

NSX 故障排除指南

Update 2

修改日期：2016 年 8 月 18 日

VMware NSX Data Center for vSphere 6.2



vmware®

您可以从 VMware 网站下载最新的技术文档：

<https://docs.vmware.com/cn/>。

VMware 网站还提供了最近的产品更新。

如果您对本文档有任何意见或建议，请将反馈信息发送至：

docfeedback@vmware.com

VMware, Inc.
3401 Hillview Ave.
Palo Alto, CA 94304
www.vmware.com

北京办公室
北京市
朝阳区新源南路 8 号
启皓北京东塔 8 层 801
www.vmware.com/cn

上海办公室
上海市
淮海中路 333 号
瑞安大厦 804-809 室
www.vmware.com/cn

广州办公室
广州市
天河路 385 号
太古汇一座 3502 室
www.vmware.com/cn

目录

- 1 NSX 故障排除指南 4**
- 2 基础架构准备 5**
 - NSX 基础架构准备步骤 7
 - 检查通信通道运行状况 19
 - 解决 NSX Manager 问题 20
 - 从 NSX Controller 故障恢复 22
 - 使用 NSX 仪表板 23
 - 使用 `show host health-status` 命令 25
 - 设置 NSX 组件的日志记录级别 26
 - vSphere ESX Agent Manager 27
 - NSX CLI 速查表 29
- 3 跟踪流 38**
 - 关于跟踪流 38
 - 使用跟踪流进行故障排除 39
- 4 NSX 路由 47**
 - 了解分布式逻辑路由器 48
 - 了解 Edge 服务网关提供的路由 51
 - ECMP 数据包流 52
 - NSX 路由：必备条件和注意事项 54
 - DLR 和 ESG UI 56
 - 新的 NSX Edge (DLR) 58
 - 典型的 ESG 和 DLR UI 操作 62
 - NSX 路由故障排除 65
- 5 Edge 设备故障排除 94**
- 6 分布式防火墙 107**
 - 如何使用 `show dfw CLI` 107
 - Distributed Firewall 故障排除 109
- 7 负载均衡 116**
 - 场景：配置单臂负载均衡器 117
 - 使用 UI 的负载均衡器故障排除 122
 - 使用 CLI 的负载均衡器故障排除 123
 - 常见的负载均衡器问题 126

NSX 故障排除指南

《NSX 故障排除指南》介绍了如何使用 NSX Manager 用户界面、vSphere Web Client 和其他 NSX 组件（如果需要）监控 VMware® NSX™ 系统和进行故障排除。

目标读者

本手册专供要在 VMware vCenter 环境中安装或使用 NSX 的用户使用。本手册的目标读者为熟悉虚拟机技术和虚拟数据中心操作且经验丰富的系统管理员。本手册假设您熟悉 VMware Infrastructure 5.x，包括 VMware ESX、vCenter Server 和 vSphere Web Client。

VMware 技术出版物术语表

VMware 技术出版物提供了一个术语表，其中包含一些您可能不熟悉的术语。有关 VMware 技术文档中使用的术语的定义，请访问 <http://www.vmware.com/support/pubs>。

基础架构准备

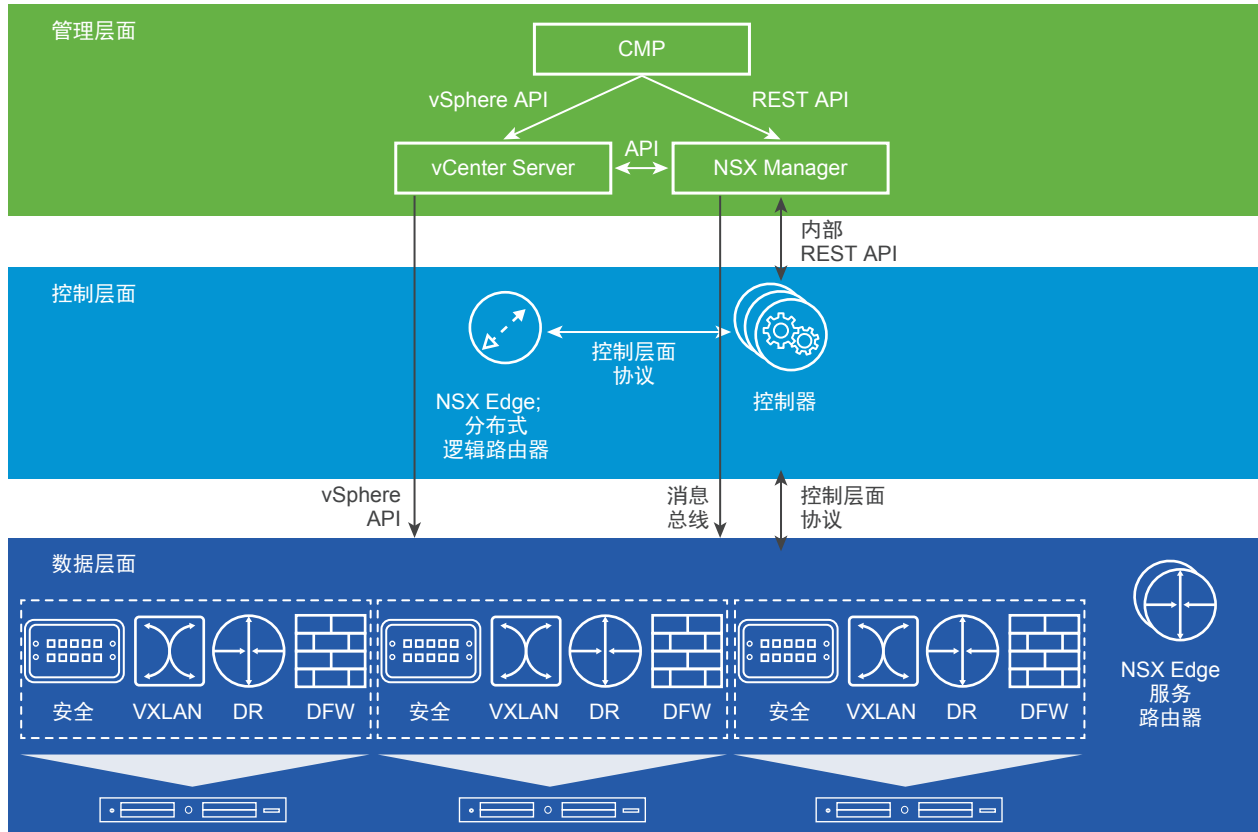
了解在准备 NSX 时使用的组件对于确定和解决常见问题非常重要。

NSX 基础架构准备组件

- vSphere ESX Agent Manager (EAM)
- NSX Manager
- 控制器群集（如果使用单播、混合模式或分布式逻辑路由）
- VTEP（ESXi 管理程序）
- 用户环境代理 (UWA)
- vSphere Distributed Switch

NSX Manager 和 ESXi 管理程序主机之间的控制层面通信是由基于 RabbitMQ 的消息服务提供的。控制器群集和 ESXi 管理程序主机之间的控制层面通信取决于在主机上作为客户端运行的 netcpa 用户环境代理。

图 2-1. 组件及其通信的简要视图



成功准备基础架构的提示

VMware 产品互操作性列表网站提供了有关 NSX 兼容性和版本要求的信息。请参见 http://partnerweb.vmware.com/comp_guide2/sim/interop_matrix.php。

确保使用的是 vSphere Distributed Switch 5.5 或更高版本。

将单独的 DVS 用于管理、服务和网关。

请注意在使用刀片时采用的网卡绑定方法。机箱交换机支持最小集合。

在安装时，请确保在执行主机准备之前部署了控制器群集并全部处于绿色状态。

在升级时，请确保在升级之前连接了控制器并全部处于绿色状态。请参见《NSX 升级指南》。

本章讨论了以下主题：

- [NSX 基础架构准备步骤](#)
- [检查通信通道运行状况](#)
- [解决 NSX Manager 问题](#)
- [从 NSX Controller 故障恢复](#)
- [使用 NSX 仪表板](#)
- [使用 show host health-status 命令](#)

- [设置 NSX 组件的日志记录级别](#)
- [vSphere ESX Agent Manager](#)
- [NSX CLI 速查表](#)

NSX 基础架构准备步骤

NSX 准备是一个包含 4 个步骤的过程。

- 1 将 NSX Manager 连接到 vCenter Server。NSX Manager 和 vCenter Server 具有一对一关系。
 - a 在 vCenter Server 中注册
- 2 部署 NSX Controller（仅逻辑交换、分布式路由或 Edge 服务需要。如果仅使用 Distributed Firewall (DFW)，则不需要使用控制器）。
- 3 主机准备：在群集中的所有主机上为 XLAN、DFW 和 DLR 安装 VIB。配置基于 Rabbit MQ 的消息传送基础架构。启用防火墙。通知控制器已为 NSX 准备好主机。
- 4 配置 IP 池设置并配置 VXLAN：在群集中的所有主机上创建 VTEP 端口组和 VMKNIC。在该步骤期间，您可以设置传输 VLAN ID、成组策略和 MTU。

将 NSX Manager 连接到 vCenter Server

通过使用 NSX Manager 和 vCenter Server 之间的连接，NSX Manager 可以使用 vSphere API 执行一些功能，例如，部署服务虚拟机，准备主机以及创建逻辑交换机端口组。连接过程在 Web Client 服务器上为 NSX 安装 Web Client 插件。

要使连接正常工作，您必须在 NSX Manager、vCenter Server 和 ESXi 主机上配置 DNS 和 NTP。如果按名称将 ESXi 主机添加到 vSphere 清单中，请确保已在 NSX Manager 上配置 DNS 服务器并且名称解析正常工作。否则，NSX Manager 将无法解析 IP 地址。必须指定 NTP 服务器，以使 SSO 服务器时间和 NSX Manager 时间保持同步。在 NSX Manager 上，`/etc/ntp.drift` 中的偏移文件包含在 NSX Manager 的技术支持包中。

此外，用于将 NSX Manager 连接到 vCenter Server 的帐户必须具有 vCenter “管理员”角色。具有“管理员”角色还允许 NSX Manager 在 Security Token Service 服务器中注册其自身。在使用特定用户帐户将 NSX Manager 连接到 vCenter 时，还会在 NSX Manager 上为该用户创建一个“企业管理员”角色。

与将 NSX Manager 连接到 vCenter Server 有关的常见问题

- 未在 NSX Manager、vCenter Server 或 ESXi 主机上正确配置 DNS。
- 未在 NSX Manager、vCenter Server 或 ESXi 主机上正确配置 NTP。
- 使用没有 vCenter “管理员”角色的用户帐户将 NSX Manager 连接到 vCenter。
- 在 NSX Manager 和 vCenter Server 之间出现网络连接问题。
- 用户使用在 NSX Manager 上没有角色的帐户登录到 vCenter。

您需要先通过用于将 NSX Manager 链接到 vCenter Server 的帐户登录到 vCenter。然后，您可以使用 **vCenter 主页 > 网络和安全 > NSX Manager > {NSX Manager IP} > 管理 > 用户 (vCenter Home > Networking & Security > NSX Managers > {IP of NSX Manager} > Manage > Users)** API 在 NSX Manager 上创建其他用户角色。

首次登录可能需要最多 4 分钟的时间，在此期间，vCenter 将加载并部署 NSX UI 包。

验证从 NSX Manager 到 vCenter Server 的连接

要验证连接，请从 NSX 虚拟设备中执行 Ping 操作，然后查看 ARP 和路由表。

```
nsxmgr# show arp
```

IP address	HW type	Flags	HW address	Mask	Device
192.168.110.31	0x1	0x2	00:50:56:ae:ab:01	*	mgmt
192.168.110.2	0x1	0x2	00:50:56:01:20:a5	*	mgmt
192.168.110.1	0x1	0x2	00:50:56:01:20:a5	*	mgmt
192.168.110.33	0x1	0x2	00:50:56:ae:4f:7c	*	mgmt
192.168.110.32	0x1	0x2	00:50:56:ae:50:bf	*	mgmt
192.168.110.10	0x1	0x2	00:50:56:03:19:4e	*	mgmt
192.168.110.51	0x1	0x2	00:50:56:03:30:2a	*	mgmt
192.168.110.22	0x1	0x2	00:50:56:01:21:f9	*	mgmt
192.168.110.55	0x1	0x2	00:50:56:01:23:21	*	mgmt
192.168.110.26	0x1	0x2	00:50:56:01:21:ef	*	mgmt
192.168.110.54	0x1	0x2	00:50:56:01:22:ef	*	mgmt
192.168.110.52	0x1	0x2	00:50:56:03:30:16	*	mgmt

```
nsxmgr# show ip route
Codes: K - kernel route, C - connected, S - static,
       > - selected route, * - FIB route

S>* 0.0.0.0/0 [1/0] via 192.168.110.1, mgmt
C>* 192.168.110.0/24 is directly connected, mgmt
```

在 NSX Manager 日志中查找错误，以找出未连接到 vCenter Server 的原因。用于查看日志的命令是 `show log manager follow`。

```
2014-02-26 12:53:23.815 GMT INFO VcEventsReaderThread DefaultRequestDirector:491 - I/O exception (org.apache.http.NoHttpResponseException: The target server failed to respond)
2014-02-26 12:53:23.815 GMT INFO VcEventsReaderThread DefaultRequestDirector:498 - Retrying request
2014-02-26 12:53:23.815 GMT WARN ViInventoryThread ViInventory:1482 - We received error from VC, probably lost connection
2014-02-26 12:53:23.817 GMT INFO VcEventsReaderThread VcEventsReader$VcEventsReaderThread:347 - Caught exception:com.vmware.vim.client.exception.ConnectionException: org.apache.http.conn.HttpHostConnectException: Connection to https://vc-1-01a.corp.local refused
2014-02-26 12:53:23.821 GMT DEBUG VcEventsReaderThread VcEventsReader$VcEventsReaderThread:348 - Caught exception during ping:com.vmware.vim.vimomi.client.exception.ConnectionException: org.apache.http.conn.HttpHostConnectException: Connection to https://vc-1-01a.corp.local refused
```

登录到 NSX Manager CLI 控制台，然后运行 `debug connection IP_of_ESXi_or_VC` 命令并检查输出。

在 NSX Manager 上执行数据包捕获以查看连接

使用 `debug packet` 命令：`debug packet [capture|display] interface interface filter`

NSX Manager 上的接口名称是 `mgmt`。

筛选器语法采用以下形式：“`port_80_or_port_443`”

该命令仅在特权模式下运行。要进入特权模式，请运行 `enable` 命令并提供管理员密码。

数据包捕获示例：

```
nsxmgr# en
nsxmgr# debug packet display interface mgmt port_80_or_port_443
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on mgmt, link-type EN10MB (Ethernet), capture size 262144 bytes
23:40:25.321085 IP 192.168.210.15.54688 > 192.168.210.22.443: Flags [P.], seq 2645022162:2645022199, ack
2668322748, win 244, options [nop,nop,TS val 1447550948 ecr 365097421], length 37
...
```

在 NSX Manager 上验证网络配置

show running-config 命令显示管理接口、NTP 和默认路由设置的基本配置。

```
nsxmgr# show running-config
Building configuration...

Current configuration:
!
ntp server 192.168.110.1
!
ip name server 192.168.110.10
!
hostname nsxmgr
!
interface mgmt
 ip address 192.168.110.15/24
!
ip route 0.0.0.0/0 192.168.110.1
!
web-manager
```

NSX Manager 证书

NSX Manager 支持使用两种方法生成证书。

- NSX Manager 生成的 CSR：由于基本 CSR 而受到限制的功能
- KCS#12：建议将其用于生产环境

存在一个已知问题：在 CMS 无法执行 API 调用时，不显示任何提示。

在调用方不知道证书颁发者时，将会发生这种情况，因为这是不可信的根证书颁发机构，或者证书是自签名证书。要解决该问题，请使用浏览器导航到 NSX Manager IP 地址或主机名并接受证书。

部署 NSX Controller

NSX Controller 是 NSX Manager 使用 OVA 格式部署的。具有控制器群集可以提供高可用性。

部署控制器要求 NSX Manager、vCenter Server 和 ESXi 主机配置了 DNS 和 NTP。

必须使用静态 IP 池为每个控制器分配 IP 地址。

建议您实施 DRS 反关联性规则以使 NSX Controller 位于单独的主机上。

您必须部署三个 NSX Controller。

常见的控制器问题

在部署 NSX Controller 期间，可能遇到的典型问题如下所示：

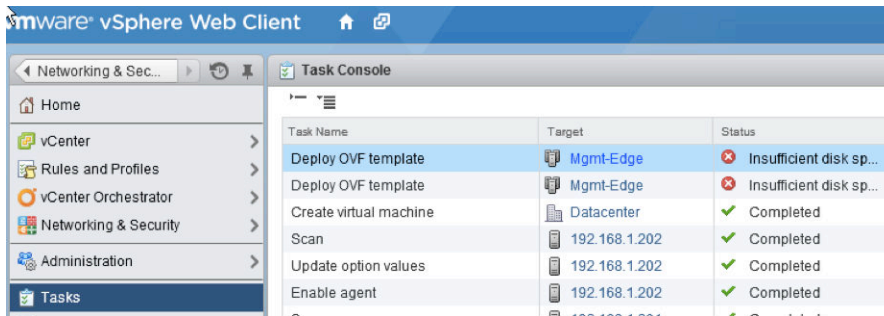
- **NSX Controller 运行速度缓慢。**这可能是资源不足造成的。要检测 NSX Controller 系统要求问题，请运行 `request system compatibility-report` 命令。

```
nsx-controller # request system compatibility-report
Testing: Number of CPUs. Done.
Testing: Aggregate CPU speed. Done.
Testing: Memory. Done.
Testing: Management NIC speed. Done.
Testing: NTP configured. Done.
Testing: /var disk partition size. Done.
Testing: /var disk speed. Done.
Testing: pserver-log disk size. Done.
Testing: pserver-log disk speed. Done.
Testing: pserver-data disk size. Done.
Testing: pserver-data disk speed. Done.
Testing: logging disk size. Done.
Testing: logging disk speed. Done.
```

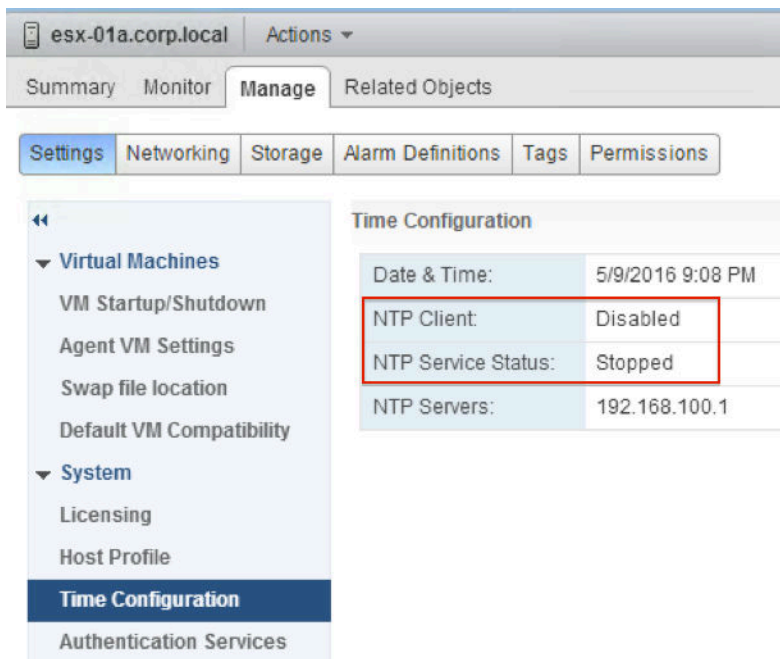
	Detected	Supported	Required
Number of CPUs	2	NO	>=8
Aggregate CPU speed	5.6 GHz	NO	>=13
Memory	1.835 GB	NO	>=63
Management NIC speed	10000 Mb/s	YES	>=1000
NTP configured	No	NO	Yes
/var disk partition size	- GB	NO	>=128
/var disk speed	- MB/s	NO	>=40
pserver-log disk size	- GB	NO	>=128
pserver-log disk speed	- MB/s	NO	>=40
pserver-data disk size	- GB	NO	>=128
pserver-data disk speed	- MB/s	NO	>=40
logging disk size	- GB	NO	>=128
logging disk speed	- MB/s	NO	>=40

- **NSX Manager 和 NSX Controller 之间出现 IP 连接问题。**这通常是物理网络连接问题或防火墙阻止通信造成的。

- vSphere 上没有足够的资源（如存储）以托管控制器。在控制器部署期间，可以查看 vCenter 事件和任务日志以发现这些问题。



- 出现异常的“恶意”控制器或升级的控制器处于断开连接状态。
- 未正确配置 ESXi 主机和 NSX Manager 上的 DNS。
- ESXi 主机和 NSX Manager 上的 NTP 不同步。



- 在新连接的虚拟机无法访问网络时，这可能是控制层面问题造成的。检查控制器状态。

NSX Manager			
NSX Manager	IP Address	vCenter	Version
	192.168.110.42	vc-01a.corp.local	6.0.2.2944561

NSX Controller nodes							
Name	Node	NSX Manager	Cluster/Resource...	Dataverse	Host	Software Version	Status
controller-1	192.168.110.201	192.168.1...	Resources	ds-site-a-nfs...	esx-01a.corp...	6.0	Disconnect
controller-2	192.168.110.202	192.168.1...	Resources	ds-site-a-nfs...	esx-02a.corp...	6.0	Disconnect
controller-3	192.168.110.203	192.168.1...	Resources	ds-site-a-nfs...	esx-01a.corp...	6.0	Disconnect

还要尝试在 ESXi 主机上运行 `esxcli network vswitch dvs vmware vxlan network list --vds-name <name>` 命令以检查控制层面状态。请注意，控制器连接中断。

```
/etc/vmware/netcpa # esxcli network vswitch dvs vmware vxlan network list --vds-name Compute_VDS
VXLAN ID Multicast IP Control Plane Controller Connection
ARP Entry Count MTEP Count
-----
5000 N/A (headend replication) Enabled (multicast proxy, ARP proxy) 192.168.110.203 (down)
0 0
```

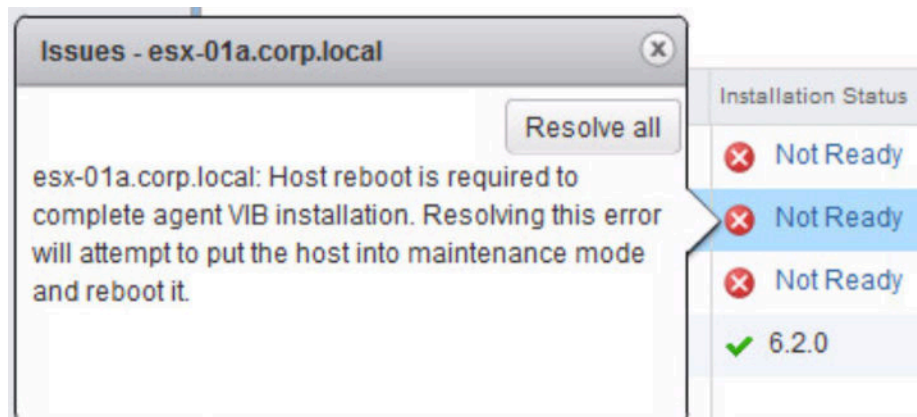
- 运行 `show log manager follow NSX Manager CLI` 命令可以找出无法部署控制器的任何其他原因。

```
2014-02-26 10:09:44.931 GMT INFO taskScheduler-25 VCConnection$VimClient:1219 - Create stub for com.vmware.vim.binding
28c5157-abf3-718e-88c5-42209f389211
2014-02-26 10:09:44.932 GMT DEBUG VcEventsReaderThread VcEventsReader$VcEventsReaderThread:301 - got prop collector up
ctReference: type = PropertyFilter, value = session[d46b86a2-7a10-c17e-6ebe-8ab252ee4efd]527420f2-bdd7-529b-8ab6-17d16
6E3-4A64-96D7-5833C287588F
2014-02-26 10:09:44.937 GMT ERROR taskScheduler-25 VCUtils:184 - Error while waiting for property collector updates.
com.vmware.vim.binding.vim.fault.NoDiskSpace:
datastore = datastore1 (1)
inherited from com.vmware.vim.binding.vim.fault.FileFault:
file = [datastore1 (1)] NSX_Controller_1c3dd18d-0cd3-4d7d-896b-51247176ae77/NSX_Controller_1c3dd18d-0cd3-4d7d-896b-512
inherited from com.vmware.vim.binding.vim.fault.VimFault:
inherited from com.vmware.vim.binding.vim.fault.NoDiskSpace: Insufficient disk space on datastore 'datastore1 (1)'.
```

主机准备

vSphere ESX Agent Manager 将 VIB 部署到 ESXi 主机上。

主机上的部署要求在主机、vCenter Server 和 NSX Manager 上配置 DNS。部署不需要重新引导 ESXi 主机，但任何 VIB 更新或移除需要重新引导 ESXi 主机。



VIB 是在 NSX Manager 上托管的，也可以作为 zip 文件提供。

可以从 <https://<NSX-Manager-IP>/bin/vdn/nwfabric.properties> 中访问该文件。可下载的 zip 文件因 NSX 和 ESXi 版本而异。例如，vSphere 6.0 主机使用 <https://<NSX-Manager-IP>/bin/vdn/vibs-6.2.3/6.0-3771165/vxlan.zip> 文件。

```
C:\Users\Administrator>curl -k https://nsxmgr-01a.corp.local/bin/vdn/nwfabric.properties
# 5.1 VDN EAM Info
VDN_VIB_PATH.1=/bin/vdn/vibs-6.2.3/5.1-2107743/vxlan.zip
VDN_VIB_VERSION.1=2107743
VDN_HOST_PRODUCT_LINE.1=embeddedEsx
VDN_HOST_VERSION.1=5.1.*

# 5.5 VDN EAM Info
VDN_VIB_PATH.2=/bin/vdn/vibs-6.2.3/5.5-3771174/vxlan.zip
VDN_VIB_VERSION.2=3771174
VDN_HOST_PRODUCT_LINE.2=embeddedEsx
VDN_HOST_VERSION.2=5.5.*

# 6.0 VDN EAM Info
VDN_VIB_PATH.3=/bin/vdn/vibs-6.2.3/6.0-3771165/vxlan.zip
VDN_VIB_VERSION.3=3771165
VDN_HOST_PRODUCT_LINE.3=embeddedEsx
VDN_HOST_VERSION.3=6.0.*

# 6.1 VDN EAM Info
VDN_VIB_PATH.4=/bin/vdn/vibs-6.2.3/6.1-3689890/vxlan.zip
VDN_VIB_VERSION.4=3689890
VDN_HOST_PRODUCT_LINE.4=embeddedEsx
VDN_HOST_VERSION.4=6.1.*

# Single Version associated with all the VIBs pointed by above VDN_VIB_PATH(s)
VDN_VIB_VERSION=6.2.3.3771501

# Legacy vib location. Used by code to discover available legacy vibs.
LEGACY_VDN_VIB_PATH_FS=/common/em/components/vdn/vibs/legacy/
LEGACY_VDN_VIB_PATH_WEB_ROOT=/bin/vdn/vibs/legacy/
```

VIB 名称是：

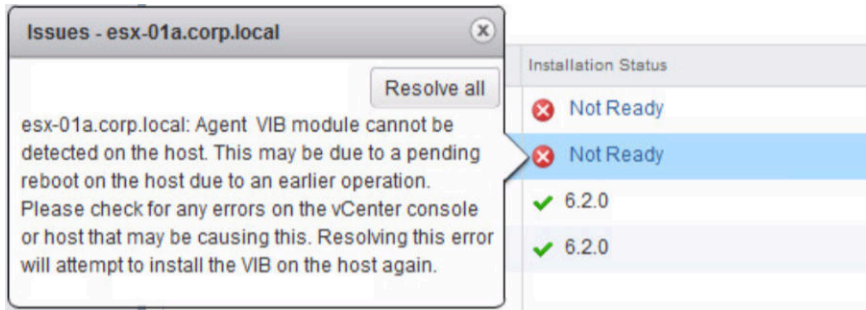
- esx-vsip
- esx-vxlan

```
[root@esx-01a:~] esxcli software vib list | grep -e vsip -e vxlan
esx-vsip                6.0.0-0.0.3771165          VMware  VMwareCertified  2016-04-20
esx-vxlan                6.0.0-0.0.3771165          VMware  VMwareCertified  2016-04-20
```

主机准备期间的常见问题

在主机准备期间，可能遇到的典型问题如下所示：

- EAM 无法部署 VIB。
 - 可能是由于未在主机上正确配置 DNS。



- 可能是由于防火墙阻止 ESXi、NSX Manager 和 vCenter Server 之间的所需端口。
- 已安装以前的旧 VIB 版本。这需要用户干预以重新引导主机。
- NSX Manager 和 vCenter Server 遇到通信问题：
 - “网络和安全”插件中的**主机准备 (Host Preparation)**选项卡未正确显示所有主机。
 - 检查 vCenter Server 是否可以枚举所有主机和群集。

主机准备 (VIB) 故障排除

- 检查主机的通信通道运行状况。请参见[检查通信通道运行状况](#)。
- 检查 vSphere ESX Agent Manager 以查找错误。

vCenter 主页 > 管理 > vCenter Server 扩展 > vSphere ESX Agent Manager (vCenter home > Administration > vCenter Server Extensions > vSphere ESX Agent Manager)

在 vSphere ESX Agent Manager 上，检查带有“VCNS160”前缀的代理机构的状态。如果某个代理机构处于错误的状态，请选择该代理机构并查看其问题。

Agency	State	Status	Optimized Deployment
_VCNS_160_Management & Edge Cl...	Enabled	✓ Normal	✓
_VCNS_160_Compute Cluster A_VMwa...	Enabled	✗ Alert	✓

Issues for the selected agencies				
Trigger Time	Agency	Issue	Host	Agent VM
Thu Apr 28 12:03:12 GMT-0...	_VCNS_160_Compute Clu...	Agent VIB module is not installed	esx-01a.corp.local	

- 在出现问题的主机上，运行 `tail /var/log/esxupdate.log` 命令。

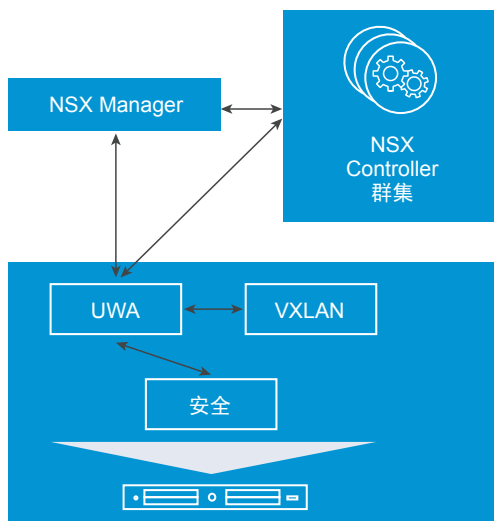
```
2016-04-28T19:02:52Z esxupdate: downloader: DEBUG: Downloading https://vcsa-  
o/tmp/tmpKT0wjN...  
2016-04-28T19:03:12Z esxupdate: esxupdate: ERROR: An esxupdate error excepti  
2016-04-28T19:03:12Z esxupdate: esxupdate: ERROR: Traceback (most recent call  
2016-04-28T19:03:12Z esxupdate: esxupdate: ERROR: File "/usr/sbin/esxupdate  
2016-04-28T19:03:12Z esxupdate: esxupdate: ERROR: cmd.Run()  
2016-04-28T19:03:12Z esxupdate: esxupdate: ERROR: File "/build/mts/release/  
site-packages/vmware/esx5update/CommandLine.py", line 106, in Run  
2016-04-28T19:03:12Z esxupdate: esxupdate: ERROR: File "/build/mts/release/  
site-packages/vmware/esximage/Transaction.py", line 73, in DownloadMetadata  
2016-04-28T19:03:12Z esxupdate: esxupdate: ERROR: MetadataDownloadError: ('h  
fd3f37ad4c', None, "('https://vcsa-01a.corp.local:443/eam/vib?id=facdb160-21  
rlopen error [Errno -3] Temporary failure in name resolution>')")  
2016-04-28T19:03:12Z esxupdate: esxupdate: DEBUG: <<<
```

- 请参见 <https://kb.vmware.com/kb/2053782>。

主机准备 (UWA) 故障排除

NSX Manager 在群集中的所有主机上配置两个用户环境代理:

- 消息总线 UWA (vsfwd)
- 控制层面 UWA (netcpa)

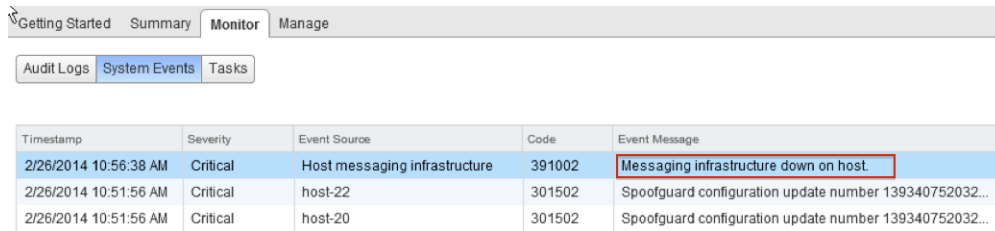


在极少数情况下，VIB 安装成功，但由于某种原因，一个或两个用户环境代理无法正常工作。这可能表现为：

- 防火墙显示错误的状态。

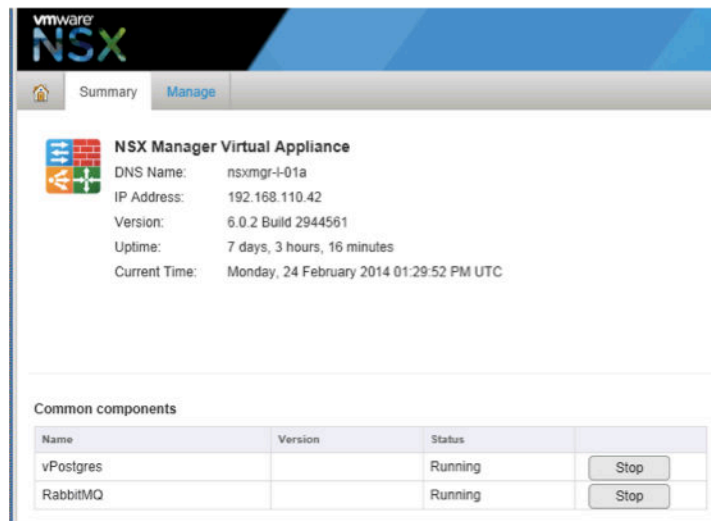
Clusters & Hosts	Installation Status	Freefall
▶  dc-1	 5.0 Uninstall	 Error

- 管理程序和控制器之间的控制层面关闭。请检查 NSX Manager 系统事件。



Timestamp	Severity	Event Source	Code	Event Message
2/26/2014 10:56:38 AM	Critical	Host messaging infrastructure	391002	Messaging infrastructure down on host.
2/26/2014 10:51:56 AM	Critical	host-22	301502	Spoofguard configuration update number 139340752032...
2/26/2014 10:51:56 AM	Critical	host-20	301502	Spoofguard configuration update number 139340752032...

