云计算技术与应用

第05章：vSphere Network and Storage

https://aitcm.hactcm.edu.cn

河南中医药大学信息技术学院（智能医疗行业学院）智能医疗教研室
河南中医药大学医疗健康信息工程技术研究所

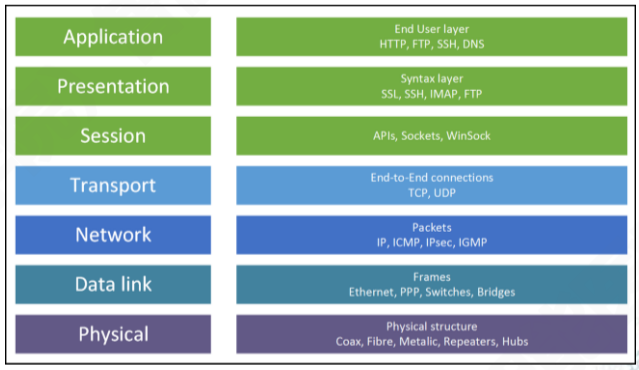
2025年3月

讨论提纲

- vSphere Network
  - 网络基本概念：OSI Model、Encapsulation、MAC、MTU、VLAN、IPv6
  - vSphere Network 基本概念：vSwitch、vDS、VMkernel NICs、Virtualize Network
- 使用 vSphere Standard Switch
- 使用 vSphere Distributed vSwitch
- vSphere Storage
  - 存储基础知识：企业级存储、本地存储和共享存储、存储阵列
  - 存储性能
- vSphere Storage 基本应用
- vSphere Storage 高级功能：SIOC 和 storage DRS、数据存储集群、VMFS 6
- vSAN：
  - 规划与设计 vSAN 方案
  - vSAN 应用

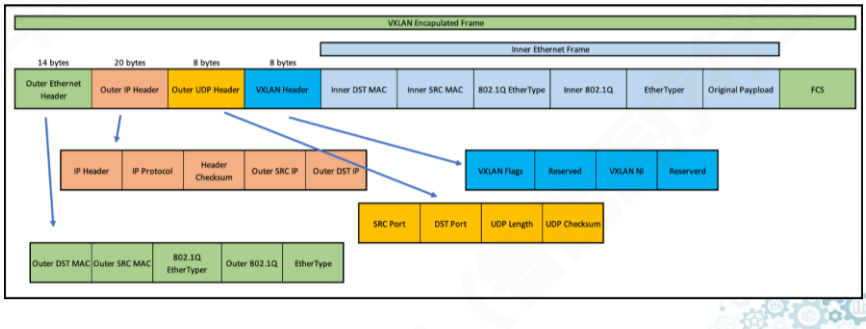
1. vSphere Network

- OSI model
  - OSI 模型是概念模型，描述网络中数据如何从一个设备流向另一个设备。



1. vSphere Network

- 封装与解封装：Encapsulation and de-encapsulation
  - 网络传输的信息必须在通信的发送端和接收端进行转换，转换过程是封装和解封装。
  - 以软件定义网络中使用 VXLAN 为例：



1. vSphere Network

- MAC转发表与自学习协议：MAC tables and MAC learning process
  - ESXi 的 VMkernel 接口有一个本地 ARP 表。
  - 查看 ESXi 的 ARP 表的命令：
    - esxcli network ip neighbor list

172.16.125.5 - PuTTY

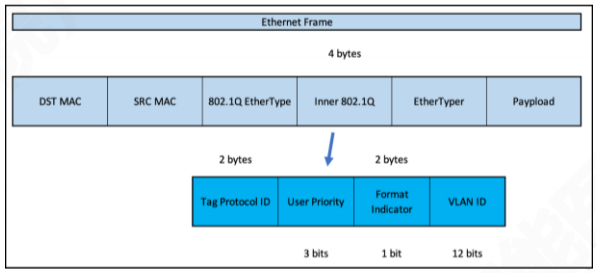
```
login as: root
Keyboard-interactive authentication prompts from server:
Password:
End of keyboard-interactive prompts from server
The time and date of this login have been sent to the system logs.

WARNING:
   All commands run on the ESXi shell are logged and may be included in
   support bundles. Do not provide passwords directly on the command line.
   Most tools can prompt for secrets or accept them from standard input.

VMware offers supported, powerful system administration tools.  Please
see www.vmware.com/go/sysadmintools for details.

The ESXi Shell can be disabled by an administrative user. See the
vSphere Security documentation for more information.
[root@C-202118122-ESXi-1:~] esxcli network ip neighbor list
Neighbor        Mac Address        Vmknic    Expiry  State  Type
--------------  -----------------  ------  --------  -----  -------
172.16.125.40   14:a0:f8:df:79:66  vmk0      96 sec         Unknown
172.16.125.8    00:50:56:a6:1c:e9  vmk0     252 sec         Unknown
172.16.125.15   00:50:56:a6:47:8d  vmk0    1154 sec         Unknown
172.16.125.30   c4:ff:1f:ac:6f:96  vmk0     384 sec         Unknown
[root@C-202118122-ESXi-1:~]
```


1. vSphere Network

- Maximum Transmission Unit (MTU)
  - 从以太网规范至今，已有多个 IEEE 标准支持其他扩展帧类型。
    - VLAN tagging (802.1Q): additional 4 bytes in the Ethernet header.
    - Provider Bridge (PB) 802.1ad: additional 8 byte.
    - FCoE frames: MTU of 2,500 bytes.
    - Multiprotocol Label Switching (MPLS):
      - This increases the maximum Ethernet frame size to 1,518 bytes + (n * 4 bytes).
    - VXLAN: adds another 50 bytes.
    - Jumbo frames:
      - These are Ethernet frames with more than 1,500 bytes of payload, typically around 9,000 bytes.
      - They are mostly used for IP-based storage traffic.
      - Should use jumbo frames (MTU 9,000) for iSCSI or NFS traffic.

1. vSphere Network

- Virtual LAN (VLAN)
  - VLAN 是一个广播域，可以使用 VLAN 对以太网广播域进行分段。
  - 网络端口配置属于一个或多个 VLAN。
  - 802.1Q 中继修改以太网帧结构，添加数字标记实现将帧转发到不同的 VLAN。



1. vSphere Network

- IPv6
  - 从 vSphere 4.1 开始支持 IPv4 和 IPv6 均支持，但 IPv6 默认禁用。
  - 从 vSphere 5.1 开始，默认情况下为 VMkernel 流量启用 IPv6。
  - 由于 Linux 和 Windows 操作系统的要求，建议不要禁用 ESXi 的 IPv6。

1. vSphere Network

- Virtual Switches 虚拟交换机
- Port Groups 端口组
- VMkernel Ports ESXi 网卡
- Virtualized NICs 虚拟网卡

1. vSphere Network

- Virtual Switch：虚拟交换机
  - vSphere 有两种类型的虚拟交换机：
    - vSphere Standard Switch（vSS）：vSphere 标准交换机
      - 运行方式与物理以太网交换机十分相似，类似于将物理交换机连接在一起以创建较大型的网络。
      - 检测与其虚拟端口进行逻辑连接的虚拟机，并使用该信息向正确的虚拟机转发流量。可使用物理以太网适配器（也称为上行链路适配器）将虚拟网络连接至物理网络，以将 vSphere 标准交换机连接到物理交换机。
      - 即使 vSphere 标准交换机的运行方式与物理交换机相似，但不具备物理交换机的一些高级功能。
    - vSphere Distributed Switch（vDS）：vSphere 分布式交换机
      - 可充当数据中心中所有关联主机的单一交换机，提供虚拟网络的集中式配置、管理以及监控。
      - 可以在 vCenter Server 系统上配置 vSphere Distributed Switch，配置将同步到与该交换机关联的所有 ESXi 主机，使虚拟机可在跨多个主机进行迁移时确保其网络配置保持一致。

对比 vSphere standard and distributed vSwitch

Feature	Standard vSwitch	Distributed vSwitch
L2 forwarding	Yes	Yes
VLAN support	Yes	Yes
NIC teaming	Yes	Yes
Outbound traffic shaping	Yes	Yes
Inbound traffic shaping	No	Yes
Centralized management	No	Yes
PVLAN support	No	Yes
Netflow export support	No	Yes
Port mirroring	No	Yes
Multicast support	No	Yes
Traffic filtering	No	Yes
Network IO control	No	Yes

流量控制

集中管理

1. vSphere Network

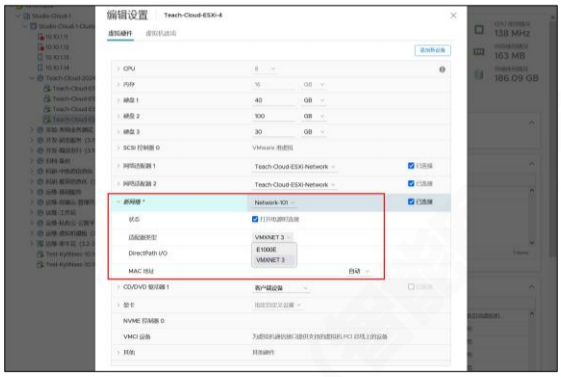
- Port Groups：端口组
  - 标准端口组
    - 网络服务通过端口组连接到标准交换机，端口组定义通过交换机连接网络的方式。
    - 单个标准交换机与一个或多个端口组关联。
    - 端口组为每个端口指定了诸如宽带限制和 VLAN 标记策略之类的端口配置选项。
  - 分布式端口
    - 连接到主机 VMkernel 或虚拟机的 vSphere Distributed Switch 上的一个端口。
  - 分布式端口组
    - 与 vSphere Distributed Switch 关联的一个端口组。
    - 分布式端口组为每个成员端口指定端口配置选项，并定义通过 vSphere Distributed Switch 连接到网络的方式。

1. vSphere Network

- VMkernel Ports：VMkernel 适配器/端口（ESXi 网卡）
  - VMkernel 适配器向主机提供网络连接并接受 vMotion、IP 存储、Fault Tolerance 日志记录、vSAN 等服务的系统流量。

1. vSphere Network

- Virtualized NICs：虚拟网卡



1. vSphere Network

- Virtualized NICs：虚拟网卡
  - E1000E
    - Intel 82574 千兆位以太网网卡的模拟版本。
    - E1000E 是 Windows 8 和 Windows Server 2012 的默认适配器。
  - E1000
    - Intel 82545EM 千兆位以太网网卡的模拟版本。
    - 其驱动程序在大多数较新的客户机操作系统中都可用，包括 Windows XP 及更高版本和 Linux 2.4.19 版及更高版本。
  - Vlance
    - AMD 79C970 PCnet32 LANCE 网卡的模拟版本，是一种较旧的 10 Mbps 网卡。
    - 其驱动程序在 32 位旧版客户机操作系统中可用。
    - 配置该网络适配器的虚拟机可立即使用其网络。

1. vSphere Network

- Virtualized NICs：虚拟网卡
  - VMXNET
    - 为在虚拟机中发挥更大的性能而进行了优化，没有物理设备为其对应。
    - 操作系统供应商没有为此卡提供内置驱动程序，必须安装 VMware Tools 以便为 VMXNET 网络适配器提供可用的驱动程序。
  - VMXNET 2 (增强型)
    - 基于 VMXNET 适配器，提供网络更高性能的功能，例如巨帧和硬件卸载。
    - VMXNET 2（增强型）只能在 ESX/ ESXi 3.5 及更高版本上可用。
  - VMXNET 3
    - 为高性能打造的准虚拟化网卡。
    - VMXNET 3 提供 VMXNET 2 中具备的所有可用功能，并且还另外添加了几项新功能，例如多队列支持（在 Windows 中也称为接收方缩放）、IPv6 卸载和 MSI/MSI-X 中断交付。
    - VMXNET 3 与 VMXNET 或 VMXNET 2 不相关。

1. vSphere Network

- Virtualized NICs：虚拟网卡
  - PVRDMA
    - 支持通过 OFED verbs API 在虚拟机之间进行远程直接内存访问 (RDMA) 的准虚拟化网卡。
    - 所有虚拟机都必须具有 PVRDMA 设备，并且应该连接到分布式交换机。
    - PVRDMA 支持 VMware vSphere vMotion 和快照技术。
    - 硬件版本为 13 且客户机操作系统为 Linux 内核 4.6 及更高版本的虚拟机中提供该设备。
  - SR-IOV 直通
    - 具有 SR-IOV 支持的物理网卡上的虚拟功能 (VF) 表示形式。
    - 虚拟机与物理适配器交换数据，而不使用 VMkernel 作为中介。
    - 此适配器类型适合延迟可能导致故障或需要更多 CPU 资源的虚拟机。

图 10-1. vSphere SR-IOV 支持中的数据路径和配置路径

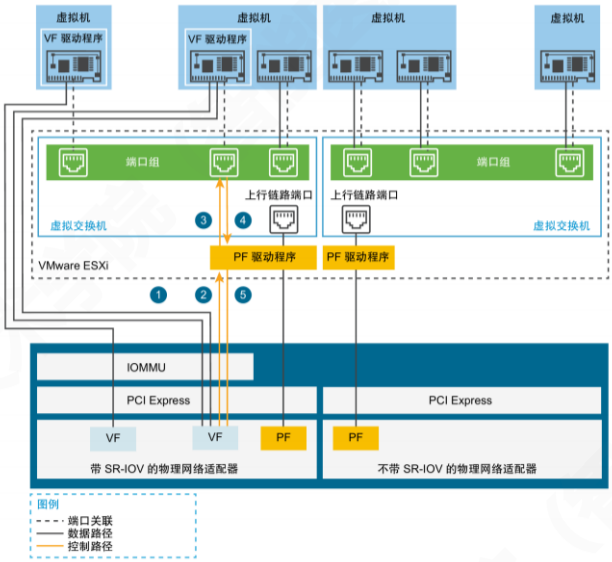
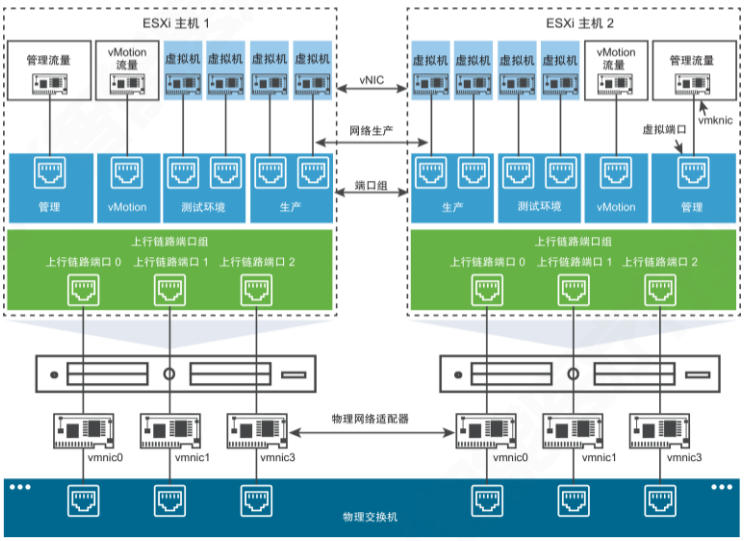


