

云计算技术与应用



第08章：Virtual Machine Resource Management

<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

- ✓ **虚拟机的资源管理**
 - 资源管理的模式：预留、限制、份额 (Reservation, Limit, Share)
 - 虚拟机资源管理：MEM、CPU、ESXi Host memory states
- ✓ **资源池与 vAPP**
- ✓ **网络与存储的资源管理**
 - NIOC: Network I/O Control
 - SIOC: Storage I/O Control
- ✓ **虚拟机迁移**
 - Compute vMonitor
 - Storage vMonitor
 - Cross-vCenter vMotion
 - DRS: Distributed Resource Scheduler, 分布式资源调度
 - DPM: Distributed Power Management, 分布式电源管理



1. 虚拟机的资源管理

1.1 资源管理的模式

- VMware vSphere 有一系列控制措施来保证虚拟机访问资源、控制资源使用、在可用资源不足时确定资源调度优先级。
- VMware vSphere 资源管理的方法有三种。（通用）
 - 预留：Reservations
 - 指定提供的最小资源，是对使用资源的保证。
 - 如果要确保无论发生什么情况，都需要确保特定 VM 或一组 VM 从启动到关闭对一定数量资源进行访问，则可以使用预留。
 - 预留的默认值是 0。
 - 发生资源争用时，预留资源也能够得到保证。
 - 预留不是独占，当预留的资源没有被使用时，也会被共享使用。



1. 虚拟机的资源管理

1.1 资源管理的模式

- VMware vSphere 有一系列控制措施来保证虚拟机访问资源、控制资源使用、在可用资源不足时确定资源调度优先级。
- VMware vSphere 资源管理的具体方法有三种。
 - 限制: Limits
 - 指定可以使用的最大资源量。
 - 如果没有指定, 则根据配置可使用最大资源量。
 - 如果进行指定, 则使用资源量不会超过限制使用的资源量。
 - 能够使用的最大资源量 = $\min(\text{配置的资源量}, \text{限制的资源量})$
 - 例如:
 - VM 配置 16 vCPU, 每个 vCPU 的运行频率为 2 GHz, 限制为 10 GHz, 则从 VM Guest OS 上看的配置是 16 vCPU * 2 GHz, 但实际可用的只有 10 GHz。

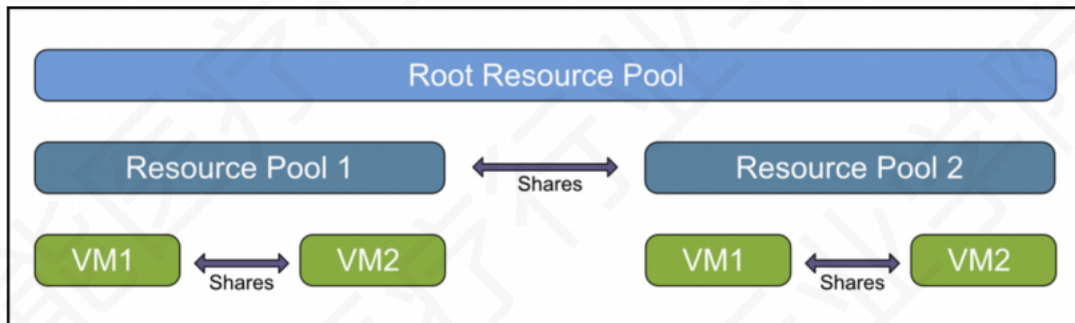
1. 虚拟机的资源管理

1.1 资源管理的模式

- VMware vSphere 有一系列控制措施来保证虚拟机访问资源、控制资源使用、在可用资源不足时确定资源调度优先级。
- VMware vSphere 资源管理的具体方法有三种。
 - 份额: Shares
 - 指定在争用期间获取资源的优先级。
 - 当 ESXi Host 中的资源有限且虚拟机争用访问资源时，配置更高份额的虚拟机将具有更高优先级访问更多资源。
 - 份额可以指定为高、正常或低，比例为 4: 2: 1，并在 vSphere 层次结构中的同级之间应用。
 - 如果不使用资源池或 vApp，则所有虚拟机都将位于层次结构中的同一级别，因此共享将在所有虚拟机之间拆分。



资源争取期按照份额分配



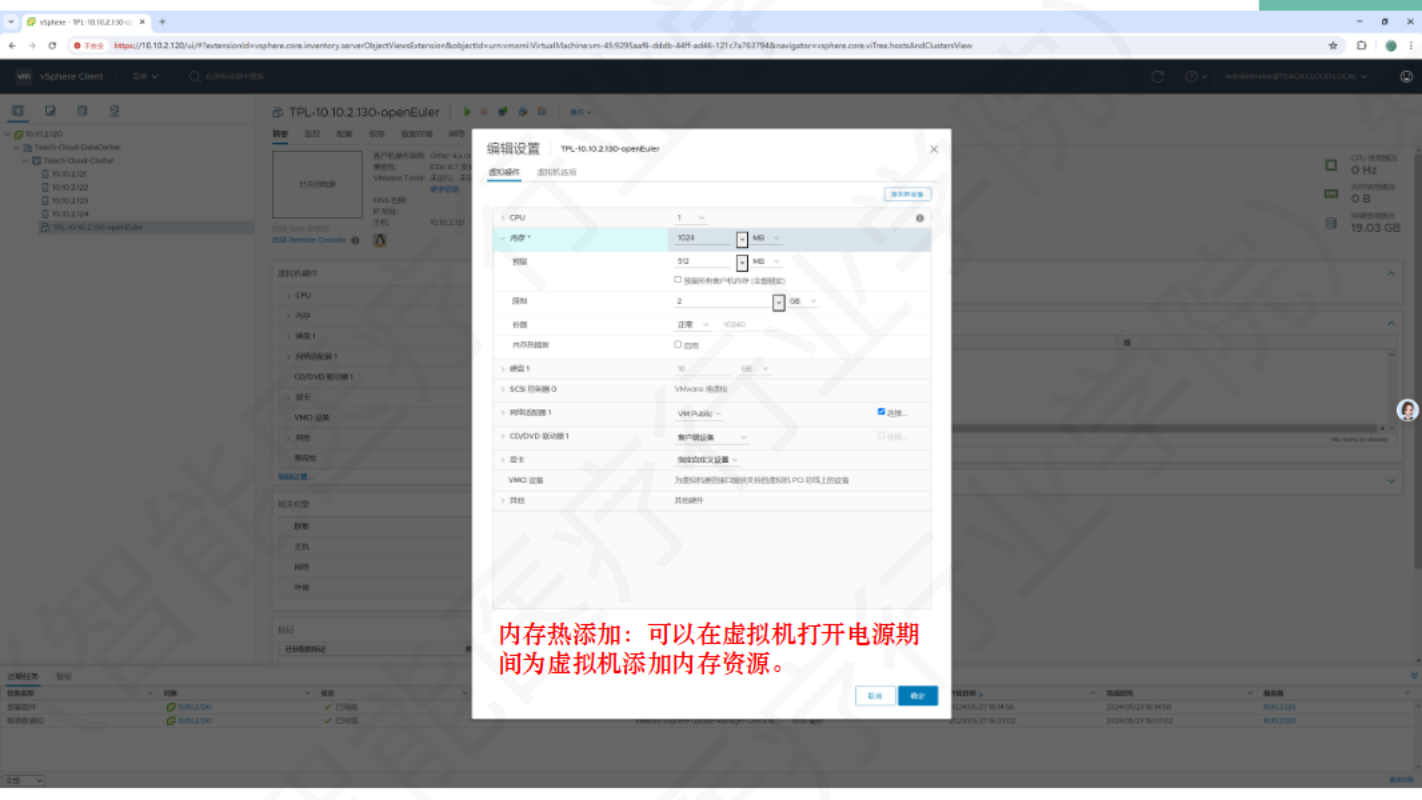
1. 虚拟机的资源管理

1.2 虚拟机内存资源的管理

□ 讨论:

- ESXi Host 配置 32GB 物理内存 2 条，共计 64GB 物理内存。
 - （假设 ESXi Host 不占用物理内存，64GB 物理内存可以全部用于 VM。）
- 问题：
 - 第 1 台 VM 分配 8GB 内存，该内存存在哪个物理内存条上？或均分在两个内存条？
 - 已经创建 8 台 VM，每台 VM 分配 8GB 内存，第 9 台 VM 是否可以启动？
 - 假设某 VM 需要 48GB 内存，该 VM 是否可以启动？





内存热添加：可以在虚拟机打开电源期间为虚拟机添加内存资源。

1. 虚拟机的资源管理

1.2 虚拟机内存资源的管理

□ 总结预留、限制、份额如何与内存配合使用

- 对于 VM 的**四类关键资源**，其预留、限制、份额的处理方式略有不同。
- 控制内存资源的分配时：
 - 预留：保证 VM 的特定内存。
 - 在虚拟机请求使用内存之前，不会分配内存，但 ESXi Host 必须有足够的可用内存来满足整个预留，然后虚拟机才能打开电源。
 - 预留内存总额不能超过 ESXi Host 的物理可用内存总量。
 - 预留内存分配给 VM 后，ESXi Host 不会交换预留内存，也不会回收预留内存。
 - 预留内存被 VM 锁定。



1. 虚拟机的资源管理

1.2 虚拟机内存资源的管理

□ 内存资源的预留、限制、份额

- 对于 VM 的**四类关键资源**，其预留、限制和共享的处理方式略有不同。
- 控制内存资源的分配时：
 - 限制：强制内存使用的上限。
 - 使用 VMware Tools 膨胀驱动程序实施限制，可能对性能产生负面影响。
 - 当 VM 接近限制时（Guest OS 是不知道），膨胀驱动程序将内存使用保持在限制以下，会导致 Guest OS 将内存交换到磁盘，进而降低 VM 的性能。
 - 份额：仅在物理内存资源争用期间有效。
 - 建立对 ESXi Host 物理内存的优先访问列表。
 - VM 的优先级基于百分比分配的额度。
 - 没有发生争用时共享不适用，不会影响 VM 内存分配或使用。

□ 四类关键资源：MEM、CPU、Pool、vAPP



1. 虚拟机的资源管理

1.3 ESXi Host 高级内存管理技术

- vSphere ESXi 有五种内存管理技术来确保 VM 高效使用内存。
 - 空闲内存税：Idle Memory Tax
 - 当 ESXi 主机上的虚拟机处于空闲状态时，其仍占用一定的内存资源。
 - 为了提高内存利用率，ESXi 会对空闲虚拟机征收一定的内存税，将这些空闲内存回收并分配给其他需要的虚拟机。
 - 透明页面共享：Transparent Page Sharing
 - 透明页面共享是一种内存共享技术，将多个虚拟机中相同的内存页面共享给其他虚拟机使用，达到减少内存浪费，提高内存利用率的目标。
 - 膨胀：Ballooning
 - 膨胀是一种内存回收技术，将虚拟机中的内存页面压缩，释放出更多内存空间。
 - 虚拟机需要更多内存时，ESXi 会将压缩的内存页面解压，满足虚拟机需求。

1. 虚拟机的资源管理

1.3 ESXi Host 高级内存管理技术

□ vSphere ESXi 有五种内存管理技术来确保 VM 高效使用内存。

■ 内存压缩：Memory Compression

- 内存压缩是一种将内存中的数据进行压缩的技术，可以将内存中的数据压缩到更小的空间中，从而释放出更多的内存空间。
- 虚拟机需要更多内存时，ESXi 会将压缩的内存数据解压，满足虚拟机需求。

■ 交换：Swapping

- 交换是一种将内存中的数据交换到磁盘上的技术，可以将内存中的数据交换到磁盘上，从而释放出更多的内存空间。
- 虚拟机需要更多内存时，ESXi 会将磁盘上的交换数据交换回内存中，满足虚拟机需求。



1. 虚拟机的资源管理

1.3 ESXi Host 高级内存管理技术

□ 空闲内存税 (Idle Memory Tax) 工作原理

- 区分空闲内存和活动内存：
 - 将 VM 内存分为空闲内存 (Idle Memory) 和活动内存 (Active Memory)。
 - 空闲内存是指没有被虚拟机当前使用, 但可以被重新利用的内存。活动内存是虚拟机正在使用的内存。
- 惩罚性计算：
 - 通过空闲内存税对空闲内存进行惩罚性的计算。如果 VM 拥有较多空闲内存, 将获得较少的内存份额。
- 税率设置：
 - 空闲内存税通过百分比来设置, 百分比介于0%到100%之间。税率为 0% 时, 按照份额分配内存。
- 份额内存比：
 - 使用份额内存比 (shares-per-page ratio) 参数计算内存分配。
- 内存回收：
 - 当 ESXi 内存不足时, 寻找份额内存比最小的虚拟机来回收内存。
- 动态调整：
 - 系统根据当前内存使用情况和空闲内存税的设置, 不断地重新分配内存资源。

1. 虚拟机的资源管理

1.3 ESXi Host 高级内存管理技术

□ 透明页面共享（Transparent Page Sharing, TPS）的工作原理

- 内存页面识别：ESXi Host 的虚拟机监控器会持续监控虚拟机的内存页面，寻找完全相同的内存页面。
- 页面哈希：每个内存页面通过计算生成唯一哈希值。这个哈希值用于快速比较不同虚拟机中的页面是否相同。
- 共享相同页面：当两个或多个虚拟机拥有相同内容的内存页面时，只保留一份物理副本并共享使用。
- 内存节省：通过共享相同的内存页面，TPS 减少了物理内存的需求。
- 动态共享：TPS 是一个动态过程，会随着虚拟机内存使用情况的变化而不断调整。
- 内存重定位：
 - 当共享页面需要被修改时，ESXi Host 会将该页面复制到请求修改的虚拟机的私有内存中，这个过程称为内存重定位（Memory Relocation）。
- 性能优化：
 - TPS 有助于提高虚拟化环境的性能，因为其减少了内存的复制和迁移，同时允许虚拟机共享内存资源。
- 内存回收：
 - 如果 ESXi Host 内存资源紧张，可以释放内存，通过回收不再被共享的页面提供更多可用内存。

1. 虚拟机的资源管理

1.3 ESXi Host 高级内存管理技术



论文: Memory Resource Management in VMware ESX Server
<https://www.waldspurger.org/carl/papers/esx-mem-osdi02.pdf>



