vSphere 资源管理指南

ESX 4.0
ESXi 4.0
vCenter Server 4.0

ZH_CN-000107-00

vmware

您可以在 VMware 的网站上找到最新的技术文档，网址为

http://www.vmware.com/cn/support/

VMware 网站还提供了最新的产品更新。

如果您对本文档有任何意见和建议，请将您的反馈提交到：

docfeedback@vmware.com

VMware, Inc.
3401 Hillview Ave.
Palo Alto, CA 94304
www.vmware.com

北京办公室
北京市海淀区科学院南路 2 号
融科资讯中心 C 座南 8 层
www.vmware.com/cn

上海办公室
上海市浦东新区浦东南路 999 号
新梅联合广场 23 楼
www.vmware.com/cn

广州办公室
广州市天河北路 233 号
中信广场 7401 室
www.vmware.com/cn

目录

关于本文档

《vSphere 资源管理指南》介绍 vSphere® 环境的资源管理。其重点讲述以下几大主题：

- 资源分配和资源管理概念
- 虚拟机属性和接入控制
- 资源池及其管理方式
- 群集、VMware® Distributed Resource Scheduler (DRS)、VMware® Distributed Power Management (DPM) 及其使用方法
- 高级资源管理选项
- 性能注意事项

《vSphere 资源管理指南》涵盖了 ESX®、ESXi 和 vCenter® Server。

目标读者

本手册专供要了解系统如何管理资源以及用户如何自定义默认行为的系统管理员使用。此外，对于要了解和使用资源池、群集、DRS 或 VMware DPM 的用户，本手册亦是必不可少的。

本手册假定您具有 VMware ESX、VMware ESXi 和 vCenter Server 的相关应用知识。

文档反馈

VMware 欢迎您提出宝贵建议，以便改进我们的文档。如有意见，请将反馈发送到 docfeedback@vmware.com。

vSphere 文档

vSphere 文档由 vCenter Server 和 ESX/ESXi 文档集组合而成。

技术支持和教育资源

您可以获取以下技术支持资源。有关本文档和其他文档的最新版本，请访问：
http://www.vmware.com/support/pubs。

在线支持和电话支持 要通过在线支持提交技术支持请求、查看产品和合同信息以及注册您的产品，请访问 http://www.vmware.com/support。

客户只要拥有相应的支持合同，就可以通过电话支持，尽快获得对优先级高的问题的答复。请访问 http://www.vmware.com/support/phone_support.html。

支持服务项目

要了解 VMware 支持服务项目如何帮助您满足业务需求，请访问 http://www.vmware.com/support/services。

VMware 专业服务

VMware 教育服务课程提供了大量实践操作环境、案例研究示例，以及用作作业参考工具的课程材料。这些课程可以通过现场指导、教室授课的方式学习，也可以通过在线直播的方式学习。关于现场试点项目及实施的最佳实践，VMware 咨询服务可提供多种服务，协助您评估、计划、构建和管理虚拟环境。要了解有关教育课程、认证计划和咨询服务的信息，请访问 http://www.vmware.com/services。

1 资源管理入门

要了解资源管理，必须清楚其组件、目标以及如何以最佳方式在群集设置中将其实现。

将讨论虚拟机的资源分配设置（份额、预留和限制），包括如何设置它们并对其进行查看。另外，还将介绍接入控制过程，系统通过该过程对照现有资源对资源分配设置进行验证。

本章讨论了以下主题：

什么是资源管理?

资源管理是将资源从资源提供方分配到资源用户的一个过程。

对于资源管理的需求来自于资源过载（即，需求大于容量）以及需求与容量随着时间的推移而有所差异的事实。通过资源管理，可以动态重新分配资源，以便更高效地使用可用容量。

资源类型

资源包括 CPU、内存、电源、存储器和网络资源。

此上下文中的资源管理着重说明 CPU 和内存资源。使用 VMware® 分布式电源管理 (DPM) 功能还可以减少电源资源的消耗。

注意 ESX/ESXi 分别使用网络流量调整和按比例分配份额机制来管理每台主机上的网络带宽和磁盘资源。

资源提供方

主机和群集是物理资源的提供方。

对于主机，可用的资源是主机的硬件规格减去虚拟化软件所用的资源。

群集是一组主机。可以使用 VMware® vCenter Server 创建群集，并将多个主机添加到群集。vCenter Server 一起管理这些主机的资源：群集拥有所有主机的全部 CPU 和内存。可以针对联合负载平衡或故障切换来启用群集。有关详细信息，请参见第 39 页，第 5 章“创建 DRS 群集”。

资源用户

虚拟机是资源用户。

创建期间分配的默认资源设置适用于大多数计算机。可以在以后编辑虚拟机设置，以便基于份额分配占资源提供方的总 CPU 和内存的百分比，或者分配所保证的 CPU 和内存预留量。启动虚拟机时，服务器检查是否有足够的未预留资源可用，并仅在有足够的资源时才允许启动虚拟机。此过程称为接入控制。

资源池是灵活管理资源的逻辑抽象。资源池可以分组为层次结构，用于对可用的 CPU 和内存资源按层次结构进行分区。相应地，资源池既可以被视为资源提供方，也可以被视为资源用户。它们向子资源池和虚拟机提供资源，但是，由于它们也消耗其父资源池和虚拟机的资源，因此它们同时也是资源用户。请参见第 33 页，第 4 章“管理资源池”。

ESX/ESXi 主机根据以下因素为每个虚拟机分配基础硬件资源的一部分：

- ESX/ESXi 主机（或群集）的可用资源总量。
- 启动的虚拟机数目和这些虚拟机的资源使用情况。
- 管理虚拟化所需的开销。
- 由用户定义的资源限制。

资源管理的目标

管理资源时，应清楚自己的目标。

除了解决资源过载问题，资源管理还可以帮助您实现以下目标：

- 性能隔离 — 防止虚拟机独占资源并保证服务率的可预测性。
- 高效使用 — 利用未过载的资源并在性能正常降低的情况下过载。
- 易于管理 — 控制虚拟机的相对重要性，提供灵活的动态分区并且符合绝对服务级别协议。

配置资源分配设置

当可用资源容量无法满足资源用户（和虚拟化开销）的需求时，管理员可能需要对分配给虚拟机或它们所驻留的资源池的资源量进行自定义。

资源分配设置（份额、预留和限制）用于确定为虚拟机提供的 CPU 和内存资源量。特别是，管理员有多个用于分配资源的选项。

- 预留主机或群集的物理资源。
- 确保 ESX/ESXi 计算机的物理内存提供一定量的虚拟机内存。
- 保证为特定虚拟机分配的物理资源百分比始终高于其他虚拟机。
- 为可以分配给虚拟机的资源量设置上限。

资源分配份额

份额指定虚拟机（或资源池）的相对优先级或重要性。如果某个虚拟机的资源份额是另一个虚拟机的两倍，则在这两个虚拟机争用资源时，第一个虚拟机有权消耗两倍于第二个虚拟机的资源。

份额通常指定为**高**、**正常**或**低**，这些值将分别按 4:2:1 的比例指定份额值。还可以选择**自定义**为各虚拟机分配特定的份额值（表示比例权重）。

指定份额仅对同级虚拟机或资源池（即在资源池层次结构中具有相同父级的虚拟机或资源池）有意义。同级将根据其相对份额值共享资源，该份额值受预留和限制的约束。为虚拟机分配份额时，始终会相对于其他已启动的虚拟机来为该虚拟机指定优先级。

下表显示了虚拟机的默认 CPU 和内存份额值。对于资源池，默认的 CPU 份额值和内存份额值是相同的，但是必须将二者相乘，就好像是资源池是具有四个 VCPU 和 16 GB 内存的虚拟机一样。

表 1–1。 份额值

设置	CPU 份额值	内存份额值
高	每个虚拟 CPU 具有 2000 个份额	所配置的虚拟机内存的每兆字节具有 20 个份额。
正常	每个虚拟 CPU 具有 1000 个份额	所配置的虚拟机内存的每兆字节具有 10 个份额。
低	每个虚拟 CPU 具有 500 个份额	所配置的虚拟机内存的每兆字节具有 5 个份额。

例如，一台具有两个虚拟 CPU 和 1GB RAM 且 CPU 和内存份额设置为**正常**的 SMP 虚拟机具有 2x1000=2000 个 CPU 份额和 10x1024=10240 个内存份额。

注意 具有一个以上虚拟 CPU 的虚拟机称为 SMP（对称多处理）虚拟机。在每个虚拟机上，ESX/ESXi 最多支持八个虚拟 CPU。这也称为 8 路 SMP 支持。

启动新的虚拟机时，每个份额所代表的相对优先级会改变。这将影响同一资源池内的所有虚拟机。所有虚拟机都具有相同数量的 VCPU。请考虑以下示例。

- 一台聚合 CPU 容量为 8 GHz 的主机上运行着两个受 CPU 约束的虚拟机。它们的 CPU 份额设置为**正常**，因此各得 4GHz。
- 现在启动了第三个受 CPU 约束的虚拟机。它的 CPU 份额设置为**高**，这意味着它拥有的份额值应该是设置为**正常**的虚拟机的两倍。新的虚拟机获得 4GHz，其他两个虚拟机各自仅获得 2GHz。如果用户为第三个虚拟机指定的自定义份额值为 2000，也会出现相同的结果。

资源分配预留

预留指定保证为虚拟机分配的最少资源量。

仅在有足够的未预留资源满足虚拟机的预留时，vCenter Server 或 ESX/ESXi 才允许您启动虚拟机。即使物理服务器负载较重，服务器也会确保该资源量。预留用具体单位（兆赫兹 (GHz) 或兆字节 (MB)）表示。