表 10-1. 要使用 SR-IOV 所需支持的配置

组件	要求
物理主机	■ 必须与 ESXi 版本兼容。 ■ 必须配备 Intel 或 AMD 处理器。 ■ 必须支持 I/O 内存管理单元 (IOMMU)，并且必须在 BIOS 中启用 IOMMU。 ■ 必须支持 SR-IOV，并且必须在 BIOS 中启用 SR-IOV。请联系服务器供应商以确定主机是否支持 SR-IOV。
物理网卡	■ 必须与 ESXi 版本兼容。 ■ 根据服务器供应商提供的技术文档，必须支持用于主机和 SR-IOV。 ■ 必须在固件中启用 SR-IOV。 ■ 必须使用 MSI-X 中断。
对于物理网卡，在 ESXi 中使用 PF 驱动程序	■ 必须经过 VMware 认证。 ■ 必须安装在 ESXi 主机上。对于某些网卡，ESXi 版本提供默认驱动程序，而对于其他版本，必须手动开启安装驱动程序。
客户机操作系统	根据网卡供应商提供的技术文档，必须受已安装的 ESXi 版本上的网卡支持。
客户机操作系统中使用 VF 驱动程序	■ 必须产品网卡兼容。 ■ 根据网卡供应商提供的技术文档，必须安装产机操作系统版本的支持。 ■ 必须由 Microsoft WLK 或 WHCK 针对 Windows 虚拟机进行认证。 ■ 必须安装在操作系统中。对于某些网卡，操作系统版本中包含默认驱动程序，而对于其他网卡，则必须从网卡供应商或主机供应商提供的位置下载并安装驱动程序。

2. 使用 vSphere Standard Switch

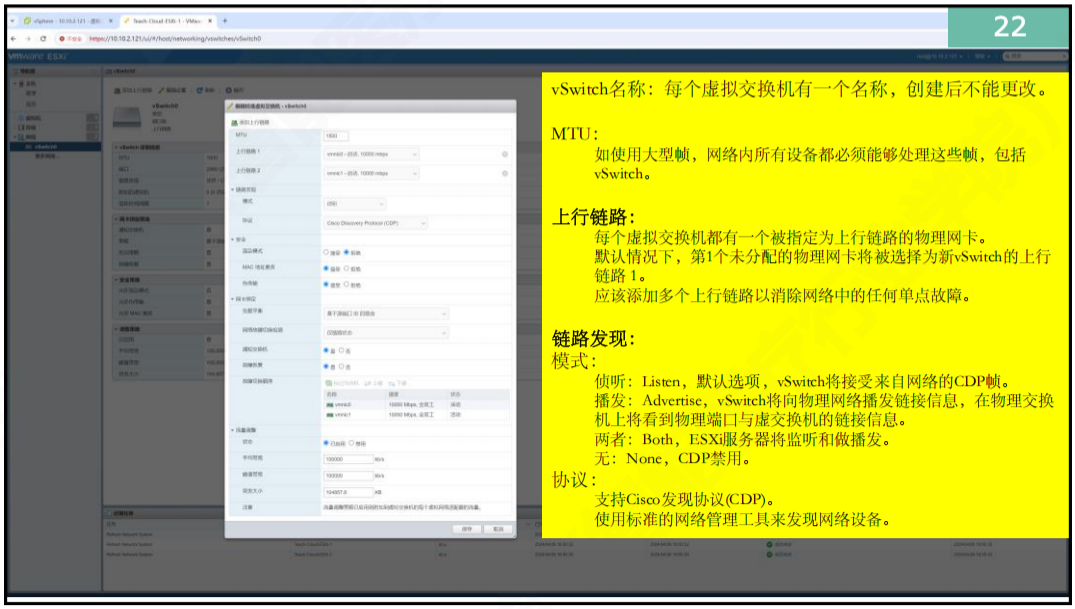
- vSphere Standard Switch：vSphere 标准交换机
  - 由 VMkernel 构建并在 VMkernel 中运行。
  - 不提供独立管理功能。
    - 无法使用 Telnet、SSH 进行管理。
    - 只能够使用 vSphere 管理。
    - 工作在 Layer 2 （数据链路层）
    - 支持 VLAN
  - 必须有上行链路和端口组
    - 没有上行链路无法与上游网络通信。
    - 没有端口组无法为 VMkernel 或 VM 提供连接。
  - 端口数将按比例自动增加和减少



vSphere Standard Switch（vSS）：vSphere 标准交换机

2. 使用 vSphere Standard Switch

- vSphere Standard Switch 的应用
  - 创建 vSphere 标准交换机
  - 虚拟机的端口组配置
    - 添加虚拟机端口组
    - 编辑标准交换机端口组
    - 从 vSphere 标准交换机移除端口组
  - vSphere 标准交换机属性
    - 更改 vSphere 标准交换机上 MTU 的大小
    - 更改物理适配器的速度
    - 添加物理适配器并使适配器成组
    - 查看 vSphere 标准交换机的拓扑图

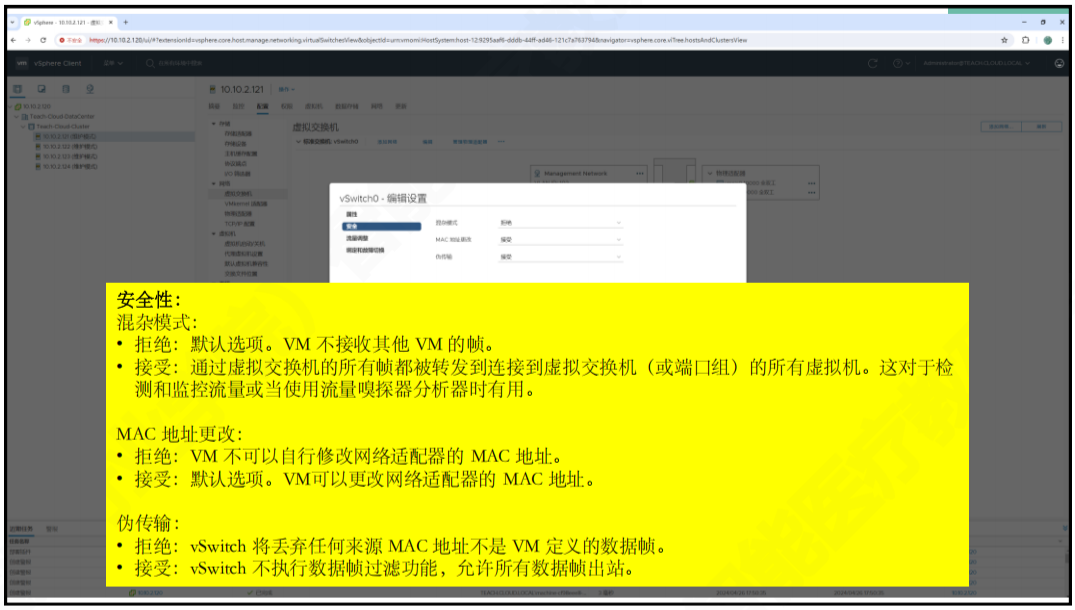


vSwitch名称：每个虚拟交换机有一个名称，创建后不能更改。

MTU：
如使用大型帧，网络内所有设备都必须能够处理这些帧，包括vSwitch。

上行链路：
每个虚拟交换机都有一个被指定为上行链路的物理网卡。
默认情况下，第1个未分配的物理网卡将被选择为新vSwitch的上行链路 1。
应该添加多个上行链路以消除网络中的任何单点故障。

链路发现：
模式：
侦听：Listen，默认选项，vSwitch将接受来自网络的CDP帧。
播发：Advertise，vSwitch将向物理网络播发链接信息，在物理交换机上将看到物理端口与虚交换机的链接信息。
两者：Both，ESXi服务器将监听和做播发。
无：None，CDP禁用。
协议：
支持Cisco发现协议(CDP)。
使用标准的网络管理工具来发现网络设备。



安全性：

混杂模式：

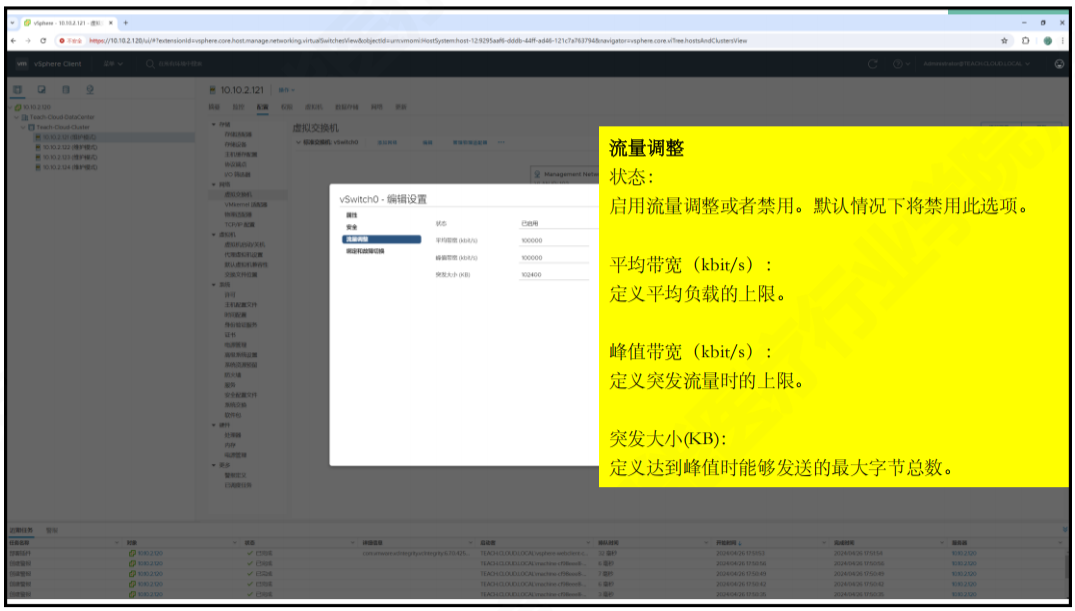
- 拒绝：默认选项。VM 不接收其他 VM 的帧。
- 接受：通过虚拟交换机的所有帧都被转发到连接到虚拟交换机（或端口组）的所有虚拟机。这对于检测和监控流量或当使用流量嗅探器分析器时有用。

MAC 地址更改：

- 拒绝：VM 不可以自行修改网络适配器的 MAC 地址。
- 接受：默认选项。VM可以更改网络适配器的 MAC 地址。

伪传输：

- 拒绝：vSwitch 将丢弃任何来源 MAC 地址不是 VM 定义的数据帧。
- 接受：vSwitch 不执行数据帧过滤功能，允许所有数据帧出站。



流量调整

状态：

启用流量调整或者禁用。默认情况下将禁用此选项。

平均带宽（kbit/s）：

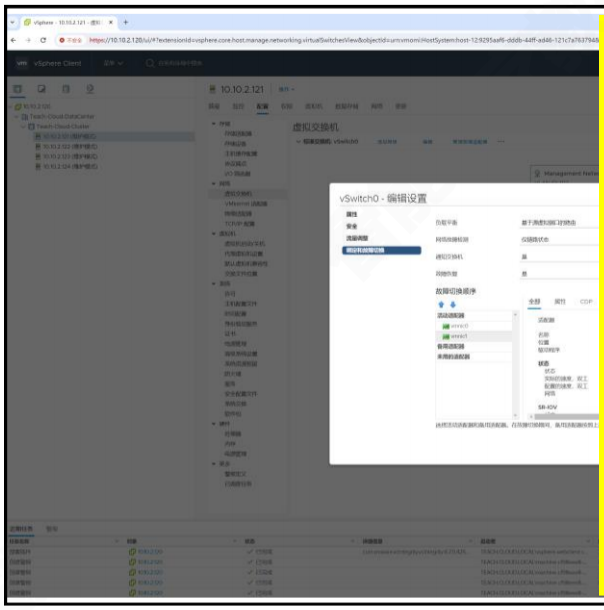
定义平均负载的上限。

峰值带宽（kbit/s）：

定义突发流量时的上限。

突发大小(KB)：

定义达到峰值时能够发送的最大字节总数。

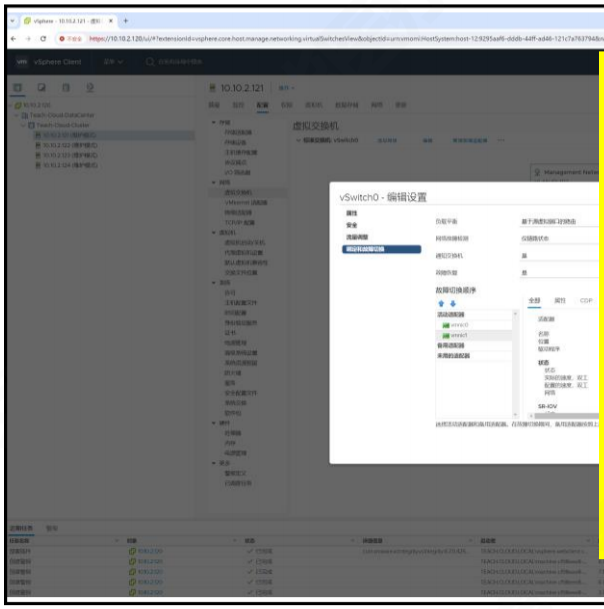


负载平衡：
决定网络通信如何在网络适配器之间分配，根据四种算法之一：
（1）基于源端口ID的路由：默认选项。
vSwitch基于VSS或VDS上VM端口ID选择上行链路。
该方法需在物理交换机进行配置，并且计算开销较低。

（2）基于源MAC哈希的路由：
vSwitch基于VM MAC地址选择一个上行链路。
vSwitch使用虚拟机MAC地址和上行链路数量计算VM上行链路。
该方法支持所有的物理交换机，而且开销较低。

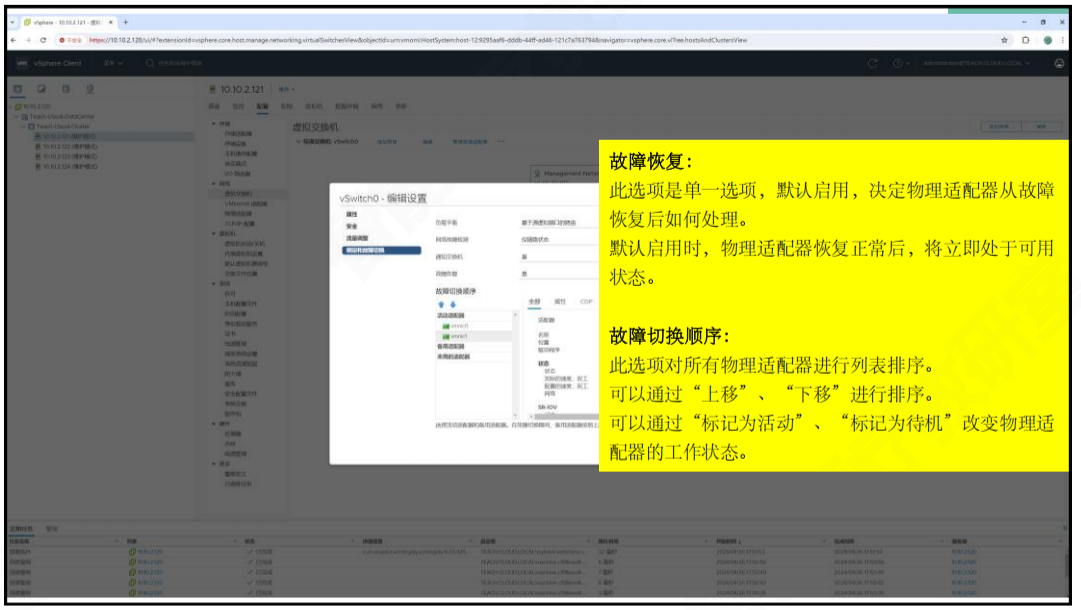
（3）基于IP哈希的路由：
vSwitch基于VMs数据包源IP地址和目的IP地址为每个数据报文选择上行链路。
该方法需要上行链路支持并开启802.3ad。