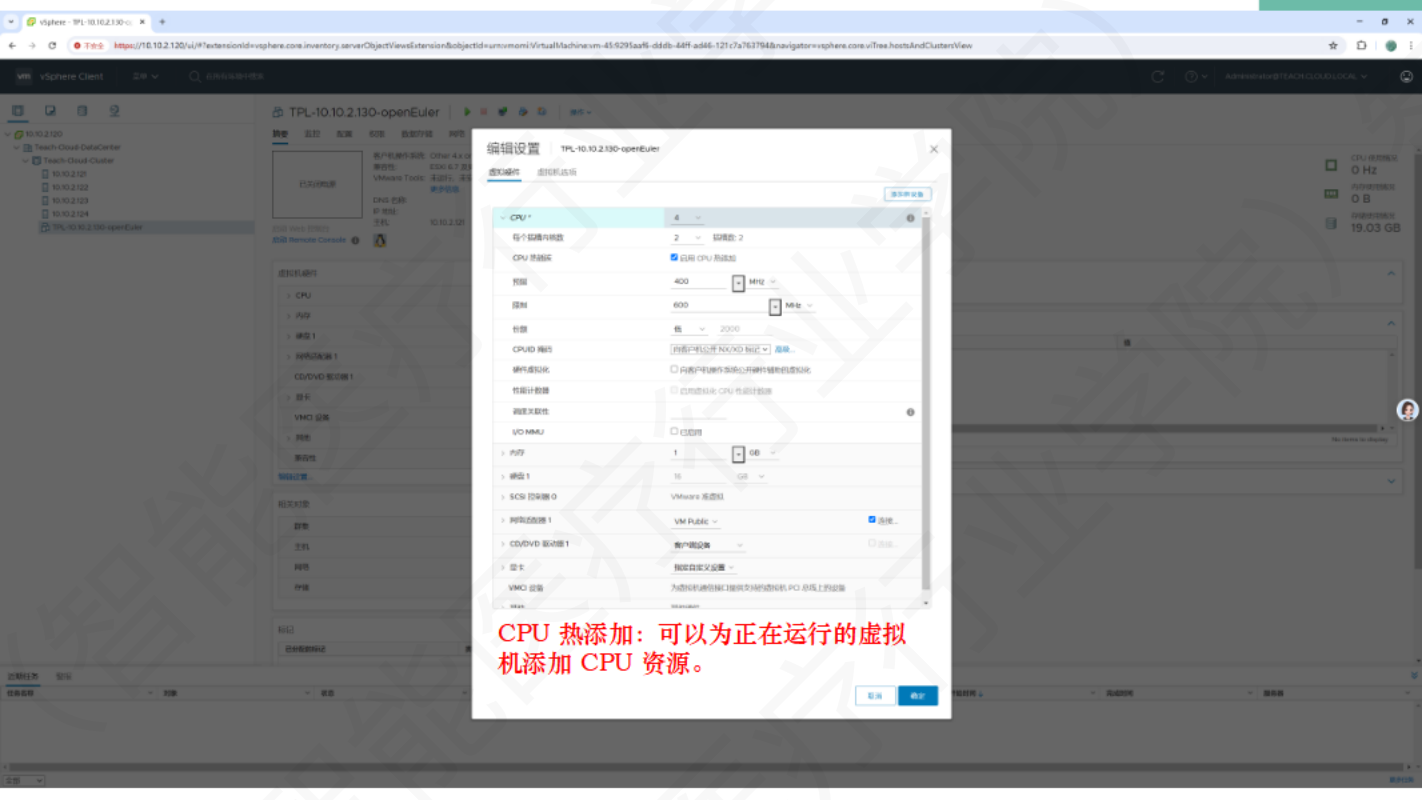
1. 虚拟机的资源管理

1.4 虚拟机计算资源的管理

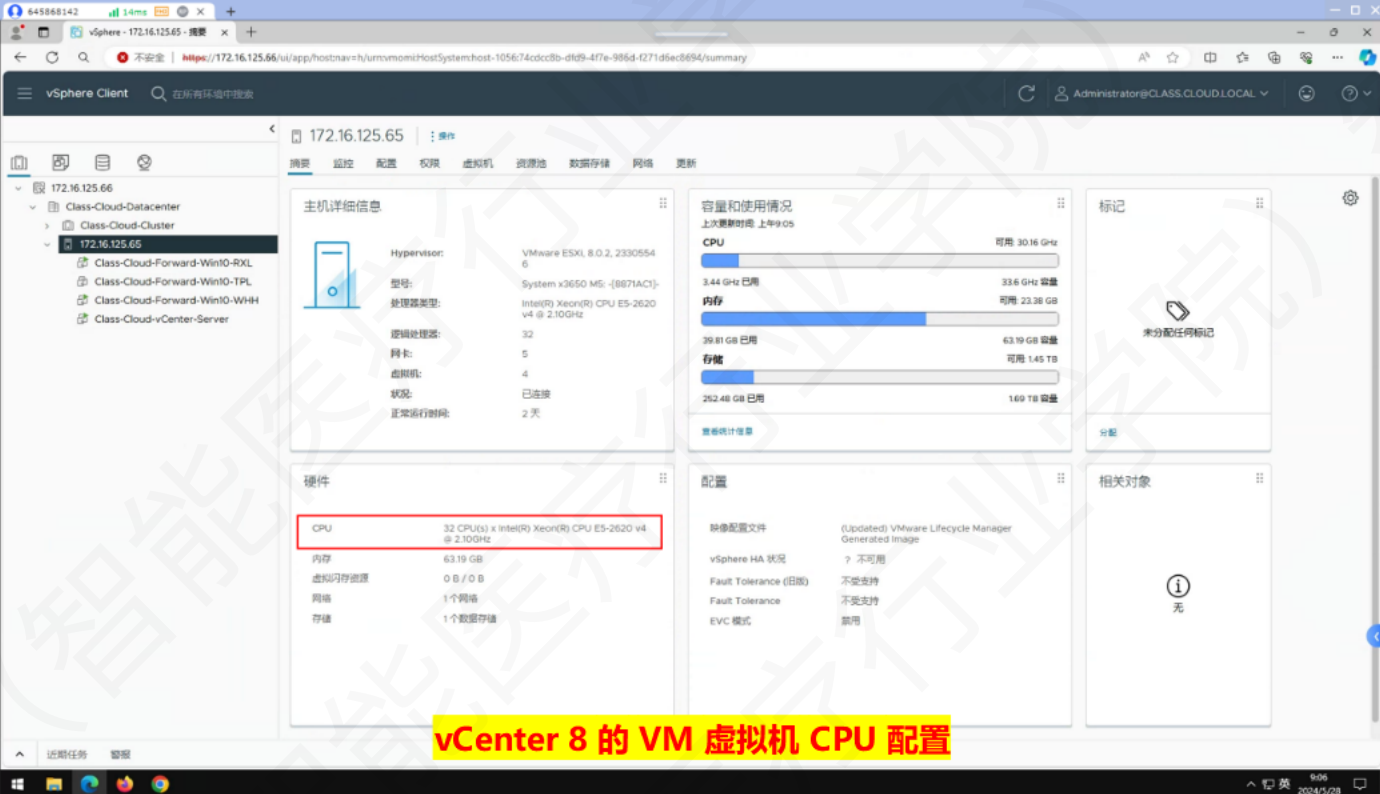
□ 讨论:

- ESXi Host 配置 2 颗 CPU，每个 CPU 有 24 个核心，主频 2.5GHz。
 - （假设 ESXi Host 不占用CPU，所有 CPU 的资源可以全部用于 VM。）
- 问题：
 - 用于 DHCP 服务器的 VM 需要配置几个 CPU？几个插槽？
 - 用于 Web 服务器的 VM 需要配置几个 CPU？几个插槽？
 - 第 1 台 VM 分配了 8 个 CPU，2 个插槽，是不是对应到两个物理 CPU 上？
 - 已经创建 10 台 VM，每台 VM 分配 6 CPU，这些 VM 是否可以启动？
 - 是否可以为 VM 配置超过物理核心总量的 CPU？超过物理 CPU 数量的插槽？



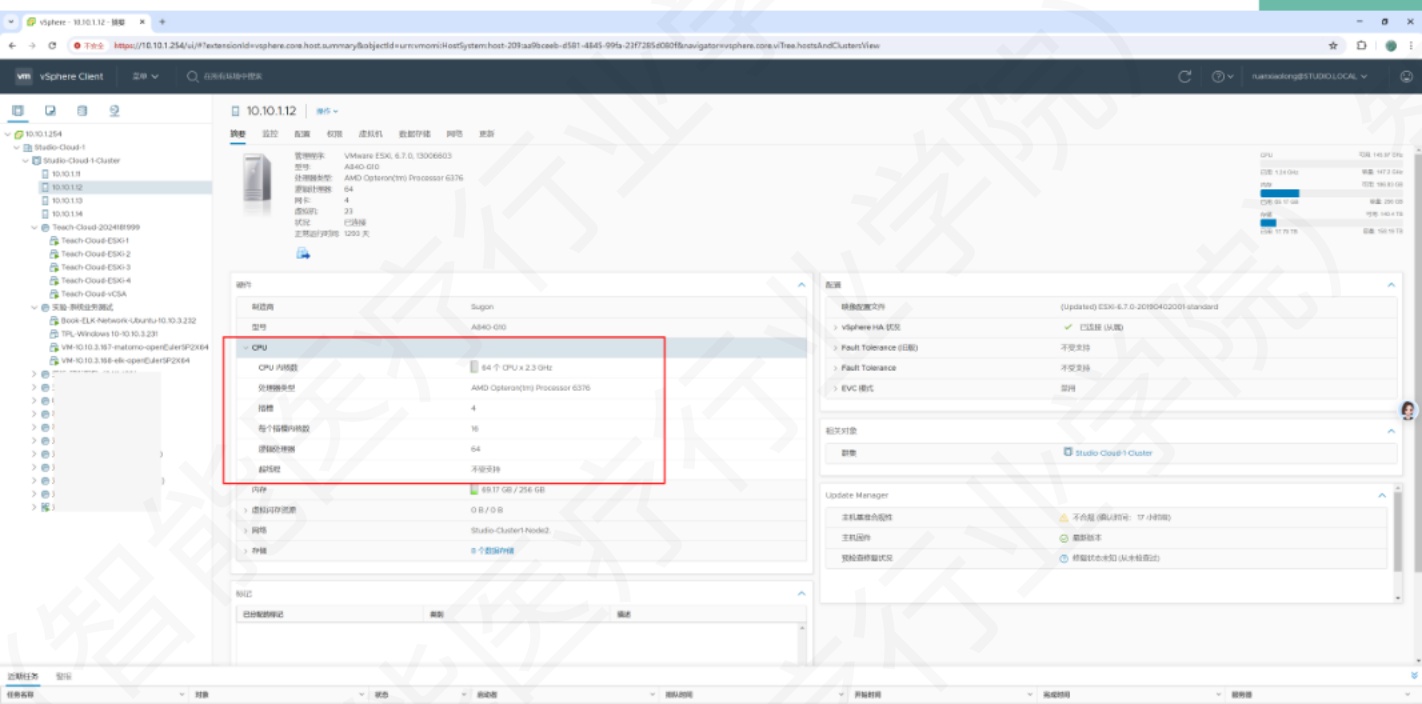


CPU 热添加：可以为正在运行的虚拟机添加 CPU 资源。

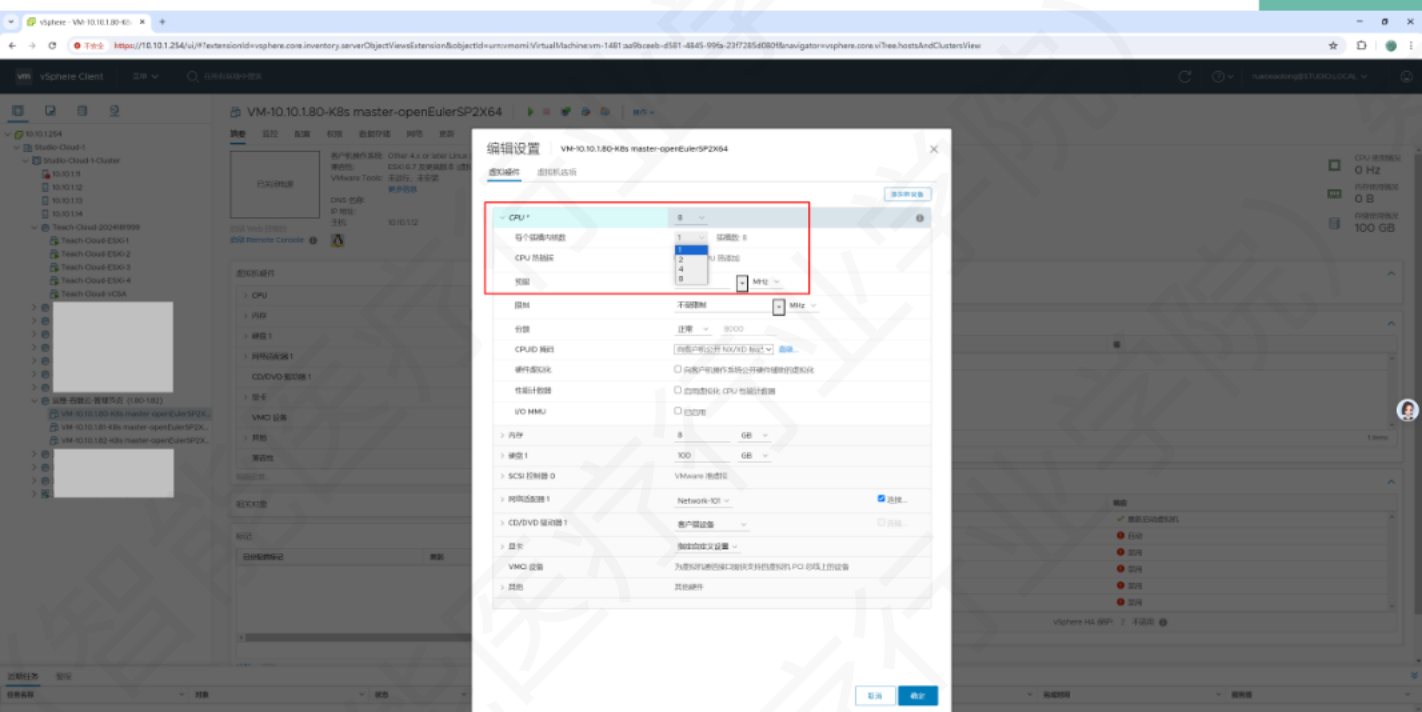




vCenter 8 的 VM 虚拟机 CPU 配置



vCenter 6.7 的 VM 虚拟机 CPU 配置



vCenter 6.7 的 VM 虚拟机 (Linux) CPU 配置

1. 虚拟机的资源管理

1.4 虚拟机计算资源的管理

□ 计算资源的预留、限制、份额

■ 控制 CPU 资源的分配时：

□ 预留：

- 为 VM 提供有保证的处理能力。
- ESXi Host 必须有足够的实际物理 CPU 资源满足 VM 预留，如不能满足，则无法打开 VM 电源。

□ 限制：

- 阻止 VM 访问额外的 CPU 周期，即使有足够的 CPU 资源也无法使用。

□ 共享：

- 确定 ESXi Host 在争用时 CPU 资源分配的优先级。
- 以百分比为基础授予 CPU 访问权限，优先级由份额决定。



2. 资源池与 vAPP

□ 什么是资源池？（Resource pools）

- 资源池是一种逻辑容器，是灵活管理资源的逻辑抽象。
- 资源池可以分组为层次结构，对可用的 CPU 和内存资源按层次结构进行分区。
 - ESXi Host 和 Cluster 都具有一个（不可见的）根资源池，对资源进行分组。
 - 处于较高级别的资源池称为父资源池，可以创建根资源池的子资源池，也可以创建子资源池的子资源池。
 - 每个子资源池都拥有部分父级资源，也可以具有各自的子资源池层次结构，每个层次结构代表更小部分的计算容量。
 - 一个资源池可包含多个子资源池和虚拟机。
 - 处于同一级别的资源池和虚拟机称为同级。



