VMware NSX设计与部署最佳实践

陈位浩
weihaoc@vmware.com
资深系统工程师

vmware

提纲

1	NSX原理、组件及组件之间的关系
2	物理网络设计的要求及VLAN规划
3	NSX安装实践
4	NSX安全部署实践
5	NSX路由及高可靠性设计
6	基于NSX的双活/灾备数据中心

NSX 架构及组件介绍

逻辑网络

云平台

- 自服务
- vRA，OpenStack, 定制化云平台

管理平面

NSX Manager

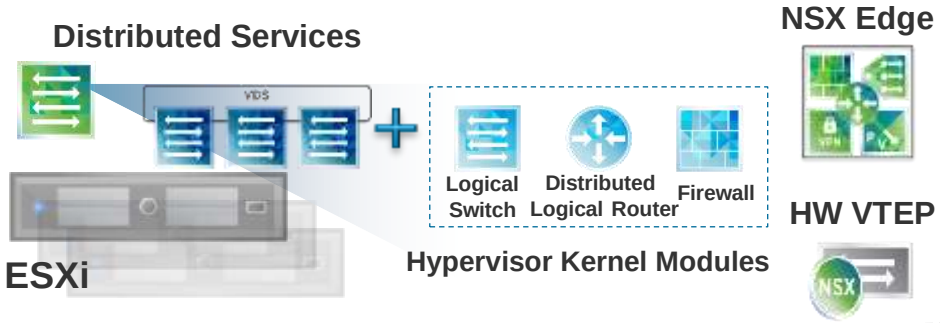
- 集中式配置界面
- REST API entry-point

控制平面

NSX Controller

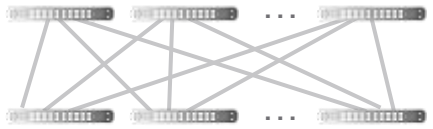
- 管理逻辑网络
- 负责控制层面
- 转发和控制分离，不在传输路径上

转发平面



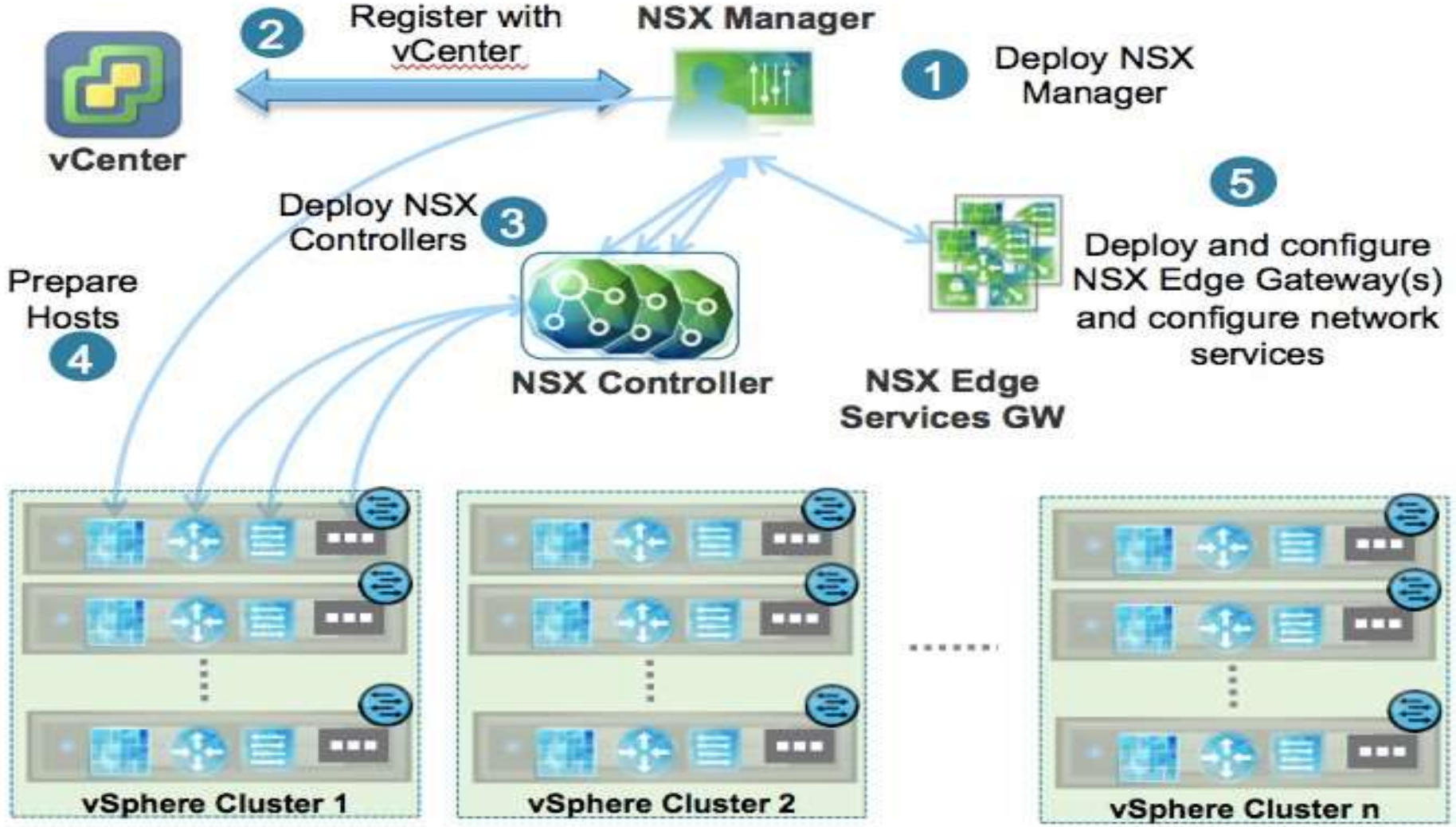
- 高性能数据转发层面
- 横向扩展全分布式转发

物理网络

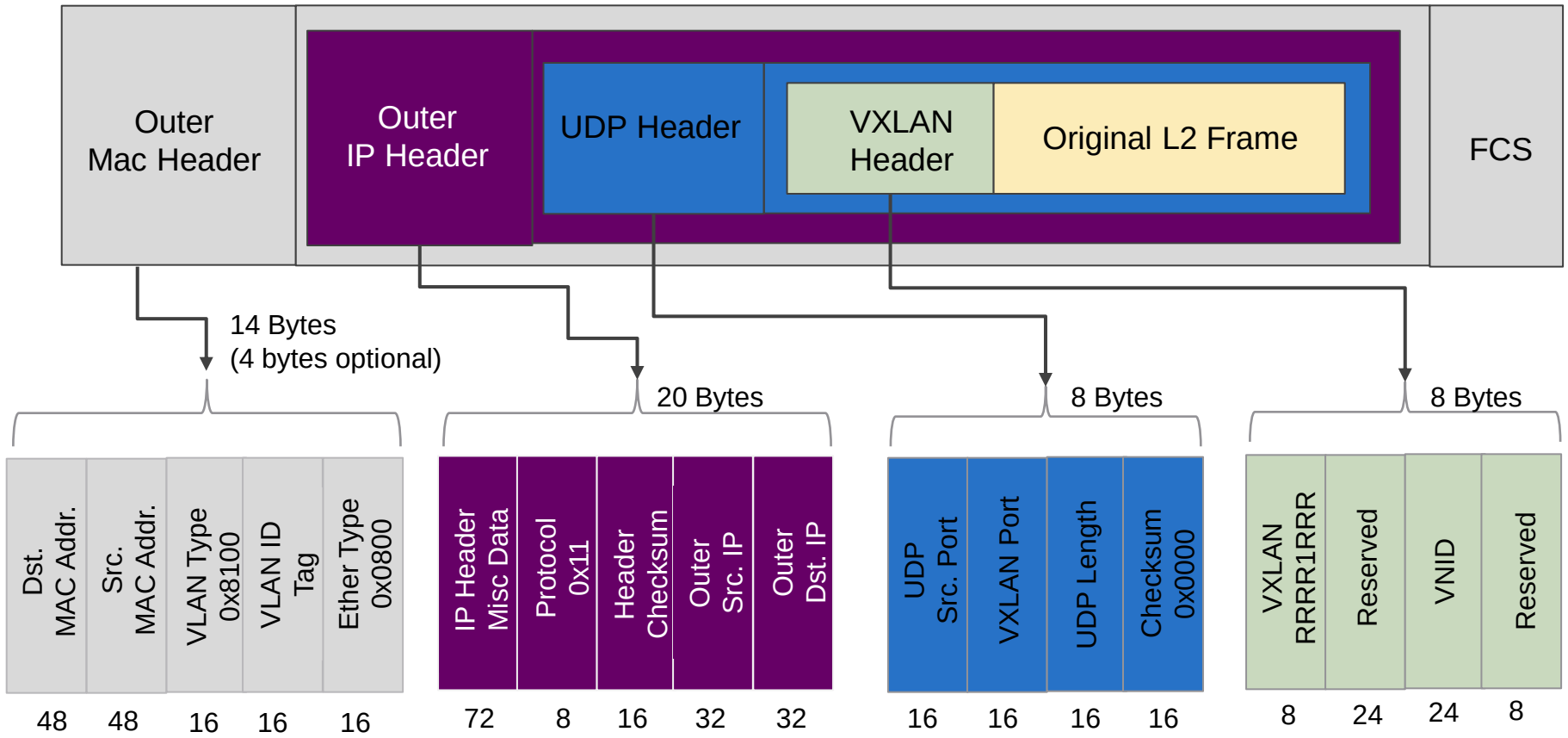


vmware

NSX组件之间的关系


2
Register with vCenter
NSX Manager
1
Deploy NSX Manager
vCenter
Deploy NSX Controllers
3
5
Deploy and configure NSX Edge Gateway(s) and configure network services
Prepare Hosts
4
NSX Controller
NSX Edge Services GW
vSphere Cluster 1
vSphere Cluster 2
vSphere Cluster n

VXLAN 简介



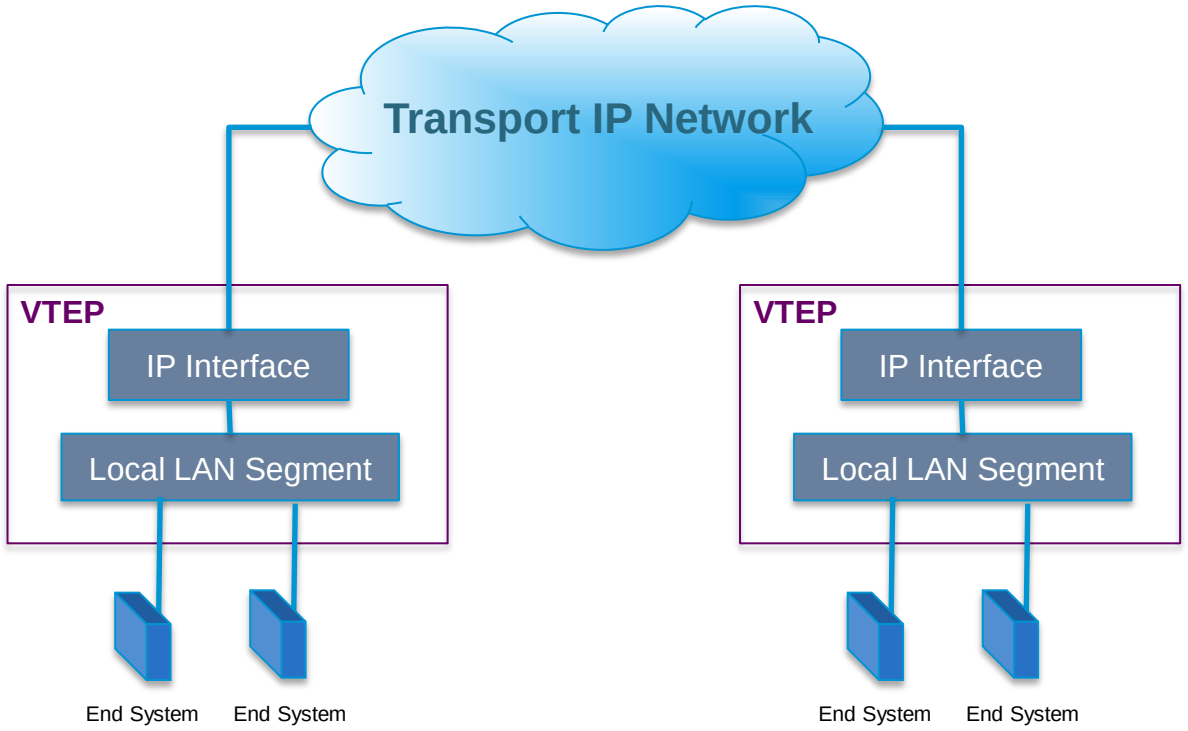
- VXLAN（Virtual eXtensible LAN，可扩展虚拟局域网络）是基于IP网络、采用“MAC in UDP”封装形式的二层技术。
- 基本格式：L2oUDP
  - 封装报头开销50字节
  - UDP目的端口为已知端口，源端口可按流分配，标准5元组方式有利于IP网络转发过程中进行负载分担
- 包含24比特 VXLAN ID
  - 能够支持16M逻辑二层网络
  - 突破4K VLAN限制的关键扩展，映射到本地二层交换域

vmware

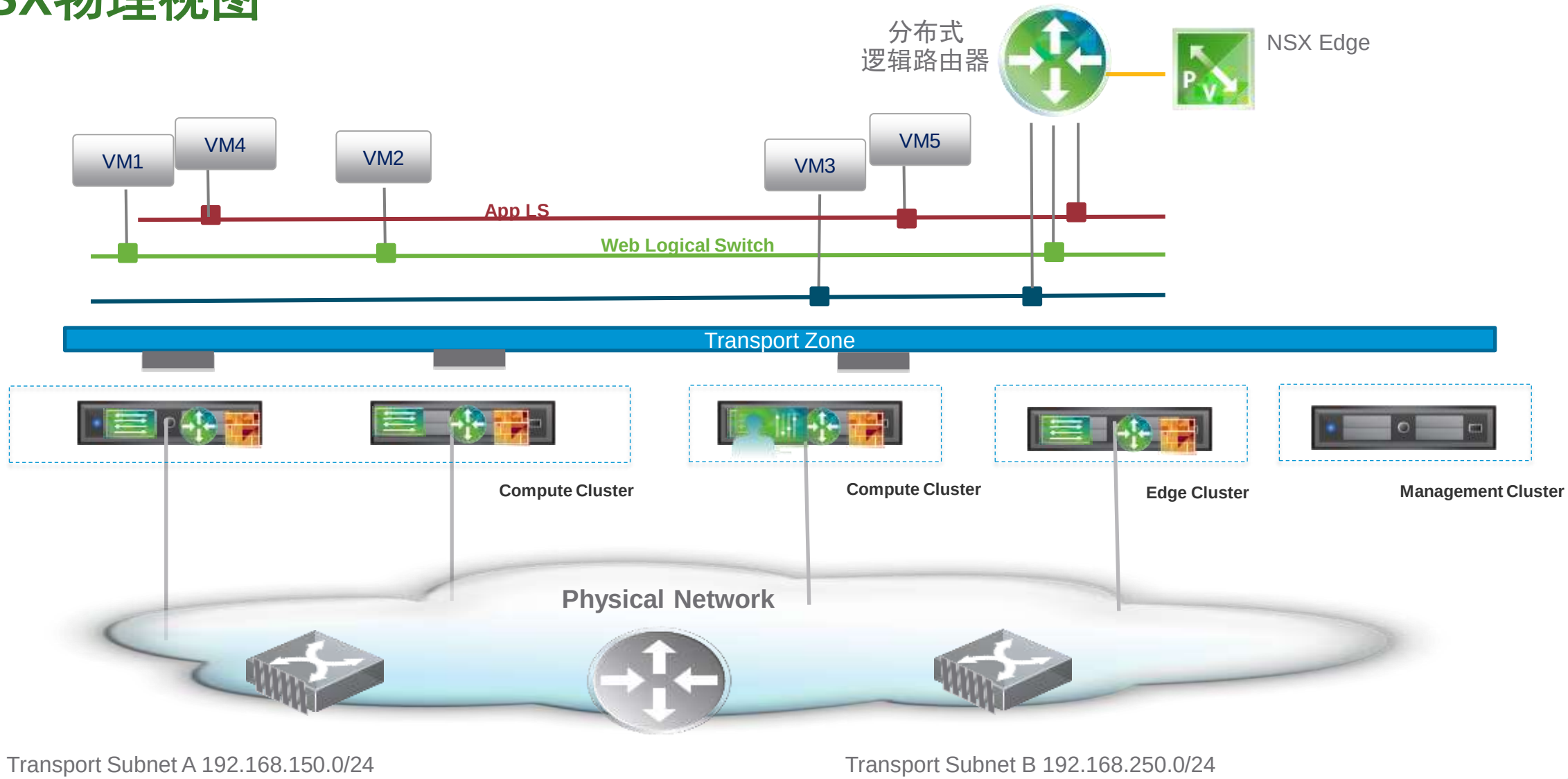
VXLAN 的VTEP

VXLAN terminates its tunnels on VTEPs (Virtual Tunnel End Point).

Each VTEP has two interfaces - one to provide bridging function for local hosts, the other has an IP identification in the core network for VxLAN encapsulation/de-encapsulation.



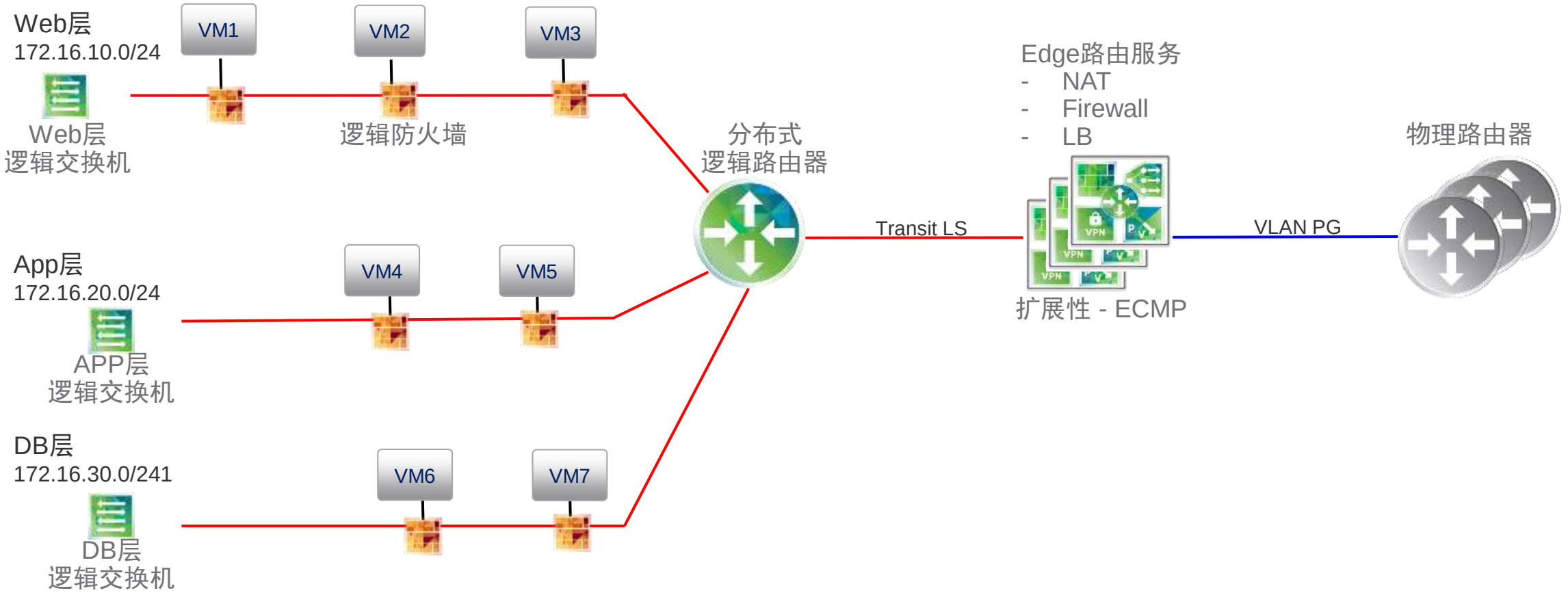
NSX物理视图



Transport Subnet A 192.168.150.0/24

Transport Subnet B 192.168.250.0/24

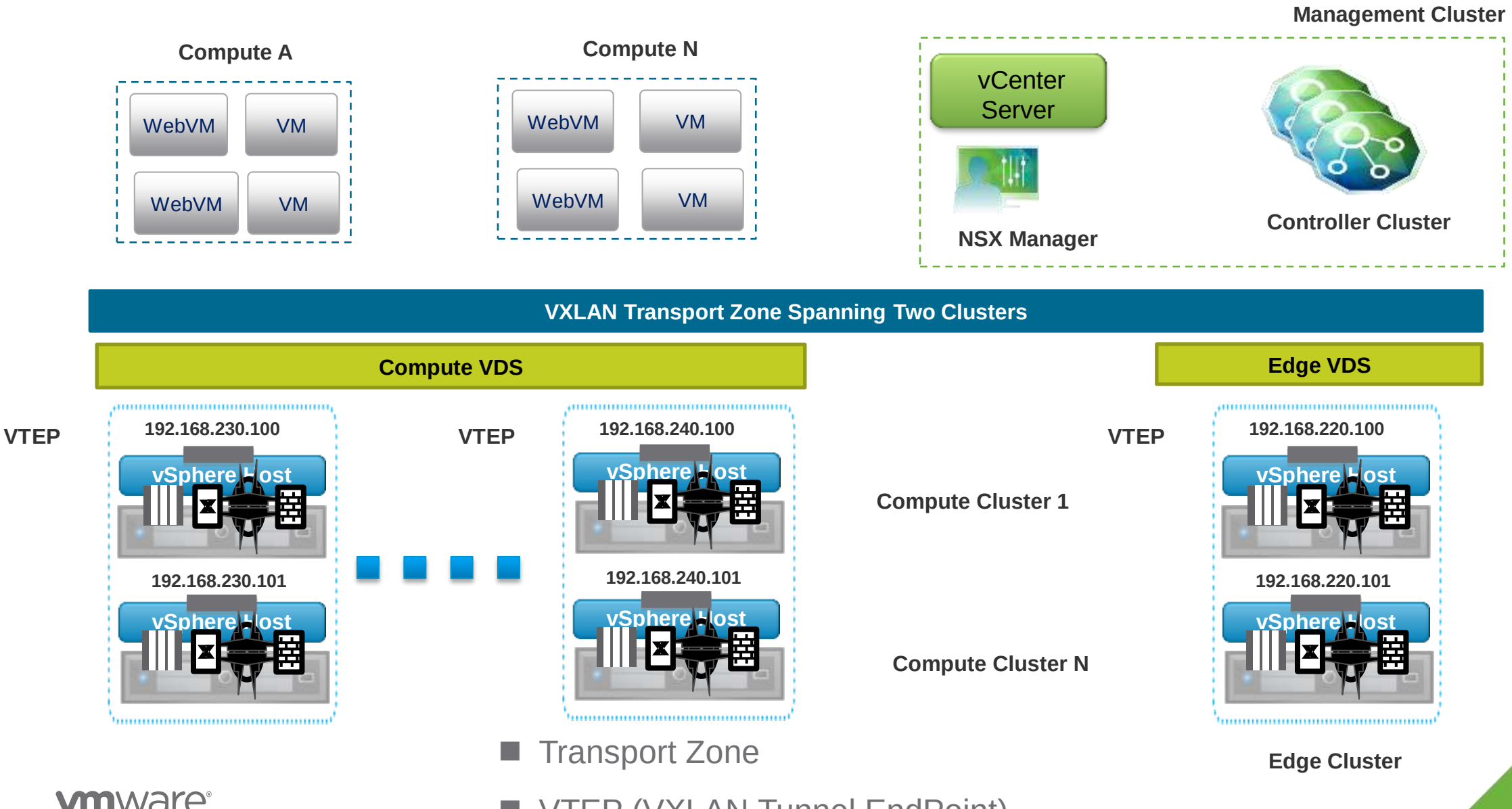
NSX逻辑视图


Web层
172.16.10.0/24
VM1
VM2
VM3
Web层
逻辑交换机
逻辑防火墙
分布式
逻辑路由器
Edge路由服务
- NAT
- Firewall
- LB
物理路由器
Transit LS
VLAN PG
扩展性 - ECMP
App层
172.16.20.0/24
VM4
VM5
APP层
逻辑交换机
DB层
172.16.30.0/241
VM6
VM7
DB层
逻辑交换机

VDS and Transport Zones (1/2)

- A Transport Zone (TZ) defines the «span» of Logical Networks
- A TZ can span Multiple VDS:
  - Distributed Logical Router LIFs considerations apply here
  - Generally do not use VLAN LIFs (see advanced training)
- Typically, only a single TZ is deployed in most environments
- A TZ must is not a security boundary!