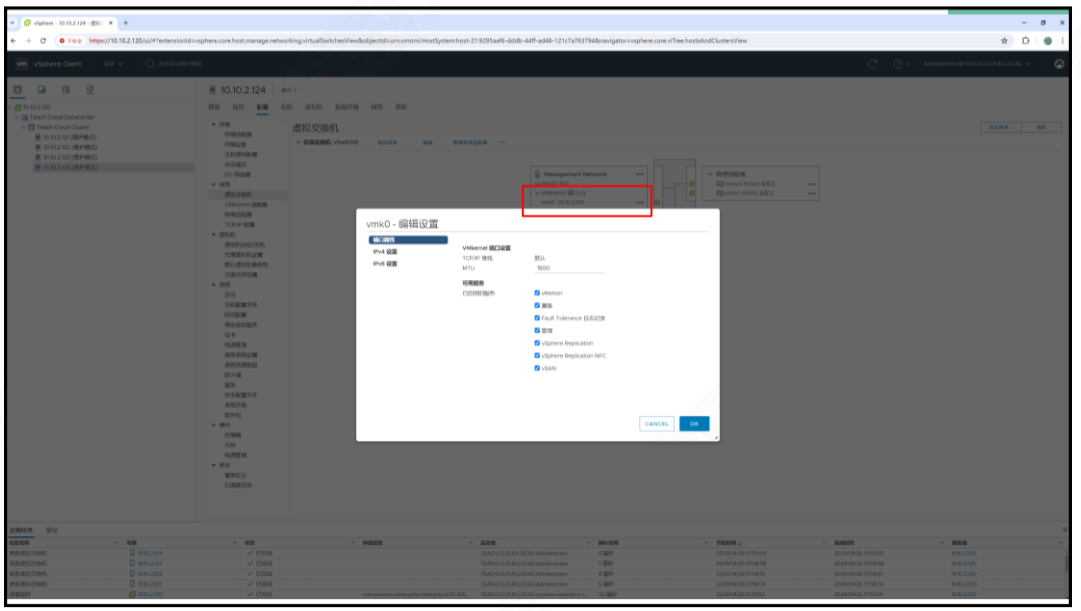
（4）使用明确故障切换顺序：
此策略没有实现负载平衡。
vSwitch使用适配器列表中的第一个可用的上行链路。
如果第一个上行链路故障，将按照列表顺序使用下一个可用的上行链路。



网络故障检测：
此选项定义如何判断一个上行链接为不可用。
判断链路不可用的方法有两种：
（1）仅链路状态：默认选项。
仅检测链路故障，例如拔掉电缆或物理连通性故障。
（2）仅信标：
使用此检测模式须至少有3个网卡来进行信标探测。
发送并侦听物理NIC发送的以太网广播帧，检测所有物理NIC中的链路故障。
ESXi主机每秒钟发送一次信标数据包。

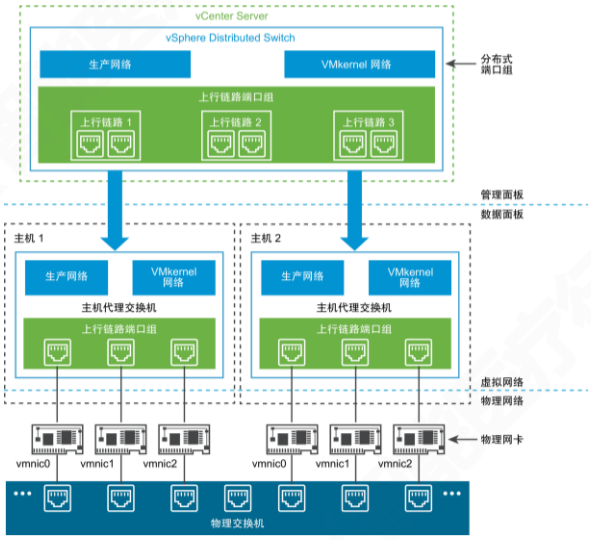
通知交换机：
此选项是单一选项，默认启用，用于加快物理交换机能够快速改变网络拓扑。
当vSwitch链接的物理交换机端口发生变化时，vSwitch通过网络发送通知，以通知物理交换机更新MAC查找表。
使用的协议是反向地址解析协议(RARP)。


vSwitch0 - 编辑设置
故障恢复：
此选项是单一选项，默认启用，决定物理适配器从故障恢复后如何处理。
默认启用时，物理适配器恢复正常后，将立即处于可用状态。
故障切换顺序：
此选项对所有物理适配器进行列表排序。
可以通过“上移”、“下移”进行排序。
可以通过“标记为活动”、“标记为待机”改变物理适配器的工作状态。

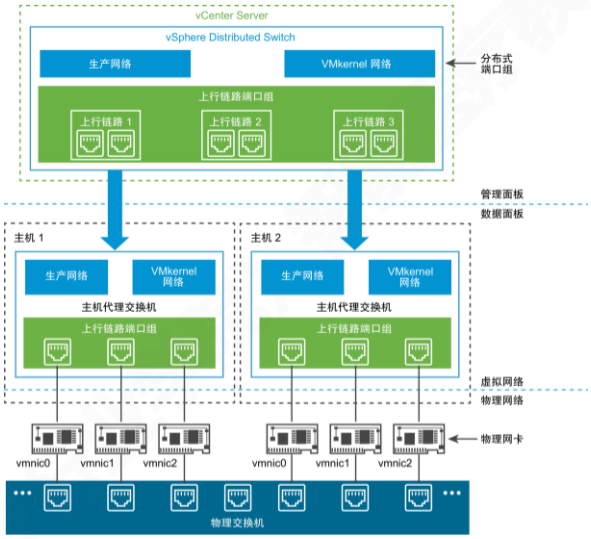

vmk0 - 编辑设置
CANCEL
OK

3. 使用 vSphere Distributed Switch

- vSphere Distributed Switch：vSphere 分布式交换机
  - vDS 为所有主机的网络连接配置提供集中化管理和监控。
  - 通过 vCenter Server 设置，并同步到与该交换机关联的所有主机上。
  - vDS 引入两个抽象概念以实现物理网卡、虚拟机和 VMkernel 网络配置一致。
    - 上行链路端口组：
      - 在创建 Distributed Switch 期间进行定义，可以具有一个或多个上行链路。
      - 可以将主机的物理网卡映射到 Distributed Switch 上的上行链路。
      - 可以对上行链路设置故障切换和负载均衡策略。
    - 分布式端口组：
      - 可向虚拟机提供网络连接并供 VMkernel 流量使用。
      - 可以在分布式端口组上配置网卡绑定、故障切换、负载均衡、VLAN、安全、流量调整和其他策略。



vSphere Distributed Switch（vDS）：vSphere 分布式交换机



vSphere Distributed Switch（vDS）：vSphere 分布式交换机

例如，假设数据中心创建一个 vSphere Distributed Switch，两个主机与其关联。

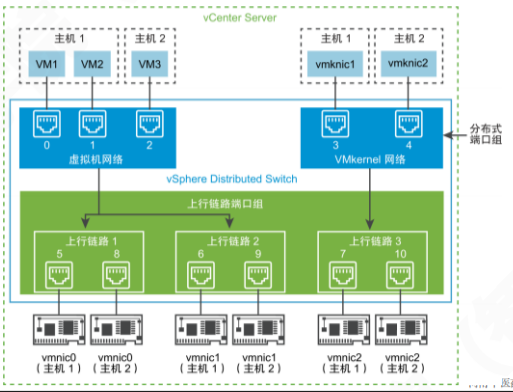
为上行链路端口组配置了三个上行链路，将每个主机的一个物理网卡连接到一个上行链路。每个上行链路都有两个物理网卡，每个物理网卡来自映射到它的每个主机。

例如，上行链路 1 配置了主机 1 和主机 2 中的 vmnic0。为虚拟机网络和 VMkernel 服务创建生产和 VMkernel 网络分布式端口组。分别在主机 1 和主机 2 上创建生产和 VMkernel 网络端口组的表示。

为生产和 VMkernel 网络端口组设置的所有策略都将传播到其在主机 1 和主机 2 上的表示。

3. 使用 vSphere Distributed Switch

- vSphere Distributed Switch 数据流
  - vSphere Distributed Switch 上的网卡成组和端口分配



假设创建分别包含 3 个和 2 个分布式端口的虚拟机网络和 VMkernel 网络分布式端口组。

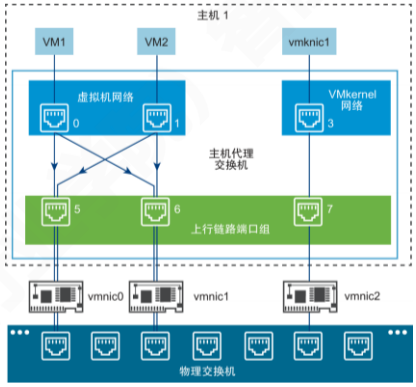
Distributed Switch 会按 ID 从 0 到 4 的顺序分配端口，该顺序与创建分布式端口组的顺序相同。然后将主机 1 和主机 2 与 Distributed Switch 关联。

Distributed Switch 会为主机上的每个物理网卡分配端口，端口将按添加主机的顺序从 5 继续编号。

要在每个主机上提供网络连接，请将 vmnic0 映射到上行链路 1、将 vmnic1 映射到上行链路 2 、将 vmnic2 映射到上行链路 3。

3. 使用 vSphere Distributed Switch

- vSphere Distributed Switch 数据流
  - vSphere Distributed Switch 上的网卡成组和端口分配



在主机端，虚拟机和 VMkernel 服务的数据包流量将通过特定端口传递到物理网络。

例如，从主机 1 上的 VM1 发送的数据包将先到达虚拟机网络分布式端口组上的端口 0。由于上行链路 1 和上行链路 2 处理虚拟机网络端口组的流量，数据包可以通过上行链路端口 5 或上行链路端口 6 继续传递。

如果数据包通过上行链路端口 5，则继续传递 vmnic0；如果数据包通过上行链路端口 6，则继续传递到 vmnic1。

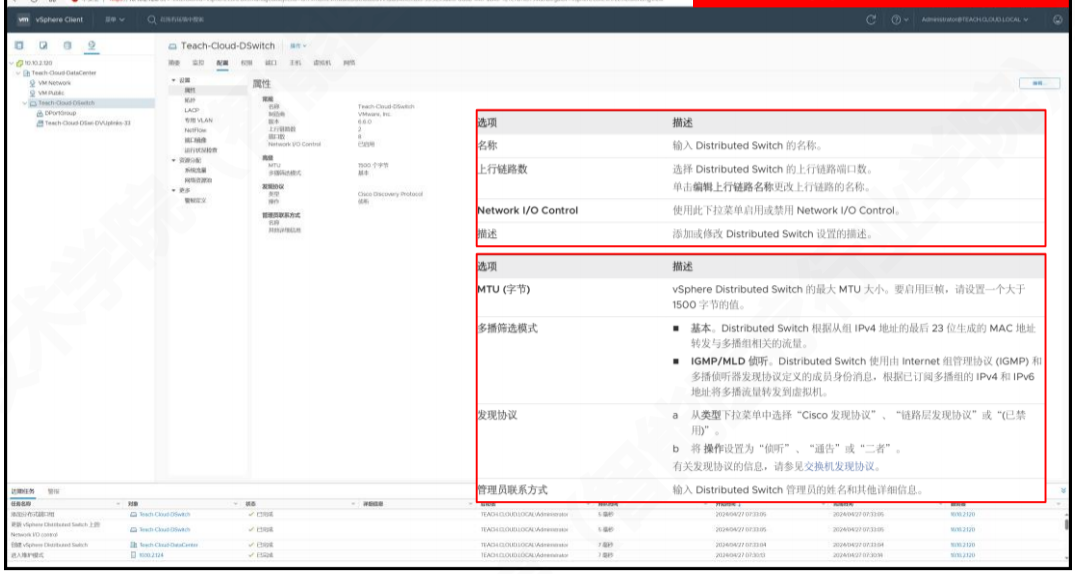
3. 使用 vSphere Distributed Switch

- vSphere Distributed Switch 的应用
  - 创建 vSphere Distributed Switch
    - 添加和移除主机
    - 配置和移除 vDS
    - 网络状态检查和网络回滚
    - 导入和导出 vDS 的配置
  - 使用分布式端口组（Distributed Port Group）
    - 创建、配置、移除、监控
  - 管理 VMkernel Adapters
    - 将 VMkernel 适配器迁移到 vDS
    - 在 vDS 上创建 VMkernel 适配器

