

第八节 vSphere-资源管理

讲师:崔应龙

邮箱: cuiyl@5iblue.com.cn

北京蓝色曙光信息技术有限公司

Beijing blue light information technology co., LTD



内存、CPU、资源池和VApp

网络和存储I/O控制

vMotion原理介绍

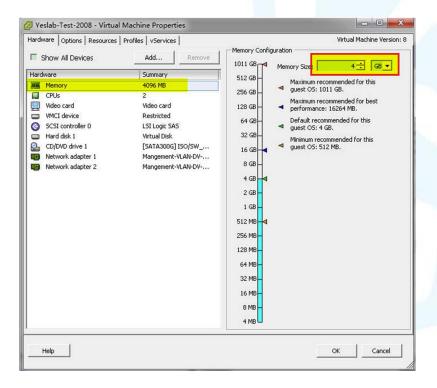
配置管理DRS集群

资源监控、报警



关联VM内存

虚拟机感觉自己拥有4G的内存,并且最多不会使用超过4G的物理内存。我们可以超额的关联内存给VM,例如: ESXi主机的物理内存只有8G,但是我们可以给三个VM分配4G的内存。



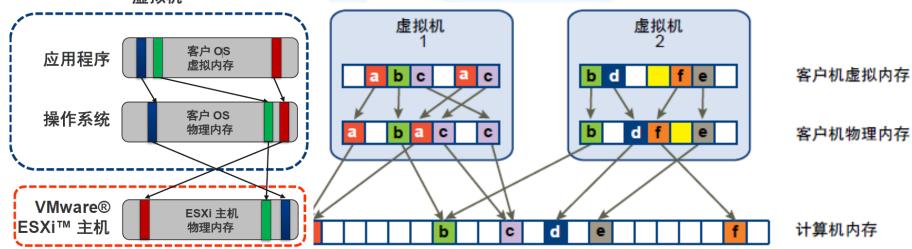




ESXi四大高级内存控制技术

1. page sharing (透明的页面共享) 内存的去重 透明页共享仅允许存储一次具有相同内容的页面 VMware vSphere® 中有 3 层内存。

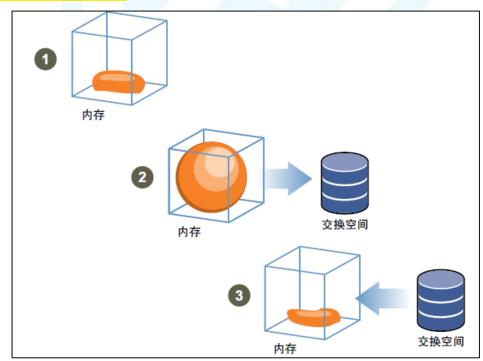
- ➤ 客户 OS 虚拟内存,该内存由操作系统提供给应用程序。
- ➤ 客户 OS 物理内存,该内存由 VMkernel 提供给虚拟机。
- ▶ 由 VMkernel 管理的主机物理内存,可向虚拟机 提供可编址的连续内存 空间。 虚拟机





ESXi四大高级内存控制技术

- 2. Ballooning(气球,内存回收)当主机没有内存分配时出发这个机制!!!
- ► ESXi 主机可以从<mark>虚拟机</mark>中回收内存 vmtools生成一个程序 vmmemctl, 占满内存, 当前活动内存进入swap 不活跃内存进入磁盘, 吹气球,把共46全挤出去,之后再把活跃的部分加入 内存虚拟增长驱动程序 (vmmemctl) 与服务器协作回收客户机操作系统认为最不
- 重要的页面
- ➤ 需要<mark>虚拟机安装VMTools</mark>



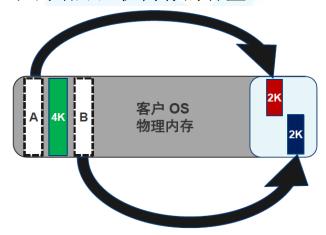


ESXi四大高级内存控制技术

- 3. Memory Compression(内存压缩) 气球机制之后,内存回收完后,内存还不够,即将使用swap空间时启用内存压缩
- ➤ ESXi 提供内存压缩缓存,可在内存过载使用时改进虚拟机性能。默认情况下已启用了内存压缩。当主机内存过载时,ESXi 会压缩虚拟页面并将其存储在内存中。
- ▶ 由于访问压缩的内存比访问交换到磁盘的内存更快,因此通过 ESXi 中的内存压缩可以使内存过载,但不会显著影响性能。
- ▶ 当需要交换虚拟页面时,ESXi 会首先尝试压缩虚拟页面。可压缩至 2 KB 或更小的页面存储在虚拟机的压缩缓存中,从而增加主机内存的容量。

内存页面已压缩为 2 KB, 并且存储在每个虚拟机的压缩缓存中。

- 用于交换到磁盘的备选内存 页面将进行压缩。
- 在内存中对压缩页面进行解 压比执行磁盘 I/O 操作的速 度快。
- 压缩操作只在存在物理内存 资源争用时执行。



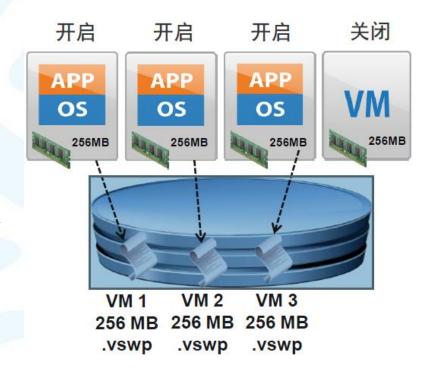
虚拟机内存已经用完,用自己的交换空间 虚拟机申请内存,主机内存用完,使用主机的交换空间 虚拟机申请内存,主机内存用完,主机的swap也用完,则虚拟机使用 开启时生成的swap(存在存储中)。

ESXi四大高级内存控制技术

4. Swapping (交换文件)

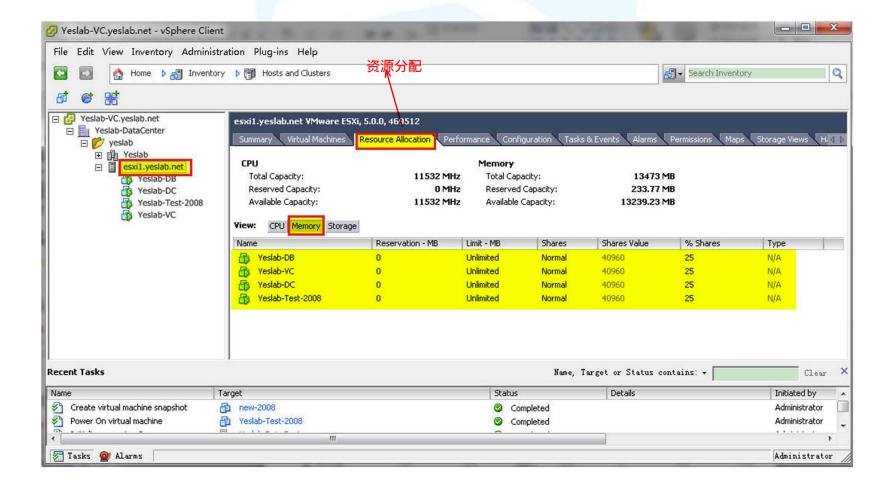
- a) 客户机(VM)交换文件
- b) 主机(ESXi)交换
- ➤ 内存的回收机制通过将数据移出内存并写入后台存储实现。 从后台存储访问数据的速度比 从内存访问数据的速度慢,因此一定要仔细选择存储交换数据的位置。建议使用SSD
- ➤ 如非必要,系统不会使用 VMkernel 交换空间,因为这种 方式的性能很差
- ▶ 交換文件的大小等于已分配内 存和预留内存的容量差