VDS and Transport Zone (2/2)



- Transport Zone
- VTEP (VXLAN Tunnel EndPoint)

提纲

1	NSX原理、组件及组件之间的关系
2	物理网络设计的要求及VLAN规划
3	NSX安装实践
4	NSX安全部署实践
5	NSX路由及高可靠性设计
6	基于NSX的双活/灾备数据中心

网络虚拟化物理架构最佳实践

NSX不关心物理网络架构

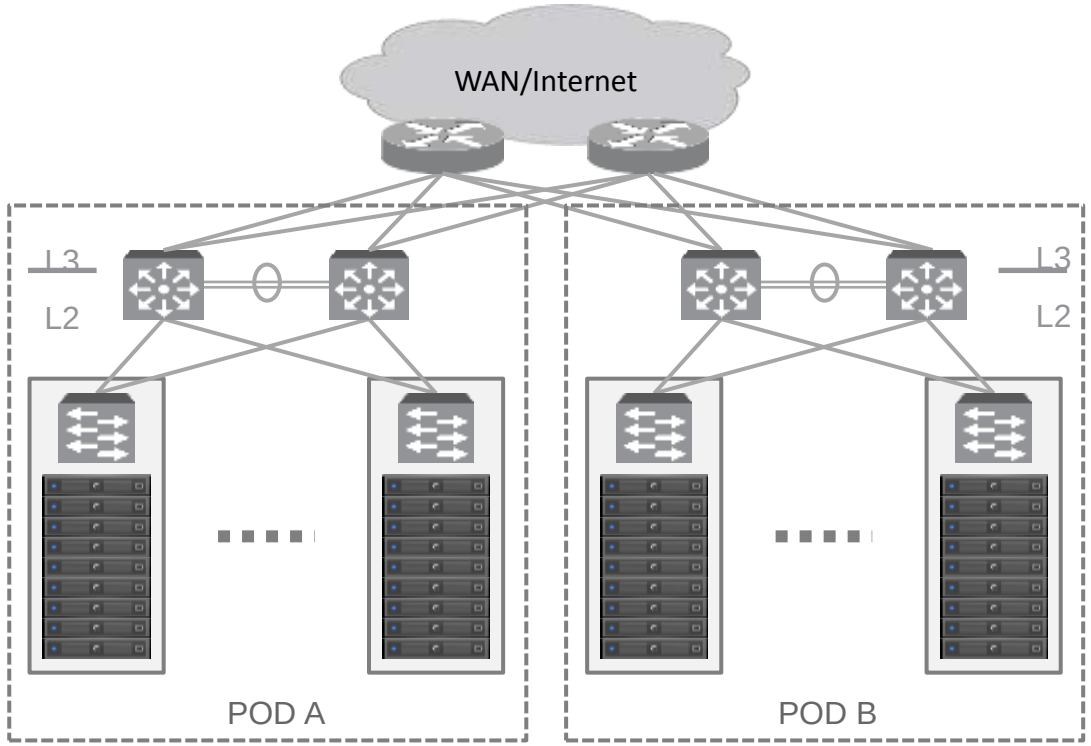
不关心物理网络中运行的L2或L3的协议

对物理网络仅有两个需求

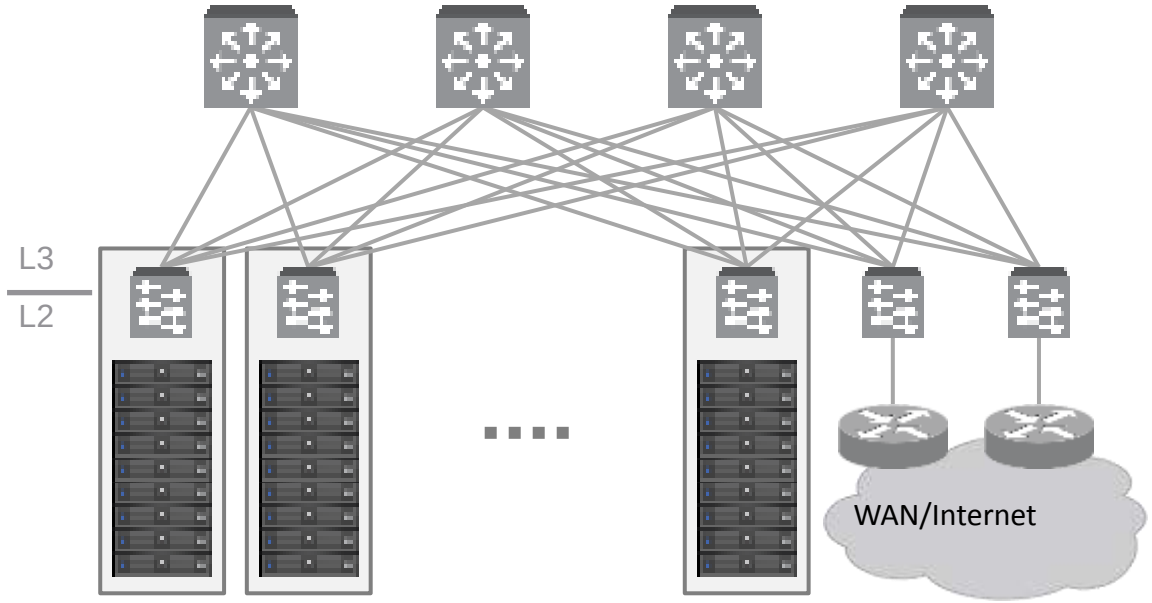
IP可达

MTU大于等于1600

物理网络任意架构



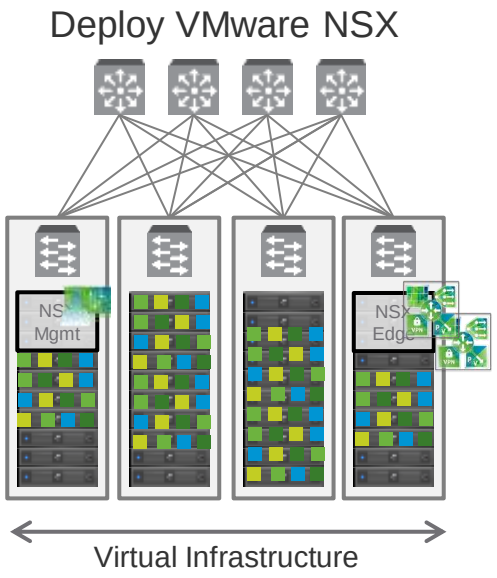
传统接入/汇聚/核心网络架构



Spine-Leaf架构

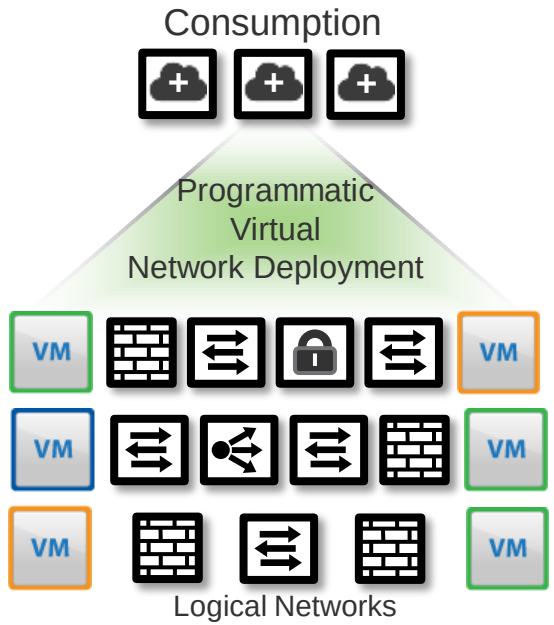
vmware

NSX简化部署和运维



One Time

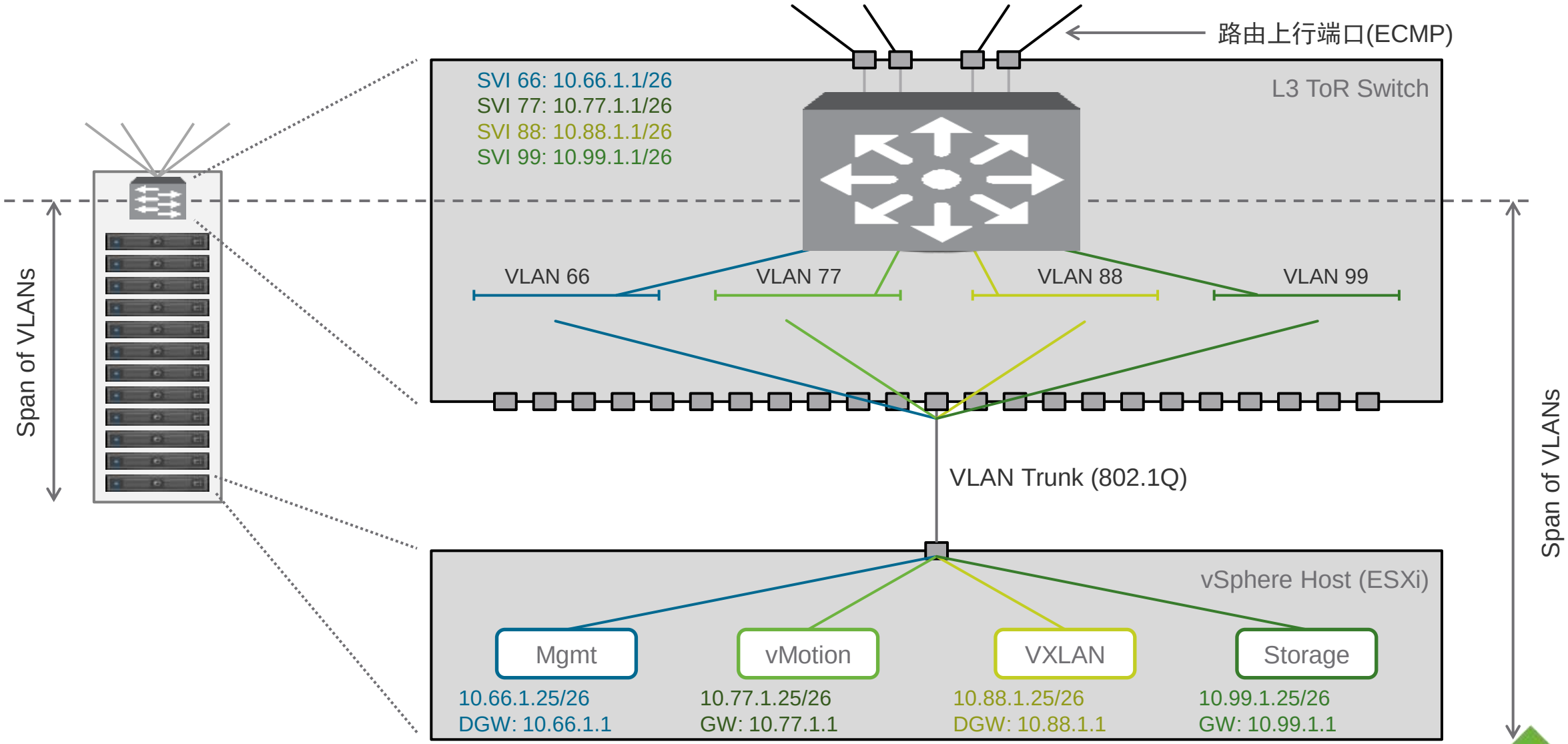
- 需要部署的组件
- 部署NSX Manager
- 部署NSX Controller Cluster
- 准备工作
- Host Preparation
- Logical Network Preparation



Recurring

- 逻辑网络/安全服务
- 部署每层应用的Logical Switches
- 部署每租户DLR
- 创建Bridged Network

TOR交换机VLAN设计原则



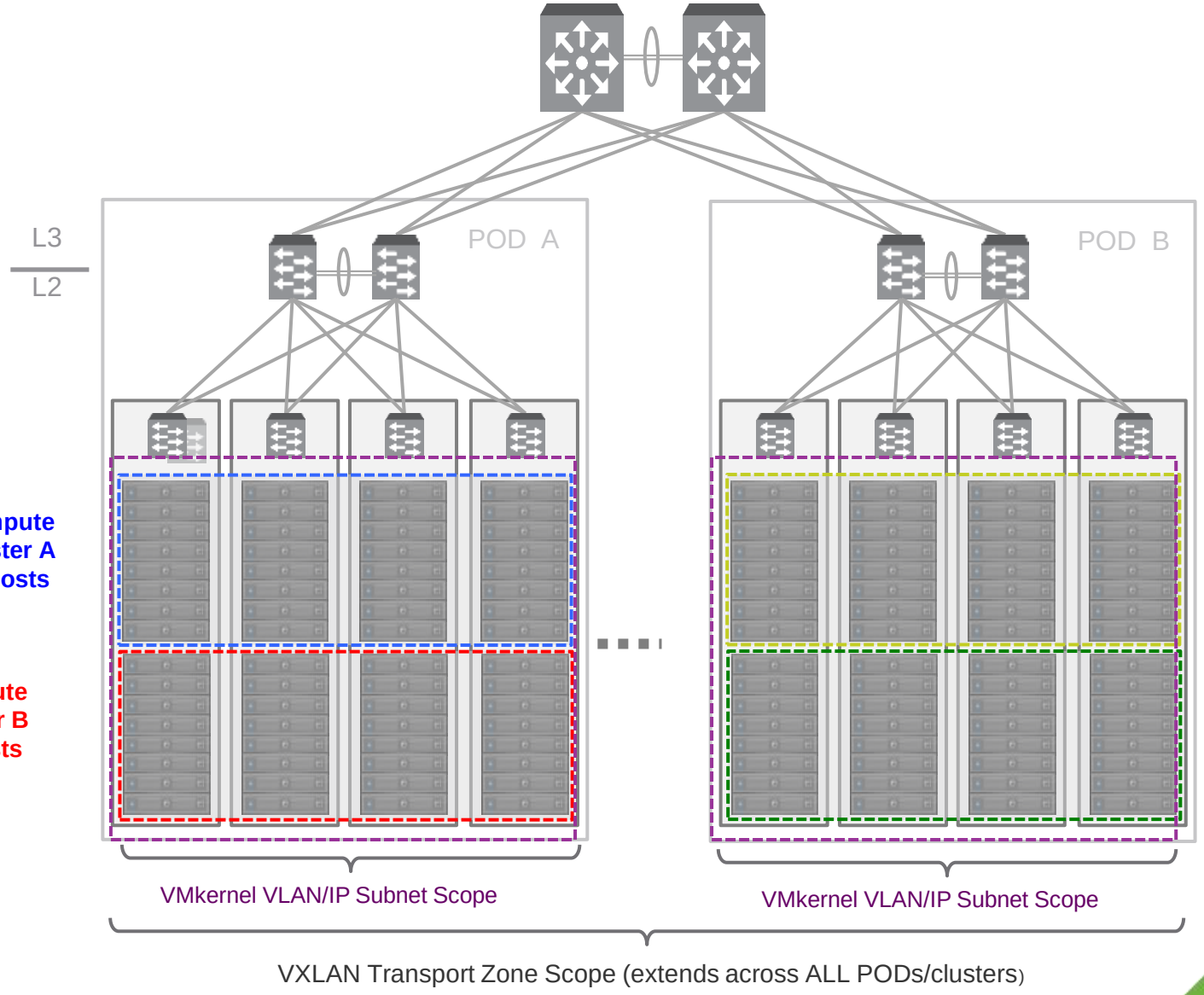
vmware

常见的IP地址/VLAN规划方案1

计算集群机架- IP 地址和VLAN划分		
功能	VLAN ID	IP 子网
Management	66	10.66.Y.0/24
vMotion	77	10.77.Y.0/24
VXLAN	88	10.88.Y.0/24
Storage	99	10.99.Y.0/24

L2 Fabric

二层域的划分 – Y代表POD编号
例如，第16个POD，管理网段为10.66.16.0/24



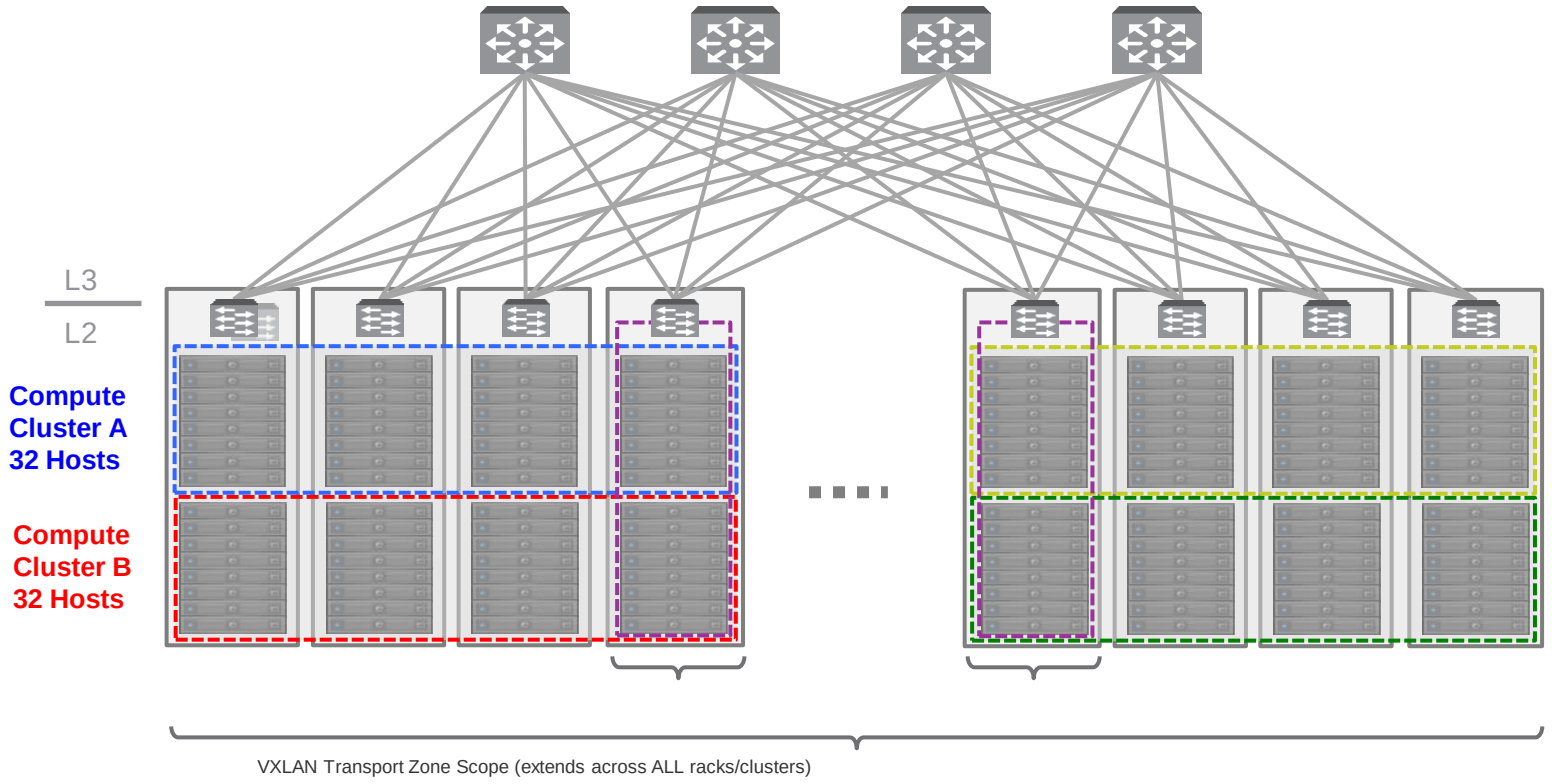
vmware

常见的IP地址/VLAN规划方案2

计算集群机架- IP 地址和VLAN划分		
功能	VLAN ID	IP 子网
Management	66	10.66.R.x/26
vMotion	77	10.77.R.x/26
VXLAN	88	10.88.R.x/26
Storage	99	10.99.R.x/26