3. 使用 vSphere Distributed Switch

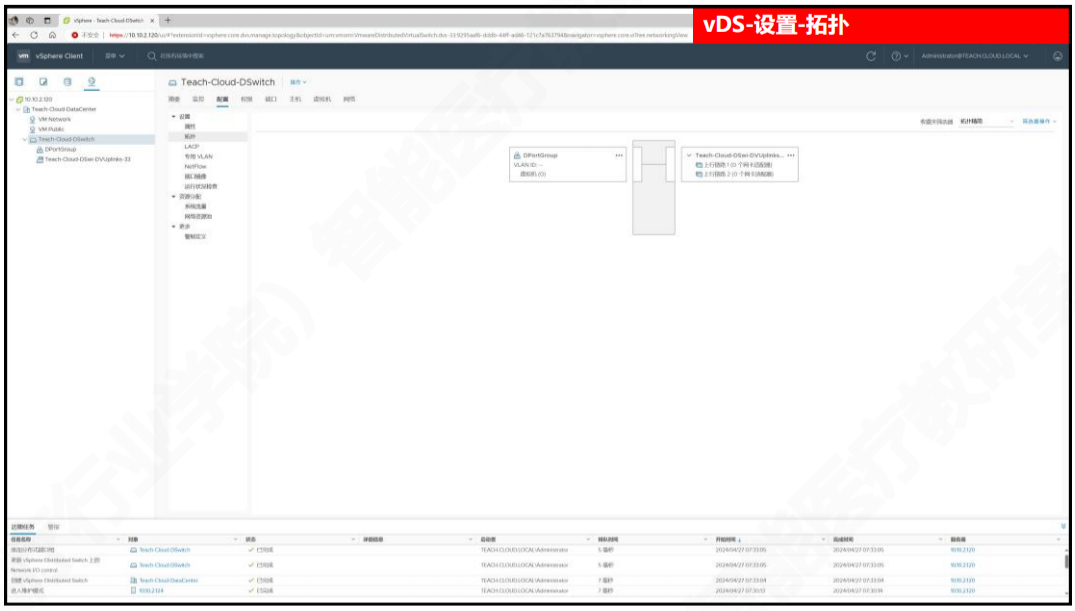
- vSphere Distributed Switch 的应用
  - vDS 网络管理功能
    - 使用 NetFlow
    - 启用 Switch Discovery Protocols（CDP）
    - 启用增强的组播（Multicast）
    - 配置 PVLAN（Private VLAN）
    - 配置 LACP（Link Aggregation Control Protocol）
  - vDS 交换机安全加固
    - 混杂模式（Promiscuous mode）
    - MAC 地址更改（MAC address changes）
    - 伪传输（Forged transmits）

vDS-设置-属性



选项	描述
名称	输入 Distributed Switch 的名称。
上行链路数	选择 Distributed Switch 的上行链路端口数。 单击**编辑上行链路名称**更改上行链路的名称。
Network I/O Control	使用此下拉菜单启用或禁用 Network I/O Control。
描述	添加或修改 Distributed Switch 设置的描述。

选项	描述
MTU (字节)	vSphere Distributed Switch 的最大 MTU 大小。要启用巨帧，请设置一个大于 1500 字节的值。
多播筛选模式	■ **基本**。Distributed Switch 根据从组 IPv4 地址的最后 23 位生成的 MAC 地址转发与多播组相关的流量。 ■ **IGMP/MLD 侦听**。Distributed Switch 使用由 Internet 组管理协议 (IGMP) 和多播侦听器发现协议定义的成员身份消息，根据已订阅多播组的 IPv4 和 IPv6 地址将多播流量转发到虚拟机。
发现协议	a 从**类型**下拉菜单中选择“Cisco 发现协议”、“链路层发现协议”或“(已禁用)”。 b 将**操作**设置为“侦听”、“通告”或“二者”。 有关发现协议的信息，请参见交换机发现协议。
管理员联系方式	输入 Distributed Switch 管理员的姓名和其他详细信息。


vDS-设置-拓扑
vSphere Client
Teach-Cloud-DSwitch


vDS-设置-LACP
模式：
活动：所有 LAG 港口都处于活动协商模式，通过 LACP 数据包与物理交换机上的 LACP 端口通道的协商。
被动：LAG 端口处于被动协商模式。它们响应收到的 LACP 数据包，但不启动 LACP 协商。
负载平衡模式：这必须与物理交换机上的配置匹配。
新建链路聚合组

vDS-设置-PVLAN

类型：

混杂：专用 VLAN 与主 VLAN 通信。

隔离：仅与混杂的 VLAN 通信。

社区：与混杂的 VLAN 以及同一辅助 VLAN 中的端口进行通信。

vDS-设置-NetFlow

进行流量分析的工具：

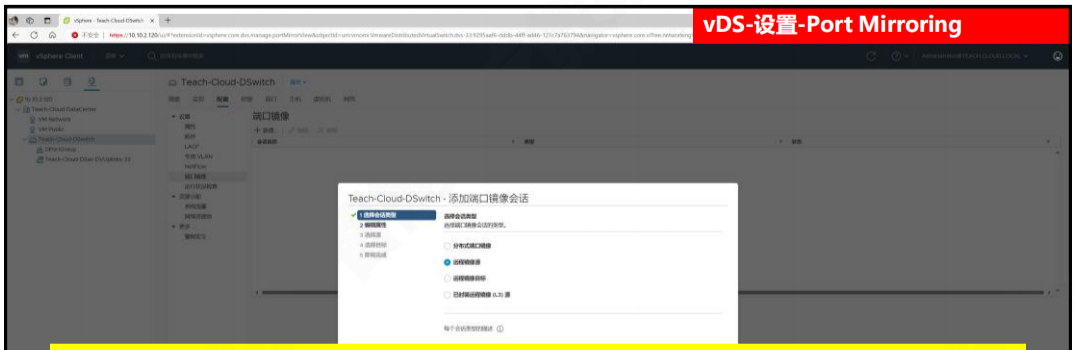
vRealize Network Insight：VMware Aria Operations for Networks

https://www.vmware.com/products/aria-operations-for-networks.html

NfSen：第三方开源

https://nfsen.sourceforge.net

vDS-设置-Port Mirroring

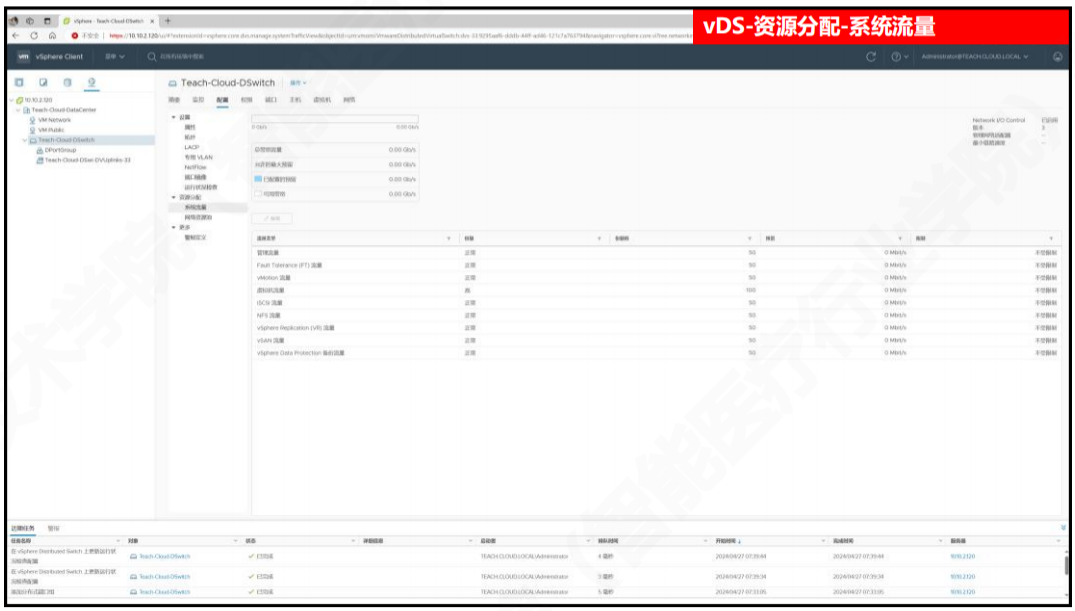


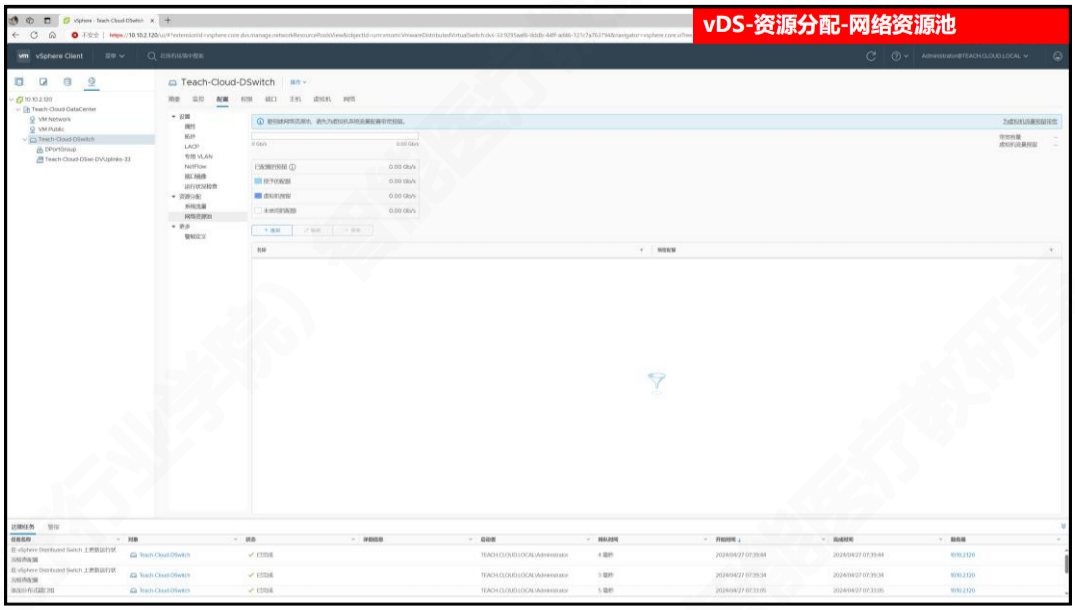
通过端口镜像，可将分布式端口流量镜像到其他分布式端口或特定物理交换机端口。
可在交换机上使用端口镜像将一个交换机端口（或整个 VLAN）上的数据包发送到另一个交换机端口上的监控连接。
端口镜像用于分析和调试数据或诊断网络上的错误，对故障排除非常有用。
若要分析数据流，可以使用 Wireshark （https://www.wireshark.org） 或 tcpdump。