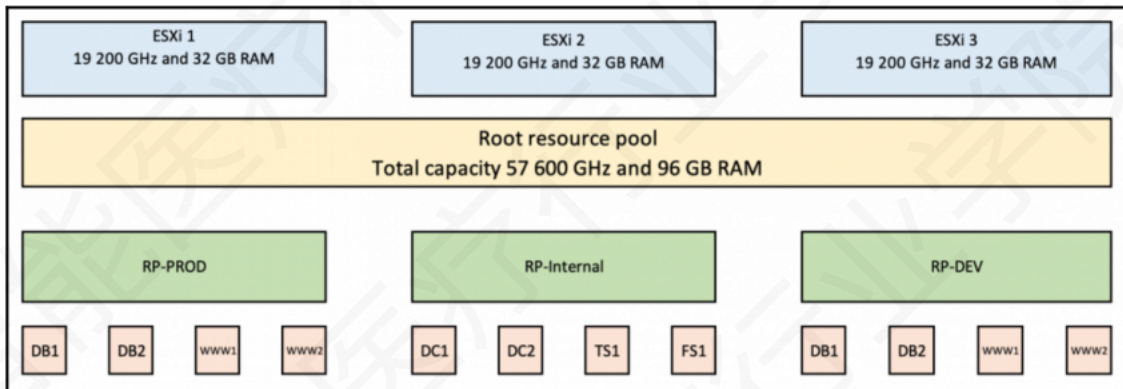
2. 资源池与 vAPP

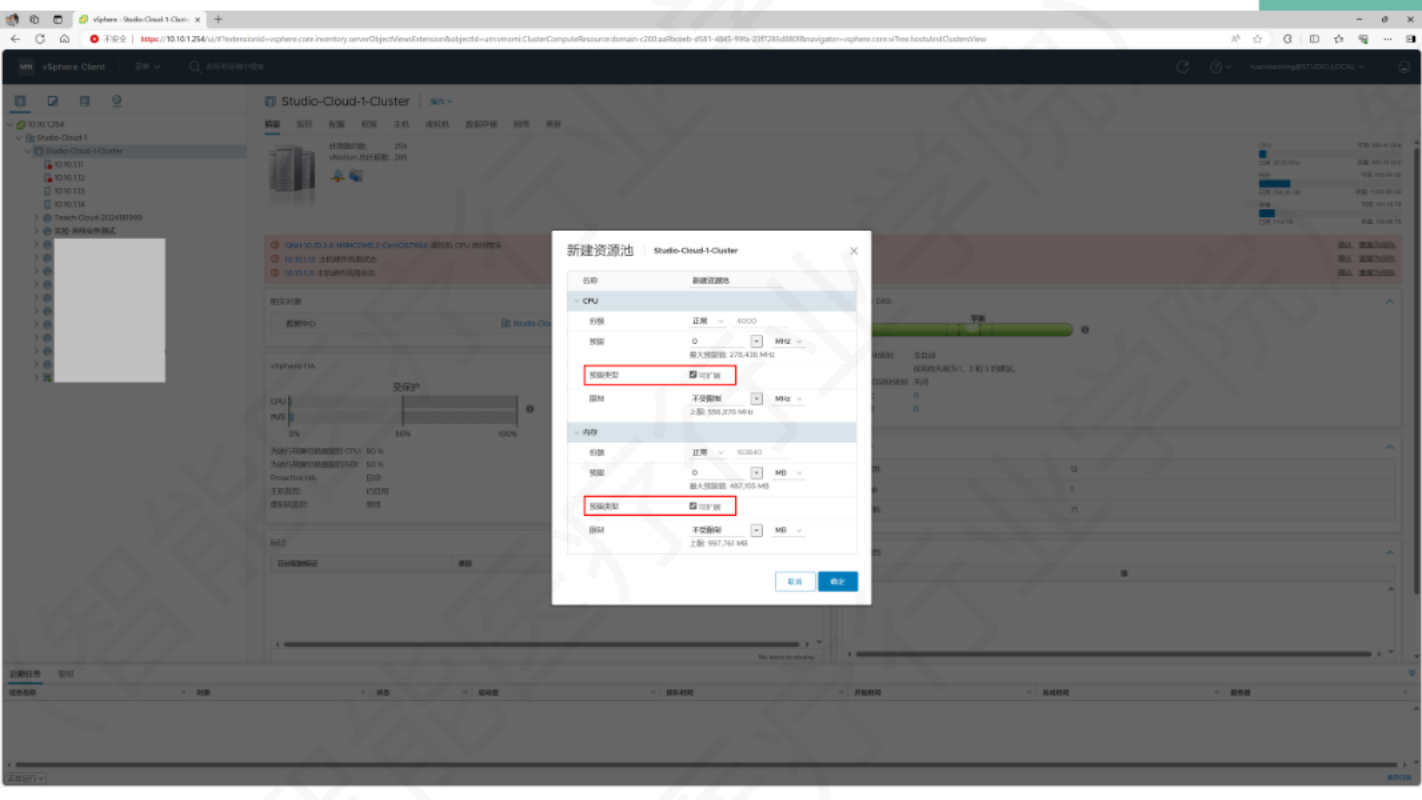
□ 资源池可以做什么？

- 通过资源池可以对 ESXi Host 或 Cluster 的资源进行集中分配与管理权限。
- 可以创建多个资源池作为 ESXi Host 或 Cluster 的直接子级，并进行配置，然后便可向其他用户分配对资源池的控制权。
- 使用资源池具有下列优点。
 - 灵活的层次结构组织。根据需要添加、移除或重组资源池，或者更改资源分配。
 - 资源池之间相互隔离，资源池内部相互共享。顶级管理员可向部门级管理员提供一个资源池。某部门资源池内部的资源分配变化不会对其他不相关的资源池造成不公平的影响。
 - 访问控制和委派。
 - 资源与硬件的分离。
 - 管理运行多层服务的各组虚拟机。



资源池的分层次结构的管理





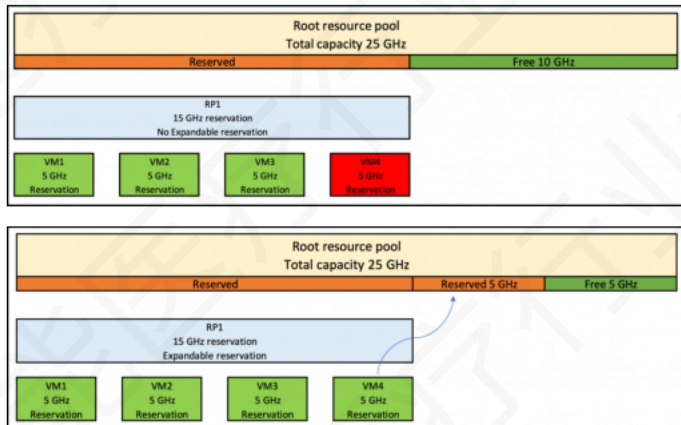
2. 资源池与 vAPP

2.1 资源池

□ 可扩展资源池

■ 可扩展选项与资源配置的预留相关。

□ 当资源池内的 VM 对预留资源总和超过资源池的总量时，可以进行扩展。



扩展资源池

2. 资源池与 vAPP

□ 对资源池的操作

■ 管理资源池

- 编辑资源池
- 删除资源池
- 添加VM
- 删除VM



2. 资源池与 vAPP

□ 什么是 vApp？（应用程序容器）

- vAPP 是一种类似于资源池的容器，可以包含一个或多个 VM。
- vAPP 可与虚拟机共享某些功能，如 vApp 电源管理和克隆等。
- 举个例子：

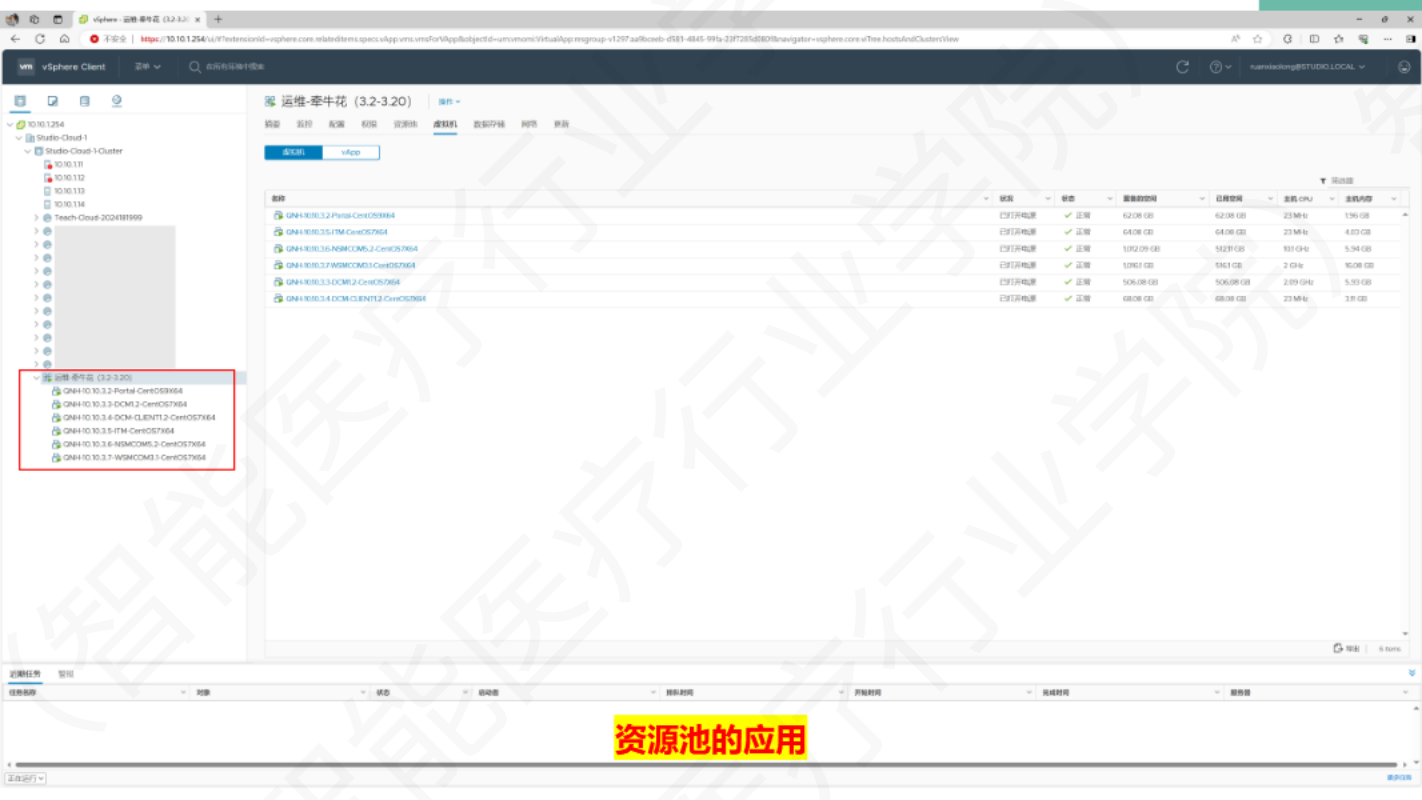
□ 某关键业务系统包含了4个 VM，分别是：

- 应用程序 VM：部署了业务系统的软件
- 数据库管理系统 VM：部署了 RDBMS
- 文件服务 VM：用于存储业务系统中的所有软件
- 用户认证 VM：提供统一认证和对外接口服务

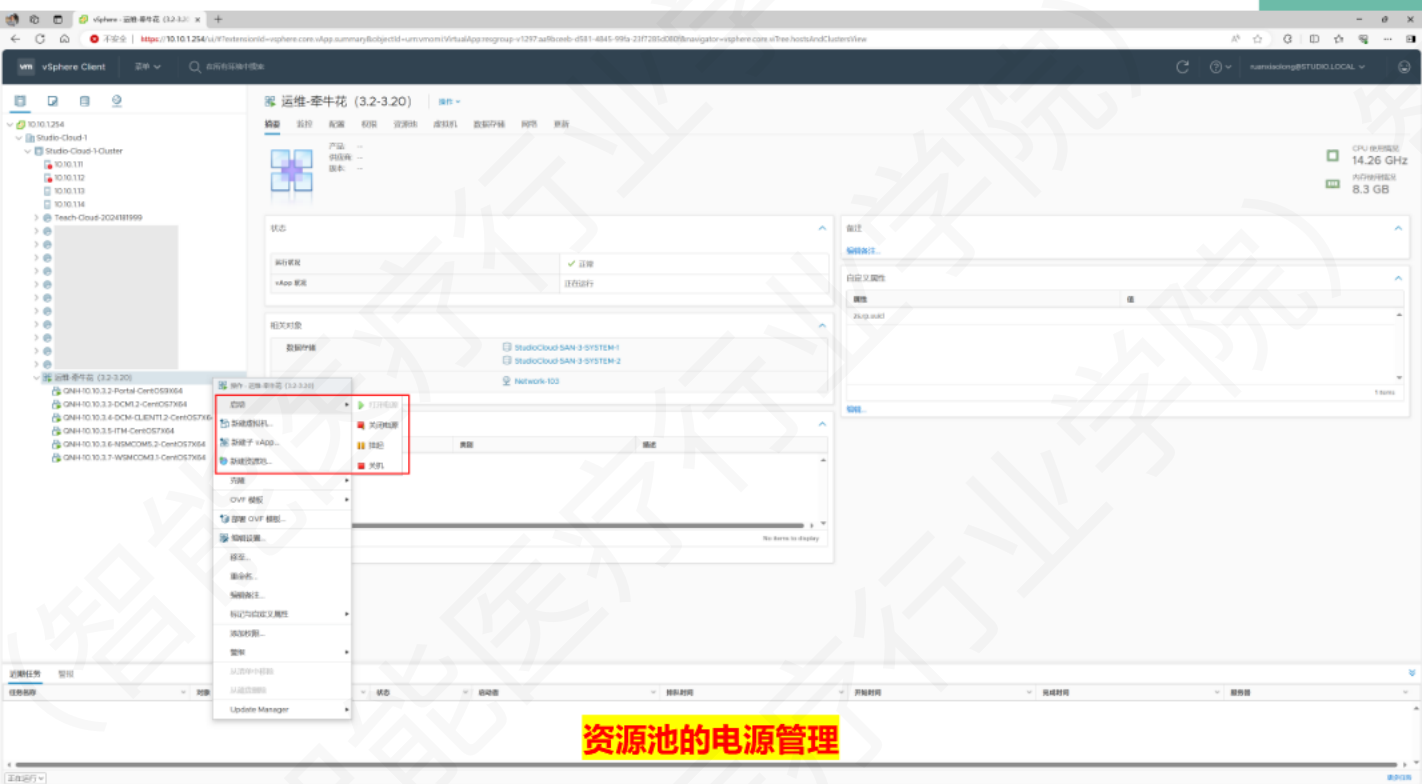
□ 可以将此4台 VM 放到一个 vAPP 中，进行统一管理。

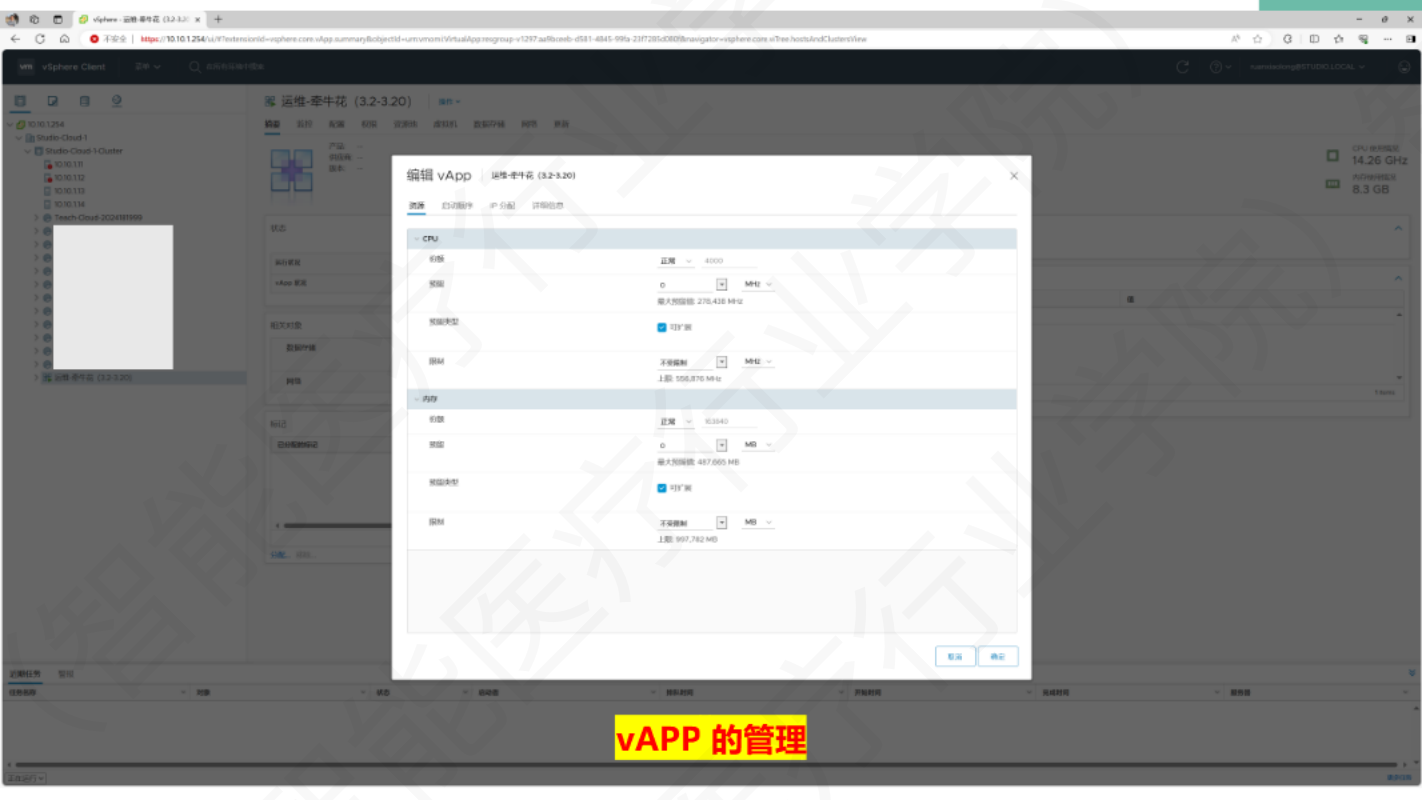
- 设置关闭和开启电源的顺序
- 设置统一的数据备份
- 设置统一的运维权限管理