如果多个 ESXi 主机受到影响，请在 NSX Manager Appliance Web UI **摘要 (Summary)** 选项卡下面检查消息总线服务的状态。如果已停止，请重新启动 RabbitMQ。



NSX Manager Virtual Appliance

DNS Name: nsxmgr-l-01a
 IP Address: 192.168.110.42
 Version: 6.0.2 Build 2944561
 Uptime: 7 days, 3 hours, 16 minutes
 Current Time: Monday, 24 February 2014 01:29:52 PM UTC

Name	Version	Status	
vPostgres		Running	Stop
RabbitMQ		Running	Stop

如果消息总线服务在 NSX Manager 上处于活动状态：

- 在 ESXi 主机上运行 `/etc/init.d/vShield-Stateful-Firewall status` 命令以检查主机上的消息总线用户环境代理状态。

```
[root@esx-01a:~] /etc/init.d/vShield-Stateful-Firewall status
vShield-Stateful-Firewall is running
```

- 检查主机上的消息总线用户环境代理日志 `/var/log/vsfwd.log`。
- 在 ESXi 主机上运行 `esxcfg-advcfg -l | grep Rmq` 命令以显示所有 Rmq 变量。应该有 16 个 Rmq 变量。

```
[root@esx-01a:~] esxcfg-advcfg -l | grep Rmq
/UserVars/RmqIpAddress [String] : Connection info for RMQ Broker
/UserVars/RmqUsername [String] : RMQ Broker Username
/UserVars/RmqPassword [String] : RMQ Broker Password
/UserVars/RmqVHost [String] : RMQ Broker VHost
/UserVars/RmqVsmRequestQueue [String] : RMQ Broker VSM Request Queue
/UserVars/RmqPort [String] : RMQ Broker Port
/UserVars/RmqVsmExchange [String] : RMQ Broker VSM Exchange
/UserVars/RmqClientPeerName [String] : RMQ Broker Client Peer Name
/UserVars/RmqHostId [String] : RMQ Broker Client HostId
/UserVars/RmqHostVer [String] : RMQ Broker Client HostVer
```



```

/UserVars/RmqClientId [String] : RMQ Broker Client Id
/UserVars/RmqClientToken [String] : RMQ Broker Client Token
/UserVars/RmqClientRequestQueue [String] : RMQ Broker Client Request Queue
/UserVars/RmqClientResponseQueue [String] : RMQ Broker Client Response Queue
/UserVars/RmqClientExchange [String] : RMQ Broker Client Exchange
/UserVars/RmqSslCertSha1ThumbprintBase64 [String] : RMQ Broker Server Certificate base64 Encoded
Sha1 Hash

```

- 在 ESXi 主机上运行 `esxcfg-advcfg -g /UserVars/RmqIpAddress` 命令。输出应显示 NSX Manager IP 地址。

```

[root@esx-01a:~] esxcfg-advcfg -g /UserVars/RmqIpAddress
Value of RmqIpAddress is 192.168.110.15

```

- 在 ESXi 主机上运行 `esxcli network ip connection list | grep 5671` 命令以查找活动消息总线连接。

```

[root@esx-01a:~] esxcli network ip connection list | grep 5671
tcp          0      0 192.168.110.51:29969      192.168.110.15:5671      ESTABLISHED    35505
newreno      vsfwd
tcp          0      0 192.168.110.51:29968      192.168.110.15:5671      ESTABLISHED    35505
newreno      vsfwd

```

要确定 netcpa 用户环境代理关闭原因，请执行以下操作：

- 在 ESXi 主机上运行 `/etc/init.d/netcpad status` 命令以检查主机上的 netcpa 用户环境代理状态。

```

[root@esx-01a:~] /etc/init.d/netcpad status
netCP agent service is running

```

- 检查 netcpa 用户环境代理配置 `/etc/vmware/netcpa/config-by-vsm.xml`。应列出 NSX Controller 的 IP 地址。

```

[root@esx-01a:~] more /etc/vmware/netcpa/config-by-vsm.xml
<config>
  <connectionList>
    <connection id="0000">
      <port>1234</port>
      <server>192.168.110.31</server>
      <sslEnabled>true</sslEnabled>
      <thumbprint>A5:C6:A2:B2:57:97:36:F0:7C:13:DB:64:9B:86:E6:EF:1A:7E:5C:36</thumbprint>
    </connection>
    <connection id="0001">
      <port>1234</port>
      <server>192.168.110.32</server>
      <sslEnabled>true</sslEnabled>
      <thumbprint>12:E0:25:B2:E0:35:D7:84:90:71:CF:C7:53:97:FD:96:EE:ED:7C:DD</thumbprint>
    </connection>
    <connection id="0002">
      <port>1234</port>
      <server>192.168.110.33</server>
      <sslEnabled>true</sslEnabled>

```

```
<thumbprint>BD:DB:BA:B0:DC:61:AD:94:C6:0F:7E:F5:80:19:44:51:BA:90:2C:8D</thumbprint>
</connection>
</connectionList>
...
```

- 运行 `esxcli network ip connection list | grep 1234` 命令以验证控制器 TCP 连接。

```
>[root@esx-01a:~] esxcli network ip connection list | grep 1234
tcp      0    0  192.168.110.51:16594    192.168.110.31:1234    ESTABLISHED    36754    newreno    netcpa-
worker
tcp      0    0  192.168.110.51:46917    192.168.110.33:1234    ESTABLISHED    36754    newreno    netcpa-
worker
tcp      0    0  192.168.110.51:47891    192.168.110.32:1234    ESTABLISHED    36752    newreno    netcpa-
worker
```

VXLAN 准备

NSX 准备 VXLAN 用户选择的 DVS。

这要求 NSX 在 DVS 上创建 DVPortgroup 以供 VTEP vmknics 使用。

成组、负载均衡方法、MTU 和 VLAN ID 是在 VXLAN 配置期间选择的。成组和负载均衡方法必须与为 VXLAN 选择的 DVS 配置相匹配。

MTU 必须设置为至少 1600，并且不小于在 DVS 上已配置的大小。

创建的 VTEP 数取决于选择的成组策略和 DVS 配置。

VXLAN 准备期间的常见问题

在 VXLAN 配置期间，可能遇到的典型问题如下所示：

- 为 VXLAN 选择的成组方法与 DVS 支持的方法不匹配。请参见《VMware NSX for vSphere 网络虚拟化设计指南》，网址为 <https://communities.vmware.com/docs/DOC-27683>。
- 为 VTEP 选择的 VLAN ID 不正确。
- 选择了 DHCP 以分配 VTEP IP 地址，但没有可用的 DHCP 服务器。
- vmknics 缺少“强制同步”配置。
- vmknics 具有错误的 IP 地址。

重要的端口号

VXLAN UDP 端口用于 UDP 封装。默认情况下，VXLAN UDP 端口号为 8472。在使用硬件 VTEP 的 NSX 6.2 和更高版本安装中，您必须改用 VXLAN UDP 端口号 4789。可以通过 REST API 修改端口号。

```
PUT /2.0/vdn/config/vxlan/udp/port/4789
```

必须从 NSX Manager 中为主机打开端口 80。这用于下载 VIB/代理。

来自、到达以及在 ESXi 主机、vCenter Server 和 NSX 数据安全之间的端口 443/TCP。

此外，必须在 NSX Manager 上打开以下端口：

- 443/TCP：需要使用该端口在 ESXi 主机上下载以部署 OVA 文件，使用 REST API 以及使用 NSX Manager 用户界面。
- 80/TCP：需要使用该端口启动到 vSphere SDK 的连接，以及在 NSX Manager 和 NSX 主机模块之间进行消息传送。
- 1234/TCP：需要使用该端口在 ESXi 主机和 NSX Controller 群集之间进行通信。
- 5671：Rabbit MQ（一种消息总线技术）需要使用该端口。
- 22/TCP：需要使用该端口通过控制台 (SSH) 访问 CLI。默认情况下，将关闭该端口。

如果将群集中的主机从 vCenter Server 5.0 版升级到 5.5，必须在这些主机上打开端口 80 和 443 才能成功安装 Guest Introspection。

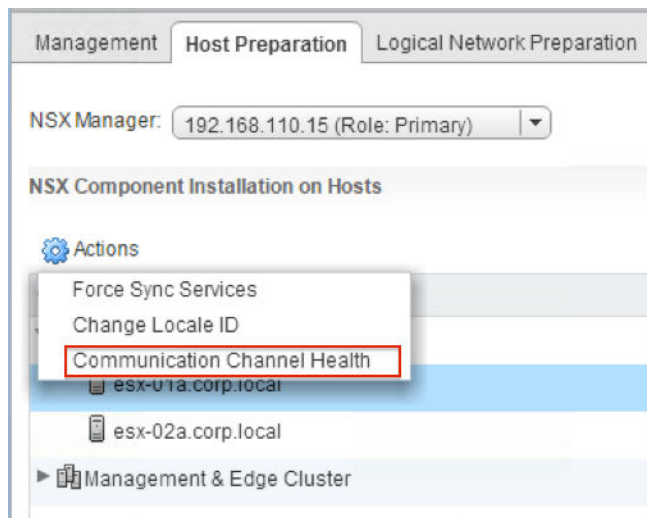
检查通信通道运行状况

从 vSphere Web Client 中，您可以检查各种组件之间的通信状态。

要检查 NSX Manager 和防火墙代理之间、NSX Manager 和控制层面代理之间以及控制层面代理和控制器之间的通信通道运行状况，请执行以下步骤：

- 1 在 vSphere Web Client 中，导航到**网络和安全 (Networking & Security) > 安装 (Installation) > 主机准备 (Host Preparation)**。
- 2 选择或展开一个群集，然后选择一个主机。单击**操作 (Actions)** (⚙️)，然后单击**通信通道运行状况 (Communication Channel Health)**。

随即显示通信通道运行状况信息。





如果主机的三个连接之一的状态发生变化，则会在日志中写入一条消息。在日志消息中，连接的状态可能是“已连接”、“已关闭”或“不可用”（在 vSphere Web Client 中显示为“未知”）。如果状态由“已连接”变为“已关闭”或“不可用”，则会生成一条警告消息。例如：

```
2016-05-23 23:36:34.736 GMT+00:00 WARN TaskFrameworkExecutor-25 VdnInventoryFacadeImpl
$HostStatusChangedEventHandler:200 - Host Connection Status Changed: Event Code: 1941, Host:
esx-04a.corp.local (ID: host-46), NSX Manager - Firewall Agent: UP, NSX Manager - Control Plane Agent:
UP, Control Plane Agent - Controllers: DOWN.
```

如果状态由“已关闭”或“不可用”变为“已连接”，则会生成一条类似于警告消息的信息消息。例如：

```
2016-05-23 23:55:12.736 GMT+00:00 INFO TaskFrameworkExecutor-25 VdnInventoryFacadeImpl
$HostStatusChangedEventHandler:200 - Host Connection Status Changed: Event Code: 1938, Host:
esx-04a.corp.local (ID: host-46), NSX Manager - Firewall Agent: UP, NSX Manager - Control Plane Agent:
UP, Control Plane Agent - Controllers: UP.
```

解决 NSX Manager 问题

问题

- 安装 VMware NSX Manager 失败。
- 升级 VMware NSX Manager 失败。
- 登录到 VMware NSX Manager 失败。
- 访问 VMware NSX Manager 失败。

解决方案

验证每个故障排除步骤是否适用于您的环境。每个步骤提供了相应说明，以消除可能的根源并在必要时采取纠正措施。这些步骤按最适当的顺序进行排列，以查找问题并确定相应的解决方案。不要跳过某个步骤。

步骤

- 1 请参阅当前版本的《NSX 发行说明》以查看是否在错误修复中解决了该问题。
- 2 确保在安装 VMware NSX Manager 时满足最低系统要求。
请参见 NSX 安装指南。
- 3 验证是否在 NSX Manager 中打开所需的所有端口。
请参见 NSX 安装指南。

4 安装问题：

- 如果配置 Lookup Service 或 vCenter Server 失败，请验证 NSX Manager 和 Lookup Service 设备上的时间是否同步。请在 NSX Manager 和 Lookup Service 上使用相同的 NTP 服务器配置。还要确保正确配置了 DNS。
- 验证是否正确安装了 OVA 文件。如果无法安装 NSX OVA 文件，vSphere Client 中的错误窗口将指出发生故障的位置。还要验证下载的 OVA/OVF 文件的 MD5 校验和。
- 验证 ESXi 主机上的时间是否与 NSX Manager 同步。
- VMware 建议您在安装 NSX Manager 后立即计划 NSX Manager 数据备份。

5 升级问题：

- 在升级之前，请参见“产品互操作性列表”页中的最新互操作性信息。
- VMware 建议在升级之前备份当前配置并下载技术支持日志。
- 在升级 NSX Manager 后，可能需要强制与 vCenter Server 重新进行同步。为此，请登录到 NSX Manager Web 界面 GUI。接下来，转到 **管理 vCenter 注册 > NSX 管理服务 > 编辑 (Manage vCenter Registration > NSX Management Service > Edit)**，然后重新输入管理用户密码。

6 性能问题：

- 确保满足最低 vCPU 要求。
- 验证根 (/) 分区是否具有足够的空间。您可以登录到 ESXi 主机并键入 `df -h` 命令以验证这种情况。

例如：

```
[root@esx-01a:~] df -h
Filesystem      Size  Used Available Use% Mounted on
NFS              111.4G  80.8G   30.5G   73% /vmfs/volumes/ds-site-a-nfs01
vfat             249.7M  172.2M   77.5M   69% /vmfs/volumes/68cb5875-d887b9c6-a805-65901f83f3d4
vfat             249.7M  167.7M   82.0M   67% /vmfs/volumes/fe84b77a-b2a8860f-38cf-168d5dfe66a5
vfat             285.8M  206.3M   79.6M   72% /vmfs/volumes/54de790f-05f8a633-2ad8-00505603302a
```

- 使用 `esxtop` 命令检查哪些进程使用大量 CPU 和内存。
- 如果 NSX Manager 在日志中遇到任何内存不足错误，请验证 `/common/dumps/java.hprof` 文件是否存在。如果该文件存在，请创建该文件的副本，并在 NSX 技术支持日志包中包含该副本。
- 验证在环境中是否存在存储延迟问题。
- 尝试将 NSX Manager 迁移到另一个 ESXi 主机。

7 连接问题：

- 如果 NSX Manager 与 vCenter Server 或 ESXi 主机之间出现连接问题，请登录到 NSX Manager CLI 控制台，然后运行 `debug connection IP_of_ESXi_or_VC` 命令并检查输出。
- 确认已启动 Virtual Center Web 管理服务，并且浏览器未处于错误状态。

- 如果未更新 NSX Manager Web 用户界面 (UI)，您可以尝试禁用并重新启用 Web 服务以解决该问题。请参见 <https://kb.vmware.com/kb/2126701>。
- 在 ESXi 主机上使用 `esxtop` 命令验证 NSX Manager 使用的端口组和上行链路网卡。有关详细信息，请参见 <https://kb.vmware.com/kb/1003893>。
- 尝试将 NSX Manager 迁移到另一个 ESXi 主机。
- 从 vSphere Web Client 的**监控 (Monitor)**选项卡中检查 NSX Manager 虚拟机设备**任务和事件 (Tasks and Events)**选项卡。
- 如果 NSX Manager 与 vCenter Server 之间出现连接问题，请尝试将 NSX Manager 迁移到运行 vCenter Server 虚拟机的相同 ESXi 主机以消除可能的底层物理网络问题。

请注意，这仅适用于两个虚拟机位于同一 VLAN/端口组的情况。

从 NSX Controller 故障恢复

当出现 NSX Controller 故障时，可能仍有两个控制器正在工作。此时保持着群集多数，并且控制层面仍继续正常工作。尽管如此，您也必须将三个控制器全部删除并添加新的控制器，以便维护完全正常工作的三节点群集。

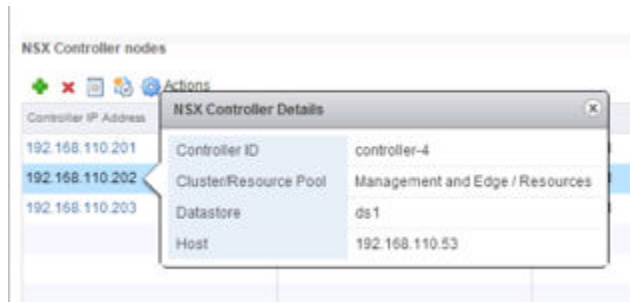
当一个或多个控制器遇到不可恢复的灾难性错误，或者一个或多个控制器虚拟机变为无法访问并且无法修复时，建议删除控制器群集。

在这种情况下，虽然部分控制器看似运行良好，我们也建议删除所有控制器。建议的过程是先创建新的控制器群集，然后在 NSX Manager 上使用“更新控制器状态”机制将状态同步到控制器。

步骤

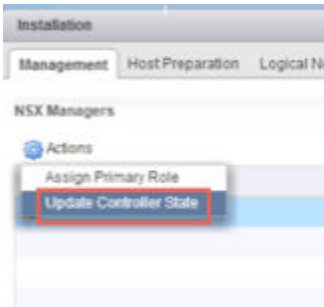
- 1 登录到 vSphere Web Client。
- 2 从 **网络和安全** 中，单击**安装 > 管理**。
- 3 在“NSX Controller 节点”部分中，单击每个控制器并获取详细信息屏幕的屏幕截图/打印屏幕，或者记下配置信息以供将来参考。

例如：



- 4 在“NSX Controller 节点”部分中，将三个节点全部删除，方法是选择每个节点并单击**删除节点 (x)**图标。
当系统中不存在任何控制器时，主机将在所谓的“无头”模式下工作。新虚拟机或已执行 vMotion 操作的虚拟机将遇到网络问题，直至部署了新的控制器并且同步已完成为止。

- 5 部署三个新的 NSX Controller 节点，方法是单击**添加节点 (+)** 图标。
- 6 在“添加控制器”对话框中，选择要添加节点的数据中心，然后配置控制器设置。
 - a 选择适当的群集。
 - b 在群集和存储中选择一个主机。
 - c 选择分布式端口组。
 - d 选择要将其中的 IP 地址分配给节点的 IP 池。
 - e 单击**确定**，等待安装完成，并确保所有节点的状态均为“正常”。
- 7 重新同步控制器状态，方法是单击**操作 > 更新控制器状态**。

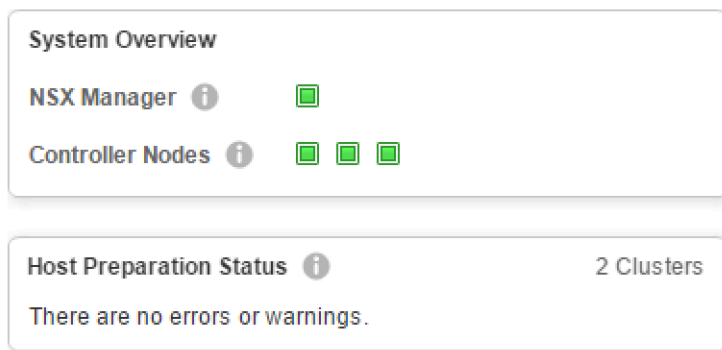


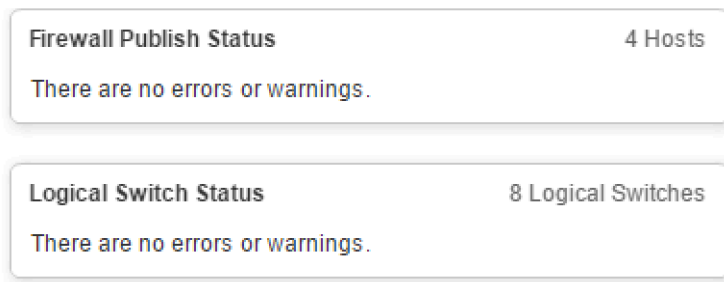
更新控制器状态将当前 VXLAN 和分布式逻辑路由器配置（包括跨 VC NSX 部署中的通用对象）从 NSX Manager 推送到控制器群集。

使用 NSX 仪表板

通过将 NSX 组件的总体运行状况显示在一个中央视图中，NSX 仪表板简化了故障排除过程。

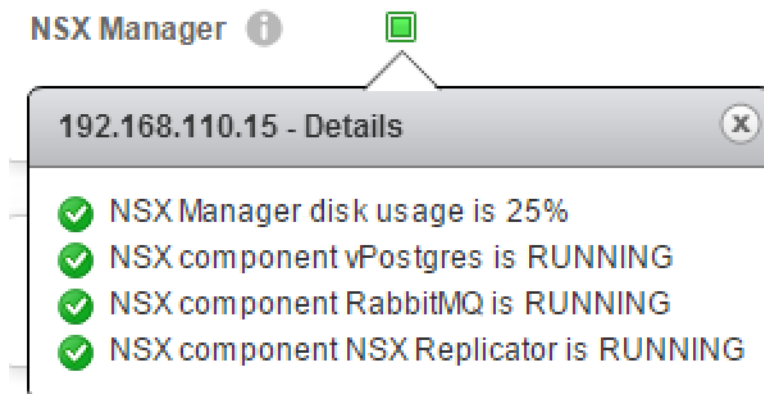
您可以从 vCenter Web Client 的**网络和安全 > 仪表板 (> Networking & Security > Dashboard)**中访问该仪表板。





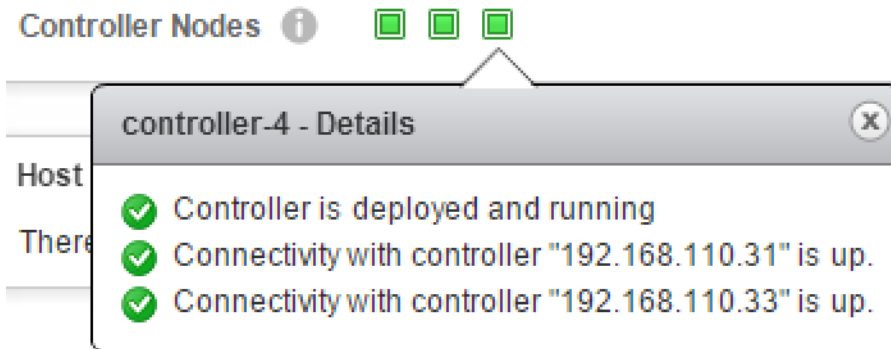
该仪表板检查以下状态：

- NSX 基础架构 - NSX Manager 状态
 - 监控以下服务的组件状态
 - 数据库服务
 - 消息总线服务
 - 复制服务 - 还会监控复制错误
 - NSX Manager 磁盘使用率：
 - 黄色（磁盘使用率 > 80%）
 - 红色（磁盘使用率 > 90%）



- NSX 基础架构 - NSX Controller 状态
 - 控制器节点状态（正在运行/正在部署/正在移除/失败/未知）
 - 控制器对等连接状态
 - 控制器虚拟机状态（已关闭电源/已删除）

- 控制器磁盘延迟警报



- NSX 基础架构 - 主机状态
 - 部署相关：
 - 具有安装失败状态的群集数
 - 需要升级的群集数
 - 正在进行安装的群集数
 - 防火墙：
 - 禁用了防火墙的群集数
 - 防火墙状态为红色/黄色的群集数
 - VXLAN：
 - 未配置 VXLAN 的群集数
 - VXLAN 状态为红色/黄色的群集数
- NSX 服务 - 防火墙发布状态
 - 防火墙发布状态为“失败”的主机数
- NSX 服务 - 逻辑网络状态
 - 具有“错误”和“警告”状态的逻辑交换机数
 - 标记是否删除虚拟线路的后备 DVS 端口组

使用 show host health-status 命令

从 NSX Manager 集中式 CLI 中，您可以检查每个 ESXi 主机的运行状态。

运行状态将报告为 **critical**、**unhealthy** 或 **healthy**。

例如：

```
nsxmgr> show host host-30 health-status
status: HEALTHY

nsxmgr> show host host-29 health-status
UNHEALTHY, Standard Switch vSwitch1 has no uplinks.
```

```
UNHEALTHY, Storage volume datastore1 has no enough free spaces: 19.% free.
status: UNHEALTHY
```

```
nsxmgr> show host host-28 health-status
CRITICAL, VXLAN VDS vds-site-a VNI 200000 multicast addr is not synchronized with VSM: 0.0.0.0.
CRITICAL, VXLAN VDS vds-site-a VNI 200003 multicast addr is not synchronized with VSM: 0.0.0.0.
CRITICAL, VXLAN VDS vds-site-a VNI 5000 multicast addr is not synchronized with VSM: 0.0.0.0.
Status: CRITICAL
```

也可以通过 NSX Manager API 调用 `host-check` 命令。

设置 NSX 组件的日志记录级别

您可以为每个 NSX 组件设置日志记录级别。

支持的级别因组件而异，如下所示。

```
nsxmgr> set
  hardware-gateway  Show Logical Switch Commands
  PACKAGE-NAME      Set log level
  controller        Show Logical Switch Commands
  host              Show Logical Switch Commands

nsxmgr> set hardware-gateway agent 10.1.1.1 logging-level
  ERROR
  WARN
  INFO
  DEBUG
  TRACE

nsxmgr-01a> set <package-name> logging-level
  OFF
  FATAL
  ERROR
  WARN
  INFO
  DEBUG
  TRACE

nsxmgr> set controller 192.168.110.31
  java-domain      Set controller node log level
  native-domain    Set controller node log level

nsxmgr> set controller 192.168.110.31 java-domain logging-level
  OFF
  FATAL
  ERROR
  WARN
  INFO
  DEBUG
  TRACE

nsxmgr> set controller 192.168.110.31 native-domain logging-level
  ERROR
```

```

WARN
INFO
DEBUG
TRACE

nsxmgr> set host host-28
  netcpa  Set host node log level by module
  vdl2    Set host node log level by module
  vdr     Set host node log level by module

nsxmgr> set host host-28 netcpa logging-level
  FATAL
  ERROR
  WARN
  INFO
  DEBUG

nsxmgr> set host host-28 vdl2 logging-level
  ERROR
  INFO
  DEBUG
  TRACE

nsxmgr> set host host-28 vdr logging-level
  OFF
  ERROR
  INFO

```

vSphere ESX Agent Manager

vSphere ESX Agent Manager (EAM) 自动完成部署和管理 vSphere ESX Agent 的过程，同时扩展 ESXi 主机功能以提供 vSphere 解决方案所需的额外服务。

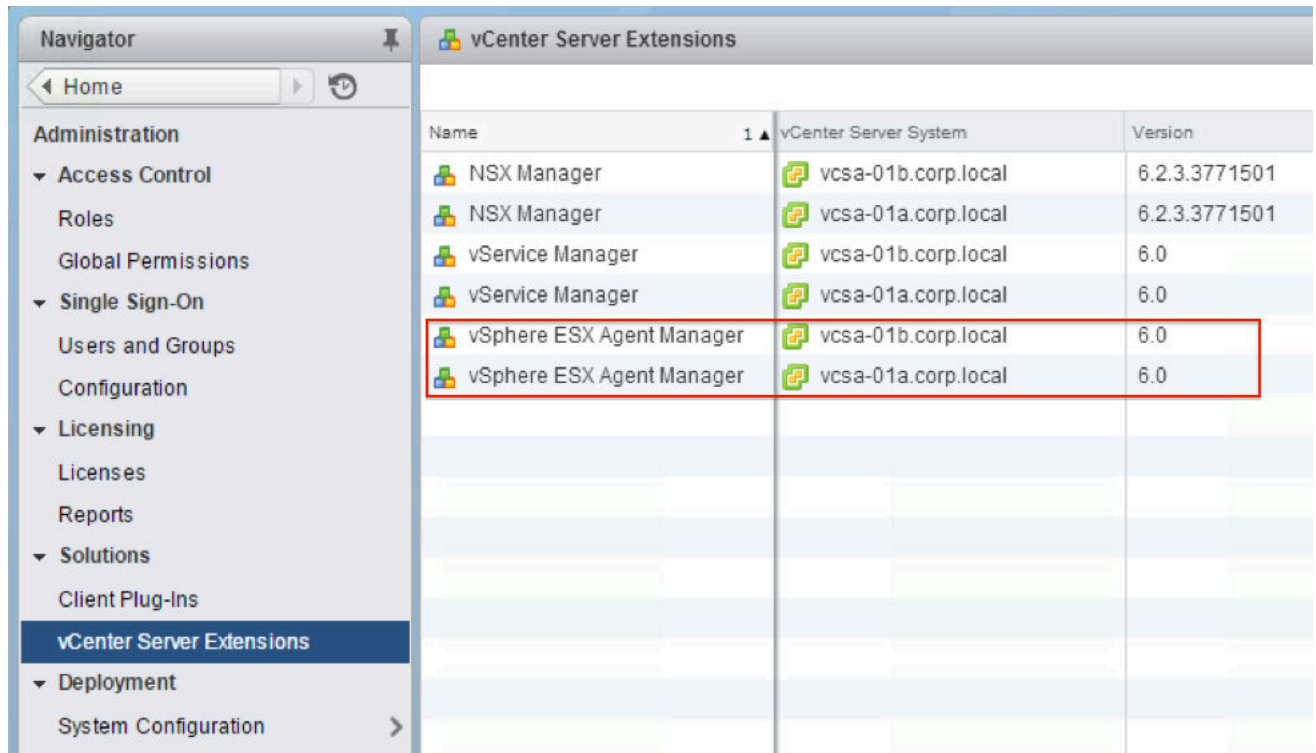
ESX 代理与 NSX 故障排除有关，因为 NSX 部署可能需要使用特定的网络筛选器或防火墙配置才能正常工作。防火墙配置可以使用 ESX 代理连接到 vSphere Hypervisor，并使用该配置特定的功能扩展主机。例如，ESX 代理可以筛选网络流量，用作防火墙或收集有关主机上的虚拟机的其他信息。

ESX 代理虚拟机类似于 Windows 或 Linux 中的服务。这些服务在操作系统启动时启动，并在操作系统关闭时停止。ESX 代理虚拟机的行为对用户是透明的。如果 ESXi 操作系统已启动，并且已置备所有 ESX 代理虚拟机并打开电源，vSphere 主机将进入就绪状态。

要将代理与 vSphere ESX Agent Manager 集成在一起并扩展 ESXi Server 功能，必须将 ESX 代理打包为 OVF 或 VIB 模块。

您可以通过 EAM 监控 ESX 代理的运行状况，并阻止用户在 ESX 代理上执行某些可能影响使用这些代理的虚拟机的操作。它还管理代理 VIB 和虚拟机的生命周期。例如，ESX Agent Manager 可能禁止关闭 ESX 代理虚拟机的电源，或者禁止从包含使用该代理的其他虚拟机的 ESXi 主机中移动该代理。

下面的屏幕截图显示了用于访问 ESX Agent Manager 的 UI。



Name	vCenter Server System	Version
NSX Manager	vcsa-01b.corp.local	6.2.3.3771501
NSX Manager	vcsa-01a.corp.local	6.2.3.3771501
vService Manager	vcsa-01b.corp.local	6.0
vService Manager	vcsa-01a.corp.local	6.0
vSphere ESX Agent Manager	vcsa-01b.corp.local	6.0
vSphere ESX Agent Manager	vcsa-01a.corp.local	6.0

vSphere ESX Agent Manager (EAM) 的日志和服务

EAM 日志包含在 vCenter 日志包中。

- Windows - C:\ProgramData\VMware\vCenterServer\logs\eam\eam.log
- VCSA - /var/log/vmware/vpx/eam.log
- ESXi - /var/log/esxupdate.log

vSphere ESX 代理和代理机构

vSphere ESX 代理机构映射到准备的 NSX 主机群集。每个 ESX 代理机构作为 ESX 代理的容器。ESX 代理机构聚合有关它们管理的代理的信息。因此，通过聚合与 ESX 代理有关的所有问题，ESX 代理机构提供它们包含的 ESX 代理的概要信息。

ESX Agent Manager 在代理机构运行时信息中报告问题。如果管理员单击 ESX Agent Manager 选项卡中的 **解决问题**，ESX Agent Manager 可以自动解决某些问题。例如，如果关闭了 ESX 代理电源，可以重新打开电源。

注 如果 ESX 代理机构的范围为空，则没有在其中部署 ESX 代理的计算资源，因此，不会部署任何 ESX 代理。在这种情况下，ESX Agent Manager 确定 ESX 代理机构是否正常运行并将状态设置为绿色。

每个代理机构的配置指定了该代理机构如何部署其代理和 VIB。请参阅

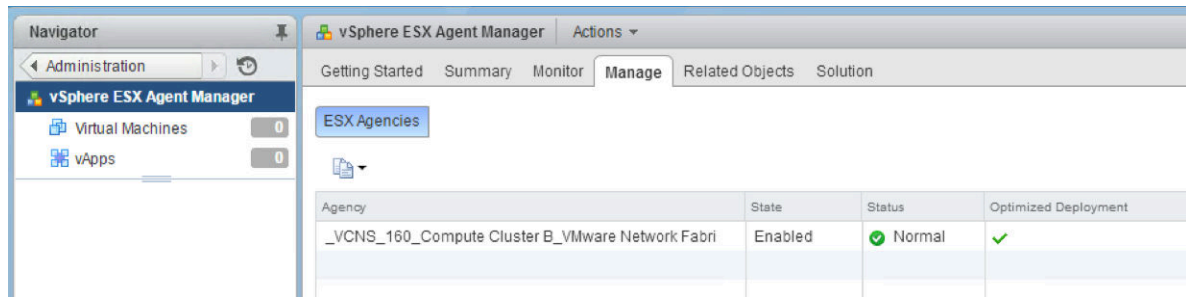
<https://pubs.vmware.com/vsphere-60/index.jsp#com.vmware.eam.apiref.doc/eam.Agency.ConfigInfo.html>

。

重要 确保在开始安装 NSX 之前将 `bypassVumEnabled` 标记设置 `True`，并在安装后将其改回到 `False`。请参阅 <https://kb.vmware.com/kb/2053782>。

要在 vSphere Web Client 中检查 EAM 状态，请转到**管理 > vCenter Server 扩展**。

EAM 的**管理**选项卡显示有关运行的代理机构的信息，列出任何孤立的 ESX 代理，并记录有关 ESX Agent Manager 管理的 ESX 代理的信息。



有关代理和代理机构的详细信息，请参见

https://pubs.vmware.com/vsphere-60/index.jsp#com.vmware.vsphere.ext_solutions.doc/GUID-40838DE9-6AD1-45E3-A1DE-B2B24A9E715A.html。

NSX CLI 速查表

表 2-1. 检查 ESXi 主机上的 NSX 安装 - 从 NSX Manager 中运行的命令

说明	NSX Manager 上的命令	备注
列出所有群集以获取群集 ID	<code>show cluster all</code>	查看所有群集信息
列出群集中的所有主机以获取主机 ID	<code>show cluster CLUSTER-ID</code>	查看群集中的主机、主机 ID 和主机准备安装状态列表
列出主机上的所有虚拟机	<code>show host HOST-ID</code>	查看特定主机信息、虚拟机、虚拟机 ID 和电源状态

表 2-2. 检查 ESXi 主机上的 NSX 安装 - 从主机中运行的命令

说明	主机上的命令	备注
加载了三个 VIB： <code>esx-vxlan</code> 、 <code>esx-vsip</code> 、 <code>esx-dvfilter-switch-security</code>	<code>esxcli software vib get --vibname <name></code>	检查安装的版本/日期 <code>esxcli software vib list</code> 显示系统上的所有 VIB 的列表
列出当前在系统中加载的所有系统模块：	<code>esxcli system module list</code>	较旧的等效命令： <code>vmkload_mod -l grep -E vdl2 vdrb vsip dvfilter-switch-security</code>
加载了四个模块： <code>vdl2</code> 、 <code>vdrb</code> 、 <code>vsip</code> 、 <code>dvfilter-switch-security</code>	<code>esxcli system module get -m <name></code>	为每个模块运行该命令

表 2-2. 检查 ESXi 主机上的 NSX 安装 - 从主机中运行的命令（续）

说明	主机上的命令	备注
两个用户环境代理 (UWA): netcpad、vsfwd	/etc/init.d/vShield-Stateful-Firewall status /etc/init.d/netcpad status	
检查 UWA 连接: 端口 1234 到控制器以及端口 5671 到 NSX Manager	esxcli network ip connection list grep 1234 esxcli network ip connection list grep 5671	控制器 TCP 连接 消息总线 TCP 连接
检查 EAM 状态	Web UI, 检查管理 > vCenter ESX Agent Manager	

表 2-3. 检查 ESXi 主机上的 NSX 安装 - 主机网络命令

说明	主机网络命令	备注
列出物理网卡/vmnic	esxcli network nic list	检查网卡类型、驱动程序类型、链路状态、MTU
物理网卡详细信息	esxcli network nic get -n vmnic#	检查驱动程序和固件版本以及其他详细信息
列出 vmk 网卡以及 IP 地址/MAC/MTU 等	esxcli network ip interface ipv4 get	确保正确实例化 VTEP
每个 vmk 网卡的详细信息, 包括 vDS 信息	esxcli network ip interface list	确保正确实例化 VTEP
每个 vmk 网卡的详细信息, 包括 VXLAN vmk 的 vDS 信息	esxcli network ip interface list --netstack=vxlan	确保正确实例化 VTEP
查找与该主机的 VTEP 关联的 VDS 名称	esxcli network vswitch dvs vmware vxlan list	确保正确实例化 VTEP
从 VXLAN 专用 TCP/IP 堆中执行 Ping 操作	ping ++netstack=vxlan -I vmk1 x.x.x.x	要解决 VTEP 通信问题: 添加 -d -s 1572 选项以确保传输网络的 MTU 适用于 VXLAN
查看 VXLAN 专用 TCP/IP 堆的路由表	esxcli network ip route ipv4 list -N vxlan	解决 VTEP 通信问题
查看 VXLAN 专用 TCP/IP 堆的 ARP 表	esxcli network ip neighbor list -N vxlan	解决 VTEP 通信问题

表 2-4. 检查 ESXi 主机上的 NSX 安装 - 主机日志文件

说明	日志文件	备注
从 NSX Manager 中	show manager log follow	跟踪 NSX Manager 日志 适用于实时故障排除
主机的任何安装相关日志	/var/log/esxupdate.log	
与主机相关的问题	/var/log/vmkernel.log	
VMkernel 警告、消息、警示和可用性报告	/var/log/vmksummary.log /var/log/vmkwarning.log	

表 2-4. 检查 ESXi 主机上的 NSX 安装 - 主机日志文件（续）

说明	日志文件	备注
捕获模块加载故障	/var/log/syslog	IXGBE 驱动程序故障 NSX 模块相关性故障是重要指标
在 vCenter 上，ESX Agent Manager 负责进行更新	在 vCenter 日志 eam.log 中	

表 2-5. 检查逻辑交换 - 从 NSX Manager 中运行的命令

说明	NSX Manager 上的命令	备注
列出所有逻辑交换机	show logical-switch list all	列出在 API、传输区域和 vdnscope 中使用的所有逻辑交换机及其 UUID

表 2-6. 逻辑交换 - 从 NSX Controller 中运行的命令

说明	控制器上的命令	备注
查找作为 VNI 所有者的控制器	show control-cluster logical-switches vni 5000	记下输出中的控制器 IP 地址并通过 SSH 访问该地址
查找连接到该 VNI 的该控制器的所有主机	show control-cluster logical-switch connection-table 5000	输出中的源 IP 地址是主机的管理接口，而端口号是 TCP 连接的源端口
查找注册以托管该 VNI 的 VTEP	show control-cluster logical-switches vtep-table 5002	
列出为该 VNI 上的虚拟机获悉的 MAC 地址	show control-cluster logical-switches mac-table 5002	指出 MAC 地址实际位于报告该地址的 VTEP 上
列出虚拟机 IP 更新填充的 ARP 缓存	show control-cluster logical-switches arp-table 5002	ARP 缓存在 180 秒后过期
对于特定的主机/控制器对，找出主机已加入的 VNI	show control-cluster logical-switches joined-vnis <host_mgmt_ip>	

表 2-7. 逻辑交换 - 从主机中运行的命令

说明	主机上的命令	备注
检查主机 VXLAN 是否同步	esxcli network vswitch dvs vmware vxlan get	显示同步状态和用于封装的端口
查看连接的虚拟机以及用于数据路径捕获的本地交换机端口 ID	net-stats -l	提供了一种更好的方法以获取特定虚拟机的虚拟机交换机端口
验证是否加载了 VXLAN 内核模块 vdl2	esxcli system module get -m vdl2	显示指定的模块的完整详细信息 验证版本
验证是否安装了正确的 VXLAN VIB 版本	esxcli software vib get --vibName esx-vxlan	显示指定的 VIB 的完整详细信息 验证版本和日期
验证主机是否了解逻辑交换机中的其他主机	esxcli network vswitch dvs vmware vxlan network vtep list --vxlan-id=5001 --vds-name=Compute_VDS	显示该主机了解并托管 vtep 5001 的所有 VTEP 的列表

表 2-7. 逻辑交换 - 从主机中运行的命令（续）

说明	主机上的命令	备注
验证逻辑交换机的控制层面是否已启动并处于活动状态	<code>esxcli network vswitch dvs vmware vxlan network list --vds-name Compute_VDS</code>	确保控制器连接已启动并且端口/Mac 数与该主机上的 LS 中的虚拟机数相匹配。
验证主机是否获悉所有虚拟机的 MAC 地址	<code>esxcli network vswitch dvs vmware vxlan network mac list --vds-name Compute_VDS --vxlan-id=5000</code>	这会列出该主机上的 VNI 5000 虚拟机的所有 MAC
验证主机是否在本机缓存远程虚拟机的 ARP 条目	<code>esxcli network vswitch dvs vmware vxlan network arp list --vds-name Compute_VDS --vxlan-id=5000</code>	验证主机是否在本机缓存远程虚拟机的 ARP 条目
验证虚拟机是否连接到 LS 并映射到本地 VMKnic 还会显示虚拟机 dvPort 映射到的 vmknic ID	<code>esxcli network vswitch dvs vmware vxlan network port list --vds-name Compute_VDS --vxlan-id=5000</code>	只要 VNI 连接到路由器，就会始终列出 vdrport
查看 vmknic ID 以及它们映射到的交换机端口/上行链路	<code>esxcli network vswitch dvs vmware vxlan vmknic list --vds-name=DSwitch-Res01</code>	

表 2-8. 检查逻辑交换 - 日志文件

说明	日志文件	备注
主机始终连接到托管其 VNI 的控制器	<code>/etc/vmware/netcpa/config-by-vsm.xml</code>	该文件应始终列出环境中的所有控制器。 <code>config-by-vsm.xml</code> 文件是由 <code>netcpa</code> 进程创建的 Vsfwd 仅为 <code>netcpa</code> 提供通道 Netcpad 连接到端口 15002 上的 vsfwd
<code>config-by-vsm.xml</code> 文件是 NSX Manager 使用 vsfwd 推送的 如果 <code>config-by-vsm.xml</code> 文件不正确，请查看 vsfwd 日志	<code>/var/log/vsfwd.log</code>	分析该文件以查找错误 要重新启动进程，请运行以下命令： <code>/etc/init.d/vShield-Stateful-Firewall stop start</code>
到控制器的连接是使用 <code>netcpa</code> 建立的	<code>/var/log/netcpa.log</code>	分析该文件以查找错误
VDL2 模块日志位于 <code>vmkernel.log</code> 中	<code>/var/log/vmkernel.log</code>	在 <code>/var/log/vmkernel.log</code> 中检查“具有 VXLAN 前缀”的 VDL2 模块日志：