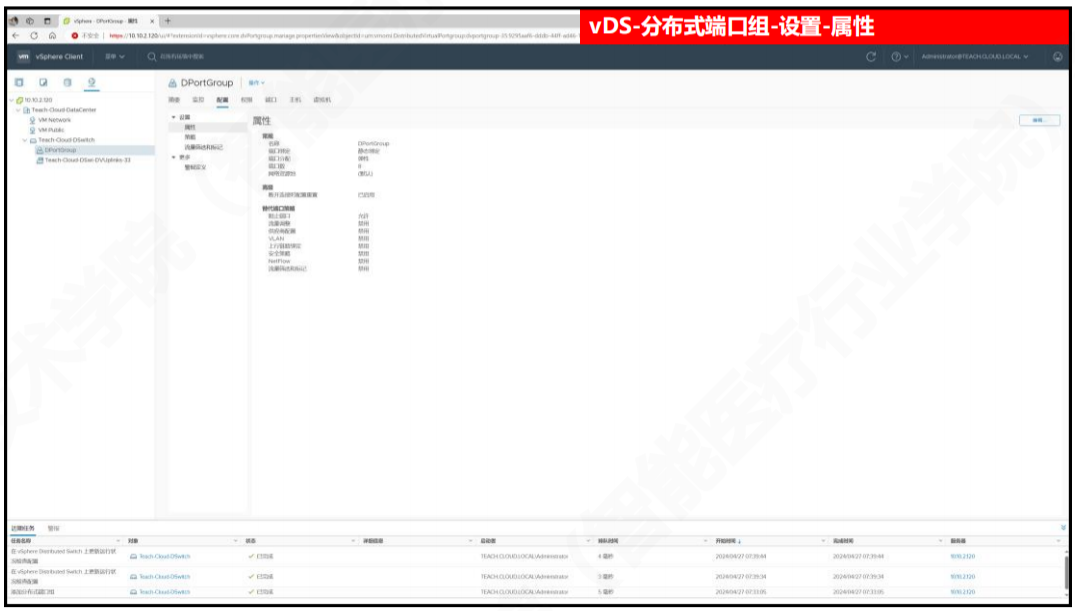
端口镜像常用的方案：

- 分布式端口镜像 （SPAN）
- 远程镜像源和远程镜像目标 （RSPAN）
- 已封装远程镜像（L3）源 （ERSPAN）

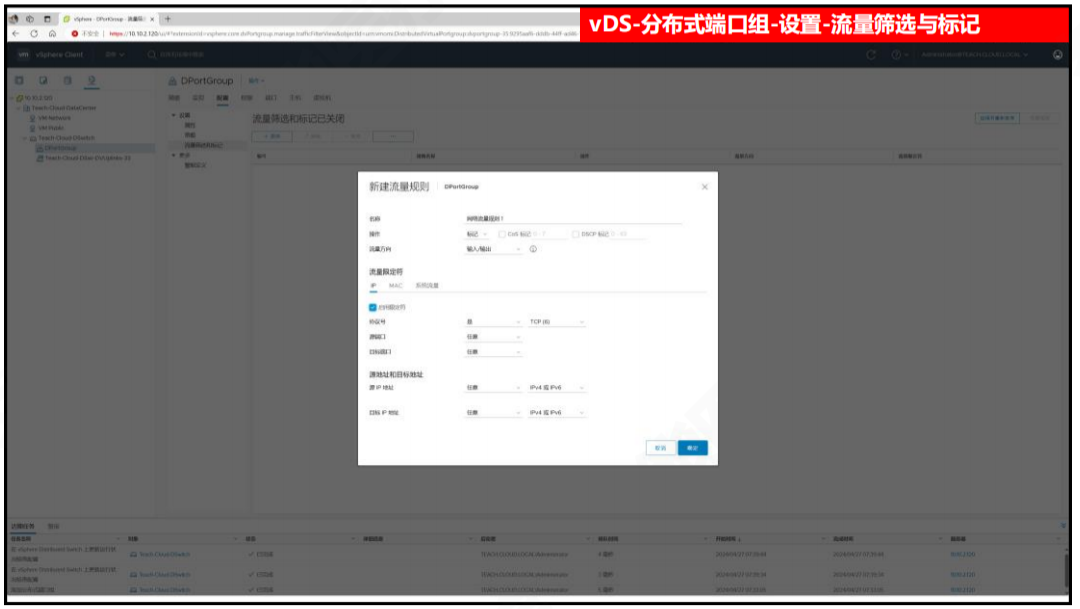
vDS-资源分配-系统流量



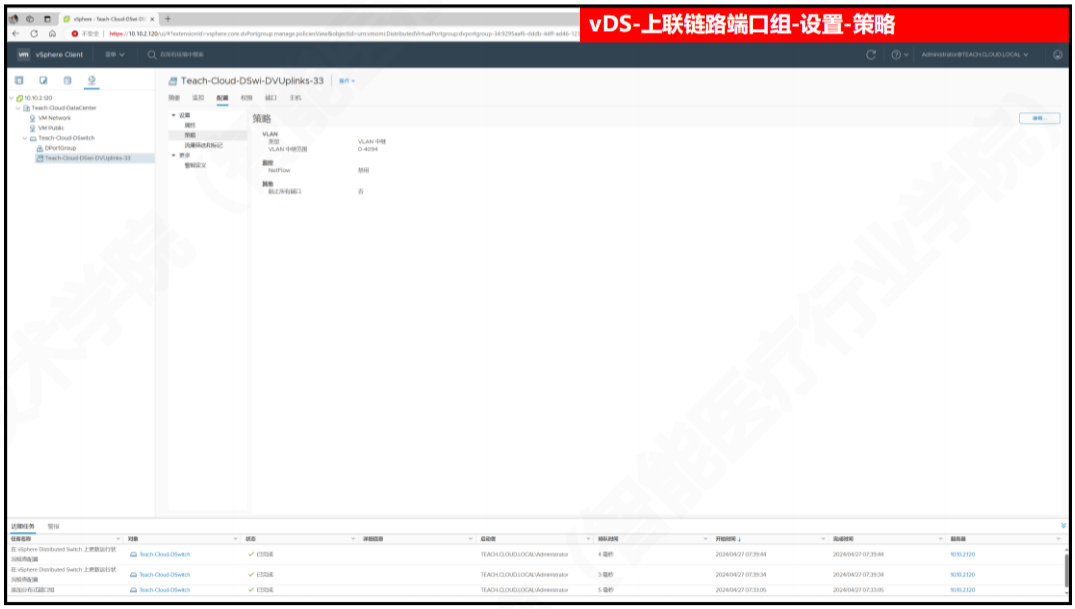

vDS-资源分配-网络资源池
vSphere Client
Teach-Cloud-DSwitch

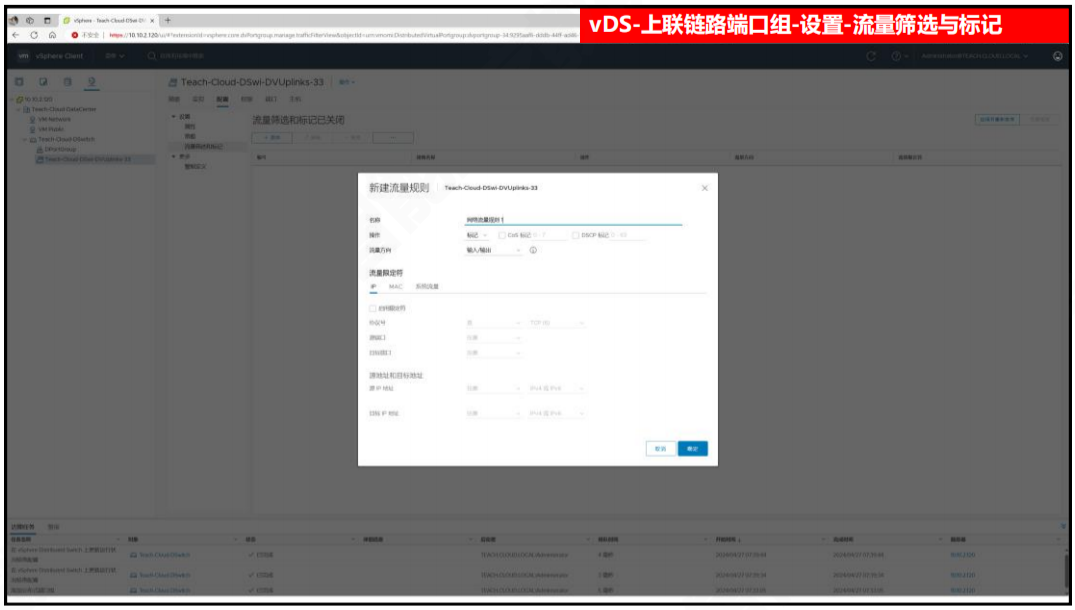

vDS-分布式端口组-设置-属性
vSphere Client
DPortGroup
属性


vDS-分布式端口组-设置-策略
vSphere Client
DPortGroup
策略


vDS-分布式端口组-设置-流量筛选与标记
vSphere Client
DPortGroup
流量筛选和标记已关闭
新建流量规则


vDS-上联链路端口组-设置-属性
vSphere Client
Teach-Cloud-DSwi-DVUplinks-33
属性


vDS-上联链路端口组-设置-策略
vSphere Client
Teach-Cloud-DSwi-DVUplinks-33
策略



VMware官方建议

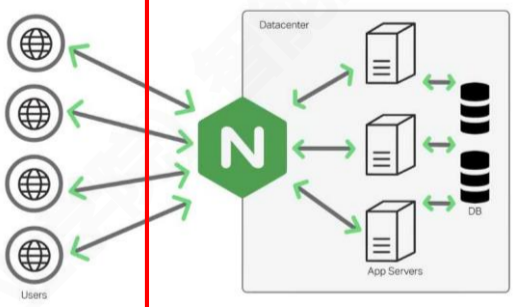
网络最佳做法

在配置网络时，请考虑这些最佳做法。

- 为了确保 vCenter Server 与 ESXi 以及其他产品和服务之间连接稳定，请不要在产品之间设置连接限制和超时。设置限制和超时可影响数据包流量，导致服务中断。
- 与用于主机管理、vSphere vMotion、vSphere FT 等的网络相互隔离，从而提高安全性和性能。
- 将单独的物理网卡专用于一组虚拟机，或使用 Network I/O Control 和流量调整以确保虚拟机的带宽。这种分离方法还可以使总网络工作负载的一部分分布到多个 CPU 上。然后，隔离的虚拟机可以从 vSphere Client 等工具更好地处理应用程序流量。
- 要以物理方式分离网络服务并且专门将一组特定的网卡用于特定的网络服务，请为每种服务创建 vSphere 标准交换机或 vSphere Distributed Switch。如果此操作无法实现，可以通过将网络服务附加到具有不同 VLAN ID 的端口组，以便在一个交换机上将它们分离开。在这两种情况下，都与网络管理员确认所选的网络或 VLAN 是否与环境中的其他部分隔离，即没有与其相连的路由器。
- 在单独的网络上保持 vSphere vMotion 连接。在进行 vMotion 迁移时，客户机操作系统内存的内容将通过该网络传输。通过使用 VLAN 对单个物理网络分段，或者使用单独的物理网络（后者为首选），可以实现这一点。

  为进行跨 IP 子网的迁移和使用单独的缓冲区和插槽池，请将 vMotion 的流量放置在 vMotion TCP/IP 堆栈上，将已关闭电源虚拟机和克隆的迁移的流量放置在置备 TCP/IP 堆栈上。请参见 VMkernel 网络层。
- 可以在不影响虚拟机或在交换机后端运行的网络服务的前提下，向标准或 Distributed Switch 添加或从中移除网络适配器。如果移除所有正在运行的硬件，虚拟机仍可互相通信。如果保留一个网络适配器原封不动，则所有的虚拟机仍然可以与物理网络相连。
- 为了保护大部分敏感的虚拟机，请在虚拟机中部署防火墙，以便在带有上行链路（连接物理网络）的虚拟网络和无上行链路的纯虚拟网络之间路由。
- 为获得最佳性能，请使用 VMXNET 3 虚拟机网卡。
- 连接到同一 vSphere 标准交换机或 vSphere Distributed Switch 的物理网络适配器还应该连接到同一物理网络。
- 在 vSphere Distributed Switch 中配置所有 VMkernel 网络适配器的相同 MTU。如果多个 VMkernel 网络适配器连接到 vSphere Distributed Switch 但配置了不同的 MTU，您可能会遇到网络连接问题。

案例 1：Nginx Proxy 发布多网站服务

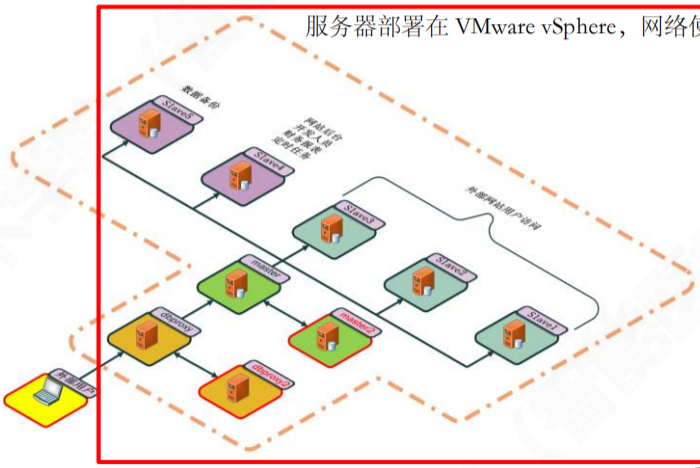


网络设计方案

服务器部署在 VMware vSphere，网络使用 vSphere Standard Switch

案例 2：MySQL/MariaDB Cluster

服务器部署在 VMware vSphere，网络使用 vSphere Standard Switch



网络设计方案

4. vSphere Storage

- 不同类型的存储具有不同协议、体系结构、扩展功能和用途。
- 虚拟化环境下能够满足性能、可弹性扩展、可靠的存储解决方案，需要使用企业级存储产品实现。
- 企业级存储（Enterprise-class storage）的分类：
  - Direct-Attached Storage (DAS)
  - Network Attached Storage (NAS)
  - Storage Area Network (SAN)

  可以运行VMs

  - Object-based storage/cloud storage

  不可以运行VMs，可用于备份等

4. vSphere Storage

- 从架构角度：企业存储有三种扩展解决方案
  - 规模内扩展或规模上扩展：Scale-in or scale-up
    - 通过添加新的磁盘架来增加存储容量，但性能没有改变。
    - 适用于中小规模存储需求，在不引入新的存储系统的情况下增加容量和性能。
  - 规模外扩展：Scale-out
    - 将更多的存储节点添加到存储集群中，形成一个更大的存储池。
    - 通过添加新的存储阵列或节点来扩展存储容量和性能，使存储系统具有更好的横向扩展性。
    - 通常与分布式存储系统或软件定义存储（SDS）解决方案结合使用。
  - 混合扩展：
    - 将规模内扩展和规模外扩展结合起来，根据实际需求采取不同的扩展策略。
    - 在已有的存储系统上进行硬件升级，在需要时添加新存储节点以实现更大规模的扩展。

55

4. vSphere Storage

4.1 企业级存储

- 从性能角度：企业级存储通常分为4个等级
  - 第 0 层：
    - 非常高性能的存储，例如全闪存阵列 （AFA），提供无与伦比的 I/O 性能。
    - 随着企业级 SSD 价格的下降，第 0 层存储越来越常见。
  - 第 1 层或主存储：
    - 通常是主存储，应用于 vShpere 中。
  - 第 2 层或辅助存储：
    - 通常不应用于生产环境中，主要存储文档、备份、冷数据等。
  - 第 3 层：
    - 长期和离线归档的存储，例如磁带或公有云存储上的备份副本。

56

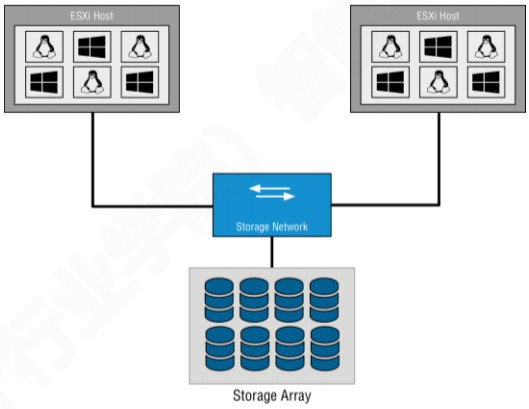
4. vSphere Storage

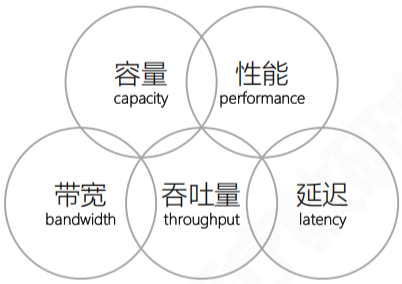
4.2 共享存储

- 存储是虚拟基础设施中最关键的部分。
  - 存储设计方案对于数据中心非常重要。
  - 存储为整个集群及其所有 VM 提供支撑，并直接决定性能。
  - 数据中心存储由本地存储和共享存储两大类。
  - 共享存储的重要性体现在三个方面：
    - 先进功能的支持：
      - vSphere 高可用性（HA）、分布式资源调度器（DRS）、容错（FT）以及 VMware Site Recovery Manager 的部分功能都依赖共享存储。
    - 性能：
      - 虚拟机（VM）和整个 vSphere 集群的整体性能都取决于共享存储。
    - 可用性：
      - 虚拟化整体可用性，以及运行的 VM，取决于共享存储基础结构。

4. vSphere Storage

4.2 共享存储





4. vSphere Storage

4.3 存储阵列

存储阵列 Storage arrays

华为存储基础课程
https://www.bilibili.com/video/BV1Zt4y1a7Rj

4. vSphere Storage

AFAs
flash memories

Hybrid
Array
flash and HDD

4. vSphere Storage

Concerning the protocols used for frontend interfaces

前端接口面板

将磁盘阵列接入到虚拟化服务器或者存储网络中

4. vSphere Storage

VMware vSphere 支持的存储接口及应用案例

Protocol type	Type of service	Interface speed	Typical usage
SAS	Block	6 or 12 Gbps	Shared storage with limited host scaling
FC	Block	8, 16, 32 Gbps	Shared storage, typically for enterprises
FCoE	Block	10, 25, 40, 50, 100 Gbps	Shared storage, typically for enterprises and mid-sized businesses
iSCSI	Block	1, 10, 25, 40, 50, 100 Gbps	Shared storage
NFS	File	1, 10, 25, 40, 50, 100 Gbps	Shared storage

4. vSphere Storage

- 存储性能最主要指标是：IOPS
  - Read and Write I/O per second
  - IOPS（Input/Output Operations Per Second）是衡量存储系统性能的一种指标，表示系统每秒钟能够执行的输入/输出操作数量。
    - 通常用于评估存储设备（如硬盘驱动器、固态硬盘或存储阵列）的性能。
    - 较高的 IOPS 值通常表示存储系统具有更快的数据访问速度和更好的性能表现。
    - IOPS 是评估存储系统性能和选择适当存储解决方案时的重要指标之一。
  - 影响存储性能的因素有很多，例如：
    - 存储使用的物理磁盘或 SSD
    - 存储控制器的成熟度
    - 物理介质
    - 使用的协议和 IO 大小

4. vSphere Storage

影响存储性能的因素

Physical disk or SSDs
physical media
the maturity of the storage controller
the IO size
the protocols used

最重要的影响因素:

- The **Redundant Array of Independent Disks** (RAID) level of the logical volume presented to the ESXi servers
- The device type that is used to form a logical volume (逻辑层设备类型)

4. vSphere Storage

- RAID （ Redundant Array of Independent Disks ）
- 独立磁盘冗余阵列，通常简称为磁盘阵列。
  - RAID 是由多个独立的高性能磁盘驱动器组成的磁盘子系统，提供比单个磁盘更高的存储性能和数据冗余的技术。
  - RAID 是一种多磁盘管理技术，其向主机环境提供了成本适中、数据可靠性高的高性能存储。
  - RAID 的初衷是为大型服务器提供高端的存储功能和冗余的数据安全。
  - RAID 的两个关键目标是提高数据可靠性和 I/O 性能。
  - RAID 的三个关键概念和技术:
    - 镜像（ Mirroring ）
    - 数据条带（ Data Stripping ）
    - 数据校验（ Data parity ）