例如，假定您有 2GHz 可用，并且为 VM1 和 VM2 各指定了 1GHz 的预留量。现在每个虚拟机都能保证在需要时获得 1GHz。但是，如果 VM1 只用了 500MHz，则 VM2 可使用 1.5GHz。

预留默认为 0。可以指定预留以保证虚拟机始终可使用最少的必要 CPU 或内存量。

资源分配限制

限制功能为可以分配到虚拟机的 CPU 或内存资源量指定上限。

服务器分配给虚拟机的资源可大于预留，但决不可大于限制，即使系统上有尚未利用的 CPU 或内存也是如此。限制用具体单位（兆赫兹 (GHz) 或兆字节 (MB)）表示。

CPU 和内存限制默认为无限。在大多数情况下，当内存限制为无限时，创建虚拟机时为其配置的内存量会成为其有效限制。

多数情况下无需指定限制。指定限制的优缺点如下：

- 优点 — 如果开始时虚拟机的数量较少，并且您想对用户期望数量的虚拟机进行管理，则分配一个限制将非常有效。但随着用户添加的虚拟机数量增加，性能将会降低。因此，您可以通过指定限制来模拟减少可用资源。
- 缺点 — 如果指定限制，可能会浪费闲置资源。系统不允许虚拟机使用的资源超过限制，即使系统未充分利用并且有闲置资源可用时也是如此。请仅在有充分理由的情况下指定限制。

资源分配设置建议

选择适合 ESX/ESXi 环境的资源分配设置（份额、预留和限制）。

遵循以下准则有助于使虚拟机获得更好性能。

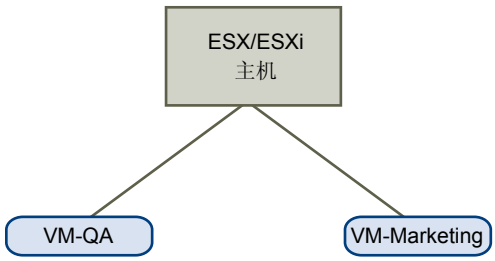
- 如需频繁更改总可用资源，可使用**份额**在虚拟机之间合理分配资源。例如，如果使用**份额**，并且升级主机，那么，即使每个份额代表较大的内存量或 CPU 量，每个虚拟机也保持相同的优先级（保持相同数量的份额）。
- 使用**预留**来指定可接受的最低 CPU 量或内存量，而不是想要使用的量。主机可以根据份额的数量、估计需求和虚拟机的限制将额外的资源指定为可用资源。预留表示的具体资源量不会随环境改变（例如添加或移除虚拟机）而变化。
- 请不要将所有资源全部指定为虚拟机的预留（请计划将至少 10% 的资源保留为未预留）。系统容量越接近于被全部预留，想要在不违反接入控制的情况下更改预留和资源池层次结构就越困难。在支持 DRS 的群集内，如果预留完全占用群集或群集内各台主机的容量，则会阻止 DRS 在主机之间迁移虚拟机。

更改资源分配设置 — 示例

以下示例说明了如何更改资源分配设置以提高虚拟机性能。

假定在某台 ESX/ESXi 主机上，您创建了两台新的虚拟机 — 一台用于 QA (VM-QA) 部门，另一台用于市场 (VM-Marketing) 部门。

图 1–1。 具有两个虚拟机的单台主机



在接下来的示例中，假定 VM-QA 占用大量内存，因此，您需要将这两个虚拟机的资源分配设置相应地更改为以下内容：

- 指定当系统内存过载时，VM-QA 可使用的内存和 CPU 量是市场部虚拟机的两倍。将 VM-QA 的内存份额和 CPU 份额设置为**高**，并将 VM-Marketing 设置为**正常**。
- 保证市场部虚拟机具有一定量的 CPU 资源。您可以使用预留设置来达到此目的。

步骤

1 启动 vSphere Client 并连接到 vCenter Server。

2 在要更改其份额的虚拟机上，右键单击 **VM-QA**，然后选择**编辑设置**。

3 选择**资源**，并在 CPU 面板的**份额**下拉菜单中选择**高**。

4 在“内存”面板的**份额**下拉菜单中选择**高**。

5 单击**确定**。

6 右键单击市场部虚拟机 (**VM-Marketing**)，然后选择**编辑设置**。

7 在 CPU 面板中，将**预留**字段中的值更改为所需值。

8 单击**确定**。

如果选择群集的**资源分配**选项卡，然后单击 **CPU**，此时应看到 **VM-QA** 的份额是另一虚拟机的两倍。另外，由于虚拟机尚未启动，因此**使用的预留**字段尚未改变。

查看资源分配信息

使用 vSphere Client，可以在“清单”面板中选择群集、资源池、独立主机或虚拟机，并通过单击**资源分配**选项卡来查看其资源的分配方式。

此信息可在以后用于帮助通知您的资源管理决定。

群集资源分配选项卡

资源分配选项卡在从清单面板中选择群集时可用。

资源分配选项卡显示有关群集中 CPU 和内存资源的信息。

CPU 区域

下面显示的是有关 CPU 资源分配的信息：

表 1–2。 CPU 资源分配

字段	描述
总容量	保证为该对象预留的 CPU 分配，以兆赫兹 (MHz) 为单位。
预留的容量	此对象正在使用的预留分配的兆赫兹 (MHz) 数。
可用容量	未预留的兆赫兹 (MHz) 数。

内存区域

下面显示的是有关内存资源分配的信息：

表 1–3。 内存资源分配

字段	描述
总容量	保证为该对象分配的内存量，以兆字节 (MB) 为单位。
预留的容量	此对象正在使用的预留分配的兆字节 (MB) 数。
开销预留	为虚拟开销预留的“预留的容量”字段的量。
可用容量	未预留的兆字节 (MB) 数。

注意 为 VMware HA 启用的群集的根资源池预留量可能大于群集内显式使用的资源总和。这些预留量不仅反映了正在运行的虚拟机的预留量和群集内以分层方式包含的（子）资源池的预留量，还反映了支持 VMware HA 故障切换所需的预留。请参见《vSphere 可用性指南》。

资源分配选项卡还会显示一个图表，该图表显示 DRS 群集内资源池和虚拟机的下列 CPU 或内存的使用情况信息。要查看 CPU 或内存信息，请分别单击 **CPU** 按钮或**内存**按钮。

表 1–4。 CPU 或内存使用情况信息

字段	描述
名称	对象的名称。
预留 - MHz	保证为该对象预留的最小 CPU 分配，以兆赫兹 (MHz) 为单位。
预留 - MB	保证为该对象预留的最小内存分配，以兆字节 (MB) 为单位。
限制 - MHz	对象可以使用的最大 CPU 量。
限制 - MB	对象可以使用的最大内存量。

表 1–4。 CPU 或内存使用情况信息 （续）

字段	描述
份额	用来分配 CPU 或内存容量的相对衡量指标。将“低”、“正常”、“高”及“自定义”值与所属资源池中的所有虚拟机的所有份额之和进行比较。
份额值	基于资源和对象设置的实际值。
% 份额	分配到该对象的群集资源百分比。
最坏情况分配	根据用户配置的资源分配策略（例如预留、份额和限制）分配给虚拟机的（CPU 或内存）资源量，并假定群集内的所有虚拟机将会完全消耗已分配的资源。此字段的值必须通过按 F5 键手动更新。
类型	预留的 CPU 或内存分配类型：“可扩展的”或“固定的”。

虚拟机资源分配选项卡

在“清单”面板中选择虚拟机后，**资源分配**选项卡可用。

此**资源分配**选项卡显示有关所选虚拟机的 CPU 和内存资源的信息。

CPU 区域

这些条显示有关主机 CPU 使用情况的下列信息：

表 1–5。 主机 CPU

字段	描述
已消耗	虚拟机实际消耗的 CPU 资源量。
活动	在没有资源争用情况下，虚拟机消耗的资源的估计量。如果已设置明确限制，则此数值不会超过该限制。

表 1–6。 资源设置

字段	描述
预留	保证为该虚拟机分配的最小 CPU 量。
限制	可为此虚拟机分配的最大 CPU 量。
份额	此虚拟机的 CPU 份额。
最坏情况分配	根据用户配置的资源分配策略（例如预留、份额和限制）分配给虚拟机的（CPU 或内存）资源量，并假定群集内的所有虚拟机将会完全消耗已分配的资源。

内存区域

这些条显示有关主机内存使用情况的下列信息：

表 1–7。 主机内存

字段	描述
已消耗	已分配给虚拟机的物理内存的实际消耗量。
开销消耗	用于虚拟化目的的已消耗内存量。“已消耗”中显示的量包括开销消耗。

这些条显示有关客户机内存使用情况的下列信息：

表 1–8。 客户机内存

字段	描述
专用	受主机内存支持且没有共享的内存量。
已共享	共享的内存量。
已交换	通过交换回收的内存量。
虚拟增长	通过虚拟增长回收的内存量。
未访问过	客户机从未访问过的内存量。
活动	最近访问过的内存量。

表 1–9。 资源设置

字段	描述
预留	保证为该虚拟机分配的内存量。
限制	该虚拟机的内存分配上限。
份额	此虚拟机的内存份额。
已配置	用户指定的客户机物理内存大小。
最坏情况分配	根据用户配置的资源分配策略（例如预留、份额和限制）分配给虚拟机的（CPU 或内存）资源量，并假定群集内的所有虚拟机将会完全消耗已分配的资源。
开销预留	为虚拟化开销预留的内存量。