资源池的应用





vAPP 的管理

2. 资源池与 vAPP

2.3 资源池与 vAPP 的对比

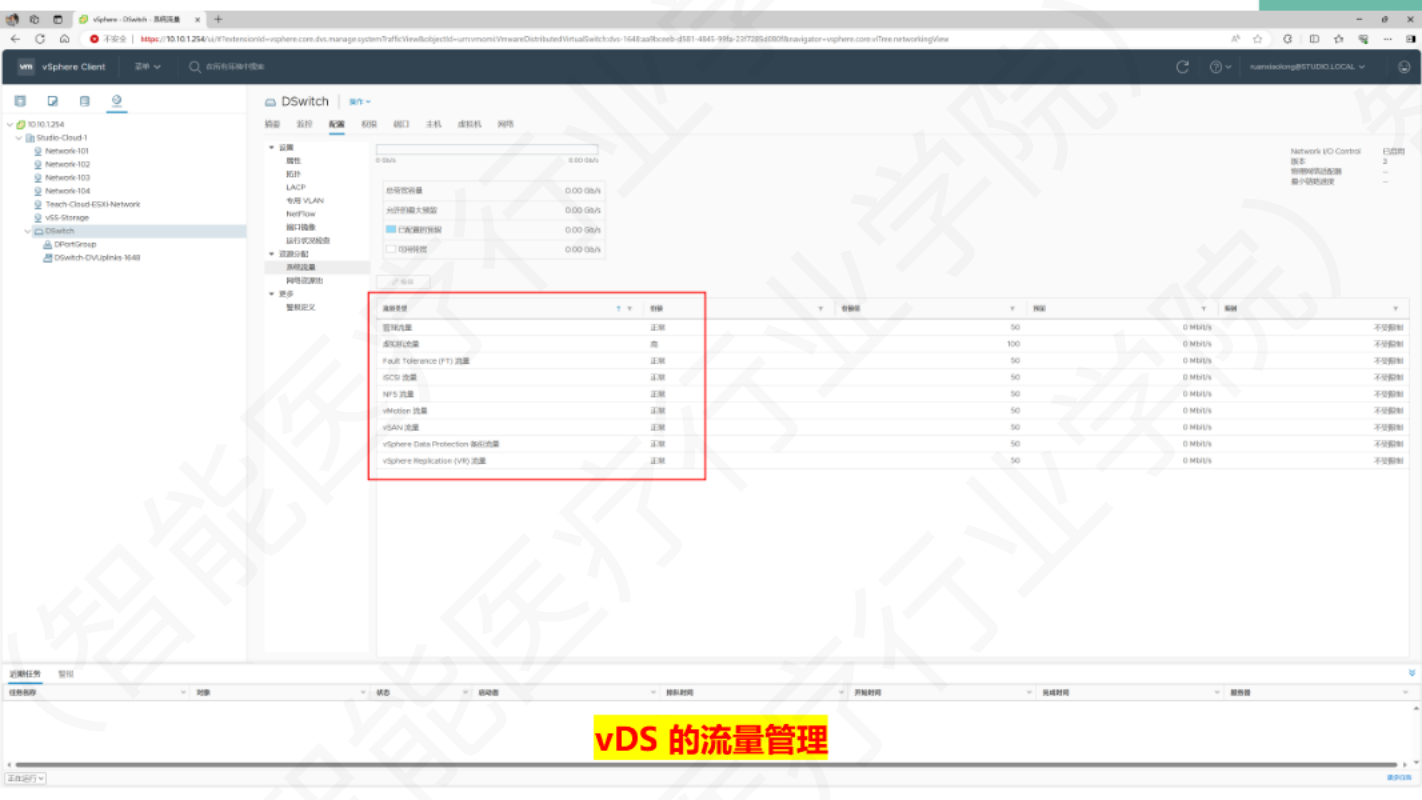
资源池的分层次结构的管理

比较项目	资源池	vApp
目的用途	对资源进行层次化分组管理，分配资源给虚拟机。	将一组相关虚拟机作为整体进行管理和部署，合并为统一的应用或业务。
层次结构	处于主机或群集层次。	通常在资源池内创建。
资源分配	根据份额、预留和限制分配资源给虚拟机。	除考虑内部虚拟机资源需求，还可进行针对整个 vApp 的资源配置。
部署迁移	无	可以作为一个整体进行部署、迁移、备份等，适合以业务为单位整体管理。

3. 网络和存储的资源管理

- vSphere 对网络资源管理功能是 vSphere Network I/O (NIOC) ，是基于 vDS 实现的。
 - 当 vDS 启用 NIOC 时，会激活九个预定义的网络资源池进行设置。
 - 管理流量
 - 虚拟机流量
 - NFS 流量
 - vSAN 流量
 - iSCSI 流量
 - vMotion 流量
 - Fault Tolerance (FT) 流量
 - vSphere Data Protection 备份流量
 - vSphere Replication (VR) 流量





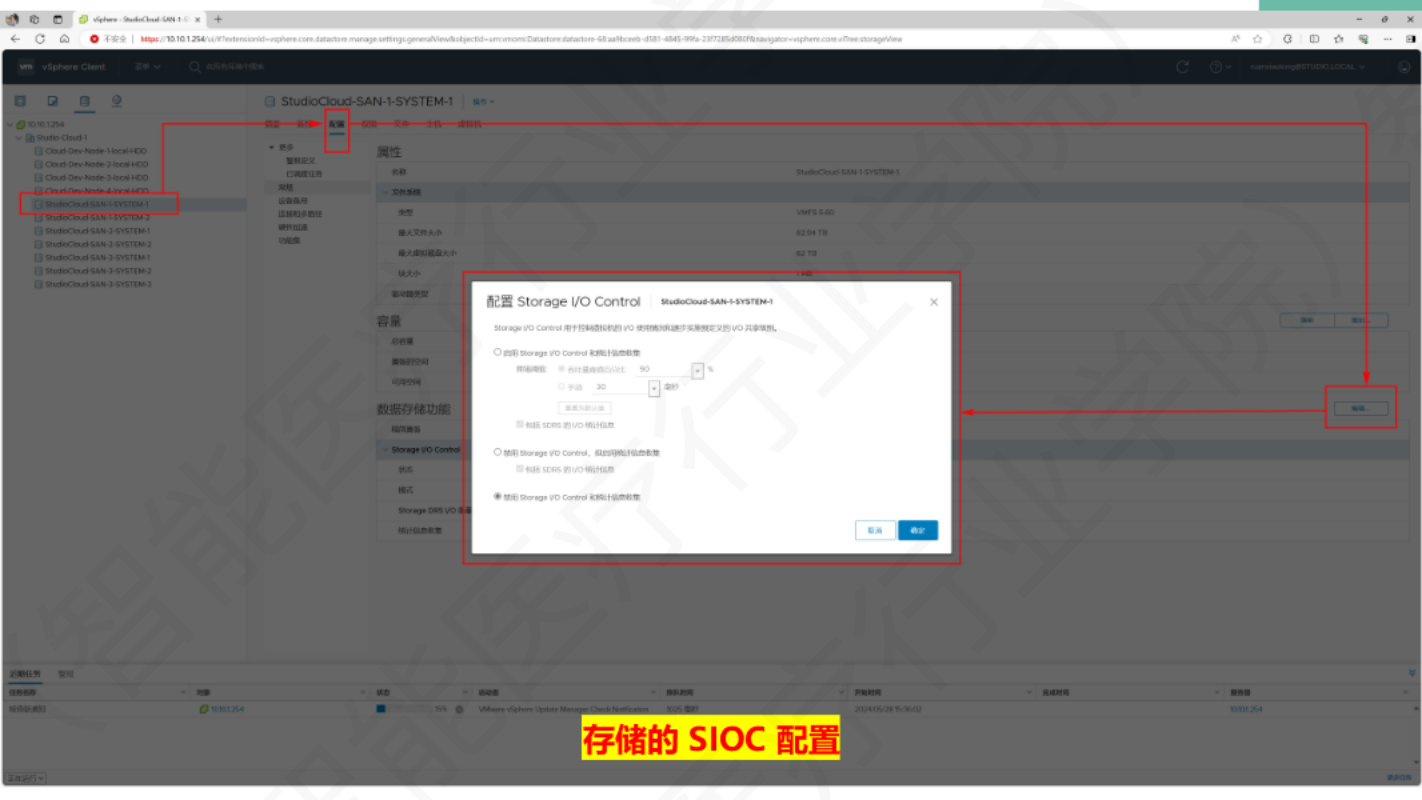
vDS 的流量管理

3. 网络和存储的资源管理

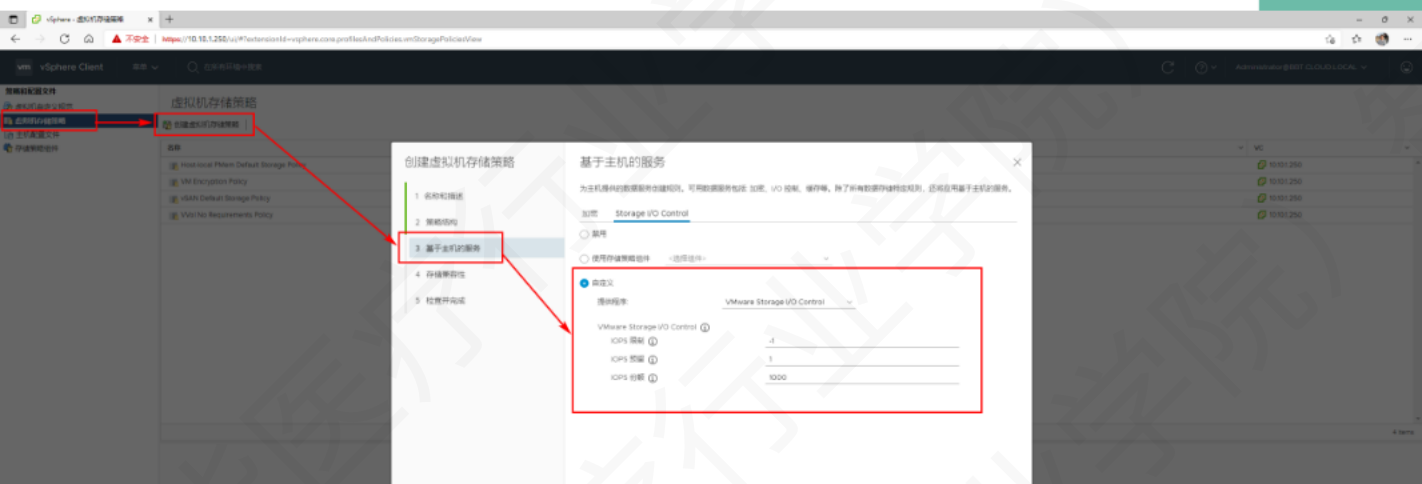
❑ vSphere 对存储资源管理功能分两个层面

- 存储设备层面是 Storage I/O Control (SIOC) 。
 - ❑ SIOC 基于每个数据存储启用，通常数据存储的 SIOC 默认是禁用的。
 - ❑ SIOC 确定存储争用的衡量指标是：延迟、峰值吞吐量。
 - 可以使用峰值吞吐量百分比作为阈值判断是否出现拥堵。
- 虚拟机层面是：份额、限制 IOPs、虚拟闪存读取缓存。
 - ❑ 份额：和 CPU、MEM 一致，可以确定在存储 IO 拥堵时的优先级。
 - ❑ 限制 IOPs：
 - 对 VM 可能生成的 IOPS 数设置限制。
 - 默认情况下，此值是无限的。
 - ❑ 虚拟闪存读取缓存：
 - 将 ESXi Host 的闪存设备用作 VM 的缓存，提升 VM 的性能。
 - 不设置不影响 VM 的使用。