4. vSphere Storage

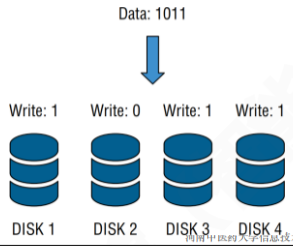
- RAID level
  - RAID使用多个物理存储设备形成逻辑卷，使用逻辑卷有两个原因：
    - 冗余：在一个逻辑卷中有多个设备（HDD 或 SSD），基于 RAID 可以在一个或多个物理存储设备故障时不影响逻辑卷的可用性。
    - 性能：RAID 形成的逻辑卷比物理存储设备具有更好的性能和容量。
- 最常用的RAID等级：

RAID 0 | RAID 1 | RAID 1+0 | RAID 5 | RAID 6

RAID 技术全解（技术文章）
https://cloud.tencent.com/developer/article/2304179

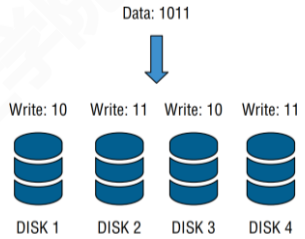
4. vSphere Storage

- RAID level
  - RAID 0：
    - 又称Stripe或Striping，条带化。
    - 将两块以上的硬盘合并成一块，数据连续地分割在每块盘上。
    - 因为带宽加倍，所以读/写速度加倍，但RAID 0在提高性能的同时，并没有提供数据保护功能，只要任何一块硬盘损坏就会丢失所有数据。



4. vSphere Storage

- RAID level
  - RAID 1 , 1+0 , 0+1 :
    - RAID 1是将一个两块硬盘所构成RAID磁盘阵列，其容量仅等于一块硬盘的容量，因为另一块只是当作数据“镜像”。
    - RAID 1是最可靠的一种阵列，总是保持一份完整的数据备份。性能不如RAID 0但高于单一硬盘。RAID 1写入速度较慢，因为数据得分别写入两块硬盘中并做比较。

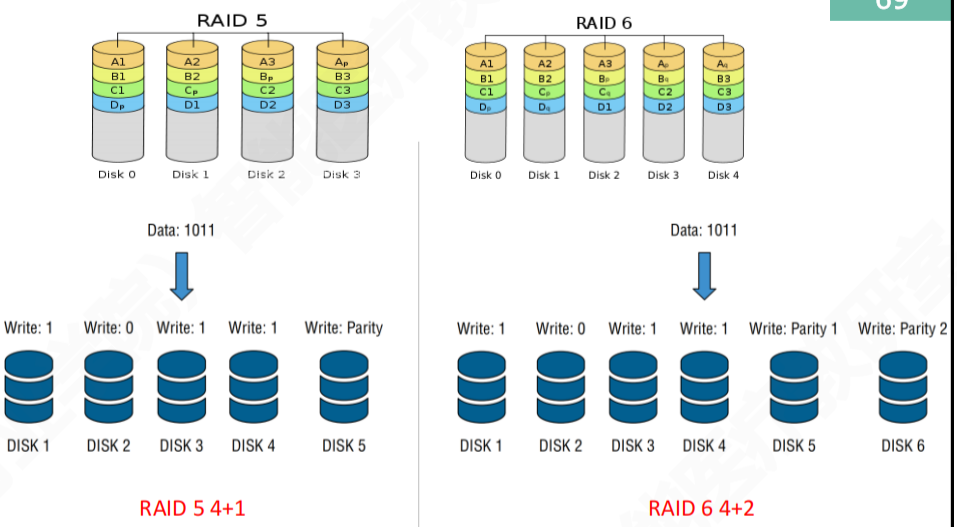


RAID 01 为 RAID 0+1：
先进行条带存放（RAID 0），再进行镜像（RAID 1）

RAID 10 为 RAID 1+0：
先进行镜像（RAID 1），再进行条带存放（RAID 0）

4. vSphere Storage

- RAID level
  - RAID 5 , 6 :
    - RAID 5 是最常见的RAID 等级。
    - RAID 5 磁盘上同时存储数据和校验数据，数据块和对应的校验信息保存在不同的磁盘上，当一个数据盘损坏时，系统可以根据同一条带的其他数据块和对应校验数据来重建损坏的数据。
    - RAID 5 具备良好扩展性，当磁盘数量增加时，并行操作量的能力也随之增长。
    - RAID 5 兼顾存储性能、数据安全和存储成本等各方面因素，可简单理解为 RAID 0 和 RAID 1 的折中方案，是目前综合性能最佳的数据保护解决方案。推荐数据中心使用RAID 5。
    - RAID 6 引入双重校验的概念，同时出现两个磁盘失效时，存储仍能够正常工作，不会发生数据丢失。
    - RAID 6 是在 RAID 5 的基础上为了进一步增强数据保护而设计的一种 RAID 方式，可以看作是一种扩展的 RAID 5 等级。



4. vSphere Storage

- 数据去重：Deduplication
  - 是一种数据存储优化技术，旨在消除数据中重复内容，减少存储空间需求。
  - 通过去除重复的数据块或文件副本，存储系统可以在物理上只保留一份数据副本，而在逻辑上仍然使得所有需要访问该数据的应用程序都能够正常工作。
  - 这种技术对于大型数据集、备份和存档等场景尤其有用，可以显著节省存储空间并提高存储效率。
  - 数据去重有两种方式：
    - In-line：数据在写入其他存储时进行重复数据删除。将数据写入持久缓存，计算哈希值并与存储的块进行比较，如果尚未存储，再将数据写入存储本身。
    - Off-line：数据按原样写入存储，并按定义的时间周期调用重复数据删除任务以删除重复数据。

4. vSphere Storage

4.4 存储性能：复制

- 复制：Replication
  - 复制通常是中端或高端存储阵列软件堆栈的一部分，用于灾难恢复。
  - 复制允许透明地在两个（有时甚至更多）物理存储阵列之间复制数据。
  - 复制复制有两种类型：
    - 同步，Synchronous：
      - 当 ESXi 服务器向底层存储发出存储命令时，数据首先写入主存储，然后再写入辅助存储。
      - 只有当辅助存储确认数据已写入时，才会将确认发送到 VM。
      - 对于同步复制，需要高性能存储区域网络，因为两个存储之间的互连导致的任何延迟都会影响 VM 的整体延迟。
      - 当主存储发生故障时，不会丢失任何数据。
    - 异步，Asynchronous：
      - 收到存储命令后，会将其写入主存储，并立即向 VM 发送确认。然后，在预定义的复制间隔之后，数据将同步到辅助存储。
      - 如果主存储发生故障，将丢失尚未复制到辅助阵列的数据。

4. vSphere Storage

4.4 存储性能：物理存储设备接口

- 物理存储设备接口也是决定存储性能的重要因素
  - 两种常见的物理存储接口标准：SATA 和 SAS
    - SATA（Serial ATA，串行ATA）：
      - SATA 是一种通用存储接口，主要用于个人计算机、消费级存储设备和一些中小型企业级存储系统。
      - SATA 相对便宜，易于使用和部署，适合于一般性的存储需求，在处理大并发时性能不及 SAS。
    - SAS（Serial Attached SCSI，串行连接 SCSI）：
      - SAS 是一种专业的存储接口，通常用于企业级存储系统和高性能计算环境。
      - SAS 接口提供了更高的带宽和更低的访问延迟，适合处理高吞吐量和低延迟的应用程序。
      - SAS 设备通常具有更高的可靠性和更长的寿命，支持更高的并发连接数和更多的同时访问。
      - SAS 支持诸多高级功能，如多路径访问、热插拔、端口多路复用等，有助于提高可用性和可管理性。
    - 对于对存储性能要求较高的企业级应用，SAS 接口往往是更好的选择，而对于一般性的存储需求，SATA 接口则是一个经济实惠的选择。

4. vSphere Storage

- 使用闪存技术提升存储性能
  - 两种使用闪存技术的存储方案：SSDs 和 AFAs
    - SSDs（固态硬盘）：
      - SSDs 是单个存储设备，使用闪存芯片来存储数据。
      - 通常用于替代传统的硬盘驱动器（HDDs），提供更快的数据访问速度、更低的延迟和更高的吞吐量。
      - 可以在个人电脑、服务器和数据中心中使用，提供更高性能和更低功耗的存储解决方案。
    - AFAs（全闪存阵列）：
      - AFA 是一种专门设计的企业级存储解决方案，完全由固态硬盘组成，旨在提供极高的性能和可靠性。
      - 采用全闪存的架构，以实现更快的数据访问速度、更低的延迟和更高的吞吐量。
      - 适用于需要大量随机 I/O 访问和快速数据处理的应用，如数据库、虚拟化、人工智能和大数据分析。
    - SSDs 提供了单个存储设备的闪存解决方案，而 AFAs 是专门设计的全闪存存储阵列，用于提供高性能、高可靠性的企业级存储解决方案。

5. vSphere Storage 基本应用

VMware vSphere 支持的存储体系结构

技术	协议	传输	接口
光纤通道	FC/SCSI、FC/NVMe	数据/LUN 的块访问	FC HBA
以太网光纤通道	FCoE/SCSI	数据/LUN 的块访问	融合网络适配器（硬件 FCoE）
iSCSI	IP/SCSI	数据/LUN 的块访问	■ iSCSI HBA 或启用 iSCSI 的网卡（硬件 iSCSI） ■ 网络适配器（软件 iSCSI）
NAS	IP/NFS	文件（无直接 LUN 访问）	网络适配器

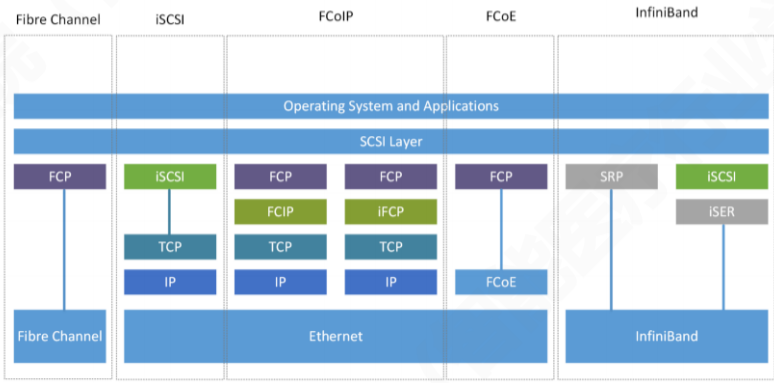
5. vSphere Storage 基本应用

不同存储体系结构支持的 VMware vSphere 特性

存储类型	引导虚拟机	vMotion	数据存储	RDM	虚拟机集群	VMware HA 和 DRS	Storage API - Data Protection
本地存储	是	否	VMFS	否	是	否	是
光纤通道	是	是	VMFS	是	是	是	是
iSCSI	是	是	VMFS	是	是	是	是
NFS 上的 NAS	是	是	NFS 3 和 NFS 4.1	否	否	是	是

5. vSphere Storage 基本应用

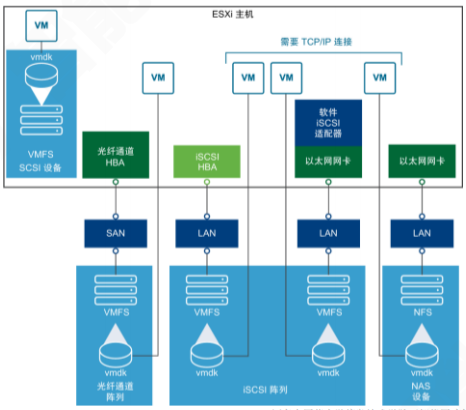
SCSI 命令从 VM Guest OS 到物理网络的流程



FCP：Fibre Channel Protocol，光纤通道协议。

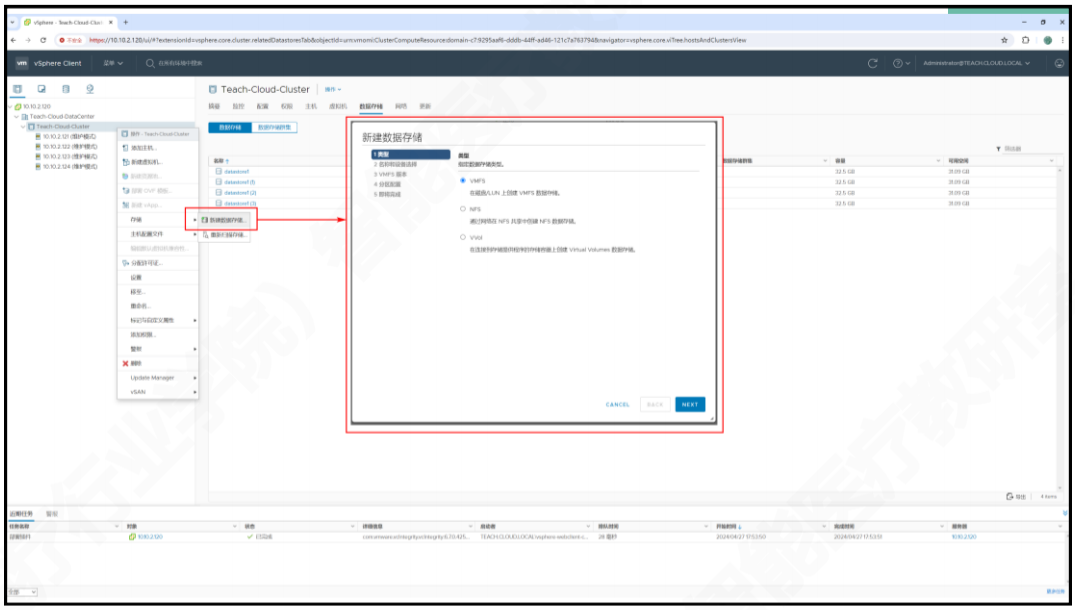
5. vSphere Storage 基本应用

SCSI 命令从 VM Guest OS 到不同体系结构存储的流程



5. vSphere Storage 基本应用

- ESXi 支持的四种数据存储类型
  - VMFS（VMware文件系统）：
    - 专门为虚拟化环境设计的文件系统，用于存储虚拟机的虚拟磁盘文件和其他相关文件。支持多路径访问、高可用性（通过存储 I/O 控制和多路径访问）、支持快照和克隆、以及支持存储 I/O 控制。
  - NFS（网络文件系统）：
    - 基于网络的文件共享协议，允许 ESXi 主机通过网络连接到远程存储设备，并将其挂载为数据存储，通常用于共享存储资源。
  - vVOL（虚拟 Volumes）：
    - 将物理硬件资源抽象到逻辑容量池来虚拟化存 储设备。
  - vSAN（虚拟 SAN）：
    - 软件定义存储解决方案，利用 ESXi 主机上的本地存储资源创建虚拟 SAN，用于存储虚拟机和其他数据。



5. vSphere Storage 基本应用

- ESXi 在物理层上通过三种方式访问存储
  - Block-based storage accessed by a hardware adapter:
    - 使用硬件适配器访问块存储
    - 例如：DAS、FC-SAN，需要服务器配置相应的存储卡
  - Block-based storage accessed by a software adapter:
    - 使用软件适配器访问块存储
    - 例如：iSCSI-SAN，需要正确配置网络连接。
  - NFS storage:
    - NFS 存储
    - 需要正确配置 IP 网络，能够访问到 NFS 数据存储。