接入控制

启动虚拟机时，系统会检查尚未预留的 CPU 和内存资源量。系统将根据可用的未预留资源确定是否可保证为虚拟机所配置的预留（如果有）。此过程称为接入控制。

如果有足够的未预留 CPU 和内存可用，或者没有预留，虚拟机将启动。否则将显示一条资源不足警告。

注意 除用户指定的内存预留外，各虚拟机还有一个开销内存量。此额外内存使用量包含在接入控制计算中。

启用了 VMware DPM 功能时，可能会将主机置于待机模式（即将其关闭）以降低功耗。这些主机所提供的未预留资源将被视为可用于接入控制的资源。如果某个虚拟机没有这些资源就无法启动，系统会建议启动足够的待机主机。

管理 CPU 资源 2

ESX/ESXi 主机支持 CPU 虚拟化。

利用 CPU 虚拟化时，应当了解其工作方式、不同类型以及特定于处理器的行为。此外，还需要了解 CPU 虚拟化的性能影响。

本章讨论了以下主题：

CPU 虚拟化基本知识

CPU 虚拟化着重于性能，只要有可能就会直接在处理器上运行。只要有可能就会使用基础物理资源，且虚拟化层仅在需要时才运行指令，使得虚拟机就像直接在物理机上运行一样。

CPU 虚拟化与仿真不同。采用仿真时，所有操作均由仿真器在软件中运行。软件仿真器允许程序在不同于最初编写时所针对的计算机系统上运行。仿真器通过接受相同的数据或输入并获得相同的结果，来模拟或再现原始计算机的行为，从而实现仿真。仿真提供了可移植能力，并在几个不同平台上运行针对一个平台而设计的软件。

CPU 资源过载时，ESX/ESXi 主机将在所有虚拟机之间对物理处理器进行时间划分，以便每个虚拟机在运行时就如同具有指定数目的虚拟处理器一样。运行多个虚拟机的 ESX/ESXi 主机会为各虚拟机分配一定份额的物理资源。如果使用默认资源分配设置，与同一主机关联的所有虚拟机都将在每个虚拟 CPU 上收到相同份额的 CPU。这意味着单处理器虚拟机分配到的资源只有双处理器虚拟机的一半。

基于软件的 CPU 虚拟化

采用基于软件的 CPU 虚拟化后，客户机应用程序代码直接在处理器上运行，同时转换客户机特权代码并在处理器上运行该代码。

转换后的代码有点大，比本机版本的执行速度通常要慢。因此，具有少量特权代码组件的客户机程序的运行速度与本机程序非常接近。而具有大量特权代码组件（如系统调用、陷阱或页面表更新）的程序在虚拟环境中的运行速度可能较慢。

硬件辅助的 CPU 虚拟化

某些处理器（例如 Intel VT 和 AMD SVM）为 CPU 虚拟化提供了硬件辅助。

使用此辅助时，客户机可以使用独立的执行模式（称为客户机模式）。应用程序代码或特权代码等客户机代码均在客户机模式中运行。出现某些事件时，处理器退出客户机模式而进入根模式。管理程序将在根模式中执行，确定退出的原因，采取任何必需的措施，并在客户机模式中重新启动客户机。

将硬件辅助用于虚拟化时，不需要再转换代码。因此，系统调用或陷阱密集型工作负载在运行时的速度非常接近本机速度。但是，诸如涉及更新页面表之类的一些工作负载会导致多次退出客户机模式而进入根模式。根据退出的次数和退出所用的总时间，这可能会明显降低执行的速度。

虚拟化和特定于处理器的行为

尽管 VMware 软件会虚拟化 CPU，虚拟机仍然能检测出它在其上运行的处理器的具体型号。

处理器型号可能在其提供的 CPU 功能方面不同，在虚拟机中运行的应用程序可以利用这些功能。因此，无法使用 VMotion® 在具有不同功能集的处理器上运行的系统之间迁移虚拟机。在某些情况下，通过将增强型 VMotion 兼容性 (EVC) 用于支持此功能的处理器，可以避免此限制。有关详细信息，请参见《基本系统管理》。

CPU 虚拟化的性能影响

根据工作负载和使用的虚拟化类型，CPU 虚拟化会增加不同的开销量。

如果应用程序的大多数时间用于执行指令而不是等待用户交互、设备输入或数据检索等外部事件，则应用程序是受 CPU 约束的。对于此类应用程序，CPU 虚拟化开销包括必须执行的额外指令。此开销消耗应用程序本身可以使用的 CPU 处理时间。CPU 虚拟化开销通常会导致整体性能下降。

对于不受 CPU 约束的应用程序，CPU 虚拟化可能会提高 CPU 利用率。如果备用 CPU 容量可用于吸收开销，则仍然可以在整体吞吐量方面提供不错的性能。

在每个虚拟机上，ESX/ESXi 最多支持八个虚拟处理器 (CPU)。

注意 在单处理器虚拟机（而不是 SMP 虚拟机）上部署单线程应用程序可获得最佳的性能和资源利用率。

单线程应用程序只能利用单个 CPU。在双处理器虚拟机中部署这些应用程序不会加快应用程序的速度。相反，这样会使得第二个虚拟 CPU 使用本该由其他虚拟机以其他方式使用的物理资源。

管理 CPU 资源

可以为虚拟机配置一个或多个虚拟处理器，每个处理器均具有自己的寄存器和控制结构集合。

当调度虚拟机时，会调度其虚拟处理器在物理处理器上运行。VMkernel 资源管理器在物理 CPU 上调度虚拟 CPU，从而管理虚拟机对物理 CPU 资源的访问。ESX/ESXi 支持最多具有八个虚拟处理器的虚拟机。

查看处理器信息

可以通过 vSphere Client 或使用 vSphere SDK 访问有关当前 CPU 配置的信息。

步骤

1 在 vSphere Client 中，选择主机，然后单击**配置**选项卡。

2 选择**处理器**。

可以查看有关物理处理器数量和类型以及逻辑处理器数量的信息。

注意 在超线程系统中，每个硬件线程都是一个逻辑处理器。例如，启用了超线程的双核处理器具有两个内核和四个逻辑处理器。

3 （可选）还可以通过单击**属性**禁用或启用超线程。

指定 CPU 配置

可以通过指定 CPU 配置来改进资源管理。但是，如果未自定义 CPU 配置，则 ESX/ESXi 主机会使用适合大多数情况的默认值。

可以按以下方式指定 CPU 配置：

- 使用可通过 vSphere Client 访问的属性和特殊功能。使用 vSphere Client 图形用户界面 (GUI) 可以连接到 ESX/ESXi 主机或 vCenter Server 系统。
- 在某些情况下使用高级设置。
- 将 vSphere SDK 用于脚本式 CPU 分配。
- 使用超线程。

多核处理器

多核处理器为执行虚拟机多任务的 ESX/ESXi 主机提供了很多优势。

Intel 和 AMD 均已开发了将两个或两个以上处理器内核组合到单个集成电路（通常称为封装件或插槽）的处理器。VMware 使用“插槽”一词来描述单个封装件，该封装件可以具有一个或多个处理器内核且每个内核具有一个或多个逻辑处理器。

例如，双核处理器通过允许同时执行两个虚拟 CPU，可以提供几乎是单核处理器两倍的性能。同一处理器中的内核通常配备由所有内核使用的最低级别的共享缓存，这有可能会减少访问较慢主内存的必要性。如果运行在逻辑处理器上的虚拟机正运行争用相同内存总线资源且占用大量内存的工作负载，则将物理处理器连接到主内存的共享内存总线可能会限制其逻辑处理器的性能。

ESX CPU 调度程序可以独立将每个处理器内核的每个逻辑处理器用于执行虚拟机，从而提供与 SMP 系统类似的功能。例如，2 路虚拟机可以让虚拟处理器运行在属于相同内核的逻辑处理器上，或运行在不同物理内核的逻辑处理器上。

ESX CPU 调度程序可以检测处理器拓扑，以及处理器内核与它上面的逻辑处理器之间的关系。它使用此信息来调度虚拟机和优化性能。

ESX CPU 调度程序可以解释处理器拓扑（包括插槽、内核和逻辑处理器之间的关系）。调度程序使用拓扑信息优化虚拟 CPU 在不同插槽上的放置位置，以最大化总体的缓存利用率，并通过最小化虚拟 CPU 迁移来改善缓存关联性。

在未过载的系统中，ESX CPU 调度程序在默认情况下将负载分配到所有插槽。这样便可通过最大化可供正在运行的虚拟 CPU 使用的缓存总量来改善性能。因此，单个 SMP 虚拟机的虚拟 CPU 在多个插槽之间分配（除非每个插槽本身还是 NUMA 节点，在这种情况下，NUMA 调度程序会限制虚拟机的所有虚拟 CPU 都驻留在同一插槽上。）

但是，在某些情况下（例如，当 SMP 虚拟机显示出其虚拟 CPU 之间存在大量数据共享时），此默认行为可能不是最佳选择。对于此类工作负载，最好是调度相同插槽（具有最低级别的共享缓存）上的所有虚拟 CPU，即使 ESX/ESXi 主机未过载也是如此。在这些情况中，通过将以下配置选项包括在虚拟机的 `.vmx` 配置文件中，可以替代在封装件之间分配虚拟 CPU 的默认行为：`sched.cpu.vsmpConsolidate="TRUE"`。

超线程

超线程技术允许单个物理处理器内核像两个逻辑处理器一样工作。处理器可以同时运行两个独立的应用程序。为了避免将逻辑处理器和物理处理器混淆，Intel 将物理处理器称为插槽，本章的讨论也使用这一术语。