已分配内存 = 512 MB 预留内存 = 256 MB





查看主机的内存





内存预留、限制和份额

- 1.没有做限制的时VM能够请求的最大内存数量就是配置的数量,限制决定了该虚拟机能够使用的最大物理内存内存
- 2. 预留决定了这个VM至少能够使用多少物理内存,其余的可能使用swap,也可能使用物理内存
- 3. 份额决定了虚拟机使用内存的比例

实例一:数值大分的细

物理主机内存2000M

VM1 预留500M 限制1000M 份额值1000

VM2 预留500M 限制1000M 份额值1000

VM1 实际得到的物理内存为1000M

VM2 实际得到的物理内存为1000M

实例二:

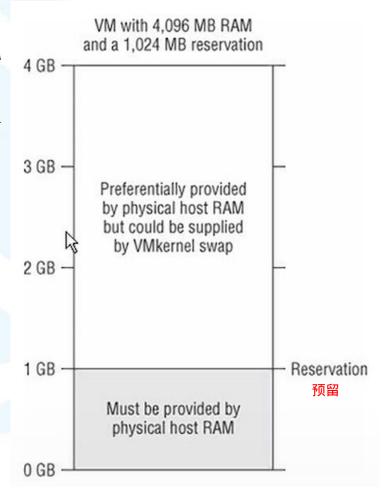
物理主机内存2000M

VM1 预留500M 限制1000M 份额值2000

VM2 预留500M 限制1000M 份额值1000

VM1 实际得到的物理内存为1166M 1000

VM2 实际得到的物理内存为833M





超线程,把一个线程分成两个

VM CPU介绍

内存有虚拟内存,然而CPU却无法设置虚拟CPU,即无法用其他资源来顶替CPU,CPU的调度只能使用物理CPU的核心或者线程(如果虚拟机开启了CPU的超线程)

多CPU调度介绍:

CPU是通过时间片的来分配的计算能力的,即每一个CPU 的核心或者线程都是通过时间片来分配给虚拟机使用,那么虚拟 机向主机申请CPU调度时间片时,和要申请的CPU时间片的个数有 关。

单个CPU的虚拟机更容易申请到CPU时间片,多个CPU的虚拟机需要同时申请到多个时间片才可以。

如果不是虚拟机必须使用多个vCPU,建议在开始创建VM的时候,都使用一个vCPU。

vSphere资源管理---CPU



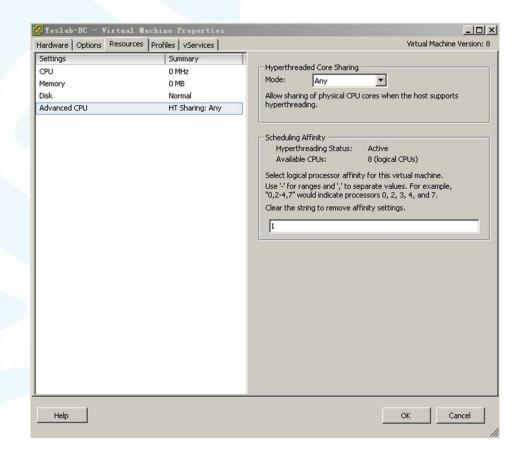
CPU亲核技术 (关联性)

CPU亲核技术 (关联性)

可以指定虚拟机使用某个物理的CPU核心。

使用CPU关联性的限制

- 1. 使用cpu亲核无法使用vMotion
- 2. CPU无法进行负载均和分配
- 3. 启用亲和技术后DRS只能设置 为手动模式



vSphere资源管理---资源池

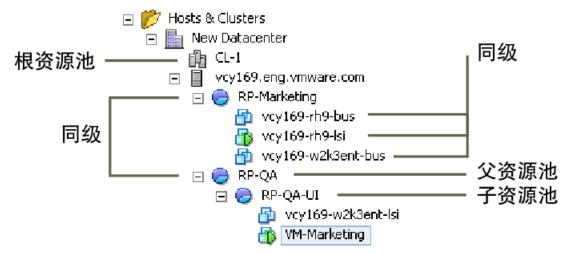


根资源池:独立的主机、集群

资源池介绍

- ➤ 资源池是一个<mark>逻辑抽象概念</mark>,用于分层管理 CPU 和 内存资源,可以基于资源池做资源 分配和权限控制。
- ➤ 每个独立主机和每个<mark>启用了DRS的群集</mark>都具有一个(不可见的)根资源池,此资源池对 该主机或群集的资源进行分组
- ▶ 用户可以创建根资源池的子资源池,也可以创建用户创建的任何子资源池的子资源池。 每个子资源池都拥有部分父级资源,然而子资源池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/或虚拟机。 处于较高级别的资源池称为父资源池。处于同一级别的资源池和虚拟机称为同级。群集 本身表示根资源池。如果不创建子资源池,则只存在根资源池

图 9-1 资源池层次结构中的父级、子级和同级

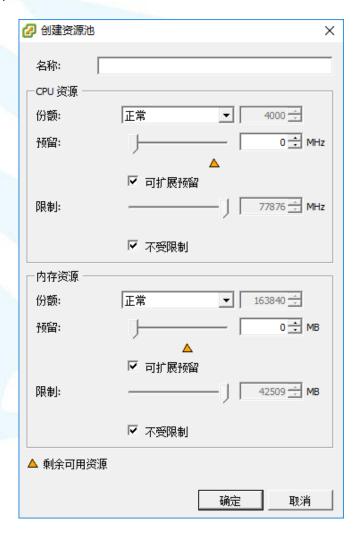




资源池介绍

资源池属性:

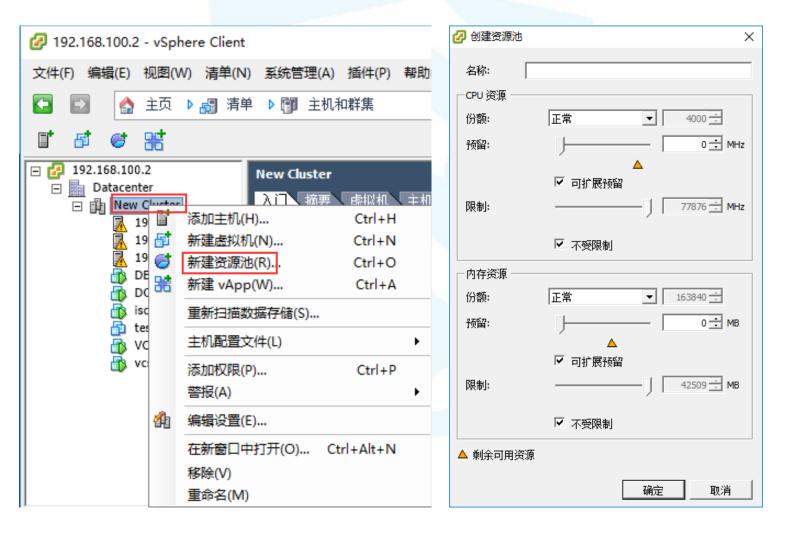
- ▶ 份额:
 - 低、正常、高、自定义
- ▶ 预留,以 MHz 和 MB 为单位
- ➤ 限制(以 MHz 和 MB 为单位):
 - 默认情况下为无限量访问(最大值为可访问的最大资源量)
- ▶ 预留是否可扩展?
 - 是 虚拟机和子资源池可以提取其父级池中的资源。
 - 否 虚拟机和子资源池仅可提取此池中的资源,即使其父级池具有可用资源,也不可提取。



vSphere资源管理---资源池



新建资源池



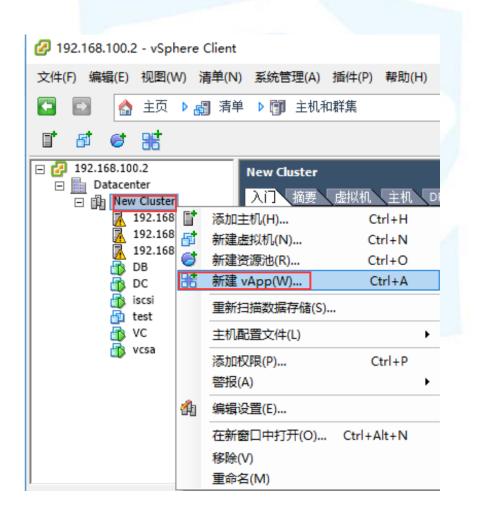


vAPP介绍

vApps可以将一组具有某种关联关系的虚拟机集中进行管理,使得某 些顺序操作或者关联操作可以按照预先设置好的规则进行 通过 vSphere vApp 可以对多个彼此交互操作的虚拟机和软件应用程序进行 打包,这些虚拟机和软件应用程序可以作为一个单元进行管理,并以 OVF 格 式进行分发。一个 vApp 可以包含一个或多个虚拟机,但对该 vApp 执行的 任何操作(如克隆或关闭电源)都会影响 vApp容器中的所有虚拟机。 可以使用vApp执行资源管理和其他某些管理活动,例如,同时为多个虚拟机 打开电源。可以将 vApp 视为虚拟机容器,您可以对该容器执行操作。 创建 vApp 后,可以将其添加到文件夹、独立主机、资源池、为 DRS 启用的群 集或另一个 vApp 中。

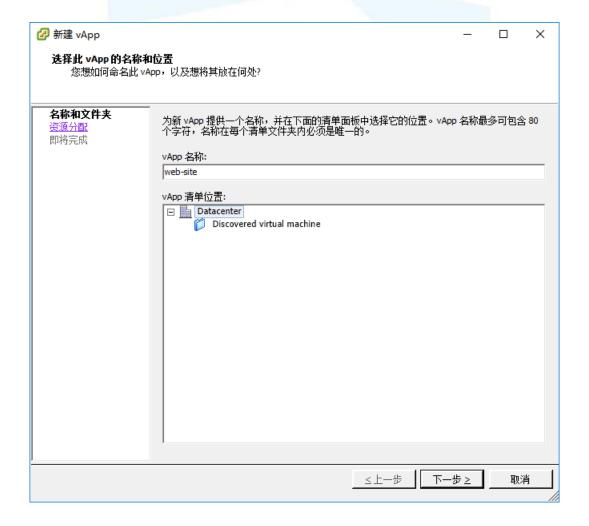


创建vAPP





创建vAPP



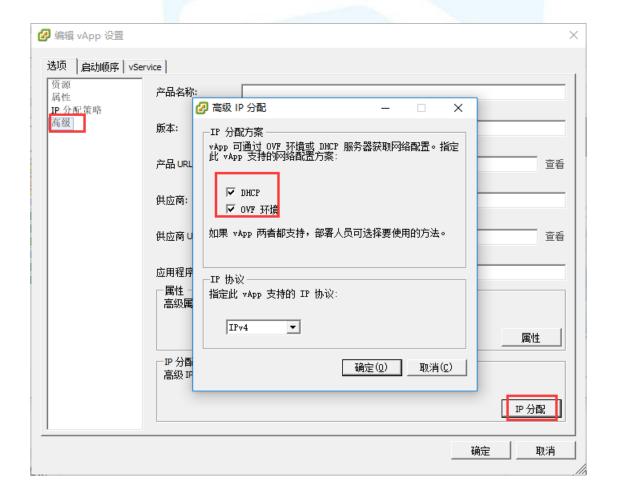


创建vAPP

🕜 新建 vApp			_		×
资源分配 您希望如何为 vAp	p 分配 CPU 和内存?				
名称和文件夹 资源分配 即将完成		正常 4000 🚉			
	预留:	○ → MHz ▼ 可扩展预留			
	限制:	77876 <u>*</u> MHz			
	一内存资源 ——	▼ 不受限制			
	份额: 预留:	正常 <u>163840</u> MB			
	限制:	✓ 可扩展预留 ————————————————————————————————————			
		▼ 不受限制			
		≤上一步	下一步≥	取	消