表 2-9. 检查逻辑路由 - 从 NSX Manager 中运行的命令

说明	NSX Manager 上的命令	备注
用于 ESG 的命令	<code>show edge</code>	用于 Edge 服务网关 (ESG) 的 CLI 命令以“ <code>show edge</code> ”开头
用于 DLR 控制虚拟机的命令	<code>show edge</code>	用于分布式逻辑路由器 (DLR) 控制虚拟机的 CLI 命令以“ <code>show edge</code> ”开头
用于 DLR 的命令	<code>show logical-router</code>	用于分布式逻辑路由器 (DLR) 的 CLI 命令以 <code>show logical-router</code> 开头
列出所有 Edge	<code>show edge all</code>	列出支持集中式 CLI 的所有 Edge
列出 Edge 的所有服务和部署详细信息	<code>show edge EDGE-ID</code>	查看 Edge 服务网关信息

表 2-9. 检查逻辑路由 - 从 NSX Manager 中运行的命令（续）

说明	NSX Manager 上的命令	备注
列出 Edge 的命令选项	show edge EDGE-ID ?	查看详细信息，例如，版本、日志、NAT、路由表、防火墙、配置、接口和服务
查看路由详细信息	show edge EDGE-ID ip ?	查看路由信息、BGP、OSPF 和其他详细信息
查看路由表	show edge EDGE-ID ip route	查看 Edge 中的路由表
查看路由邻居	show edge EDGE-ID ip ospf neighbor	查看路由邻居关系
查看逻辑路由器连接信息	show logical-router host hostID connection	验证连接的 LIF 数是否正确，成组策略是否正确以及是否使用相应的 vDS
列出在主机上运行的所有逻辑路由器实例	show logical-router host hostID dlr all	验证 LIF 和路由数 在逻辑路由器的所有主机上，控制器 IP 应该相同 Control Plane Active 应该为 yes --brief 提供了精简响应
检查主机上的路由表	show logical-router host hostID dlr dlrID route	这是控制器推送到传输区域中的所有主机的路由表 在所有主机上，该表必须是相同的 如果在少数主机上缺少某些路由，请尝试从控制器中运行前面提到的 sync 命令 E 标记表示路由是通过 ECMP 获悉的
检查主机上的 DLR 的 LIF	show logical-router host hostID dlr dlrID interface (all intName) verbose	LIF 信息将从控制器推送到主机 可以使用该命令确保主机了解应了解的所有 LIF

表 2-10. 检查逻辑路由 - 从 NSX Controller 中运行的命令

说明	NSX Controller 上的命令	备注
查找所有逻辑路由器实例	show control-cluster logical-routers instance all	这会列出逻辑路由器实例以及传输区域中具有逻辑路由器实例的所有主机 此外，还会显示为该逻辑路由器提供服务的控制器
查看每个逻辑路由器的详细信息	show control-cluster logical-routers instance 0x570d4555	IP 列显示该 DLR 所在的所有主机的 vmk0 IP 地址
查看连接到逻辑路由器的所有接口	show control-cluster logical-routers interface-summary 0x570d4555	IP 列显示该 DLR 所在的所有主机的 vmk0 IP 地址
查看该逻辑路由器获悉的所有路由	show control-cluster logical-routers routes 0x570d4555	请注意，IP 列显示该 DLR 所在的所有主机的 vmk0 IP 地址
显示建立的所有网络连接，类似于 net stat 输出	show network connections of-type tcp	检查要排除故障的主机是否建立到控制器的 netcpa 连接

表 2-10. 检查逻辑路由 - 从 NSX Controller 中运行的命令（续）

说明	NSX Controller 上的命令	备注
将接口从控制器同步到主机	<code>sync control-cluster logical-routers interface-to-host <logical-router-id> <host-ip></code>	如果新接口连接到逻辑路由器，但未同步到所有主机，这是非常有用的
将路由从控制器同步到主机	<code>sync control-cluster logical-routers route-to-host <logical-router-id> <host-ip></code>	如果在少数主机上缺少某些路由，但这些路由在大多数主机上可用，这是非常有用的

表 2-11. 检查逻辑路由 - 从 Edge 中运行的命令

说明	Edge 或逻辑路由器控制虚拟机上的命令	备注
查看配置	<code>show configuration <global bgp ospf ...></code>	
查看获悉的路由	<code>show ip route</code>	确保路由和转发表保持同步
查看转发表	<code>show ip forwarding</code>	确保路由和转发表保持同步
查看 vDR 接口	<code>show interface</code>	在输出中显示的第一个网卡是 VDR 接口 VDR 接口不是该虚拟机上的真正 vNIC 连接到 VDR 的所有子网具有“内部”类型
查看其他接口（管理）	<code>show interface</code>	管理/HA 接口是逻辑路由器控制虚拟机上的真正 vNIC 如果启用 HA 而未指定 IP 地址，则使用 169.254.x.x/30 如果为管理接口分配了 IP 地址，则会在此处显示该地址
调试协议	<code>debug ip ospf</code> <code>debug ip bgp</code>	在查看配置问题（例如，不匹配的 OSPF 区域、计时器和错误的 ASN）时，这是非常有用的 注意：只能在 Edge 控制台上查看输出（而不能通过 SSH 会话）
OSPF 命令	<code>show configuration ospf</code> <code>show ip ospf interface</code> <code>show ip ospf neighbor</code> <code>show ip route ospf</code> <code>show ip ospf database</code> <code>show tech-support</code> （并查找字符串“EXCEPTION”和“PROBLEM”）	
BGP 命令	<code>show configuration bgp</code> <code>show ip bgp neighbor</code> <code>show ip bgp</code> <code>show ip route bgp</code> <code>show ip forwarding</code> <code>show tech-support</code> （查找字符串“EXCEPTION”和“PROBLEM”）	

表 2-12. 检查逻辑路由 - 主机中的日志文件

说明	日志文件	备注
vsfwd 将 VDR 实例信息推送到主机并保存为 XML 格式	/etc/vmware/netcpa/config-by-vsm.xml	如果在主机上缺少 VDR 实例，请先查看该文件以确定是否列出该实例 如果未列出，请重新启动 vsfwd 此外，还可以使用该文件确保主机了解所有控制器
上述文件是 NSX Manager 使用 vsfwd 推送的 如果 config-by-vsm.xml 文件不正确，请查看 vsfwd 日志	/var/log/vsfwd.log	分析该文件以查找错误 要重新启动进程，请运行以下命令： /etc/init.d/vShield-Stateful-Firewall stop start
到控制器的连接是使用 netcpa 建立的	/var/log/netcpa.log	分析该文件以查找错误
VDL2 模块日志位于 vmkernel.log 中	/var/log/vmkernel.log	在 /var/log/vmkernel.log 中检查“具有 vxlan 前缀”的 VDL2 模块日志：

表 2-13. 控制器调试 - 从 NSX Manager 中运行的命令

描述	命令（在 NSX Manager 上）	备注
列出所有控制器及其状态	show controller list all	显示所有控制器及其运行状态的列表

表 2-14. 控制器调试 - 从 NSX Controller 中运行的命令

说明	命令（在控制器上）	备注
检查控制器群集状态	show control-cluster status	应始终显示“Join complete”和“Connected to Cluster Majority”
检查抖动连接的统计信息和消息	show control-cluster core stats	丢弃的数据包计数器不应发生变化
查看与最初加入群集或重新启动后有关的节点活动	show control-cluster history	这对于解决群集加入问题非常有用
查看群集中的节点列表	show control-cluster startup-nodes	请注意，该列表不需要仅包含活动群集节点 它应该是具有当前部署的所有控制器的列表 在启动控制器以联系群集中的其他控制器时，可以使用该列表
显示建立的所有网络连接，类似于 net stat 输出	show network connections of-type tcp	检查要排除故障的主机是否建立到控制器的 netcpa 连接
重新启动控制器进程	restart controller	仅重新启动主控制器进程 强制重新连接到群集
重新引导控制器节点	restart system	重新引导控制器虚拟机

表 2-15. 控制器调试 - NSX Controller 上的日志文件

说明	日志文件	备注
查看控制器历史记录以及最近的加入、重新启动，等等	<code>show control-cluster history</code>	用于解决控制器问题的极佳工具，尤其是解决群集问题
检查速度较慢的磁盘	<code>show log cloudnet/cloudnet_java-zookeeper<timestamp>.log filtered-by fsync</code>	一种检查速度较慢的磁盘的可靠方法是，在 <code>cloudnet_java-zookeeper</code> 日志中查找“fsync”消息 如果同步所需的时间超过 1 秒，ZooKeeper 将输出该消息，这很好地指明了其他程序此时正在使用该磁盘
检查速度较慢/发生故障的磁盘	<code>show log syslog filtered-by collectd</code>	有关“collectd”的示例输出中的消息往往与速度较慢或发生故障的磁盘有关
检查磁盘空间使用率	<code>show log syslog filtered-by freespace:</code>	在空间使用率达到某个阈值时，名为“freespace”的后台作业定期从磁盘中清除旧日志和其他文件。在某些情况下，如果磁盘很小以及/或者填充速度很快，则会看到大量 <code>freespace</code> 消息。这可能表明磁盘已填满
查找当前的活动群集成员	<code>show log syslog filtered-by Active cluster members</code>	列出当前的活动群集成员的节点 ID。可能需要查看较早的 <code>syslog</code> ，因为并非始终输出该消息。
查看核心控制器日志	<code>show log cloudnet/cloudnet_java-zookeeper. 20150703-165223.3702.log</code>	可能具有多个 <code>zookeeper</code> 日志，请查看具有最新时间戳的文件 该文件包含有关选择的控制器群集主控制器的信息以及与控制器的分布式特性有关的其他信息
查看核心控制器日志	<code>show log cloudnet/cloudnet.nsx-controller.root.log.INFO. 20150703-165223.3668</code>	主控制器工作日志，例如，LIF 创建时间、1234 上的连接侦听器、分片

表 2-16. 检查分布式防火墙 - 从 NSX Manager 中运行的命令

说明	NSX Manager 上的命令	备注
查看虚拟机信息	<code>show vm VM-ID</code>	DC、群集、主机、虚拟机名称、vNIC、安装的 <code>dvfilter</code> 等详细信息
查看特定的虚拟网卡信息	<code>show vnic VNIC-ID</code>	VNIC 名称、mac 地址、端口组、应用的筛选器等详细信息
查看所有群集信息	<code>show dfw cluster all</code>	群集名称、群集 ID、数据中心名称、防火墙状态
查看特定的群集信息	<code>show dfw cluster CLUSTER-ID</code>	主机名、主机 ID、安装状态
查看 dfw 相关主机信息	<code>show dfw host HOST-ID</code>	虚拟机名称、虚拟机 ID、电源状态
查看 dvfilter 中的详细信息	<code>show dfw host HOST-ID filter filterID <option></code>	列出每个 VNIC 的规则、统计信息、地址集等
查看虚拟机的 DFW 信息	<code>show dfw vm VM-ID</code>	查看虚拟机的名称、VNIC ID、筛选器等
查看 VNIC 详细信息	<code>show dfw vnic VNIC-ID</code>	查看 VNIC 名称、ID、MAC 地址、端口组、筛选器

表 2-16. 检查分布式防火墙 - 从 NSX Manager 中运行的命令（续）

说明	NSX Manager 上的命令	备注
列出为每个 vNIC 安装的筛选器	<code>show dfw host hostID summarize-dvfilter</code>	查找感兴趣的虚拟机/vNIC，并获取名称字段以在后续命令中作为筛选器
查看特定筛选器/vNIC 的规则	<code>show dfw host hostID filter filterID rules</code> <code>show dfw vnic nicID</code>	
查看地址集的详细信息	<code>show dfw host hostID filter filterID addrsets</code>	这些规则仅显示地址集，可以使用该命令扩充地址集包含的内容
每个 vNIC 的 spoofguard 详细信息	<code>show dfw host hostID filter filterID spoofguard</code>	检查是否启用了 spoofguard 以及当前的 IP/MAC
查看流量记录详细信息	<code>show dfw host hostID filter filterID flows</code>	如果启用了流量监控，主机定期将流量信息发送到 NSX Manager 可以使用该命令查看每个 vNIC 的流量
查看 vNIC 的每个规则的统计信息	<code>show dfw host hostID filter filterID stats</code>	在查看是否命中规则时，这是非常有用的

表 2-17. 检查分布式防火墙 - 从主机中运行的命令

说明	主机上的命令	备注
列出在主机上下载的 VIB	<code>esxcli software vib list grep vsip</code>	检查以确保下载了正确的 VIB 版本
有关当前加载的系统模块的详细信息	<code>esxcli system module get -m vsip</code>	检查以确保安装/加载了模块
进程列表	<code>ps grep vsfwd</code>	查看是否使用多个线程运行 vsfwd 进程
守护程序命令	<code>/etc/init.d/vShield-Stateful-Firewall {start stop status restart}</code>	检查是否正在运行守护程序，并根据需要重新启动
查看网络连接	<code>esxcli network ip connection list grep 5671</code>	检查主机是否具有到 NSX Manager 的 TCP 连接

表 2-18. 检查分布式防火墙 - 主机上的日志文件

说明	日志	备注
进程日志	<code>/var/log/vsfwd.log</code>	vsfwd 守护程序日志，对 vsfwd 进程、NSX Manager 连接和 RabbitMQ 故障排除非常有用
数据包日志专用文件	<code>/var/log/dfwpktlogs.log</code>	数据包日志的专用日志文件
dvfilter 中的数据包捕获	<code>pktcap-uw --dvfilter nic-1413082-eth0-vmware-sfw.2 --outfile test.pcap</code>	

跟踪流

跟踪流是一个故障排除工具，能够插入数据包并在数据包通过物理网络和逻辑网络时观察该数据包出现的位置。通过观察，可以确定网络的相关信息，例如识别已关闭的节点或阻止数据包被目标接收的防火墙规则。

本章讨论了以下主题：

- [关于跟踪流](#)
- [使用跟踪流进行故障排除](#)

关于跟踪流

当数据包通过覆盖网络和底层网络中的物理及逻辑实体（如 ESXi 主机、逻辑交换机和逻辑路由器）时，跟踪流会将数据包注入到 **vSphere Distributed Switch (VDS)** 端口，并沿着数据包的路径提供各个观察点。这样，您就可以标识数据包到达其目标所经过的一个或多个路径，或者反过来，标识数据包沿着哪个路径时被丢弃。每个实体都会报告输入和输出上的数据包处理，因此您可以确定接收数据包或转发数据包时是否出现问题。

请记住，跟踪流与在客户机虚拟机堆栈之间传送的 **ping** 请求/响应不同。跟踪流执行的操作是在标记的数据包通过覆盖网络时观察这些数据包。在每个数据包通过覆盖网络时，将监视该数据包，直至到达并可传输到目标客户机虚拟机。不过，绝不会将注入的跟踪流数据包实际传输到目标客户机虚拟机。这意味着即使客户机虚拟机电源关闭，跟踪流仍能够成功。

跟踪流支持以下流量类型：

- 第 2 层单播
- 第 3 层单播
- 第 2 层广播
- 第 2 层多播

您可以使用自定义标头字段和数据包大小构造数据包。跟踪流的源始终为虚拟机的虚拟网卡 (vNIC)。目标端点可以是 **NSX** 覆盖或底层中的任何设备。不过，您无法选择位于 **NSX Edge** 服务网关 (ESG) 北部的目标。目标必须位于相同的子网上，或者必须能够通过 **NSX** 分布式逻辑路由器访问目标。

如果源和目标虚拟网卡位于同一第 2 层域中，跟踪流操作将被视为第 2 层。在 **NSX** 中，这意味着它们位于同一 **VXLAN** 网络标识符 (VNI 或分段 ID) 上。例如，当两个虚拟机连接到同一逻辑交换机时，会发生这种情况。

如果配置了 NSX 桥接，未知第 2 层数据包将始终发送到网桥。通常，网桥会将这些数据包转发到 VLAN 并将跟踪流数据包报告为“已传送”。报告为“已递送”的数据包不一定表示跟踪数据包已传送到指定的目标。

对于第 3 层跟踪流单播流量，两个端点位于不同的逻辑交换机上，且具有连接到分布式逻辑路由器 (DLR) 的不同 VNI。

对于多播流量，源为虚拟机的虚拟网卡，目标为多播组地址。

跟踪流观察可能包括广播的跟踪流数据包观察。如果不知道目标主机的 MAC 地址，ESXi 主机将广播跟踪流数据包。对于广播流量，源为虚拟机的虚拟网卡。广播流量的第 2 层目标 MAC 地址为 FF:FF:FF:FF:FF:FF。要为防火墙检测创建有效的数据包，广播跟踪流操作需要子网前缀长度。子网掩码使 NSX 可以计算数据包的 IP 网络地址。



小心 根据部署中逻辑端口的数量，多播和广播跟踪流操作可能会生成大量流量。

使用跟踪流的方式有两种：通过 API 和通过 GUI。API 即为 GUI 使用的 API，但是 API 允许您在数据包中指定确切设置，而 GUI 具有的设置更有限。

GUI 允许您设置以下值：

- 协议---TCP、UDP、ICMP。
- 活动时间 (TTL)。默认值为 64 个跃点。
- TCP 和 UDP 源及目标端口号。默认值为 0。
- TCP 标记。
- ICMP ID 和序列号。两者的默认值均为 0。
- 跟踪流操作的过期超时，以毫秒 (ms) 为单位。默认值为 10,000 ms。
- 以太网帧大小。默认值为每帧 128 字节。最大帧大小为每帧 1000 字节。
- 负载编码。默认值为 Base64。
- 负载值。

使用跟踪流进行故障排除

跟踪流在多种场景中非常有用。

跟踪流在以下场景下有用：

- 排除网络故障以查看流量流经的准确路径
- 监控性能以了解链接使用率
- 规划网络以了解网络在生产环境中的运行方式

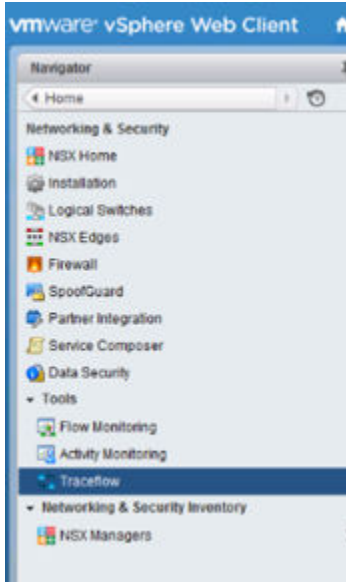
前提条件

- 跟踪流操作要求 vCenter、NSX Manager、NSX Controller 群集以及主机上的 netcpa 用户环境代理之间能够通信。

- 要使跟踪流按预期工作，请确保控制器群集已连接且处于正常状态。

步骤

- 1 在 vCenter Web Client 中，导航到主页 > 网络和安全 > 跟踪流 (Home > Networking & Security > Traceflow)。



- 2 选择流量类型：单播、广播或多播。
- 3 选择源虚拟机的虚拟网卡。

如果虚拟机托管于运行跟踪流的 vCenter Server 中，则可以从列表中选择虚拟机和虚拟网卡。

- 4 对于单播跟踪流，请输入目标虚拟网卡的信息。

目标可以是 NSX 覆盖网络或底层网络中任意设备的虚拟网卡，例如主机、虚拟机、逻辑路由器或 Edge 服务网关。如果目标是运行 VMware Tools 的虚拟机，并且该虚拟机在从中运行跟踪流的同一 vCenter Server 中进行管理，则可以从列表中选择虚拟机和虚拟网卡。

否则，必须输入目标 IP 地址（对于单播第 2 层跟踪流，还需输入 MAC 地址）。可以在设备控制台或 SSH 会话中从设备自身收集此信息。例如，如果目标是 Linux 虚拟机，则可以通过在 Linux 终端中运行 `ifconfig` 命令来获取其 IP 和 MAC 地址。对于逻辑路由器或 Edge 服务网关，可以通过 `show interface` CLI 命令收集信息。

- 5 对于第 2 层广播跟踪流，请输入子网前缀的长度。

数据包仅基于 MAC 地址进行交换。目标 MAC 地址为 FF:FF:FF:FF:FF:FF。

需要提供源和目标 IP 地址，才能使 IP 数据包对防火墙监测有效。

- 6 对于第 2 层多播跟踪流，请输入多播组地址。

数据包仅基于 MAC 地址进行交换。

需要提供源和目标 IP 地址，才能使 IP 数据包有效。在多播情况下，MAC 地址是根据 IP 地址推导出来的。

7 配置其他必选和可选设置。

8 单击跟踪 (Trace)。

示例：场景

下例展示的第 2 层跟踪流涉及在同一台 ESXi 主机上运行的两个虚拟机。这两个虚拟机连接到同一个逻辑交换机。

Traceflow

NSX Manager: 192.168.110.15 (Role: Primary)

Trace Parameters

Traffic Type: Unicast

Source: * web-01a - Network adapter 1 Change...
IP: 172.16.10.11, MAC: 00:50:56:ae:3e:3d

Destination: * web-02a - Network adapter 1 Change...
IP: 172.16.10.12, MAC: 00:50:56:ae:f8:6b

Advanced Options

Protocol: TCP

Source Port: 0

Destination Port: 0

TCP Flags: ☐ FIN ☒ SYN ☐ RST

Timeout (ms): 10000

Frame Size: 128

TTL: 64

Trace

Trace Result: Traceflow delivered observation(s) reported

1 Delivered

Sequence	Observation Type	Host	Component Type	Component Name
0	Injected	esx-01a.corp.local	vNIC	vNIC
1	Received	esx-01a.corp.local	Firewall	Firewall
2	Forwarded	esx-01a.corp.local	Firewall	Firewall
3	Received	esx-01a.corp.local	Firewall	Firewall
4	Forwarded	esx-01a.corp.local	Firewall	Firewall
5	Delivered	esx-01a.corp.local	vNIC	vNIC

下例展示的第 2 层跟踪流涉及分别在两台不同的 ESXi 主机上运行的两个虚拟机。这两个虚拟机连接到同一个逻辑交换机。

Traceflow

NSX Manager: 192.168.110.15 (Role: Primary)

Trace Parameters

Traffic Type: Unicast

Source: * web-01a - Network adapter 1 Change...
IP: 172.16.10.11, MAC: 00:50:56:ae:3e:3d

Destination: * web-03a - Network adapter 1 Change...
IP: 172.17.10.11, MAC: 00:50:56:ae:cf:88

Advanced Options

Protocol: TCP

Source Port: 0

Destination Port: 0

TCP Flags: ☐ FIN ☒ SYN ☐ RST

Timeout (ms): 10000

Frame Size: 128

TTL: 64

Trace

Trace Result: Traceflow delivered observation(s) reported

1 Delivered

Sequence	Observation Type	Host	Component Type	Component Name
0	Injected	esx-01a.corp.local	vNIC	vNIC
1	Received	esx-01a.corp.local	Firewall	Firewall
2	Forwarded	esx-01a.corp.local	Firewall	Firewall
3	Forwarded	esx-01a.corp.local	Physical	esx-01a.corp.local
3	Forwarded	esx-01a.corp.local	Physical	esx-01a.corp.local
4	Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4	Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4	Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4	Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4	Received	esx-02a.corp.local	Physical	esx-02a.corp.local
4	Received	esx-02a.corp.local	Physical	esx-02a.corp.local
5	Received	esx-02a.corp.local	Firewall	Firewall
6	Forwarded	esx-02a.corp.local	Firewall	Firewall
7	Delivered	esx-02a.corp.local	vNIC	vNIC

下例展示了一个第 3 层跟踪流。其中的两个虚拟机分别连接到由逻辑路由器分隔的两个不同逻辑交换机。

Traceflow

NSX Manager: 192.168.110.15 (Role: Primary)

Trace Parameters

Traffic Type: Unicast

Source: * web-01a - Network adapter 1 Change...
IP: 172.16.10.11, MAC: 00:50:56:ae:3e:3d

Destination: * db-01a - Network adapter 1 Change...
IP: 172.16.30.11, MAC: 00:50:56:ae:d4:2b

► Advanced Options

Trace

Trace Result: Traceflow delivered observation(s) reported

1 Delivered


Sequence	1 ▲	Observation Type	Host	Component Type	Component Name
0		Injected	esx-01a.corp.local	vNIC	vNIC
1		Received	esx-01a.corp.local	Firewall	Firewall
2		Forwarded	esx-01a.corp.local	Firewall	Firewall
3		Forwarded	esx-01a.corp.local	Logical Switch	Web-Tier-01
4		Received	esx-01a.corp.local	Logical Router	Local-Distributed-Router
5		Forwarded	esx-01a.corp.local	Logical Router	Local-Distributed-Router
6		Received	esx-01a.corp.local	Logical Switch	DB-Tier-01
7		Forwarded	esx-01a.corp.local	Physical	esx-01a.corp.local
8		Received	esx-02a.corp.local	Physical	esx-02a.corp.local
8		Received	esx-02a.corp.local	Physical	esx-02a.corp.local
9		Received	esx-02a.corp.local	Firewall	Firewall
10		Forwarded	esx-02a.corp.local	Firewall	Firewall
11		Delivered	esx-02a.corp.local	vNIC	vNIC


下例展示了三个虚拟机连接到同一个逻辑交换机的部署中的广播跟踪流。其中的两个虚拟机位于一个主机 (esx-01a) 上，第三个虚拟机位于另一个主机 (esx-02a) 上。广播是从主机 192.168.210.53 上的其中一个虚拟机发送的。

Traceflow

NSX Manager: 192.168.110.15 (Role: Primary)

Trace Parameters

Traffic Type: L2 Broadcast  High volume of traffic may get generated for this traffic type.

Source:  web-01a - Network adapter 1 [Change...](#) Subnet Prefix Length: * 24











































IP: 172.16.10.11, MAC: 00:50:56:ae:3e:3d IP: 172.16.10.255, MAC: FF:FF:FF:FF:FF:FF

► Advanced Options

[Trace](#)

Trace Result: Traceflow delivered observation(s) reported

3 Delivered

Sequence	1 ▲	Observation Type	Host	Component Type	Component Name
0		Injected	esx-01a.corp.local	vNIC	 vNIC
1		Received	esx-01a.corp.local	Firewall	 Firewall
2		Forwarded	esx-01a.corp.local	Firewall	 Firewall
3		Forwarded	esx-01a.corp.local	Logical Switch	 Web-Tier-01
3		Received	esx-01a.corp.local	Firewall	 Firewall
3		Forwarded	esx-01a.corp.local	Physical	esx-01a.corp.local
3		Forwarded	esx-01a.corp.local	Physical	esx-01a.corp.local
4		Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4		Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4		Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4		Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4		Forwarded	esx-01a.corp.local	Firewall	 Firewall
4		Received	esx-02a.corp.local	Physical	esx-02a.corp.local
4		Received	esx-02a.corp.local	Physical	esx-02a.corp.local
5		Forwarded	esxmgt-02a.corp.local	Logical Switch	 Web-Tier-01
5		Forwarded	esxmgt-02a.corp.local	Logical Switch	 Web-Tier-01
5		Forwarded	esxmgt-02a.corp.local	Logical Switch	 Web-Tier-01
5		Forwarded	esxmgt-02a.corp.local	Logical Switch	 Web-Tier-01
5		Delivered	esxmgt-02a.corp.local	vNIC	 vNIC
5		Delivered	esx-01a.corp.local	vNIC	 vNIC
5		Forwarded	esx-02a.corp.local	Logical Switch	 Web-Tier-01
5		Forwarded	esx-02a.corp.local	Logical Switch	 Web-Tier-01
5		Received	esx-02a.corp.local	Firewall	 Firewall
6		Forwarded	esx-02a.corp.local	Firewall	 Firewall
7		Delivered	esx-02a.corp.local	vNIC	 vNIC

下例显示了在配置多播的部署中发送多播流量时会出现的情况。

Traceflow

NSX Manager: 192.168.110.15 (Role: Primary)

Trace Parameters

Traffic Type: L2 Multicast ⚠ High volume of traffic may get generated for this traffic type.

Source: web-01a - Network adapter 1 [Change...](#)
IP: 172.16.10.11, MAC: 00:50:56:ae:3e:3d

Destination IP: 239.0.0.1 [Change...](#) e.g. 239.0.0.1
IP: 239.0.0.1, MAC: 01:00:5e:00:00:01

▶ Advanced Options

[Trace](#)

Trace Result: Traceflow delivered observation(s) reported

3 Delivered

Sequence	1 ▲	Observation Type	Host	Component Type	Component Name
0		Injected	esx-01a.corp.local	vNIC	vNIC
1		Received	esx-01a.corp.local	Firewall	Firewall
2		Forwarded	esx-01a.corp.local	Firewall	Firewall
3		Received	esx-01a.corp.local	Firewall	Firewall
3		Forwarded	esx-01a.corp.local	Physical	esx-01a.corp.local
3		Forwarded	esx-01a.corp.local	Physical	esx-01a.corp.local
4		Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4		Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4		Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4		Received	esxmgt-02a.corp.local	Physical	esxmgt-02a.corp.local
4		Forwarded	esx-01a.corp.local	Firewall	Firewall
4		Received	esx-02a.corp.local	Physical	esx-02a.corp.local
4		Received	esx-02a.corp.local	Physical	esx-02a.corp.local
5		Delivered	esxmgt-02a.corp.local	vNIC	vNIC
5		Delivered	esx-01a.corp.local	vNIC	vNIC
5		Received	esx-02a.corp.local	Firewall	Firewall
6		Forwarded	esx-02a.corp.local	Firewall	Firewall
7		Delivered	esx-02a.corp.local	vNIC	vNIC

下例显示了因应用阻止将 ICMP 流量发送到目标地址的分布式防火墙规则而丢弃跟踪流时出现的情况。请注意，流量永远不会离开原始主机，即使目标虚拟机位于另一个主机上也是如此。

Traceflow

NSX Manager: 192.168.110.15 (Role: Primary)

Trace Parameters

Traffic Type: Unicast

Source: web-02a - Network adapter 1 [Change...](#)
IP: 172.16.10.12, MAC: 00:50:56:ae:f8:6b

Destination: web-03a - Network adapter 1 [Change...](#)
IP: 172.17.10.11, MAC: 00:50:56:ae:cf:88

▶ Advanced Options

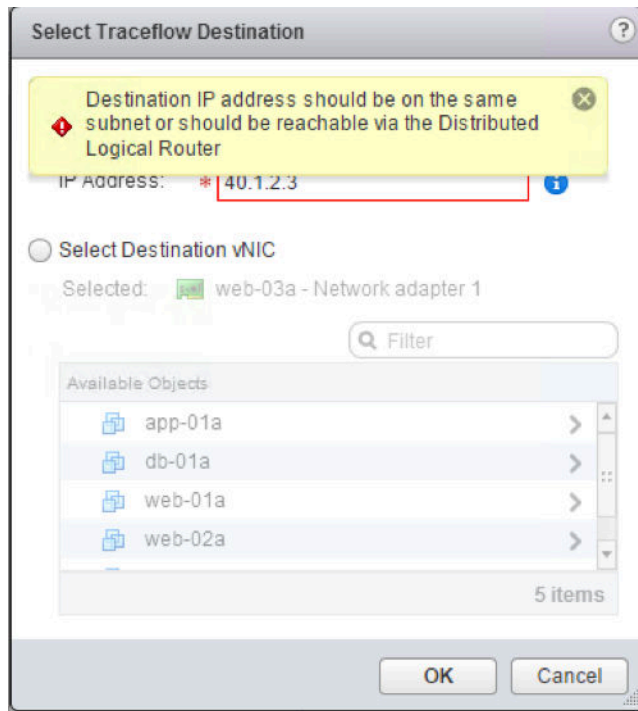
[Trace](#)

Trace Result: Traceflow dropped observation(s) reported

1 Dropped

Sequence	1 ▲	Observation Type	Host	Component Type	Component Name
0		Injected	esx-01a.corp.local	vNIC	vNIC
1		Received	esx-01a.corp.local	Firewall	Firewall
2		Dropped	esx-01a.corp.local	Firewall	Firewall (Rule - 1013)

下例显示了跟踪流目标位于 **Edge** 服务网关的另一端时出现的情况，例如 **Internet** 或必须通过 **Edge** 服务网关路由的任何内部目标上的 **IP** 地址。从设计上不允许使用跟踪流，因为仅位于同一子网上或可通过分布式逻辑路由器 (DLR) 访问的目标支持跟踪流。



下例显示了跟踪流目标是位于另一个子网中已关闭电源的虚拟机时出现的情况。

Traceflow

NSX Manager: 192.168.110.15 (Role: Primary)

Trace Parameters

Traffic Type: Unicast

Source: * app-01a - Network adapter 1 Change...
IP: 172.16.20.11, MAC: 00:50:56:ae:23:b9

Destination: * db-01a - Network adapter 1 Change...
IP: 172.16.30.11, MAC: 00:50:56:ae:d...

Advanced Options

Trace

Trace Result: No delivered or dropped observations reported

Sequence	Observation Type	Host	Component Type	Component Name
0	Injected	esx-02a.corp.local	vNIC	vNIC
1	Received	esx-02a.corp.local	Firewall	Firewall
2	Forwarded	esx-02a.corp.local	Firewall	Firewall
3	Forwarded	esx-02a.corp.local	Logical Switch	App-Tier-01
4	Received	esx-02a.corp.local	Logical Router	Local-Distributed-Router
5	Forwarded	esx-02a.corp.local	Logical Router	Local-Distributed-Router
6	Received	esx-02a.corp.local	Logical Switch	DB-Tier-01

NSX 路由

NSX 具有两种类型的路由子系统，它们针对两种主要需求进行了优化。

NSX 路由子系统是：

- 逻辑空间中的路由；也称为“东-西”路由，这是由分布式逻辑路由器 (DLR) 提供的；
- 物理和逻辑空间之间的路由；也称为“北-南”路由，这是由 Edge 服务网关 (ESG) 提供的。

它们都提供了水平扩展选项。

您可以通过 DLR 横向扩展分布式东-西路由。

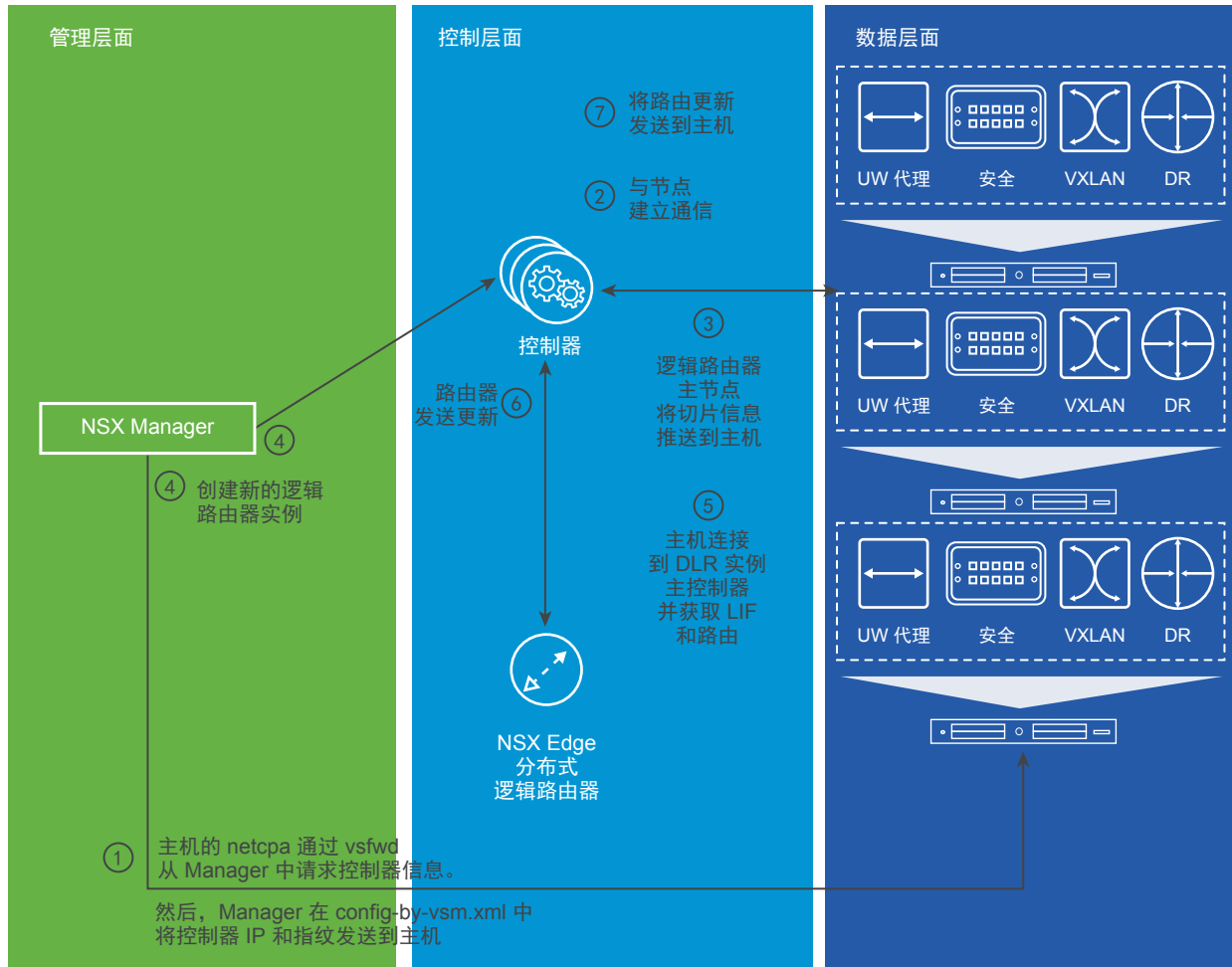
DLR 支持每次运行单个动态路由协议（OSPF 或 BGP）；而 ESG 支持同时运行两个路由协议。这样做的原因是，DLR 设计为“末端”路由器并具有单个输出路径，这意味着通常不需要使用更高级的路由配置。

DLR 和 ESG 均支持静态路由和动态路由组合。

DLR 和 ESG 均支持 ECMP 路由。

它们提供了 L3 域隔离，这意味着，每个分布式逻辑路由器或 Edge 服务网关实例具有自己的 L3 配置，类似于 L3VPN VRF。

图 4-1. 创建 DLR



本章讨论了以下主题：

- 了解分布式逻辑路由器
- 了解 Edge 服务网关提供的路由
- ECMP 数据包流
- NSX 路由：必备条件和注意事项
- DLR 和 ESG UI
- 新的 NSX Edge (DLR)
- 典型的 ESG 和 DLR UI 操作
- NSX 路由故障排除