Intel Corporation 开发了超线程技术来增强 Pentium IV 和 Xeon 处理器系列的性能。超线程技术允许单个处理器内核同时执行两个独立的线程。

虽然超线程不会使系统的性能加倍，但是它可以通过更好地利用空闲资源来提高性能，使得某些重要的工作负载类型产生更大的吞吐量。如果应用程序运行在忙碌内核的一个逻辑处理器上，则与单独运行在非超线程处理器上相比，预期获得的吞吐量会稍高于一半。超线程性能改进情况与应用程序有很大关系，有些应用程序使用超线程可能会出现性能下降的情况，因为两个逻辑处理器之间会共享许多处理器资源（例如缓存）。

注意 在具有 Intel 超线程技术的处理器上，每个内核可以具有两个逻辑处理器，这两个逻辑处理器共享大多数内核资源（如内存缓存和功能单元）。此类逻辑处理器通常称为线程。

许多处理器都不支持超线程，因此每个内核仅具有一个线程。对于此类处理器，内核数目还与逻辑处理器的数目相匹配。以下处理器支持超线程，并且每个内核具有两个线程。

- 基于 Intel Xeon 5500 处理器微架构的处理器。
- Intel Pentium 4（支持 HT）
- Intel Pentium EE 840（支持 HT）

超线程和 ESX/ESXi 主机

支持超线程的 ESX/ESXi 主机应具有与没有超线程的主机类似的行为。但是，如果启用超线程，则可能需要考虑某些因素。

ESX/ESXi 主机以智能方式管理处理器时间，保证负载均匀分布在系统的多个处理器内核上。相同内核上的逻辑处理器具有连续的 CPU 编号，因此 CPU 0 和 1 一起在第一个内核上，而 CPU 2 和 3 在第二个内核上，依此类推。优先在两个不同的内核上调度虚拟机，然后才选择在同一内核的两个逻辑处理器上调度虚拟机。

如果逻辑处理器没有工作，则将其置于暂停状况，从而释放其执行资源并允许在同一内核的另一个逻辑处理器上运行的虚拟机使用该内核的全部执行资源。VMware 调度程序会正确地考虑此暂停时间，因此使用全部内核资源运行的虚拟机的效率要高于在半个内核上运行的虚拟机。按这种方法管理处理器可确保服务器不会违反任何标准的 ESX/ESXi 资源分配规则。

在使用超线程的主机上启用 CPU 关联性之前，请考虑资源管理需求。例如，如果将高优先级虚拟机绑定到 CPU 0，并将另一个高优先级虚拟机绑定到 CPU 1，则这两个虚拟机必须共享相同的物理内核。这种情况下，可能无法满足这些虚拟机的资源需求。请确保所有的自定义关联性设置对超线程系统都有意义。

启用超线程

要启用超线程，必须首先在系统的 BIOS 设置中将其启用，然后在 vSphere Client 中打开它。超线程在默认情况下处于启用状态。

部分 Intel 处理器（如 Xeon 5500 处理器或基于 P4 微架构的处理器）支持超线程。请查阅系统文档，确定您的 CPU 是否支持超线程。ESX/ESXi 无法在具有 32 个以上物理内核的系统上启用超线程，因为 ESX/ESXi 的逻辑限制是 64 个 CPU。

步骤

1 请确保您的系统支持超线程技术。

2 在系统 BIOS 中启用超线程。

有些制造商将该选项标记为**逻辑处理器**，而有些制造商则称之为**启用超线程**。

3 确保为 ESX/ESXi 主机打开了超线程。

a 在 vSphere Client 中，选择主机，然后单击**配置**选项卡。

b 选择**处理器**并单击**属性**。

c 在该对话框中，可以查看超线程状态，还可以开启（默认）或关闭超线程。

现在，超线程处于启用状态。

为虚拟机设置超线程共享选项

可以指定虚拟机的虚拟 CPU 如何在超线程系统上共享物理内核。

如果两个虚拟 CPU 同时在内核的逻辑 CPU 上运行，则这两个虚拟 CPU 共享内核。可以为各个虚拟机设置此选项。

步骤

1 在 vSphere Client“清单”面板中，右键单击虚拟机并选择**编辑设置**。

2 单击**资源**选项卡，然后单击**高级 CPU**。

3 从**模式**下拉菜单选择此虚拟机的超线程模式。

超线程内核共享选项

可以使用 vSphere Client 为虚拟机设置超线程内核共享模式。

此模式提供以下选项。

表 2–1。 超线程内核共享模式

选项	描述
任意	超线程系统上所有虚拟机的默认值。具有该设置的虚拟机的虚拟 CPU 可与该虚拟机或任何其他虚拟机的其他虚拟 CPU 随时共享内核。
无	虚拟机的虚拟 CPU 不应彼此共享内核，或不应与其他虚拟机的虚拟 CPU 共享内核。即，该虚拟机的每个虚拟 CPU 本身始终应获得完整的内核，而该内核上的另一个逻辑 CPU 则置于暂停状况。
内部	该选项类似于“无”。该虚拟机的虚拟 CPU 不能与其他虚拟机的虚拟 CPU 共享内核。这些虚拟 CPU 可以与同一虚拟机的其他虚拟 CPU 共享内核。 只能为 SMP 虚拟机选择此选项。如果应用于单处理器虚拟机，则系统会将该选项更改为“无”。

这些选项不会影响公平性或 CPU 时间分配。无论虚拟机的超线程设置如何，它仍然会得到与 CPU 份额成比例的 CPU 时间，且会受到 CPU 预留和 CPU 限制值的约束。

对于典型的工作负载，自定义超线程设置并非必要设置。对于与超线程交互不良的非常见工作负载，该选项很有用。例如，具有缓存颠簸问题的应用程序可能会让共享其物理内核的应用程序降低速度。可以将运行该应用程序的虚拟机置于“无”或“内部”超线程状态，以将其与其他虚拟机隔离开。

如果虚拟 CPU 具有超线程限制，不允许该虚拟 CPU 与其他虚拟 CPU 共享内核，那么，当其他虚拟 CPU 有资格消耗处理器时间时，系统可能取消对该虚拟 CPU 的调度。如果没有超线程限制，则可以在同一内核上调度这两个虚拟 CPU。

对于（每个虚拟机）内核数有限的系统，问题会变得更糟。这些情况下，可能没有内核来让取消调度的虚拟机进行迁移。因此，超线程设置为“无”或“内部”的虚拟机性能可能会降低，这一点对于内核数有限的系统而言尤其明显。

隔离

在某些极少数情况下，ESX/ESXi 主机可能会检测到应用程序正在与 Pentium IV 超线程技术（不适用于基于 Intel Xeon 5500 处理器微架构的系统）进行不良交互。在这种情况下，对用户透明的隔离可能是必要的。

例如，对于与问题代码共享一个内核的应用程序，某些类型的自修改代码可能中断 Pentium IV 跟踪缓存的正常行为，导致速度显著降低（最多 90%）。在这些情况下，ESX/ESXi 主机隔离运行该代码的虚拟 CPU，并将其虚拟机相应地置于“无”或“内部”模式。

将主机的“Cpu.MachineClearThreshold”高级设置配置为“0”可禁用隔离。

使用 CPU 关联性

通过为每个虚拟机指定 CPU 关联性设置，可以仅将虚拟机只分配给多处理器系统中的某个可用处理器子集。通过使用此功能，可以将每个虚拟机分配到指定关联性集合中的处理器。

在这个上下文中，术语“CPU”指的是超线程系统上的逻辑处理器，而不是非超线程系统上的内核。

某个虚拟机的 CPU 关联性设置不仅应用到与该虚拟机关联的所有虚拟 CPU，还会应用到与该虚拟机关联的所有其他线程（也称为“环境”）。这些虚拟机线程执行仿真鼠标、键盘、屏幕、CD-ROM 和其他老设备所需的处理。

在某些情况下（例如，占用大量显示资源的工作负载），可能会在虚拟 CPU 和其他虚拟机线程之间出现大量通信。如果虚拟机的关联性设置阻止这些额外的线程同时由虚拟机的虚拟 CPU 调度（例如，单处理器虚拟机与单个 CPU 关联，或 2 路 SMP 虚拟机仅与两个 CPU 关联），则性能可能会降低。

为了获得最佳性能，在使用手动关联性设置时，VMware 建议在关联性设置中至少包括一个额外的物理 CPU，以允许至少有虚拟机的一个线程与其虚拟 CPU 同时调度（例如，单处理器虚拟机至少与两个 CPU 关联，或 2 路 SMP 虚拟机至少与三个 CPU 关联）。

注意 CPU 关联性指定虚拟机到处理器的放置位置的限制，与基于 DRS 规则的关联性不同，基于 DRS 规则的关联性指定虚拟机到虚拟机主机的放置位置的限制。

向特定处理器分配虚拟机

使用 CPU 关联性，可以向特定处理器分配虚拟机。通过此操作，可以将虚拟机只分配给多处理器系统中特定的可用处理器。

步骤

1 在 vSphere Client“清单”面板中，选择一个虚拟机并选择**编辑设置**。

2 选择**资源**选项卡，然后选择**高级 CPU**。

3 单击**在处理器上运行**按钮。

4 选择要在骑上运行虚拟机的处理器，然后单击**确定**。

CPU 关联性的潜在问题

使用 CPU 关联性之前，可能需要考虑某些问题。

CPU 关联性的潜在问题包括：

- 对于多处理器系统，ESX/ESXi 系统执行自动负载平衡。避免手动指定虚拟机关联性，以改进调度程序跨处理器平衡负载的能力。
- 关联性可能会干扰 ESX/ESXi 主机满足为虚拟机指定的预留和份额的能力。
- 因为 CPU 接入控制不考虑关联性，所以具有手动关联性设置的虚拟机可能不会始终得到其完整的预留量。