5. vSphere Storage 基本应用

5.2 ESXi 支持的物理存储和逻辑存储

- 使用物理存储适配器时，ESXi 支持五种存储通讯协议和技术
  - Fibre Channel Host Bus Adapter (FC HBA)：https://www.unicaca.com/info/detail/244.html
  - iSCSI HBA：https://sniansfblog.org/tag/iscsi
    - 专用的PCIe卡，可完全在硬件中实现整个iSCSI堆栈，从而减少主机CPU的负载。
  - CNA adapters for FCoE or iSCSI：https://blog.csdn.net/wuzhimang/article/details/52399064
    - 主要是 10 Gbps（或更高）以太网适配器。
    - 在融合（或专用）网络上提供硬件（或硬件辅助）FCoE 或 iSCSI 功能。
  - RDMA over Converged Ethernet (RoCE)：https://zhuanlan.zhihu.com/p/361740115
    - 通过以太网进行远程直接内存访问 （RDMA）。
  - InfiniBand HCA：https://www.cnblogs.com/D-Tec/p/3157582.html
    - Mellanox Technologies InfiniBand HCA。

5. vSphere Storage 基本应用

5.3 VM 支持的物理存储和逻辑存储

- Thick Provision Lazy Zeroed：厚置备延迟置零
- Thick Provision Eager Zeroed：厚置备快速置零
- Thin Provision：精简置备

5. vSphere Storage 基本应用

- VM 逻辑存储级别（即虚拟磁盘）有四种类型
  - Thick Provision Lazy Zeroed（厚置备延迟置零）：
    - 在创建虚拟磁盘时，会预先分配所需的存储空间，并将磁盘初始化为零值，但只有在虚拟机首次写入数据时才会执行初始化。
    - 提供了最佳的性能，并且避免了创建磁盘时的延迟，但可能会浪费一些存储空间。
  - Thick Provision Eager Zeroed（厚置备快速置零）：
    - 在创建虚拟磁盘时，会预先分配所需的存储空间，并将磁盘立即初始化为零值。
    - 提供了最佳的性能，并且确保磁盘的内容是安全的，但可能会消耗一些额外的时间来初始化磁盘。

5. vSphere Storage 基本应用

- VM 逻辑存储级别（即虚拟磁盘）有四种类型
  - Thin Provision（精简置备）：
    - 在创建虚拟磁盘时，只会预留所需的存储空间，并不会立即分配实际的存储空间。存储空间只有在虚拟机写入数据时才会被动态分配。
    - 可以节省存储空间，并且可以根据需要动态增长，但可能会对性能产生一定的影响，尤其是在存储空间不足时。
  - Raw Device Mapping (RDM)（原始设备映射）：
    - 允许虚拟机直接访问存储阵列上的物理磁盘或逻辑单元（逻辑卷），而不是使用虚拟磁盘文件（VMDK）。
    - RDM 提供了更接近物理硬件的访问方式，并且可以通过兼容性模式或原始模式进行配置，以满足不同的需求。

以默认的厚格式创建虚拟磁盘。
创建过程中为虚拟磁盘分配所需空间。
创建时不会擦除物理设备上保留的任何数据，但是以后从虚拟机首次执行写操作时会按需要将其置零。

简单的说：
立刻分配指定大小空间，空间内数据暂时不清空，以后按需清空。

创建支持群集功能的厚磁盘。
在创建时为虚拟磁盘分配所需的空间。
在创建过程中会将物理设备上保留的数据置零。
创建这种格式的磁盘所需的时间可能会比创建其他类型的磁盘长。

简单的说：
就是立刻分配指定大小的空间，并将该空间内所有数据清空。

- vSphere VM 磁盘规格
  - 厚置备
    - 厚置备快速置零：zeroed thick
    - 厚置备延时置零：eager zeroed thick
  - 精简置备
    - thin

使用精简置备格式。
精简置备的磁盘只使用该磁盘最初所需要的数据存储空间。
如果以后精简磁盘需要更多空间，则它可以增长到为其分配的最大容量。

简单的说：
为该磁盘文件指定增长的最大空间，需要增长的时候检查是否超过限额。

5. vSphere Storage 基本应用

5.3 VM 支持的物理存储和逻辑存储

5. vSphere Storage 基本应用

- VM 支持的存储控制器类型
  - BusLogic 并行：
    - 最早在VMware ESX中可用的模拟 SCSI 控制器之一。
    - 主要用于旧版操作系统，不支持大于 2TB 的 VMDK。
  - LSI Logic 并行：
    - 是一个可用的 SCSI 虚拟控制器。
    - 用于操作系统，如 Windows Server 2003。
  - LSI Logic SAS：
    - vSphere 4.0 引入，并且是并行驱动程序的演进。
    - 作为SAS虚拟控制器工作，用于 Windows Server 2008 或更新版本。

5. vSphere Storage 基本应用

- VM 支持的存储控制器类型
  - VMware paravirtual（或PVSCSI）：
    - VMware 准虚拟化，vSphere 4.0 引入的。
    - 是一个 SCSI 虚拟控制器，旨在以最小的处理成本支持非常高的吞吐量，工作不是在模拟模式下，而是在 paravirtual 模式下（需要VMware工具才能识别）。
  - NVMe：
    - 虚拟 NVMe 设备降低了客户端 I/O 处理开销。
    - 允许每个主机承载更多的虚拟机或每分钟执行更多的事务。
    - 每个VM支持四个 NVMe 控制器，每个控制器最多支持 15 个设备。
  - SATA：
    - 通常用于桌面虚拟化和测试环境，不支持高性能和高可用性需求。

Others virtual controllers are also possible in a VM, such as AHCI SATA (introduced in vSphere 5.5), IDE, and also USB controllers, but usually for specific cases (for example, SATA or IDE are usually used for virtual DVD drives).

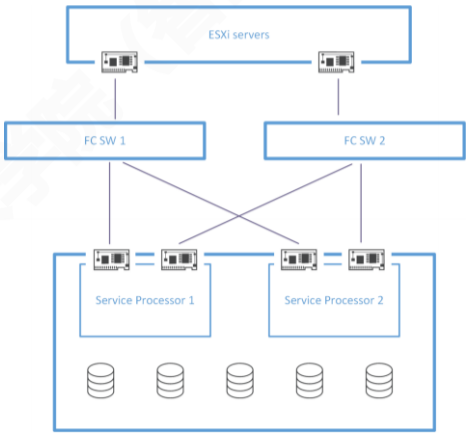
VMware 存储控制器兼容性

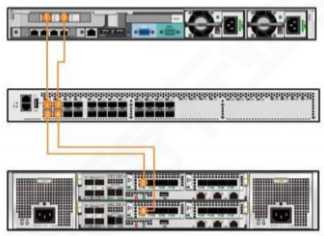
现有控制器	添加的控制器						
	BusLogic 并行	LSI Logic	LSI Logic SAS	VMware 准虚拟 SCSI	AHCI SATA	IDE	NVMe
BusLogic 并行	是	是	是	是	是	是	是
LSI Logic	是	是	是	是	是	是	是
LSI Logic SAS	需要 BIOS 设置	需要 BIOS 设置	通常生效	通常生效	需要 BIOS 设置	是	通常生效
VMware 准虚拟 SCSI	需要 BIOS 设置	需要 BIOS 设置	通常生效	通常生效	需要 BIOS 设置	是	通常生效
AHCI SATA	需要 BIOS 设置	需要 BIOS 设置	是	是	是	是	是
IDE	是	是	是	是	是	不适用	是
NVMe	需要 BIOS 设置	需要 BIOS 设置	通常生效	通常生效	需要 BIOS 设置	是	通常生效

SCSI、SATA 和 NVMe 存储控制器条件、限制和兼容性：
https://docs.vmware.com/cn/VMware-vSphere/7.0/com.vmware.vsphere.vm_admin.doc/GUID-5872D173-A076-42FE-8D0B-9DB0EB0E7362.html

5. vSphere Storage 基本应用

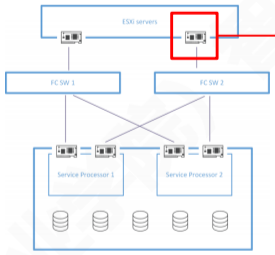
5.4 vSphere 存储的配置：FC





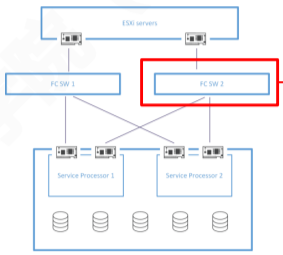
5. vSphere Storage 基本应用

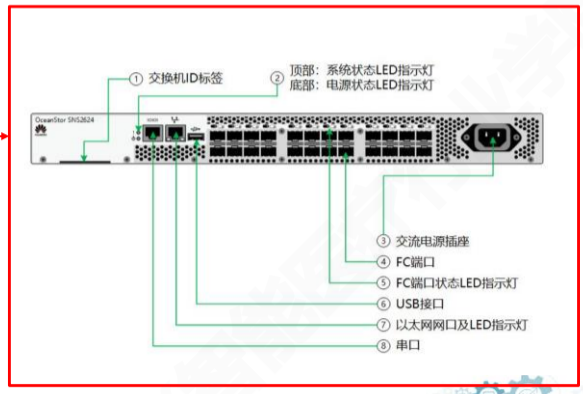
5.4 vSphere 存储的配置：FC



5. vSphere Storage 基本应用

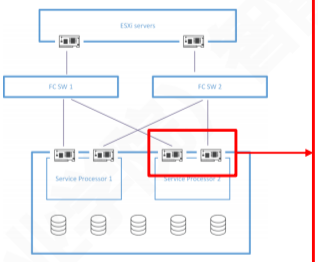
5.4 vSphere 存储的配置：FC

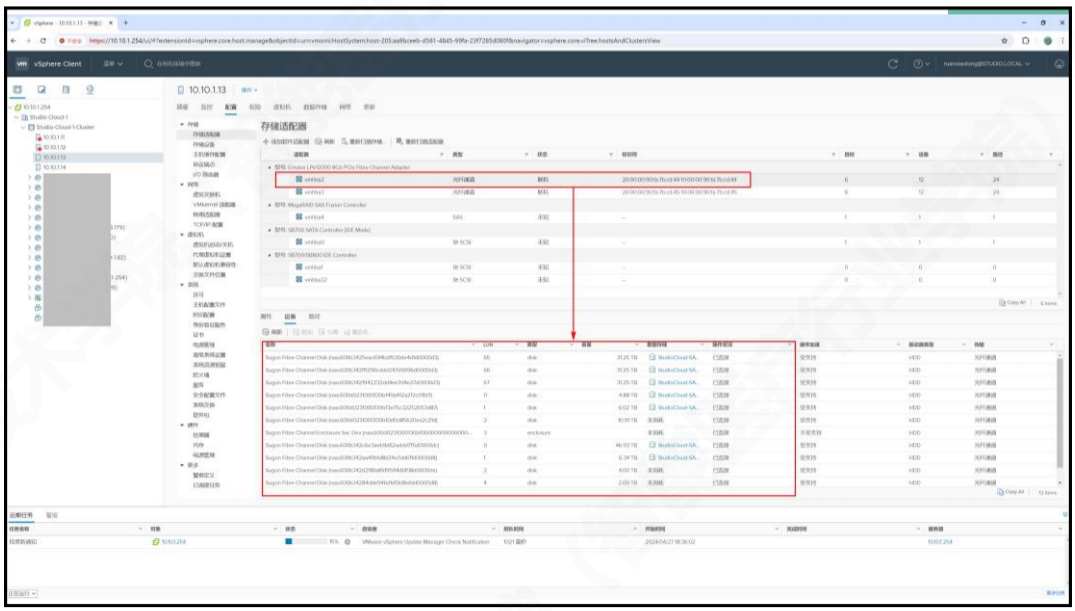




5. vSphere Storage 基本应用

5.4 vSphere 存储的配置：FC



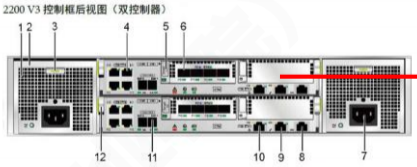


2U 12 盘位控制框前视图

1	硬盘模块	2	保险箱盘标识
3	信息牌（带ESN号）	4	控制框ID显示器
5	电源指示灯/电源开关按钮	6	硬盘模块卡扣

存储阵列

2200 V3 控制框后视图（双控制器）



1	电源模块	2	电源模块拉手
3	电源模块卡扣	4	GE电端口
5	接口模块拉手	6	SmartIO端口
7	电源模块插座	8	串口
9	维护网口	10	管理网口
11	mini SAS HD级联端口	12	控制器拉手

FC 存储交换机

FC HBA（Fibre Channel Host Bus Adapter）

服务器（ESXi 主机）

NVIDIA Quantum-2 Switches

Orderable Part Number (OPN)	Description
MQM9790-NS2F	NVIDIA Quantum-2-based 400Gb/s InfiniBand switch, 64 400Gb/s ports, 32 OSFP ports, non-blocking switching capacity of 51.2Tb/s, two power supplies (AC), standard depth, unmanaged, power-to-connector (P2C) airflow, rail kit
MQM9790-NS2R	NVIDIA Quantum-2-based 400Gb/s InfiniBand switch, 64 400Gb/s ports, 32 OSFP ports, non-blocking switching capacity of 51.2Tb/s, two power supplies (AC), standard depth, unmanaged, connector-to-power (C2P) airflow, rail kit
MQM9700-NS2F	NVIDIA Quantum-2-based 400Gb/s InfiniBand switch, 64 400Gb/s ports, 32 OSFP ports, non-blocking switching capacity of 51.2Tb/s, two power supplies (AC), standard depth, managed, P2C airflow, rail kit
MQM9700-NS2R	NVIDIA Quantum-2-based 400Gb/s InfiniBand switch, 64 400Gb/s ports, 32 OSFP ports, non-blocking switching capacity of 51.2Tb/s, two power supplies (AC), standard depth, managed, C2P airflow, rail kit

ConnectX-7 Adapters

PCIe Standup Adapters

Orderable Part Number (OPN)	Description
MCX755106AC-HEAT	NVIDIA ConnectX-7 HHHL adapter card, 200GbE (default mode) / NDR200 IB, dual-port QSFP112, PCIe 5.0 x16 with x16 PCIe extension option, crypto enabled, Secure Boot enabled, tall bracket
MCX755106AS-HEAT	NVIDIA ConnectX-7 HHHL adapter card, 200GbE (default mode) / NDR200 IB, dual-port QSFP112, PCIe 5.0 x16 with x16 PCIe extension option, crypto disabled, Secure Boot enabled, tall bracket
MCX75310AAC-NEAT	NVIDIA ConnectX-7 HHHL adapter card, 400GbE / NDR IB (default mode), single-port OSFP, PCIe 5.0 x16, crypto enabled, Secure Boot enabled, tall bracket

