



链滴

HCIP-VxLAN

作者: [shuaiqijun](#)

原文链接: <https://ld246.com/article/1615108720121>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

VXLAN (Virtual eXtensible Local Area Network, 虚拟扩展局域网) 是一种网络虚拟化技术, 可改进大型云计算在部署时的扩展问题, 是对VLAN的一种扩展。VXLAN是一种功能强大的工具, 可以透三层网络对二层进行扩展。它可通过封装流量并将其扩展到第三层网关, 以此来解决VMS (虚拟内系统) 的可移植性限制, 使其可以访问在外部IP子网上的服务器。

RFC7348定义了VLAN扩展方案VXLAN (Virtual eXtensible Local Area Network, 虚拟扩展局域网)。VXLAN采用MAC in UDP (User Datagram Protocol) 封装方式, 是NVO3 (Network Virtualization over Layer 3) 中的一种网络虚拟化技术。

起源:

随着网络技术的发展, 云计算凭借其在系统利用率高、人力/管理成本低、灵活性/可扩展性强等方面表现出的优势, 已经成为目前企业IT建设的新趋势。而服务器虚拟化作为云计算的核心技术之一, 得到了越来越多的应用。

服务器虚拟化技术的广泛部署, 极大地增加了数据中心的计算密度; 同时, 为了实现业务的灵活变更, 虚拟机VM (Virtual Machine) 需要能够在网络中不受限迁移, 这给传统的“二层+三层”数据中心网络带来了新的挑战。

- 虚拟机规模受网络设备表项规格的限制

在传统二层网络环境下, 数据报文是通过查询MAC地址表进行二层转发。服务器虚拟化后, VM的数量比原有的物理机发生了数量级的增长, 伴随而来的便是VM网卡MAC地址数量的空前增加。而接入侧二层设备的MAC地址表规格较小, 无法满足快速增长的VM数量。

- 网络隔离能力有限

VLAN作为当前主流的网络隔离技术, 在标准定义中只有12比特, 因此可用的VLAN数量仅4096个。对于公有云或其它大型虚拟化云计算服务这种动辄上万甚至更多租户的场景而言, VLAN的隔离能力无法满足。

- 虚拟机迁移范围受限

由于服务器资源等问题 (如CPU过高, 内存不够等), 虚拟机迁移已经成为了一个常态化业务。

虚拟机迁移是指将虚拟机从一个物理机迁移到另一个物理机。为了保证虚拟机迁移过程中业务不中断, 则需要保证虚拟机的IP地址、MAC地址等参数保持不变, 这就要求虚拟机迁移必须发生在一个二层网络中。而传统的二层网络, 将虚拟机迁移限制在了一个较小的局部范围内。

- 针对虚拟机规模受设备表项规格限制

VXLAN将管理员规划的同一区域内的VM发出的原始报文封装成新的UDP报文, 并使用物理网络的IP和MAC地址作为外层头, 这样报文对网络中的其他设备只表现为封装后的参数。因此, 极大降低了大二层网络对MAC地址规格的需求。

- 针对网络隔离能力限制

VXLAN引入了类似VLAN ID的用户标识, 称为VXLAN网络标识VNI (VXLAN Network Identifier), 由24比特组成, 支持多达16M的VXLAN段, 有效得解决了云计算中海量租户隔离的问题。

- 针对虚拟机迁移范围受限

VXLAN将VM发出的原始报文进行封装后通过VXLAN隧道进行传输, 隧道两端的VM不需感知传输网络的物理架构。这样, 对于具有同一网段IP地址的VM而言, 即使其物理位置不在同一个二层网络中, 但从逻辑上看, 相当于处于同一个二层域。即VXLAN技术在三层网络之上, 构建出了一个虚拟的大二层网络, 只要虚拟机路由可达, 就可以将其规划到同一个大二层网络中。这就解决了虚拟机迁移范围受限问题。

传统数据中心网络所面临的挑战

数据中心的基本概念及特点:

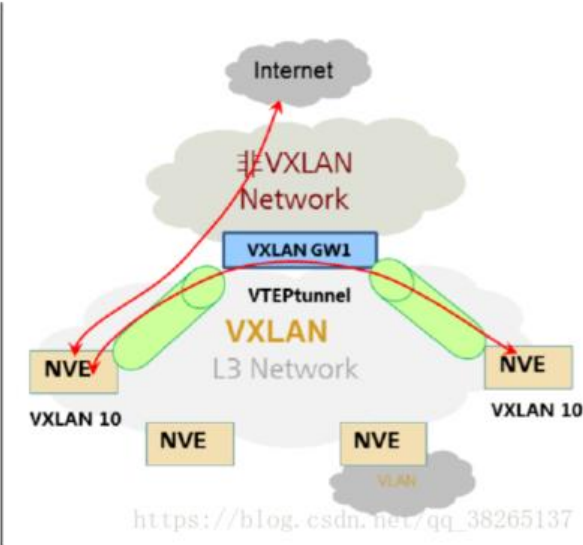
- 数据中心 (Data Center) 是一套完整、复杂的集合系统, 它不仅包括计算机系统和其它与之配套的设备 (例如通信和存储系统), 还包含数据通信系统、环境控制设备、监控设备以及各种安全装置。
- 数据中心通常是指在一个物理空间内实现信息集中处理、存储、传输、交换、管理的场所。
- 服务器、网络设备、存储设备等通常都是数据中心的关键设备。
- 设备运行所需要的环境因素, 如供电系统、制冷系统、机柜系统、消防系统、监控系统等通常都被认为是关键物理基础设施。
- 互联网数据中心 (Internet Data Center, IDC) 是互联网中数据存储和处理的中心, 是互联网中数据交互最为集中的地方。
- 数据中心的四大焦点: 可靠、灵活、绿色、高效。

传统数据中心网络结构:

- 传统网络模型在很长一段时间内, 支撑了各种类型的数据中心。原文链接: [HCIP-VxLAN](#)
- 按照功能模块划分, 传统数据中心可分为核心区、外网服务器区、内网服务器区、互联网服务器区、数据中心管理区、数据交换&测试服务器区、数据存储功能区、数据容灾功能区等。

- 适合云业务：可实现千万级别的租户间隔离，有力地支持了云业务的大规模部署。
- 技术优势：VXLAN利用了现有通用的UDP进行传输，成熟性极高。

VXLAN网关：



图：VXLAN网关结构图

NVE:目前有软件NVE（一般安装在服务器上；例如OVS）和硬件NVE（一般集成在交换机上，例如CE6850）。由于软件NVE是在原设备中安装一个软件包，硬件NVE是在原设备中增加一个硬件模块，而原设备多数是VLAN的二层设备，所以，NVE又是VXLAN的二层网关，主要实现VXLAN与VLAN、MAC等的二层映射。

VXLAN网关：

- 与VXLAN NVE类似，是地位更高一些的另一个VXLAN角色，即VXLAN三层网关，简称VXLAN GW，主要实现VXLAN报文头与IP报文头的映射。
- 不管二层VXLAN网关还是三层VXLAN网关，都是主要实现了VXLAN网络和非VXLAN网络之间的连接。

- NVE是服务器虚拟化层的一个功能模块，虚拟机通过虚拟化软件直接建立VTEP隧道。
- NVE也可以是一台支持VXLAN的接入交换机集中为多租户提供VXLAN网关服务。
- VXLAN网关可以实现不同VXLAN下租户间通信，也能实现VXLAN用户与非VXLAN用户间通信，这和VLANIF接口的功能是类似的。