存储的 SIOC 配置



Limits 限制

是对资源的硬性上限。

VM（或虚拟磁盘）将不会消耗超过限制所配置的资源量。

Reservations 预留

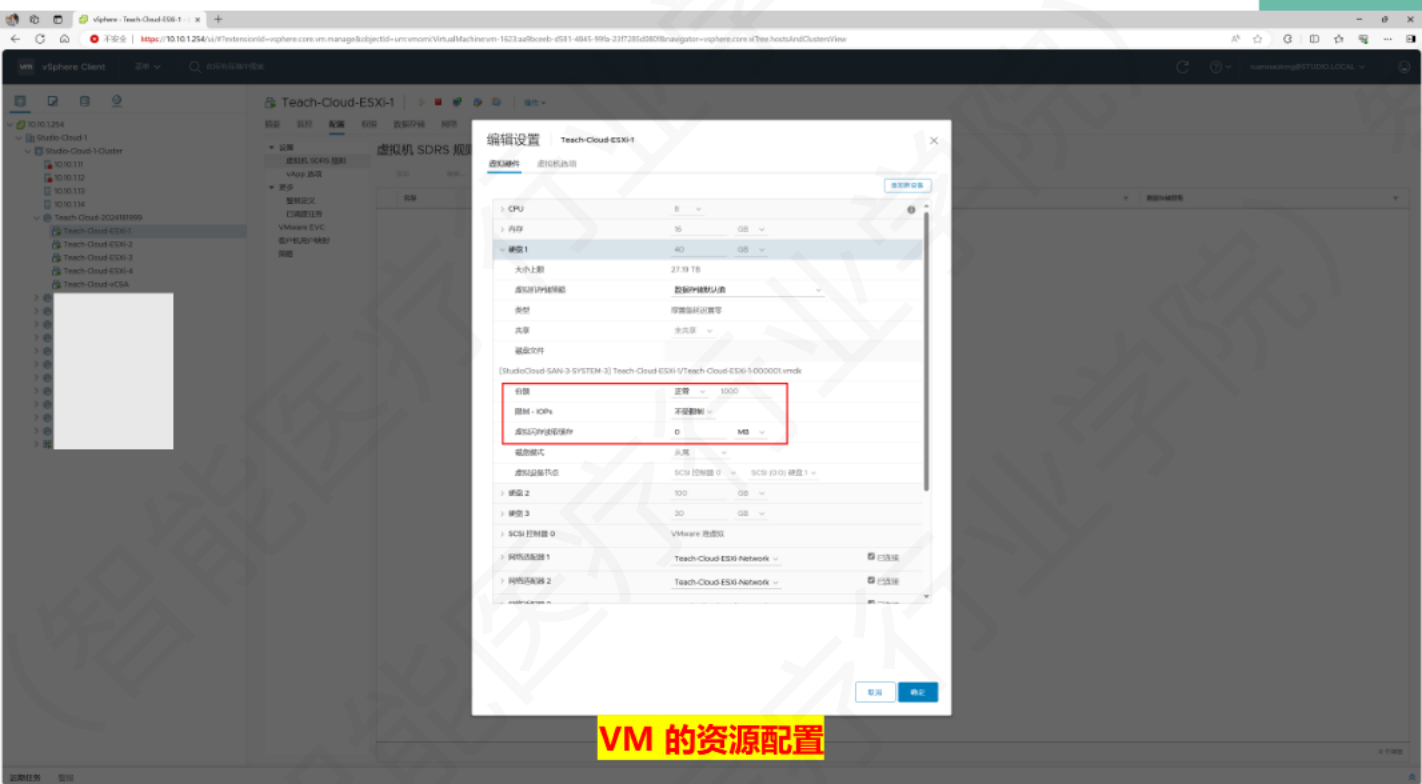
是 VM（或在这种情况下是虚拟磁盘）始终可用的资源数量。

如果未使用预留的资源，剩余的资源将被用于满足其他 VM 的需求。一旦 VM 要访问预留资源，即使其他 VM 正在使用，也会保证其可用性。

Shares 份额

只在发生拥塞时才起作用。

当 VM 想要访问的资源超过存储所能提供的资源时，就会出现这种情况。在这种情况下，每个 VM 将获得一定数量的资源份额。



编辑设置 Teach-Cloud-ESXi-1

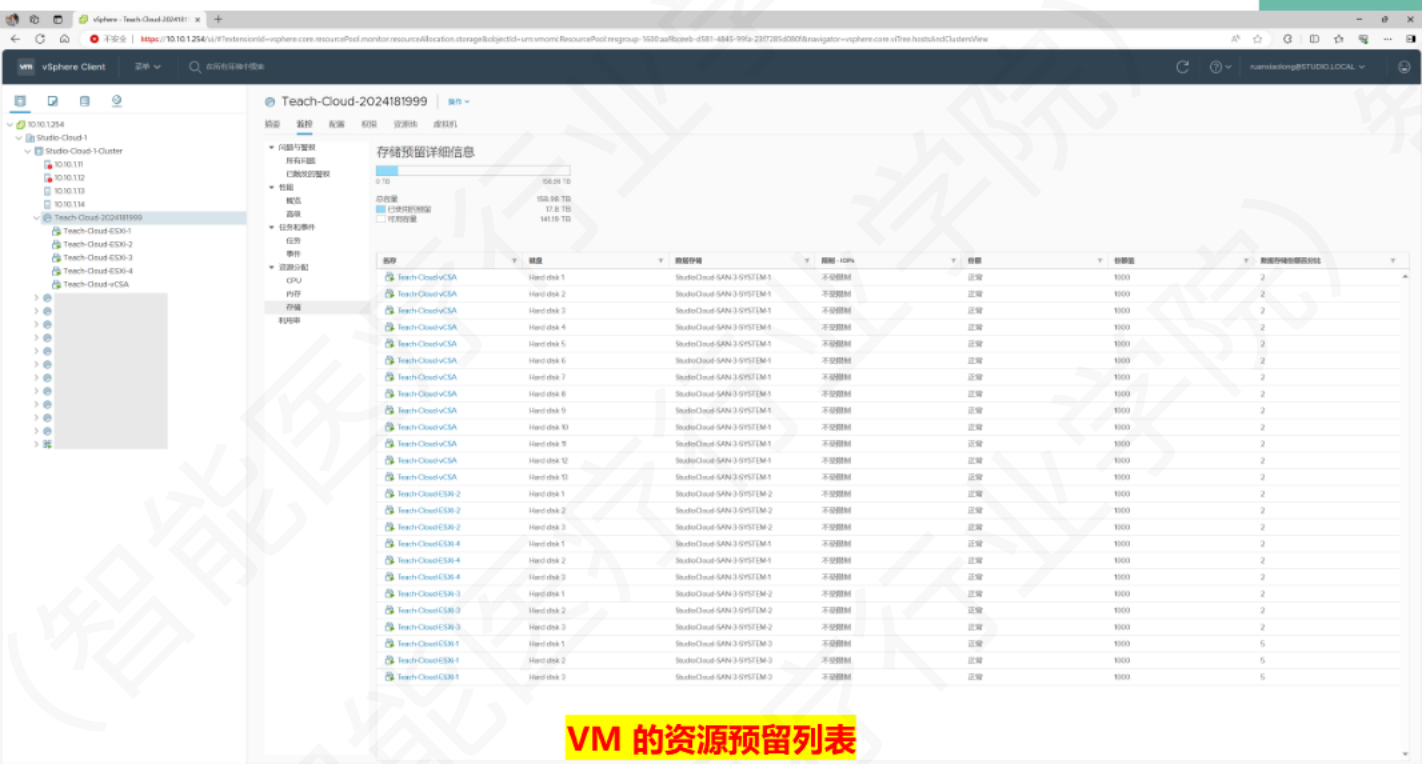
虚拟机设置

保存设置

CPU	16	GB
内存	16	GB
硬盘 1	27.19 TB	
大小上限	27.19 TB	
虚拟机内存策略	正常	
类型	厚置备延迟置零	
共享	全共享	
磁盘文件	[StudioCloud-SAN-3-SYSTEM-3] Teach-Cloud-ESXi-1/Teach-Cloud-ESXi-1-000001.vmdk	
份额	正常	1000
限制 - IOPS	不受限制	
虚拟机内存保留	0	MB
磁盘格式	直通	
虚拟机设备节点	SCSI 控制器 0	SCSI (0:0) 硬盘 1
硬盘 2	100	GB
硬盘 3	20	GB
SCSI 控制器 0	VMware 直通	
网络适配器 1	Teach-Cloud-ESXi-Network	已连接
网络适配器 2	Teach-Cloud-ESXi-Network	已连接

取消 确定

VM 的资源配置



vSphere 对资源管理途径与措施

计算

VM
CPU + MEM
预留、限制、份额

存储

存储设备
SIOC
延迟、峰值吞吐量

VM 磁盘
份额、IOPs
虚拟闪存读取缓存

网络

vDS
NIOC

4. 虚拟机迁移

4.1 资源分配和资源利用率

对资源相关的两种描述

分配
Allocation

分配是指如何分配资源，CPU、内存、存储 I/O 和网络带宽指派给 VM 的过程。

利用
Utilization

利用是指在 ESXi Host 将资源分配给 VM 后，资源是如何使用及使用情况。（利用率）

4. 虚拟机迁移

4.1 资源分配和资源利用率

- vSphere 通过预留、限制、份额三种机制进行资源分配。
- vSphere 在调度和平衡资源应用上采用五种措施进行：
 - vMotion（实时迁移）：手动平衡 ESXi 主机和群集之间的计算资源利用率。
 - Storage vMotion：手动平衡数据存储之间的存储利用率。
 - Cross-vCenter vMotion：
 - 用于在 vSphere 环境之间（如数据中心之间）迁移工作负载，实现数据中心之间的利用率平衡。
 - vSphere Distributed Resource Scheduler：
 - 自动平衡 vSphere 群集中 ESXi Host 之间的计算资源利用率。
 - Storage DRS：
 - SDRS 是 DRS 在存储上的应用，用于在创建 VM 文件之前和之后，自动平衡数据存储群集中多个数据存储之间的存储利用率。

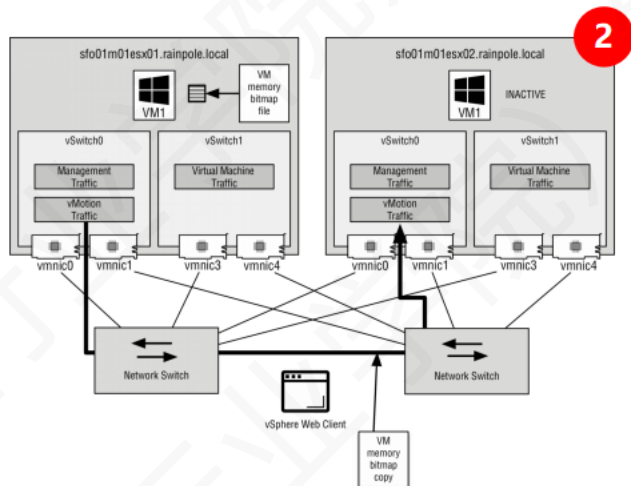
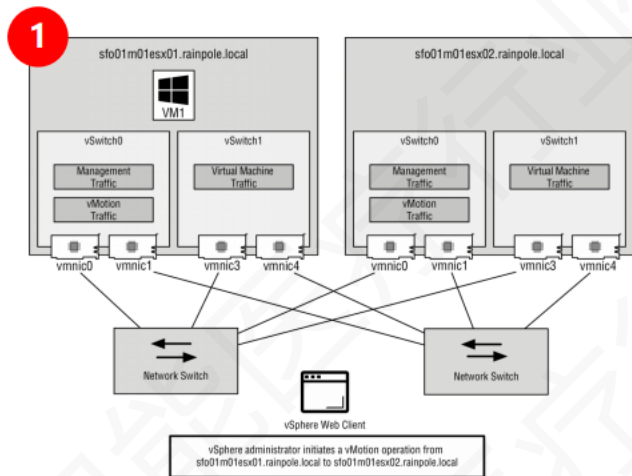
4. 虚拟机迁移

□ vMotion

- 用于在两个 ESXi Host 之间手动平衡计算资源利用率。
- vMotion 可以将已打开电源的 VM 从一台 ESXi Host 实时迁移到另一台 ESXi Host，且不会中断服务，是零停机的迁移操作。
 - 网络连接不会断开，网络连接的会话不会中断。
 - 应用程序或服务不间断地运行。
- vMotion 用于手动对 ESXi Host 进行负载平衡，通过迁移 VM 消除数据中心内高负载 ESXi Host，提升资源平衡和 VM 性能。



vMotion 的工作原理



管理员进行操作:

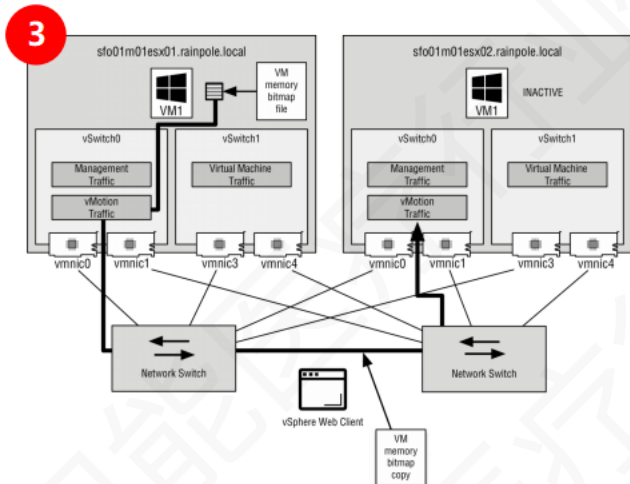
- ESXi Host 2台, VM1 已经开机。
- VM1 需从 ESXi Host 01 迁移到 ESXi Host 02。

后续使用设备名称: VM1、ESXi01、ESXi02

预复制阶段, 通过 vMotion 网络进行迁移:

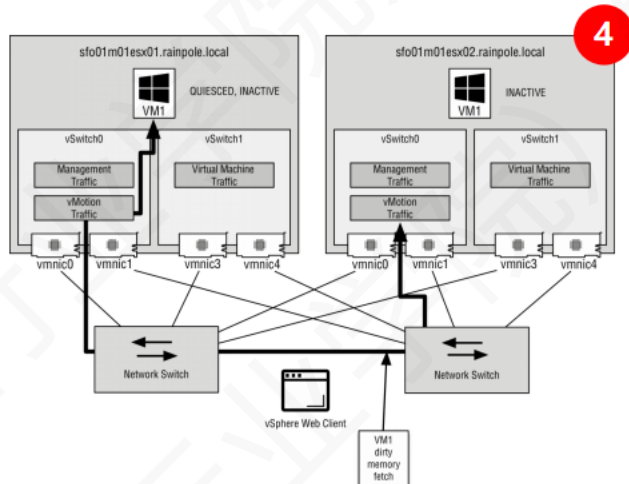
- VM1 在 ESXi01 上的活动内存页复制到 ESXi02。
- VM1 在 ESXi01 上持续提供服务, 其内存页仍可改变。
- ESXi01 记录内存页复制到 ESXi02 后的 VM1 的内存变化信息, 称为内存位图 (VM Memory Bitmap File)。

vMotion 的工作原理



VM1 的活动内存页全部传送到 ESXi02 后:

- ESXi02 读取到内存位图文件的地址。
- ESXi02 请求获取内存位图文件。

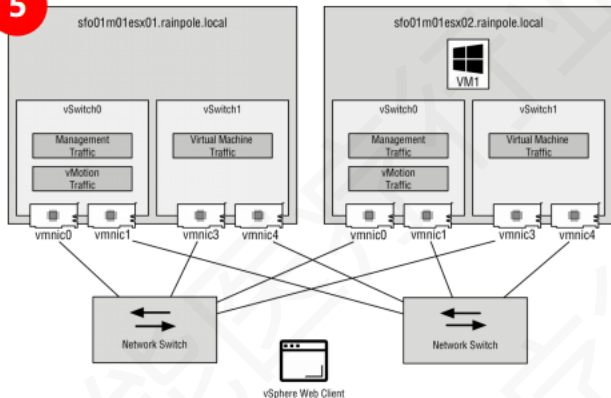


ESXi02 获取到内存位图文件地址后:

- ESXi01 上的 VM1 状态变为静默状态，但不删除，也不再提供服务。
- ESXi01 上的 VM1 的内存位图文件传送到 ESXi02。

vMotion 的工作原理

5



内存位图文件传送到 ESXi02 后:

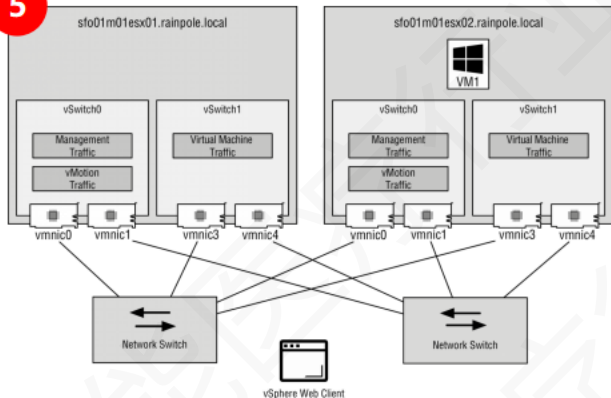
- VM1 在 ESXi02 上启动, 注意不是重启 VM1, 因为 VM1 本来就在 ESXi02 的内存中。
- ESXi02 发送反向地址解析 (RAPR) 消息, 让交换机更改 VM1 的 IP 地址对应的 MAC Address。(转发表)
- 交换机会将发往 VM1 的所有请求, 目的 MAC Address 改为新的地址, VM1 的业务恢复。

VM1 在 ESXi02 上提供服务后:

- 删除 VM1 在 ESXi01 上的内存。

vMotion 的工作原理

5



内存位图文件传送到 ESXi02 后:

- VM1 在 ESXi02 上启动, 注意不是重启 VM1, 因为 VM1 本来就在 ESXi02 的内存中。
- ESXi02 发送反向地址解析 (RAPR) 消息, 让交换机更改 VM1 的 IP 地址对应的 MAC Address。(转发表)
- 交换机会将发往 VM1 的所有请求, 目的 MAC Address 改为新的地址, VM1 的业务恢复。

VM1 在 ESXi02 上提供服务后:

- 删除 VM1 在 ESXi01 上的内存。

业务不是不中断, 而是中断的时间非常短,
短于应用服务可以接受的长度。

高可用就是时间的游戏。

4. 虚拟机迁移

□ vMotion

- 使用 vMotion 实时迁移虚拟机的易用性非常高，是非常领先的技术。
- 启用 vMotion 需要一定的条件：
 - 依赖 vCenter Server 实现。
 - 参与 vMotion 的所有 ESXi Host 都必须获得 vMotion 的授权。
 - VM 对应的虚拟机文件必须存放在共享存储上，且迁移的源 ESXi Host 和目标 ESXi Host 均能够访问到共享存储。
 - ESXi Host 在网络接口卡（NIC）上需要启用为 vMotion 流量定义和启用 VMkernel 端口，并建议不低于 1 Gb E。



4. 虚拟机迁移

□ vMotion

- 使用 vMotion 实时迁移虚拟机的易用性非常高，是非常领先的技术。
- 启用 vMotion 必须满足的条件：
 - 基于 vCenter Server 实现，所有 ESXi Host 都必须获得 vMotion 的授权。
 - 参与迁移的源 ESXi Host 和目标 ESXi Host 的 CPU 必须兼容。
 - 建议使用相同型号的 CPU，至少应是同一 CPU 系列，保证 CPU 指令集的一致或兼容。
 - CPU 必须支持相同的功能，例如支持 SSE2、SSE3 和 SSE4 以及 NX 或 XD 等。
 - 迁移 64 位的 VM，ESXi Host 的 CPU 必须启用虚拟化技术（Intel VT 或 AMD-V）。
 - VM 对应的虚拟机文件必须存放在共享存储上，参与迁移的源 ESXi Host 和目标 ESXi Host 均能够访问到共享存储。
 - ESXi Host 在网络接口卡（NIC）上需要启用为 vMotion 流量定义和启用 VMkernel 端口。
 - 建议不低于 1 Gb E。
 - 建议使用专用网络用于 vMotion 通信。