了解分布式逻辑路由器

DLR 经过优化以在支持 VXLAN 或 VLAN 的端口组上的虚拟机之间的逻辑空间中转发。

DLR 具有以下属性：

- 高性能、低开销的第一跃点路由：
- 随主机数呈线性扩展
- 在上行链路上支持 8 向 ECMP
- 每个主机最多 1,000 个 DLR 实例
- 每个 DLR 上最多 999 个逻辑接口 (LIF) (8 个上行链路 + 991 个内部) + 1 个管理
- 在所有 DLR 实例中分布的每个主机最多 10,000 个 LIF (NSX Manager 未强制实施)

请注意以下问题：

- 无法将多个 DLR 连接到任何给定的 VLAN 或 VXLAN。
- 无法在每个 DLR 上运行多个路由协议。
- 如果使用 OSPF，则无法在多个 DLR 上行链路上运行 OSPF。
- 要在 VXLAN 和 VLAN 之间路由，传输区域必须跨单个 DVS。

概括来说，DLR 设计在以下方面与模块化路由器机箱类似：

- ESXi 主机类似于线路卡：
 - 它们具有连接了终端站（虚拟机）的端口。
 - 这是进行转发决策的位置。
- DLR 控制虚拟机类似于路由处理器引擎：
 - 它运行动态路由协议，以便与网络的其余部分交换路由信息。
 - 它根据接口配置、静态路由和动态路由信息计算“线路卡”的转发表。
 - 它将这些转发表写入到“线路卡”（通过控制器群集）以支持扩展和弹性。
- 将 ESXi 主机连接在一起的物理网络类似于背板：
 - 它在“线路卡”之间传送 VLAN 或 VXLAN 封装的数据。

简要 DLR 数据包流

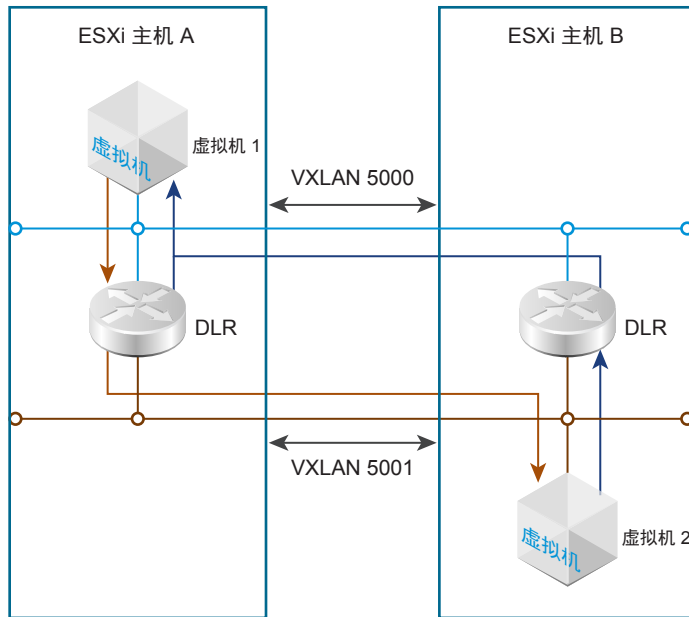
对于每个配置的 DLR 实例，每个 ESXi 主机具有自己的副本。每个 DLR 实例具有自己的一组独特的表，其中包含转发数据包所需的信息。将在该 DLR 实例所在的所有主机之间同步该信息。不同主机上的单个 DLR 实例具有完全相同的信息。

路由始终是由运行源虚拟机的相同主机上的 DLR 实例处理的。这意味着，在源和目标虚拟机位于不同的主机上时，在它们之间提供路由的 DLR 实例仅在一个方向（从源虚拟机到目标虚拟机）上看到数据包。仅目标虚拟机的相同 DLR 的相应实例看到返回流量。

如果源和目标虚拟机位于不同的主机上，在 DLR 完成路由后，DVS 负责通过 L2（VXLAN 或 VLAN）传送到最终目标；如果源和目标虚拟机位于相同的主机上，则 DVS 在本地进行传送。

图 4-2 说明了在不同主机上运行并连接到两个不同逻辑交换机（VXLAN 5000 和 VXLAN 5001）的两个虚拟机（虚拟机 1 和虚拟机 2）之间的数据流。

图 4-2. 简要 DLR 数据包流



数据包流（跳过 ARP 解析）：

- 1 虚拟机 1 向虚拟机 2 发送一个数据包，它将发送到虚拟机 2 子网的虚拟机 1 网关（或默认位置）。该网关是 DLR 上的 VXLAN 5000 LIF。
- 2 ESXi 主机 A 上的 DVS 将数据包传送到该主机上的 DLR，将在其中执行查找并确定输出 LIF（此处为 VXLAN 5001 LIF）。
- 3 然后，从该目标 LIF 中发出数据包，这实际上将数据包返回到 DVS，但位于不同的逻辑交换机 (5001) 上。
- 4 接下来，DVS 执行 L2 传送以将该数据包传送到目标主机（ESXi 主机 B），DVS 在其中将数据包转发到虚拟机 2。

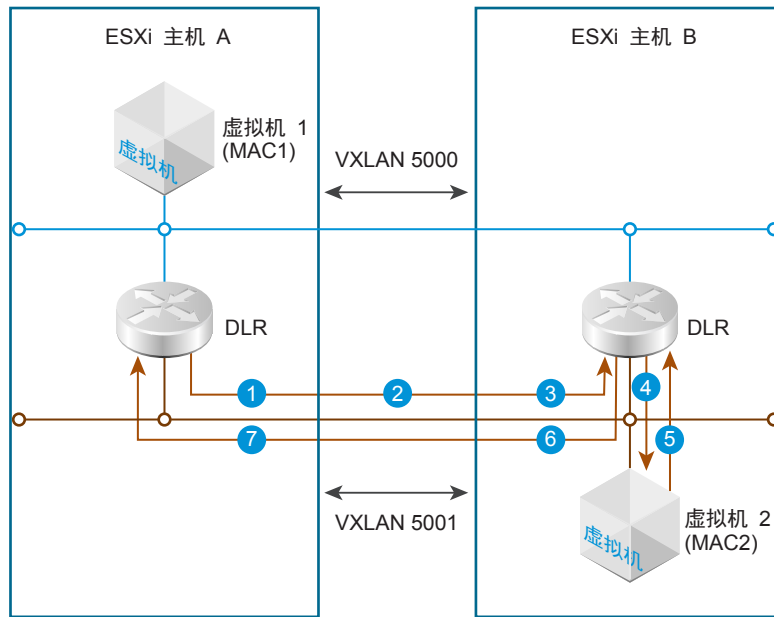
返回流量将采用相同的顺序，来自虚拟机 2 的流量转发到 ESXi 主机 B 上的 DLR 实例，然后通过 VXLAN 5000 上的 L2 进行传送。

DLR ARP 解析过程

在虚拟机 1 中的流量到达虚拟机 2 之前，DLR 需要获悉虚拟机 2 的 MAC 地址。在获悉虚拟机 2 的 MAC 地址后，DLR 可以为出站数据包创建正确的 L2 标头。

图 4-3 显示了 DLR 的 ARP 解析过程。

图 4-3. DLR ARP 过程



要获悉 MAC 地址，DLR 将执行以下步骤：

- 1 主机 A 上的 DLR 实例生成一个 ARP 请求数据包（源 MAC = vMAC，目标 MAC = 广播）。主机 A 上的 VXLAN 模块在输出 VXLAN 5001 上查找所有 VTEP，然后为每个 VTEP 发送该广播帧的一个副本。
- 2 在该帧通过 VXLAN 封装过程离开主机时，源 MAC 将从 vMAC 更改为 pMAC A，以便返回流量可以在主机 A 上找到源 DLR 实例。该帧现在为源 MAC = pMAC A，目标 MAC = 广播。
- 3 在主机 B 上收到并解封该帧时，将检查该帧并发现它来自于与 VXLAN 5001 上的本地 DLR 实例的 LIF 匹配的 IP 地址。这会将该帧标记为 **abrequest** 以执行代理 ARP 功能。目标 MAC 将从广播更改为 vMAC，以便该帧可以到达本地 DLR 实例。
- 4 主机 B 上的本地 DLR 实例收到 ARP 请求帧（源 MAC = pMAC A，目标 MAC = vMAC），然后查看自己的 LIF IP 地址以请求该信息。它保存源 MAC 并生成新的 ARP 请求数据包（源 MAC = vMAC，目标 MAC = 广播）。该帧将标记为“DVS 本地”，以防止它通过 dvUplink 发生洪泛。DVS 将该帧传送到虚拟机 2。
- 5 虚拟机 2 发送一个 ARP 回复（源 MAC = MAC2，目标 MAC = vMAC）。DVS 将其传送到本地 DLR 实例。
- 6 主机 B 上的 DLR 实例将目标 MAC 替换为在步骤 4 中保存的 pMAC A，然后将数据包发送回 DVS 以传送回主机 A。
- 7 在 ARP 回复到达主机 A 后，目标 MAC 将更改为 vMAC，并且 ARP 回复帧（源 MAC = MAC2，目标 MAC = vMAC）到达主机 A 上的 DLR 实例。

ARP 解析过程已完成，主机 A 上的 DLR 实例现在可以开始将流量发送到虚拟机 2。

了解 Edge 服务网关提供的路由

NSX 路由的第二个子系统是由 Edge 服务网关提供的。

ESG 实际上是虚拟机中的路由器。其外形尺寸与设备类似并具有四个尺寸，并由 NSX Manager 管理其整个生命周期。ESG 的主要用途是作为外围路由器，它部署在多个 DLR 之间以及物理环境和虚拟网络之间。

ESG 具有以下属性：

- 每个 ESG 最多可以具有 10 个 vNIC 接口或 200 个中继子接口。
- 每个 ESG 支持 8 向 ECMP 以提供路径冗余和可扩展性。

ECMP 数据包流

假设部署了两个 ESG，以便在物理环境中提供具有 2 向 ECMP 上行链路的 DLR 实例。

图 4-4 显示了在两个 ESG 和物理基础架构之间启用等价多路径 (ECMP) 路由时的 ESG 和 DLR 数据包流。

因此，与具有单个 ESG 的部署相比，虚拟机 1 可以获得 2 倍的双向吞吐量。

VM1 连接到具有 VNI 5000 的逻辑交换机。

DLR 具有两个 LIF - VNI 5000 上的内部 LIF 以及 VNI 5001 上的上行链路 LIF。

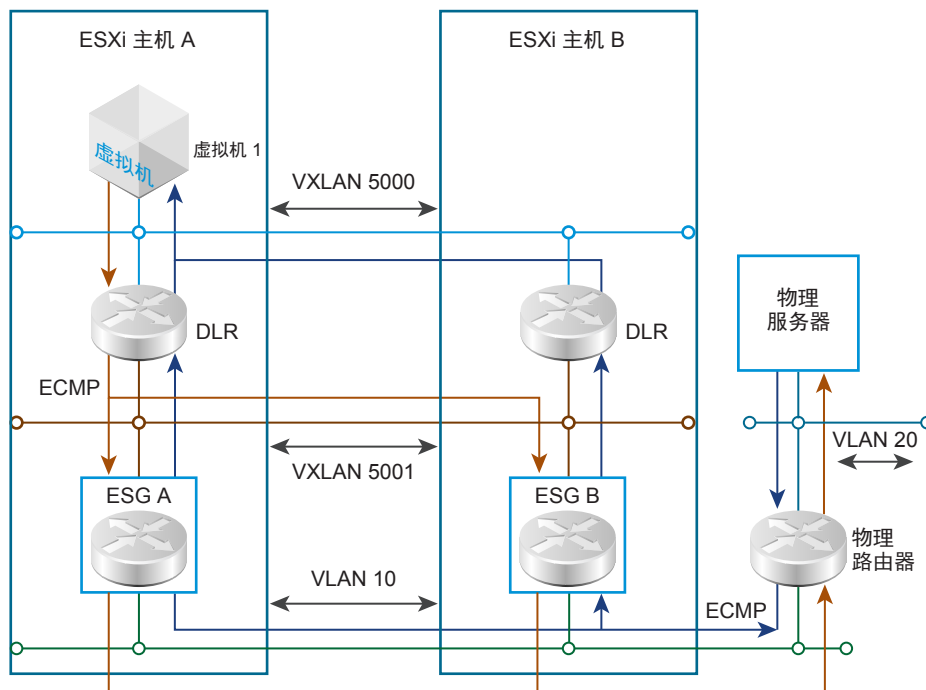
DLR 启用了 ECMP，并通过动态路由协议（BGP 或 OSPF）从一对 ESG（ESG A 和 ESG B）中接收到 VLAN 20 的 IP 子网的等价路由。

两个 ESG 连接到与 VLAN 10 关联且支持 VLAN 的 dvPortgroup，还会在其中连接提供到 VLAN 20 的连接物理路由器。

ESG 通过动态路由协议从物理路由器中接收 VLAN 20 的外部路由。

进行交换的物理路由器从两个 ESG 中获悉与 VXLAN 5000 关联的 IP 子网，并对传输到该子网中的虚拟机的流量执行 ECMP 负载平衡。

图 4-4. 具有 ECMP 的简要 ESG 和 DLR 数据包流



DLR 可以接收最多 8 个等价路由并在这些路由之间平衡流量。图中的 ESG A 和 ESG B 提供了两个等价路由。ESG 可以执行到物理网络的 ECMP 路由，假设存在多个物理路由器。为简单起见，该图显示单个物理路由器。

不需要在 ESG 上配置到 DLR 的 ECMP，因为所有 DLR LIF 位于 ESG 所在的同一主机“本地”。在 DLR 上配置多个上行链路接口并不会带来额外的好处。

在需要更多北-南带宽的情况下，可以将多个 ESG 放在不同的 ESXi 主机上以通过 8 个 ESG 纵向扩展到大约 80Gbps。

ECMP 数据包流（不包括 ARP 解析）：

- 1 虚拟机 1 将数据包发送到物理服务器，数据包将发送到 ESXi 主机 A 上的虚拟机 1 IP 网关（它是 DLR LIF）。
- 2 DLR 为物理服务器的 IP 执行路由查找，并发现它不是直接连接的，而是与从 ESG A 和 ESG B 中收到的两个 ECMP 路由相匹配。
- 3 DLR 计算 ECMP 哈希，确定下一跃点（可能是 ESG A 或 ESG B），然后将数据包从 VXLAN 5001 LIF 中发出。
- 4 DVS 将数据包传送到选定的 ESG。
- 5 ESG 执行路由查找，并发现可以通过 VLAN 10 上的物理路由器 IP 地址访问物理服务器的子网，它直接连接到 ESG 的某个接口。
- 6 数据包是通过 DVS 发出的，在使用 VLAN ID 10 的相应 801.Q 标记标记后，DVS 将数据包传送到物理网络。
- 7 数据包穿过物理交换基础架构以到达物理路由器，物理路由器执行查找以发现物理服务器直接连接到 VLAN 20 上的接口。
- 8 物理路由器将数据包发送到物理服务器。

在相反方向上：

- 1 物理服务器将数据包发送到虚拟机 1，并将物理路由器作为下一跃点。
- 2 物理路由器为虚拟机 1 的子网执行查找，并发现到该子网的两个等价路径，下一跃点分别为 ESG A 和 ESG B 的 VLAN 10 接口。
- 3 物理路由器选择其中的一个路径，然后将数据包发送到相应的 ESG。
- 4 物理网络将数据包传送到 ESG 所在的 ESXi 主机，然后将其传送到 DVS，DVS 解封数据包并在与 VLAN 10 关联的 dvPortgroup 上将其转发到 ESG。
- 5 ESG 执行路由查找，并发现可以通过与 VXLAN 5001 关联的接口访问虚拟机 1 的子网，下一跃点是 DLR 的上行链路接口 IP 地址。
- 6 ESG 将数据包发送到与 ESG 相同的主机上的 DLR 实例。
- 7 DLR 执行路由查找，并发现可以通过 VXLAN 5000 LIF 访问虚拟机 1。
- 8 DLR 将数据包从其 VXLAN 5000 LIF 发送到 DVS，DVS 执行最终传送。

NSX 路由：必备条件和注意事项

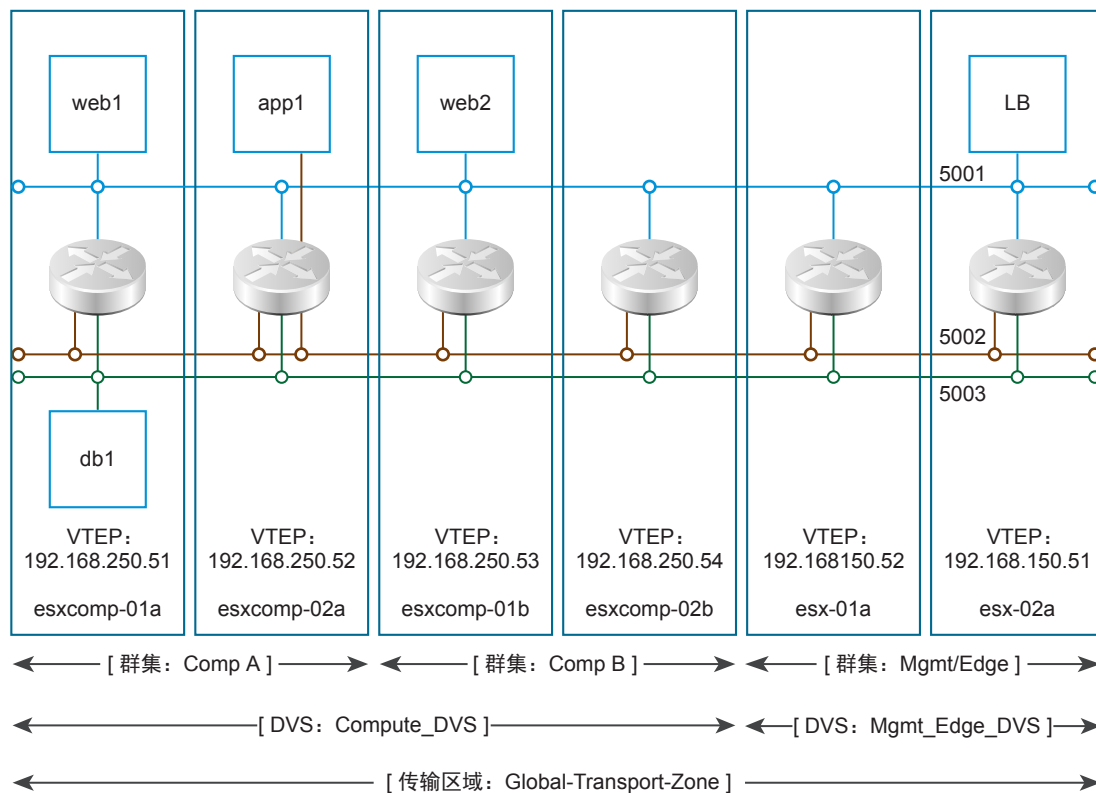
DLR 和 ESG 依靠 DVS 为 dvPortgroup 提供 L2 转发服务（基于 VXLAN 和 VLAN）以使端到端连接正常工作。

这意味着，必须配置连接到 DLR 或 ESG 的 L2 转发服务并且正常运行。在 NSX 安装过程中，“主机准备”和“逻辑网络准备”提供了这些服务。

在多群集 DVS 配置中创建传输区域时，请确保选定的 DVS 中的所有群集包含在传输区域中。这可确保 DLR 在提供了 DVS dvPortgroup 的所有群集上可用。

在传输区域与 DVS 边界对齐时，将正确创建 DLR 实例。

图 4-5. 传输区域与 DVS 边界正确对齐



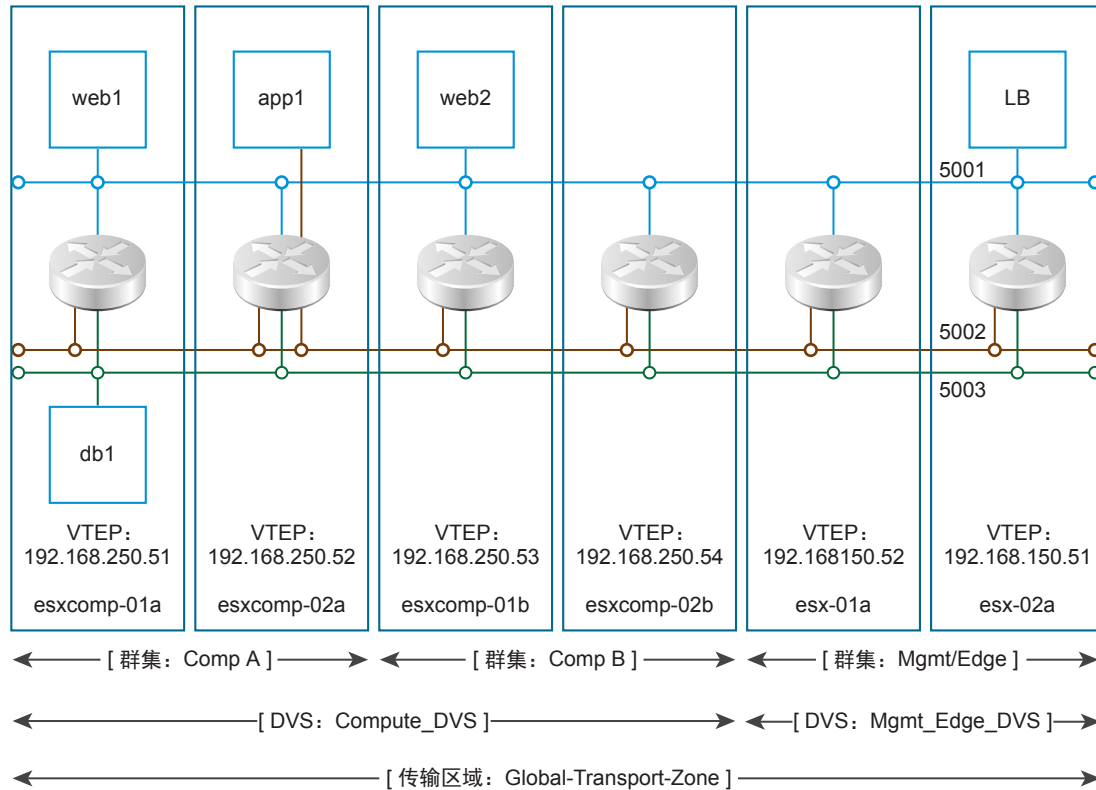
在传输区域与 DVS 边界不对应时，逻辑交换机（5001、5002 和 5003）的范围和这些逻辑交换机连接到的 DLR 实例将断开连接，从而导致 Comp A 群集中的虚拟机无法访问 DLR LIF。

在上图中，DVS “Compute_DVS” 包括两个群集：“Comp A” 和 “Comp B”。“Global-Transport-Zone” 包括 “Comp A” 和 “Comp B”。

这会导致逻辑交换机（5001、5002 和 5003）的范围与在所有群集中这些逻辑交换机所在的所有主机上创建的 DLR 实例正确对齐。

现在，让我们看一下另一种情况：未将传输区域配置为包含 “Comp A” 群集。

图 4-6. 传输区域与 DVS 边界不对齐



在这种情况下，在“Comp A”群集上运行的虚拟机具有所有逻辑交换机的完全访问权限。这是因为逻辑交换机是由主机上的 dvPortgroup 表示的，而 dvPortgroup 是 DVS 范围的结构。在我们的示例环境中，“Compute_DVS”包括“Comp A”和“Comp B”。

不过，创建的 DLR 实例与传输区域范围严格对齐，这意味着，不会在“Comp A”中的主机上创建任何 DLR 实例。

因此，“web1”虚拟机可以访问“web2”和“LB”虚拟机，因为它们位于相同的逻辑交换机上，但“app1”和“db1”虚拟机无法与任何设备进行通信。

DLR 依靠控制器群集才能正常工作，而 ESG 不依靠控制器群集。在创建或更改 DLR 配置之前，请确保控制器群集已启动并且可用。

如果要将 DLR 连接到 VLAN dvPortgroup，请确保配置了 DLR 的 ESXi 主机可以在 UDP/6999 上相互访问以使基于 DLR VLAN 的 ARP 代理正常工作。

注意事项：

- 给定的 DLR 实例无法连接到位于不同传输区域的逻辑交换机。这可确保所有逻辑交换机和 DLR 实例对齐。
- 如果 DLR 连接到跨多个 DVS 的逻辑交换机，则该 DLR 无法连接到支持 VLAN 的端口组。如上所述，这可确保 DLR 实例与主机中的逻辑交换机和 dvPortgroup 正确对齐。
- 在选择 DLR 控制虚拟机位置时，应使用 DRS 反关联性规则以避免将其放在与一个或多个上游 ESG 相同的主机上（如果它们位于同一群集中）。这可降低主机故障对 DLR 转发的影响。

- 只能在单个上行链路上启用 **OSPF**（但支持多个邻接）。另一方面，可以在多个上行链路接口上启用 **BGP**（如有必要）。

DLR 和 ESG UI

DLR 和 ESG UI 指示系统工作状态。

NSX 路由 UI

vSphere Web Client UI 提供了两个与 NSX 路由有关的主要部分。

这些部分包含 L2 和控制层面基础架构依赖关系和路由子系统配置。

NSX 分布式路由需要使用控制器群集提供的功能。下面的屏幕截图显示了处于正常状态的控制器群集。

Name	Controller Node	NSX Manager	Managed By	DNS Name	Status	Peers	Software Version
	192.168.110.31 controller-1	192.168.110.15	192.168.110.15		✓ Connected		6.2.46893
	192.168.110.32 controller-2	192.168.110.15	192.168.110.15		✓ Connected		6.2.46893
	192.168.110.33 controller-3	192.168.110.15	192.168.110.15		✓ Connected		6.2.46893

请注意以下事项：

- 部署了三个控制器。
- 所有控制器的“状态”为“已连接”。
- 所有控制器的软件版本相同。
- 每个控制器节点具有两个对等项。

分布式路由的主机内核模块是作为主机上的 **VXLAN** 配置的一部分安装和配置的。这意味着，分布式路由要求准备 **ESXi** 主机并在这些主机上配置 **VXLAN**。

Clusters & Hosts	Installation Status	Firewall	VXLAN
▶ Compute Cluster A	✓ 6.2.3.3771501	✓ Enabled	✓ Configured
▶ Management & Edge Cluster	✓ 6.2.3.3771501	✓ Enabled	✓ Configured

请注意以下事项：

- “安装状态”为绿色。
- “VXLAN”为“已配置”。

确保正确配置了 **VXLAN** 传输组件。

VLAN Transport		Segment ID	Transport Zones				
Clusters & Hosts	Configuration Status	Switch	VLAN	MTU	VMKNic IP Addressing	Teaming Policy	VTEP
▼ Compute Cluster A	Unconfigure	vds-site-a	0	1600	IP Pool	Fail Over	1
esx-02a.corp.local	Ready				vmk3: 192.168.130.51		
esx-01a.corp.local	Ready				vmk3: 192.168.130.52		
▼ Management & Edge	Unconfigure	vds-mgt-edge	0	1600	IP Pool	Fail Over	1
esxmgmt-02a.corp.l	Ready				vmk3: 192.168.120.52		
esxmgmt-01a.corp.l	Ready				vmk3: 192.168.120.51		

请注意以下事项：

- VTEP 传输 VLAN 的 VLAN ID 必须正确无误。请注意，在上面的屏幕截图中，它为“0”。在大多数实际部署中，情况并非如此。
- MTU 配置为 1600 或更大。确保 MTU 未设置为 9000，预计虚拟机上的 MTU 也设置为 9000。DVS 最大 MTU 为 9000；如果虚拟机也是 9000，则没有为 VXLAN 标头留出空间。
- VMKNic 必须具有正确的地址。确保它们未设置为 169.254.x.x 地址，以表明节点无法从 DHCP 中获取地址。
- 对于相同 DVS 的所有群集成员，成组策略必须是一致的。
- VTEP 数必须与 dvUplink 数相同。确保列出了有效/预期的 IP 地址。

传输区域必须与 DVS 边界正确对齐，以避免出现在某些群集上缺少 DLR 的情况。

Name	NSX vSwitch	Status
Compute Cluster A	vds-site-a	Normal
Management & Edge...	vds-mgt-edge	Normal

“NSX Edge” UI

NSX 路由子系统是在 UI 的“NSX Edge”部分中配置和管理的。

在选择 UI 的该部分时，将显示以下视图。

Home		NSX Manager: 192.168.110.15 (Role: Primary)					
Networking & Security		0 Installing 0 Failed					
NSX Home	Dashboard	Installation	Logical Switches	NSX Edges	Firewall	SpooftGuard	
Id	Name	Type	Version	Status	Tenant	Interfaces	Size
edge-2	Local-Distributed-Router	Logical Router	6.2.3	Deployed	Default	4	Compact
edge-3	Perimeter-Gateway-01	NSX Edge	6.2.3	Deployed	Default	2	Compact
edge-4	OneArm-LoadBalancer-01	NSX Edge	6.2.3	Deployed	Default	1	Compact
edge-5	Perimeter-Gateway-02	NSX Edge	6.2.3	Deployed	Default	2	Compact
edge-6	OneArm-LoadBalancer-02	NSX Edge	6.2.3	Deployed	Default	1	Compact
edge-9178...	Universal-Distributed-Router	Universal Distributed Router	6.2.3	Deployed	Default	4	Compact

显示所有当前部署的 DLR 和 ESG，并分别显示以下信息：

- “Id” 显示 ESG 或 DLR Edge 设备 ID，可用于任何 API 调用以引用该 ESG 或 DLR。
- “租户” + “Id” 构成了 DLR 实例名称。可以在 NSX CLI 中看到和使用该名称。
- 对于 DLR，“大小”始终为“精简”；对于 ESG，它是操作员选择的大小。

除了表中的信息以外，还可以通过按钮或“操作”访问上下文菜单。

表 4-1. NSX Edge 上下文菜单

图标	操作
	“强制同步”操作清除 ESG 或 DLR 的控制虚拟机配置，重新引导，然后重新推送该配置。
	“重新部署”删除 ESG 或 DLR，然后使用相同的配置创建新的 ESG 或 DLR。将保留现有的 ID。
	“更改自动规则配置”适用于在 ESG 上启用服务时创建的 ESG 内置防火墙规则（例如，需要使用 TCP/179 的 BGP）。
	“下载技术支持日志”从 ESG 或 DLR 控制虚拟机中创建日志包。 对于 DLR，主机日志未包含在技术支持包中，需要单独进行收集。
	“更改设备大小”仅适用于 ESG。这会使用新设备执行“重新部署”（vNIC MAC 地址将发生变化）。
	通过使用“更改 CLI 凭据”，操作员可以强制更新 CLI 凭据。 如果在 5 次登录失败后在 ESG 或 DLR 控制虚拟机上锁定 CLI，这不会解除锁定。您需要等待 5 分钟，或“重新部署”ESG/DLR 以使用正确的凭据重新登录。
	“更改日志级别”更改发送到 ESG/DLR syslog 的详细信息级别。
	“配置高级调试”在启用核心转储的情况下重新部署 ESG 或 DLR，并连接额外的虚拟磁盘以存储核心转储文件。
	在创建而未部署 ESG 时，可以使用“部署”。 该选项仅执行部署步骤（部署 OVF，配置接口以及将配置推送到创建的设备）。
	如果 DLR/ESG 版本比 NSX Manager 早，则可以使用“升级版本”选项。
	“筛选器”可以按“名称”搜索 ESG/DLR。

新的 NSX Edge (DLR)

在操作员创建新的 DLR 时，将使用以下向导收集所需的信息。

New NSX Edge

1 Name and description

Name and description

Install Type: ☐ Edge Services Gateway
Provides common gateway services such as DHCP, Firewall, VPN, NAT, Routing and Load Balancing.

☒ Logical (Distributed) Router
Provides Distributed Routing and Bridging capabilities.

☐ Universal Logical (Distributed) Router
Provides Distributed Routing capabilities for Universal Logical Switches.

Name: * DLR-01

Hostname: dlr-01

Description:

Tenant: Tenant01

☒ Deploy Edge Appliance
Deploys NSX Edge Appliance to support Firewall and Dynamic routing.

☐ Enable High Availability
Enable HA, for enabling and configuring High Availability.

在“名称和描述”屏幕上，将收集以下信息：

- “名称”显示在“NSX Edge”UI 中。
- “主机名”用于设置 ESG 或 DLR 控制虚拟机的 DNS 名称，该名称显示在 SSH/控制台会话、syslog 消息以及 vCenter 的 ESG/DLR 虚拟机“摘要”页中的“DNS 名称”下面。
- “描述”位于 UI 中，它显示 NSX Edge 列表。
- “租户”用于组成 NSX CLI 使用的 DLR 实例名称。外部云管理平台也可能会使用该名称。

在“设置”屏幕上：

New NSX Edge

2 Settings

Settings

CLI credentials will be set on the NSX Edge appliance(s). These credentials can be used to login to the read only command line interface of the appliance.

User Name: * admin

Password: *

Confirm password: *

☒ Enable SSH access

Edge Control Level Logging: EMERGENCY

Set the Edge Control Level Logging

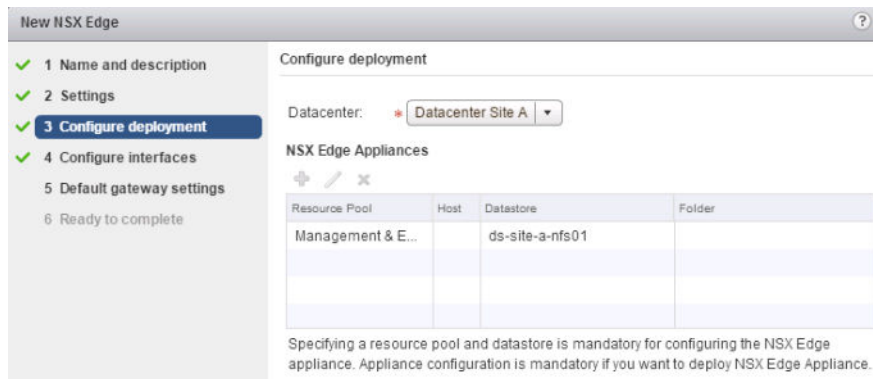
- “用户名”和“密码”设置 CLI/虚拟机控制台凭据以访问 DLR 控制虚拟机。NSX 在 ESG 或 DLR 控制虚拟机上不支持 AAA。该帐户具有 ESG/DLR 控制虚拟机的完全权限；但无法通过 CLI/虚拟机控制台更改 ESG/DLR 配置。
- “启用 SSH 访问”允许启动 DLR 控制虚拟机上的 SSH 守护程序。
 - 需要调整控制虚拟机防火墙规则以允许 SSH 网络访问。

- 操作员可以从控制虚拟机管理接口的子网上的主机中连接到 DLR 控制虚拟机，或者通过 OSPF/BGP “协议地址” 进行连接而没有此类限制（如果配置了协议地址）。

注 无法在 DLR 控制虚拟机和属于在该 DLR 的任何“内部”接口上配置的任何子网的任何 IP 地址之间建立网络连接。这是因为 DLR 控制虚拟机上的这些子网的输出接口指向伪接口“VDR”（它未连接到数据层面）。

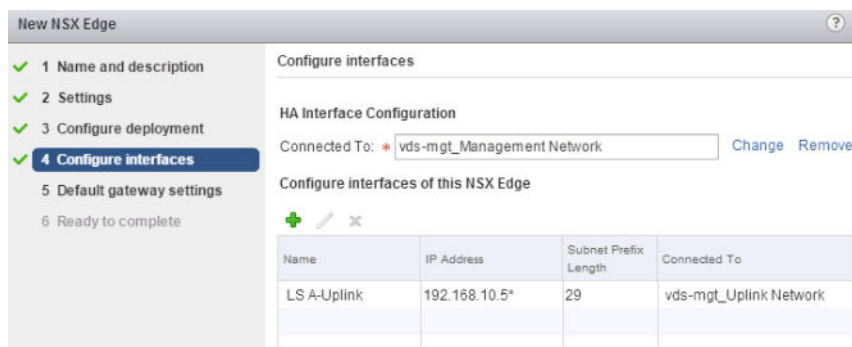
- “启用 HA” 将控制虚拟机部署为活动/备用 HA 对。
- “Edge 控制级别日志记录” 在 Edge 设备上设置 syslog 级别。

在“配置部署”屏幕上：



- “数据中心” 选择在其中部署控制虚拟机的 vCenter 数据中心。
- “NSX Edge 设备” 指的是 DLR 控制虚拟机，并允许定义恰好一个控制虚拟机（如图所示）。
 - 如果启用了“HA”，将在相同的群集、主机和数据存储上部署备用 Edge。将为活动和备用 DLR 控制虚拟机创建 DRS “单独虚拟机” 规则。

在“配置接口”屏幕上：



- “HA 接口”
 - 未创建为可路由的 DLR 逻辑接口。它仅是控制虚拟机上的 vNIC。
 - 该接口不需要使用 IP 地址，因为 NSX 通过 VMCI 管理 DLR 配置。
 - 如果在“名称和描述”屏幕上选中 DLR “启用高可用性”，该接口将用于 HA 检测信号。

- “此 NSX Edge 的接口” 指的是 DLR 逻辑接口 (LIF)
 - DLR 为 “已连接到” 的 dvPortgroup 或逻辑交换机上的虚拟机提供 L3 网关服务，这些虚拟机具有相应子网中的 IP 地址。
 - “上行链路” 类型的 LIF 在控制虚拟机上创建为 vNIC，因此，最多支持 8 个 vNIC；最后两个可用的 vNIC 分配给 HA 接口并保留一个 vNIC。
 - 动态路由需要使用 “上行链路” 类型的 LIF 才能在 DLR 上正常工作。
 - “内部” 类型的 LIF 在控制虚拟机上创建为伪 vNIC，最多可以具有 991 个伪 vNIC。

在 “默认网关设置” 屏幕上：

- 如果选定，“配置默认网关” 在 DLR 上创建静态默认路由。如果在上一屏幕中创建了 “上行链路” 类型的 LIF，则可以使用该选项。
- 如果在上行链路上使用 ECMP，建议将该选项保持禁用状态，以防止在下一跃点失败时数据层面中断。

注 右上角的双右箭头可以 “暂停” 正在执行的向导，以便以后可以恢复该向导。

ESG 和 DLR 差异

与 DLR 相比，在部署 ESG 时，向导屏幕存在一些差异。

第一个差异是 “配置部署” 屏幕：

对于 ESG，可以在 “配置部署” 中选择 Edge 大小。如果 ESG 仅用于路由，“中型” 是适用于大多数情况的典型大小。选择较大的大小并不会为 ESG 的路由进程提供更多 CPU 资源，并且不会导致更高的吞吐量。

也可以创建而不部署 ESG，这仍需配置 Edge 设备。

以后，可以通过 API 调用或使用“部署”UI 操作部署“未部署的”Edge。

如果选择 Edge HA，您必须创建至少一个“内部”接口，否则，HA 将静默失败，从而导致“脑裂”情况。

NSX UI 和 API 允许操作员移除最后一个“内部”接口，这会导致 HA 静默失败。

典型的 ESG 和 DLR UI 操作

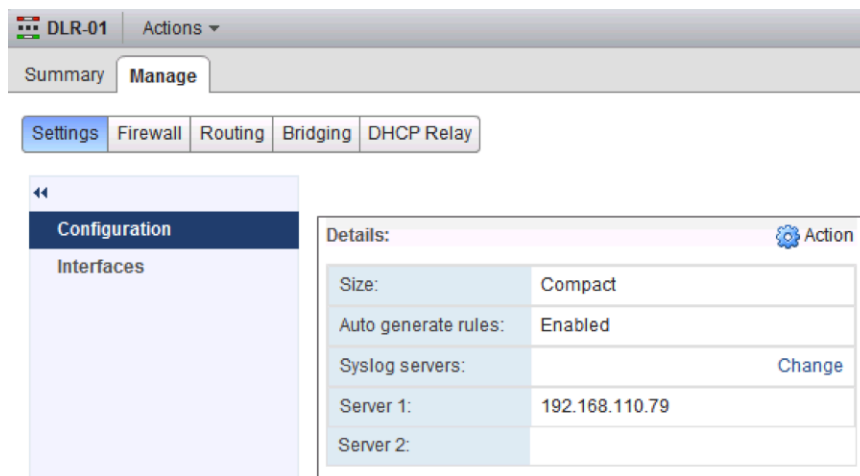
除了创建以外，在初始部署后通常还会执行一些配置操作。

其中包括：

- syslog 配置
- 静态路由管理
- 路由协议配置和路由重新分发

syslog 配置

配置 ESG 或 DLR 控制虚拟机以将日志条目发送到远程 syslog 服务器。



备注：

- 必须将 syslog 服务器配置为 IP 地址，因为 ESG/DLR 控制虚拟机没有配置 DNS 解析器。
 - 对于 ESG，可以“启用 DNS 服务”（DNS 代理），ESG 本身可以使用该服务解析 DNS 名称，但通常在具有较少依赖项的更可靠方法中将 syslog 服务器指定为 IP 地址。
- 无法在 UI 中指定 syslog 端口（它始终为 514），但可以指定协议 (UDP/TCP)。
- syslog 消息来自于 Edge 接口的 IP 地址，Edge 的转发表选择该接口以作为 syslog 服务器 IP 的输出。
 - 对于 DLR，syslog 服务器的 IP 地址不能位于在 DLR 的任何“内部”接口上配置的任何子网上。这是因为 DLR 控制虚拟机上的这些子网的输出接口指向伪接口“VDR”（它未连接到数据层面）。

默认情况下，将禁用 ESG/DLR 路由引擎的日志记录。如果需要，请单击“动态路由配置”的“编辑”按钮以通过 UI 启用日志记录。

DLR-01 Actions ▾

Summary Manage

Settings Firewall Routing Bridging DHCP Relay

Global Configuration
Static Routes
OSPF
BGP
Route Redistribution

Routing Configuration : Reset

ECMP : Disabled Enable

Default Gateway : Edit Delete

Interface :
Gateway IP :
MTU :
Description :

Dynamic Routing Configuration : Edit

Router ID :
OSPF : Disabled
BGP : Disabled
Logging : Disabled
Log Level :

您还必须配置路由器 ID，它通常是上行链路接口的 IP 地址。

静态路由

静态路由必须将下一跃点设置为与 DLR 的某个 LIF 或 ESG 的某个接口关联的子网上的 IP 地址。否则，配置将失败。

如果未选定，则自动将下一跃点与某个直接连接的子网匹配以设置“接口”。

Add Static Route ?