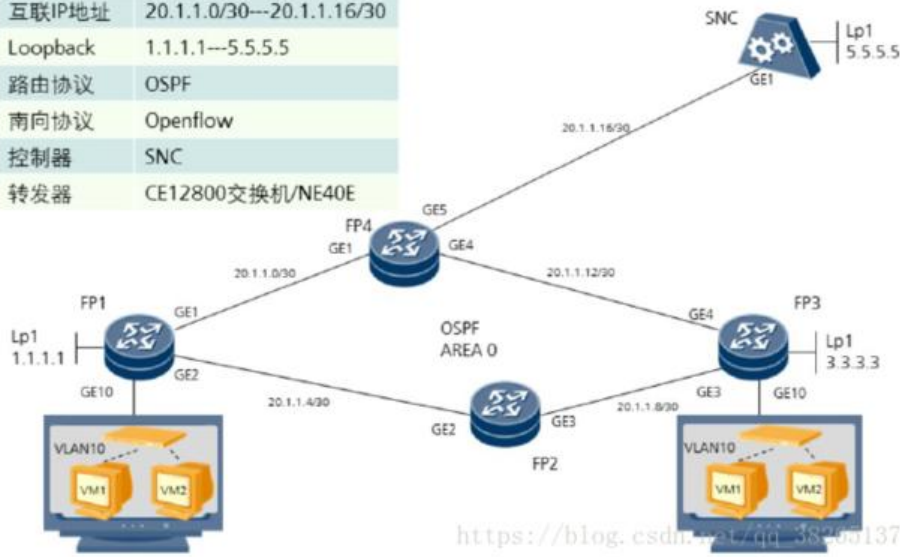
NVO3标准术语：

VN	虚拟网络
VNI	虚拟网络实例（Virtual Network Instance），可以为L2或L3网络，一个租户可以对应一个或多个VNI。
	虚拟网络ID（Virtual Network Identifier），标识一个虚拟网络。
	虚拟网络边缘（Virtual Network Edge），可以位于物理网络边缘设备，也可以位于Hypervisor，可以是二层转发或三层转发。
VN Context	该字段位于Overlay封装头部，用于Egress NVE设备确定VNI。
	运行在物理服务器内的虚拟化软件，为服务的VN提供共享计算资源、内存、存储，而且Hypervisor内经常内嵌Virtual Switch。
Tenant End System	租户终端系统，可以是物理服务器也可以是VM。

业界其他技术实现-NVO3技术背景：

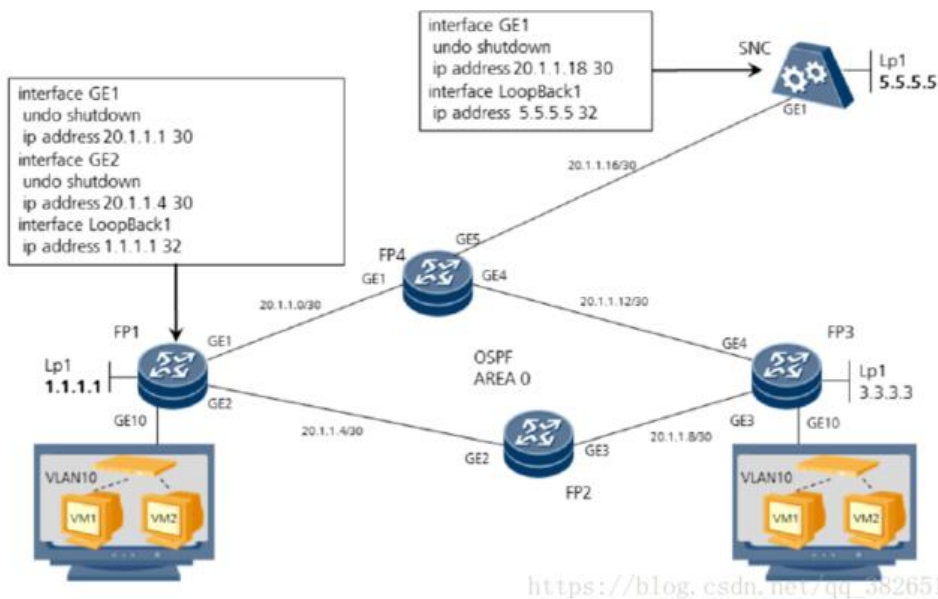
类别	简单描述	主导厂商
NVGRE	通过Mac In GRE封装实现虚拟二层网络的方法。	Microsoft&Intel、Arista、Broadcom、Dell、Emulex、HP。

规划与描述	
互联IP地址	20.1.1.0/30---20.1.1.16/30
Loopback	1.1.1.1---5.5.5.5
路由协议	OSPF
南向协议	Openflow
控制器	SNC
转发器	CE12800交换机/NE40E