4. 虚拟机迁移

□ vMotion

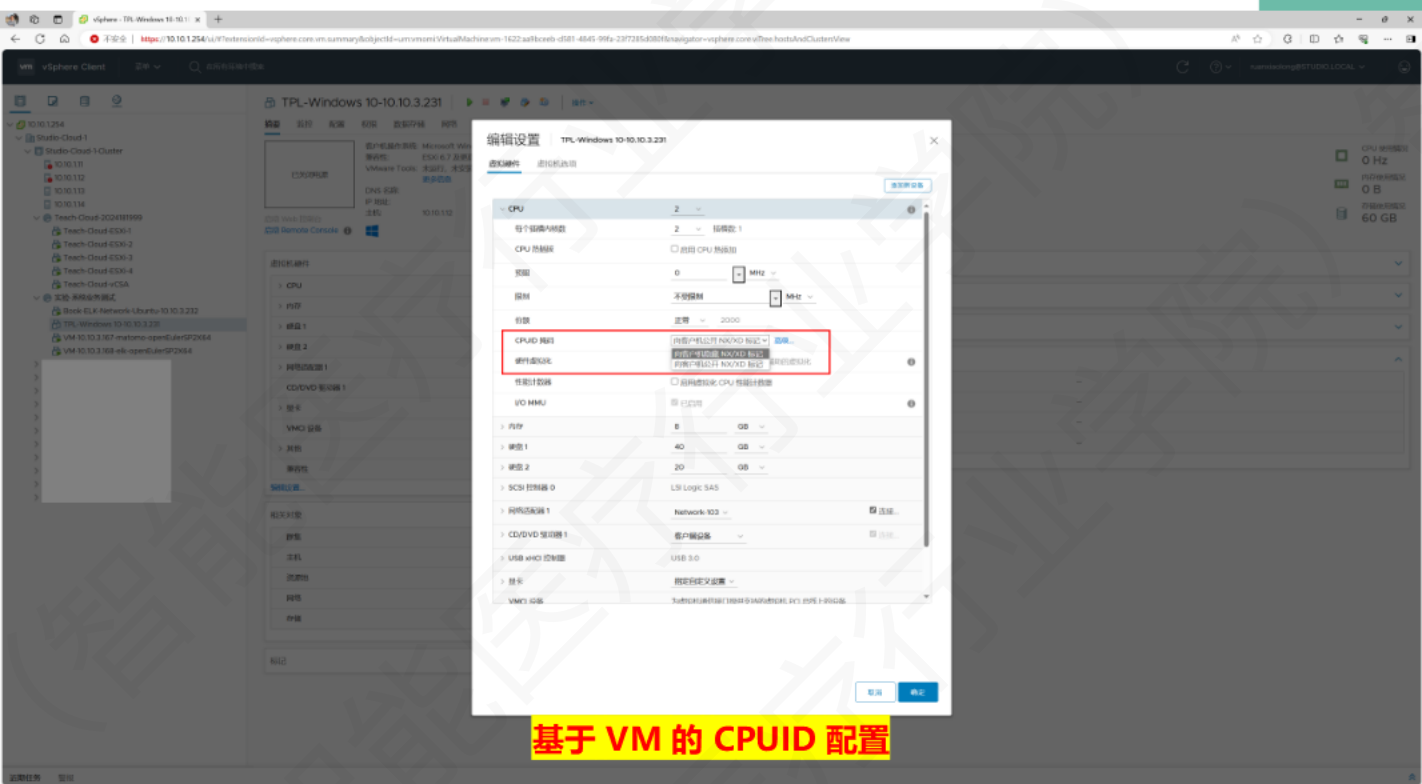
- 使用 vMotion 实时迁移虚拟机的易用性非常高，是非常领先的技术。
- 使用 vMotion 必须满足的要求：
 - VM 不得连接任何物理上仅供源主机使用的设备。
 - 包括磁盘存储和 CD/DVD 驱动器。
 - 如果 VM 具有这些映射之一，要取消选中有问题设备旁边的“已连接”复选框即可。
 - VM 不得连接到仅限内部访问的虚拟交换机。
 - VM 的 CPU 关联性不得设置为特定 CPU。
 - VM 必须将所有磁盘、配置、日志和非易失性随机存取存储器（NVRAM）文件存储在源和目标 ESXi Host 均可访问的数据存储上。
- 不满足条件时，vMotion 会报错或者发出警告，且迁移任务不能执行。



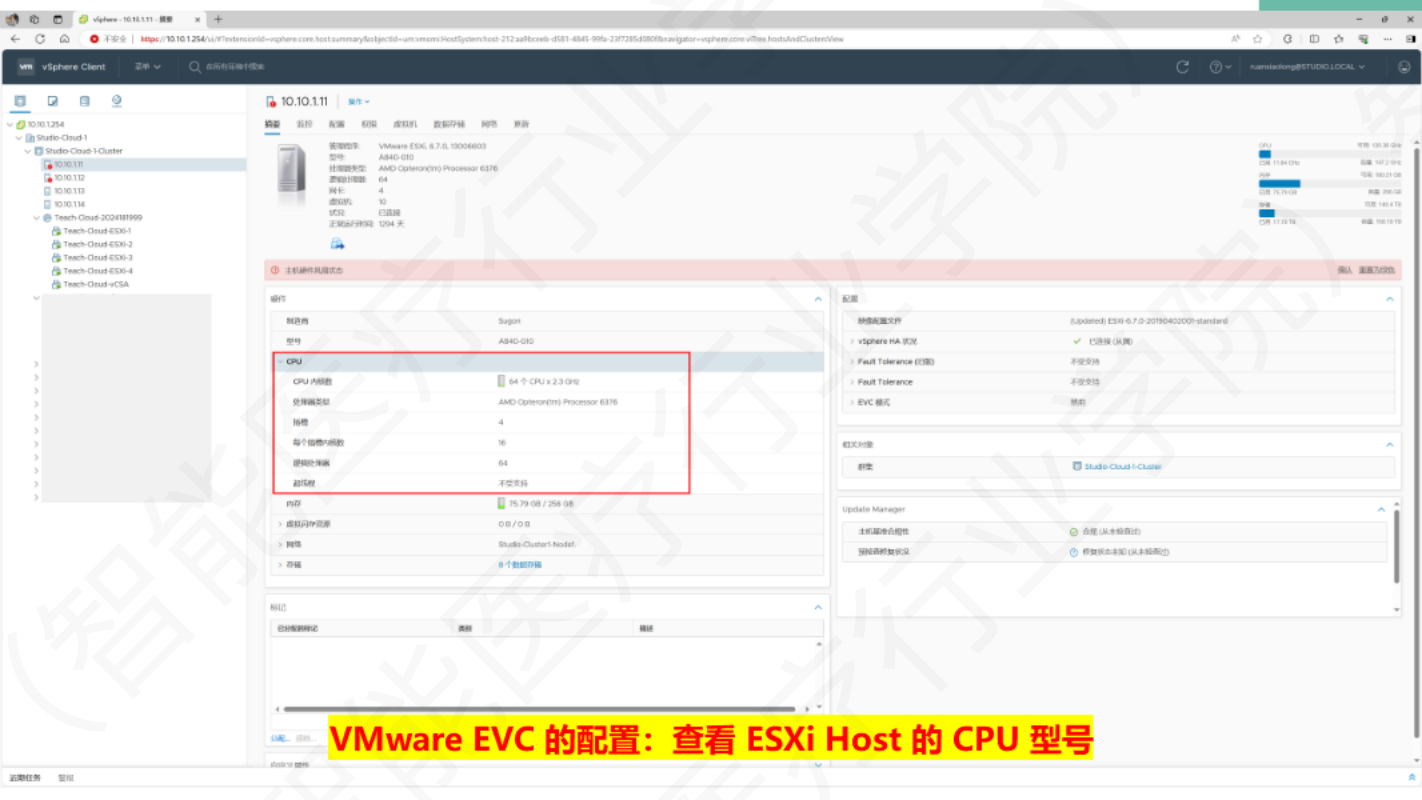
4. 虚拟机迁移

□ vMotion 的兼容性

- 使用 vMotion 对 ESXi Host 的 CPU 有先决条件的要求。
- vSphere 提供两个方案来提升 vMotion 对 CPU 的兼容性：
 - 基于每个 VM: CPUID 掩码
 - 通过屏蔽或隐藏 VM 的 CPU 功能，使虚拟机在不同的 ESXi Host 之间进行迁移，提高了 VM 的可移植性和兼容性。
 - 通过 CPUID 掩码，禁用 CPU 的某些功能，让 VM 无法使用，继而实现向其他 ESXi Host 的迁移。
 - EVC: Enhanced vMotion Compatibility, 增强的 vMotion 兼容性
 - VMware、Intel、AMD 通过合作的结果。
 - 硬件方面: Intel、AMD 在 CPU 中允许修改 CPUID 的功能，Intel 将该功能叫做 Flex 迁移，AMD 将此功能引入 AMD-V 扩展。
 - 软件方面: VMware 利用硬件功能实现集群中所有服务器的通用 CPUID 标准，称为 EVC。
 - 在虚拟层面上支持 CPUID 掩码功能，可屏蔽 CPU 功能，通过强制 CPU 指令集兼容来实现不同 CPU 系列之间的 vMotion。



基于 VM 的 CPUID 配置



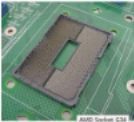
VMware EVC 的配置: 查看 ESXi Host 的 CPU 型号

CPU Database > AMD Opteron > AMD Opteron 6376 Specs

AMD Opteron 6376

Report an Error

16 CORES 16 THREADS 115 W TDP 2.3 GHz FREQUENCY 3.2 GHz BOOST Abu Dhabi CODENAME Socket G34 SOCKET

AMD Socket G34

The AMD Opteron 6376 is a server/workstation processor with 16 cores, launched in November 2012. It is part of the Opteron lineup, using the Abu Dhabi architecture with Socket G34. To further increase overall system performance, up to four Opteron 6376 CPUs can link up in a multi-processor (SMP) configuration. Opteron 6376 has 8 MB of L3 cache per die and operates at 2.3 GHz by default, but can boost up to 3.2 GHz, depending on the workload. AMD is building the Opteron 6376 on a 32 nm production process using 2,400 million transistors. The multiplier is locked on Opteron 6376, which limits its overclocking capabilities. With a TDP of 115 W, the Opteron 6376 consumes a lot of power, so good cooling is definitely needed. AMD's processor supports DDR3 memory. For communication with other components in the machine, Opteron 6376 uses a PCI-Express Gen 2 connection. This processor lacks integrated graphics, you might need a graphics card. Hardware virtualization is available on the Opteron 6376, which greatly improves virtual machine performance. Programs using Advanced Vector Extensions (AVX) will run on this processor, boosting performance for calculation-heavy applications.

Physical	Processor	Performance	Architecture
Socket: AMD Socket G34	Market: Server/Workstation	Frequency: 2.3 GHz	Codename: Abu Dhabi
Process Size: 32 nm	Production Status: unknown	Turbo Clock: up to 3.2 GHz	Generation: Opteron (Abu Dhabi)
Transistors: 2,400 million	Release Date: Nov 5th, 2012	Base Clock: 200 MHz	Memory Support: DDR3
Die Size: 315 mm²	Part#: OS6376WKTGGH-R OS6376WKTGGH-WXDF	Multiplier: 11.5x	ECC Memory: No
Package:		Multiplier Unlocked: No	PCI-Express: Gen 2
		TDP: 115 W	
Cache	Core Config	Features	
Cache L1: 768 KB	# of Cores: 16	- MMX - SSE - SSE2	
Cache L2: 16 MB	# of Threads: 16	- SSE3 - SSE3.1 - SSE4.1	
Cache L3: 8 MB (per die)	SMP # CPUs: 4	- AMD-V - AVX - AVX 5.1 - XOP	
	Integrated Graphics: N/A	- FMA3 - FMA4 - CVT16 - F16C	
Notes			

vmware by Broadcom

多云服务 产品 解决方案 合作伙伴 资源

Home / Resources / Compatibility Guides / Detail

VMware Compatibility Guide

CPU SUPPORT

CPU 系列: AMD Opteron E1xx Series
 每个插槽最大内核数: 12
 最大线程数: 12

增强型 vMotion 兼容性模式: AMD Opteron™ Generation 1
 AMD Opteron™ Generation 2
 AMD Opteron™ Generation 3
 AMD Opteron™ Generation 3 without 3D Now!

CPUID 标志: 16.9
 CPUID: D40130F90
 别名名称: "Magna Court" G34
 发布日期: 2013-03-28
 支持 SMP-PT: No
 支持超频: Yes
 增强型兼容性标志: AMD Opteron™ Generation 3

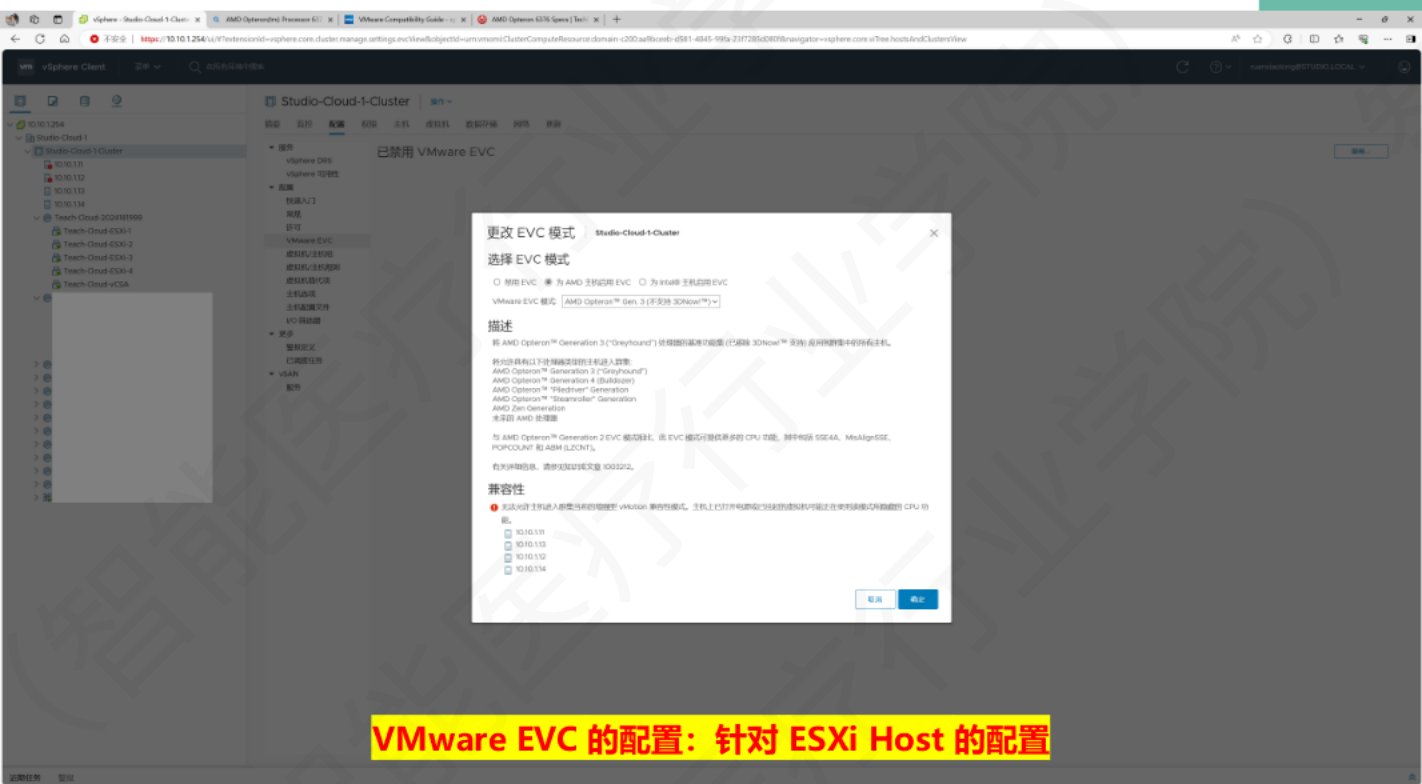
ESXi 版本: ESXi 6.5 U3
 ESXi 6.5 U2
 ESXi 6.5 U1
 ESXi 6.5

有关 EVC 结构的详细情况以及 ESX 版本支持的 EVC 模式，请参考 VMware 知识库文章 1033212。
 有关兼容性指令集的详细情况，请参考 VMware 知识库文章 1006027。
 最新固件供应商提供的 ESXi CPU 使用的所有最大内核数可能小于本页列出的值。

公司 关于我们 联系我们 支持 VMware Customer Connect 帮助 支持资源

客户案例 办公地址 合作伙伴

VMware EVC 的配置：检索了解 ESXi Host 的 CPU 型号支持的 vMotion 增强型模式



VMware EVC 的配置：针对 ESXi Host 的配置

4. 虚拟机迁移

4.3 Storage vMotion

Storage vMotion

- vMotion 将正在运行的虚拟机从一个主机迁移到另一个主机，在主机之间移动 CPU 和内存资源使用情况，但虚拟机的存储保持不变。
- Storage vMotion 将正在运行的虚拟机的虚拟磁盘从一个数据存储迁移到另一个数据存储，而虚拟机在同一主机上继续使用 CPU 和内存资源，达到手动平衡数据存储的利用率的目的。
- Storage vMotion 是实时迁移，将虚拟磁盘从一个数据存储迁移到另一个数据存储时，虚拟机不会产生任何停机时间。