Network: *

10.10.10.0/24

*Network should be entered in CIDR format
e.g. 192.169.1.0/24*

Next Hop: *

192.168.10.1

Interface:

i

MTU:

1500

Description:

OK

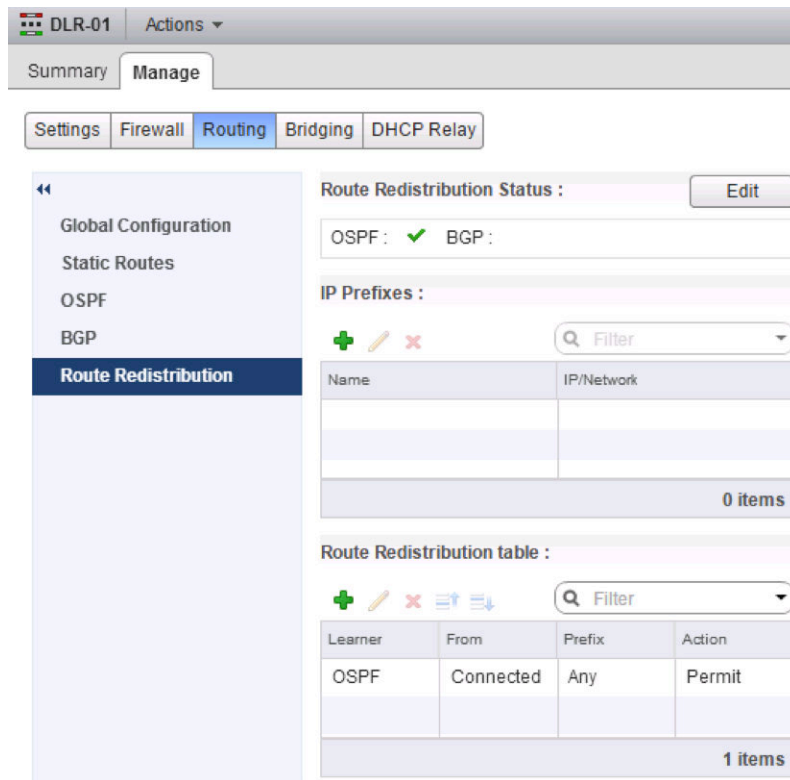
Cancel

路由重新分发

将条目添加到“路由重新分发表”并不会自动为选定的“学习者协议”启用重新分发。必须通过“路由重新分发状态”的“编辑”按钮明确完成该操作。

默认情况下，DLR 配置为将连接的路由重新分发到 OSPF，而 ESG 不是这样。

“路由重新分发表”是按从上到下的顺序处理的，并在首次匹配后停止处理。要从重新分发中排除某些前缀，请在顶部包含更具体的条目。



NSX 路由故障排除

NSX 提供了多种工具以确保路由正常工作。

NSX 路由 CLI

通过使用一组 CLI 命令，操作员可以检查 NSX 路由子系统的各个部分的运行状态。

由于 NSX 路由子系统的分布式特性，可以在各种 NSX 组件上访问一些 CLI。从 NSX 6.2 版开始，NSX 还具有一个集中式 CLI，可帮助缩短访问和登录到各种分布式组件所需的“行程时间”。它可以从一个位置中访问大多数信息：NSX Manager shell。

检查必备条件

每个 ESXi 主机必须满足两个主要的必备条件：

- 连接到 DLR 的任何逻辑交换机正常工作。
- 已成功为 VXLAN 准备 ESXi 主机。

逻辑交换机运行状况检查

NSX 路由与 NSX 逻辑交换配合使用。要验证连接到 DLR 的逻辑交换机是否正常工作，请执行以下操作：

- 查找连接到相关 DLR 的每个逻辑交换机的分段 ID (VXLAN VNI)，例如，5004..5007。

Name	Status	Transport Zone	Segment ID	Control Plane Mode	Description
LS A	Normal	Global-Transport-Zone	5004	Unicast	
LS B	Normal	Global-Transport-Zone	5005	Unicast	
LS C	Normal	Global-Transport-Zone	5006	Unicast	
LS D	Normal	Global-Transport-Zone	5007	Unicast	

- 在运行该 DLR 提供服务的虚拟机的 ESXi 主机上，检查连接到该 DLR 的逻辑交换机的 VXLAN 控制层面的状态。

```
# esxcli network vswitch dvs vmware vxlan network list --vds-name=Compute_VDS
```

VXLAN ID	Multicast IP	Control Plane	Controller Connection	Port Count
5004	N/A (headend replication)	Enabled (multicast proxy, ARP proxy)	192.168.110.201 (up)	2
5005	N/A (headend replication)	Enabled (multicast proxy, ARP proxy)	192.168.110.202 (up)	0
5006	N/A (headend replication)	Enabled (multicast proxy, ARP proxy)	192.168.110.203 (up)	1
5007	N/A (headend replication)	Enabled (multicast proxy, ARP proxy)	192.168.110.202 (up)	0

检查每个相关 VXLAN 的以下内容：

- 对于混合或单播模式下的逻辑交换机：
 - Control Plane 为 “Enabled”。
 - 列出了 “multicast proxy” 和 “ARP proxy”；即使禁用了 IP 发现，也会列出 “ARP proxy”。
 - 在 “Controller” 下面列出了有效的控制器 IP 地址，并且 “Connection” 为 “up”。
- “Port Count” 正确无误 - 至少为 1 个，即使在连接到相关逻辑交换机的该主机上没有虚拟机。该端口为 vdrPort，这是连接到 ESXi 主机上的 DLR 内核模块的特殊 dvPort。
- 运行以下命令以确保 vdrPort 连接到每个相关的 VXLAN。

```
~ # esxcli network vswitch dvs vmware vxlan network port list --vds-name=Compute_VDS --vxlan-id=5004
```

Switch Port ID	VDS Port ID	VMKNIC ID
50331656	53	0
50331650	vdrPort	0

```
~ # esxcli network vswitch dvs vmware vxlan network port list --vds-name=Compute_VDS --vxlan-id=5005
Switch Port ID  VDS Port ID  VMKNIC ID
-----
50331650      vdrPort      0
```

- 在上面的示例中，VXLAN 5004 具有一个虚拟机和一个 DLR 连接，而 VXLAN 5005 仅具有一个 DLR 连接。
- 检查是否将相应的虚拟机正确连接到对应的 VXLAN，例如，VXLAN 5004 上的 web-sv-01a。

```
~ # esxcfg-vswitch -l
DVS Name      Num Ports  Used Ports  Configured Ports  MTU      Uplinks
Compute_VDS   1536      10          512              1600     vmnic0

  DVPort ID      In Use      Client
[.skipped..]
  53              1           web-sv-01a.eth0
```

VXLAN 准备检查

作为 ESXi 主机的 VXLAN 配置的一部分，还会安装和配置 DLR 内核模块，并将其连接到为 VXLAN 准备的 DVS 上的 dvPort。

- 1 运行 `show cluster all` 以获取群集 ID。
- 2 运行 `show cluster cluster-id` 以获取主机 ID。
- 3 运行 `show logical-router host hostID connection` 以获取状态信息。

```
nsxmgr-01a# show logical-router host <hostID> connection

Connection Information:
-----

DvsName      VdrPort      NumLifs  VdrVmac
-----
Compute_VDS  vdrPort      4        02:50:56:56:44:52
  Teaming Policy: Default Teaming
  Uplink   : dvUplink1(50331650): 00:50:56:eb:41:d7(Team member)

Stats : Pkt Dropped      Pkt Replaced      Pkt Skipped
Input : 0                 0                 1968734458
Output : 303             7799             31891126
```

- 启用了 VXLAN 的 DVS 将创建一个 vdrPort，它由该 ESXi 主机上的所有 DLR 实例共享。
- “NumLifs” 指的是位于该主机上的所有 DLR 实例中的 LIF 总和。
- “VdrVmac” 是 DLR 在所有实例中的所有 LIF 上使用的 vMAC。该 MAC 在所有主机上是相同的。在 ESXi 主机外部的物理网络中传输的任何帧中，不会看到该内容。

- 对于启用了 VXLAN 的 DVS 的每个 dvUplink，具有一个匹配的 VTEP；但使用 LACP/以太网通道成组模式时除外，此时，仅创建一个 VTEP，而与 dvUplink 数无关。
 - 在离开主机时 DLR 路由的流量（源 MAC = vMAC）将源 MAC 更改为相应 dvUplink 的 pMAC。
 - 请注意，将使用原始虚拟机的源端口或源 MAC 确定 dvUplink（这是在 DVS 元数据中为每个数据包保留的）。
 - 如果在主机上具有多个 VTEP 并且某个 dvUplink 发生故障，与故障 dvUplink 关联的 VTEP 以及绑定到该 VTEP 的所有虚拟机将移动到剩下的某个 dvUplink。这样做是为了避免与将虚拟机移动到不同 VTEP 有关的控制层面更改发生洪泛。
- 每个“dvUplinkX”旁边的“()”中的数字是 dvPort 编号。这对于单个上行链路上的数据包捕获非常有用。
- 为每个“dvUplinkX”显示的 MAC 地址是与该 dvUplink 关联的“pMAC”。该 MAC 地址用于来自于 DLR 的流量，例如，DLR 生成的 ARP 查询以及在离开 ESXi 主机时 DLR 路由的任何数据包。可以在物理网络上看到该 MAC 地址：直接（如果 DLR LIF 具有 VLAN 类型）或从 VXLAN LIF 的 VXLAN 数据包中。
- Pkt Dropped/Replaced/Skipped 指的是与 DLR 内部实施详细信息有关的计数器，通常不用于故障排除或监控。

路由简要概述

为了有效地解决路由问题，了解路由的工作方式和查看相关的信息表是非常有用的。

- 1 收到一个数据包以发送到目标 IP 地址。
- 2 检查路由表并确定下一跃点的 IP 地址。
- 3 确定可访问该地址的网络接口。
- 4 获取该下一跃点的 MAC 地址（通过 ARP）。
- 5 生成一个 L2 帧。
- 6 将该帧从接口中发出。

因此，要进行路由，您需要使用：

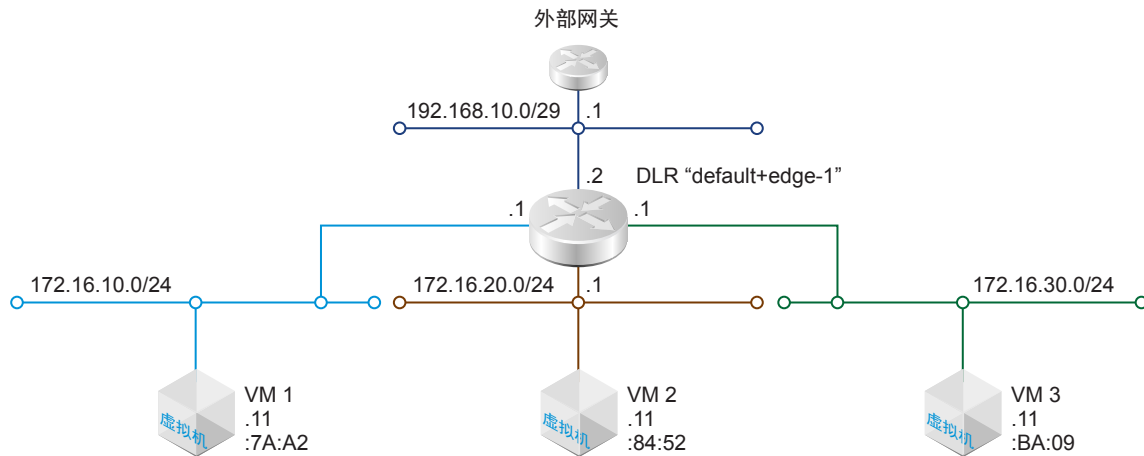
- 接口表（包含接口 IP 地址和子网掩码）
- 路由表
- ARP 表

使用示例路由拓扑验证 DLR 状态

本节讨论了如何获取 DLR 路由数据包所需的信息。

让我们使用示例路由拓扑，并创建一组逻辑交换机和 DLR 以在 NSX 中进行创建。

图 4-7. 示例路由拓扑



该图显示：

- 4 个逻辑交换机，每个交换机具有自己的子网
- 3 个虚拟机，每个逻辑交换机连接一个虚拟机
 - 每个虚拟机具有自己的 IP 地址和 IP 网关
 - 每个虚拟机具有 MAC 地址（显示了最后两个八位字节）
- 一个连接到 4 个逻辑交换机的 DLR；一个逻辑交换机用于“上行链路”，其余逻辑交换机是内部交换机
- 一个外部网关，它可能是 ESG 以作为 DLR 的上游网关

为上面的 DLR 显示了“即将完成”向导屏幕。

New NSX Edge

Ready to complete

1 Name and description
2 CLI credentials
3 Configure deployment
4 Configure interfaces
5 Configure HA
6 Ready to complete

Name and description

Name: DLR1

Install Type: Logical (Distributed) Router

Tenant:

HA: Disabled

Management Interface Configuration

Connected To: Mgmt_Edge_VDS - Mgmt

IP Address	Subnet Prefix Length

NSX Edge Appliances

Resource Pool	Host	Datastore	Folder
Management and Edge Cluster		ds-site-a-nfs01	

Interfaces

Name	IP Address	Subnet Prefix Length	Connected To
LS A	172.16.10.1*	24	LS A
LS B	172.16.20.1*	24	LS B
LS C	172.16.30.1*	24	LS C
LS D	192.168.10.2*	29	LS D

Back Next Finish Cancel

在 DLR 部署完成后，可以使用 ESXi CLI 命令查看和验证涉及的主机上的相关 DLR 的分布状态。

确认 DLR 实例

首先要确认的是，是否创建了 DLR 实例以及其控制层面是否处于活动状态。

- 1 从 NSX Manager shell 中，运行 `show cluster all` 以获取群集 ID。
- 2 运行 `show cluster cluster-id` 以获取主机 ID。
- 3 运行 `show logical-router host hostID dlr all verbose` 以获取状态信息。

```
nsxmgr# show logical-router host host-id dlr all verbose
```

```
VDR Instance Information :
```

```
-----

Vdr Name:                default+edge-1
Vdr Id:                  1460487509
Number of Lifs:          4
Number of Routes:        5
State:                   Enabled
Controller IP:           192.168.110.201
Control Plane Active:    Yes
Control Plane IP:        192.168.210.51
Edge Active:             No
```

请注意以下几点：

- 该命令显示位于给定 ESXi 主机上的所有 DLR 实例。
- “Vdr Name” 由 “租户” 和 “Edge Id” 组成。在该示例中，未指定 “租户”，因此，使用 “default” 一词。“Edge Id” 是 “edge-1”，可以在 NSX UI 中看到该 ID。
 - 如果主机上具有多个 DLR 实例，一种查找正确实例的方法是查找在 UI “NSX Edge” 中显示的 “Edge ID”。
- “Vdr Id” 对于进一步查找非常有用，包括日志。
- “Number of Lifs” 指的是位于该单个 DLR 实例上的 LIF。
- 此处，“Number of Routes” 为 5，它包含 4 个直接连接的路由（每个 LIF 一个）和一个默认路由。
- “State”、“Controller IP” 和 “Control Plane Active” 指的是 DLR 的控制层面状态，必须列出正确的控制器 IP 并且 Control Plane Active 为 Yes。请记住，DLR 功能需要使用正常工作的控制器；上面的输出显示正常 DLR 实例所需的设置。
- “控制层面 IP” 指的是 ESXi 主机用于与控制器通信的 IP 地址。该 IP 始终是与 ESXi 主机的管理 vmknic 关联的 IP 地址，在大多数情况下，该 IP 为 vmk0。
- “Edge Active” 显示该主机是否为运行该 DLR 实例的控制虚拟机的主机以及是否处于活动状态。
 - 活动 DLR 控制虚拟机的位置决定了用于执行 NSX L2 桥接（如果启用）的 ESXi 主机。

- 还提供了上述命令的“brief”版本，以便生成压缩的输出以提供概要信息。请注意，此处以十六进制格式显示“Vdr Id”：

```
nsxmgr# show logical-router host host-id dlr all brief

VDR Instance Information :
-----

State Legend: [A: Active], [D: Deleting], [X: Deleted], [I: Init]
State Legend: [SF-R: Soft Flush Route], [SF-L: Soft Flush LIF]

Vdr Name          Vdr Id      #Lifs  #Routes State      Controller Ip    CP Ip
-----
default+edge-1    0x570d4555 4      5      A          192.168.110.201  192.168.210.51
```

“Soft Flush”状态指的是 LIF 生命周期的短暂过渡状态，通常在正常 DLR 中看不到该状态。

DLR 的逻辑接口

在确定已创建 DLR 后，请确保 DLR 的所有逻辑接口存在并具有正确的配置。

- 1 从 NSX Manager shell 中，运行 `show cluster all` 以获取群集 ID。
- 2 运行 `show cluster cluster-id` 以获取主机 ID。
- 3 运行 `show logical-router host hostID dlr all brief` 以获取 dlrID（Vdr 名称）。
- 4 运行 `show logical-router host hostID dlr dlrID interface all brief` 以获取所有接口的摘要状态信息。
- 5 运行 `show logical-router host hostID dlr dlrID interface (all | intName) verbose` 以获取所有接口或特定接口的状态信息。

```
nsxmgr# show logical-router host hostID dlr dlrID interface all verbose

VDR default+edge-1:1460487509 LIF Information :

Name:                570d455500000000a
Mode:                Routing, Distributed, Internal
Id:                  Vxlan:5000
Ip(Mask):             172.16.10.1(255.255.255.0)
Connected Dvs:       Compute_VDS
VXLAN Control Plane: Enabled
VXLAN Multicast IP:  0.0.0.1
State:               Enabled
Flags:               0x2388
DHCP Relay:          Not enabled

Name:                570d455500000000c
Mode:                Routing, Distributed, Internal
Id:                  Vxlan:5002
Ip(Mask):             172.16.30.1(255.255.255.0)
Connected Dvs:       Compute_VDS
VXLAN Control Plane: Enabled
```

```

VXLAN Multicast IP: 0.0.0.1
State: Enabled
Flags: 0x2288
DHCP Relay: Not enabled

Name: 570d45550000000b
Mode: Routing, Distributed, Internal
Id: Vxlan:5001
Ip(Mask): 172.16.20.1(255.255.255.0)
Connected Dvs: Compute_VDS
VXLAN Control Plane: Enabled
VXLAN Multicast IP: 0.0.0.1
State: Enabled
Flags: 0x2388
DHCP Relay: Not enabled

Name: 570d455500000002
Mode: Routing, Distributed, Uplink
Id: Vxlan:5003
Ip(Mask): 192.168.10.2(255.255.255.248)
Connected Dvs: Compute_VDS
VXLAN Control Plane: Enabled
VXLAN Multicast IP: 0.0.0.1
State: Enabled
Flags: 0x2208
DHCP Relay: Not enabled

```

请注意以下几点：

- LIF 的 “Name” 在主机上的所有 DLR 实例中是唯一的。它在主机和 DLR 的主控制器节点上是相同的。
- LIF 的 “Mode” 显示 LIF 是路由还是桥接，以及它是内部还是上行链路。
- “Id” 显示 LIF 类型和相应的服务 ID（VXLAN 和 VNI 或 VLAN 和 VID）。
- 将为 “Routing” LIF 显示 “Ip(Mask)”。
- 如果 LIF 在混合或单播模式下连接到 VXLAN，则 “VXLAN 控制层面” 为 “Enabled”。
- 对于 VXLAN 处于单播模式的 VXLAN LIF，“VXLAN Multicast IP” 显示为 “0.0.0.1”，否则，显示实际多播 IP 地址。
- 对于路由的 LIF，“State” 应该为 “Enabled”。对于桥接 LIF，在执行桥接的主机上为 “Enabled”，在所有其他主机上为 “Init”。
- “Flags” 是 LIF 状态的摘要表示形式，并显示 LIF 是：
 - 路由还是桥接
 - VLAN LIF 是否为 DI
 - 它是否启用了 DHCP 中继
 - 请注意 0x0100 标记，它是在 DLR 导致 VXLAN VNI 加入时设置的（与在该 VXLAN 上具有虚拟机的主机相对）

- 在“brief”模式下，将以更便于阅读的格式显示标记

```
nsxmgr# show logical-router host hostID dlr dlrID interface all brief
```

VDR default+edge-1 LIF Information :

State Legend: [A:Active], [d:Deleting], [X:Deleted], [I:Init],[SF-L:Soft Flush LIF]
 Modes Legend: [B:Bridging],[E: Empty], [R:Routing],[S:Sedimented],[D:Distributed]
 Modes Legend: [In:Internal],[Up:Uplink]

Lif Name	Id	Mode	State	Ip(Mask)
-----	--	----	-----	-----
570d45550000000a	Vxlan:5001	R,D,In	A	172.16.10.1(255.255.255.0)
570d45550000000c	Vxlan:5003	R,D,In	A	172.16.30.1(255.255.255.0)
570d45550000000b	Vxlan:5002	R,D,In	A	172.16.20.1(255.255.255.0)
570d455500000002	Vxlan:5000	R,D,Up	A	192.168.10.5(255.255.255.248)

DLR 的路由

在确定 DLR 存在、正常工作并具有所有 LIF 后，接下来应检查路由表。

- 1 从 NSX Manager shell 中，运行 `show cluster all` 以获取群集 ID。
- 2 运行 `show cluster cluster-id` 以获取主机 ID。
- 3 运行 `show logical-router host hostID dlr all brief` 以获取 dlrID（Vdr 名称）。
- 4 运行 `show logical-router host hostID dlr dlrID route` 以获取所有接口的状态信息。

```
nsxmgr# show logical-router host hostID dlr dlrID route
```

VDR default+edge-1:1460487509 Route Table

Legend: [U: Up], [G: Gateway], [C: Connected], [I: Interface]
 Legend: [H: Host], [F: Soft Flush] [!: Reject] [E: ECMP]

Destination	GenMask	Gateway	Flags	Ref	Origin	UpTime	Interface
-----	-----	-----	-----	---	-----	-----	-----
0.0.0.0	0.0.0.0	192.168.10.1	UG	1	AUTO	10068944	570d455500000002
172.16.10.0	255.255.255.0	0.0.0.0	UCI	1	MANUAL	10068944	570d45550000000a
172.16.20.0	255.255.255.0	0.0.0.0	UCI	1	MANUAL	10068944	570d45550000000b
172.16.30.0	255.255.255.0	0.0.0.0	UCI	1	MANUAL	10068944	570d45550000000c
192.168.10.0	255.255.255.248	0.0.0.0	UCI	1	MANUAL	10068944	570d455500000002

请注意以下几点：

- “Interface” 显示为相应的 “Destination” 选择的输出 LIF。它设置为 DLR 的某个 LIF 的 “Lif Name”。
- 对于 ECMP 路由，存在多个具有相同 Destination、GenMask 和 Interface 但具有不同 Gateway 的路由。标记还包括 “E” 以反映这些路由的 ECMP 特性。

DLR 的 ARP 表

对于 DLR 转发的数据包，它必须能够解析下一跃点 IP 地址的 ARP 请求。该解析过程的结果存储在各个主机的 DLR 实例本地。

控制器在该过程中不起任何作用，并且不会使用控制器将生成的 **ARP** 条目分发到其他主机。

非活动缓存条目保留 600 秒，然后将其移除。有关 **DLR ARP** 解析过程的详细信息，请参见 [DLR ARP 解析过程](#)。

- 1 从 NSX Manager shell 中，运行 `show cluster all` 以获取群集 ID。
- 2 运行 `show cluster cluster-id` 以获取主机 ID。
- 3 运行 `show logical-router host hostID dlr all brief` 以获取 dlrID（Vdr 名称）。
- 4 运行 `show logical-router host hostID dlr dlrID arp` 以获取所有接口的状态信息。

```
nsxmgr# show logical-router host hostID dlr dlrID arp

VDR default+edge-1:1460487509 ARP Information :
Legend: [S: Static], [V: Valid], [P: Proxy], [I: Interface]
Legend: [N: Nascent], [L: Local], [D: Deleted]
```

Network	Mac	Flags	Expiry	SrcPort	Interface	Refcnt
-----	---	-----	-----	-----	-----	-----
172.16.10.1	02:50:56:56:44:52	VI	permanent	0	570d45550000000a	1
172.16.10.11	00:50:56:a6:7a:a2	VL	147	50331657	570d45550000000a	2
172.16.30.1	02:50:56:56:44:52	VI	permanent	0	570d45550000000c	1
172.16.30.11	00:50:56:a6:ba:09	V	583	50331650	570d45550000000c	2
172.16.20.11	00:50:56:a6:84:52	VL	568	50331658	570d45550000000b	2
172.16.20.1	02:50:56:56:44:52	VI	permanent	0	570d45550000000b	1
192.168.10.2	02:50:56:56:44:52	VI	permanent	0	570d455500000002	1
192.168.10.1	00:50:56:8e:ee:ce	V	147	50331650	570d455500000002	1

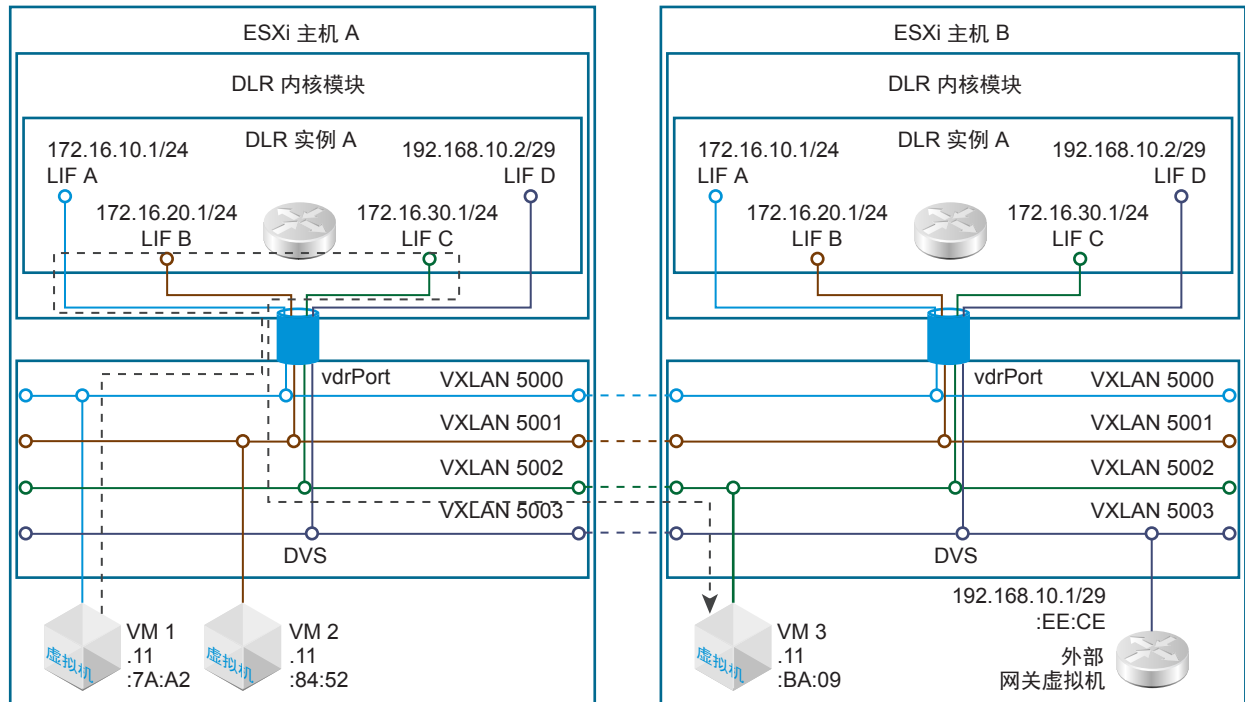
请注意以下事项：

- DLR 自己的 LIF 的所有 **ARP** 条目（“I” 标记）是相同的，并显示 [VXLAN 准备检查](#) 中讨论的相同 vMAC。
- 具有“L”标记的 **ARP** 条目对应于在运行 CLI 命令的主机上运行的虚拟机。
- “SrcPort”显示 **ARP** 条目来自的 dvPort ID。如果 **ARP** 条目来自于另一个主机，将显示 dvUplink 的 dvPort ID。该 dvPort ID 可以与 [VXLAN 准备检查](#) 中讨论的 dvUplink dvPort ID 交叉引用。
- 通常不会看到“Nascent”标记。在 DLR 等待 **ARP** 回复到达时，将设置该标记。设置了该标记的任何条目可能表明 **ARP** 解析出现问题。

可视化 DLR 及其相关主机组件

下图显示了两个主机（ESXi 主机 A 和 ESXi 主机 B），其中配置了示例“DLR 实例 A”并连接到四个 **VXLAN** LIF。

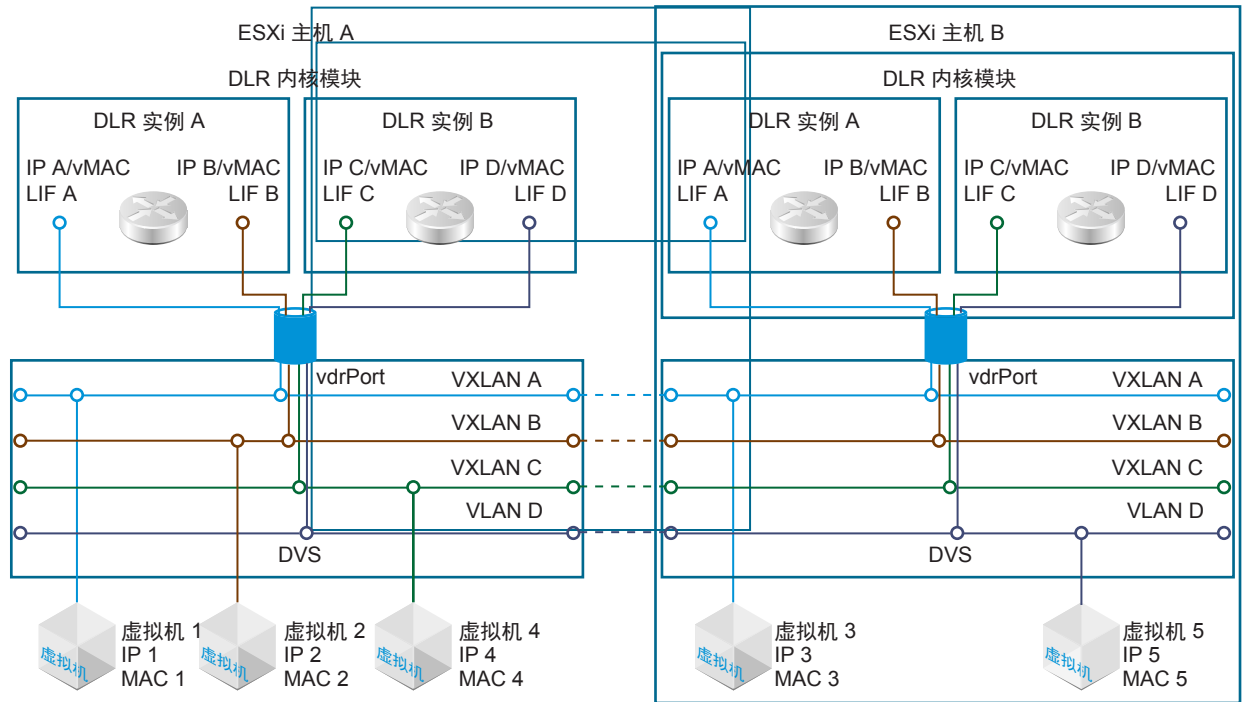
图 4-8. 两个具有单个 DLR 实例的主机



- 每个主机具有一个“L2 交换机” (DVS) 和一个“单臂路由器” (DLR 内核模块)，该路由器通过“中继”接口 (vdrPort) 连接到该“交换机”。
 - 请注意，该“中继”接口可以传输 VLAN 和 VXLAN，但在通过 vdrPort 传输的数据包中不包含 801.Q 或 UDP/VXLAN 标头。相反，DVS 使用内部元数据标记方法将该信息传送到 DLR 内核模块。
- 在看到目标 MAC = VMAC 的帧时，DVS 知道应将其发送到 DLR 并将该帧转发到 vdrPort。
- 在数据包通过 vdrPort 到达 DLR 内核模块后，将检查其元数据以确定它们所属的 VXLAN VNI 或 VLAN ID。然后，使用该信息确定数据包所属的 DLR 实例的 LIF。
 - 该系统的不足之处是，无法将多个 DLR 实例连接到给定的 VLAN 或 VXLAN。