L3 Fabric

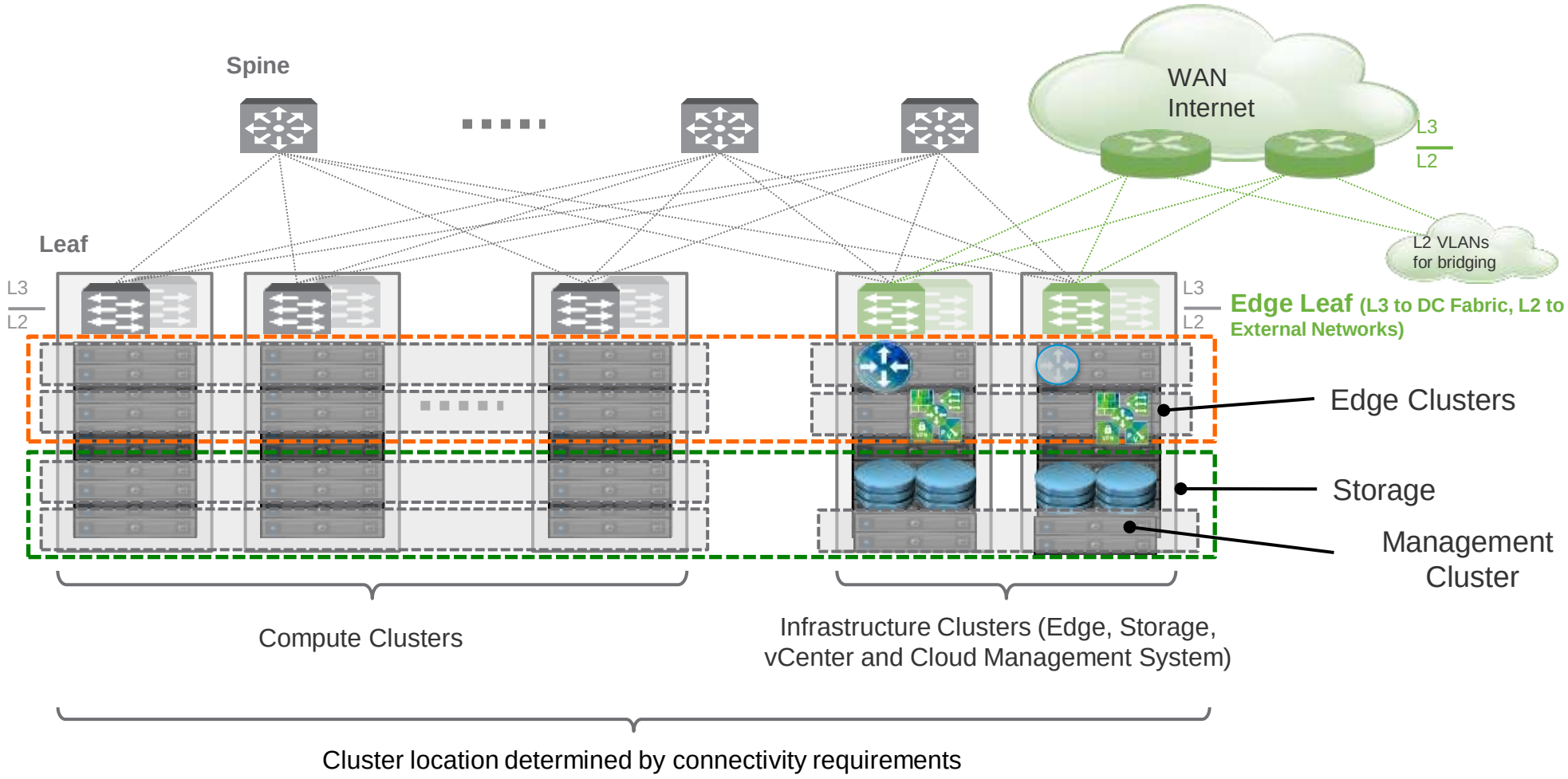
三层架构 – R代表 Rack的编号



提纲

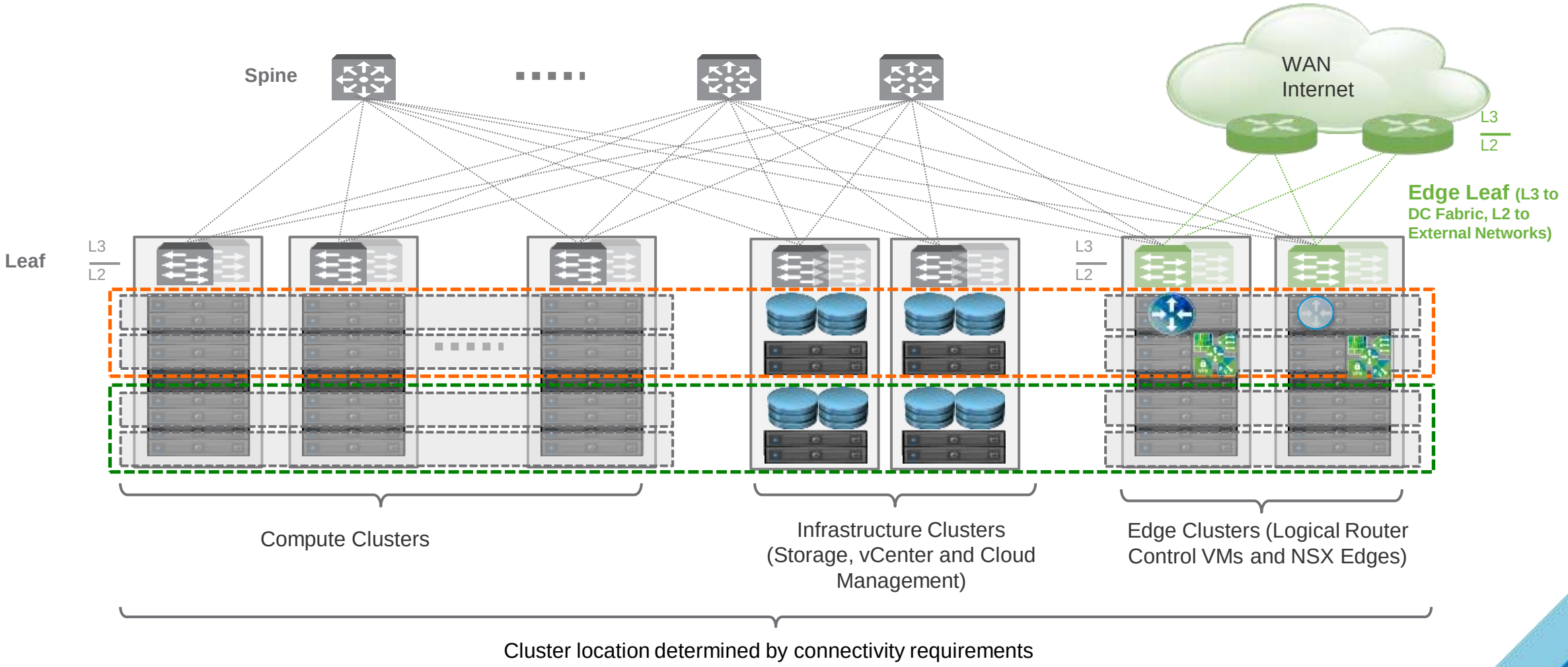
1	NSX原理、组件及组件之间的关系
2	物理网络设计的要求及VLAN规划
3	NSX安装实践
4	NSX安全部署实践
5	NSX路由及高可靠性设计
6	基于NSX的双活/灾备数据中心

NSX集群设计——整合Edge/Management（中小规模部署）



vmware

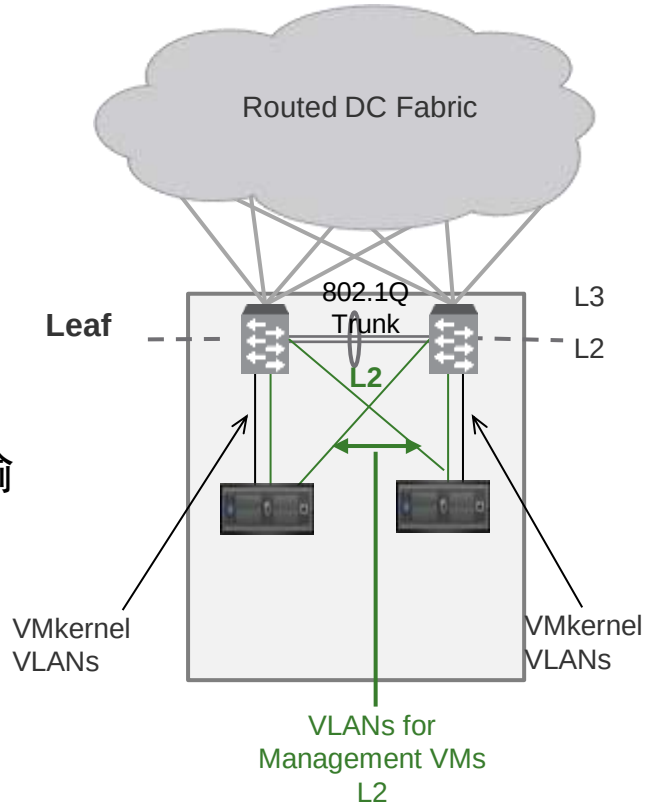
NSX集群设计——分离Edge和Management（大规模部署）



vmware

管理集群部署最佳实践

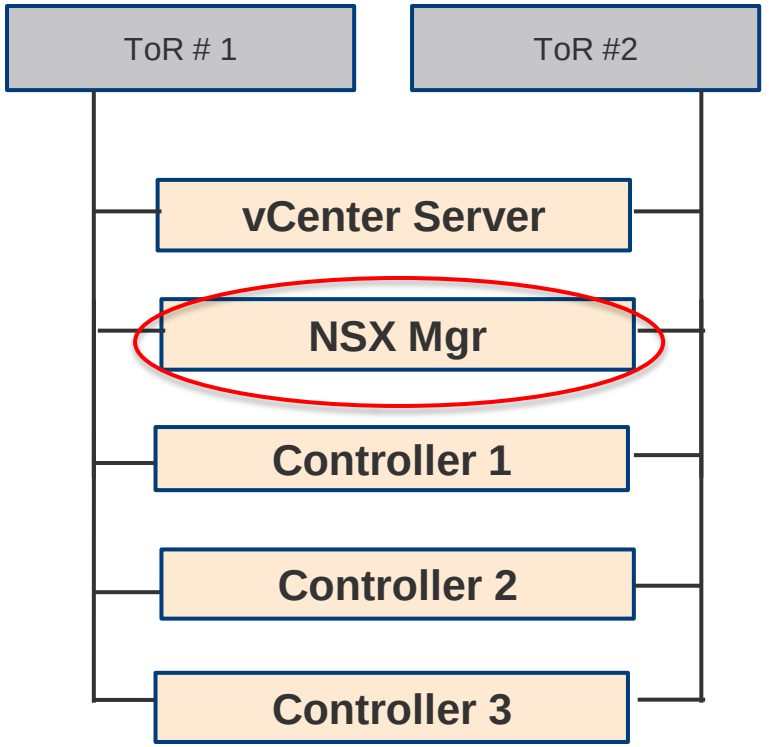
- 管理集群部署考虑:
  - 管理组件的CPU & 内存消耗
    - VC, NSX manager, NSX Controller, vRA, vROPs etc.
  - 静态分配, 空闲资源可以用于其他
- 机架设计需要主机到TOR的冗余上行链路，保证管理VLAN传输
  - 支持LACP teaming模式，管理数据流二层连通
- 大型部署使用专用的管理集群
  - 不启用VXLAN
- 小型部署环境中，管理集群和Edge集群可以合并
  - NSX manager和Controllers自动DFW excluded
  - 将vCenter server放入DFW exclusion list
  - 使用资源预留机制保护Edge虚拟机



Management Cluster Connectivity

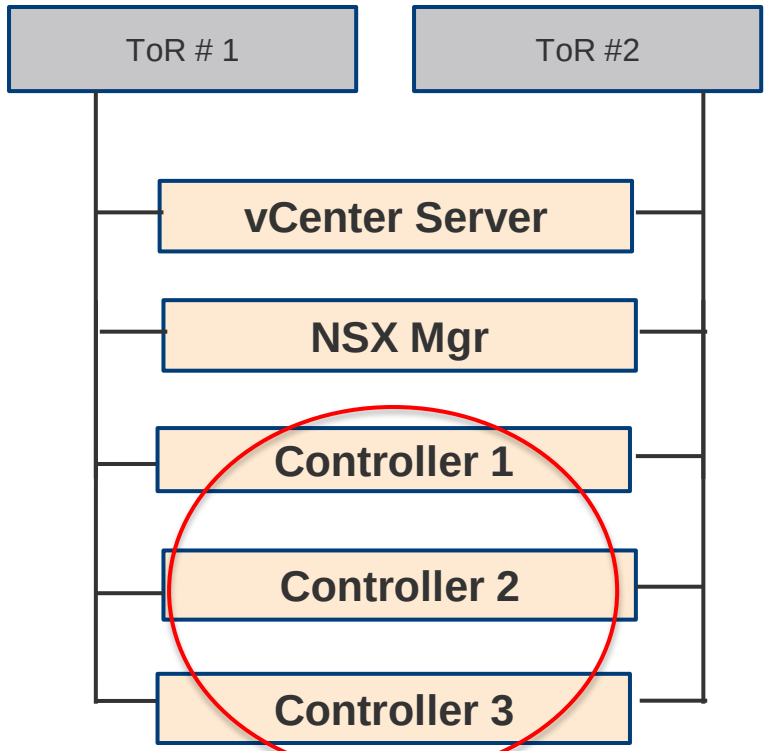
NSX Manager部署最佳实践

- NSX Manager ova虚拟机部署
  - 最小16 GB内存，4 vCPU
  - 大型环境部署24 GB内存，8 vCPU
- 为vCenter预留内存确保Web客户端速度
- 推荐定期备份NSX Manager
  - NSX Manager 管理平面故障不影响数据转发



NSX Controllers部署最佳实践

- Controller节点也是虚拟机
  - 4 vCPU, 4GB每节点
  - CPU预留2048 MHz
  - 不支持修改配置虚拟机配置
- 部署在管理或Edge集群
- 控制器集群支持3个Controller节点
- 控制器集群支持N+1冗余，即使Controller全部失效，数据转发不受影响
- 部署的时候默认不打开DRS和anti-affinity rules
  - 推荐手工开启DRS和anti-affinity rules
  - 最少需要3台物理主机



NSX Edge Services Gateway Sizing

Edge Services Gateway Form	vCPU	Memory MB	Specific Usage
X-Large	6	8192	Suitable for L7 High Performance LB
Quad-Large	4	1024	Suitable for high performance ECMP or FW deployment
Large	2	1024	Small to Medium DC or multi-tenant
Compact	1	512	Small Deployments or single service use or PoC

安装NSX Manager

- NSX Manager 安装包为OVA文件
- 通过VC部署
- 勾选复选框接受其他配置选项 **(Accept extra configuration options)**
- 选择部署位置

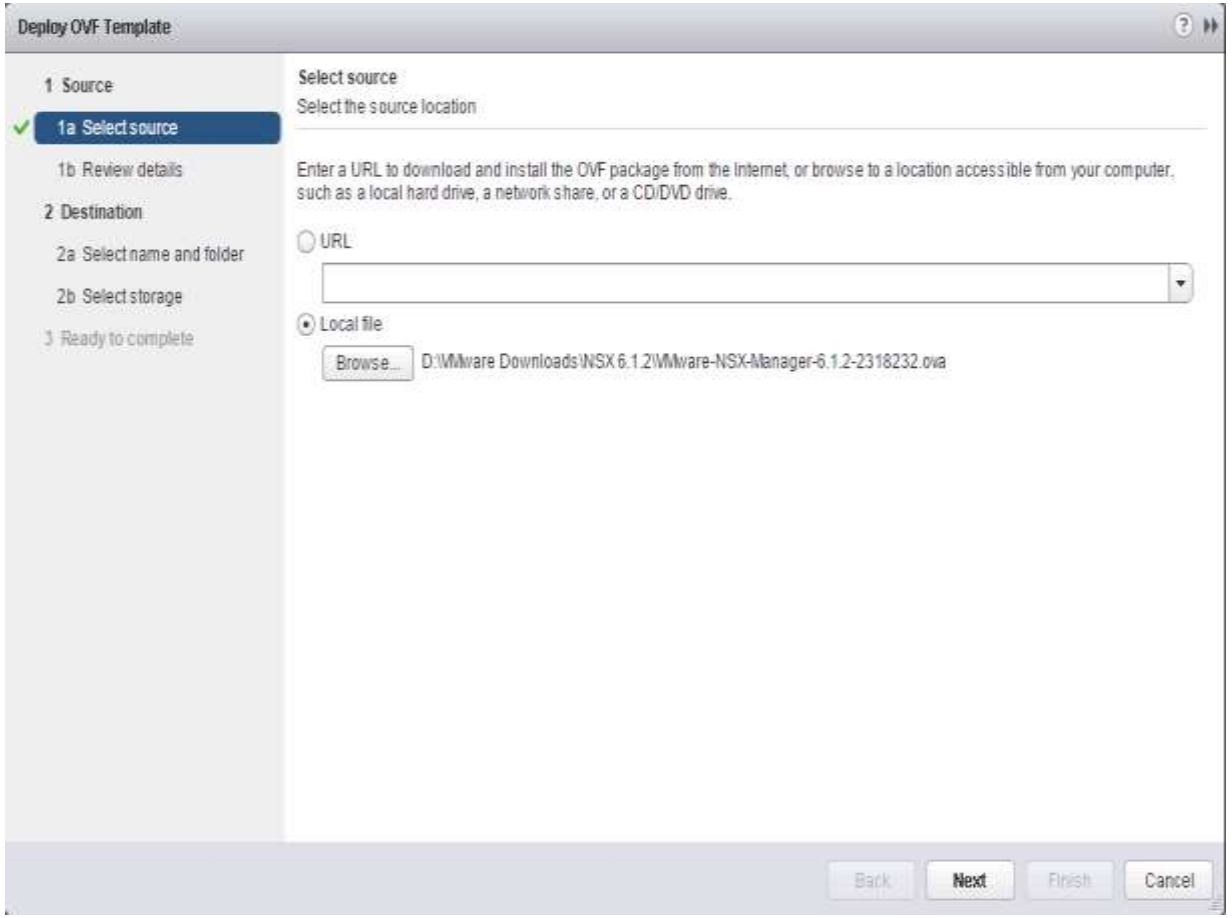
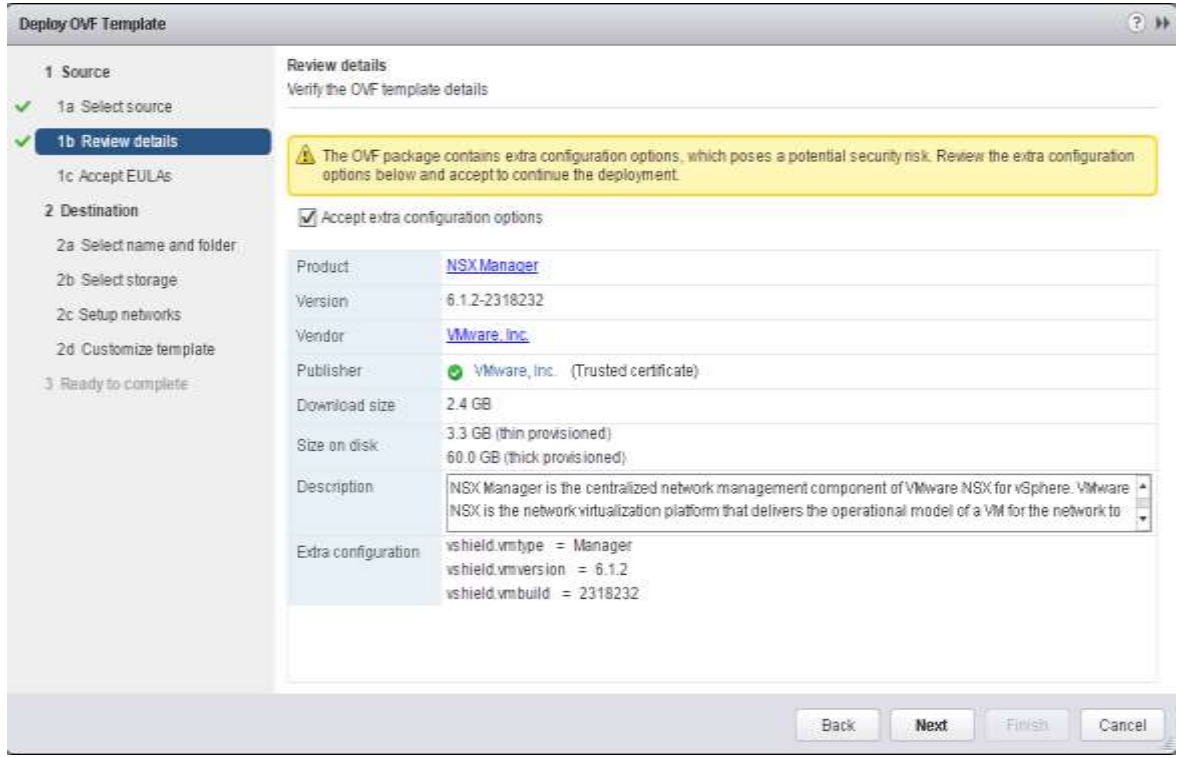


