network orchestration-network \
--subnet-range 192.168.100.0/24 orchestration-subnet
```

- Sub-rede de gerenciamento:

```
openstack subnet create --network management-network \
--subnet-range 10.10.10.0/24 management-subnet
```

Etapa 1.3. Criar redes SR-IOV

- Rede 1 SR-IOV:

```
openstack network create --provider-network-type vlan \
--provider-physical-network sriov0 --provider-segment 101 sriov-network-1
```

- Rede SR-IOV 2:

```
openstack network create --provider-network-type vlan \
--provider-physical-network sriov1 --provider-segment 102 sriov-network-2
```

Etapa 2. Opções do OpenStack

Etapa 2.1. Criar um Tipo Cross-NUMA

Para garantir que o VNF possa acessar as placas de rede SR-IOV de ambos os nós NUMA, crie um tipo com suporte a cross-NUMA:

```
openstack flavor create --ram 8192 --vcpus 4 --disk 40 cross-numa-flavor
```

Etapa 2.2. Configurar Propriedades NUMA

Definir propriedades específicas de NUMA:

```
openstack flavor set cross-numa-flavor \
--property hw:numa_nodes=2 \
--property hw:cpu_policy=dedicated \
--property hw:mem_page_size=large
```

Etapa 3. Configurar a vinculação no modo de backup ativo

Depois de iniciar o VNF, configure a interface de ligação para portas SR-IOV (eth2 e eth3) no VNF.

Etapa 3.1. Configuração da interface de ligação

Crie uma interface de vínculo (bond0) no modo de Backup Ativo:

```
vi /etc/sysconfig/network-scripts/ifcfg-bond0
```

```
DEVICE=bond0
BOOTPROTO=static
ONBOOT=yes
BONDING_OPTS="mode=active-backup miimon=100"
IPADDR=172.16.1.10
NETMASK=255.255.255.0
GATEWAY=172.16.1.1
```

Etapa 3.2. Configurar interfaces slave

- eth2:

```
vi /etc/sysconfig/network-scripts/ifcfg-eth2
```

```
DEVICE=eth2
ONBOOT=yes
MASTER=bond0
SLAVE=yes
```

- eth3:

```
vi /etc/sysconfig/network-scripts/ifcfg-eth3
```

```
DEVICE=eth3
ONBOOT=yes
MASTER=bond0
SLAVE=yes
```

Etapa 3.3. Aplicar configuração

Reinicie o serviço de rede para aplicar a configuração:

```
systemctl restart network
```

Verificar

Após implementar o VNF, verifique sua funcionalidade usando estas etapas:

1. Verificar o status do VNF

Verifique se a instância do VNF está ativa:

```
openstack server show cpnr-instance
```

Verifique se o status é ATIVE.

2. Verificar a Conectividade da Rede

- Teste de ping: Verifique se o VNF pode se comunicar por todas as redes:

  ```
  ping
  ```

  ```
  ping
  ```

- Interface de ligação:
  - Confirme se bond0 está ativo:

    ```
    cat /proc/net/bonding/bond0
    ```

    Procure:
    - Escravo Ativo Atualmente: Indica a interface ativa.
    - Interface slave: Confirma que ambos2 e 3 fazem parte do vínculo.

3. Verificar Posicionamento NUMA

Verifique se o VNF está usando recursos de ambos os nós NUMA:

```
nova show

        --human | grep numa
```

Melhores práticas

- Monitoramento e solução de problemas: Use ferramentas como cpdumpandethtoolpara monitorar as interfaces SR-IOV.
- Segurança: Gerencie cuidadosamente o acesso à rede física e aplique o isolamento estrito entre os usuários.
- Dimensionamento: Planeje a capacidade física da placa de rede ao escalar implantações de SR-IOV, pois o número de VFs disponíveis é limitado pelo hardware da placa de rede.

Troubleshooting

Se a implantação não funcionar como esperado, consulte estas etapas de Troubleshooting:

1. Verificar a Configuração do SR-IOV

- Verifique se o SR-IOV está habilitado no BIOS:

```
dmesg | grep -i "SR-IOV"
```

- Confirme se os VFs são criados em NICs:

```
lspci | grep Ethernet
```

2. Verificar Posicionamento NUMA

Se o VNF não puder acessar ambas as NICs, verifique se o modo cross-NUMA está habilitado:

- Verifique as propriedades NUMA do tipo:

```
openstack flavor show cross-numa-flavor
```

3. Questões relativas à interface das ligações

- Verifique o status do vínculo:

```
cat /proc/net/bonding/bond0
```

- Se a obrigação não estiver funcionando:
  - Certifique-se de que as interfaces slave (eth2e th3) estejam configuradas corretamente como parte da ligação.
  - Reinicie o serviço de rede:

```
systemctl restart network
```

4. Problemas de conectividade de rede

- Verifique as vinculações de porta do OpenStack:

```
openstack port list --server cpnr-instance
```

- Verifique a configuração IP correta dentro do VNF:

```
ip addr show
```

Conclusão

A implantação de CPNR VNF no OpenStack com portas SR-IOV requer o modo cross-NUMA para permitir que o VNF se conecte a NICs de ambos os nós NUMA. Isso é essencial porque o OpenStack restringe VNFs no modo de NUMA único para acessar apenas recursos (NICs, CPUs, memória) dentro do nó NUMA onde o VNF é iniciado. A combinação do modo cross-NUMA com a vinculação de backup ativo garante alta disponibilidade, tolerância a falhas e utilização eficiente de recursos, tornando essa implantação altamente resiliente e eficiente.

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.