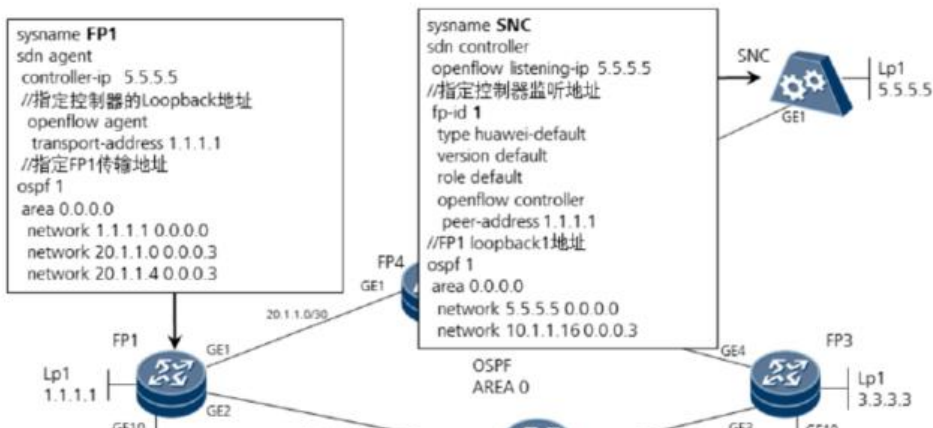
图：拓扑图

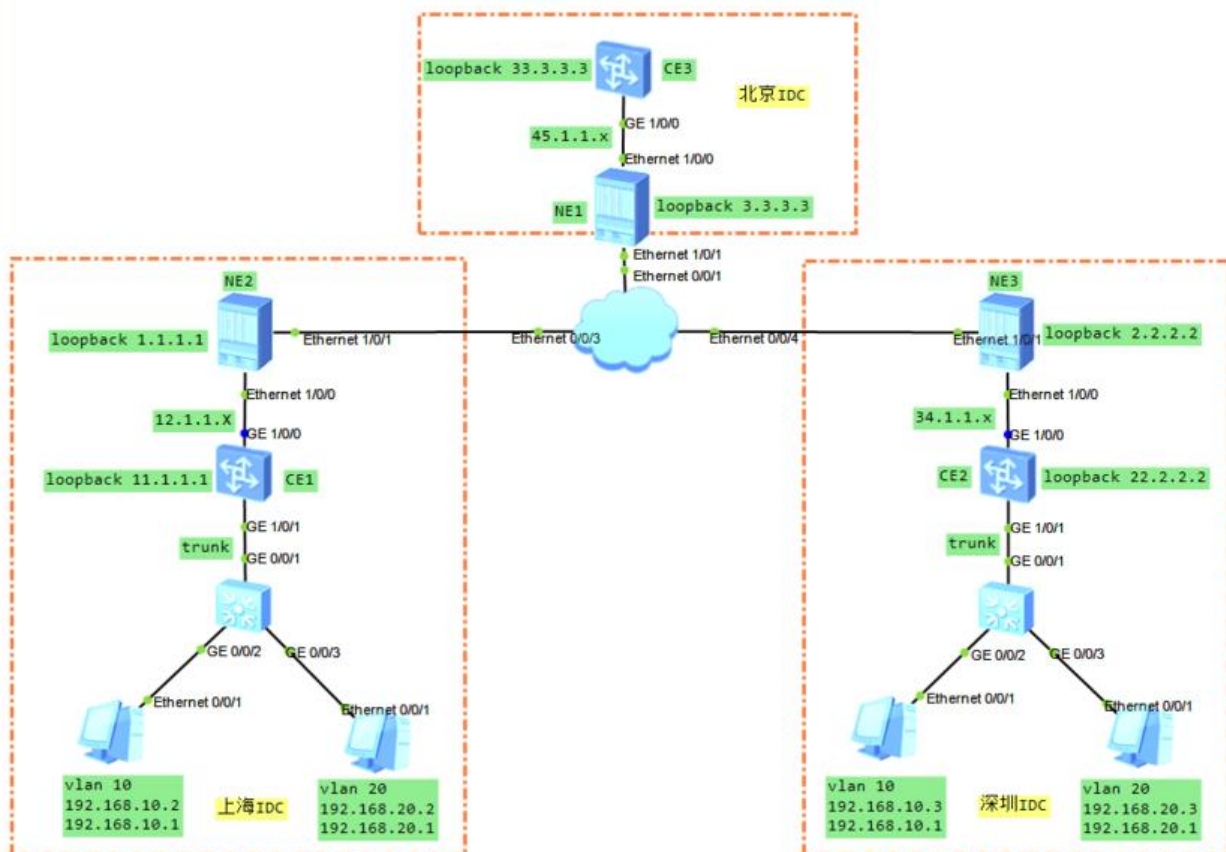
接口配置：



图：基于SDN的VLAN基本组网-接口配置

协议配置：





[Huawei]vlan 10

[Huawei-vlan2]vlan 20

[Huawei-vlan20]qu

[Huawei]interface g0/0/1

[Huawei-GigabitEthernet0/0/1]port link-type trunk

[Huawei-GigabitEthernet0/0/1]port trunk allow-pass vlan all

[Huawei-GigabitEthernet0/0/1]qu

[Huawei]interface g0/0/2

[Huawei-GigabitEthernet0/0/2]port link-type access

[Huawei-GigabitEthernet0/0/2]port default vlan 10

[Huawei-GigabitEthernet0/0/2]interface g0/0/3

[Huawei-GigabitEthernet0/0/3]port link-type access

```
[Huawei-GigabitEthernet0/0/3]port default vlan 20
[Huawei]vlan batch 10 20
[Huawei]interface g0/0/1
[Huawei-GigabitEthernet0/0/1]port link-type trunk
[Huawei-GigabitEthernet0/0/1]port trunk allow-pass vlan all
[Huawei-GigabitEthernet0/0/1]interface g0/0/2
[Huawei-GigabitEthernet0/0/2]port link-type access
[Huawei-GigabitEthernet0/0/2]port default vlan 10
[Huawei-GigabitEthernet0/0/2]interface g0/0/3
[Huawei-GigabitEthernet0/0/3]port link-type access
[Huawei-GigabitEthernet0/0/3]port default vlan 20
```

```
[*HUAWEI]bridge-domain 10    //标识一个二层广播域，BD和VNI 1:1映射（VNI在BD里配置）
[*HUAWEI-bd10]vxlan vni 10
[*HUAWEI-bd10]q
[*HUAWEI]bridge-domain 20
[*HUAWEI-bd20]vxlan vni 20
[*HUAWEI-bd20]q
[*HUAWEI]commit
[~HUAWEI]interface g1/0/1.10 mode l2
[*HUAWEI-GE1/0/1.10]encapsulation dot1q vid 10
[*HUAWEI-GE1/0/1.10]bridge-domain 10
[*HUAWEI-GE1/0/1.10]q
[*HUAWEI]commit
[~HUAWEI]interface g1/0/1.20 mode l2
[*HUAWEI-GE1/0/1.20]encapsulation dot1q vid 20
[*HUAWEI-GE1/0/1.20]bridge-domain 20