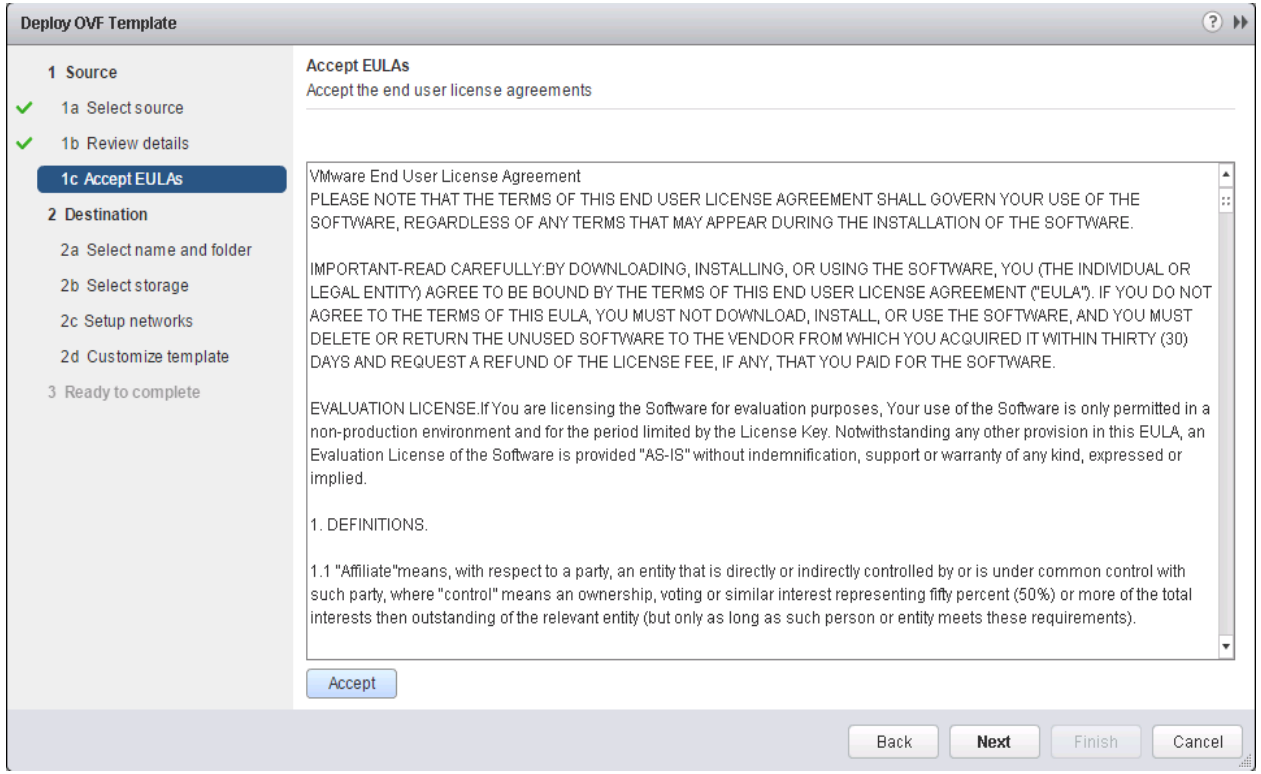
表 2-1 硬件要求

组件	最低要求
内存	■ NSX Manager：16 GB（对于某些 NSX 部署规模*为 24 GB） ■ NSX Controller：4 GB ■ NSX Edge 小型：512 MB，大型：1 GB，四倍大尺寸：1 GB，超大型：8 GB ■ Guest Introspection：1 GB ■ NSX Data Security：512 MB
磁盘空间	■ NSX Manager：60 GB ■ NSX Controller：20 GB ■ NSX Edge 小型、大型和四倍大尺寸：512 MB，超大型：4.5 GB（含 4 GB 交换） ■ Guest Introspection：4GB ■ NSX Data Security：每个 ESX 主机 6 GB
vCPU	■ NSX Manager：4（对于某些 NSX 部署规模*为 8） ■ NSX Controller：4 ■ NSX Edge 小型：1，大型：2，四倍大尺寸：4，超大型：6 ■ Guest Introspection：2 ■ NSX Data Security：1

检查硬件资源表

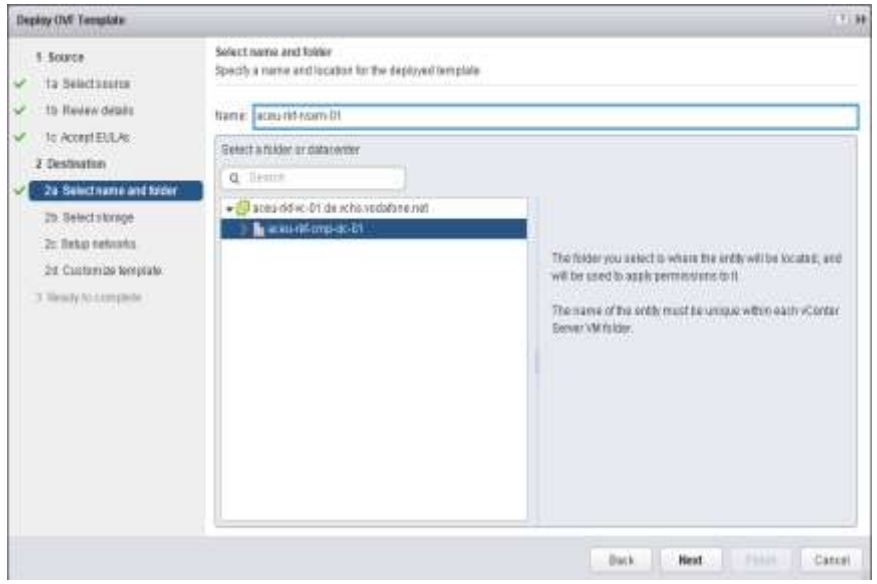
安装NSX Manager步骤



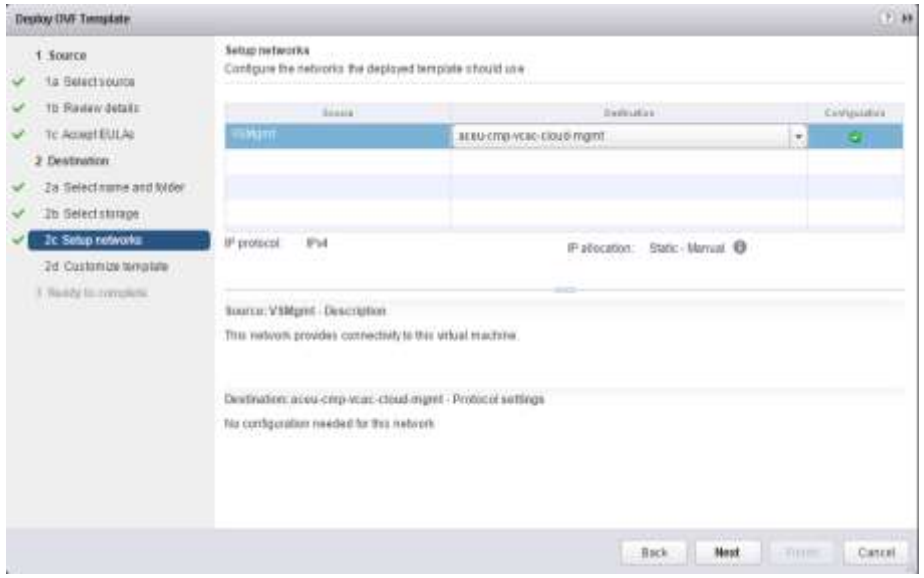


校验安装包并软件同意许可

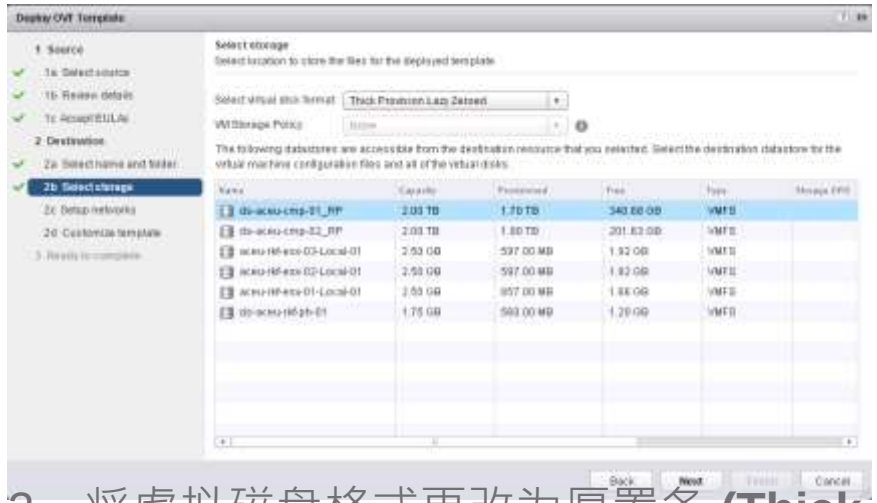
vmware



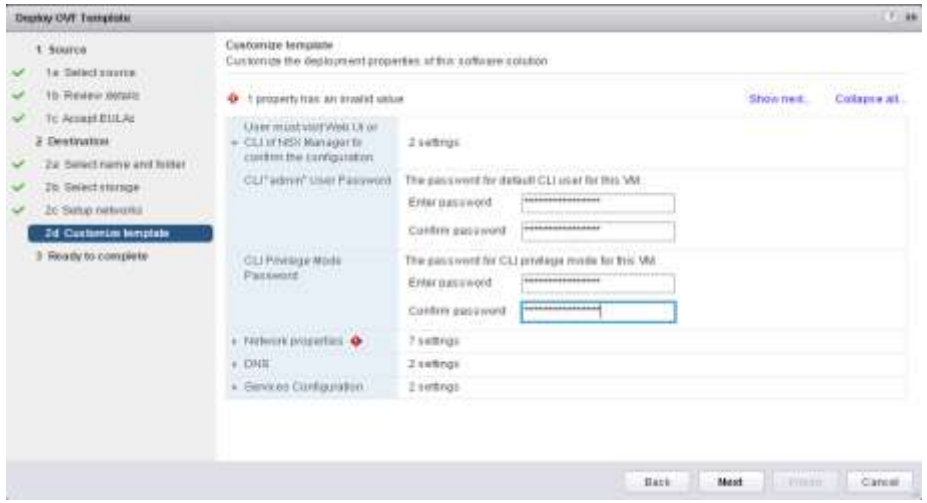
1、选择要在其上部署 NSX Manager 设备的主机或群集



3、选择 NSX Manager 安装的的端口组



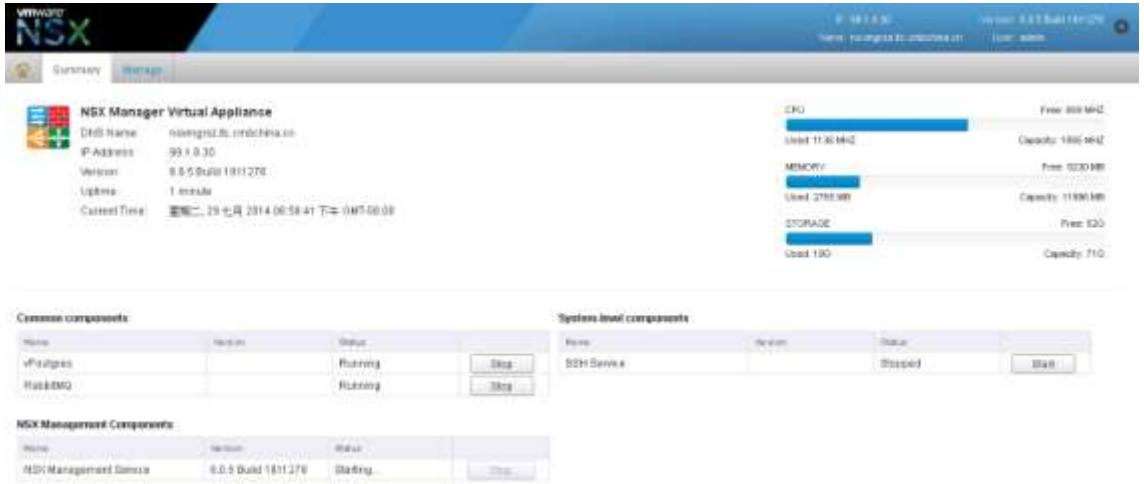
2、将虚拟磁盘格式更改为厚置备 **(Thick Provision)**，并为虚拟机配置文件和虚拟磁盘选择目标数据存储



4、NSX 管理员设置和NSX Manager网络配置

vmware

配置NSX Manager



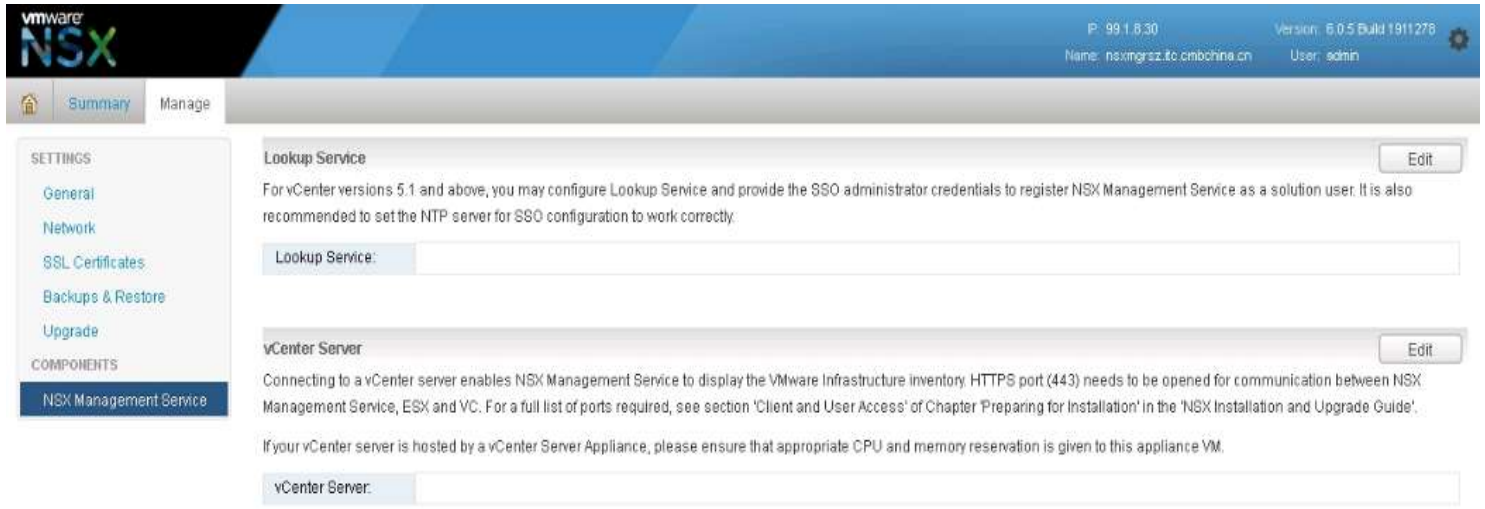
使用安装期间设置的密码以 **admin** 身份登录之后,

单击查看汇总 **(View Summary)**,并确保以下服务正在运行:

- vPostgres
- RabbitMQ
- NSX 管理服务

vmware

NSX注册到VCenter

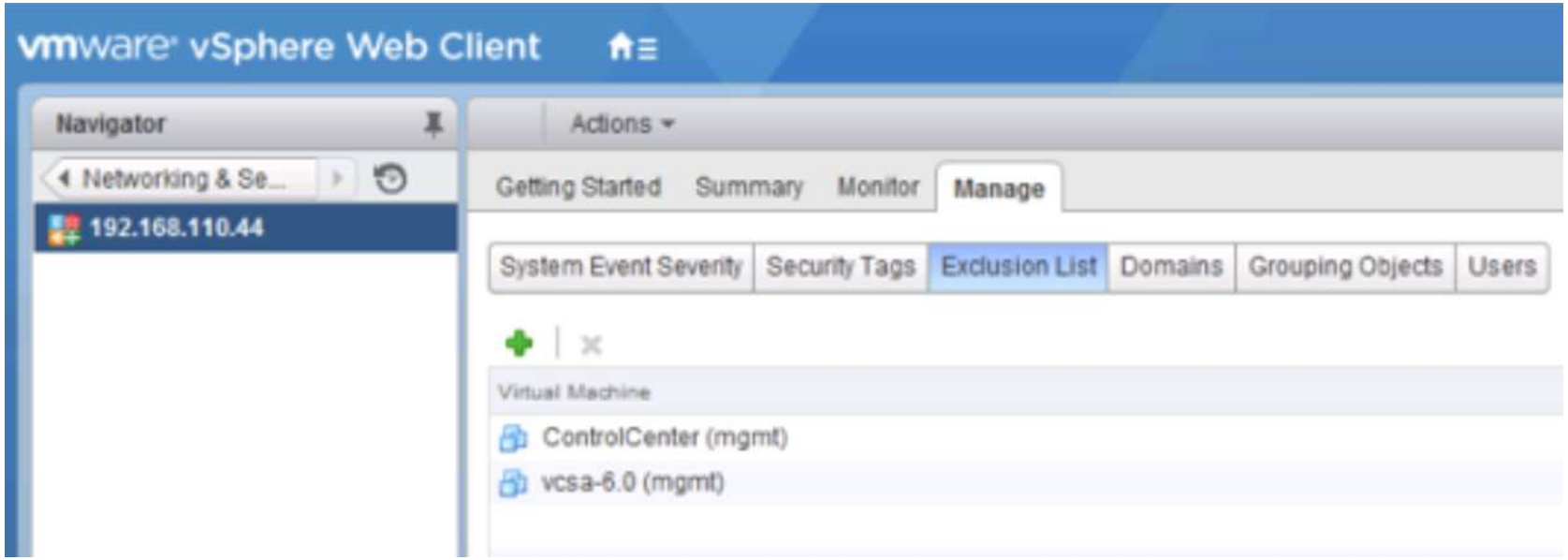


将NSX注册到VCenter Server：

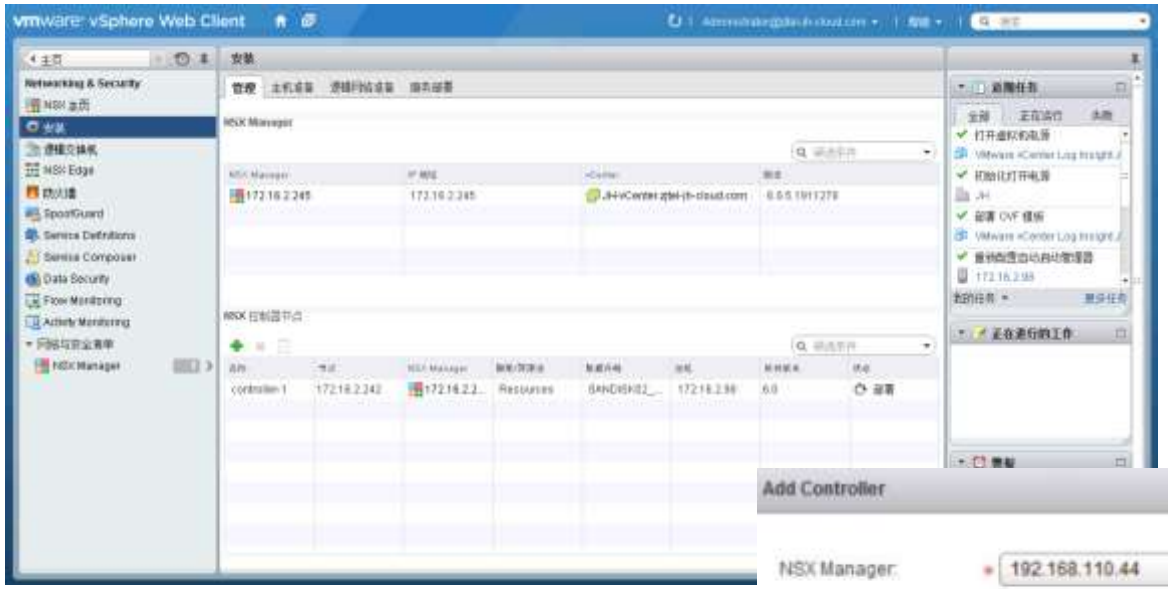
- 请输入相关的ID和PWD
- 通过同一ID登录VC

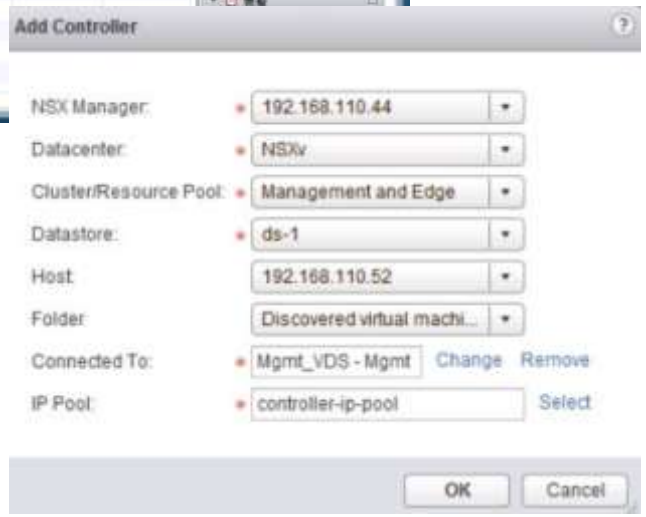
从防火墙保护中排除虚拟机

* 非常重要



部署NSX控制器



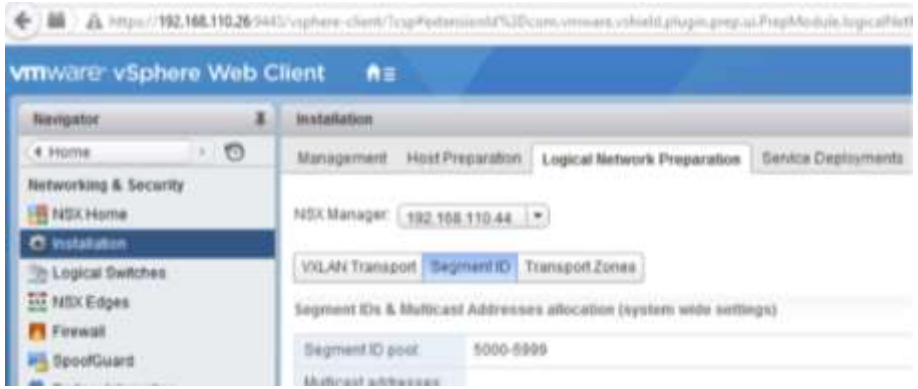


- 单击添加节点 （ + ） 图标
- 在完全部署第一个控制器后,部署其他两个控制器。 必须具有 3 个控制器，并设置反相关
- 后续完成 主机准备
- 逻辑网络准备