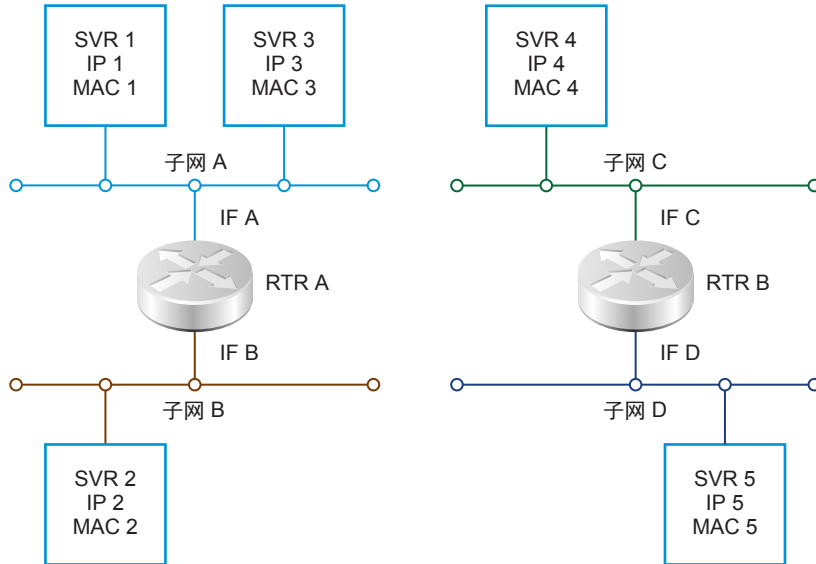
如果存在多个 DLR 实例，上图将如下所示：

图 4-9. 两个具有两个 DLR 实例的主机



这对应于具有两个独立路由域的网络拓扑，这两个域彼此完全隔离，并且可能具有重叠的 IP 地址。

图 4-10. 与两个主机和两个 DLR 实例对应的网络拓扑



分布式路由子系统架构

ESXi 主机上的 DLR 实例可以访问执行 L3 路由所需的所有信息。

- 直接连接网络（从接口的配置中了解）
- 每个子网的下一跃点（在路由表中查找）

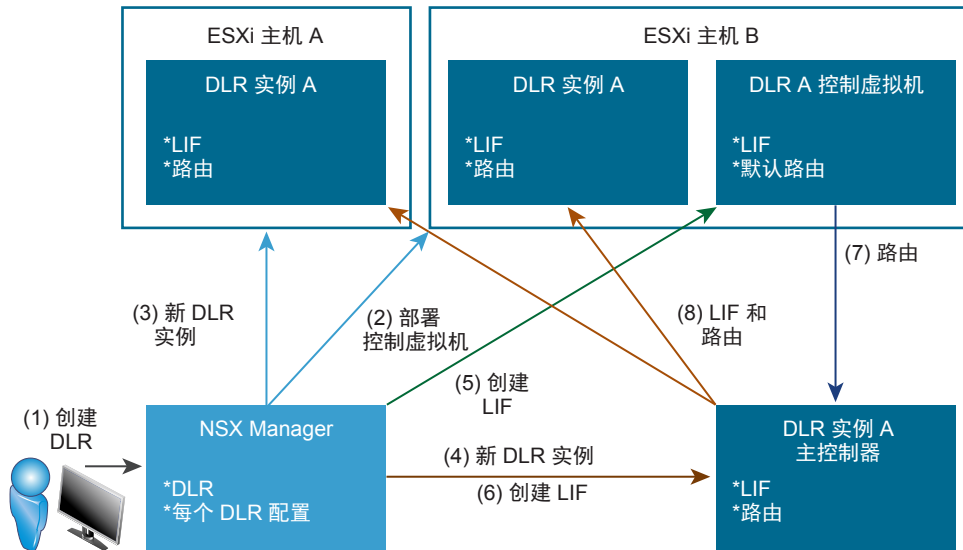
- 要插入到输出帧以到达下一跃点的 MAC 地址（ARP 表）

该信息将传送到在多个 ESXi 主机中分配的实例。

DLR 创建过程

下图简要说明了 NSX 在创建新的 DLR 时执行的过程。

图 4-11. DLR 创建过程



在使用“完成”按钮提交 UI 向导或进行 API 调用以部署新的 DLR 时，系统将执行以下步骤：

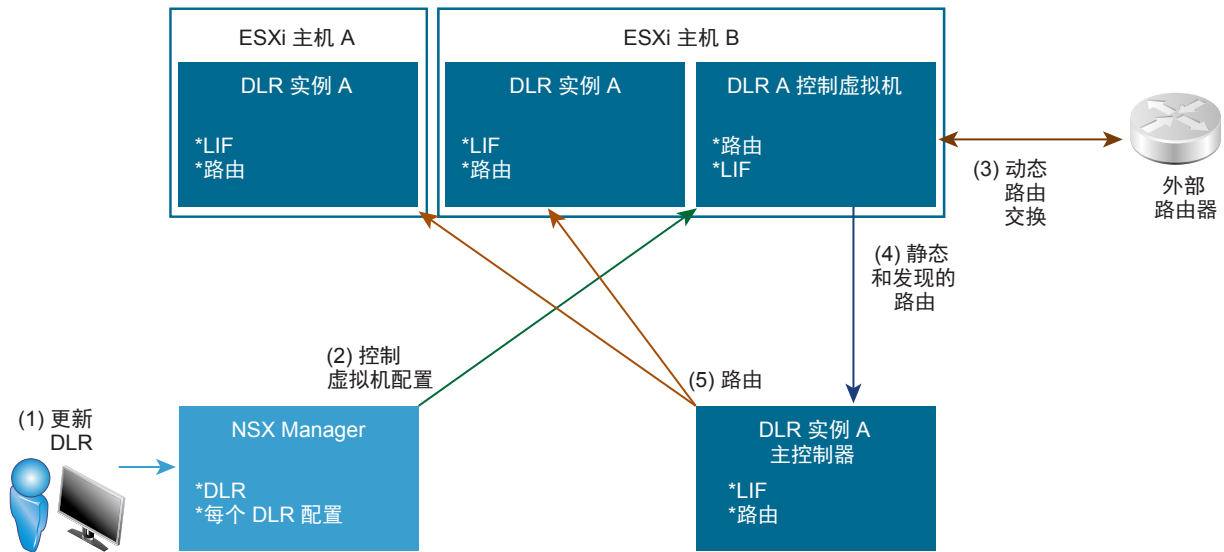
- 1 NSX Manager 收到 API 调用以部署新的 DLR（直接或从 UI 向导调用的 vSphere Web Client 中）。
- 2 NSX Manager 调用其链接的 vCenter Server 以部署一个或一对（如果请求 HA）DLR 控制虚拟机。
 - a 打开 DLR 控制虚拟机电源并连接回 NSX Manager 以准备接收配置。
 - b 如果部署了 HA 对，NSX Manager 配置反关联性规则以使 HA 对在不同的主机上运行。然后，DRS 采取措施以将它们分开。
- 3 NSX Manager 在主机上创建 DLR 实例：
 - a NSX Manager 查找要连接到新 DLR 的逻辑交换机，以确定它们属于哪个传输区域。
 - b 然后，它查找在该传输区域中配置的一组群集，并在这些群集中的每个主机上创建新的 DLR。
 - c 此时，主机仅知道新的 DLR ID，而没有任何相应的信息（LIF 或路由）。
- 4 NSX Manager 在控制器群集上创建新的 DLR 实例。
 - a 控制器群集将某个控制器节点分配为该 DLR 实例的主节点。
- 5 NSX Manager 将配置（包括 LIF）发送到 DLR 控制虚拟机。
 - a ESXi 主机（包括运行 DLR 控制虚拟机的主机）从控制器群集中接收切片信息，确定负责新 DLR 实例的控制器节点，然后连接到该控制器节点（如果没有现有的连接）。
- 6 在 DLR 控制虚拟机上创建 LIF 后，NSX Manager 在控制器群集上创建新 DLR 的 LIF。

- 7 DLR 控制虚拟机连接到新 DLR 实例的控制器节点，然后将路由发送到该控制器节点：
 - a 首先，DLR 将其路由表转换为转发表（通过将前缀解析为 LIF）。
 - b 然后，DLR 将生成的表发送到该控制器节点。
- 8 通过在步骤 5.a 中建立的连接，控制器节点将 LIF 和路由推送到新 DLR 实例所在的其他主机。

将动态路由添加到 DLR 中

在通过“直接”API 调用（与使用 vSphere Web Client UI 相对）创建 DLR 时，可能会为其提供包含动态路由的完整配置 (1)。

图 4-12. DLR 上的动态路由



- NSX Manager 收到 API 调用以更改现有 DLR 的配置，此处指的是添加动态路由。
- NSX Manager 将新配置发送到 DLR 控制虚拟机。
- DLR 控制虚拟机应用该配置并执行以下过程：建立路由邻接，交换路由信息，等等。
- 在交换路由后，DLR 控制虚拟机计算转发表，并将其发送到 DLR 的主控制器节点。
- 然后，DLR 的主控制器节点将更新的路由分发到 DLR 实例所在的 ESXi 主机。

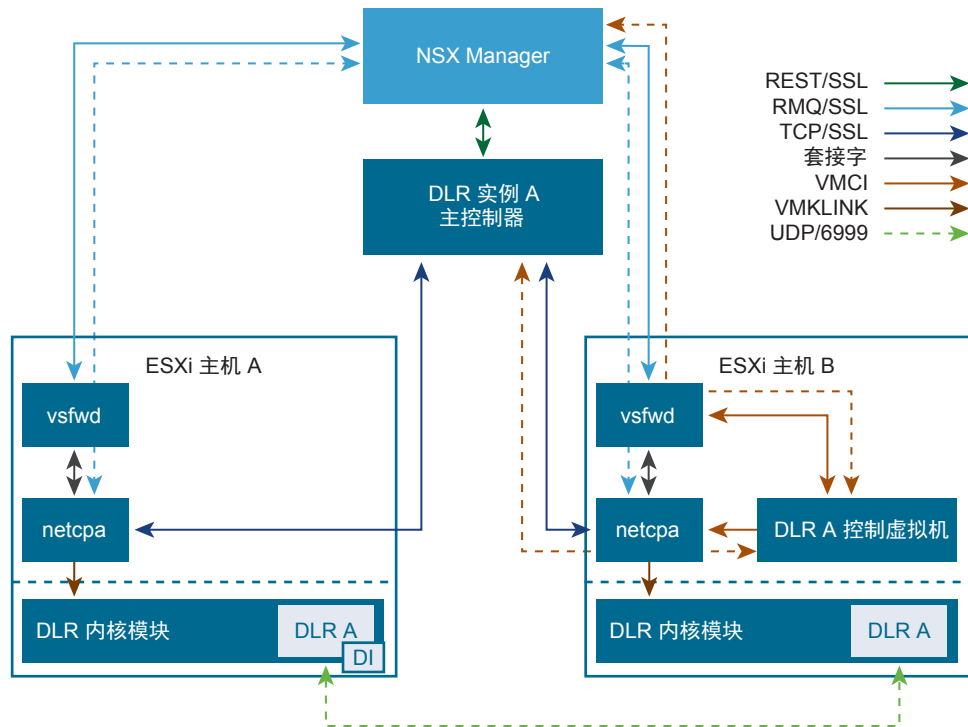
请注意，运行 DLR 控制虚拟机的 ESXi 主机上的 DLR 实例收到其 LIF，并且仅从 DLR 的主控制器节点中路由，而从不直接从 DLR 控制虚拟机或 NSX Manager 中路由。

DLR 控制和管理层面组件和通信

本节简要说明了 DLR 控制和管理层面的组件。

该图显示了这些组件以及它们之间的相应通信通道。

图 4-13. DLR 控制和管理层面组件



- **NSX Manager:**
 - 具有与控制器群集的直接通信
 - 具有到为 NSX 准备的每个主机上运行的消息总线客户端 (vsfwd) 进程的直接永久连接
- 对于每个 DLR 实例，将一个控制器节点（共有 3 个可用的节点）选择为主节点
 - 如果原始控制器节点发生故障，主节点功能可以移动到其他控制器节点
- 每个 ESXi 主机运行两个用户环境代理 (UWA): 消息总线客户端 (vsfwd) 和控制层面代理 (netcpa)
 - netcpa 需要使用 NSX Manager 中的信息才能正常工作（例如，在何处查找控制器以及如何在控制器中进行身份验证）；可以通过 vsfwd 提供的消息总线连接访问该信息
 - netcpa 还会与 DLR 内核模块通信，以使用从控制器中收到的相关信息对其进行编程
- 对于每个 DLR 实例，具有一个 DLR 控制虚拟机，它在某个 ESXi 主机上运行；DLR 控制虚拟机具有两个通信通道：
 - 通过 vsfwd 到 NSX Manager 的 VMCi 通道，用于配置控制虚拟机
 - 通过 netcpa 到 DLR 主控制器的 VMCi 通道，用于将 DLR 的路由表发送到该控制器
- 如果 DLR 具有一个 VLAN LIF，则控制器将涉及的某个 ESXi 主机指定为指定实例 (DI)。其他 ESXi 主机上的 DLR 内核模块请求 DI 在关联的 VLAN 上执行代理 ARP 查询。

NSX 路由子系统组件

NSX 路由子系统是由多个组件实现的。

- NSX Manager
- 控制器群集
- ESXi 主机模块（内核和 UWA）
- DLR 控制虚拟机
- ESG

NSX Manager

NSX Manager 提供与 NSX 路由有关的以下功能：

- 作为集中式管理层面，从而为所有 NSX 管理操作提供统一的 API 访问点
- 在主机上安装分布式路由内核模块和用户环境代理，以做好准备以提供 NSX 功能
- 创建/破坏 DLR 和 DLR LIF
- 通过 vCenter 部署/删除 DLR 控制虚拟机和 ESG
- 通过 REST API 配置控制器群集，并通过消息总线配置主机：
 - 为主机控制层面代理提供控制器 IP 地址
 - 生成证书并将其分发到主机和控制器以保护控制层面通信安全
- 通过消息总线配置 ESG 和 DLR 控制虚拟机
 - 请注意，可以在未准备的主机上部署 ESG，在这种情况下，将使用 VIX 代替消息总线

控制器群集

NSX 分布式路由需要使用群集的控制器以扩展和提高可用性，从而提供以下功能：

- 支持 VXLAN 和分布式路由控制层面
- 提供 CLI 接口以获取统计信息和运行时状态
- 为每个 DLR 实例选择主控制器节点
 - 主节点从 DLR 控制虚拟机中接收路由信息，并将其分发到主机
 - 将 LIF 表发送到主机
 - 跟踪 DLR 控制虚拟机所在的主机
 - 为 VLAN LIF 选择指定的实例，并将该信息传送到主机；通过控制层面保持活动（超时为 30 秒，检测时间可以是 20-40 秒）监控 DI 主机；为主机发送更新（如果选定的 DI 主机消失）

ESXi 主机模块

NSX 路由直接利用两个用户环境代理 (UWA) 和路由内核模块，并且还依靠 VXLAN 内核模块建立 VXLAN 连接。

下面简要说明了其中的每个组件的功能：

- 控制层面代理 (netcpa) 是 TCP (SSL) 客户端，它使用控制层面协议与控制器通信。它可能会连接到多个控制器。netcpa 与消息总线客户端 (vsfwd) 通信，以便从 NSX Manager 中检索控制层面相关信息。
- netcpa 打包和部署：
 - 该代理打包为 VXLAN VIB (vSphere 安装包)
 - 在主机准备期间，NSX Manager 通过 EAM (ESX Agency Manager) 进行安装
 - 在 ESXi netcpa 上作为服务守护程序运行
 - 可以通过其启动脚本 /etc/init.d/netcpad 启动/停止/查询
 - 可以通过“网络和安全”用户界面中的“安装”->“主机准备”->“安装状态”在单个主机或整个群集上远程重新启动
- DLR 内核模块 (vdrb) 与 DVS 集成在一起以启用 L3 转发
 - 由 netcpa 配置
 - 作为 VXLAN VIB 部署的一部分进行安装
 - 通过名为“vdrPort”的特殊中继（支持 VLAN 和 VXLAN）连接到 DVS
 - 保留有关 DLR 实例的信息以及每个实例的：
 - LIF 和路由表
 - 主机本地 ARP 缓存
- netcpa、ESG 和 DLR 控制虚拟机使用消息总线客户端 (vsfwd) 与 NSX Manager 通信
 - vsfwd 通过 vpxa/hosd 从 vCenter 设置的 /UserVars/RmqIpAddress 中获取 NSX Manager 的 IP 地址，然后使用在其他 /UserVars/Rmq* 变量中存储的每个主机的凭据登录到消息总线服务器
- 在 ESXi 主机上运行的 netcpa 依靠 vsfwd 执行以下操作：
 - 从 NSX Manager 中获取主机的控制层面 SSL 私钥和证书。然后，将这些信息存储在 /etc/vmware/ssl/rui-for-netcpa.* 中
 - 从 NSX Manager 中获取控制器的 IP 地址和 SSL 指纹。然后，将这些信息存储在 /etc/vmware/netcpa/config-by-vsm.xml 中
 - 在主机上根据 NSX Manager 指令创建和删除 DLR 实例
- 打包和部署
 - 与 netcpa 相同，它是 VXLAN VIB 的一部分
 - 在 ESXi vsfwd 上作为服务守护程序运行
 - 可以通过其启动脚本 /etc/init.d/vShield-Stateful-Firewall 启动/停止/查询
- ESG 和 DLR 控制虚拟机使用到 vsfwd 的 VMCI 通道从 NSX Manager 中接收配置

DLR 控制虚拟机和 ESG

- DLR 控制虚拟机是其 DLR 实例的“路由处理器”
 - 具有每个 DLR LIF 的“占位符”或“真正 vNIC 接口”以及 IP 配置
 - 可以运行两个可用的动态路由协议（BGP 或 OSPF）之一以及/或者使用静态路由
 - 至少需要一个“上行链路”LIF 才能运行 OSPF 或 BGP
 - 通过直接连接的 (LIF) 子网、静态和动态路由计算转发表，然后通过到 netcpa 的 VMCI 链路将其发送到 DLR 实例的主控制器
 - 在活动/备用虚拟机对配置中支持 HA
- ESG 是虚拟机中的自包含路由器
 - 完全独立于 NSX DLR 路由子系统（无 NSX 控制层面集成）
 - 通常作为一个或多个 DLR 的上游网关
 - 支持多个同时运行的动态路由协议

NSX 路由控制层面 CLI

除了主机组件以外，NSX 路由还使用控制器群集和 DLR 控制虚拟机的服务，它们都是 DLR 控制层面信息来源，并具有自己的 CLI 以用于检查这些信息。

DLR 实例主控制器

每个 DLR 实例由某个控制器节点提供服务。可以使用以下 CLI 命令查看 DLR 实例的主控制器节点包含的信息：

```
nsx-controller # show control-cluster logical-routers instance 1460487509
LR-Id      LR-Name      Hosts[]      Edge-Connection Service-Controller
1460487509 default+edge-1 192.168.210.57
              192.168.210.51
              192.168.210.52
              192.168.210.56
              192.168.110.51
              192.168.110.52

nsx-controller # show control-cluster logical-routers interface-summary 1460487509
Interface      Type  Id      IP[]
570d455500000002  vxlan 5003    192.168.10.2/29
570d45550000000b  vxlan 5001    172.16.20.1/24
570d45550000000c  vxlan 5002    172.16.30.1/24
570d45550000000a  vxlan 5000    172.16.10.1/24
```

```
nsx-controller # show control-cluster logical-routers routes 1460487509
LR-Id      Destination      Next-Hop
1460487509  0.0.0.0/0        192.168.10.1
```

- “show control-cluster logical-routers” 命令的 “instance” 子命令显示连接到该 DLR 实例的该控制器的主机列表。在正常工作的环境中，该列表包含所有群集中 DLR 所在的所有主机。
- “interface-summary” 显示控制器从 NSX Manager 中获悉的 LIF。该信息将发送到主机。
- “routes” 显示该 DLR 的控制虚拟机发送到该控制器的路由表。请注意，与 ESXi 主机上不同，该表不包含任何直接连接的子网，因为该信息是 LIF 配置提供的。

DLR 控制虚拟机

DLR 控制虚拟机具有 LIF 和路由/转发表。DLR 控制虚拟机的生命周期的主要输出是 DLR 路由表，这是 Interfaces 和 Routes 生成的。

```
edge-1-0> show ip route

Codes: 0 - OSPF derived, i - IS-IS derived, B - BGP derived,
C - connected, S - static, L1 - IS-IS level-1, L2 - IS-IS level-2,
IA - OSPF inter area, E1 - OSPF external type 1, E2 - OSPF external type 2

Total number of routes: 5

S      0.0.0.0/0      [1/1]      via 192.168.10.1
C      172.16.10.0/24 [0/0]      via 172.16.10.1
C      172.16.20.0/24 [0/0]      via 172.16.20.1
C      172.16.30.0/24 [0/0]      via 172.16.30.1
C      192.168.10.0/29 [0/0]      via 192.168.10.2

edge-1-0> show ip forwarding
Codes: C - connected, R - remote,
      > - selected route, * - FIB route
R>* 0.0.0.0/0 via 192.168.10.1, vNic_2
C>* 172.16.10.0/24 is directly connected, VDR
C>* 172.16.20.0/24 is directly connected, VDR
C>* 172.16.30.0/24 is directly connected, VDR
C>* 192.168.10.0/29 is directly connected, vNic_2
```

- 转发表的用途是显示选择哪个 DLR 接口以作为给定目标子网的输出。
 - 将为所有“内部”类型的 LIF 显示“VDR”接口。“VDR”接口是与 vNIC 不对应的伪接口。

DLR 控制虚拟机的接口可能如下所示：

```
edge-1-0> show interface
Interface VDR is up, line protocol is up
index 2 metric 1 mtu 1500 <UP,BROADCAST,RUNNING,NOARP>
HWaddr: be:3d:a1:52:90:f4
inet6 fe80::bc3d:a1ff:fe52:90f4/64
inet 172.16.10.1/24
inet 172.16.20.1/24
```

```

inet 172.16.30.1/24
proxy_arp: disabled
Auto-duplex (Full), Auto-speed (2460Mb/s)
  input packets 0, bytes 0, dropped 0, multicast packets 0
  input errors 0, length 0, overrun 0, CRC 0, frame 0, fifo 0, missed 0
  output packets 0, bytes 0, dropped 0
  output errors 0, aborted 0, carrier 0, fifo 0, heartbeat 0, window 0
  collisions 0

Interface vNic_0 is up, line protocol is up
index 3 metric 1 mtu 1500 <UP,BROADCAST,RUNNING,MULTICAST>
HWaddr: 00:50:56:8e:1c:fb
inet6 fe80::250:56ff:fe8e:1cfb/64
inet 169.254.1.1/30
inet 10.10.10.1/24
proxy_arp: disabled
Auto-duplex (Full), Auto-speed (2460Mb/s)
  input packets 582249, bytes 37339072, dropped 49, multicast packets 0
  input errors 0, length 0, overrun 0, CRC 0, frame 0, fifo 0, missed 0
  output packets 4726382, bytes 461202852, dropped 0
  output errors 0, aborted 0, carrier 0, fifo 0, heartbeat 0, window 0
  collisions 0

Interface vNic_2 is up, line protocol is up
index 9 metric 1 mtu 1500 <UP,BROADCAST,RUNNING,MULTICAST>
HWaddr: 00:50:56:8e:ae:08
inet 192.168.10.2/29
inet6 fe80::250:56ff:fe8e:ae08/64
proxy_arp: disabled
Auto-duplex (Full), Auto-speed (2460Mb/s)
  input packets 361446, bytes 30167226, dropped 0, multicast packets 361168
  input errors 0, length 0, overrun 0, CRC 0, frame 0, fifo 0, missed 0
  output packets 361413, bytes 30287912, dropped 0
  output errors 0, aborted 0, carrier 0, fifo 0, heartbeat 0, window 0
  collisions 0

```

感兴趣的注意事项：

- “VDI” 接口没有关联的虚拟机网卡 (vNIC)。这是单个“伪接口”，将为其配置 DLR 的所有“内部” LIF 的所有 IP 地址。
- 此示例中的 Nic_0 接口是 HA 接口。
 - 上面的输出是从启用了 HA 的部署 DLR 中提取的，并为 HA 接口分配一个 IP 地址。这显示为两个 IP 地址：169.254.1.1/30（为 HA 自动分配的）和 10.10.10.1/24（为 HA 接口手动分配的）。
 - 在 ESG 上，操作员可以手动将某个 vNIC 指定为 HA，或者保留默认设置以让系统自动从可用的“内部”接口中进行选择。具有“内部”类型是一项要求，否则，HA 将失败。
- vNic_2 接口具有“上行链路”类型；因此，它表示为“真实” vNIC。
 - 请注意，在该接口上看到的 IP 地址与 DLR 的 LIF 相同；但 DLR 控制虚拟机不会应答 LIF IP 地址（此处为 192.168.10.2/29）的 ARP 查询。可以为该 vNIC 的 MAC 地址应用一个 ARP 筛选器以进行应答。

- 在 DLR 上配置动态路由协议之前，上述观点是正确的，此时，将 IP 地址与 ARP 筛选器一起移除，并替换为在动态路由协议配置期间指定的“协议 IP”地址。
- 在 DLR 控制虚拟机上运行的动态路由协议使用该 vNIC 与其他路由器通信以发布和获悉路由。

NSX 路由子系统故障模式和影响

本章介绍了可能会影响 NSX 路由子系统组件的典型故障场景，并简要说明了这些故障的影响。

NSX Manager

表 4-2. NSX Manager 故障模式和影响

故障模式	故障影响
到 NSX Manager 虚拟机的网络连接中断	<ul style="list-style-type: none"> ■ 所有 NSX Manager 功能完全中断，包括用于 NSX 路由/桥接的 CRUD ■ 不会丢失配置数据 ■ 数据层面或控制层面不会中断
NSX Manager 和 ESXi 主机之间的网络连接中断，或者 RabbitMQ 服务器发生故障	<ul style="list-style-type: none"> ■ 如果 DLR 控制虚拟机或 ESG 在受影响的主机上运行，这些主机上的 CRUD 操作将失败 ■ 在受影响的主机上创建和删除 DLR 实例失败 ■ 不会丢失配置数据 ■ 数据层面或控制层面不会中断 ■ 任何动态路由更新继续正常工作
NSX Manager 和控制器之间的网络连接中断	<ul style="list-style-type: none"> ■ NSX 分布式路由和桥接的创建、更新和删除操作失败 ■ 不会丢失配置数据 ■ 数据层面或控制层面不会中断
NSX Manager 虚拟机已破坏（数据存储故障）	<ul style="list-style-type: none"> ■ 所有 NSX Manager 功能完全中断，包括用于 NSX 路由/桥接的 CRUD ■ 如果 NSX Manager 还原为较旧的配置，一部分路由/桥接实例可能会变为孤立实例，从而需要手动进行清理和协调 ■ 数据层面或控制层面不会中断，除非需要进行协调

控制器群集

表 4-3. NSX Controller 故障模式和影响

故障模式	故障影响
控制器群集与 ESXi 主机之间的网络连接中断	<ul style="list-style-type: none"> ■ DLR 控制层面功能（创建、更新和删除路由，包括动态路由）完全中断 ■ DLR 管理层面功能（在主机上创建、更新和删除 LIF）中断 ■ 将影响 VXLAN 转发，这可能会导致端到端 (L2+L3) 转发过程也会失败 ■ 根据最后已知状态，数据层面继续正常工作
一个或两个控制器与 ESXi 主机之间的连接中断	<ul style="list-style-type: none"> ■ 如果受影响的控制器仍然可以访问群集中的其他控制器，该控制器控制的任何 DLR 实例将受到上面所述的相同影响。其他控制器不会自动接管

表 4-3. NSX Controller 故障模式和影响（续）

故障模式	故障影响
一个控制器与其他控制器之间的网络连接中断（或完全中断）	<ul style="list-style-type: none"> 两个剩下的控制器接管隔离的控制器处理的 VXLAN 和 DLR 受影响的控制器进入只读模式，丢弃到主机的会话并拒绝新的会话
控制器之间的连接中断	<ul style="list-style-type: none"> 所有控制器将进入只读模式，关闭到主机的连接并拒绝新的连接 DLR 的所有 LIF 和路由（包括动态路由）的创建、更新和删除操作失败 NSX 路由配置 (LIF) 可能在 NSX Manager 和控制器群集之间不同步，从而需要手动干预以重新同步 主机将继续在最后已知控制层面状态下运行
一个控制器虚拟机丢失	<ul style="list-style-type: none"> 控制器群集缺少冗余 管理/控制层面继续正常运行
两个控制器虚拟机丢失	<ul style="list-style-type: none"> 其余控制器将进入只读模式；受到的影响与控制器之间的连接中断时相同（如上所述）。可能需要手动恢复群集

主机模块

netcpa 依靠主机 SSL 密钥和证书以及 SSL 指纹与控制器建立安全通信。这些信息是通过消息总线（由 vsfwd 提供）从 NSX Manager 中获取的。

如果证书交换过程失败，netcpa 将无法成功连接到控制器。

注意：本节不涉及内核模块故障，因为这种故障的影响非常严重 (PSOD) 并且很少会发生。

表 4-4. 主机模块故障模式和影响

故障模式	故障影响
vsfwd 使用用户名/密码身份验证访问消息总线服务器（可能会过期）	<ul style="list-style-type: none"> 如果新准备的 ESXi 主机上的 vsfwd 在两小时内无法访问 NSX Manager，在安装期间提供的临时登录名/密码将过期，并且该主机上的消息总线无法运行
消息总线客户端 (vsfwd) 的故障影响取决于时间。	
如果它在 NSX 控制层面的其他部分进入稳定运行状态之前发生故障	<ul style="list-style-type: none"> 主机上的分布式路由停止工作，因为主机无法与控制器通信 主机无法从 NSX Manager 中获悉 DLR 实例
如果它在主机进入稳定状态后发生故障	<ul style="list-style-type: none"> 在主机上运行的 ESG 和 DLR 控制虚拟机无法接收配置更新 主机未获悉新的 DLR，并且无法删除现有的 DLR 根据主机在发生故障时具有的配置，主机数据路径将继续运行

表 4-5. netcpa 故障模式和影响

故障模式	故障影响
控制层面代理 (netcpa) 的故障影响取决于时间。	
如果它在 NSX 数据路径内核模块进入稳定运行状态之前发生故障	<ul style="list-style-type: none"> 主机上的分布式路由停止工作
如果它在主机进入稳定状态后发生故障	<ul style="list-style-type: none"> 在主机上运行的 DLR 控制虚拟机无法将其转发表更新发送到控制器 分布式路由数据路径不会从控制器中收到任何 LIF 或路由更新，但根据故障前具有的状态继续运行

DLR 控制虚拟机

表 4-6. DLR 控制虚拟机故障模式和影响

故障模式	故障影响
DLR 控制虚拟机丢失或关闭电源	<ul style="list-style-type: none"> 该 DLR 的 LIF 和路由的创建、更新和删除操作失败 不会将任何动态路由更新发送到主机（包括撤消通过现在断开的邻接收到的前缀）
DLR 控制虚拟机与 NSX Manager 和控制器之间的连接中断	<ul style="list-style-type: none"> 影响与上面相同，所不同的是，如果 DLR 控制虚拟机及其路由邻接仍然启动，与以前获悉的前缀之间的流量将不会受到影响
DLR 控制虚拟机与 NSX Manager 之间的连接中断	<ul style="list-style-type: none"> 该 DLR 的 LIF 和路由的 NSX Manager 创建、更新和删除操作失败，并且不会重试 动态路由更新继续进行传播
DLR 控制虚拟机与控制器之间的连接中断	<ul style="list-style-type: none"> 该 DLR 的任何路由更改（静态或动态）不会传播到主机

与路由有关的 NSX 日志

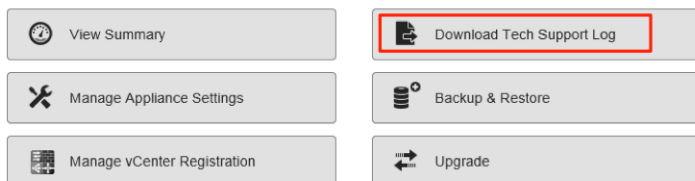
最佳做法是，配置 NSX 的所有组件以将其日志发送到集中式收集器，以便在一个地方检查这些日志。

如有必要，您可以更改 NSX 组件的日志级别。有关详细信息，请参见[设置 NSX 组件的日志记录级别](#)。

NSX Manager 日志

- NSX Manager CLI 中的 `show log`
- 通过 NSX Manager UI 收集的技术支持日志包

NSX Manager Virtual Appliance Management



NSX Manager 日志包含与管理层面有关的信息，其中包括创建、读取、更新和删除 (CRUD) 操作。

控制器日志

控制器包含多个模块，很多模块具有自己的日志文件。可以使用 `show log <log file> [filtered-by <string>]` 命令访问控制器日志。与路由有关的日志文件如下所示：

- `cloudnet/cloudnet_java-vnet-controller.<start-time-stamp>.log`
- `cloudnet/cloudnet_cpp.log.INFO`
- `cloudnet/cloudnet_cpp.log.nvp-controller.root.log.INFO.<start-time-stamp>`
- `cloudnet/cloudnet_cpp.log.ERROR`（如果出现任何错误，则包含该文件。）

控制器日志非常详细，在大多数情况下，只有在请求 VMware 工程团队帮助解决更困难的问题时，才需要使用这些日志。

除了 `show log CLI` 以外，还可以使用 `watch log <logfile> [filtered-by <string>]` 命令在更新各个日志文件时实时观察这些文件。

这些日志包含在控制器支持包中，可以在 NSX UI 中选择一个控制器节点并单击 **下载技术支持日志 (Download tech support logs)** 图标以生成并下载该支持包。

ESXi 主机日志

在 ESXi 主机上运行的 NSX 组件写入几个日志文件：

- VMkernel 日志： `/var/log/vmkernel.log`
- 控制层面代理日志： `/var/log/netcpa.log`
- 消息总线客户端日志： `/var/log/vsfwd.log`

也可以将这些日志作为从 vCenter Server 中生成的虚拟机支持包的一部分进行收集。

ESG/DLR 控制虚拟机日志

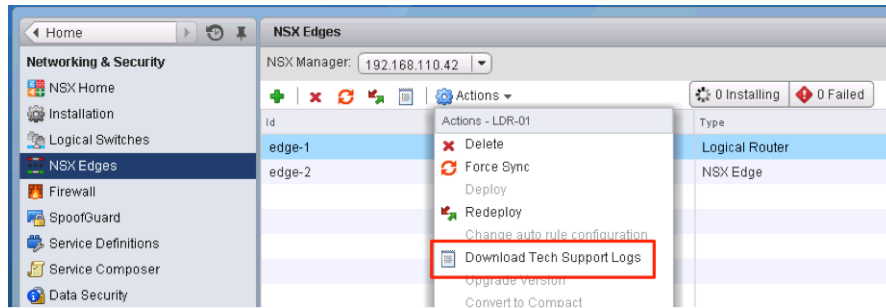
可以使用两种方法访问 ESG 和 DLR 控制虚拟机上的日志文件：使用 CLI 显示这些文件，或者使用 CLI 或 UI 下载技术支持包。

用于显示日志的 CLI 命令是 `show log [follow | reverse]`。

要下载技术支持包，请执行以下操作：

- 从 CLI 中，进入 `enable` 模式，然后运行 `export tech-support <[scp | ftp]> <URI>` 命令。

- 从 vSphere Web Client 中，在**操作 (Actions)**菜单中选择**下载技术支持日志 (Download Tech Support Logs)**选项。



其他有用的文件及其位置

虽然严格来说很多文件并不是日志，但它们可以帮助了解和解决 NSX 路由问题。

- 控制层面代理配置 `/etc/vmware/netcpa/config-by-vsm.xml` 包含有关以下组件的信息：
 - 控制器 IP 地址、TCP 端口、证书指纹、SSL 启用/禁用
 - 启用了 VXLAN 的 DVS 上的 dvUplink（成组策略、名称、UUID）
 - 主机了解的 DLR 实例（DLR ID、名称）
- 控制层面代理配置 `/etc/vmware/netcpa/netcpa.xml` 包含各种 netcpa 配置选项，包括日志记录级别（默认为 **info**）。
- 控制层面证书文件：`/etc/vmware/ssl/rui-for-netcpa.*`
 - 两个文件：主机证书和主机私钥
 - 用于验证到控制器的主机连接

所有这些文件是 netcpa 使用从 NSX Manager 中收到的信息创建的（通过 vsfwd 提供的消息总线连接）。

常见故障情况和修复

最常见的故障情况分为两类。

它们是配置和控制层面问题。也可能是管理层面问题，但并不常见。

配置问题和修复

表 4-7 中介绍了常见配置问题及其影响。

表 4-7. 常见配置问题和影响

问题	影响
动态路由的协议和转发 IP 地址是相反的	没有建立动态协议邻接
传输区域与 DVS 边界不对齐	分布式路由在一部分 ESXi 主机上无法正常工作（在传输区域中缺少这些主机）

表 4-7. 常见配置问题和影响（续）

问题	影响
动态路由协议配置不匹配（计时器、MTU、BGP ASN、密码、接口到 OSPF 区域的映射）	没有建立动态协议邻接
为 DLR HA 接口分配了 IP 地址并允许重新分发连接的路由	DLR 控制虚拟机可能会吸收 HA 接口子网的流量并产生流量黑洞

要解决这些问题，请查看配置并根据需要进行更正。

如果需要，请使用 `debug ip ospf` 或 `debug ip bgp` CLI 命令，并观察 DLR 控制虚拟机或 ESG 控制台（而不是通过 SSH 会话）上的日志以检测协议配置问题。

控制层面问题和修复

发现的控制层面问题通常是以下问题造成的：

- 主机控制层面代理 (netcpa) 无法通过 vsfwd 提供的消息总线通道连接到 NSX Manager
- 控制器群集在处理 DLR/VXLAN 实例的主角色时出现问题

通常，可以重新启动某个 NSX Controller（控制器的 CLI 上的 `restart controller`）以解决与处理主角色有关的问题。

有关解决控制层面问题的详细信息，请参见 <http://kb.vmware.com/kb/2125767>。

收集故障排除数据

本节简要说明了通常用于排除 NSX 路由故障的 CLI 命令。

NSX Manager

从 NSX 6.2 开始，以前从 NSX Controller 和其他 NSX 组件中运行以解决 NSX 路由问题的命令现在从 NSX Manager 中直接运行。

- DLR 实例列表
- 每个 DLR 实例的 LIF 列表
- 每个 DLR 实例的路由列表
- 每个 DLR 桥接实例的 MAC 地址列表
- 接口
- 路由和转发表
- 动态路由协议（OSPF 或 BGP）状态
- NSX Manager 发送到 DLR 控制虚拟机或 ESG 的配置

DLR 控制虚拟机和 ESG

DLR 控制虚拟机和 ESG 提供在其接口上捕获数据包的功能。数据包捕获可以帮助解决路由协议问题。

- 1 运行 `show interfaces` 以列出接口名称。

2 运行 `debug packet [display | capture] interface <interface name>`。

- 如果使用捕获，数据包将保存在 `.pcap` 文件中。

3 运行 `debug show files` 以列出保存的捕获文件。

4 运行 `debug copy [scp | ftp] ...` 以下载捕获包进行脱机分析。

```
dlr-01-0> debug packet capture interface vNic_2
tcpdump: listening on vNic_2, link-type EN10MB (Ethernet), capture size 65535 bytes
43 packets captured
48 packets received by filter
0 packets dropped by kernel
```

```
dlr-01-0> debug show files
total 4.0K
-rw----- 1 3.6K Mar 30 23:49 tcpdump_vNic_2.0
```

```
dlr-01-0> debug copy
  scp  use scp to copy
  ftp  use ftp to copy
```

```
dlr-01-0> debug copy scp
  URL  user@<remote-host>:<path-to>
```

`debug packet` 命令在后台使用 `tcpdump` 并且可以接受筛选修饰符，这些修饰符的格式类似于 UNIX 上的 `tcpdump` 筛选修饰符。唯一需要注意的是，将筛选表达式中的任何空格替换为下划线（“_”）。

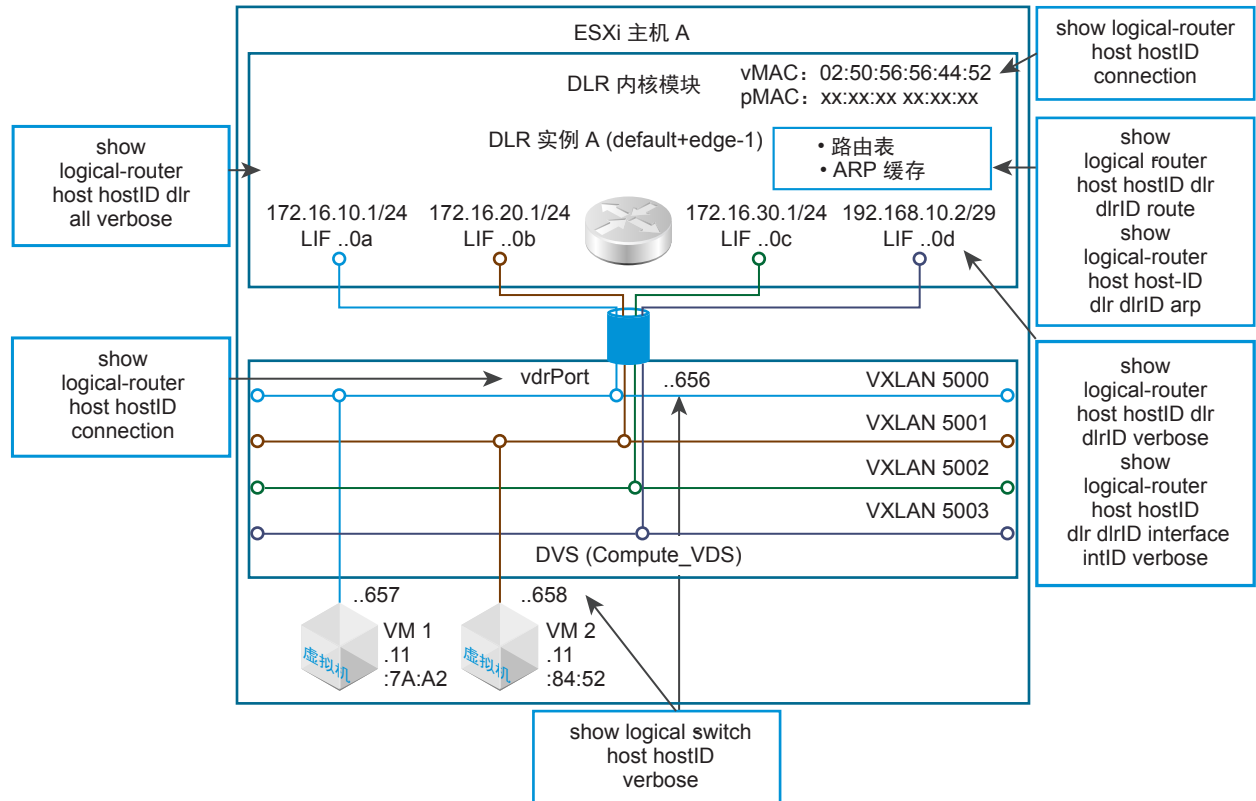
例如，以下命令显示通过 `vNic_0` 的所有流量（SSH 除外），以避免查看属于交互式会话本身的流量。

```
plr-02-0> debug packet display interface vNic_0 port_not_22
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on vNic_0, link-type EN10MB (Ethernet), capture size 65535 bytes
04:10:48.197768 IP 192.168.101.3.179 > 192.168.101.2.25698: Flags [P.], seq 4191398894:4191398913, ack
2824012766, win 913, length 19: BGP, length: 19
04:10:48.199230 IP 192.168.101.2.25698 > 192.168.101.3.179: Flags [.] , ack 19, win 2623, length 0
04:10:48.299804 IP 192.168.101.2.25698 > 192.168.101.3.179: Flags [P.], seq 1:20, ack 19, win 2623,
length 19: BGP, length: 19
04:10:48.299849 IP 192.168.101.3.179 > 192.168.101.2.25698: Flags [.] , ack 20, win 913, length 0
04:10:49.205347 IP 192.168.101.3.179 > 192.168.101.2.25698: Flags [P.], seq 19:38, ack 20, win 913,
length 19: BGP, length: 19
```