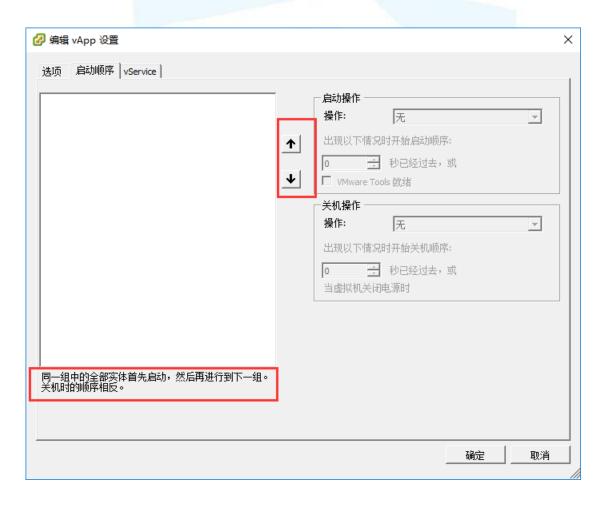


编辑vAPP



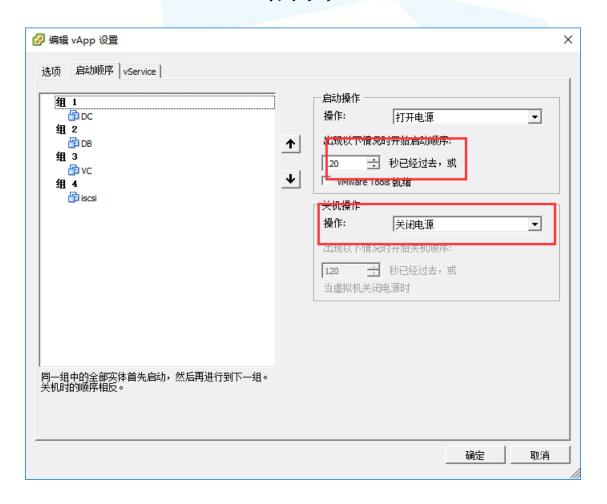


编辑vAPP





编辑vAPP





内存、CPU、资源池和vApp

网络和存储I/O控制

vMotion原理介绍

配置管理DRS集群

资源监控、报警



网络I/O控制介绍

- > 网络I/O控制利用<mark>网络资源池</mark>来控制网络资源的使用率
- ▶ 能够控制出向的流量和入向的流量
- ▶ <mark>只有在分布式交换机</mark>上才能进行网络I/O控制 企业增强版
- ▶ 要新建资源池,需要先设置系统虚拟机流量的预留

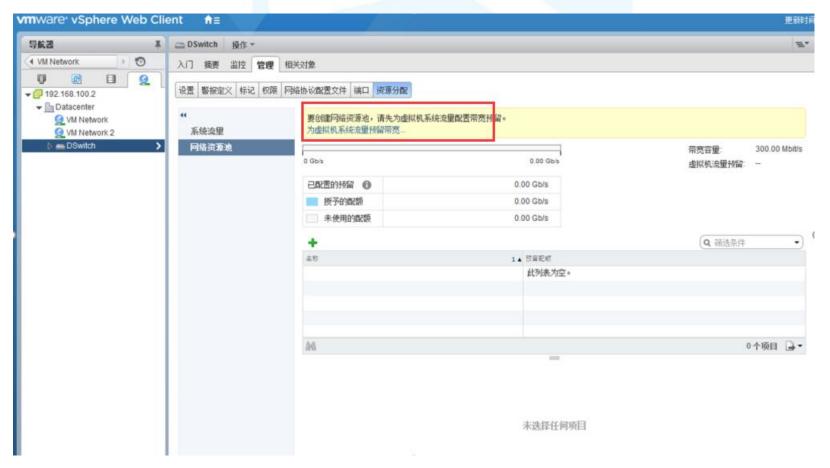
vSphere 可以识别一下类型的网络流量:

流量类型	1 🛦	份额	份额值	預留	限制
Fault Tolerance (FT) 流量		正常	50	0 Mbit/s	
NFS 流里		正常	50	0 Mbit/s	
Virtual SAN 流量		正常	50	0 Mbit/s	
iSCSI流里		正常	50	0 Mbit/s	
vMotion 流里		正常	50	0 Mbit/s	
vSphere Data Protection 备份流量		正常	50	0 Mbit/s	
vSphere Replication (VR) 流里		正常	50	0 Mbit/s	
管理流量		正常	50	0 Mbit/s	
虚拟机流量		高	100	0 Mbit/s	