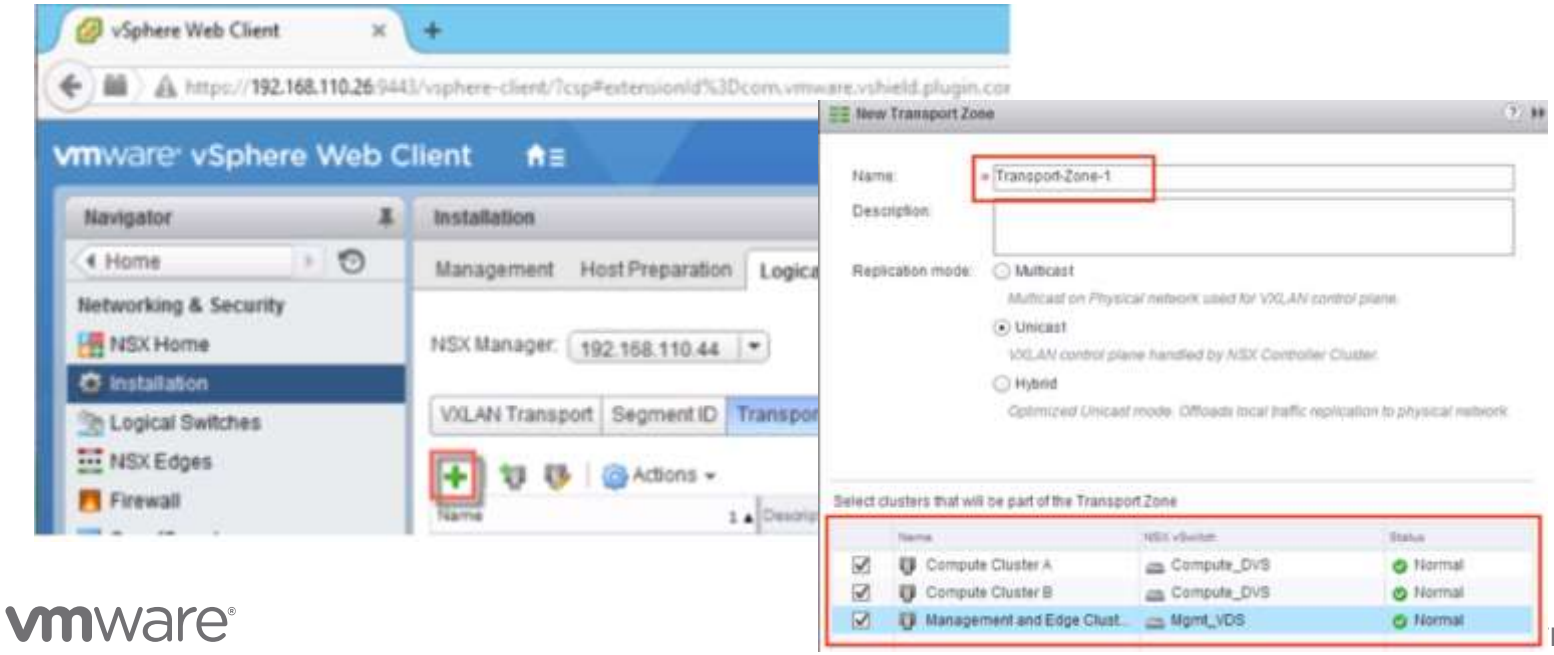
部署NSX 组件

- 主机准备
  - 开始NSX主机准备流程之前,务必要确保群集处于已解决状态—这意味着群集的操作**(Actions)**列表中不 显示解决**(Resolve)** 选项
  - 单击所有群集的齿轮图标,然后单击安装 **(Install)**
  - 点击配置 配置VxLAN
- 逻辑网络准备

部署NSX 组件



<一配置网段 ID



<一配置传输区域

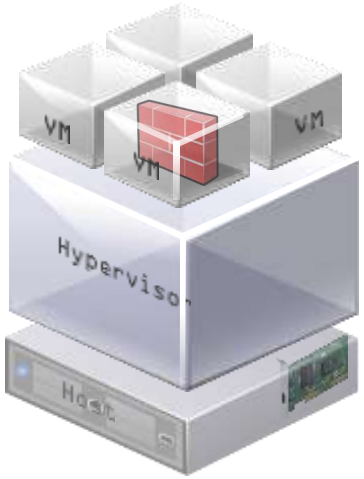
提纲

1	NSX原理、组件及组件之间的关系
2	物理网络设计的要求及VLAN规划
3	NSX安装实践
4	NSX安全部署实践
5	NSX路由及高可靠性设计
6	基于NSX的双活/灾备数据中心

分布式防火墙与传统火墙区别

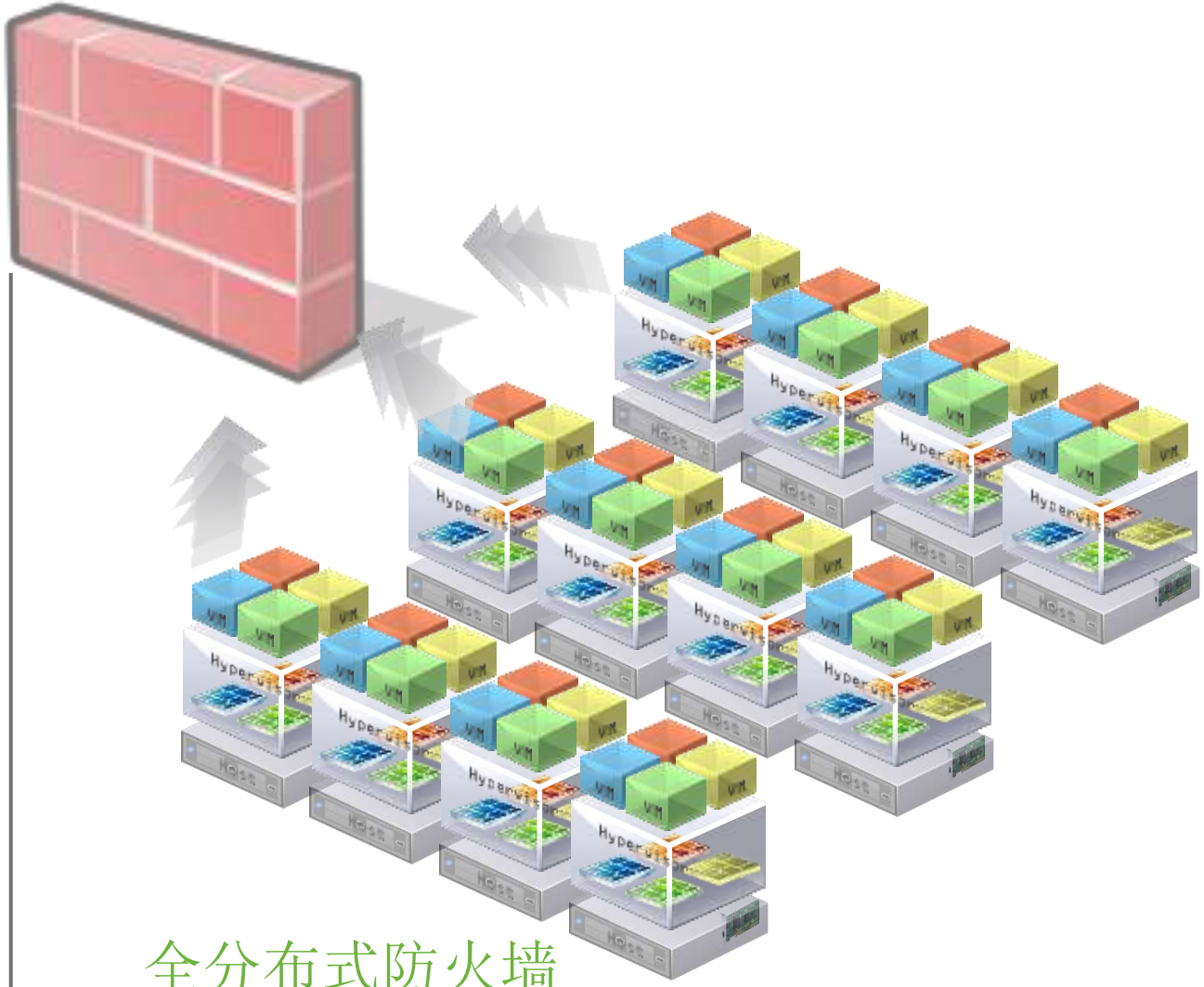
物理防火墙

- 传统防火墙规则管理和运维方式
- 集中式的瓶颈点
- 物理防火墙性能~100G



虚拟防火墙

- 传统防火墙规则管理和运维方式
- 集中式的瓶颈点
- **虚拟**防火墙性能~1-3G



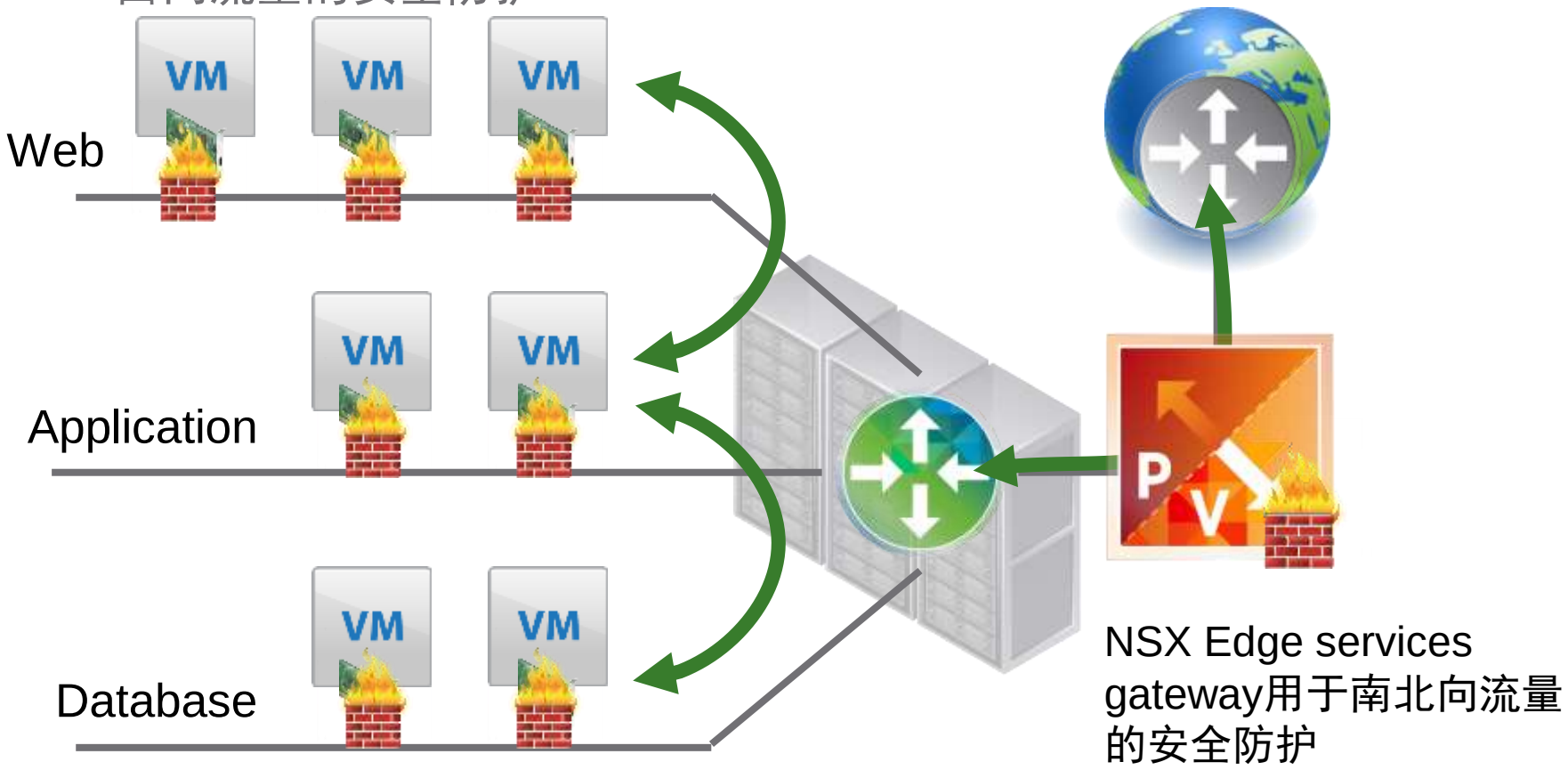
全分布式防火墙

- 自动化的策略管理和运维
- 全分布式的执行策略
- 基于vSphere内核的高性能
- 分布式横向扩展容量（20Gbps/每主机）

vmware

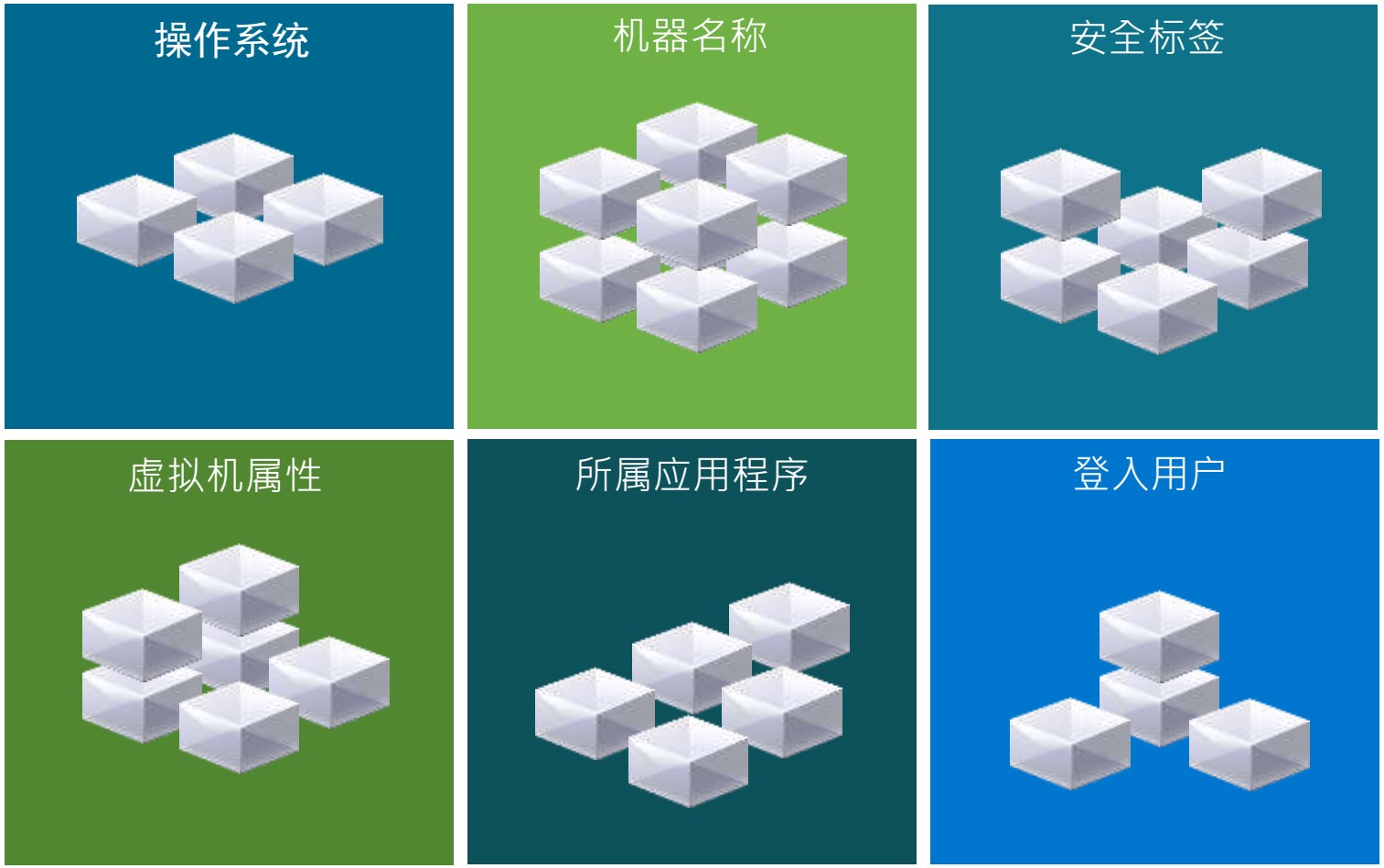
NSX内核级分布式防火墙

分布式内核级防火墙实现东西向流量的安全防护



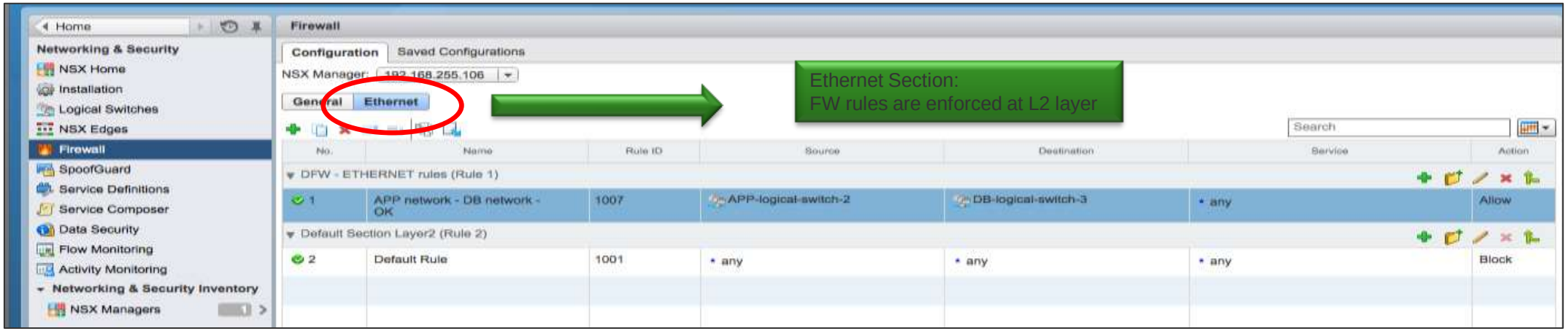
NSX Edge services gateway用于南北向流量的安全防护

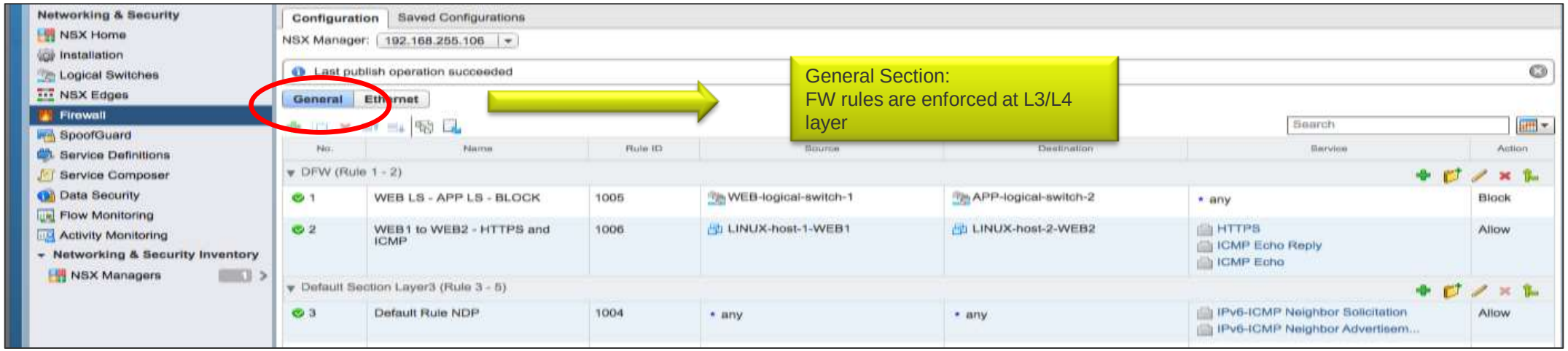
安全策略的灵活性



使用NSX微分段，安全管理员可以用多样性的动态条件来建立面向对象的安全策略，与网络拓扑无关

策略


Home
Networking & Security
NSX Home
Installation
Logical Switches
NSX Edges
Firewall
SpoofGuard
Service Definitions
Service Composer
Data Security
Flow Monitoring
Activity Monitoring
Networking & Security Inventory
NSX Managers
Firewall
Configuration
Saved Configurations
NSX Manager: 192.168.255.106
General
Ethernet
Ethernet Section:
FW rules are enforced at L2 layer
Search
No.
Name
Rule ID
Source
Destination
Service
Action
DFW - ETHERNET rules (Rule 1)
1
APP network - DB network - OK
1007
APP-logical-switch-2
DB-logical-switch-3
any
Allow
Default Section Layer2 (Rule 2)
2
Default Rule
1001
any
any
any
Block