  没有手动关联性设置的虚拟机不会受到具有手动关联性设置的虚拟机的负面影响。
- 将虚拟机从一个主机移动到另一个主机时，因为新的主机可能具有不同的处理器数，所以关联性可能不再适用。
- NUMA 调度程序可能无法管理已经借助于关联性分配到某些处理器的虚拟机。
- 关联性可能会影响 ESX/ESXi 主机在多核或超线程处理器上调度虚拟机以充分利用在这些处理器上共享资源的能力。

CPU 电源管理

要改进 CPU 电源效率，可以将 ESX/ESXi 主机配置为根据工作负载需求来动态切换 CPU 频率。这种类型的电源管理称为动态电压和频率缩放 (DVFS)。它使用 VMkernel 通过 ACPI 接口使用的处理器性能状况（P-状况）。

ESX/ESXi 支持增强型 Intel SpeedStep 和增强型 AMD PowerNow! CPU 电源管理技术。为了让 VMkernel 利用这些技术所提供的电源管理功能，可能需要首先在 BIOS 中启用电源管理（有时称为“基于需求的切换” (DBS)）。

要设置 CPU 电源管理策略，请使用高级主机属性 Power.CpuPolicy。此属性设置保存在主机配置中，可以在引导时再次使用，您可以随时更改它，而无需重新引导服务器。可以将此属性设置为以下值。

静态 默认值。VMkernel 可以检测到主机上可用的电源管理功能，但不主动使用它们，除非 BIOS 由于电源上限或热事件提出了请求。

动态 VMkernel 优化每个 CPU 的频率以满足需求，从而提高电源效率，但不影响性能。当 CPU 需求增加时，此策略设置可确保 CPU 频率同样增加。

管理内存资源 3

所有现代的操作系统均提供对虚拟内存的支持，并允许软件使用的内存要多于计算机实际拥有的内存。同样，ESX/ESXi 管理程序提供对过载虚拟机内存的支持，所有为虚拟机配置的客户机内存量可能大于物理主机内存量。

如果要使用内存虚拟化，则应当了解 ESX/ESXi 主机分配、消耗和回收内存的方式。此外，还需要了解虚拟机引起的内存开销。

本章讨论了以下主题：

内存虚拟化基本知识

在管理内存资源之前，应当了解 ESX/ESXi 是如何虚拟化和使用这些内存资源的。

VMkernel 管理所有的计算机内存。（一种例外情况是在 ESX 中分配给服务控制台的内存。）VMkernel 会将这种受管计算机内存的一部分拿来自己使用。剩余的内存可供虚拟机使用。虚拟机将计算机内存用于两个用途：每个虚拟机均需要有自己的内存，且 VMM 需要一些内存和动态开销内存用于其代码和数据。

虚拟内存空间划分为块，每个块通常为 4 KB，块也称为页。物理内存也划分为块，每个块通常也是 4 KB。当物理内存占满时，不在物理内存中的虚拟页的数据将存储到磁盘上。ESX/ESXi 还提供对大页 (2 MB) 的支持。请参见第 86 页，“高级内存属性”。

虚拟机内存

每个虚拟机均会根据其配置大小消耗内存，还会消耗额外开销内存以用于虚拟化。

配置大小

配置大小是一种由虚拟机的虚拟化层来维持的构造。它是提供给客户机操作系统的内存量，但独立于分配给虚拟机的物理 RAM 量，这取决于下文所述的资源设置（份额、预留和限制）。

例如，请考虑配置大小为 1GB 的虚拟机。当客户机操作系统引导时，系统会检测到它正运行在具有 1 GB 物理内存的专用计算机上。分配给虚拟机的物理主机内存的实际数量取决于其内存资源设置和 ESX/ESXi 主机的内存争用情况。有些情况下，可能向虚拟机分配全部内容（即 1GB）。在其他情况下，可能会得到较小的分配量。无论实际分配如何，客户机操作系统都会继续运行，就好像正运行在具有 1 GB 物理内存的专用计算机上一样。

份额	如果可用量超过预留，则会为虚拟机指定相对优先级。
预留	主机保证为虚拟机预留的物理内存量下限，即使内存过载也是如此。将预留设置为确保虚拟机高效运行的足够内存水平，这样就不会有过多的内存分页。 在虚拟机访问了其全部预留后，会允许其保留该内存量，并且不会将其回收，即使该虚拟机闲置也是如此。例如，某些客户机操作系统（例如 Linux）在引导之后可能不会立即访问所配置的全部内存。在虚拟机访问其全部预留之前，VMkernel 可以将其预留的任何未使用部分分配给其他虚拟机。但是，在客户机的工作负载增加并消耗其全部预留之后，允许其保留此内存。
限制	主机可分配给虚拟机的物理内存量的上限。虚拟机的内存分配还受其配置大小的隐式限制。 开销内存包括为虚拟机框架缓冲区和各种虚拟化数据结构预留的空间。

内存过载

对于每个正在运行的虚拟机，系统会为虚拟机的预留（如果有）和虚拟化开销预留物理内存。

由于 ESX/ESXi 主机使用内存管理技术，因此虚拟机可以使用的内存大于物理机（主机）可用的内存。例如，您有一个内存为 2 GB 的主机，其上运行四个虚拟机，每个虚拟机的内存为 1 GB。这种情况下，内存会过载。

过载有一定的意义，因为通常情况下有些虚拟机负载较轻，而有些虚拟机负载较重，相对活动水平会随着时间的推移而有所差异。

为了改善内存利用率，ESX/ESXi 主机将闲置虚拟机的内存转移给需要更多内存的虚拟机。使用“预留”或“份额”参数可优先向重要的虚拟机分配内存。如果这部分内存未使用，可以用于其他虚拟机。

内存共享

许多工作负载存在跨虚拟机共享内存的机会。

例如，几个虚拟机可能正在运行同一客户机操作系统的多个实例，加载了相同的应用程序或组件，或包含公用数据。ESX/ESXi 系统使用专用的分页共享技术安全地消除了内存页的冗余副本。

采用内存共享，由多个虚拟机组成的工作负载消耗的内存通常要少于其在物理机上运行时所需的内存。因此，系统可以高效地支持更高级别的过载。

内存共享保存的内存量取决于工作负载特性。许多几乎相同的虚拟机的工作负载可能释放 30% 以上的内存，而有较大差异的工作负载可以节省的内存少于 5%。

基于软件的内存虚拟化

ESX/ESXi 通过添加附加级别的地址转换来虚拟化客户机物理内存。

- 每个虚拟机的 VMM 保持了从客户机操作系统的物理内存页到基础计算机上物理内存页的映射。（VMware 将基础主机物理页称为“计算机”页，将客户机操作系统的物理页称为“物理”页。）

  每个虚拟机均有连续的可寻址物理内存空间，该空间从零开始。每个虚拟机使用的服务器上的基础计算机内存不一定是连续的。

- VMM 侦听对客户机操作系统内存管理结构进行操作的虚拟机指令，以便虚拟机不会直接更新处理器上的实际内存管理单元 (MMU)。
- ESX/ESXi 主机将虚拟-计算机页映射保持在卷影页表中，该表与 VMM 所维护的物理-计算机映射保持同步。
- 卷影页表由处理器的分页硬件直接使用。

这种地址转换方法允许在设置卷影页表之后，执行虚拟机中的正常内存访问，而不会增加地址转换开销。因为处理器上的转换旁视缓冲区 (TLB) 缓存从卷影页表中读取的直接虚拟-计算机映射，所以 VMM 访问内存时不会增加额外开销。

性能注意事项

使用两个页表具有以下性能影响。

- 对于常规客户机内存访问不会产生开销。
- 在虚拟机中映射内存需要额外时间，这可能意味着：
  - 虚拟机操作系统正在设置或更新虚拟地址到物理地址的映射。
  - 虚拟机操作系统从一个地址空间切换到另一个地址空间（上下文切换）。
- 与 CPU 虚拟化相似，内存虚拟化开销取决于工作负载。

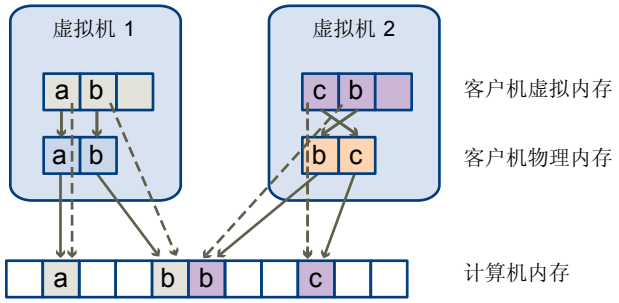
硬件辅助的内存虚拟化

类似于 AMD SVM-V 和 Intel Xeon 5500 系列之类的部分 CPU 通过使用两层页表来提供对内存虚拟化的硬件支持。

第一层页表存储客户机虚拟-物理转换，而第二层页表存储客户机物理-计算机转换。TLB（translation look-aside buffer，转换旁视缓冲区）是由处理器的内存管理单元 (MMU) 硬件维护的转换缓存。TLB 缺失是此缓存中的缺失，而且硬件需要访问内存（可能是多次）来查找所需转换。如果 TLB 中没有某个客户机虚拟地址，则硬件会查看这两个页表，将客户机虚拟地址转换成主机物理地址。