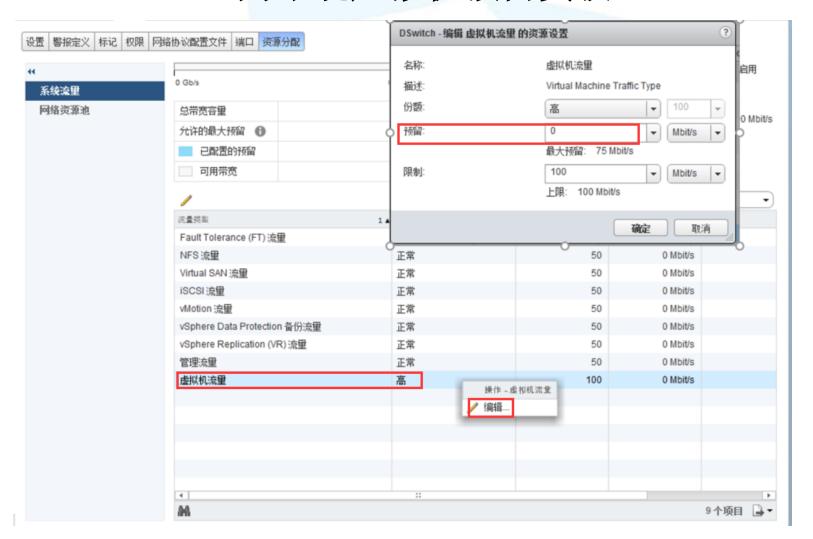
网络I/O控制介绍

> 要新建资源池,需要先设置系统虚拟机流量的预留



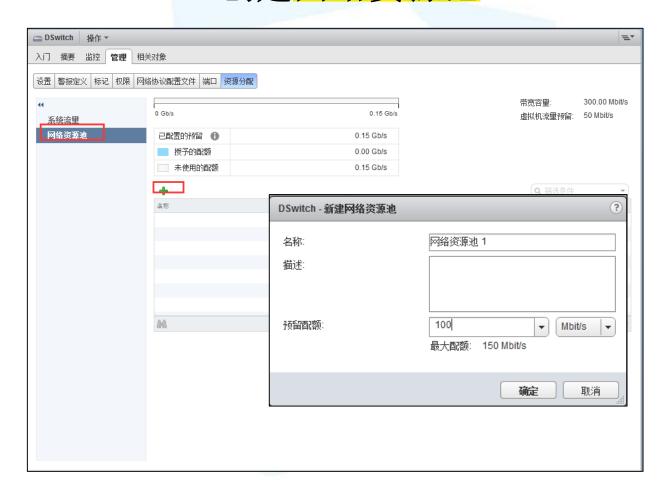


为系统虚拟机预留贷款



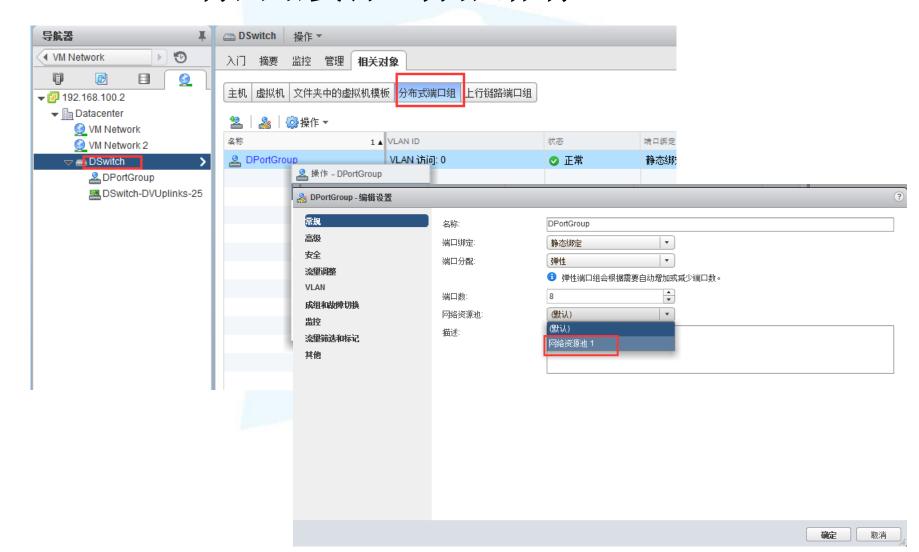


创建网络资源池





将网络资源池分配给端口组





i/ops到达80% 启用i/o控制 延迟达到

存储I/O控制介绍

通过 vSphere Storage I/O Control,可以优化群集范围内的存储 I/O,从而可更好地让存储进行工作负载整合。

Storage I/O Control 可以像CPU和内存一样来设置限制和预留以处理存储 I/O 资源。也可以控制在 I/O 拥堵期间分配给虚拟机的存储 I/O 量,从而确保更重要的虚拟机优先于重要性较低的虚拟机获得 I/O 资源分配。

当对数据存储启用 Storage I/O Control 时,ESXi 会开始监控主机与该数据存储通信时主机的设备滞后时间。当设备滞后时间超出阈值时,数据存储会被视为已出现拥堵,访问该数据存储的每个虚拟机都会按其份额比率分配相应的I/O 资源。还可以可以按虚拟机设置份额,并且可以根据需要调整每个虚拟机的份额。



存储I/O控制前提条件

- 1.必须在数据存储器上启用存储I/O控制
- 2.启用了 Storage I/O Control 的数据存储必须由单个 vCenter Server 系统管理
- 4.Storage I/O Control 不支持具有<u>多个数据区的数据存储</u>,即VMFS必须只有一个LUN组成。就是说不支持扩展过的VMFS



两种判断拥堵的指标

ESXi中的存储I/O注射器探测存储的容量和压力程度通过以下两种方式来 判断存储是否已经出现紧张性

- 1.吞吐量峰值百分比
- 2.存储的延迟时间如何设置延迟时间:

SSD: decrease to 10 ms.

10K/15K FC and SAS: 30 ms.

7.2K SATA/NL-SAS: 50 ms.

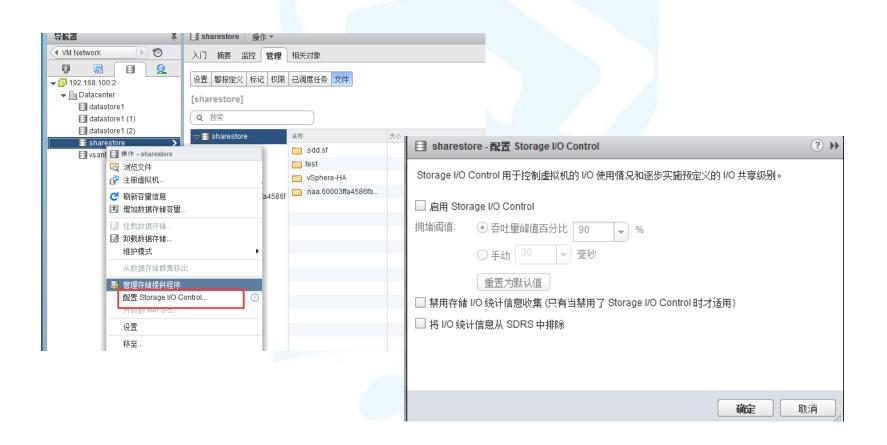
混合行存储: 至少30 ms.

企业级的flash 1ms





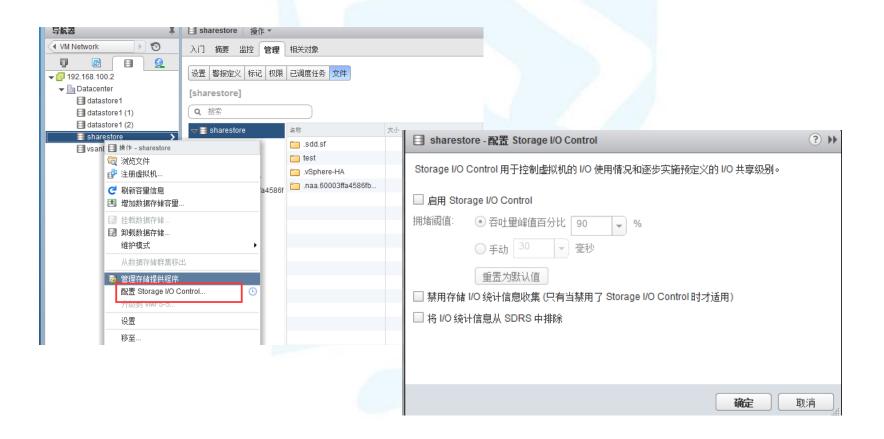
启用存储I/O控制





配置存储I/O控制

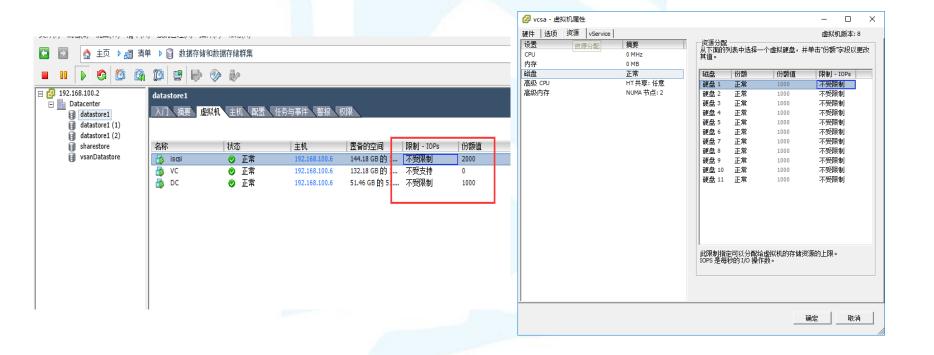
第一步: 在存储上启用storage I/O control并且设置判断启用存储控制的阈值





配置存储I/O控制

第二部: 在虚拟机上设置限制和份额





内存、CPU、资源池和vApp

网络和存储I/O控制

vMotion原理介绍

配置管理DRS集群

资源监控、报警

vSphere资源管理---vMotion原理介绍



vMotion原理介绍

vMotion是vSphere一个十分重要功能,可以实现开机的虚拟机在两台ES系主 机之间在线迁移, 切不会影响业务。

- 一下原因可能是要进行vMotion的主要原因:
- 全自动DRS才会自动迁移,否则只能手动

- ✓ 平衡平均 CPU 负载或预留。
 vmotion vmkernel 能诵讯 ✓ 平衡平均内存负载或预留。虚拟机要求两台esxi主机要有相同的网络配置, portgroup
- ✓ 满足资源池预留。
- ✓ 满足关联性规则。
- 共享存储(drs) cpu要兼容(1相同的 2可以通过evc兼容) 不要在业务高峰期迁移(不然会出现静默失败)
- ✓ 主机正在进入维护模式或待机模式。

vMotion的要求:

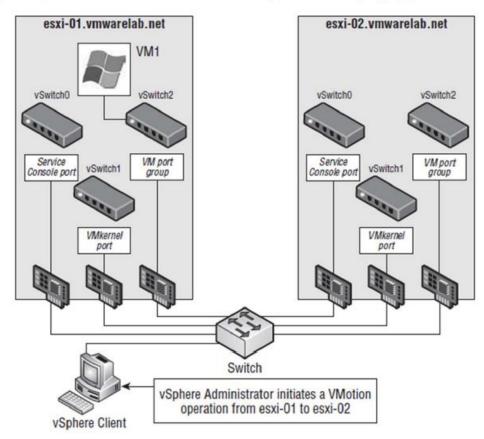
- ✓ 5.5之前需要有共享存储支持,之后的版本可以没有共享存储web端做
- ✔ 必须制定vmotion Vmkernel,并且有<mark>相同</mark>的标准交换机配置和分布式交换机配置, 即两台ESXi主机网络一直。
- ✓ CPU必须是同一厂商同一家族的CPU(同一个家族测CPU可以开启EVC)
- ✓ 64的操作系统必须启用Intel VT or AM-v

vSphere资源管理---vMotion原理介绍



vMotion迁移原理

Step 1:在VM处于"Power ON"状态,发起VMotion迁移.

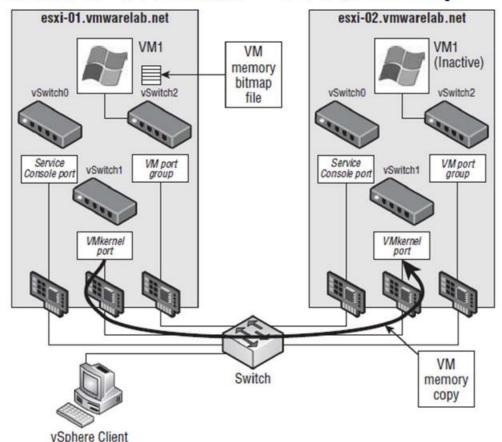




vMotion迁移原理

记录迁移过程中新的内存在那个页面(位置

Step 2:开始内存拷贝,并且创建一个内存bitmap.

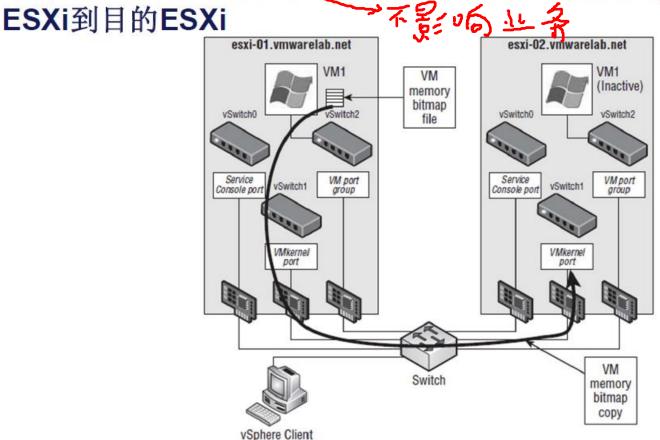




vMotion迁移原理 ///

mover进程

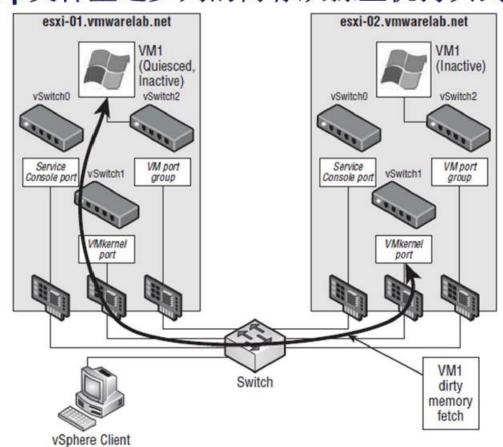
Step 3:VMotion迁移静默VM1并且传输内存bitmap文件从源





vMotion迁移原理

Step4:在bitmap文件里边罗列的内存从源主机拷贝到目的

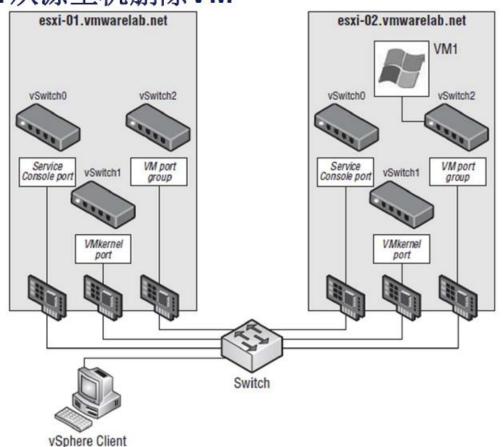




注销:从清单删除 注册:打开配置文件,右击添加到 清单

vMotion迁移原理

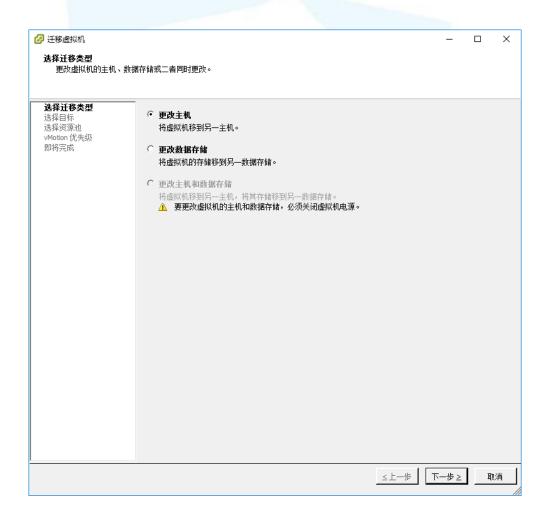
Step5:vCenter从源主机删除VM







实验:迁移虚拟机演示





内存、CPU、资源池和vApp

网络和存储I/O控制

vMotion原理介绍

集群和DRS

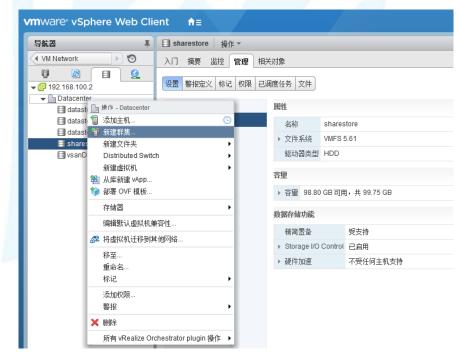
资源监控、报警



集群简介

集群是一个多台ESXi组成的一个管理组,群集在vSphere 中的和vCenter的重要性相当。只有创建了集群很多非常有用的功能才能被使用:不如说,vSphere High Availability (HA), vSphere Distributed Resource Scheduler(DRS), and vSphere Fault Tolerance(FT)、EVC、VSAN都是