4. 虚拟机迁移

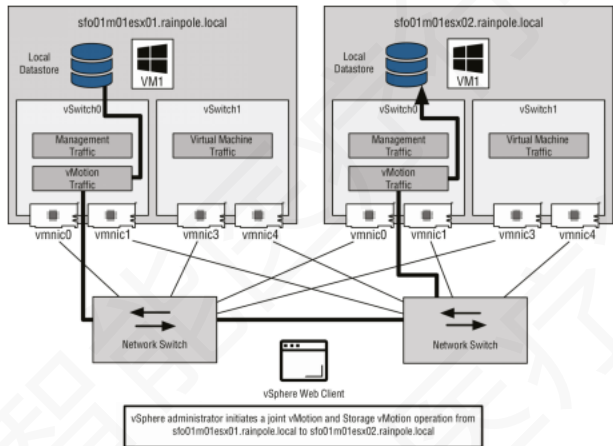
4.3 Storage vMotion

Storage vMotion 的工作过程

- 通过 vSphere 复制功能创建 VM 的非易失性文件：配置文件、VMkernel 交换区、日志文件和快照等。
- 在目标数据存储上创建影子虚拟机，且设置为闲置等待状态。
- Storage vMotion 将源虚拟磁盘单程复制到目标数据存储上。
 - 复制过程中对源进行的更改，将同步到目标数据存储上的虚拟磁盘文件。
- 当虚拟磁盘文件复制完成后，vSphere 快速暂停并恢复 VM，将 VM 的存储对应到目标数据存储上的影子虚拟机。
 - 与 vMotion 一样，这个过程非常快，以至于不会对服务造成中断。
- 将源数据存储上的 VM 文件删除。

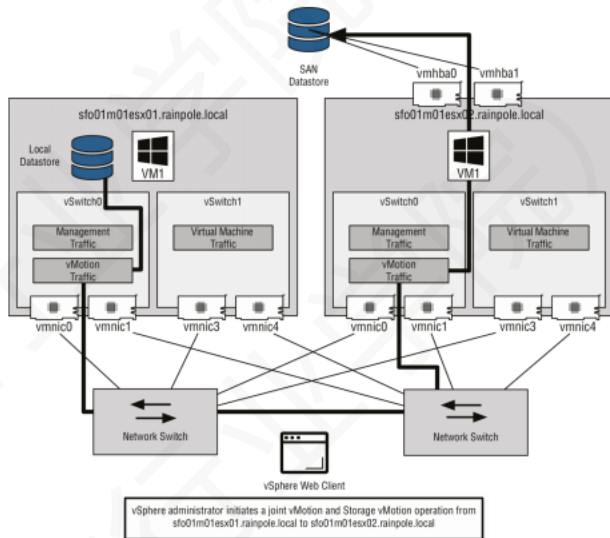


Storage vMotion 的应用场景举例



两台 ESXi Host:

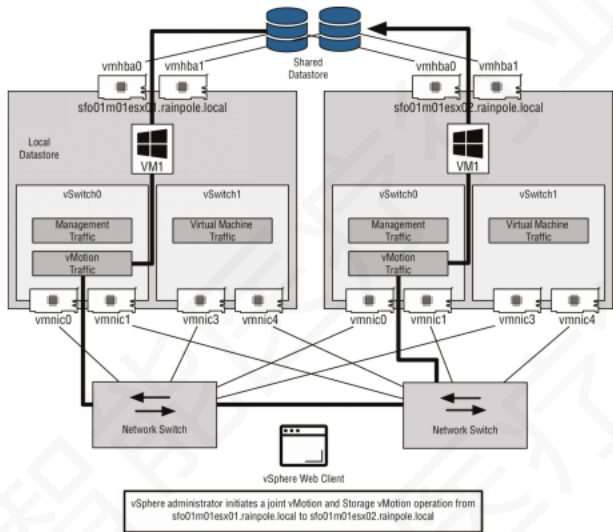
单个 vMotion 网络，两台主机均使用本地数据存储。
将 VM 迁移时启动两个数据流操作，首先进行存储复制，其次进行内存复制，所有数据流经同一个 vMotion 网络。



两台 ESXi Host:

单个 vMotion 网络，一本地数据存储，一台 SAN 存储。
将 VM 迁移时启动两个数据流操作，首先进行存储复制，其次进行内存复制，所有数据流经同一个 vMotion 网络。
第二台主机通过存储网络将收到的数据推送到 SAN。

Storage vMotion 的应用场景举例



该模式是数据中心的常见模式和推荐方案。

两台 ESXi Host:

单个 vMotion 网络，均使用 SAN 存储。SAN 或者 NAS 连接所有 ESXi Host。

将 VM 迁移时启动两个数据流操作：

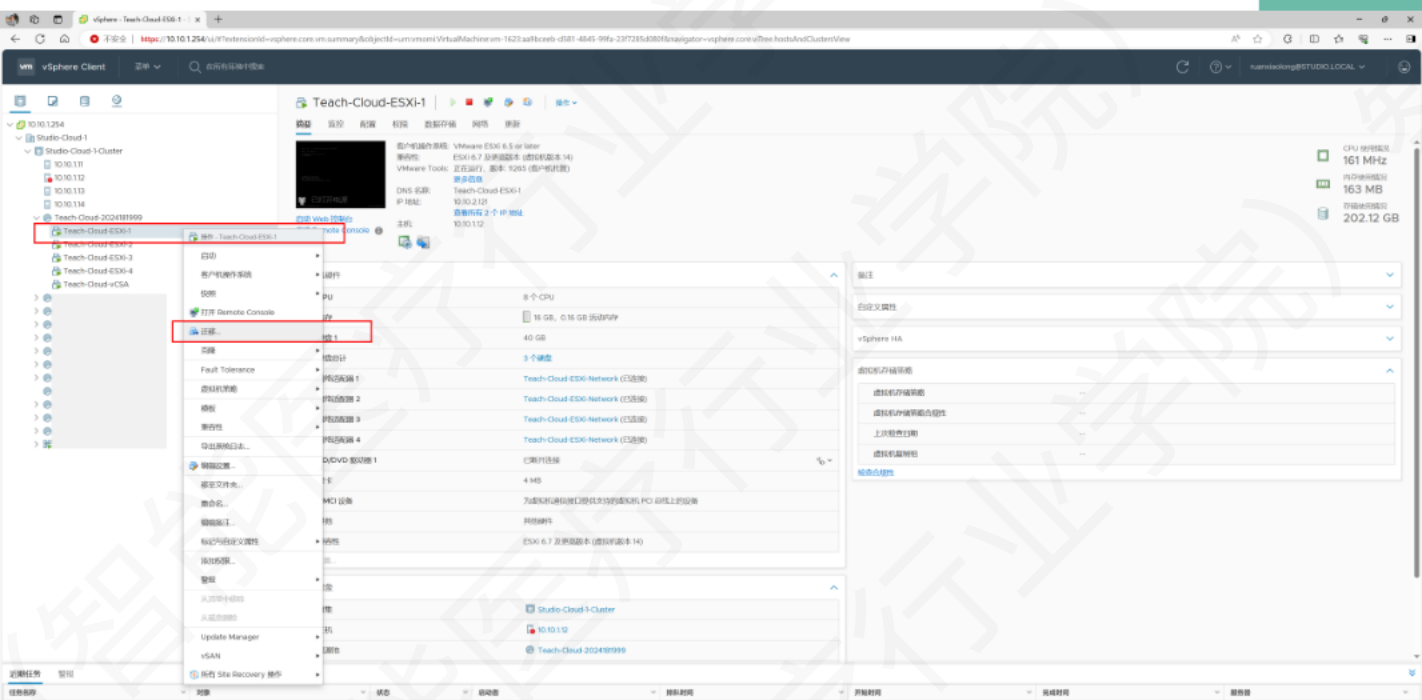
vMotion 网络承担内存数据流。

Storage vMotion 会检测两台 ESXi Host 是否可以同时查看源数据存储和目的数据存储，使用存储网络进行数据迁移。

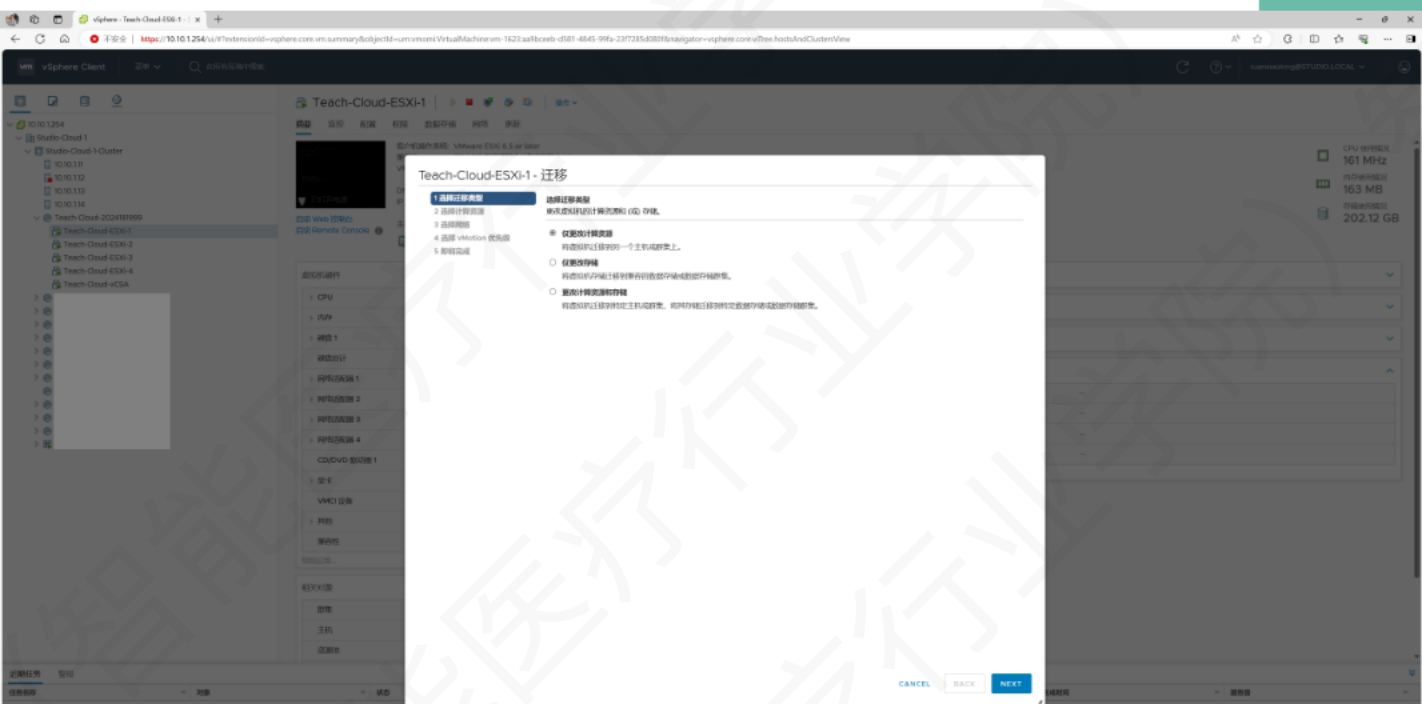
将 vMotion 和 Storage vMotion 组合后，首先会执行存储迁移。

原因1：虚拟磁盘比内存更大，速度也更慢。Storage vMotion 将比 vMotion 花费的时间长得多。

原因2：基于虚拟磁盘的存储的更改速率通常远低于内存。如果先执行 vMotion 操作，则在等待 Storage vMotion 任务完成时，内存位图文件会变得更大。



VM 的迁移



VM 的迁移类型

4. 虚拟机迁移

4.4 Cross-vCenter vMotion

□ Cross-vCenter vMotion, 跨 vCenter 的 vMotion

- 自 vSphere 6 开始, VMware 进一步增强了 vMotion 的功能, 使得 vMotion 能够跨群集、跨 vCenter 进行工作。
- 实现在不同的 vCenter 之间迁移正在运行的虚拟机, 实现虚拟机在跨 vCenter 域的动态移动, 并确保迁移过程中虚拟机的状态、配置和数据的完整性。
- 其工作流程是:
 - 在源 vCenter 和目标 vCenter 之间建立信任和连接。
 - 识别要迁移的虚拟机及其相关资源。
 - 在迁移过程中, 对虚拟机的状态、内存数据等进行同步传输, 同时在目标 vCenter 中进行相应的资源配置和适配。



4. 虚拟机迁移

4.4 Cross-vCenter vMotion

□ Cross-vCenter vMotion, 跨 vCenter 的 vMotion

■ 应用价值:

□ 资源灵活调配:

- 可以在不同的数据中心或 vCenter 管理域之间根据业务需求灵活地调配虚拟机资源, 提高资源利用率。

□ 业务连续性保障:

- 方便在出现故障或需要进行维护时, 将虚拟机快速迁移到其他 vCenter 中, 确保业务的持续运行。

□ 整合与扩展:

- 有助于在进行数据中心整合或扩展时, 更便捷地管理和迁移虚拟机。

□ 多站点协作:

- 支持不同站点的 vCenter 之间进行虚拟机迁移, 便于跨地域的协作和资源共享。

■ 应用场景:

- 企业合并与整合、数据中心迁移、多地多中心容灾与备份、跨地域分支机构协作 等

4. 虚拟机迁移

4.4 Cross-vCenter vMotion

□ Cross-vCenter vMotion, 跨 vCenter 的 vMotion

- 跨 vCenter vMotion 需要满足 vMotion、Storage vMotion 的条件。
- 还有一些其他特殊的要求：
 - 参与迁移的 vCenter Server 至少为 vCenter Server 6.0 版本
 - 参与迁移的 vCenter Server 必须加入同一 Single Sign-On (SSO) 域
 - 参与迁移的 ESXi Host 至少为 vSphere 6.0 版本
 - 参与迁移的 ESXi Host 之间的 RTT 必须小于 150 毫秒
- 为了确保迁移后业务可持续, 还需要:
 - 迁移后的网络环境能够保证 VM 的需要
 - 需要考虑数据备份, 基于 WAN 的备份方案建议不要采用
 - 业务迁移后的访问策略和性能, 需要测试和评估