图 3–1 中的插图说明了 ESX/ESXi 如何实施内存虚拟化。

图 3–1。 ESX/ESXi 内存映射



- 方框表示页，而箭头表示不同的内存映射。
- 从客户机虚拟内存到客户机物理内存的箭头表示客户机操作系统中的页表所保持的映射。（未显示 x86 架构处理器从虚拟内存到线性内存的映射。）
- 从客户机物理内存到计算机内存的箭头表示由 VMM 保持的映射。
- 虚线箭头表示从客户机虚拟内存到计算机内存的映射，该映射也由 VMM 保持。运行虚拟机的基础处理器使用卷影页表映射。

因为虚拟化引入了额外级别的内存映射，所以 ESX/ESXi 可以跨所有的虚拟机管理内存。虚拟机的一些物理内存可能映射到共享页面或未映射或换出的页面。

ESX/ESXi 主机执行虚拟内存管理时无需了解客户机操作系统，也不会干涉客户机操作系统自身的内存管理子系统。

性能注意事项

使用硬件辅助时，会消除软件内存虚拟化的开销。特别是，硬件辅助消除了使卷影页表与客户机页表保持同步所需的开销。但是，使用硬件辅助时 TLB 缺失滞后时间明显较长。因此，工作负载是否受益于硬件辅助主要取决于在使用软件内存虚拟化时由内存虚拟化引起的开销。如果工作负载涉及少量页表活动（例如进程创建、映射内存或上下文切换），则软件虚拟化不会引起显著开销。相反，具有大量页表活动的工作负载可能会因使用硬件辅助而受益。

管理内存资源

使用 vSphere Client，可以查看有关内存分配设置的信息并对其进行更改。为了有效管理内存资源，还必须熟悉内存开销、闲置内存消耗以及 ESX/ESXi 主机回收内存的方式。

当管理内存资源时，可以指定内存分配。如果未自定义内存分配，则 ESX/ESXi 主机使用适合大多数情况的默认值。

可以通过几种方式指定内存分配。

- 使用可通过 vSphere Client 访问的属性和特殊功能。使用 vSphere Client GUI 可以连接到 ESX/ESXi 主机或 vCenter Server 系统。
- 使用高级设置。
- 将 vSphere SDK 用于脚本式内存分配。

查看内存分配信息

可以使用 vSphere Client 查看有关当前内存分配的信息。

可以查看有关总内存的信息和可用于虚拟机的内存信息。在 ESX 中，还可以查看分配给服务控制台的内存。

步骤

1 在 vSphere Client 中，选择一台主机，然后单击**配置**选项卡。

2 单击**内存**。

可以查看第 26 页，“主机内存信息”中显示的信息。

主机内存信息

vSphere Client 显示有关主机内存分配的信息。

表 3–1 对主机内存字段进行了论述。

表 3–1。 主机内存信息

字段	描述
总计	该主机的总物理内存。
系统	ESX/ESXi 系统使用的内存。 ESX/ESXi 至少使用 50 MB 系统内存用于 VMkernel，并使用额外内存用于设备驱动程序。该内存在 ESX/ESXi 已加载且无法配置时分配。 虚拟化层实际所需的内存取决于主机上 PCI（外围组件互连）设备的数量和类型。有些驱动程序需要 40 MB，几乎是基本系统内存的两倍。 ESX/ESXi 主机还尝试使一些内容一直保持可用，以便高效处理动态分配请求。ESX/ESXi 设置大约 6% 的内存来供正在运行的虚拟机使用。 ESXi 主机针对在 ESX 主机的服务控制台中运行的管理代理使用额外的系统内存。
虚拟机	由选定主机上运行的虚拟机使用的内存。 主机的大多数内存都用于正在运行的虚拟机。ESX/ESXi 主机根据管理参数和系统负载，管理此内存到虚拟机的分配。 可供虚拟机使用的物理内存量总是低于物理主机的内存量，因为虚拟化层会占用一些资源。例如，具有 2 GB 内存和两个 3.2 GHz CPU 的主机可能只有 1.5 GB 内存和 6 GHz CPU 资源供虚拟机使用。
服务控制台	为服务控制台预留的内存。 单击**属性**以更改可用于服务控制台的内存量。此字段仅出现在 ESX 中。ESXi 不提供服务控制台。

了解内存开销

内存资源的虚拟化会涉及一些相关开销。

ESX/ESXi 虚拟机可以引起两种内存开销。

- 在虚拟机内访问内存所需的额外时间。
- 超出向每个虚拟机分配的内存后，ESX/ESXi 主机自身代码和数据结构所需的额外空间。

ESX/ESXi 内存虚拟化为内存访问增加很少的时间开销。因为处理器分页硬件直接使用页表（基于软件的卷影页表方法或硬件辅助的嵌套页表方法），所以虚拟机中的大多数内存访问在执行时没有地址转换开销。

内存空间开销有两部分：

- VMkernel 和服务控制台（仅用于 ESX）的系统范围的固定开销。
- 每个虚拟机的额外开销。

对于 ESX，服务控制台通常使用 272MB，而 VMkernel 则使用更少的内存。所使用的内存量取决于正在使用的设备驱动程序的数量和大小。

开销内存包括为虚拟机框架缓冲区和各种虚拟化数据结构（如卷影页表）预留的空间。开销内存取决于虚拟 CPU 数量以及为客户机操作系统配置的内存。

ESX/ESXi 还提供了内存共享等优化措施来减少基础服务器上使用的物理内存量。这些优化措施可以节省的内存多于开销占用的内存。

虚拟机上的开销内存

虚拟机会引起开销内存。您应当了解此开销量。

表 3–2 列出了每种数量的 VCPU 的开销内存（以 MB 为单位）。

表 3–2。 虚拟机上的开销内存

内存 (MB)	1 个 VCPU	2 个 VCPU	3 个 VCPU	4 个 VCPU	5 个 VCPU	6 个 VCPU	7 个 VCPU	8 个 VCPU
256	113.17	159.43	200.53	241.62	293.15	334.27	375.38	416.50
512	116.68	164.96	206.07	247.17	302.75	343.88	385.02	426.15
1024	123.73	176.05	217.18	258.30	322.00	363.17	404.34	445.52
2048	137.81	198.20	239.37	280.53	360.46	401.70	442.94	484.18
4096	165.98	242.51	283.75	324.99	437.37	478.75	520.14	561.52
8192	222.30	331.12	372.52	413.91	591.20	632.86	674.53	716.19
16384	334.96	508.34	550.05	591.76	900.44	942.98	985.52	1028.07
32768	560.27	863.41	906.06	948.71	1515.75	1559.42	1603.09	1646.76
65536	1011.21	1572.29	1616.19	1660.09	2746.38	2792.30	2838.22	2884.14
131072	1912.48	2990.05	3036.46	3082.88	5220.24	5273.18	5326.11	5379.05
262144	3714.99	5830.60	5884.53	5938.46	10142.83	10204.79	10266.74	10328.69

ESX/ESXi 主机如何分配内存

ESX/ESXi 主机将 `Limit` 参数所指定的内存分配给每个虚拟机，除非内存过载。ESX/ESXi 主机向虚拟机分配的内存决不会超过为它指定的物理内存大小。

例如，1 GB 虚拟机可能具有默认的限制（无限）或用户指定的限制（例如 2 GB）。在这两种情况下，ESX/ESXi 主机分配的内存决不会超过 1 GB，即不会超过为其指定的物理内存大小。

当内存过载时，向每个虚拟机分配的内存量介于**预留**和**限制**指定的内存量之间。授予虚拟机的高于预留量的内存量会因当前的内存负载而异。

ESX/ESXi 主机根据分配给虚拟机的份额数和对最近工作集大小的估计，确定每个虚拟机的分配量。

- 份额 — ESX/ESXi 主机使用经过修改的按比例份额内存分配策略。内存份额给予虚拟机一部分可用物理内存。
- 工作集大小 — ESX/ESXi 主机通过在连续的虚拟机执行时间周期监控内存活动，来估计工作集。采用快速响应工作集大小增加且慢速响应工作集大小减小的技术，在几个时间周期内进行平稳估计。

  该方法确保虚拟机开始更活跃地使用其内存时，已经回收闲置内存的虚拟机可以快速达到基于完整份额的分配量。

  在默认情况下将对内存活动监控 60 秒以估计工作集大小。要修改此默认值，请调整 `Mem.SamplePeriod` 高级设置。请参见第 85 页，“设置高级主机属性”。

闲置虚拟机的内存消耗

如果虚拟机未在使用当前为其分配的所有内存，则 ESX/ESXi 对闲置内存的消耗量大于对正在使用的内存的消耗量。这样有助于防止虚拟机累积闲置内存。

闲置内存消耗以渐进方式应用。随着虚拟机闲置内存与活动内存的比率的提高，有效消耗率将增加。（在不支持分层资源池的早期版本 ESX 中，虚拟机的所有闲置内存是以同等比率消耗的）。

`Mem.IdleTax` 高级设置允许您修改闲置内存消耗率。使用该选项以及 `Mem.SamplePeriod` 高级属性可控制系统如何确定虚拟机的目标内存分配。请参见第 85 页，“设置高级主机属性”。