[*HUAWEI-GE1/0/1.20]q
```

```
[*HUAWEI]commit
[~HUAWEI]interface g1/0/1
[~HUAWEI-GE1/0/1]undo shutdown
[*HUAWEI-GE1/0/1]commit
[~HUAWEI-GE1/0/1]qu
[~HUAWEI]interface loopback 0
[*HUAWEI-LoopBack0]ip add 11.1.1.1 24
[*HUAWEI]interface g1/0/0
[*HUAWEI-GE1/0/0]undo shutdown
[*HUAWEI-GE1/0/0]undo portswitch
[*HUAWEI-GE1/0/0]ip add 12.1.1.2 24
[*HUAWEI-GE1/0/0]qu
[*HUAWEI]ospf 1 router-id 11.1.1.1
[*HUAWEI-ospf-1]area 0
[*HUAWEI-ospf-1-area-0.0.0.0]network 11.1.1.0 0.0.0.255
[~HUAWEI-ospf-1-area-0.0.0.0]network 12.1.1.0 0.0.0.255
[*HUAWEI-ospf-1-area-0.0.0.0]commit
[~HUAWEI]interface NVE 1
[*HUAWEI-Nve1]source 11.1.1.1
[*HUAWEI-Nve1]vni 10 head-end peer-list 22.2.2.2
[*HUAWEI-Nve1]vni 20 head-end peer-list 22.2.2.2

[~HUAWEI]bridge-domain 10
[*HUAWEI-bd10]vxlan vni 10
[*HUAWEI]bridge-domain 20
[*HUAWEI-bd20]vxlan vni 20
[*HUAWEI-bd20]qu
```



```
[*HUAWEI]interface g1/0/1.10 mode l2
[*HUAWEI-GE1/0/1.10]encapsulation dot1q vid 10
[*HUAWEI-GE1/0/1.10]qu
[*HUAWEI]interface g1/0/1.20 mode l2
[*HUAWEI-GE1/0/1.20]encapsulation dot1q vid 20
[*HUAWEI-GE1/0/1.20]bridge-domain 20
[*HUAWEI-GE1/0/1.20]interface g1/0/1.10 mode l2
[*HUAWEI-GE1/0/1.10]bridge-domain 10
[*HUAWEI-GE1/0/1.10]qu
[*HUAWEI]interface nve 1
[*HUAWEI-Nve1]source 22.2.2.2
[*HUAWEI-Nve1]vni 10 head-end peer-list 11.1.1.1
[*HUAWEI-Nve1]vni 20 head-end peer-list 11.1.1.1
[*HUAWEI-Nve1]qu
[*HUAWEI]commit

[*HUAWEI]interface loopback 0
[*HUAWEI-LoopBack0]ip add 33.3.3.3 24
[*HUAWEI-LoopBack0]interface g1/0/0
[*HUAWEI-GE1/0/0]undo shutdown
[*HUAWEI-GE1/0/0]undo portswitch
[*HUAWEI-GE1/0/0]ip add 45.1.1.2 24