工作在集群级别





创建集群

web端

都 新建群集		? ₩
名称	新建群集	
位置	Datacenter Datacenter	
▼ DRS	□打开	
自动化级别	全自动 ▼	
迁移阈值	保守 ——— 激进	
▶ vSphere HA	□打开	
▼ EVC		
EVC 模式	禁用	
描述	-	
▼ Virtual SAN	□打开	
向存储中添加磁盘	自动 ▼ Virtual SAN 将自动声明所包含的主机上的所有空磁盘。将不会在自动模式中声明远程磁盘。	
许可	⚠ 必须向群集分配许可证,以便自动创建磁盘组或消耗磁盘空间。	
	确定	取消



EVC技术介绍

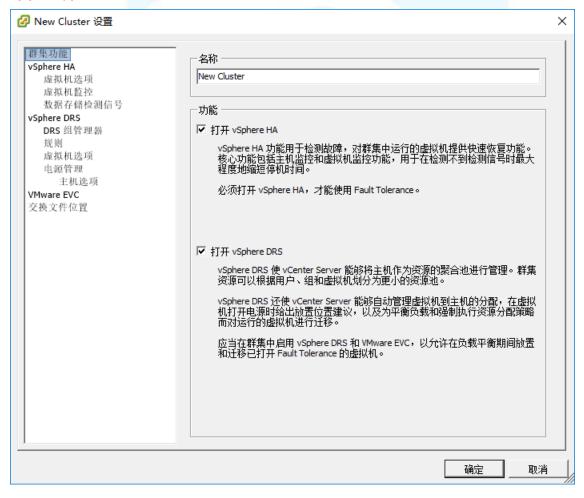
ESXi并不对CPU进行虚拟,如果把一个VM从一个安装志强CPU的服务器,迁移(vMotion)到另外一个安装i7 CPU的服务器上,会出现兼容性问题。所以VMware提出了EVC(Enhanced vMotionCompatibility)技术,它使用CPU Mask技术来模拟CPU,例如:多个ESXi主机安装了不同型号的Intel CPU,我们可以使用EVC技术,把这些CPU都模拟成为一个固定的Intel志强CPU型号。这样在这些主机之间,就可以使用vMotion技术来迁移VM了。

注意:如果在某一个ESXi主机内,存在开机状态的VM,那么这个ESXi不能启用EVC技术。后续要介绍的FT技术,必须在Cluster上启用EVC技术。



在集群上开启DRS

客户端

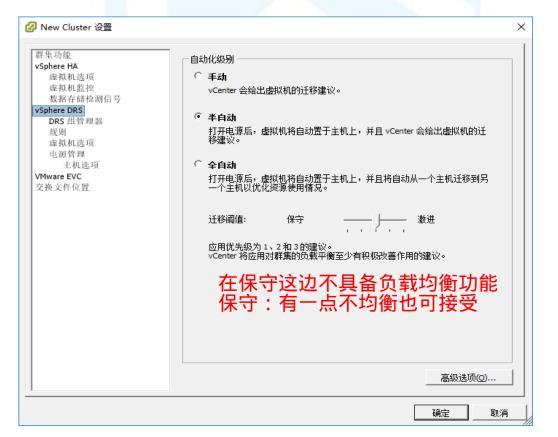


vSphere资源管理---集群和DRS DRS DRS 分布式资源调度



DRS的三种自动化级别

- 1.手动(开机的部署,在线迁移都是手动操作)
- 2.半自动(开机的部署自动完成,在线迁移需要手动操作)
- 3.全自动(开机的部署,在线迁移都是自动完成)