注意 大多数情况下，没有必要更改 `[Mem.IdleTax]`，甚至于如果更改的话，反而不合适。

内存回收

ESX/ESXi 主机可以从虚拟机中回收内存。

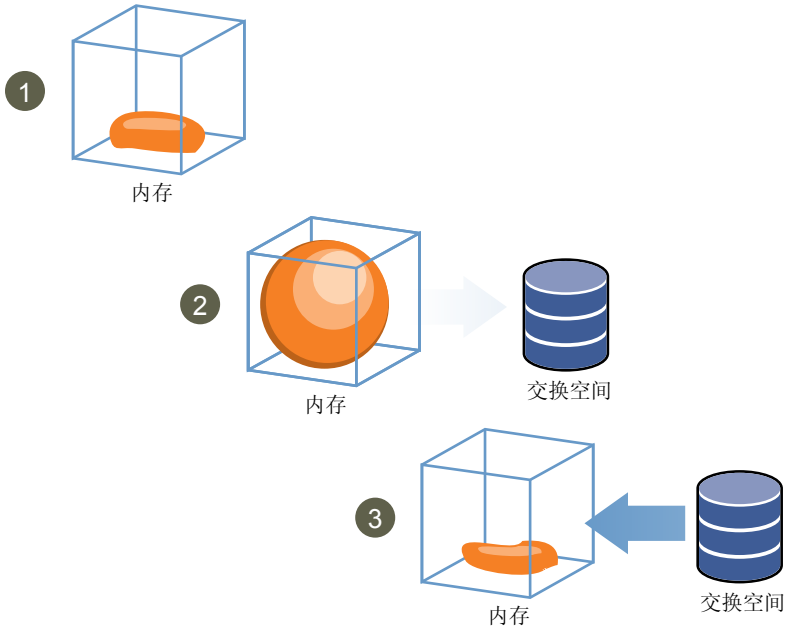
ESX/ESXi 主机会将预留功能指定的内存量直接分配给虚拟机。超出预留的任何部分都使用主机的物理资源进行分配，如果物理资源不可用，则使用虚拟增长或交换等特殊技术进行处理。主机可使用两种技术来动态增加或减少分配给虚拟机的内存量：

- ESX/ESXi 系统使用已加载到虚拟机中所运行的客户机操作系统的内存虚拟增长驱动程序 (`vmmemctl`)。请参见第 28 页，“内存虚拟增长驱动程序”。
- ESX/ESXi 系统从虚拟机分页到服务器交换文件，无需客户机操作系统参与。每个虚拟机均有自己的交换文件。

内存虚拟增长驱动程序

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。该驱动程序使用专用虚拟增长技术，提供在类似的内存限制下与本机系统的行为极为相近的可预测性能。该技术可增加或减少客户机操作系统的内存压力，使得客户机能够使用自己的本机内存管理算法。当内存很紧张时，客户机操作系统决定要回收哪些页面，并在必要时将这些页面换到自己的虚拟磁盘上。请参见图 3–2。

图 3–2。 客户机操作系统中的内存虚拟增长



注意 必须使用足够的交换空间来配置客户机操作系统。某些客户机操作系统具有其他限制。

如有必要，通过为特定虚拟机设置 **sched.mem.maxmemctl** 参数，可以限制由 `vmmemctl` 回收的内存量。该选项指定了可以从虚拟机中回收的最大内存量，以兆字节 (MB) 为单位。请参见第 87 页，“设置高级虚拟机属性”。

使用交换文件

可以指定交换文件的位置、当内存过载时预留交换空间以及删除交换文件。

当 `vmmemctl` 驱动程序不可用或未响应时，ESX/ESXi 主机会使用交换从虚拟机中强制回收内存。

- 从未安装。
- 明确禁用。
- 未运行（例如，客户机操作系统正在引导时）。
- 暂时无法以足够快的速度回收内存来满足当前系统需求。
- 正常工作，但是已经达到最大虚拟增长大小。

当虚拟机需要页面时，标准需求分页技术会重新换入页面。

注意 为了获得最佳性能，只要有可能，ESX/ESXi 主机就会使用虚拟增长方法（通过 `vmmemctl` 驱动程序实施）。交换是只有必须回收内存时主机才会使用的最后可靠机制。

交换文件位置

默认情况下，在与虚拟机配置文件相同的位置中创建交换文件。

当虚拟机启动时，ESX/ESXi 主机会创建交换文件。如果无法创建该文件，则无法启动虚拟机。除了接受默认值以外，您还可以：

- 使用每个虚拟机配置选项将数据存储更改为另一个共享的存储位置。
- 使用主机-本地交换在主机上指定存储在本地的数据库存储。这样就可以在每个主机级别上进行交换，从而节省 SAN 上的空间。但是，对于 VMware VMotion，可能会导致性能稍有下降，因为交换到源主机上的本地交换文件的页面必须通过网络传输到目标主机。

为 DRS 群集启用主机-本地交换

主机-本地交换允许将存储在主机本地的数据存储指定为交换文件位置。可以为 DRS 群集启用主机-本地交换。

步骤

1 右键单击 vSphere Client“清单”面板中的群集，然后单击**编辑设置**。

2 在“群集设置”对话框的左窗格中，单击**交换文件位置**。

3 选中**将交换文件存储在主机指定的数据存储中**选项，然后单击**确定**。

4 在 vSphere Client“清单”面板中选择群集内的某个主机，然后单击**配置**选项卡。

5 选择**虚拟机交换文件位置**。

6 单击**交换文件数据存储**选项卡。

7 在提供的列表中，选择要使用的本地数据存储，然后单击**确定**。

8 对群集内的每台主机重复步骤 4 到步骤 7。

现在已为 DRS 群集启用主机-本地交换。

为独立主机启用主机-本地交换

主机-本地交换允许将存储在主机本地的数据存储指定为交换文件位置。可以为独立主机启用主机-本地交换。

步骤

1 在 vSphere Client“清单”面板中选择主机，然后单击**配置**选项卡。

2 选择**虚拟机交换文件位置**。

3 在“虚拟机交换文件位置”对话框的**交换文件位置**选项卡下，选择**将交换文件存储到交换文件数据存储中**选项。

4 单击**交换文件数据存储**选项卡。

5 在提供的列表中，选择要使用的本地数据存储，然后单击**确定**。

现在已为独立主机启用主机-本地交换。

交换空间和内存过载

必须在每个虚拟机交换文件中为任何未预留的虚拟机内存预留交换空间（预留和配置内存大小之间的差值）。

需要该交换预留来确保 ESX/ESXi 系统在任何情况下均能预留虚拟机内存。实际上，只有一小部分主机级别的交换空间可能会用到。

如果正在通过 ESX/ESXi 使内存过载以支持由虚拟增长导致的客户机内部交换，请确保客户机操作系统还有足够的交换空间。该客户机级别交换空间必须大于或等于虚拟机配置内存大小与其“预留”之间的差值。

小心 如果内存过载且客户机操作系统配置的交换空间不足，则虚拟机中的客户机操作系统可能会出现故障。

为了避免虚拟机出现故障，请增加虚拟机中交换空间的大小。

- Windows 客户机操作系统 — Windows 操作系统将其交换空间称为分页文件。如果有足够的可用磁盘空间，一些 Windows 操作系统会尝试自动增加分页文件的大小。

  请查看 Microsoft Windows 文档或搜索 Windows 帮助文件来了解“分页文件”。按照说明更改虚拟内存分页文件的大小。

- Linux 客户机操作系统 — Linux 操作系统将其交换空间称为交换文件。有关增加交换文件的信息，请参见以下 Linux 手册页：
  - `mkswap` — 设置 Linux 交换区。
  - `swapon` — 针对分页和交换启用设备和文件。

具有大量内存和较小虚拟磁盘的客户机操作系统（例如，具有 8 GB RAM 和 2 GB 虚拟磁盘的虚拟机）更容易出现交换空间不足的情况。

删除交换文件

如果 ESX/ESXi 主机发生故障，并且该主机上正在运行的虚拟机使用交换文件，则这些交换文件将继续存在并占用磁盘空间，即使 ESX/ESXi 主机重新启动以后也是如此。这些交换文件可能消耗数千兆字节的磁盘空间，因此请确保正确删除这些交换文件。

步骤

1 重新启动故障主机上的虚拟机。

2 停止该虚拟机。

该虚拟机的交换文件即会删除。

在虚拟机之间共享内存

许多 ESX/ESXi 工作负载存在跨虚拟机（以及在单个虚拟机中）共享内存的机会。

例如，几个虚拟机可能正在运行同一客户机操作系统的多个实例，加载了相同的应用程序或组件，或包含公用数据。这些情况下，ESX/ESXi 主机使用专用的透明页共享技术安全地消除内存页的冗余副本。采用内存共享，在虚拟机中运行的工作负载消耗的内存通常要少于其在物理机上运行时所需的内存。因此，可以高效地支持更高级别的过载。

使用 `Mem.ShareScanTime` 和 `Mem.ShareScanGHz` 高级设置可控制系统扫描内存以确定内存共享机会的速率。

通过将 `sched.mem.pshare.enable` 选项设置为**无效**（该选项默认为**有效**），还可以针对单个虚拟机禁用共享。请参见第 87 页，“设置高级虚拟机属性”。

ESX/ESXi 内存共享作为后台活动运行，随着时间的推移而扫描共享机会。节省的内存量随着时间而变化。对于相当固定的工作负载，在使用所有共享机会之前，内存量一般会缓慢增加。

要确定给定工作负载内存共享的有效性，请尝试运行工作负载，并使用 `resxtop` 或 `esxtop` 观察实际节省的内存量。此信息可在“内存”页面中交互模式的 `PSHARE` 字段中找到。