5. vSphere Storage 基本应用

- FCoE：
  - 以太网上封装FC帧的技术。
  - FCoE 网络在第1层和第2层使用 10Gbps（或更高）的以太网网络，但其余仍然是FC协议栈。
    - FC 是一个完整的网络栈，不使用 IP、UDP 或 TCP 协议。
    - FCoE 仍支持 FC 的特性。
  - 如何使用FCoE呢？
    - ESXi Host 需要安装 CAN 卡，使用 NPAR 协议 (Network partition)。
    - 需要使用 FCoE 专用交换机。
    - 存储阵列和通信网络可以不做改变直接接入 FCoE。
    - 在 vSphere 中进行 FCoE 的配置。

5. vSphere Storage 基本应用

华为CloudEngine系列交换机 FC 和 FCoE 特性介绍
https://e.huawei.com/cn/videolist/video/25b1591c674c4881a4f8ac371a2ac199

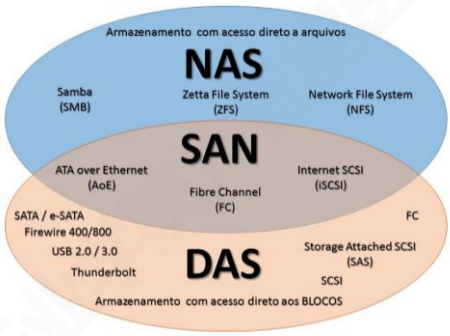
5. vSphere Storage 基本应用

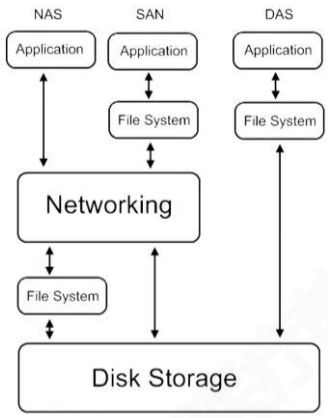
- iSCSI
  - ESXi 支持以下三种 iSCSI
    - Software iSCSI adapter：
      - 利用一个或多个 VMkernel 网络接口和虚拟交换机来管理整个iSCSI流量。
      - 使用软件 iSCSI 适配器，无需购买专用硬件。
    - Dependent hardware iSCSI adapter：
      - 使用硬件的 iSCSI 适配器（iSCSI 卡），但由 VMware vSphere 进行管理和配置。
    - Independent hardware iSCSI adapter or iSCSI HBA：
      - 如同 FC HBA。
      - 所有网络堆栈都是在适配器内部的硬件中实现的。
      - 在 ESXi 端，将看到一个或多个 vmhba，就像其他块存储适配器一样。
      - 必须使用 BIOS 管理或特定工具在卡级别执行网络配置。

5. vSphere Storage 基本应用

- NFS：
  - ESXi 支持的 NAS 协议是 NFS。
    - 不支持 SMB。
  - 支持NFS协议版本是：
    - NFS 3
    - NFS 4.1（vSphere6.0 及以后版本）

101





102

SAN

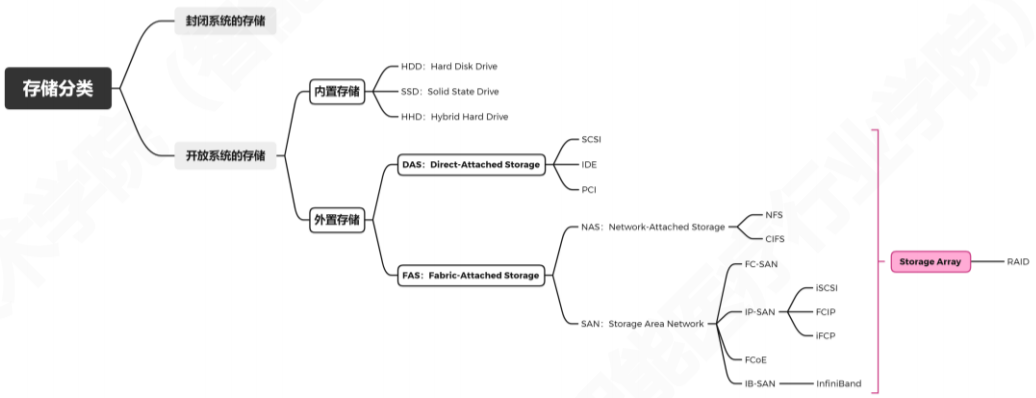
- 存储区域网络 (Storage Area Network, SAN) 是企业最常用的存储网络架构，要求高吞吐量和低延迟的业务关键型应用往往采用这类架构运行。
- SAN 将数据存储在集中式共享存储中，使企业能够运用一致的方法和工具来实施安全防护、数据保护和灾难恢复。
- SAN 是一种基于块的存储，利用高速架构将服务器与其逻辑磁盘单元 (Logical Disk Unit, LUN) 相连。LUN 是一系列通过共享存储池配置的块，以逻辑磁盘的形式呈现给服务器。服务器会对这些块进行分区和格式化，通常使用文件系统，以便可以像在本地磁盘存储上一样在 LUN 上存储数据。
- SAN 约占网络存储市场总额三分之二。SAN 的设计可消除单点故障，具有极高的可用性和故障恢复能力，设计完善的 SAN 可以轻松承受多个组件或设备的故障。

103

NAS

- NAS（Network Attached Storage）网络附加存储。
- 在NAS存储结构中，存储系统不再通过I/O总线附属于某个服务器或客户机，而直接通过网络接口与网络直接相连，由用户通过网络访问。
- NAS实际上是一个带有瘦服务器的存储设备，其作用类似于一个专用的文件服务器。这种专用存储服务器去掉了通用服务器原有的不适用的大多数计算功能，而仅仅提供文件系统功能。
- 与传统以服务器为中心的存储系统（如DAS）相比，数据不再通过服务器内存转发，直接在客户机和存储设备间传送，服务器仅起控制管理的作用。

104



6. vSphere Storage 高级功能

- SIOC 和 Storage DRS是解决存储资源在高负载情况下的争用冲突。
  - 启用SIOC，可以在VM对存储资源使用发生争用时，对虚拟机访问存储资源的优先级和访问压力进行限制管理。
  - 由于存储在所有虚拟机之间是共享的，当特定虚拟机大量且无限制的进行I/O操作视，可能会大量占用存储容量和资源，导致所有虚拟机性能变差。

6. vSphere Storage 高级功能

- SIOC 和 Storage DRS是解决存储资源在高负载情况下的争用冲突。
  - Storage DRS：Storage Distributed Resource Scheduler (SDRS)
  - 启用 SDRS 后，VM 优化和资源分配依据两个指标：space（容量）和 I/O
  - Storage DRS 管理数据存储群集的聚合资源。
    - 如果启用 Storage DRS，会对虚拟机磁盘放置和迁移提出建议，以平衡数据存储群集中各个数据存储之间的空间和 I/O 资源。
    - 启用 Storage DRS 时，可以使用如下存储集群的功能：
      - 数据存储群集中数据存储之间的空间负载平衡
      - 数据存储群集中数据存储之间的 I/O 负载平衡
      - 基于空间和 I/O 工作负载的虚拟磁盘的初始放置
    - Storage DRS 仅应用于数据存储集群，不能在单个数据存储中使用。

6. vSphere Storage 高级功能

6.2 数据存储集群

- 数据存储集群是一组数据存储
  - 是一种逻辑上的存储资源池，将多个数据存储（Datastore）组合成一个集群，并提供统一的管理和分配。
  - 为虚拟机提供了更高的存储可用性和性能，同时简化了存储管理的复杂性。
  - 优点与特性：
    - 高可用性：可以跨多个物理存储设备分布数据，从而提高了存储资源的可用性。
    - 负载均衡：可以在多个存储设备之间动态平衡虚拟机的负载，确保每个存储设备的负载均衡，并提供了更高的存储性能。
    - 简化管理：可以在一个集群级别上管理和配置存储资源，简化了存储管理。
    - 存储策略应用：可以与存储策略（Storage Policy）结合使用，根据业务需求定义不同的存储策略，从而实现对存储资源的灵活管理和配置。

6. vSphere Storage 高级功能

6.3 VMFS 6

- VMFS版本
  - VMFS 1：
    - 用于 ESX Server 1.x，不支持集群。不支持对多个服务器的并发访问。集群不可用，一次只能在一台服务器上使用。
  - VMFS 2：
    - 用于ESX Server 2.x 或 3.x，不支持目录结构。
  - VMFS 3：
    - 适用于vSphere 3.x 或 4.x，支持目录结构。至多50TB容量。VMFS 3 不适用 ESXi 7.0。
  - VMFS 5：
    - 用于vSphere 5.x 和 6.x。单个盘区大小增加到 64TB，VMDK文件大小增加到 62TB。
  - VMFS 6：
    - 在 vSphere 6.5 发布，vSphere 6.7 可以使用。支持4Kn和512n存储设备和本地设备。

6. vSphere Storage 高级功能

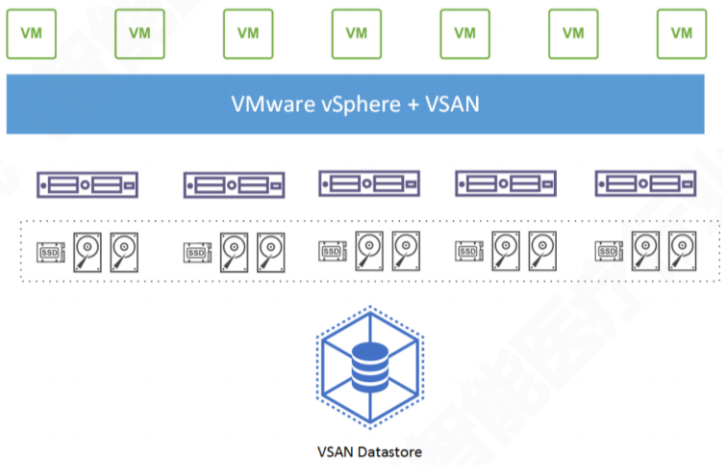
- VMFS版本
  - VMFS 6 的特点和优势
    - 64位文件系统：
      - 引入 64 位文件系统，支持更大的文件和存储容量，可以有效管理大规模的虚拟化环境。
    - 增强的存储性能：
      - 通过优化和改进了存储访问算法，提供了更高的存储性能和更低的延迟，提升了虚拟机的性能和响应速度。
    - 更强的可靠性和一致性：
      - 引入了新的存储校验和功能，可以检测和纠正数据损坏，提高了存储的可靠性和一致性。
    - 增强的管理功能：
      - 提供更丰富的管理，包括在线扩容、在线存储迁移、存储I/O控制和存储策略应用等，简化了存储管理的操作和流程。
    - 与其他VMFS版本的兼容性：
      - 与早期版本的 VMFS 文件系统兼容，并且可以无缝地升级文件系统到 VMFS 6，而不会影响虚拟机的正常运行。

6. vSphere Storage 高级功能

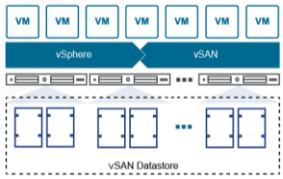
- vSphere 存储管理上新增了诸多高级特性
  - Instant clones versus linked clones：快速克隆与链接克隆的存储支持
  - Permanent Device Loss (PDL) and All-Paths Down (APD)：
    - VM Component Protection (VMCP) 能够管理的两种 VM 存储风险
  - Flash Read Cache：vFlash，闪存读缓存
  - Storage integration：存储集成
  - Multipathing：多路径访问LUN
  - VMware vStorage API for Array Integration (VAAI)：用于和存储设备进行深度整合
  - VMware vSphere APIs for I/O Filtering (VAIO)：用于I/O 过滤 vSphere API框架

7. VMware vSAN

- VMware vSAN 是软件定义的企业存储解决方案，支持超融合基础架构 (Hyper-Converged Infrastructure, HCI) 系统。
  - vSAN 与 VMware vSphere 完全集成在一起，是 ESXi 内置的功能。
  - vSAN 使用本地存储设备或直接连接的存储设备，创建 vSAN 集群中所有主机之间共享的单个存储池。
    - 混合 vSAN 集群将闪存设备用于缓存层，将磁盘驱动器用于容量层。
    - 全闪存 vSAN 集群将闪存设备用于缓存层和容量层。
    - 可用于软件定义数据中心（SDDC），且是经过闪存优化的弹性共享数据存储。
  - vSAN 无需外部共享存储，并通过基于存储策略的管理 (Storage Policy-Based Management, SPBM) 简化了存储配置。



7. VMware vSAN



VMware vSAN：
https://docs.vmware.com/cn/VMware-vSAN/index.html

7. VMware vSAN

VMware vSAN 介绍
https://www.bilibili.com/video/BV1j4411K7ZV（3分40秒）
https://www.bilibili.com/video/BV1Ls411K7AJ（31分51秒）

115

7. VMware vSAN

7.1 vSAN的方案

- vSAN实现的两种设备选型方案：
  - Hybrid deployment：混合部署
    - 使用固态硬盘作为缓存层，磁盘作为容量层。
    - 不需要使用 RAID 5/6 进行去重和压缩，以及存储容灾。
  - All-flash：全闪存（推荐方案）
    - 使用 High-performance SSD （例如 NVMe）作为缓存层，使用 SSDs 作为容量层。
    - All-flash 能够提供非常高的 I/O 性能。
  - 网络要求：
    - 建议使用成对的 10 GbE Network Cards。
    - 将 vSAN 流量进行隔离，并为 vSAN 提供充足带宽。

116

7. VMware vSAN

7.2 vSAN的实现

现场演示：
vSAN 的实现与应用

117

SDDC的虚拟化数据中心设计方案

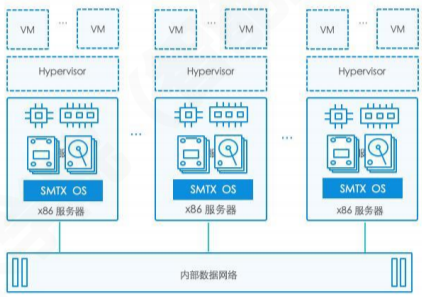
中央存储

虚拟化服务器
存储阵列
虚拟化软件

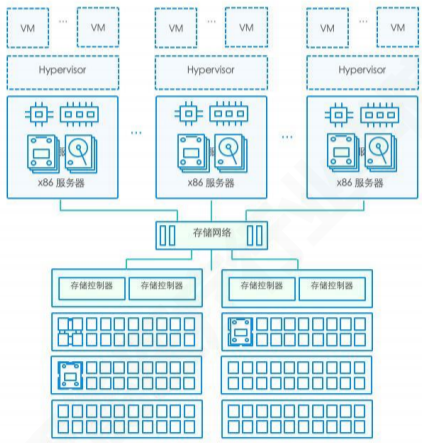
超融合

虚拟化服务器
虚拟化软件(vSAN)

118



超融合的架构



传统架构

信创智能医疗系统研发课程体系

河南中医药大学信息技术学院（智能医疗行业学院）

河南中医药大学信息技术学院（智能医疗行业学院）智能医疗教研室

河南中医药大学医疗健康信息工程技术研究所