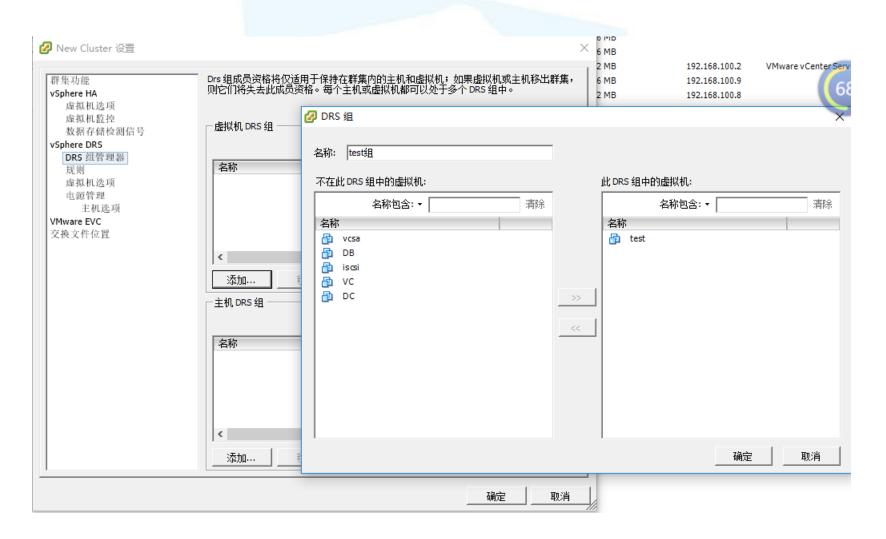




DRS关联规则,

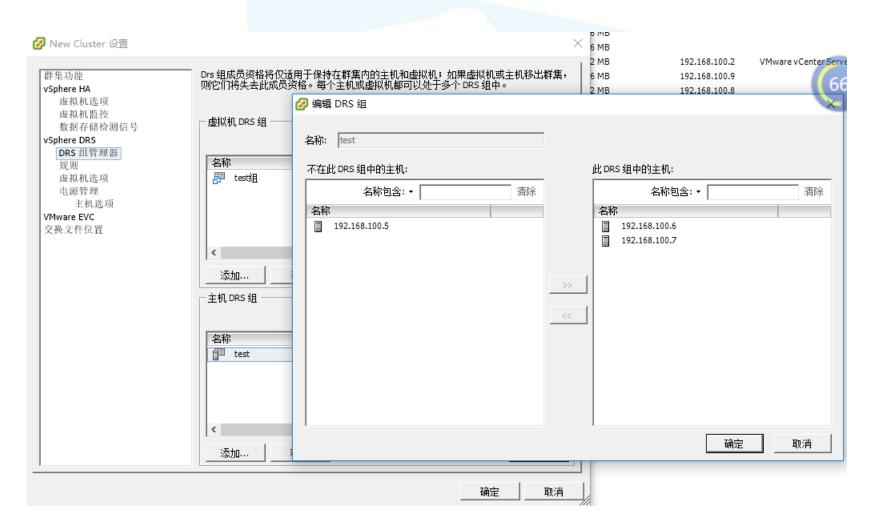
DRS定期运行,不是一直运行

创建虚拟机组组



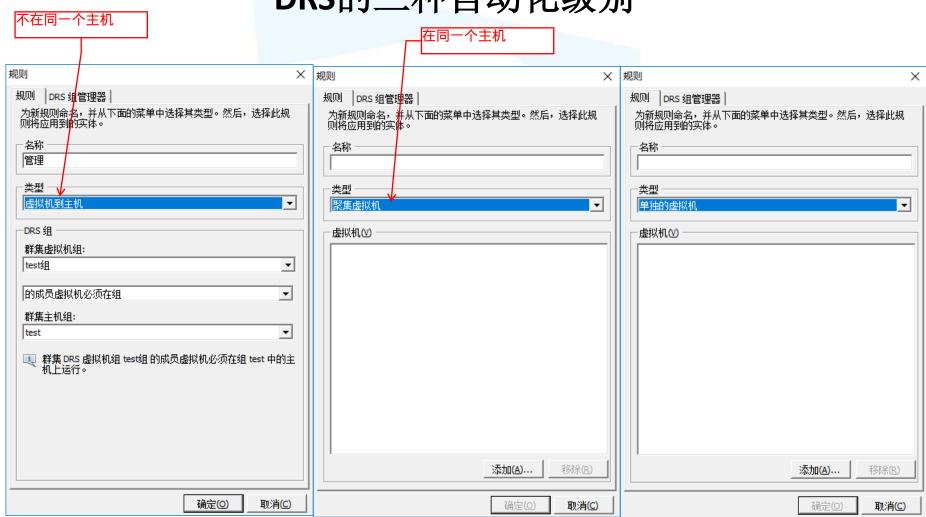


创建主机组





DRS的三种自动化级别





内存、CPU、资源池和vApp

网络和存储I/O控制

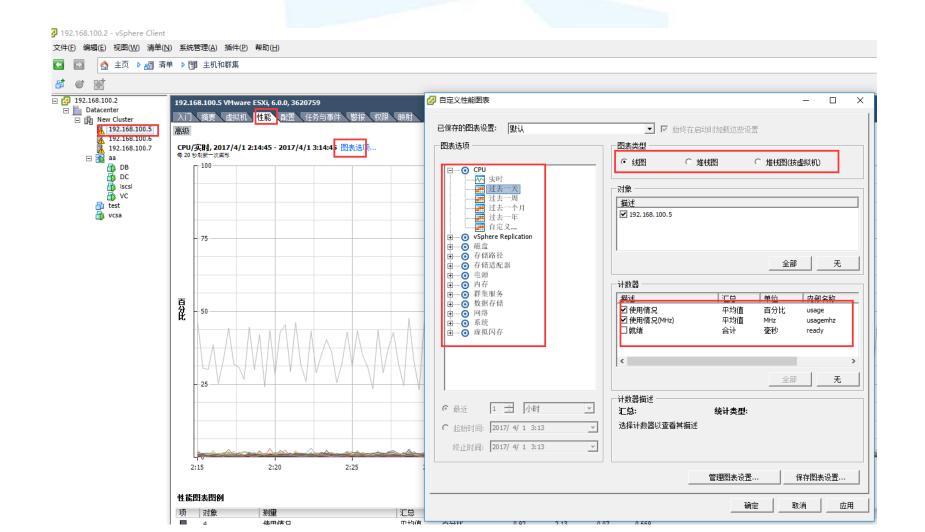
vMotion原理介绍

集群和DRS

资源监控、报警



使用vCenter监控主机性能



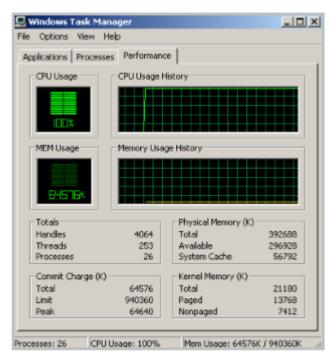


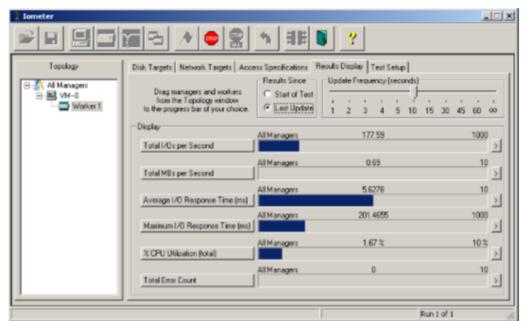
使用vCenter监控虚拟机





使用客户端工具监控虚拟机性能





Iometer

任务管理器



什么是报警

警报是为了响应清单中的对象所发生的事件或情况而生成的通知。

各种清单对象都有默认警报:

- ✓ 主机和虚拟机具有很多默认警报 您可以为各种清单对象创建自定义 警报:
- ✓ 虚拟机、主机、集群、数据中心、 数据存储、网络、分布式交换机 和 分布式端口组

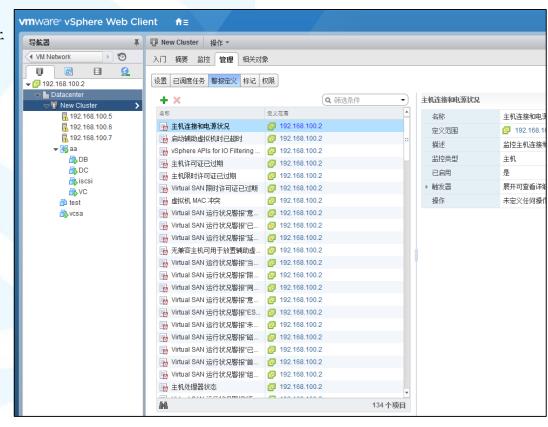
Name Datastore usage on disk Exit standby error Health status changed alarm Health status monitoring Host Baseboard Management Controller status Host battery status Host connection and power state Host connection failure Host cpu usage 默认的数据 Host error 中心警报 Host hardware fan status (部分列表) Host hardware power status Host hardware system board status Host hardware temperature status Host hardware voltage Host IPMI System Event Log status Host memory status Host memory usage Host processor status Host service console swap rates Host storage status Insufficient vSphere HA failover resources License capacity monitoring License error License inventory monitoring



创建报警

警报需要使用触发器。触发器类型:

- 条件或状态触发器 监视当前条件 或状态。例如:
 - ✓ 某个虚拟机当前快照的大小超过 2 GB。
 - ✓ 某个主机使用的内存占其总内存的 90%。
 - ✓ 某个数据存储已经与所有主机断 开连接。
- 事件 监视事件。例如:
 - ✓ 某个主机硬件的运行状况发生了变化。数据中心许可证已过期。
 - ✓ 主机已断开与 vNetwork 分布 式交换机的连接





查看对象事件和任务