ESXi 主机

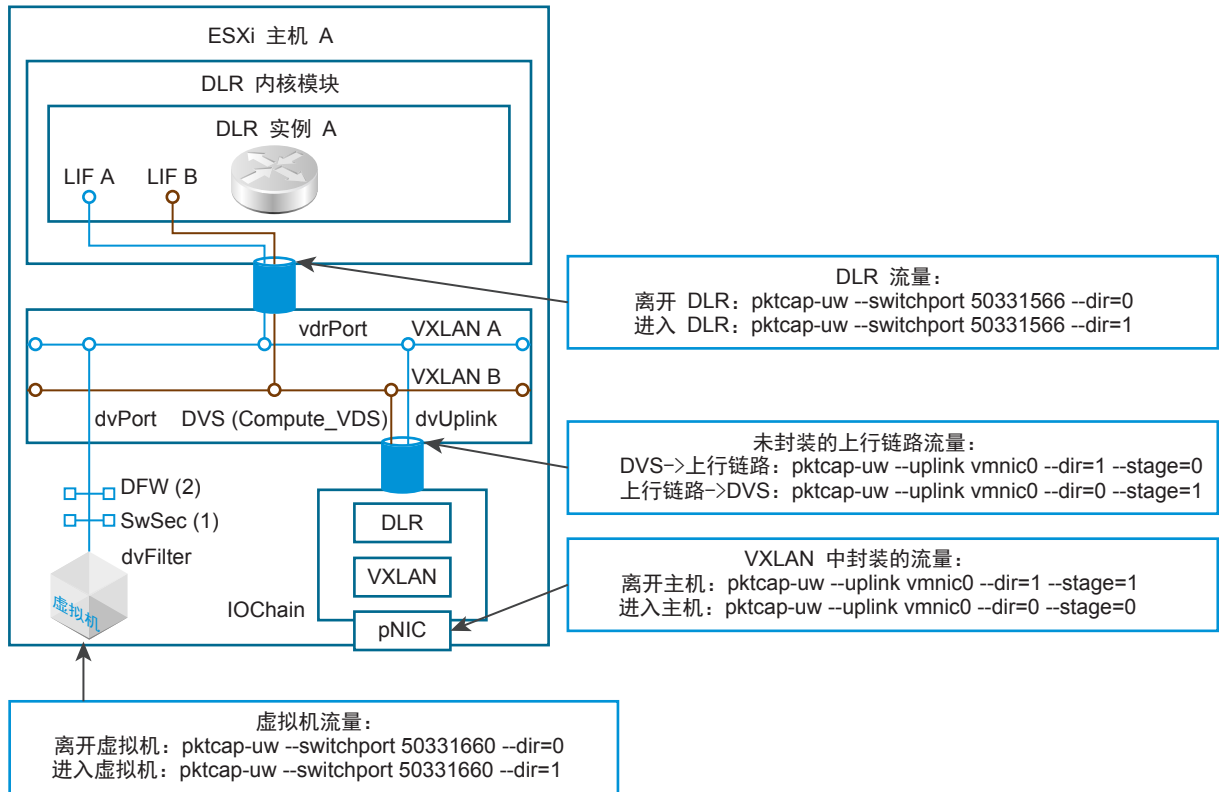
主机与 NSX 路由密切相关。图 4-14 直观地显示了路由子系统涉及的组件以及用于显示相关信息的 NSX Manager CLI 命令：

图 4-14. 与解决 NSX 路由问题有关的主机组件



在数据路径中捕获的数据包可以帮助找出在数据包转发的各个阶段出现的问题。图 4-15 涵盖了主要捕获点和使用的相应 CLI 命令。

图 4-15. 捕获点和相关的 CLI 命令



Edge 设备故障排除

本主题提供了有助于您了解 VMware NSX Edge 设备和进行故障排除的信息。

要解决 NSX Edge 设备问题，请验证下面的每个故障排除步骤是否适用于您的环境。每个步骤提供了相应说明或指向文档的链接，以消除可能的根源并在必要时采取纠正措施。这些步骤按最适当的顺序进行排列，以查找问题并确定相应的解决方案。不要跳过某个步骤。

请参阅当前版本的发行说明以查看是否解决了该问题。

确保在安装 VMware NSX Edge 时满足最低系统要求。请参阅《NSX 安装指南》。

安装和升级问题

- 验证遇到的问题是否与“Would Block”问题无关。有关详细信息，请参阅 <https://kb.vmware.com/kb/2107951>。
- 如果升级或重新部署成功，但 Edge 接口没有连接，请验证后端 2 层交换机上的连接。请参阅 <https://kb.vmware.com/kb/2135285>。
- 如果 Edge 部署或升级失败并出现以下错误：

```
/sbin/ifconfig vNic_1 up failed : SIOCSIFFLAGS: Invalid argument
```

或

- 如果部署或升级成功，但在 Edge 接口上没有连接：
- 运行 `show interface` 命令以及 Edge 支持日志将显示类似下面的条目：

```
vNic_0: <NO-CARRIER,BROADCAST,MULTICAST,UP> mtu 1500 qdisc mq state DOWN qlen 1000  
link/ether 00:50:56:32:05:03 brd ff:ff:ff:ff:ff:ff  
inet 21.12.227.244/23 scope global vNic_0  
inet6 fe80::250:56ff:fe32:503/64 scope link tentative dadfailed  
valid_lft forever preferred_lft forever
```

在这两种情况下，主机交换机未就绪或出现某些问题。要解决该问题，请调查主机交换机。

配置问题

- 收集 NSX Edge 诊断信息。请参阅 <https://kb.vmware.com/kb/2079380>。

搜索字符串 `vse_die` 以筛选 NSX Edge 日志。包含该字符串的日志条目可能会提供有关配置错误的信息。

防火墙问题

- 如果出现非活动超时问题，且应用程序闲置很长时间，请使用 REST API 增加非活动超时设置。请参阅 <https://kb.vmware.com/kb/2101275>。

Edge 防火墙数据包丢弃问题

- 1 使用 `show firewall` 命令检查防火墙规则表。`usr_rules` 表显示配置的规则。

```
nsxedge> show firewall
Chain PREROUTING (policy ACCEPT 3146M packets, 4098G bytes)
rid  pkts bytes target    prot opt in     out     source        destination

Chain INPUT (policy ACCEPT 0 packets, 0 bytes)
rid  pkts bytes target    prot opt in     out     source        destination
0    78903 16M ACCEPT    all  --  lo      *        0.0.0.0/0      0.0.0.0/0
0      0 0 DROP      all  --  *      *        0.0.0.0/0      0.0.0.0/0      state INVALID
0    140K 9558K block_in  all  --  *      *        0.0.0.0/0      0.0.0.0/0
0    23789 1184K ACCEPT    all  --  *      *        0.0.0.0/0      0.0.0.0/0      state
RELATED,ESTABLISHED
0    116K 8374K usr_rules all  --  *      *        0.0.0.0/0      0.0.0.0/0
0      0 0 DROP      all  --  *      *        0.0.0.0/0      0.0.0.0/0

Chain FORWARD (policy ACCEPT 3146M packets, 4098G bytes)
rid  pkts bytes target    prot opt in     out     source        destination

Chain OUTPUT (policy ACCEPT 173K packets, 22M bytes)
rid  pkts bytes target    prot opt in     out     source        destination

Chain POSTROUTING (policy ACCEPT 0 packets, 0 bytes)
rid  pkts bytes target    prot opt in     out     source        destination
0    78903 16M ACCEPT    all  --  *      lo      0.0.0.0/0      0.0.0.0/0
0    679K 41M DROP      all  --  *      *        0.0.0.0/0      0.0.0.0/0      state INVALID
0    3146M 4098G block_out all  --  *      *        0.0.0.0/0      0.0.0.0/0
0      0 0 ACCEPT    all  --  *      *        0.0.0.0/0      0.0.0.0/0      PHYSDEV
match --physdev-in tap0 --physdev-out vNic_+
0      0 0 ACCEPT    all  --  *      *        0.0.0.0/0      0.0.0.0/0      PHYSDEV
match --physdev-in vNic_+ --physdev-out tap0
0      0 0 ACCEPT    all  --  *      *        0.0.0.0/0      0.0.0.0/0      PHYSDEV
match --physdev-in na+ --physdev-out vNic_+
0      0 0 ACCEPT    all  --  *      *        0.0.0.0/0      0.0.0.0/0      PHYSDEV
match --physdev-in vNic_+ --physdev-out na+
0    3145M 4098G ACCEPT    all  --  *      *        0.0.0.0/0      0.0.0.0/0      state
RELATED,ESTABLISHED
0    221K 13M usr_rules all  --  *      *        0.0.0.0/0      0.0.0.0/0
```

0	0	0	DROP	all	--	*	*	0.0.0.0/0	0.0.0.0/0	
Chain block_in (1 references)										
rid	pkts	bytes	target	prot	opt	in	out	source	destination	
Chain block_out (1 references)										
rid	pkts	bytes	target	prot	opt	in	out	source	destination	
Chain usr_rules (2 references)										
rid	pkts	bytes	target	prot	opt	in	out	source	destination	
131074	70104	5086K	ACCEPT	all	--	*	*	0.0.0.0/0	0.0.0.0/0	match-
set 0_131074-os-v4-1 src										
131075	116K	8370K	ACCEPT	all	--	*	*	0.0.0.0/0	0.0.0.0/0	match-
set 1_131075-ov-v4-1 dst										
131073	151K	7844K	ACCEPT	all	--	*	*	0.0.0.0/0	0.0.0.0/0	

在 `show firewall` 命令的 `POST_ROUTING` 部分中检查 `DROP invalid` 规则的递增值。典型的原因包括非对称路由问题，或基于 `TCP` 的应用程序已处于非活动状态多个小时。非对称路由问题的进一步证据包括：

- Ping 在一个方向上正常工作，而在另一个方向上失败
- Ping 正常工作，而 `TCP` 无法正常工作

2 收集 `show ipset` 命令输出。

```
nsxedge> show ipset
Name: 0_131074-os-v4-1
Type: bitmap:if (Interface Match)
Revision: 3
Header: range 0-64000
Size in memory: 8116
References: 1
Number of entries: 1
Members:
vse (vShield Edge Device)

Name: 0_131074-os-v6-1
Type: bitmap:if (Interface Match)
Revision: 3
Header: range 0-64000
Size in memory: 8116
References: 1
Number of entries: 1
Members:
vse (vShield Edge Device)

Name: 1_131075-ov-v4-1
Type: hash:oservice (Match un-translated Ports)
Revision: 2
Header: family inet hashsize 64 maxelem 65536
Size in memory: 704
References: 1
Number of entries: 2
Members:
```



```

Proto=6, DestPort=179, SrcPort=Any      (encoded: 0.6.0.179,0.6.0.0/16)
Proto=89, DestPort=Any, SrcPort=Any      (encoded: 0.89.0.0/16,0.89.0.0/16)

Name: 1_131075-ov-v6-1
Type: hash:oservice (Match un-translated Ports)
Revision: 2
Header: family inet hashsize 64 maxelem 65536
Size in memory: 704
References: 1
Number of entries: 2
Members:
Proto=89, DestPort=Any, SrcPort=Any      (encoded: 0.89.0.0/16,0.89.0.0/16)
Proto=6, DestPort=179, SrcPort=Any      (encoded: 0.6.0.179,0.6.0.0/16)

```

- 3 使用 REST API 或 Edge 用户界面在特定防火墙规则上启用日志记录，然后使用 `show log follow` 命令监视日志。

如果看不到日志，请使用以下 REST API 在 DROP Invalid 规则上启用日志记录。

```

URL : https://NSX_Manager_IP/api/4.0/edges/{edgeId}/firewall/config/global

PUT Method
Input representation
<globalConfig>    <!-- Optional -->
<tcpPickOngoingConnections>false</tcpPickOngoingConnections>    <!-- Optional. Defaults to false -->
<tcpAllowOutOfWindowPackets>false</tcpAllowOutOfWindowPackets>    <!-- Optional. Defaults to false -->
<tcpSendResetForClosedVsePorts>true</tcpSendResetForClosedVsePorts>    <!-- Optional. Defaults to true -->
<dropInvalidTraffic>true</dropInvalidTraffic>    <!-- Optional. Defaults to true -->
<logInvalidTraffic>true</logInvalidTraffic>    <!-- Optional. Defaults to false -->
<tcpTimeoutOpen>30</tcpTimeoutOpen>    <!-- Optional. Defaults to 30 -->
<tcpTimeoutEstablished>3600</tcpTimeoutEstablished>    <!-- Optional. Defaults to 3600 -->
<tcpTimeoutClose>30</tcpTimeoutClose>    <!-- Optional. Defaults to 30 -->
<udpTimeout>60</udpTimeout>    <!-- Optional. Defaults to 60 -->
<icmpTimeout>10</icmpTimeout>    <!-- Optional. Defaults to 10 -->
<icmp6Timeout>10</icmp6Timeout>    <!-- Optional. Defaults to 10 -->
<ipGenericTimeout>120</ipGenericTimeout>    <!-- Optional. Defaults to 120 -->
</globalConfig>
Output representation
No payload

```

使用 `show log follow` 命令查找类似下面的日志：

```

2016-04-18T20:53:31+00:00 edge-0 kernel: nf_ct_tcp: invalid TCP flag combination IN= OUT=
SRC=172.16.1.4 DST=192.168.1.4 LEN=40 TOS=0x00 PREC=0x00 TTL=64 ID=43343 PROTO=TCP
SPT=5050 DPT=80 SEQ=0 ACK=1572141176 WINDOW=512 RES=0x00 URG PSH FIN URGP=0
2016-04-18T20:53:31+00:00 edge-0 kernel: INVALID IN= OUT=vNic_1 SRC=172.16.1.4
DST=192.168.1.4 LEN=40 TOS=0x00 PREC=0x00 TTL=63 ID=43343 PROTO=TCP SPT=5050 DPT=80
WINDOW=512 RES=0x00 URG PSH FIN URGP=0

```

- 4 使用 `show flowtable rule_id` 命令在 Edge 防火墙状态表中查找匹配的连接。

```
nsxedge> show flowtable
1: tcp 6 21554 ESTABLISHED src=192.168.110.10 dst=192.168.5.3 sport=25981
d port=22 pkts=52 bytes=5432 src=192.168.5.3 dst=192.168.110.10 sport=22 dport=259
81 pkts=44 bytes=7201 [ASSURED] mark=0 rid=131073 use=1
2: tcp 6 21595 ESTABLISHED src=127.0.0.1 dst=127.0.0.1 sport=53194
dport=10 001 pkts=33334 bytes=11284650 src=127.0.0.1 dst=127.0.0.1 sport=10001 dport=5319
4 pkts=33324 bytes=1394146 [ASSURED] mark=0 rid=0 use=1
```

使用 `show flowstats` 命令将活动连接数与最大允许连接数进行比较：

```
nsxedge> show flowstats
Total Flow Capacity: 65536
Current Statistics :
cpu=0 searched=3280373 found=3034890571 new=52678 invalid=659946 ignore=77605
delete=52667 delete_list=49778 insert=49789 insert_failed=0 drop=0 early_drop=0
error=0 search_restart=0
```

- 5 使用 `show log follow` 命令检查 Edge 日志并查找任何 ALG 丢弃问题。搜索类似于 `tftp_alg`、`msrpc_alg` 或 `oracle_tns` 的字符串。有关其他信息，请参见：

- <https://kb.vmware.com/kb/2126674>
- <https://kb.vmware.com/kb/2137751>

Edge 路由连接问题

- 1 使用 `ping <destination_IP_address>` 命令从客户端中启动控制的流量。
- 2 在两个接口上同时捕获流量，将输出写入到一个文件中，然后使用 SCP 导出该文件。

例如：

使用以下命令在输入接口上捕获流量：

```
debug packet display interface vNic_0 -n_src_host_1.1.1.1
```

使用以下命令在输出接口上捕获流量：

```
debug packet display interface vNic_1 -n_src_host_1.1.1.1
```

对于同时的数据包捕获，请在 ESXi 中使用 ESXi 数据包捕获实用程序 `pktcap-uw` 工具。请参阅 <https://kb.vmware.com/kb/2051814>。

如果数据包丢弃一贯出现，请查找与以下内容相关的配置错误：

- IP 地址和路由
- 防火墙规则或 NAT 规则
- 非对称路由

- RP 筛选器检查

- a 使用 `show interface` 命令检查接口 IP/子网。
- b 如果在数据层面中缺少路由，请运行以下命令：
 - `show ip route`
 - `show ip route static`
 - `show ip route bgp`
 - `show ip route ospf`
- c 运行 `show ip forwarding` 命令以在路由表中查找所需的路由。
- d 如果具有多个路径，请运行 `show rpfilter` 命令。

```
nsxedge> show rpfilter
net.ipv4.conf.VDR.rp_filter = 0
net.ipv4.conf.all.rp_filter = 0
net.ipv4.conf.br-sub.rp_filter = 1
net.ipv4.conf.default.rp_filter = 1
net.ipv4.conf.lo.rp_filter = 0
net.ipv4.conf.vNic_0.rp_filter = 1
net.ipv4.conf.vNic_1.rp_filter = 1
net.ipv4.conf.vNic_2.rp_filter = 1
net.ipv4.conf.vNic_3.rp_filter = 1
net.ipv4.conf.vNic_4.rp_filter = 1
net.ipv4.conf.vNic_5.rp_filter = 1
net.ipv4.conf.vNic_6.rp_filter = 1
net.ipv4.conf.vNic_7.rp_filter = 1
net.ipv4.conf.vNic_8.rp_filter = 1
net.ipv4.conf.vNic_9.rp_filter = 1

nsxedge> show rpfstats
RPF drop packet count: 484
```

要查找 RPF 统计信息，请运行 `show rpfstats` 命令。

```
nsxedge> show rpfstats
RPF drop packet count: 484
```

如果数据包丢弃是随机出现的，请检查资源限制：

- a 对于 CPU 或内存使用率，请运行以下命令：
 - `show system cpu`
 - `show system memory`
 - `show system storage`
 - `show process monitor`
 - `top`

对于 ESXi，请运行 `esxtop n` 命令。

```
6:26:46pm up 28 days 20:01, 548 worlds, 3 VMs, 3 vCPUs; CPU load average: 0.14, 0.12, 0.12
PCPU USED(%): 7.2 32 AVG: 19
PCPU UTIL(%): 6.2 37 AVG: 21
```

ID	GID	NAME	NWLD	%USED	%RUN	%SYS	%WAIT	%VMWAIT	%RDY	%IDLE	%OVRP	%CSTP	%MLMTD	%SWPW
2	2	system	131	5.43	28.79	0.00	12908.50	-	35.03	0.00	24.03	0.00	0.00	0.00
88295638	88295638	esxtop.11413506	1	3.05	2.52	0.01	95.50	-	0.32	0.00	0.03	0.00	0.00	0.00
371958	371958	web-02a	6	1.18	0.84	0.34	588.66	0.00	0.27	97.90	0.00	0.00	0.00	0.00
368736	368736	web-01a	6	0.92	0.92	0.04	591.45	0.00	0.44	98.26	0.05	0.00	0.00	0.00
362728	362728	app-02a	6	0.60	0.62	0.01	589.15	0.89	0.23	96.68	0.00	0.00	0.00	0.00
14826	14826	netcpa.35043	21	0.30	0.30	0.00	2063.89	-	0.28	0.00	0.00	0.00	0.00	0.00
793	793	vmtoolsd.32996	5	0.28	0.27	0.00	491.39	-	0.16	0.00	0.00	0.00	0.00	0.00
8176	8176	hostd.34168	34	0.16	0.27	0.00	3340.26	-	0.42	0.00	0.00	0.00	0.00	0.00
19890	19890	vmtoolsd.35736	2	0.08	0.08	0.00	196.31	-	0.15	0.00	0.00	0.00	0.00	0.00
17967	17967	logchannellogge	1	0.07	0.07	0.00	98.31	-	0.05	0.00	0.00	0.00	0.00	0.00
6024	6024	storageRM.33890	1	0.07	0.01	0.05	98.36	-	0.00	0.00	0.00	0.00	0.00	0.00

较高的 CPU 占用率

如果 NSX Edge 上的 CPU 占用率较高，请在 ESXi 主机上使用 `esxtop` 命令验证设备的性能。请参阅以下知识库文章：

- <https://kb.vmware.com/kb/1008205>
- <https://kb.vmware.com/kb/1008014>
- <https://kb.vmware.com/kb/1010071>
- <https://kb.vmware.com/kb/2096171>

另请参见 <https://communities.vmware.com/docs/DOC-9279>。

较高的 `ksoftirqd` 进程值表示传入数据包率较高。检查是否在数据路径上启用日志记录，例如，为防火墙规则启用。运行 `show log follow` 命令以确定是否记录了大量的日志命中数。

NSX Manager 与 Edge 通信问题

NSX Manager 通过 VIX 或消息总线与 NSX Edge 通信。NSX Manager 在部署 Edge 时选择 VIX 或消息总线，并且从不发生变化。

VIX

- 如果未准备 ESXi 主机，则将 VIX 用于 NSX Edge。
- NSX Manager 先从 vCenter Server 中获取主机凭据以连接到 ESXi 主机。
- NSX Manager 使用 Edge 凭据登录到 Edge 设备。
- Edge 上的 `vmtoolsd` 进程处理 VIX 通信。

发生 VIX 故障的原因如下所示：

- NSX Manager 无法与 vCenter Server 通信。
- NSX Manager 无法与 ESXi 主机通信。
- 存在 NSX Manager 内部问题。
- 存在 Edge 内部问题。

VIX 调试

在 NSX Manager 日志中查找 VIX 错误 `VIX_E_<error>` 以缩小原因范围。查找类似下面的错误：

```
Vix Command 1126400 failed, reason com.vmware.vshield.edge.exception.VixException: vShield
Edge:10013:Error code 'VIX_E_FILE_NOT_FOUND' was returned by VIX API.:null

Health check failed for edge edge-13 VM vm-5025 reason:
com.vmware.vshield.edge.exception.VixException: vShield Edge:10013:Error code
'VIX_E_VM_NOT_RUNNING' was returned by VIX API.:null
```

通常，如果很多 Edge 同时发生相同的故障，则问题不是出在 Edge 上。

Edge 诊断

- 使用以下命令检查 `vmtoolsd` 是否正在运行。

```
nsxedge> show process list
Perimeter-Gateway-01-0> show process list
%CPU %MEM    VSZ   RSZ STAT  STARTED    TIME COMMAND
  0.0  0.1   4244   720 Ss     May 16 00:00:15 init [3]
...
  0.0  0.1   4240   640 S      May 16 00:00:00 logger -p daemon debug -t vserrdd
  0.2  0.9  57192  4668 S      May 16 00:23:07 /usr/local/bin/vmtoolsd --plugin-pa
  0.0  0.4   4304  2260 SLs   May 16 00:01:54 /usr/sbin/watchdog
...
```

- 运行以下命令以检查 Edge 是否处于正常状态：

```
nsxedge> show eventmgr
-----
messagebus      : enabled
debug           : 0
profiling       : 0
cfg_rx          : 1
cfg_rx_msgbus   : 0
...
```

此外，还可以使用 `show eventmgr` 命令验证是否收到并处理查询命令。

```
nsxedge> show eventmgr
-----
messagebus      : enabled
debug           : 0
profiling       : 0
cfg_rx          : 1
cfg_rx_msgbus   : 0
cfg_rx_err      : 0
cfg_exec_err    : 0
cfg_resp        : 0
```

```

cfg_resp_err    : 0
cfg_resp_ln_err: 0
fastquery_rx    : 0 fastquery_err : 0
clearcmd_rx     : 0
clearcmd_err    : 0
ha_rx           : 0
ha_rx_err       : 0
ha_exec_err     : 0
status_rx       : 16
status_rx_err   : 0
status_svr      : 10
status_evt      : 0
status_evt_push: 0
status_ha       : 0
status_ver      : 1
status_sys      : 5
status_cmd      : 0
status_svr_err  : 0
status_evt_err  : 0
status_sys_err  : 0
status_ha_err   : 0
status_ver_err  : 0
status_cmd_err  : 0
evt_report      : 1
evt_report_err  : 0
hc_report       : 10962
hc_report_err   : 0
cli_rx          : 2
cli_resp        : 1
cli_resp_err    : 0
counter_reset   : 0
----- Health Status -----
system status   : good
ha state        : active
cfg version     : 7
generation      : 0
server status   : 1
syslog-ng       : 1
haproxy         : 0
ipsec           : 0
sslvpn         : 0
l2vpn          : 0
dns             : 0
dhcp            : 0
heartbeat       : 0
monitor         : 0
gslb            : 0
----- System Events -----

```

如果 `vmtoolsd` 未运行或 `Edge` 处于错误状态，请重新引导 `Edge`。

您还可以检查 `Edge` 日志。请参阅 <https://kb.vmware.com/kb/2079380>。

消息总线调试

在准备 ESXi 主机时，将使用消息总线进行 NSX Edge 通信。在遇到问题时，NSX Manager 日志可能包含类似下面的条目：

```
GMT ERROR taskScheduler-6 PublishTask:963 - Failed to configure VSE-vm index 0, vm-id vm-117,
edge edge-5. Error: RPC request timed out
```

在以下情况下，将出现该问题：

- Edge 处于错误状态
- 消息总线连接中断

要在 Edge 上诊断该问题，请执行以下操作：

- 要检查 rmq 连接，请运行以下命令：

```
nsxedge> show messagebus messages
-----
Message bus is enabled
cmd conn state : listening
init_req       : 1
init_resp      : 1
init_req_err    : 0
...
```

- 要检查 vmci 连接，请运行以下命令：

```
nsxedge> show messagebus forwarder
-----
Forwarder Command Channel
vmci_conn       : up
app_client_conn : up
vmci_rx         : 3649
vmci_tx         : 3648
vmci_rx_err     : 0
vmci_tx_err     : 0
vmci_closed_by_peer: 8
vmci_tx_no_socket : 0
app_rx         : 3648
app_tx         : 3649
app_rx_err     : 0
app_tx_err     : 0
app_conn_req    : 1
app_closed_by_peer : 0
app_tx_no_socket : 0
-----
Forwarder Event Channel
vmci_conn       : up
app_client_conn : up
vmci_rx         : 1143
vmci_tx         : 13924
```

```

vmci_rx_err      : 0
vmci_tx_err      : 0
vmci_closed_by_peer: 0
vmci_tx_no_socket : 0
app_rx           : 13924
app_tx           : 1143
app_rx_err       : 0
app_tx_err       : 0
app_conn_req     : 1
app_closed_by_peer : 0
app_tx_no_socket : 0
-----
cli_rx           : 1
cli_tx           : 1
cli_tx_err       : 0
counters_reset   : 0

```

在该示例中，输出 `vmci_closed_by_peer: 8` 表示主机代理关闭连接的次数。如果该数字不断增加并且 `vmci conn` 为 `down`，则主机代理无法连接到 `RMQ` 代理。在 `show log follow` 中，在 `Edge` 日志中查找重复的错误：`VmciProxy: [daemon.debug] VMCi Socket is closed by peer`

要在 `ESXi` 主机上诊断该问题，请执行以下操作：

- 要检查 `ESXi` 主机是否连接到 `RMQ` 代理，请运行以下命令：

```

esxcli network ip connection list | grep 5671

```

tcp	0	0	10.32.43.4:43329	10.32.43.230:5671	ESTABLISHED	35854	newreno	vsfwd
tcp	0	0	10.32.43.4:52667	10.32.43.230:5671	ESTABLISHED	35854	newreno	vsfwd
tcp	0	0	10.32.43.4:20808	10.32.43.230:5671	ESTABLISHED	35847	newreno	vsfwd
tcp	0	0	10.32.43.4:12486	10.32.43.230:5671	ESTABLISHED	35847	newreno	vsfwd

显示数据包丢弃统计信息

从 `NSX for vSphere 6.2.3` 开始，您可以使用 `show packet drops` 命令显示以下位置的数据包丢弃统计信息：

- 接口
- 驱动程序
- L2
- L3
- 防火墙

要运行该命令，请登录到 `NSX Edge CLI` 并进入基本模式。有关详细信息，请参见《`NSX 命令行界面参考`》。例如：

```

show packet drops

vShield Edge Packet Drop Stats:

Driver Errors

```



```
=====
              TX      TX      TX      RX      RX      RX
Interface Dropped Error Ring Full Dropped Error Out Of Buf
vNic_0      0        0      0      0      0      0
vNic_1      0        0      0      0      0      0
vNic_2      0        0      0      0      0      2
vNic_3      0        0      0      0      0      0
vNic_4      0        0      0      0      0      0
vNic_5      0        0      0      0      0      0
```

Interface Drops

```
=====
Interface RX Dropped TX Dropped
vNic_0              4          0
vNic_1             2710          0
vNic_2              0          0
vNic_3              2          0
vNic_4              2          0
vNic_5              2          0
```

L2 RX Errors

```
=====
Interface length crc frame fifo missed
vNic_0          0  0      0  0      0
vNic_1          0  0      0  0      0
vNic_2          0  0      0  0      0
vNic_3          0  0      0  0      0
vNic_4          0  0      0  0      0
vNic_5          0  0      0  0      0
```

L2 TX Errors

```
=====
Interface aborted fifo window heartbeat
vNic_0          0  0      0      0
vNic_1          0  0      0      0
vNic_2          0  0      0      0
vNic_3          0  0      0      0
vNic_4          0  0      0      0
vNic_5          0  0      0      0
```

L3 Errors

```
=====
IP:
  ReasmFails : 0
  InHdrErrors : 0
  InDiscards : 0
  FragFails : 0
  InAddrErrors : 0
  OutDiscards : 0
  OutNoRoutes : 0
  ReasmTimeout : 0
ICMP:
  InTimeExcds : 0
  InErrors : 227
  OutTimeExcds : 0
```

```

OutDestUnreachs : 152
OutParmProbs : 0
InSrcQuenchs : 0
InRedirects : 0
OutSrcQuenchs : 0
InDestUnreachs : 151
OutErrors : 0
InParmProbs : 0

Firewall Drop Counters
=====

Ipv4 Rules
=====
Chain - INPUT
rid pkts bytes target prot opt in out source destination
0 119 30517 DROP all -- * * 0.0.0.0/0 0.0.0.0/0 state INVALID
0 0 0 DROP all -- * * 0.0.0.0/0 0.0.0.0/0
Chain - POSTROUTING
rid pkts bytes target prot opt in out source destination
0 101 4040 DROP all -- * * 0.0.0.0/0 0.0.0.0/0 state INVALID
0 0 0 DROP all -- * * 0.0.0.0/0 0.0.0.0/0

Ipv6 Rules
=====
Chain - INPUT
rid pkts bytes target prot opt in out source destination
0 0 0 DROP all * * ::/0 ::/0 state INVALID
0 0 0 DROP all * * ::/0 ::/0
Chain - POSTROUTING
rid pkts bytes target prot opt in out source destination
0 0 0 DROP all * * ::/0 ::/0 state INVALID
0 0 0 DROP all * * ::/0 ::/0

```

管理 NSX Edge 时的预期行为

在 vSphere Web Client 中，在 NSX Edge 上配置 L2 VPN 以及添加、移除或修改站点配置详细信息时，该操作将导致断开并重新连接所有现有的连接。这是预期的行为。

例如：

```
nsxmgr> show cluster all
No. Cluster Name Cluster Id Datacenter Name Firewall Status
1 Compute Cluster A domain-c33 Datacenter Site A Enabled
2 Management & Edge Cluster domain-c41 Datacenter Site A Enabled

nsxmgr> show cluster domain-c33
Datacenter: Datacenter Site A
Cluster: Compute Cluster A
No. Host Name Host Id Installation Status
1 esx-02a.corp.local host-32 Enabled
2 esx-01a.corp.local host-28 Enabled

nsxmgr> show host host-28
Datacenter: Datacenter Site A
Cluster: Compute Cluster A
Host: esx-01a.corp.local
No. VM Name VM Id Power Status
1 web-02a vm-219 on
2 web-01a vm-216 on
3 win8-01a vm-206 off
4 app-02a vm-264 on

nsxmgr> show vm vm-264
Datacenter: Datacenter Site A
Cluster: Compute Cluster A
Host: esx-01a.corp.local
Host-ID: host-28
VM: app-02a
Virtual Nics List:
1.
Vnic Name app-02a - Network adapter 1
Vnic Id 502ef2fa-62cf-d178-cb1b-c825fb300c84.000
Filters nic-79396-eth0-vmware-sfw.2

nsxmgr> show dfw vnic 502ef2fa-62cf-d178-cb1b-c825fb300c84.000
Vnic Name app-02a - Network adapter 1
Vnic Id 502ef2fa-62cf-d178-cb1b-c825fb300c84.000
Mac Address 00:50:56:ae:6c:6b
Port Group Id dvportgroup-385
Filters nic-79396-eth0-vmware-sfw.2

nsxmgr> show dfw host host-28 filter nic-79396-eth0-vmware-sfw.2 rules
ruleset domain-c33 {
# Filter rules
rule 1012 at 1 inout protocol any from addrset ip-securitygroup-10 to addrset ip-securitygroup-10 drop
with log;
rule 1013 at 2 inout protocol any from addrset src1013 to addrset src1013 drop;
rule 1011 at 3 inout protocol tcp from any to addrset dst1011 port 443 accept;
rule 1011 at 4 inout protocol icmp icmptype 8 from any to addrset dst1011 accept;
rule 1010 at 5 inout protocol tcp from addrset ip-securitygroup-10 to addrset ip-securitygroup-11 port
8443 accept;
rule 1010 at 6 inout protocol icmp icmptype 8 from addrset ip-securitygroup-10 to addrset ip-
securitygroup-11 accept;
```

```

    rule 1009 at 7 inout protocol tcp from addrset ip-securitygroup-11 to addrset ip-securitygroup-12 port
3306 accept;
    rule 1009 at 8 inout protocol icmp icmp-type 8 from addrset ip-securitygroup-11 to addrset ip-
securitygroup-12 accept;
    rule 1003 at 9 inout protocol ipv6-icmp icmp-type 136 from any to any accept;
    rule 1003 at 10 inout protocol ipv6-icmp icmp-type 135 from any to any accept;
    rule 1002 at 11 inout protocol udp from any to any port 67 accept;
    rule 1002 at 12 inout protocol udp from any to any port 68 accept;
    rule 1001 at 13 inout protocol any from any to any accept;
}

ruleset domain-c33_L2 {
    # Filter rules
    rule 1004 at 1 inout ethertype any from any to any accept;
}

```

Distributed Firewall 故障排除

本主题提供了有助于您了解 VMware NSX 6.x Distributed Firewall (DFW) 和进行故障排除的信息。

问题

发布 Distributed Firewall 规则失败。

更新 Distributed Firewall 规则失败。

原因

NSX Distributed Firewall 是嵌入在管理程序内核中的防火墙，可以查看和控制虚拟化的工作负载和网络。您可以根据 VMware vCenter 对象（如数据中心和群集）、虚拟机名称和标记、网络结构（如 IP/VLAN/VXLAN 地址）以及 Active Directory 中的用户组标识创建访问控制策略。现在，在通过 vMotion 在不同物理主机之间移动虚拟机时，将会强制实施一致的访问控制策略，而无需重写防火墙规则。由于 Distributed Firewall 嵌入在管理程序中，它可以提供接近线路速率的吞吐量，从而在物理服务器上支持更高的工作负载整合。防火墙的分布式特性使架构具有向外扩展性，可在向数据中心添加更多主机时自动扩展防火墙功能。

ESXi 主机上的 NSX Manager Web 应用程序和 NSX 组件通过 RabbitMQ 代理进程相互通信，该进程在与 NSX Manager Web 应用程序相同的虚拟机上运行。使用的通信协议为 AMQP（高级消息队列协议），并使用 SSL 保护通道安全。在 ESXi 主机上，VSFW（vShield 防火墙守护程序）进程建立并维护到代理的 SSL 连接，并代表其他组件发送和接收消息，这些组件通过 IPC 与其进行通信。

验证下面的每个故障排除步骤是否适用于您的环境。每个步骤提供了相应说明或指向文档的链接，以消除可能的根源并在必要时采取纠正措施。这些步骤按最适当的顺序进行排列，以查找问题并确定相应的解决方案。在执行每个步骤后，再次尝试更新/发布 Distributed Firewall 规则。

解决方案

1 验证是否满足运行 Distributed Firewall (DFW) 的必备条件。

- VMware vCenter Server 5.5
- VMware ESXi 5.1 或 ESXi 5.5

- VMware NSX 6.0 和更高版本

- 2 验证是否在群集中的每个 ESXi 主机上成功安装 DFW VIB。为此，在群集中的每个 ESXi 主机上运行以下命令。

例如：

```
# esxcli software vib list | grep esx-vsip

esx-vsip                    5.5.0-0.0.2318233  VMware  VMwareCertified  2015-01-24

# esxcli software vib list | grep dvfilter

esx-dvfilter-switch-security 5.5.0-0.0.2318233  VMware  VMwareCertified  2015-01-24
```

- 3 在 ESXi 主机上，验证 vShield-Stateful-Firewall 服务是否处于运行状态。

例如：

```
# /etc/init.d/vShield-Stateful-Firewall status

vShield-Stateful-Firewall is running
```