Networking & Security
NSX Home
Installation
Logical Switches
NSX Edges
Firewall
SpoofGuard
Service Definitions
Service Composer
Data Security
Flow Monitoring
Activity Monitoring
Networking & Security Inventory
NSX Managers
Configuration
Saved Configurations
NSX Manager: 192.168.255.106
Last publish operation succeeded
General
Ethernet
General Section:
FW rules are enforced at L3/L4 layer
Search
No.
Name
Rule ID
Source
Destination
Service
Action
DFW (Rule 1 - 2)
1
WEB LS - APP LS - BLOCK
1005
WEB-logical-switch-1
APP-logical-switch-2
any
Block
2
WEB1 to WEB2 - HTTPS and ICMP
1006
LINUX-host-1-WEB1
LINUX-host-2-WEB2
HTTPS
ICMP Echo Reply
ICMP Echo
Allow
Default Section Layer3 (Rule 3 - 5)
3
Default Rule NDP
1004
any
any
IPv6-ICMP Neighbor Solicitation
IPv6-ICMP Neighbor Advertisem...
Allow

vmware

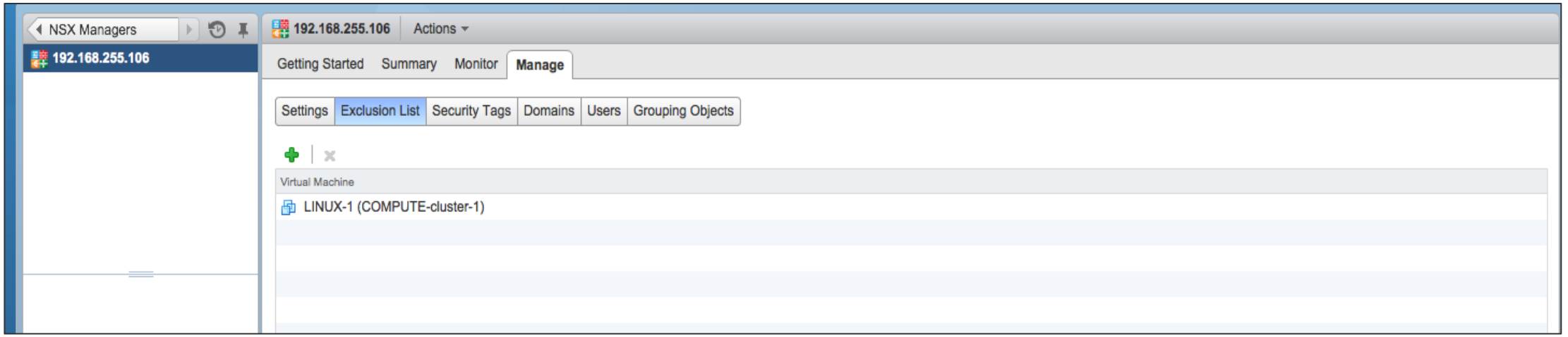
Applied To

Rule ID	Rule Name	Source	Destination	Service	Action	Applied To

Applied To 字段	描述
DataCenter	VMware Datacenter attribute
Cluster	VMware Cluster attribute
Distributed Virtual Port Group	Port Group of a dVS
Network	Network attribute
Virtual Machine	VM attribute
vNIC	vNIC attribute
Logical Switch	VXLAN logical switch

- 减轻负载
- 多租户IP地址重复

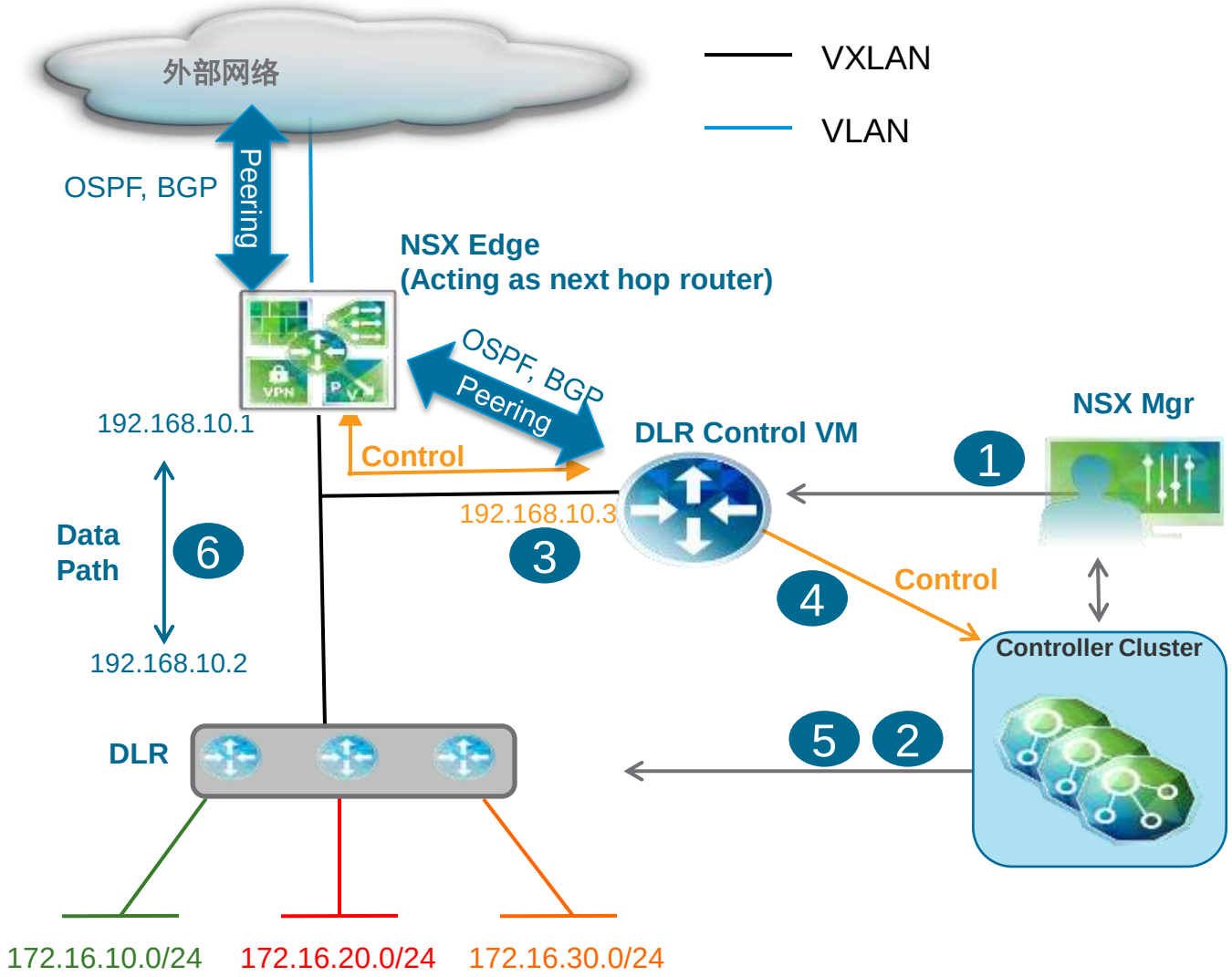
NSX 分布式防火墙 - Exclusion List



提纲

1	NSX原理、组件及组件之间的关系
2	物理网络设计的要求及VLAN规划
3	NSX安装实践
4	NSX安全部署实践
5	NSX路由及高可靠性设计
6	基于NSX的双活/灾备数据中心

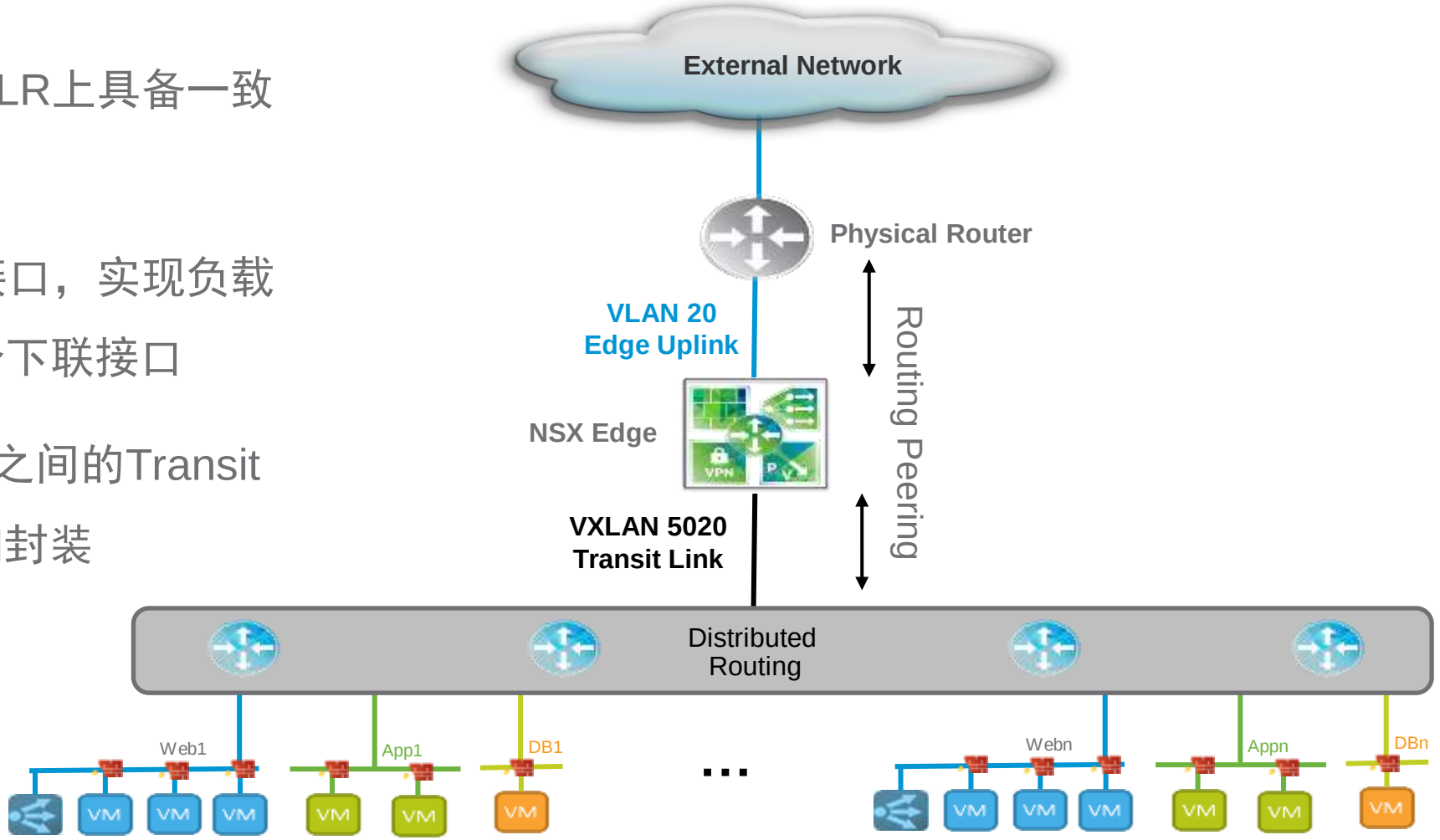
NSX逻辑路由的实现



1. 通过NSX Manager创建DLR实例，同时配置路由
2. Controller向DLR推送配置信息，包括DLR上的接口信息（LIFs）
3. 在NSX Edge和DLR control VM之间创建OSPF/BGP的peer邻居关系
4. DLR Control VM将学习到的的路由信息推送给Controller Cluster
5. Controller发送路由更新到所有的ESXi hosts
6. ESXi host根据路由信息指导数据包的转发

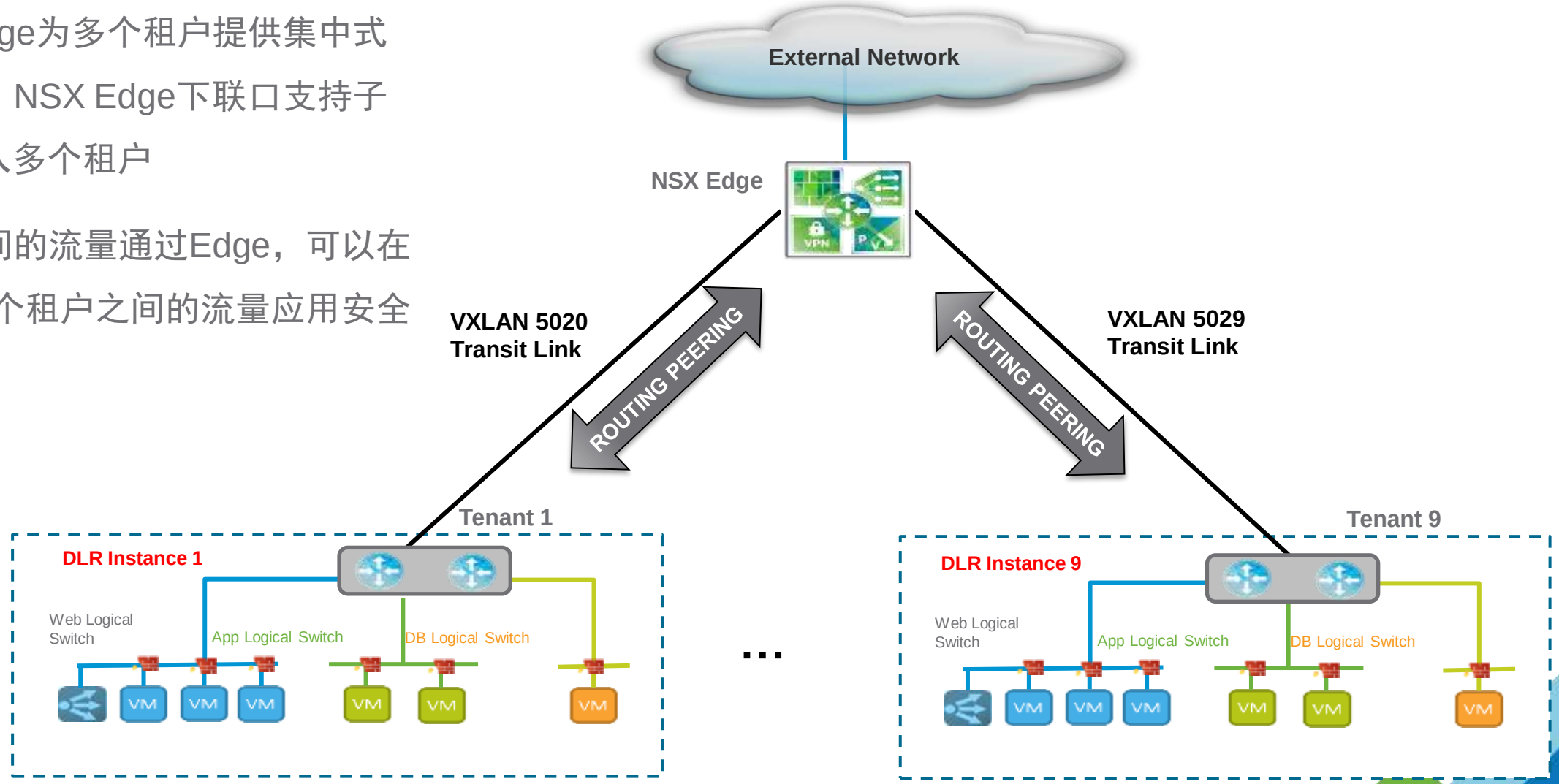
NSX单租户路由拓扑

- 优化东西向流量，DLR上具备一致的路由转发表向
- DLR具备多个上联接口，实现负载均衡，同时具有多个下联接口
- 在DLR和NSX Edge之间的Transit Link建议使用VxLAN封装

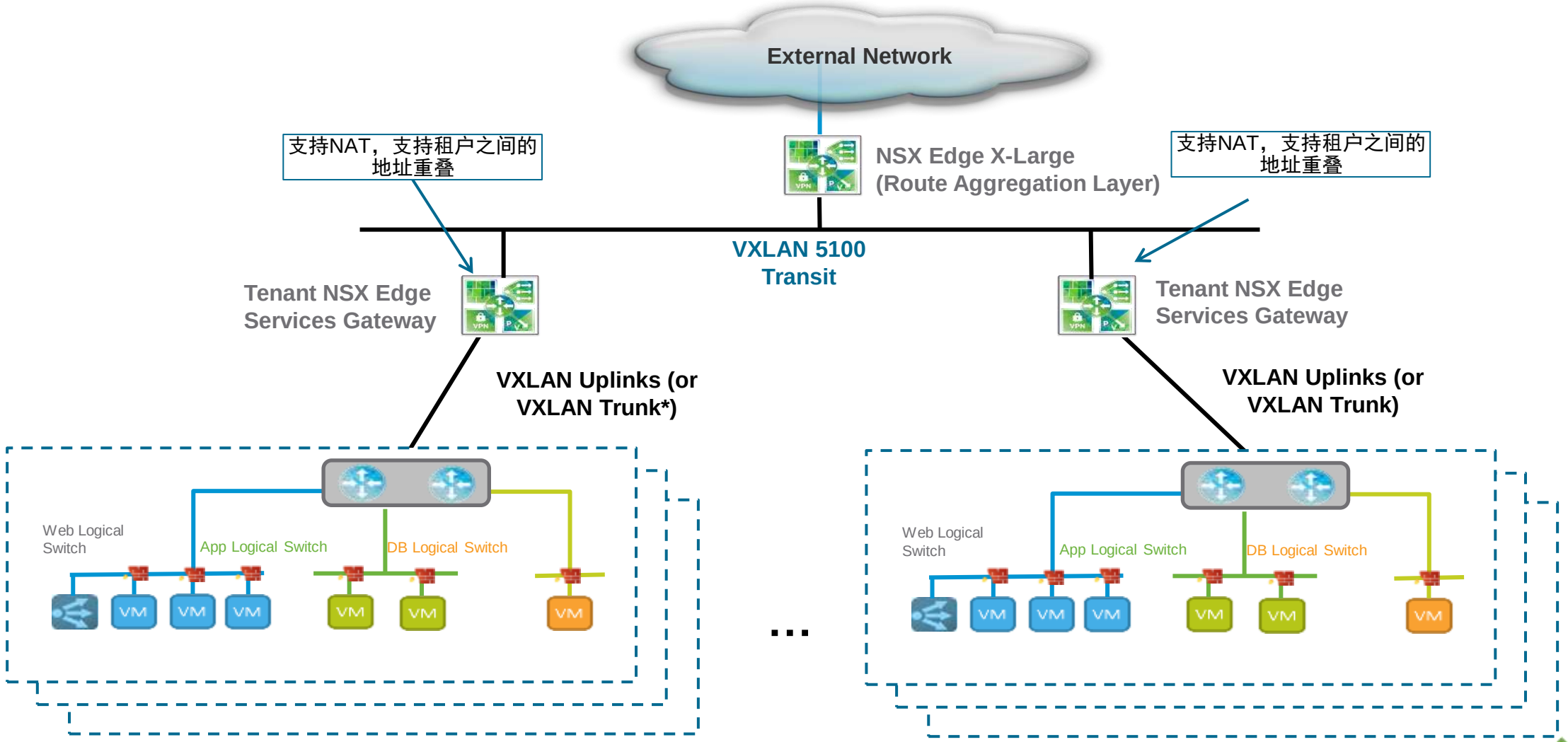


NSX多租户路由拓扑

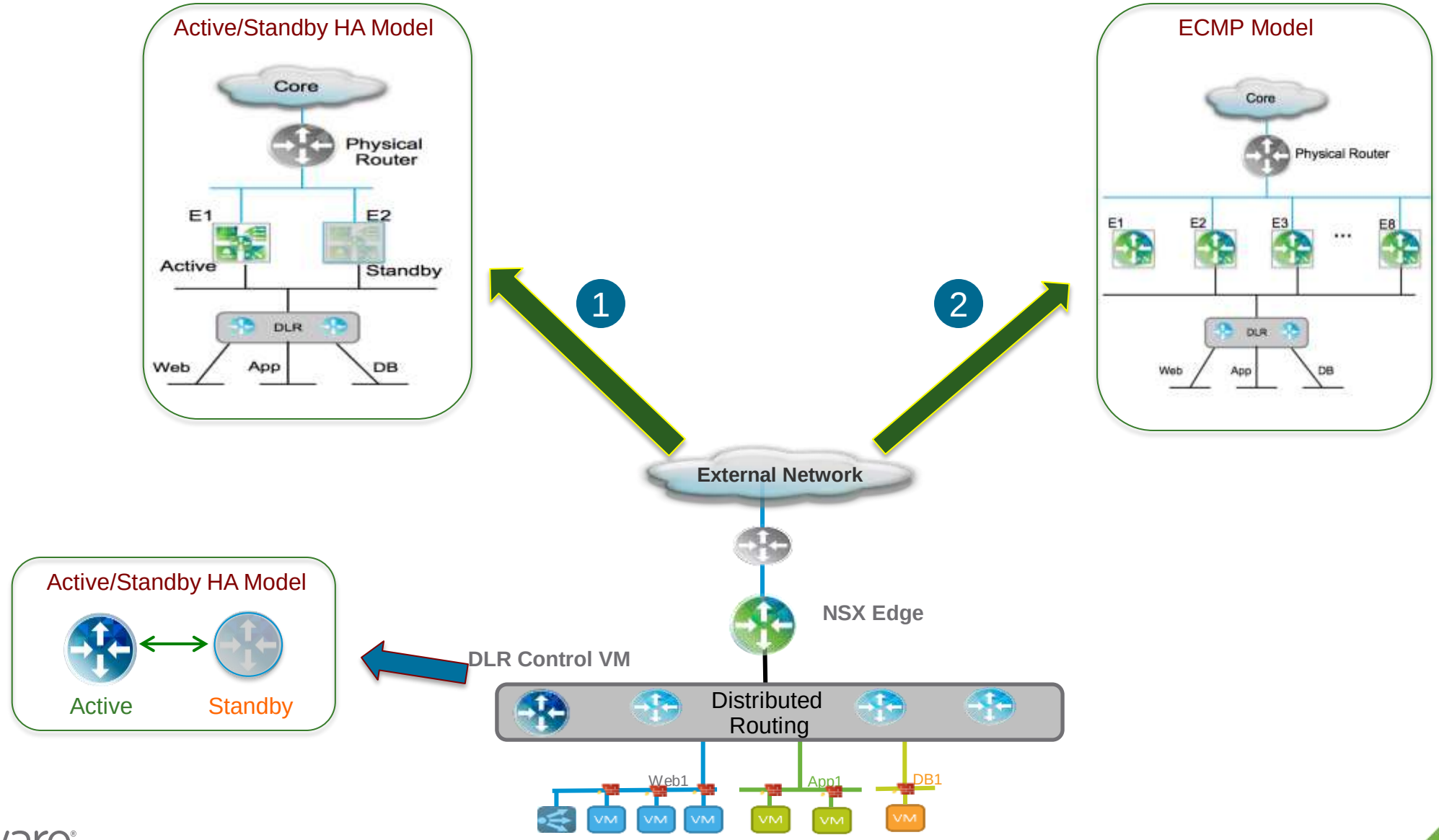
- 一个NSX Edge为多个租户提供集中式的路由功能，NSX Edge下联口支持子接口方式接入多个租户
- 每个租户之间的流量通过Edge，可以在Edge上为每个租户之间的流量应用安全策略