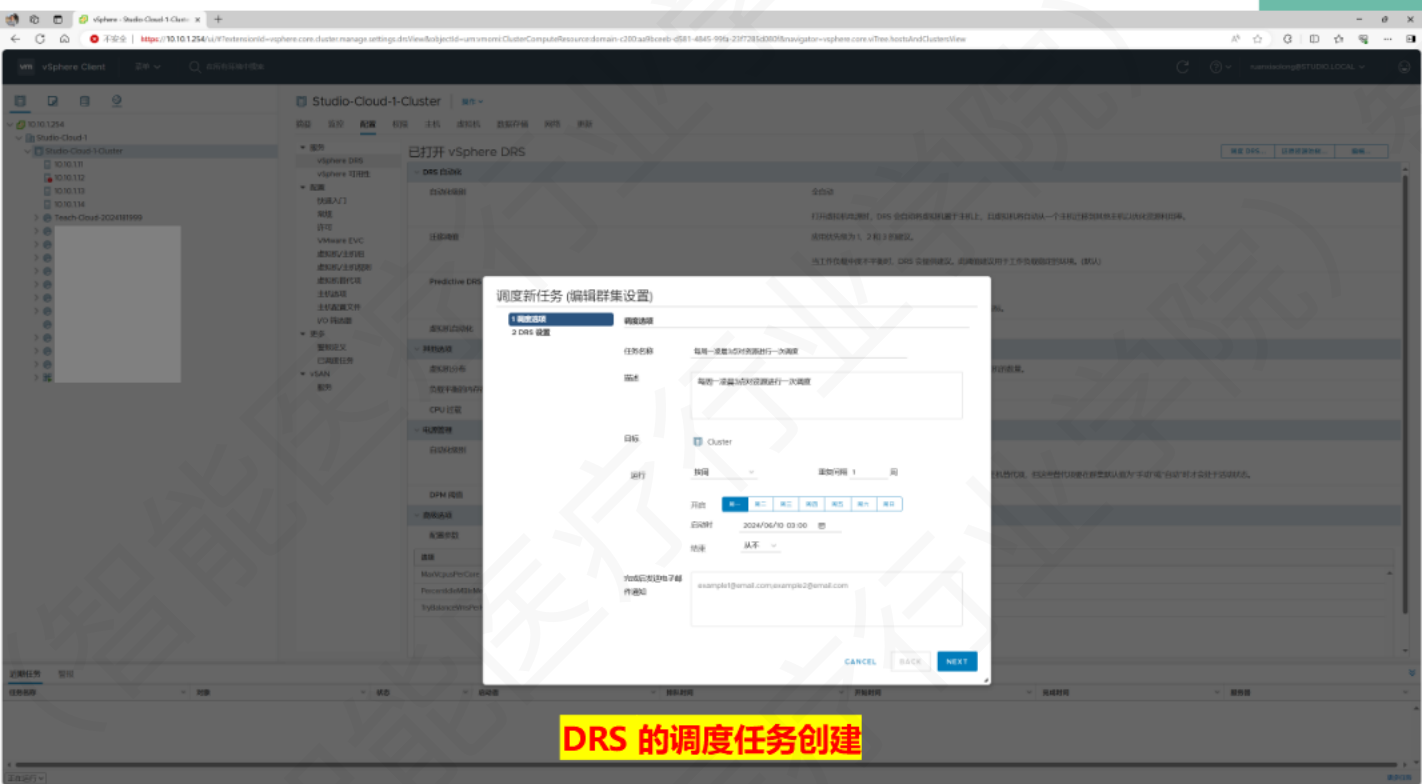
4. 虚拟机迁移

4.5 vSphere Distributed Resource Scheduler, DRS

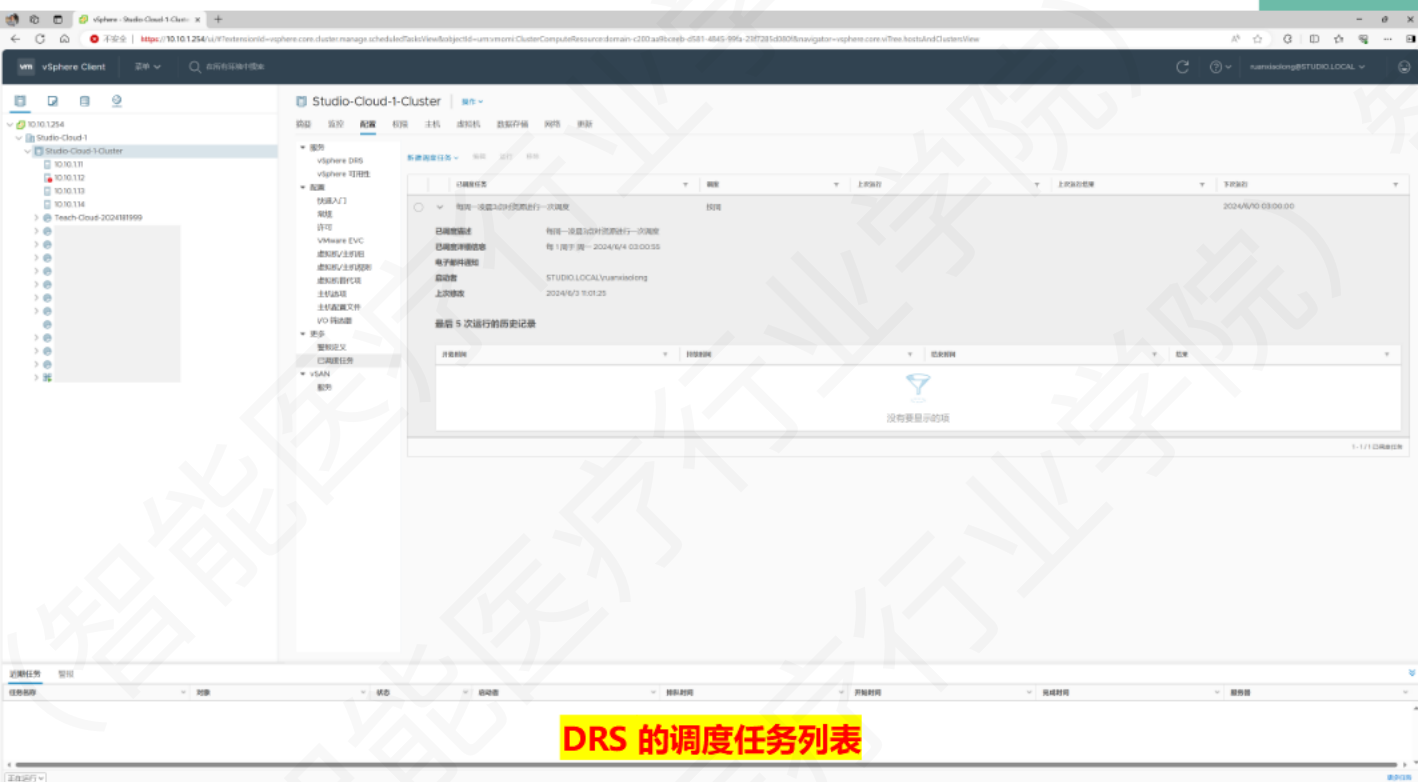
□ DRS，分布式资源调度

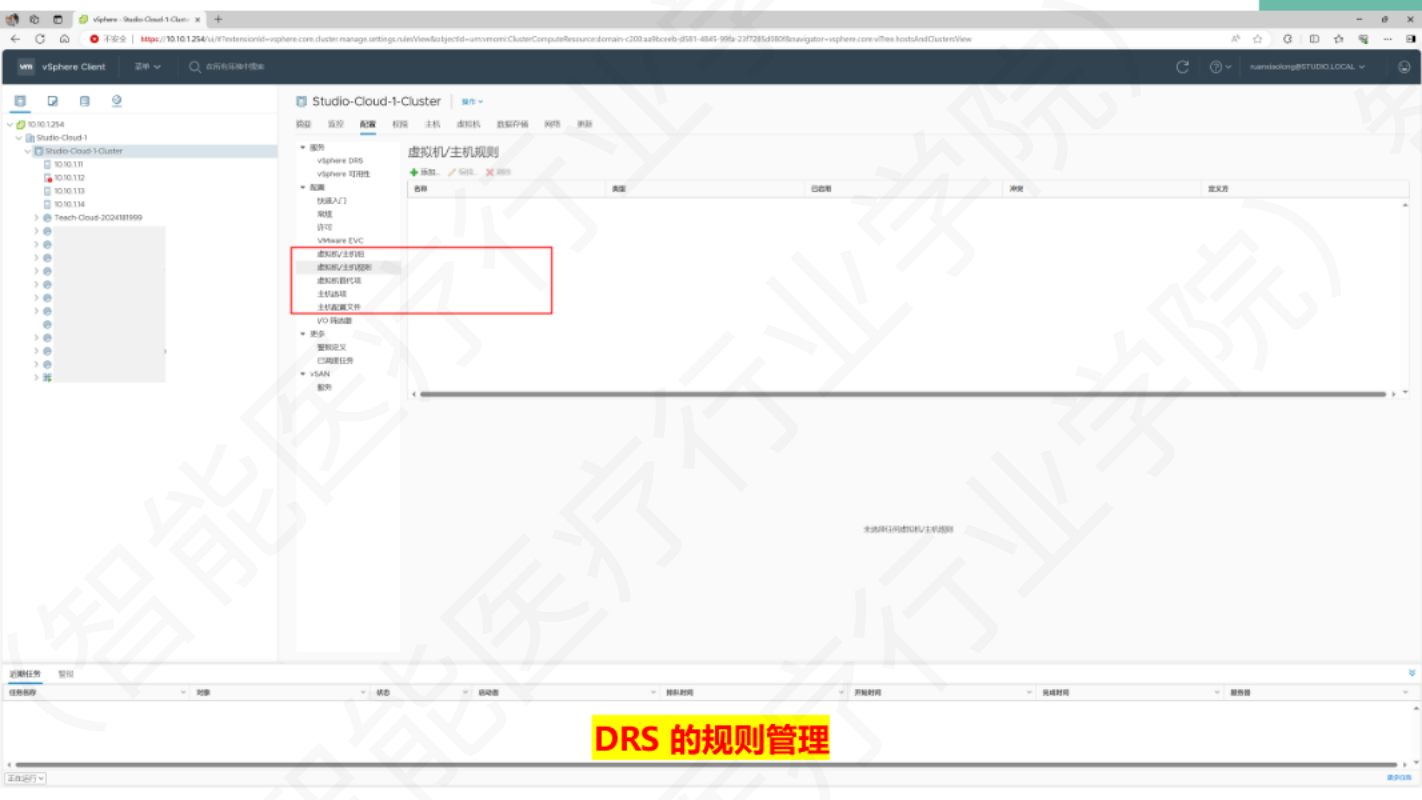
- DRS 需要在集群的基础上进行工作。
- DRS 周期性地评估集群中主机的资源使用情况和虚拟机的资源需求。
- DRS 根据定义好的规则和算法决定是否进行虚拟机迁移，实现自动迁移。
 - 当某主机负载过高，而其他主机有足够资源时，DRS 会发起虚拟机的迁移操作。
 - 迁移过程中，VM 的运行状态不受影响，并尽量减少对业务的干扰。
 - DRS 与其他功能（如 vMotion）协同实现高效的资源管理，实现对资源的智能调度和优化。
- 通过 DRS 可以实现：
 - 资源分配和优化
 - 负载均衡
 - 资源规划





DRS 的调度任务创建





DRS 的规则管理

4. 虚拟机迁移

4.6 vSphere Distributed Power Management, DPM

□ DPM，分布式电源管理

- 根据群集资源利用率，启用了 DRS 的群集可以通过 DPM 打开或关闭 ESXi Host 的电源来降低数据中心的能耗。
- 实现原理：
 - 群集中 VM 所需总内存和 CPU 资源，与群集中 ESXi Host 的总资源量进行比较。
 - 如果群集提供过多资源，则 DPM 将一个或多个 ESXi Host 置于维护模式，并在将其上的 VM 迁移到其他 ESXi Host 后，将关闭 ESXi Host 的电源。
 - 当提供的容量不足时，DRS 会打开 ESXi Host 的电源，使其退出维护模式，通过 vMotion 重新进行资源调度。



4. 虚拟机迁移

4.6 vSphere Distributed Power Management, DPM

□ DPM，分布式电源管理

■ vSphere DPM 支持三种协议进行 ESXi Host 的电源管理：

□ Intelligent Platform Management Interface (IPMI)

- IPMI 是一组交互标准管理规范，由 Intel、HP、Dell 和 NEC 公司于 1998 年 9 月 16 日共同提出。
- 目前已经有超过 200 家计算机制造商/供应商支持。

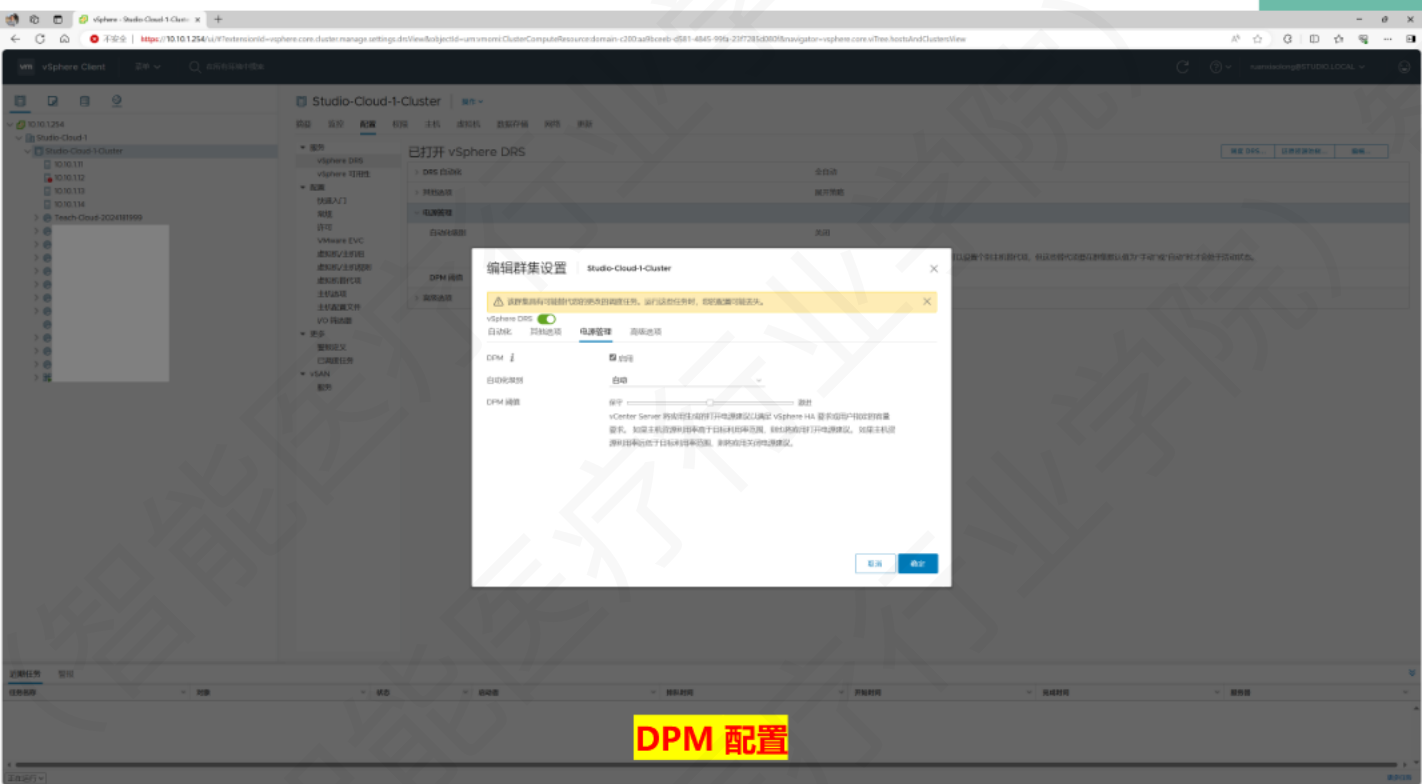
□ Hewlett Packard Integrated Lights-Out (iLO)

- 惠普服务器特有的远程管理技术。

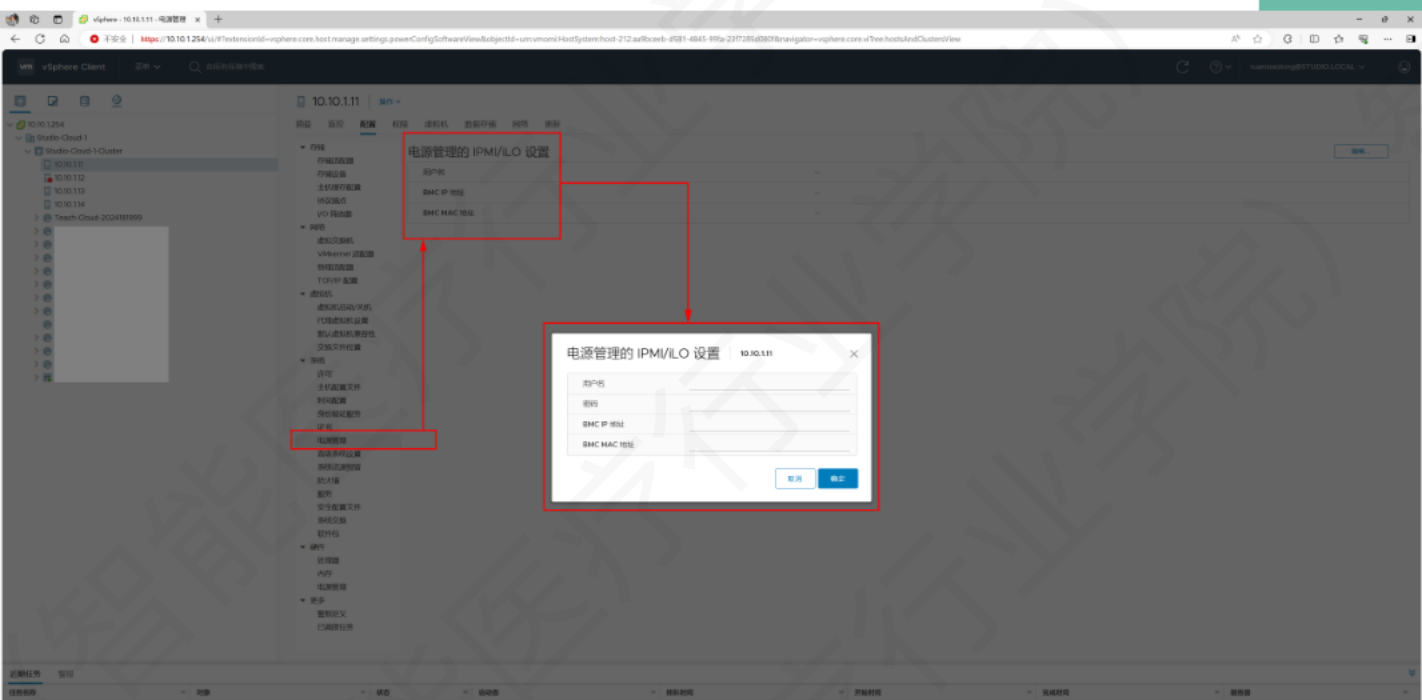
□ Wake-On-LAN (WOL)

- 网络唤醒功能是大多数主板具有的功能，通过 BIOS 进行设置。





DPM 配置



DPM 功能需要对每台 ESXi Host 配置 IPMI/iLO

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