[*HUAWEI-GE1/0/0]q
[*HUAWEI]commit

[~HUAWEI]ospf 1 router-id 33.3.3.3
[*HUAWEI-ospf-1]area 0
[*HUAWEI-ospf-1-area-0.0.0.0]network 33.3.3.0 0.0.0.255
[*HUAWEI-ospf-1-area-0.0.0.0]network 45.1.1.0 0.0.0.255
```

```
[*HUAWEI-ospf-1-area-0.0.0.0]commit
```

```
[*HUAWEI]interface Ethernet 1/0/0
```

```
[*HUAWEI-Ethernet1/0/0]ip add 12.1.1.1 24
```

```
[*HUAWEI-Ethernet1/0/0]interface Ethernet 1/0/1
```

```
[*HUAWEI-Ethernet1/0/1]ip add 100.1.1.1 24
```

```
[*HUAWEI-Ethernet1/0/1]commit
```

```
[~HUAWEI-Ethernet1/0/1]q
```

```
[~HUAWEI]interface loopback 0
```

```
[*HUAWEI-LoopBack0]ip add 1.1.1.1 24
```

```
[*HUAWEI-LoopBack0]commit
```

```
[~HUAWEI-LoopBack0]qu
```

```
[~HUAWEI]ospf 1 router-id 1.1.1.1
```

```
[*HUAWEI-ospf-1]area 0
```

```
[*HUAWEI-ospf-1-area-0.0.0.0]network 1.1.1.1 0.0.0.0
```

```
[*HUAWEI-ospf-1-area-0.0.0.0]network 12.1.1.0 0.0.0.255
```

```
[*HUAWEI-ospf-1-area-0.0.0.0]network 100.1.1.0 0.0.0.255
```

```
[*HUAWEI-ospf-1-area-0.0.0.0]commit
```

```
[*HUAWEI]interface loopback 0
```

```
[*HUAWEI-LoopBack0]ip add 3.3.3.3 24
```

```
[*HUAWEI-LoopBack0]interface ethernet 1/0/0
```

```
[*HUAWEI-Ethernet1/0/0]ip add 45.1.1.1 24
```

```
[*HUAWEI-Ethernet1/0/0]interface ethernet 1/0/1
```

```
[*HUAWEI-Ethernet1/0/1]ip add 100.1.1.3 24
```

```
[*HUAWEI-Ethernet1/0/1]qu
```

```
[*HUAWEI]ospf 1 router-id 3.3.3.3
```

```
[*HUAWEI-ospf-1]area 0
```

```
[*HUAWEI-ospf-1-area-0.0.0.0]network 3.3.3.3 0.0.0.0
```

```
[*HUAWEI-ospf-1-area-0.0.0.0]network 100.1.1.0 0.0.0.255
```

```
[*HUAWEI-ospf-1-area-0.0.0.0]network 45.1.1.0 0.0.0.255
```