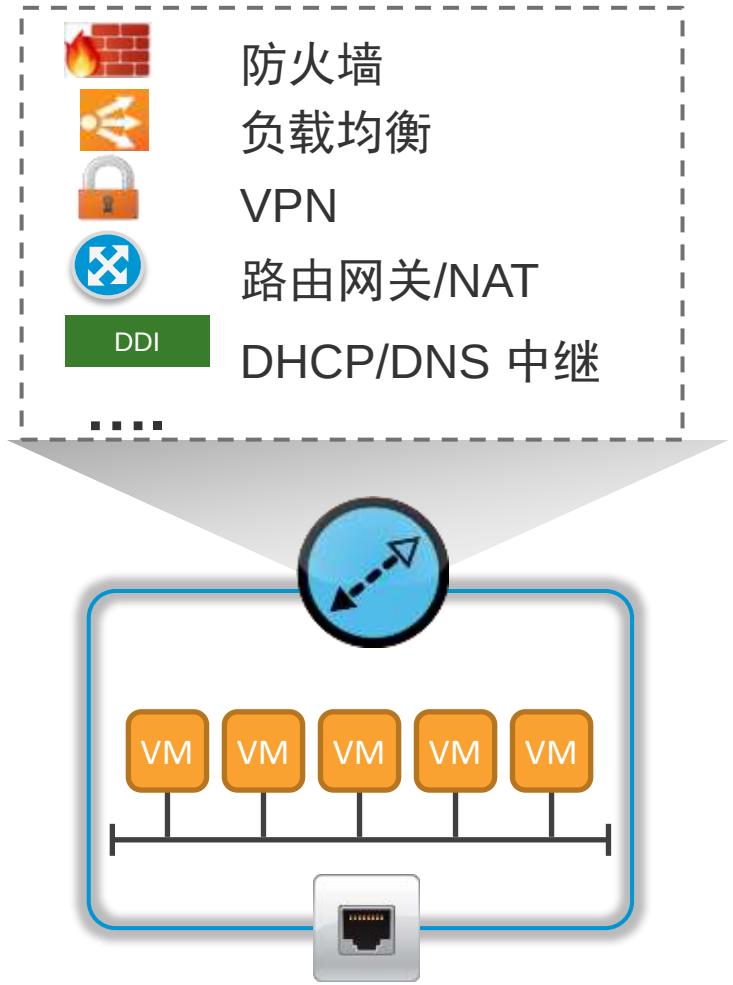
NSX高密度多租户拓扑



逻辑路由高可靠性模型


Active/Standby HA Model
Core
Physical Router
E1
E2
Active
Standby
DLR
Web
App
DB
ECMP Model
Core
Physical Router
E1
E2
E3
...
E8
DLR
Web
App
DB
1
2
External Network
NSX Edge
DLR Control VM
Distributed Routing
Web1
App1
DB1
VM
Active/Standby HA Model
Active
Standby

Edge网关：集成化多功能的服务



设计考虑

- 集群设计：通过配置Edge的数量，实现能力的扩展，通常部署在Edge集群
- Edge的HA或ECMP：可以实现两个Edge的HA和多个edge的ECMP
- 多租户服务：在每个租户内部可以有自己的Edge
- Edge与DLR control VM的路由设计以及上联物理网络路由，支持OSPF和BGP

Edge主备冗余模式

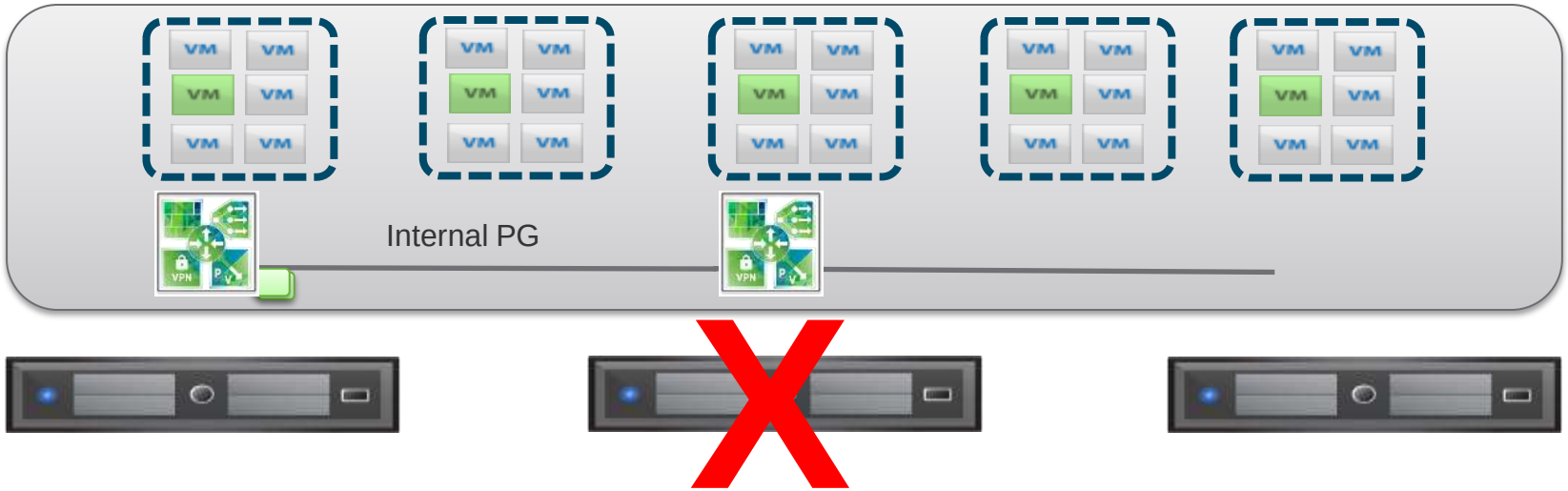
心跳和同步

- 使用internal vNic实现心跳和同步
- PG L2连通(VLAN 或 VXLAN)
- 状态切换:

  FW - connection tracking; LB - Sticky table;
  Routing - Graceful restart extensions to OSPF/BGP plus NSF via FIB sync

Anti-affinity

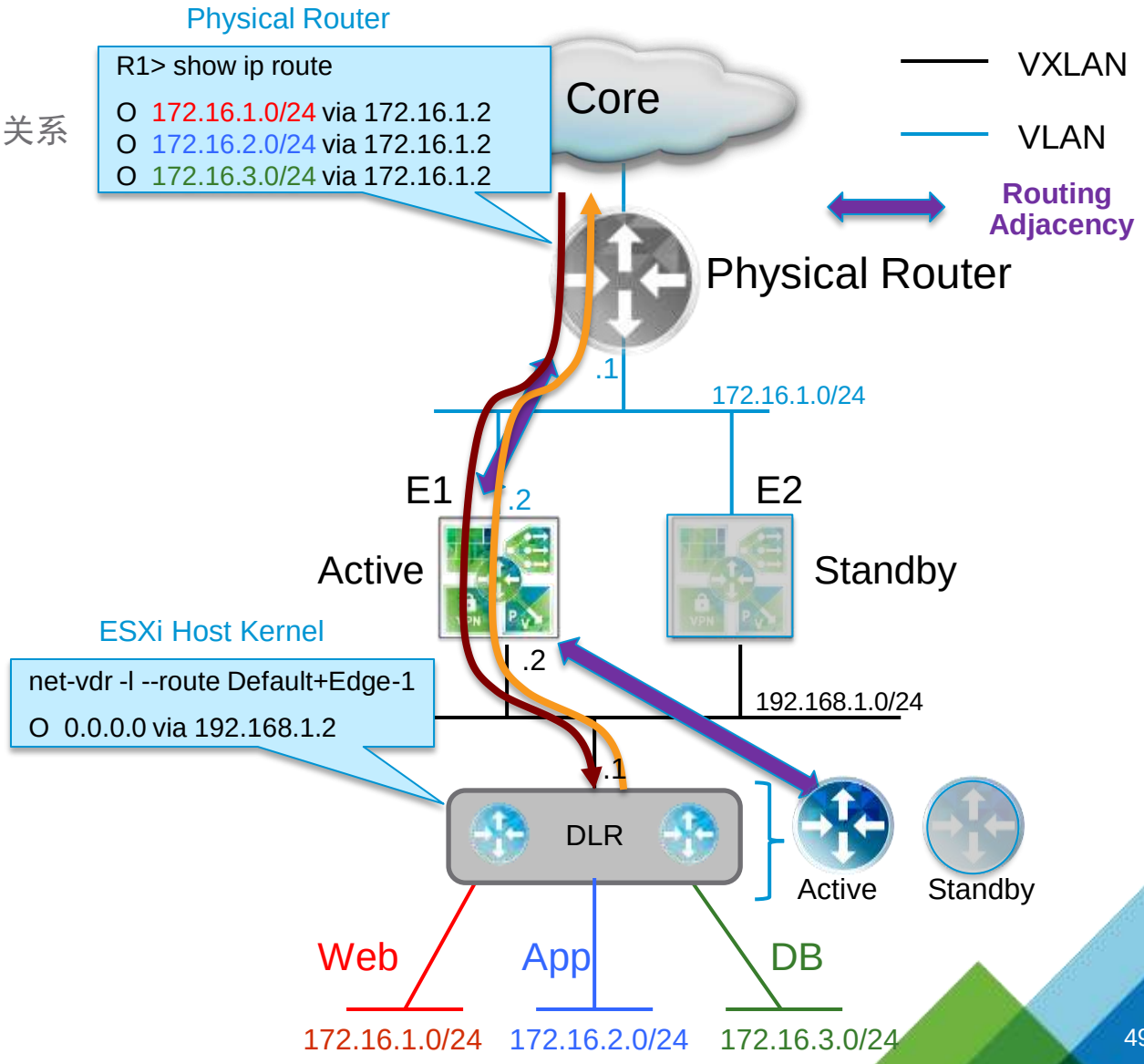
- 主备Edges放置于不同的ESXi主机
- 一旦Edge所在ESXi出现故障，NSX Manager将Edges放置于不同的主机



Active/Standby HA Model

- 所有南北向的流量都经Active NSX Edge转发

  Active NSX Edge仅与DLR Control VM和物理路由器之间建立邻居关系



Active/Standby HA Model

- 所有南北向的流量都经Active NSX Edge转发

  Active NSX Edge仅与DLR Control VM和物理路由器之间建立邻居关系

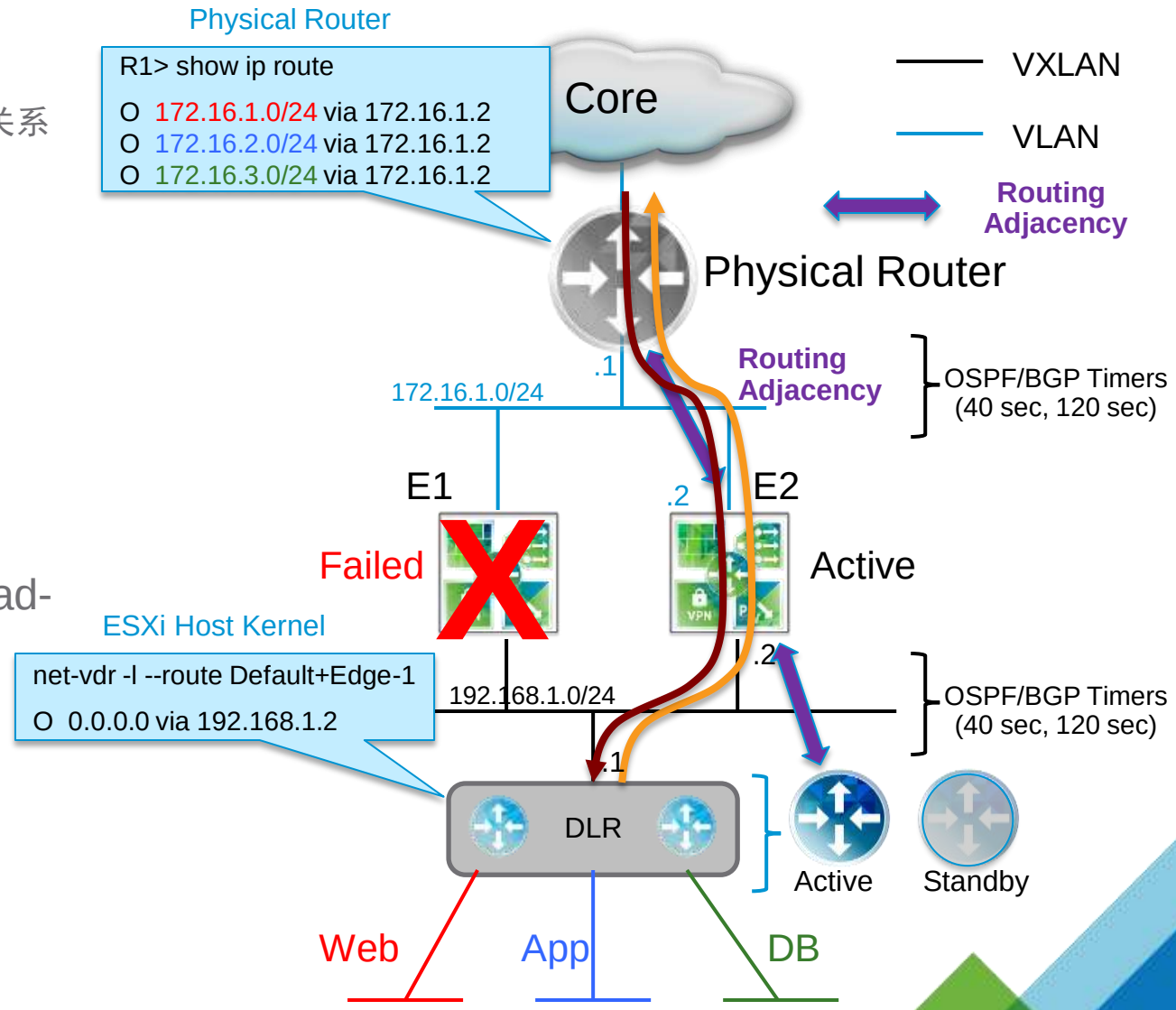
- Active NSX Edge E1 Failed:

  E2检测到E1 Failed，将自己的状态转变为Active

  流量转发利用E2与E1同步的表项进行转发

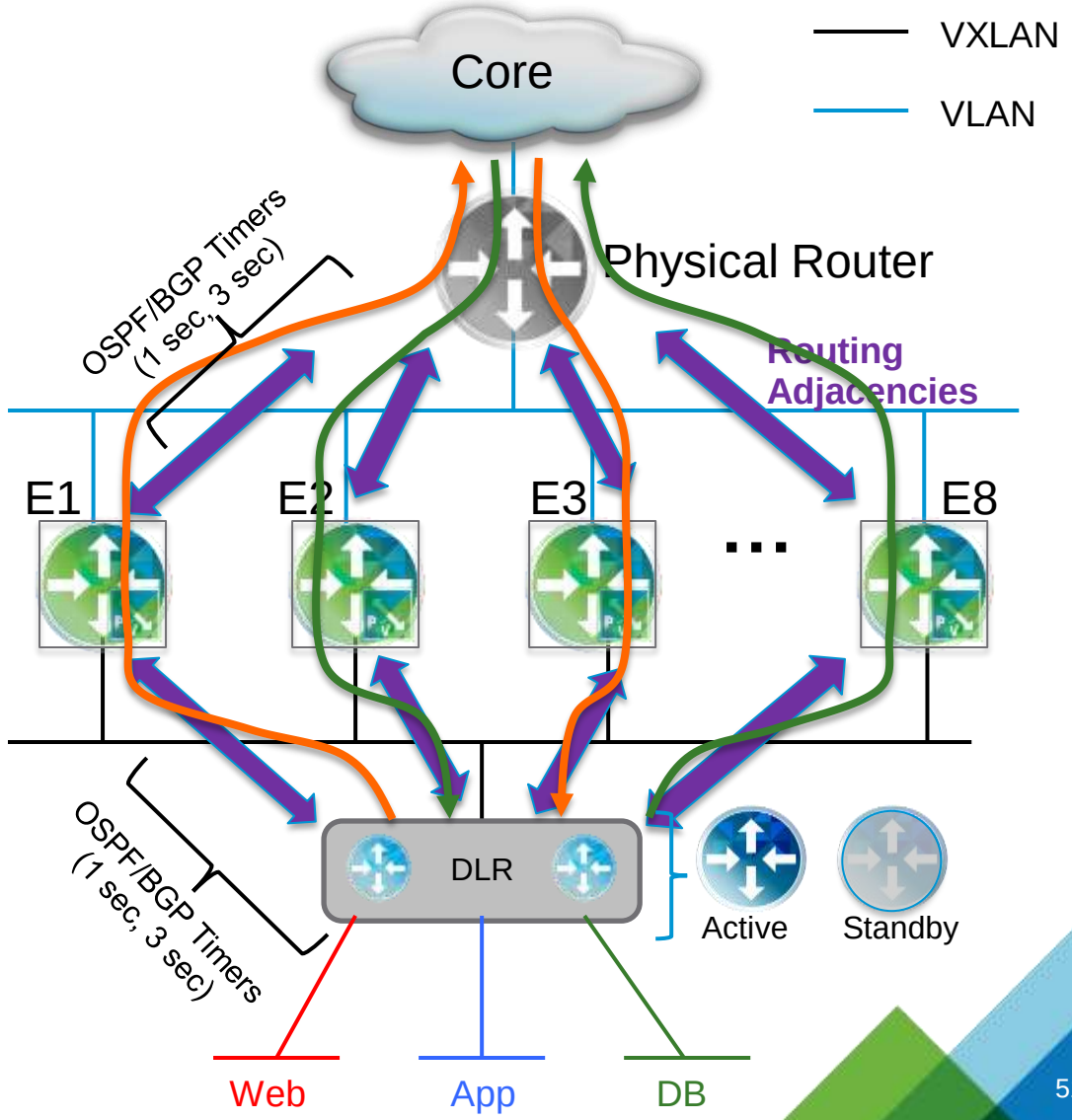
- vSphere HA用于HA的模型

- NSX Edge Stateful services是支持的，例如FW，Load-Balancing，NAT



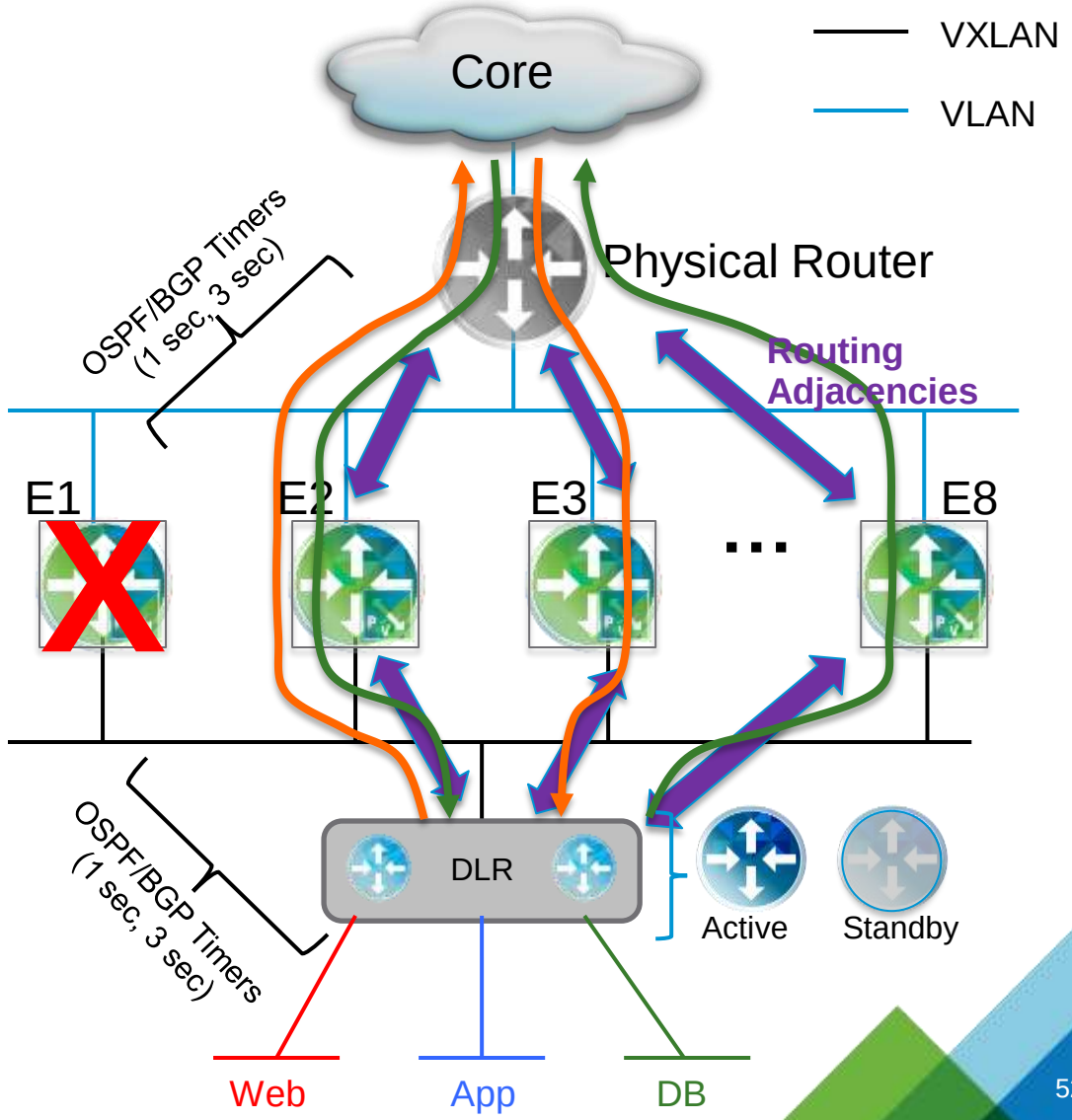
ECMP模型

- 所有NSX Edge都是Active，所有南北向流量都经过Edge
  - DLR Control VM与所有的Edge建立邻居关系，同时所有的Edge与物理路由器之间建立邻居关系
  - 南北向的流量在不同的Edge之间负载均衡