- 4 验证消息总线是否正确与 NSX Manager 通信。

该进程是由监视程序脚本自动启动的；如果由于未知原因终止，将重新启动该进程。在群集中的每个 ESXi 主机上运行以下命令。

例如：

```
# ps | grep vsfwd

107557 107557 vsfwd /usr/lib/vmware/vsfw/vsfwd
107574 107557 vsfwd /usr/lib/vmware/vsfw/vsfwd
107575 107557 vsfwd /usr/lib/vmware/vsfw/vsfwd
107576 107557 vsfwd /usr/lib/vmware/vsfw/vsfwd
107577 107557 vsfwd /usr/lib/vmware/vsfw/vsfwd
107578 107557 vsfwd /usr/lib/vmware/vsfw/vsfwd
```

5 验证是否在防火墙配置中打开端口 5671 以进行通信。

以下命令显示到 RabbitMQ 代理的 VSFWD 连接。请在 ESXi 主机上运行以下命令，以查看从 ESXi 主机上的 vsfwd 进程到 NSX Manager 的连接列表。确保在环境中的任何外部防火墙上打开端口 5671 以进行通信。此外，在端口 5671 上应具有至少两个连接。可能在端口 5671 上具有更多连接，因为在 ESXi 主机上部署的 NSX Edge 虚拟机也会建立到 RMQ 代理的连接。

例如：

```
# esxcli network ip connection list |grep 5671

tcp          0      0 192.168.110.51:30133      192.168.110.15:5671    ESTABLISHED  10949155
newreno      vsfwd
tcp          0      0 192.168.110.51:39156      192.168.110.15:5671    ESTABLISHED  10949155
newreno      vsfwd
```

6 验证是否配置了 VSFWD。

以下命令应显示 NSX Manager IP 地址。

```
# esxcfg-advcfg -g /UserVars/RmqIpAddress
```

7 如果在该 ESXi 主机中使用主机配置文件，请确认未在主机配置文件中设置 RabbitMQ 配置。

请参见：

- <https://kb.vmware.com/kb/2092871>
- <https://kb.vmware.com/kb/2125901>

8 验证 ESXi 主机的 RabbitMQ 凭据是否与 NSX Manager 不同步。请下载 NSX Manager 技术支持日志。在收集所有 NSX Manager 技术支持日志后，在所有日志中搜索类似下面的条目：

将 host-420 替换为可疑主机的 mo-id。

```
PLAIN login refused: user 'uw-host-420' - invalid credentials.
```

9 如果在可疑 ESXi 主机的日志中找到这些条目，请重新同步消息总线。

要重新同步消息总线，请使用 REST API。为了更好地了解该问题，请在重新同步消息总线后立即收集日志。

```
HTTP Method : POST
Headers ,
Authorization : base64encoded value of username password
Accept : application/xml
Content-Type : application/xml
Request:

POST https://NSX_Manager_IP/api/2.0/nwfabric/configure?action=synchronize

Request Body:

<nwFabricFeatureConfig>
```

```
<featureId>com.vmware.vshield.vsm.messagingInfra</featureId>
<resourceConfig>
<resourceId>{HOST/CLUSTER M0ID}</resourceId>
</resourceConfig>
</nwFabricFeatureConfig>
```

- 10** 使用 `export host-tech-support <host-id> scp <uid@ip:/path>` 命令收集主机特定的防火墙日志。

例如：

```
nsxmgr# export host-tech-support host-28 scp Administrator@192.168.110.10
Generating logs for Host: host-28...
```

- 11** 使用 `show dfw host host-id summarize-dvfilter` 命令验证是否在主机上部署了防火墙规则并将其应用于虚拟机。

在输出中，`module: vsip` 显示已加载 DFW 模块并正在运行。`name` 显示在每个 vNic 上运行的防火墙。

您可以运行 `show dfw cluster all` 命令以获取群集域 ID，然后运行 `show dfw cluster domain-id` 以获取主机 ID。

例如：

```
# show dfw host host-28 summarize-dvfilter

Fastpaths:
agent: dvfilter-faulter, refCount: 1, rev: 0x1010000, apiRev: 0x1010000, module: dvfilter
agent: ESXi-Firewall, refCount: 5, rev: 0x1010000, apiRev: 0x1010000, module: esxfw
agent: dvfilter-generic-vmware, refCount: 1, rev: 0x1010000, apiRev: 0x1010000, module: dvfilter-
generic-fastpath
agent: dvfilter-generic-vmware-swsec, refCount: 4, rev: 0x1010000, apiRev: 0x1010000, module:
dvfilter-switch-security
agent: bridgelearningfilter, refCount: 1, rev: 0x1010000, apiRev: 0x1010000, module: vdrb
agent: dvfg-igmp, refCount: 1, rev: 0x1010000, apiRev: 0x1010000, module: dvfg-igmp
agent: vmware-sfw, refCount: 4, rev: 0x1010000, apiRev: 0x1010000, module: vsip

Slowpaths:

Filters:
world 342296 vmm0:2-vm_RHEL63_srv_64-shared-846-3f435476-8f54-4e5a-8d01-59654a4e9979 vcUuid:'3f 43
54 76 8f 54 4e 5a-8d 01 59 65 4a 4e 99 79'
port 50331660 2-vm_RHEL63_srv_64-shared-846-3f435476-8f54-4e5a-8d01-59654a4e9979.eth1
vNic slot 2
  name: nic-342296-eth1-vmware-sfw.2
  agentName: vmware-sfw
  state: IOChain Attached
  vmState: Detached
  failurePolicy: failClosed
  slowPathID: none
  filter source: Dynamic Filter Creation
vNic slot 1
  name: nic-342296-eth1-dvfilter-generic-vmware-swsec.1
  agentName: dvfilter-generic-vmware-swsec
```



```

state: IOChain Attached
vmState: Detached
failurePolicy: failClosed
slowPathID: none
filter source: Alternate Opaque Channel
port 50331661 (disconnected)
vNic slot 2
name: nic-342296-eth2-vmware-sfw.2
agentName: vmware-sfw          <===== DFW filter
state: IOChain Detached
vmState: Detached
failurePolicy: failClosed
slowPathID: none
filter source: Dynamic Filter Creation
port 33554441 2-vm-RHEL63_srv_64-shared-846-3f435476-8f54-4e5a-8d01-59654a4e9979
vNic slot 2
name: nic-342296-eth0-vmware-sfw.2
agentName: vmware-sfw          <===== DFW filter
state: IOChain Attached
vmState: Detached
failurePolicy: failClosed
slowPathID: none
filter source: Dynamic Filter Creation

```

12 运行 show dfw host hostID filter filterID rules 命令。

例如：

```

# show dfw host host-28 filter nic-79396-eth0-vmware-sfw.2 rules

ruleset domain-c33 {
  # Filter rules
  rule 1012 at 1 inout protocol any from addrset ip-securitygroup-10 to addrset ip-securitygroup-10
  drop with log;
  rule 1013 at 2 inout protocol any from addrset src1013 to addrset src1013 drop;
  rule 1011 at 3 inout protocol tcp from any to addrset dst1011 port 443 accept;
  rule 1011 at 4 inout protocol icmp icmptype 8 from any to addrset dst1011 accept;
  rule 1010 at 5 inout protocol tcp from addrset ip-securitygroup-10 to addrset ip-securitygroup-11
  port 8443 accept;
  rule 1010 at 6 inout protocol icmp icmptype 8 from addrset ip-securitygroup-10 to addrset ip-
  securitygroup-11 accept;
  rule 1009 at 7 inout protocol tcp from addrset ip-securitygroup-11 to addrset ip-securitygroup-12
  port 3306 accept;
  rule 1009 at 8 inout protocol icmp icmptype 8 from addrset ip-securitygroup-11 to addrset ip-
  securitygroup-12 accept;
  rule 1003 at 9 inout protocol ipv6-icmp icmptype 136 from any to any accept;
  rule 1003 at 10 inout protocol ipv6-icmp icmptype 135 from any to any accept;
  rule 1002 at 11 inout protocol udp from any to any port 67 accept;
  rule 1002 at 12 inout protocol udp from any to any port 68 accept;
  rule 1001 at 13 inout protocol any from any to any accept;
}

```

```
ruleset domain-c33_L2 {
  # Filter rules
  rule 1004 at 1 inout ethertype any from any to any accept;
```

- 13** 运行 `show dfw host hostID filter filterID addrsets` 命令。

例如：

```
# show dfw host host-28 filter nic-342296-eth2-vmware-sfw.2 addrsets

addrset dst1011 {
ip 172.16.10.10,
ip 172.16.10.11,
ip 172.16.10.12,
ip fe80::250:56ff:feae:3e3d,
ip fe80::250:56ff:feae:f86b,
}
addrset ip-securitygroup-10 {
ip 172.16.10.11,
ip 172.16.10.12,
ip fe80::250:56ff:feae:3e3d,
ip fe80::250:56ff:feae:f86b,
}
addrset ip-securitygroup-11 {
ip 172.16.20.11,
ip fe80::250:56ff:feae:23b9,
}
addrset ip-securitygroup-12 {
ip 172.16.30.11,
ip fe80::250:56ff:feae:d42b,
}
addrset src1013 {
ip 172.16.10.12,
ip 172.17.10.11,
ip fe80::250:56ff:feae:cf88,
ip fe80::250:56ff:feae:f86b,
}
```

- 14** 如果已验证上面的每个故障排除步骤，并且无法将防火墙规则发布到主机虚拟机中，请通过 NSX Manager UI 或以下 REST API 调用执行主机级别强制同步。

```
URL : [https:]https://<nsx-mgr-ip>/api/4.0/firewall/forceSync/<host-id>
HTTP Method : POST
Headers ,
Authorization : base64encoded value of username password
Accept : application/xml
Content-Type : application/xml
```

备注：

- 如果防火墙规则不使用 IP 地址，请确保 VMware Tools 正在虚拟机上运行。有关详细信息，请参见 <https://kb.vmware.com/kb/2084048>。

VMware NSX 6.2.0 引入了一个使用 DHCP 侦听或 ARP 侦听获悉虚拟机 IP 地址的选项。通过使用这些新的发现机制，NSX 可以在未安装 VMware Tools 的虚拟机上强制实施基于 IP 地址的安全规则。有关详细信息，请参见 NSX 6.2.0 发行说明。

在完成主机准备过程后，将会立即激活 DFW。如果虚拟机根本不需要使用 DFW 服务，可以将其添加到排除列表功能中（默认情况下，自动从 DFW 功能中排除 NSX Manager、NSX Controller 和 Edge 服务网关）。在 DFW 中创建“全部拒绝”规则后，可能会阻止 vCenter Server 访问。有关详细信息，请参见 <https://kb.vmware.com/kb/2079620>。

- 在请求 VMware 技术支持部门排除 VMware NSX 6.x Distributed Firewall (DFW) 故障时，需要提供以下信息：
 - 群集中的每个 ESXi 主机上的 `show dfw host hostID summarize-dvfilter` 命令输出。
 - 从网络和安全 > 防火墙 > 常规 (Networking and Security > Firewall > General) 选项卡中选择 Distributed Firewall 配置，然后单击导出配置 (Export Configuration)。这会将 Distributed Firewall 配置导出为 XML 格式。
 - NSX Manager 日志。有关详细信息，请参见 <https://kb.vmware.com/kb/2074678>。
 - vCenter Server 日志。有关详细信息，请参见 <https://kb.vmware.com/kb/1011641>。

负载均衡

NSX Edge 负载均衡器使网络流量可以沿着多个路径流向特定目标。它将入站服务请求均匀分布在多个服务器中，从方式上确保负载分配对用户透明。可以在 **NSX** 中配置两种类型的负载均衡服务：单臂模式（也称为代理模式）或内嵌模式（也称为透明模式）。

NSX 负载均衡功能如下所示：

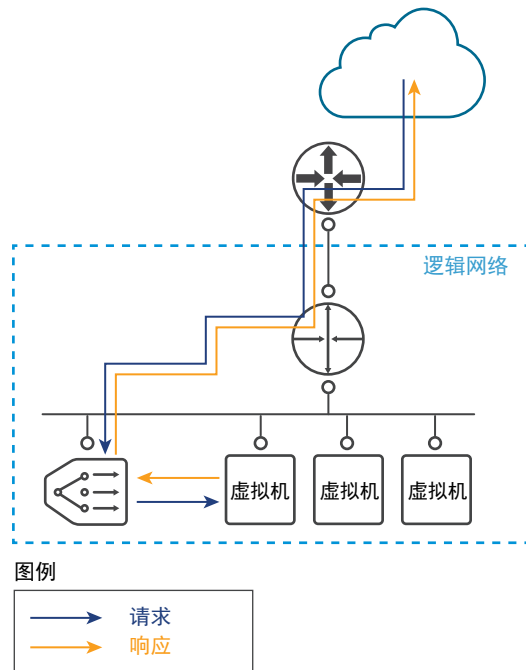
- 协议：TCP、HTTP、HTTPS
- 算法：加权循环、IP 哈希、URI、最少连接
- 具有 AES-NI 加速的 SSL 终止
- SSL 桥接（客户端 SSL + 服务器端 SSL）
- SSL 证书管理
- 用于客户端识别的 X 标头转发
- L4/L7 透明模式
- 连接限制
- 启用/禁用单个服务器（池成员）以进行维护
- 运行状况检查方法（TCP、HTTP、HTTPS）
- 增强的运行状况检查监控
- 保持/粘性方法：SourceIP、MSRDP、COOKIE、SSLSESSIONID
- 单臂模式
- URL 重写和重定向
- 用于 iRule 类型流量调整或内容交换的应用程序规则
- 用于 L7 代理负载均衡的 HA 会话粘性支持
- IPv6 支持
- 用于故障排除的增强负载均衡器 CLI
- 适用于所有类型的 Edge
- 用于高性能 SLB 的优化超大类型

本章讨论了以下主题：

- 场景：配置单臂负载均衡器
- 使用 UI 的负载均衡器故障排除
- 使用 CLI 的负载均衡器故障排除
- 常见的负载均衡器问题

场景：配置单臂负载均衡器

可以将 Edge 服务网关 (ESG) 视为传入客户端流量的代理。



在代理模式下，负载均衡器将自己的 IP 地址作为源地址，以将请求发送到后端服务器。后端服务器查看从负载均衡器中发送的所有流量，并直接响应负载均衡器。这种模式也称为 **SNAT** 模式或非透明模式。

典型的 **NSX** 单臂负载均衡器部署在具有其后端服务器的相同子网上，与逻辑路由器分开。**NSX** 负载均衡器虚拟服务器侦听虚拟 IP 以查找来自客户端的传入请求，并将这些请求发送到后端服务器。对于返回流量，需要使用反向 **NAT** 以将源 IP 地址从后端服务器更改为虚拟 IP (**VIP**) 地址，然后将虚拟 IP 地址发送到客户端。如果不执行该操作，到客户端的连接将中断。

在 **ESG** 收到流量后，它执行两个操作：目标网络地址转换 (**DNAT**) 以将 **VIP** 地址更改为某个负载均衡计算机的 IP 地址；以及源网络地址转换 (**SNAT**) 以将客户端 IP 地址与 **ESG** IP 地址调换。

然后，**ESG** 服务器将流量发送到负载均衡服务器，负载均衡服务器将响应发送回 **ESG**，然后发送回客户端。该选项比内嵌模式容易配置得多，但具有两个潜在问题。第一个问题是，该模式需要使用专用的 **ESG** 服务器，第二个问题是，负载均衡器服务器不知道原始客户端 IP 地址。**HTTP/HTTPS** 应用程序的一种解决方法是，在 **HTTP** 应用程序配置文件中启用“插入 **X-Forwarded-For**”，以便在发送到后端服务器的请求的 **X-Forwarded-For** **HTTP** 标头中包含客户端 IP 地址。

如果 HTTP/HTTPS 以外的应用程序要求在后端服务器上看到客户端 IP 地址，您可以将 IP 池配置为透明的。如果客户端没有位于与后端服务器相同的子网上，建议使用内嵌模式。否则，您必须将负载均衡器 IP 地址作为后端服务器的默认网关。

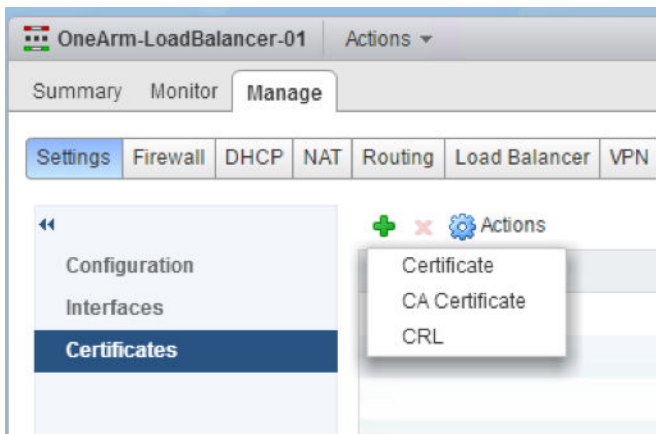
注 通常，可以使用两种方法确保连接完整性：

- SNAT/代理/非透明模式（如上所述）
- 直接服务器返回 (DSR)

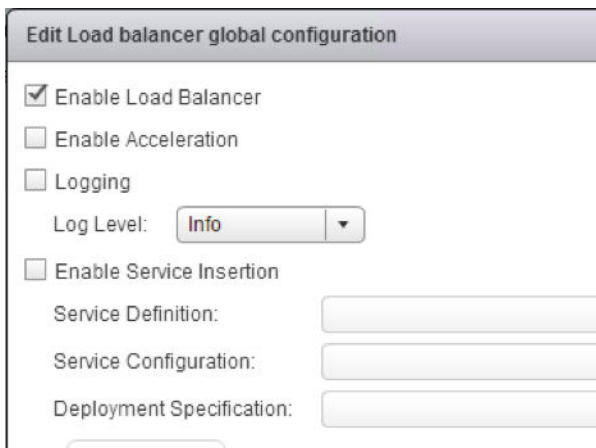
在 DSR 模式下，后端服务器直接响应客户端。目前，NSX 负载均衡器不支持 DSR。

步骤

- 1 双击 Edge，然后选择**管理 > 设置 > 证书 (Manage > Settings > Certificate)**以创建一个证书。



- 2 选择**管理 > 负载均衡器 > 全局配置 > 编辑 (Manage > Load Balancer > Global Configuration > Edit)**以启用负载均衡器服务。



- 3 选择**管理 > 负载均衡器 > 应用程序配置文件 (Manage > Load Balancer > Application Profiles)**以创建一个 HTTPS 应用程序配置文件。

New Profile

Name:

Type:

☐ Enable SSL Passthrough

HTTP Redirect URL:

Persistence:

Cookie Name:

Mode:

Expires in (Seconds):

☐ Insert X-Forwarded-For HTTP header

☐ Enable Pool Side SSL

Virtual Server Certifica... **Pool Certificates**

Service Certificates CA Certificates CRL

☒ Configure Service Certificate

	Common Name	Issuer	Validity
<input checked="" type="radio"/>	VSM_SOLUTION_71	VSM_SOLUTION_71	Tue Sep 8 2015 - Thu
<input type="radio"/>	VSM_SOLUTION_71	VSM_SOLUTION_71	Tue Sep 8 2015 - Thu

注 上面的屏幕截图使用自签名证书以仅用于说明目的。

- 4 (可选) 单击**管理 > 负载均衡器 > 服务监控 (Manage > Load Balancer > Service Monitoring)**, 然后编辑默认服务监控以将其从基本 HTTP/HTTPS 更改为特定的 URL/URI (如果需要)。

5 选择**管理 > 负载均衡器 > 池 (Manage > Load Balancer > Pools)**以创建服务器池。

要使用 **SNAT** 模式，请在池配置中取消选中**透明 (Transparent)**复选框。

Edit Pool

Name: * Web-Tier-Pool-01

Description:

Algorithm: ROUND-ROBIN

Algorithm Parameters:

Monitors: default_https_monitor

Members:

Enabled	Name	IP Address / VC Container	Weight	Monitor Port	Port	Max Connections	Min Connections
✓	web-01a	172.16.10.11	1	443	443	0	0
✓	web-02a	172.16.10.12	1	443	443	0	0

☐ Transparent

OK Cancel

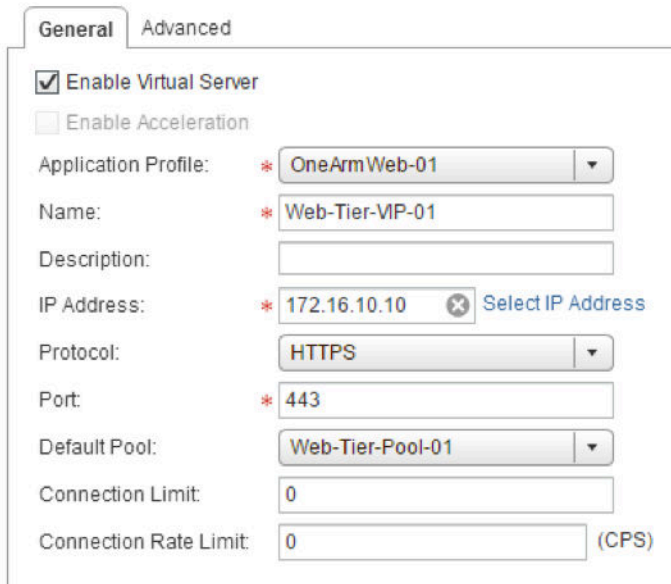
确保已列出并启用虚拟机。

6 （可选）单击**管理 > 负载均衡器 > 池 > 显示池统计信息 (Manage > Load Balancer > Pools > Show Pool Statistics)**以检查状态。

确保成员处于已启动状态。

- 7 选择**管理 > 负载均衡器 > 虚拟服务器 (Manage > Load Balancer > Virtual Servers)**以创建一个虚拟服务器。

如果要将 L4 负载均衡器用于 UDP 或更高性能的 TCP，请选中**启用加速 (Enable Acceleration)**。如果选中**启用加速 (Enable Acceleration)**，请确保负载均衡器 NSX Edge 上的防火墙状态为已启用 (**Enabled**)，因为 L4 SNAT 需要使用防火墙。



General Advanced

☒ Enable Virtual Server

☐ Enable Acceleration

Application Profile: * OneArmWeb-01

Name: * Web-Tier-VIP-01

Description:

IP Address: * 172.16.10.10 Select IP Address

Protocol: HTTPS

Port: * 443

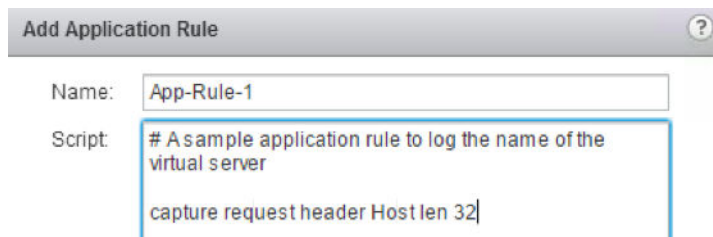
Default Pool: Web-Tier-Pool-01

Connection Limit: 0

Connection Rate Limit: 0 (CPS)

确保 IP 地址绑定到服务器池。

- 8 (可选) 如果使用一个应用程序规则，请在**管理 > 负载均衡器 > 应用程序规则 (Manage > Load Balancer > Application Rules)**中检查配置。



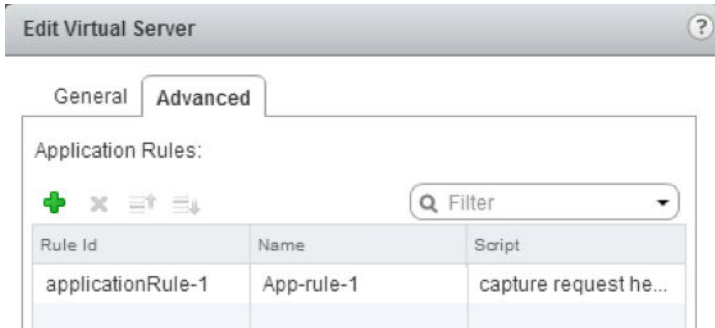
Add Application Rule ?

Name: App-Rule-1

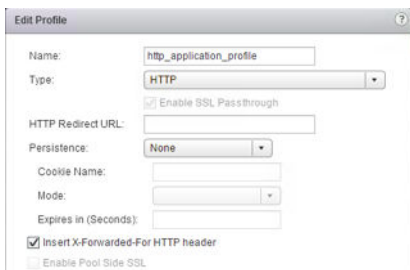
Script: # A sample application rule to log the name of the virtual server
capture request header Host len 32

- 9 如果使用一个应用程序规则，请在**管理 > 负载均衡器 > 虚拟服务器 > 高级 (Manage > Load Balancer > Virtual Servers > Advanced)**中确保该应用程序规则与虚拟服务器相关联。

有关支持的示例，请参见：<https://communities.vmware.com/docs/DOC-31772>。



在非透明模式下，后端服务器无法看到客户端 IP，但可以看到负载均衡器内部 IP 地址。作为 HTTP/HTTPS 流量的解决方法，请选中**插入 X-Forwarded-For HTTP 标头 (Insert X-Forwarded-For HTTP header)**。如果选中该选项，Edge 负载均衡器将添加 “X-Forwarded-For” 标头并且值为客户端源 IP 地址。



使用 UI 的负载均衡器故障排除

您可以使用 UI 执行某些负载均衡器故障排除。

问题

故障排除未正常工作。

解决方案

- 1 通过 UI 验证配置。
- 2 通过 UI 检查池成员状态。
- 3 确保其他服务（如 SSL VPN）未使用默认 HTTP/HTTPS 端口 80/443。
- 4 检查成员端口和监控端口配置。

不正确的配置可能会导致运行状况检查失败。

- 5 如果使用第 4 层负载均衡器引擎，请确保：
 - a 流量使用 TCP 协议。
 - b 未配置保持或第 7 层设置。
 - c 在负载均衡器全局配置中将启用加速 (**Enable Acceleration**) 设置为“有效”。
- 6 如果池处于透明（内嵌）模式，请确保 Edge 位于返回路径中。如果虚拟工作负载的默认网关指向负载均衡器 ESG 以外的 ESG，Edge 可能位于返回路径之外。

使用 CLI 的负载均衡器故障排除

您可以使用 NSX CLI 执行某些负载均衡器故障排除。

问题

负载均衡未正常工作。

解决方案

- 1 显示配置和统计信息。

```
nsxedge> show configuration loadbalancer
nsxedge> show configuration loadbalancer virtual [virtual-server-name]
nsxedge> show configuration loadbalancer pool [pool-name]
nsxedge> show configuration loadbalancer monitor [monitor-name]
nsxedge> show configuration loadbalancer profile [profile-name]
nsxedge> show configuration loadbalancer rule [rule-name]
```

- 2 检查负载均衡器引擎状态 (L4/L7)。

```
nsxedge> show service loadbalancer
haIndex:          0
-----
Loadbalancer Services Status:

L7 Loadbalancer      : running
-----
L7 Loadbalancer Statistics:
STATUS   PID      MAX_MEM_MB MAX SOCK  MAX_CONN  MAX_PIPE  CUR_CONN  CONN_RATE  CONN_RATE_LIMIT
MAX_CONN_RATE
running  1580      0          2081    1024      0         0         0         0           0
-----
L4 Loadbalancer Statistics:
MAX_CONN  ACT_CONN  INACT_CONN TOTAL_CONN
0         0         0         0
-----
Prot LocalAddress:Port Scheduler Flags
-> RemoteAddress:Port      Forward Weight ActiveConn InActConn
```

3 检查负载均衡器池状态 (L4/L7)。

```
nsxedge> show service loadbalancer pool

-----
Loadbalancer Pool Statistics:

POOL Web-Tier-Pool-01
| LB METHOD round-robin
| LB PROTOCOL L7
| Transparent disabled
| SESSION (cur, max, total) = (0, 0, 0)
| BYTES in = (0), out = (0)
+-->POOL MEMBER: Web-Tier-Pool-01/web-01a, STATUS: UP
| | HEALTH MONITOR = BUILT-IN, default_https_monitor:L7OK
| | | LAST STATE CHANGE: 2016-05-16 07:02:00
| | SESSION (cur, max, total) = (0, 0, 0)
| | BYTES in = (0), out = (0)
+-->POOL MEMBER: Web-Tier-Pool-01/web-02a, STATUS: UP
| | HEALTH MONITOR = BUILT-IN, default_https_monitor:L7OK
| | | LAST STATE CHANGE: 2016-05-16 07:02:01
| | SESSION (cur, max, total) = (0, 0, 0)
| | BYTES in = (0), out = (0)
```

4 检查负载均衡器对象统计信息 (VIP、池、成员)。

指定虚拟服务器的名称。

```
nsxedge> show service loadbalancer virtual Web-Tier-VIP-01

-----
Loadbalancer VirtualServer Statistics:

VIRTUAL Web-Tier-VIP-01
| ADDRESS [172.16.10.10]:443
| SESSION (cur, max, total) = (0, 0, 0)
| RATE (cur, max, limit) = (0, 0, 0)
| BYTES in = (0), out = (0)
+-->POOL Web-Tier-Pool-01
| LB METHOD round-robin
| LB PROTOCOL L7
| Transparent disabled
| SESSION (cur, max, total) = (0, 0, 0)
| BYTES in = (0), out = (0)
+-->POOL MEMBER: Web-Tier-Pool-01/web-01a, STATUS: UP
| | HEALTH MONITOR = BUILT-IN, default_https_monitor:L7OK
| | | LAST STATE CHANGE: 2016-05-16 07:02:00
| | SESSION (cur, max, total) = (0, 0, 0)
| | BYTES in = (0), out = (0)
+-->POOL MEMBER: Web-Tier-Pool-01/web-02a, STATUS: UP
| | HEALTH MONITOR = BUILT-IN, default_https_monitor:L7OK
```

```
| | | LAST STATE CHANGE: 2016-05-16 07:02:01
| | SESSION (cur, max, total) = (0, 0, 0)
| | BYTES in = (0), out = (0)
```

```
nsxedge> show service loadbalancer pool Web-Tier-VIP-01
TIMESTAMP          SESSIONS      BYTESIN      BYTESOUT      SESSIONRATE    HTTPREQS
2016-04-27 19:56:40    00           00           00           00            00
2016-04-27 19:55:00    00           32           100          00            00
```

5 检查服务监控状态（正常、警告、严重）。

```
nsxedge> show service loadbalancer monitor
-----
Loadbalancer Health Check Statistics:

MONITOR PROVIDER  POOL          MEMBER      HEALTH STATUS
built-in          Web-Tier-Pool-01 web-01a     default_https_monitor:L70K
built-in          Web-Tier-Pool-01 web-02a     default_https_monitor:L70K
```

6 检查日志。

```
nsxedge> show log
2016-04-20T20:15:36+00:00 vShieldEdge kernel: Initializing cgroup subsys cpuset
2016-04-20T20:15:36+00:00 vShieldEdge kernel: Initializing cgroup subsys cpu
2016-04-20T20:15:36+00:00 vShieldEdge kernel: Initializing cgroup subsys cpuacct
...
```

7 检查负载均衡器会话表。

```
nsxedge> show service loadbalancer session
-----
L7 Loadbalancer Statistics:
STATUS    PID      MAX_MEM_MB MAX SOCK  MAX_CONN  MAX_PIPE  CUR_CONN  CONN_RATE  CONN_RATE_LIMIT
MAX_CONN_RATE
running   1580      0          2081     1024      0         0         0         0           0

-----L7 Loadbalancer Current
Sessions:

0x2192df1f300: proto=unix_stream src=unix:1 fe=GLOBAL be=<NONE> srv=<none> ts=09 age=0s calls=2
rq[f=c08200h,i=0,an=00h,rx=20s,wx=,ax=] rp[f=008000h,i=0,an=00h,rx=,wx=,ax=] s0=[7,8h,fd=1,ex=]
s1=[7,0h,fd=-1,ex=] exp=19s

-----
L4 Loadbalancer Statistics:
MAX_CONN  ACT_CONN  INACT_CONN TOTAL_CONN
0          0         0          0
```

```
L4 Loadbalancer Current Sessions:
```

```
pro expire state      source      virtual      destination
```

8 检查负载均衡器第 7 层粘性表状态。

```
nsxedge> show service loadbalancer table
```

```
-----
```

```
L7 Loadbalancer Sticky Table Status:
```

```
TABLE    TYPE    SIZE(BYTE)  USED(BYTE)
```

常见的负载均衡器问题

本主题讨论了一些问题以及如何解决这些问题。

在使用 NSX 负载均衡时，通常会出现以下问题：

- TCP 端口 443 上的负载均衡无法正常工作。
- 未利用负载均衡池的某个成员。
- Edge 流量未实现负载均衡。
- 第 7 层负载均衡引擎停止。
- 运行状况监控引擎停止。
- 池成员监控状态为“警告/严重”。
- 池成员处于“非活动”状态。
- 第 7 层粘性表与备用 Edge 不同步。

基本故障排除

- 1 在 vSphere Web Client 中检查负载均衡器配置状态：
 - a 单击**网络和安全 > NSX Edge (Networking & Security > NSX Edges)**。
 - b 双击一个 NSX Edge。
 - c 单击**管理 (Manage)**。
 - d 单击**负载均衡器 (Load Balancer)**选项卡。
 - e 检查负载均衡器状态和配置的日志记录级别。
- 2 在解决负载均衡器服务问题之前，请在 NSX Manager 上运行以下命令以确保该服务已启动并正在运行：

```
nsxmgr> show edge edge-4 service loadbalancer
```

```
haIndex:          0
```

```
-----
```

```
Loadbalancer Services Status:
```

```

L7 Loadbalancer      : running
-----
L7 Loadbalancer Statistics:
STATUS      PID      MAX_MEM_MB MAX SOCK  MAX_CONN  MAX_PIPE  CUR_CONN  CONN_RATE  CONN_RATE_LIMIT
MAX_CONN_RATE
running     1580      0          2081     1024      0          0          0          0
-----
L4 Loadbalancer Statistics:
MAX_CONN  ACT_CONN  INACT_CONN  TOTAL_CONN
0          0          0           0

Prot LocalAddress:Port Scheduler Flags
-> RemoteAddress:Port      Forward Weight ActiveConn InActConn

```

注 您可以运行 `show edge all` 以查找 NSX Edge 名称。

解决配置问题

如果 NSX 用户界面或 REST API 调用拒绝负载平衡器配置操作，可将其归类为配置问题。

解决数据层面问题

NSX Manager 接受了负载平衡器配置，但在客户端 Edge 与负载平衡服务器之间出现连接或性能问题。数据层面问题还包括负载平衡器运行时 CLI 问题以及负载平衡器系统事件问题。

- 1 通过使用该 REST API 调用，将 NSX Manager 中的 Edge 日志记录级别从“信息”更改为“跟踪”或“调试”。

```

URL: https://NSX_Manager_IP/api/1.0/services/debug/loglevel/com.vmware.vshield.edge?level=TRACE
Method: POST

```

- 2 在 vSphere Web Client 中检查池成员状态。
 - a 单击**网络和安全 > NSX Edge (Networking & Security > NSX Edges)**。
 - b 双击一个 NSX Edge。
 - c 单击**管理 (Manage)**。
 - d 单击**负载平衡器 (Load Balancer)**选项卡。
 - e 单击**池 (Pools)**以查看配置的负载平衡器池摘要。
 - f 选择负载平衡器池。单击**显示池统计信息 (Show Pool Statistics)**，然后验证池状态是否为“已启动”。
- 3 通过使用该 REST API 调用，您可以从 NSX Manager 中获取更详细的负载平衡器池配置统计信息：

```

URL: https://NSX_Manager_IP/api/4.0/edges/{edgeId}/loadbalancer/statistics
Method: GET

```

```

<?xml version="1.0" encoding="UTF-8"?>
<loadBalancerStatusAndStats>

```

```

<timeStamp>1463507779</timeStamp>
<pool>
  <poolId>pool-1</poolId>
  <name>Web-Tier-Pool-01</name>
  <member>
    <memberId>member-1</memberId>
    <name>web-01a</name>
    <ipAddress>172.16.10.11</ipAddress>
    <status>UP</status>
    <lastStateChangeTime>2016-05-16 07:02:00</lastStateChangeTime>
    <bytesIn>0</bytesIn>
    <bytesOut>0</bytesOut>
    <curSessions>0</curSessions>
    <httpReqTotal>0</httpReqTotal>
    <httpReqRate>0</httpReqRate>
    <httpReqRateMax>0</httpReqRateMax>
    <maxSessions>0</maxSessions>
    <rate>0</rate>
    <rateLimit>0</rateLimit>
    <rateMax>0</rateMax>
    <totalSessions>0</totalSessions>
  </member>
  <member>
    <memberId>member-2</memberId>
    <name>web-02a</name>
    <ipAddress>172.16.10.12</ipAddress>
    <status>UP</status>
    <lastStateChangeTime>2016-05-16 07:02:01</lastStateChangeTime>
    <bytesIn>0</bytesIn>
    <bytesOut>0</bytesOut>
    <curSessions>0</curSessions>
    <httpReqTotal>0</httpReqTotal>
    <httpReqRate>0</httpReqRate>
    <httpReqRateMax>0</httpReqRateMax>
    <maxSessions>0</maxSessions>
    <rate>0</rate>
    <rateLimit>0</rateLimit>
    <rateMax>0</rateMax>
    <totalSessions>0</totalSessions>
  </member>
  <status>UP</status>
  <bytesIn>0</bytesIn>
  <bytesOut>0</bytesOut>
  <curSessions>0</curSessions>
  <httpReqTotal>0</httpReqTotal>
  <httpReqRate>0</httpReqRate>
  <httpReqRateMax>0</httpReqRateMax>
  <maxSessions>0</maxSessions>
  <rate>0</rate>
  <rateLimit>0</rateLimit>
  <rateMax>0</rateMax>
  <totalSessions>0</totalSessions>
</pool>
<virtualServer>
  <virtualServerId>virtualServer-1</virtualServerId>

```



```

<name>Web-Tier-VIP-01</name>
<ipAddress>172.16.10.10</ipAddress>
<status>OPEN</status>
<bytesIn>0</bytesIn>
<bytesOut>0</bytesOut>
<curSessions>0</curSessions>
<httpReqTotal>0</httpReqTotal>
<httpReqRate>0</httpReqRate>
<httpReqRateMax>0</httpReqRateMax>
<maxSessions>0</maxSessions>
<rate>0</rate>
<rateLimit>0</rateLimit>
<rateMax>0</rateMax>
<totalSessions>0</totalSessions>
</virtualServer>
</loadBalancerStatusAndStats>

```

- 4 要从命令行中检查负载均衡器统计信息，请在 NSX Edge 上运行这些命令。

对于特定的虚拟机：先运行 `show service loadbalancer virtual` 以获取虚拟机名称，然后运行 `show statistics loadbalancer virtual <virtual-machine-name>`。

对于特定的 TCP 池：先运行 `show service loadbalancer pool` 以获取池名称，然后运行 `show statistics loadbalancer pool <pool-name>`。

- 5 查看负载均衡器统计信息以查找故障迹象。