ECMP HA Model

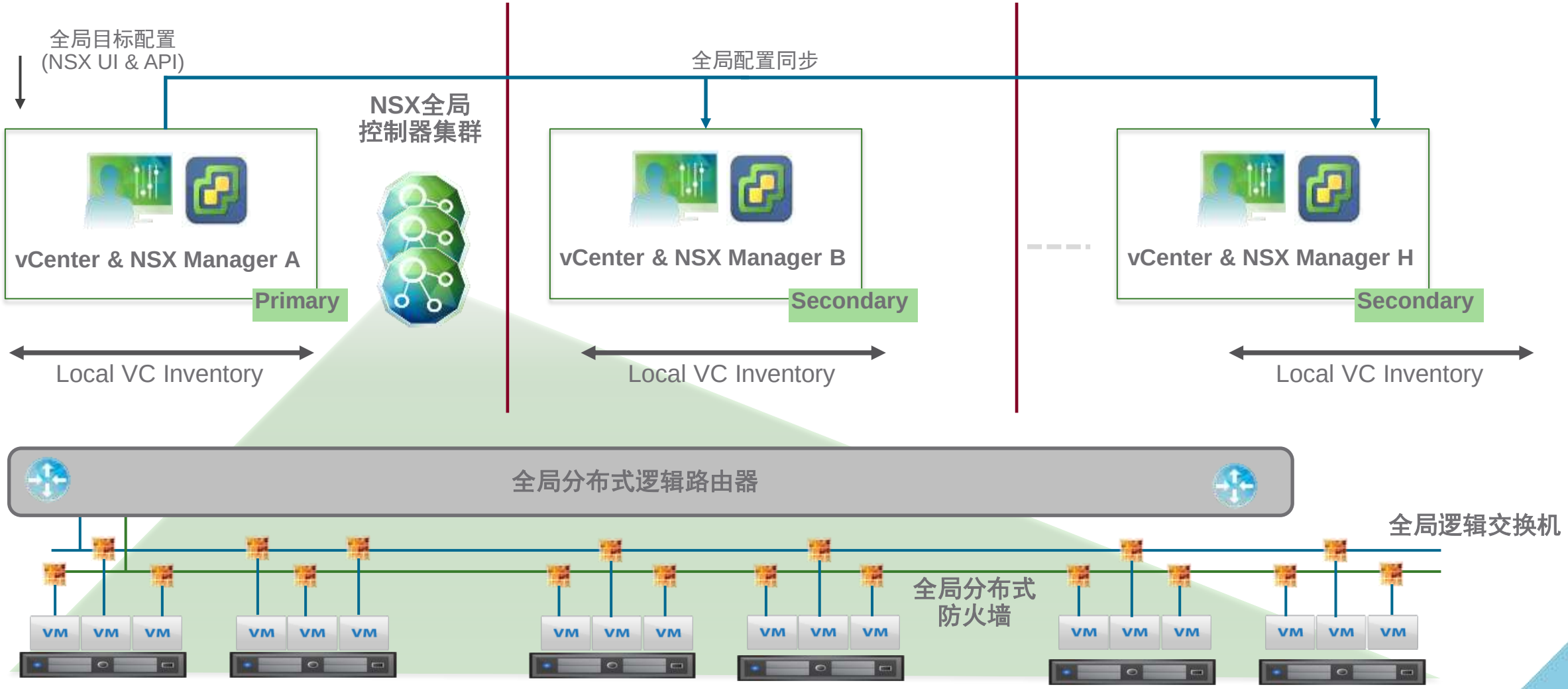
- 所有NSX Edge都是Active，所有南北向流量都经过Edge
  - DLR Control VM与所有的Edge建立邻居关系，同时所有的Edge与物理路由器之间建立邻居关系
  - 南北向的流量在不同的Edge之间负载均衡
- 当其中一个NSX Edge失效后，南北向流量通过其它Edge实现负载均衡



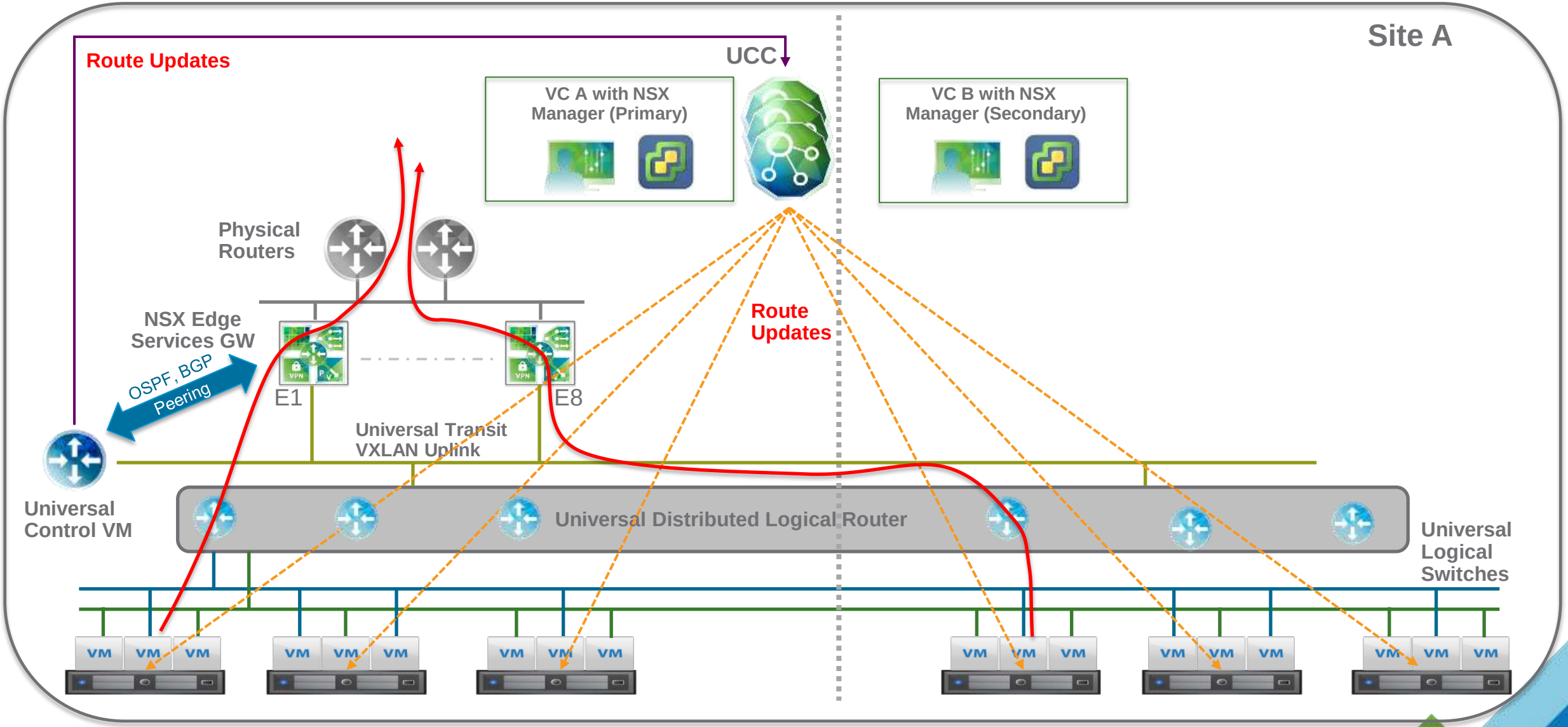
提纲

1	NSX原理、组件及组件之间的关系
2	物理网络设计的要求及VLAN规划
3	NSX安装实践
4	NSX安全部署实践
5	NSX路由及高可靠性设计
6	基于NSX的双活/灾备数据中心

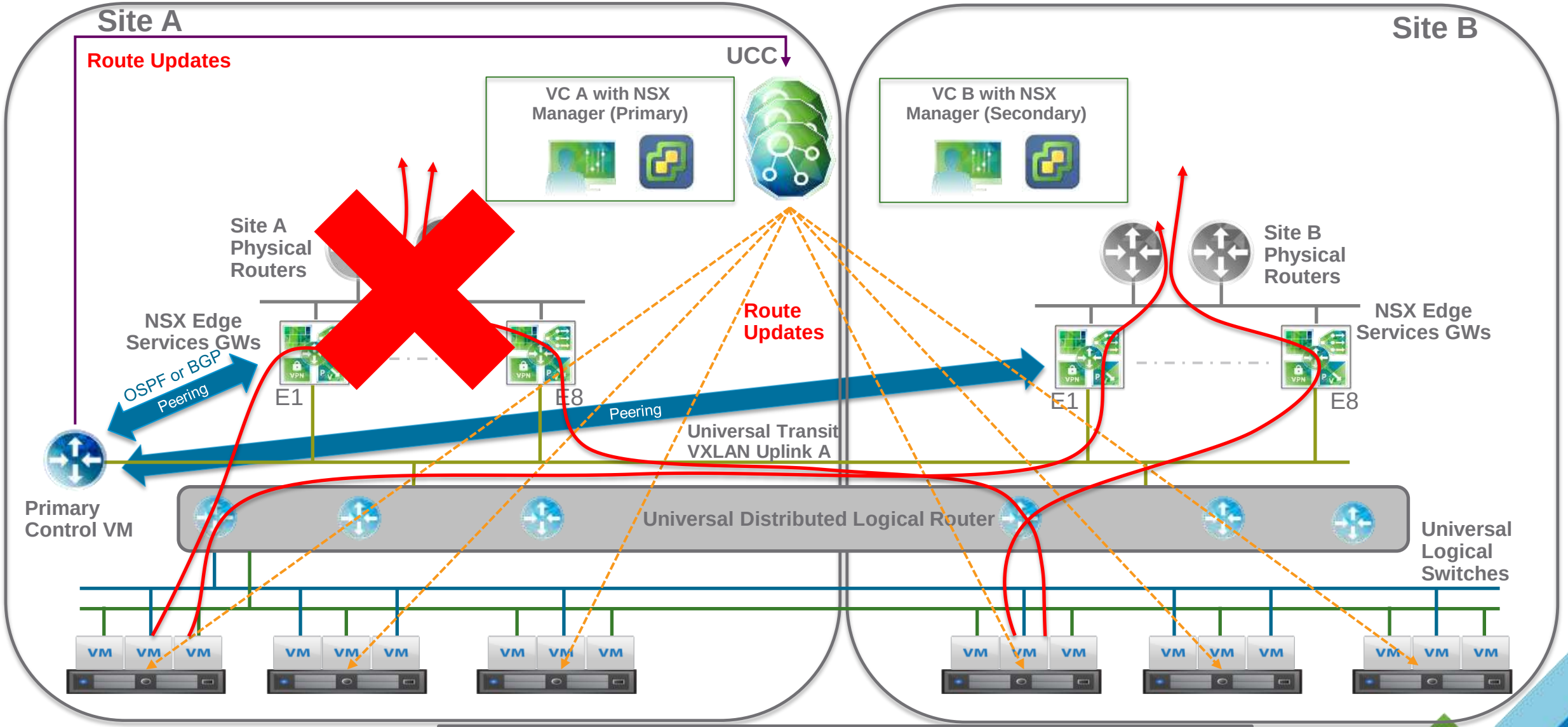
多vCenter、多站点逻辑网络 (NSX 6.2)



Cross-VC NSX– 部署场景1：单个站点多个vCenter A/P N-S

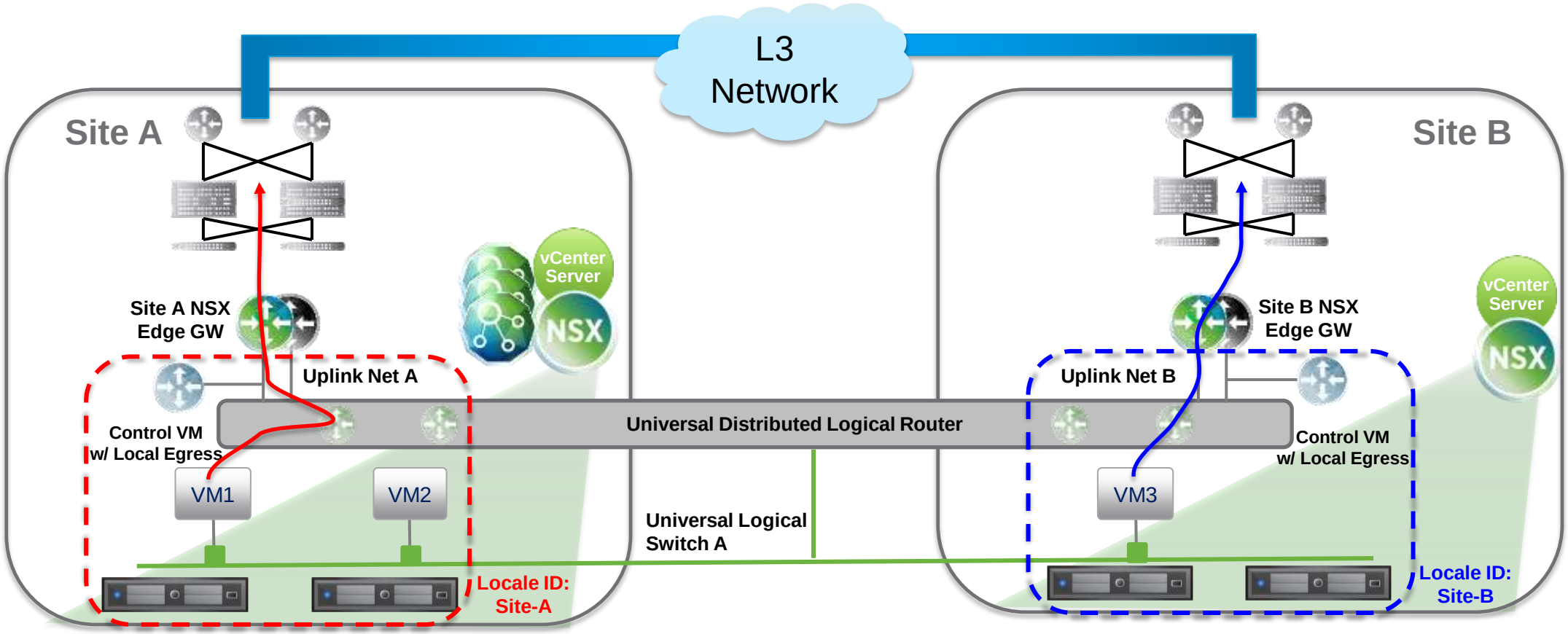


Cross-VC NSX部署场景2: 多个站点多个vCenter A/P N-S



Example UDLR Routing Config (also set matching priorities in Physical Network):
SiteA-Edges BGP Weight 10 or OSPF Cost 1
SiteB-Edges BGP Weight 1 or OSPF Cost 10

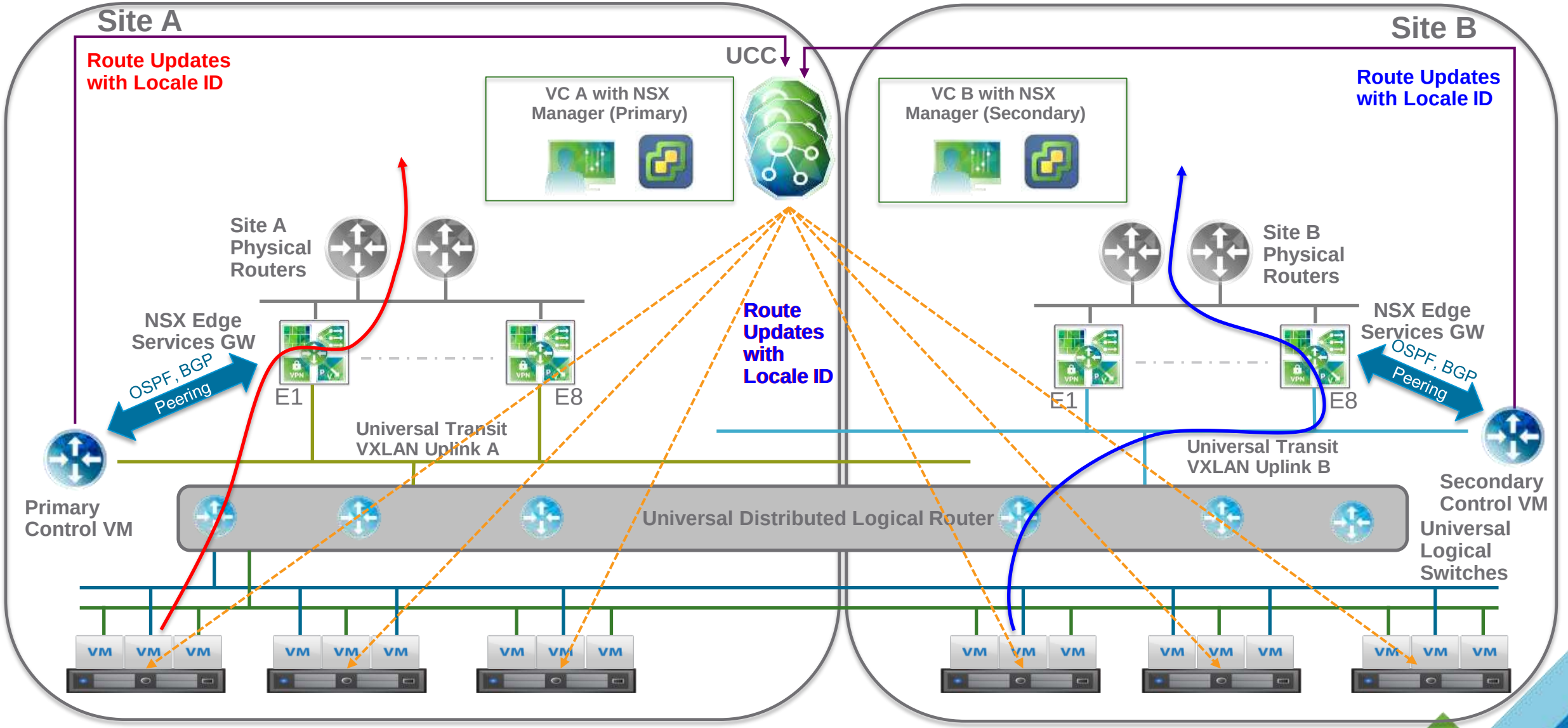
Cross-VC NSX 出向路由优化



Multi-Site Enhancement: Locale ID

- NSX 6.2 引进了一个新的locale id的概念，当路由发送给controller时会带上locale id。缺省情况locale id等于NSX manager的UUID。
- 如果UDLR没有启用local egress的功能，Locale ID的值被忽略。
- 当启用了Local Egress功能, NSX Controller只会将路由发送给匹配Locale ID的esxi host。
- 每个站点可以有自己独立的路由配置
- Locale ID 支持按照UDLR, Cluster 以及Host来设置，如果跨site采用的是同一个NSX manager。

Cross-VC NSX– 部署场景3：多个站点多个vCenter A/A N-S



总结

- NSX部署与物理网络解耦，适合于任何物理网络架构
- NSX部署和后期运维简单化，只需增加LS、DLR
- NSX提供任意虚机之间的安全隔离和多租户隔离的能力
- NSX Edge/DLR支持OSPF、BGP路由
- NSX Edge支持Active/Standby和ECMP模型
- Cross-VC NSX支持双活/灾备数据中心场景

谢谢！

vmware®