衡量和区分各种内存使用情况

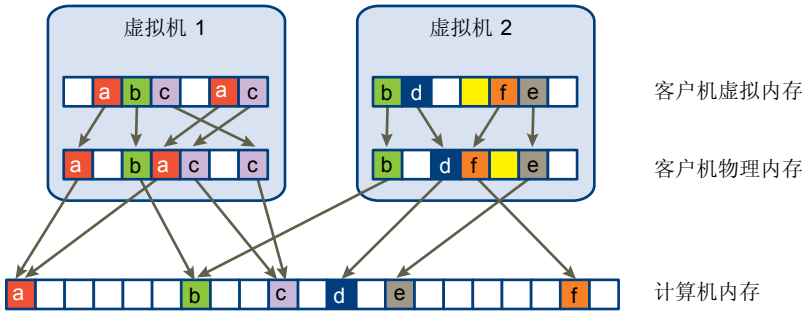
vSphere Client 的**性能**选项卡显示了可用于分析内存使用情况的多个衡量指标。

某些内存衡量指标用于衡量客户机物理内存，而另一些衡量指标用于衡量计算机内存。例如，可以使用性能衡量指标检查的两种内存使用情况是客户机物理内存和计算机内存。可以使用“已分配的内存”衡量指标（对于虚拟机）或“共享的内存”（对于 ESX/ESXi 主机）衡量客户机物理内存。但是，要衡量计算机内存，需要使用“消耗的内存”（对于虚拟机）或“共享的公用内存”（对于 ESX/ESXi 主机）。了解这些类型的内存使用情况之间的概念性差异对知道这些衡量指标的衡量对象以及如何对其进行解释十分重要。

VMkernel 会将客户机物理内存映射到计算机内存，但是它们不总是一对一映射。它可能会将客户机物理内存的多个区域映射到计算机内存的同一区域（当存在内存共享时），或者未将客户机物理内存的特定区域映射到计算机内存（当 VMkernel 换出或虚拟增长客户机物理内存时）。在这些情况中，单个虚拟机或 ESX/ESXi 主机的客户机物理内存使用情况和计算机内存使用情况的计算有所不同。

请考虑下图中的示例。两个虚拟机正在 ESX/ESXi 主机上运行。每块代表 4 KB 内存，每个颜色/字母代表相应块上的数据集。

图 3–3。 内存使用情况示例



可以按照如下方式确定虚拟机的性能衡量指标：

- 要确定虚拟机 1 的“已分配的内存”（映射到计算机内存的客户机物理内存量），请计算虚拟机 1 的客户机物理内存中的块（含有指向计算机内存的箭头）的数量并乘以 4 KB。由于有 5 个块含有箭头，因此“已分配的内存”是 20 KB。
- “消耗的内存”是分配给虚拟机的计算机内存量，包括从共享的内存中节省的内存量。首先，计算计算机内存中的块（含有从虚拟机 1 的客户机物理内存指出的箭头）的数量。这样的块有三个，但有一个块与虚拟机 2 共享。因此，计算两个完整的块加上半个第三个块并乘以 4 KB，得到总计 10 KB 的“消耗的内存”。

这两个衡量指标之间的重要差异是：“已分配的内存”计算带箭头的客户机物理内存级块的数量，“消耗的内存”计算带箭头的计算机内存级块的数量。由于内存共享，这两个级别的块的数量不同，因此“已分配的内存”和“消耗的内存”也不同。这并不表示存在问题，而是表示内存是通过共享或其他回收技术节省的。

在确定 ESX/ESXi 主机的“共享的内存”和“共享的公用内存”时，会获得类似的结果。

- 主机的“共享的内存”是每个虚拟机“共享的内存”的总和。通过查看每个虚拟机的客户机物理内存并计算含有指向计算机内存块（计算机内存块本身也含有多个指向自己的箭头）的箭头的块数量，可计算此内存总和。在本例中，这样的块有六个，因此主机的“已共享的内存”是 24 KB。
- “共享的公用内存”是由虚拟机共享的计算机内存量。要确定此内存量，可查看计算机内存，并计算有多个箭头指向自身的块数量。这样的块有三个，因此“共享的公用内存”是 12 KB。

“共享的内存”涉及到客户机物理内存，即作为箭头起始点。而“共享的公用内存”涉及到计算机内存，即作为箭头的目标点。

用于衡量客户机物理内存和计算机内存的内存衡量指标可能会出现矛盾。事实上，它们衡量的是虚拟机内存使用情况的不同方面。通过了解这些衡量指标之间的差异，可以更好地利用它们来诊断性能问题。

管理资源池

4

资源池是灵活管理资源的逻辑抽象。资源池可以分组为层次结构，用于对可用的 CPU 和内存资源按层次结构进行分区。

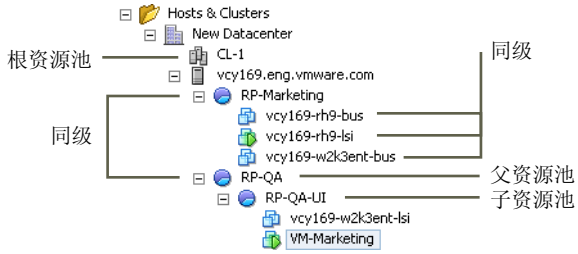
每台独立主机和每个 DRS 群集都具有一个（不可见的）根资源池，此资源池对该主机或群集的资源进行分组。根资源池之所以不显示，是因为主机（或群集）与根资源池的资源总是相同的。

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。您可以创建共享资源的层次结构。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。

在图 4–1 中，RP-QA 是 RP-QA-UI 的父资源池。RP-Marketing 与 RP-QA 是同级。紧靠 RP-Marketing 下面的三个虚拟机也是同级。

图 4–1。 资源池层次结构中的父级、子级和同级



对于每个资源池，均可指定预留、限制、份额以及预留是否应为可扩展。随后该资源池的资源将可用于子资源池和虚拟机。

本章讨论了以下主题：

为什么使用资源池?

通过资源池可以委派对主机（或群集）资源的控制权，在使用资源池划分群集内的所有资源时，其优势非常明显。可以创建多个资源池作为主机或群集的直接子级，并对它们进行配置。然后便可向其他个人或组织委派对资源池的控制权。

使用资源池具有下列优点。

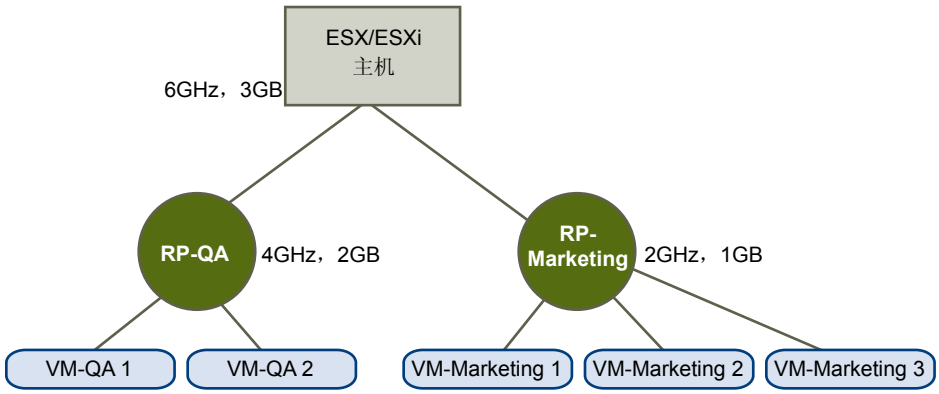
- 灵活的层次结构组织 - 根据需要添加、移除或重组资源池，或者更改资源分配。
- 资源池之间相互隔离，资源池内部相互共享 - 顶级管理员可向部门级管理员提供一个资源池。某部门资源池内部的资源分配变化不会对其他不相关的资源池造成不公平的影响。
- 访问控制和委派 — 顶级管理员使资源池可供部门级管理员使用后，该管理员可以在当前的份额、预留和限制设置向该资源池授予的资源范围内进行所有的虚拟机创建和管理操作。委派通常结合权限设置一起执行。
- 资源与硬件的分离 — 如果使用的是已启用 DRS 的群集，则所有主机的资源始终会分配给群集。这意味着管理员可以独立于提供资源的实际主机来进行资源管理。如果将三台 2GB 主机替换为两台 3GB 主机，您无需对资源分配进行更改。

  这一分离可使管理员更多地考虑聚合计算能力而非各个主机。
- 管理运行多层服务的各组虚拟机 — 为资源池中的多层服务进行虚拟机分组。您无需对每个虚拟机进行资源设置，而是可以通过更改该组虚拟机所属资源池的设置来控制对这些虚拟机的总体资源分配。

例如，假定一台主机拥有多个虚拟机。营销部门使用其中的三个虚拟机，QA 部门使用两个虚拟机。由于 QA 部门需要更多的 CPU 和内存，管理员为每组创建了一个资源池。管理员将 QA 部门资源池和营销部门资源池的 **CPU 份额**分别设置为**高**和**正常**，以便 QA 部门的用户可以运行自动测试。CPU 和内存资源较少的第二个资源池足以满足营销工作人员的较低负载要求。只要 QA 部门未完全利用所分配到的资源，营销部门就可以使用这些可用资源。

图 4–2 中演示了此应用场景。这些数字显示了向资源池的有效分配。

图 4–2。 向资源池分配资源



创建资源池

您可以创建任何 ESX/ESXi 主机、资源池或 DRS 群集的子资源池。

注意 如果已将某台主机添加到群集，将无法创建该主机的子资源池。如果已针对群集启用 DRS，则可以创建该群集的子资源池。

创建子资源池时，系统将提示您输入资源池属性信息。系统使用接入控制确保您不能分配不可用的资源。

步骤

1 选择所需的父级，然后选择**文件 > 新建 > 资源池**（或在**摘要**选项卡的“命令”面板中单击**新建资源池**）。

2 在“创建资源池”对话框中，为资源池提供所需信息。

3 完成所有选择操作后，请单击**确定**。

此时 vCenter Server 将创建该资源池，并将其显示在“清单”面板中。如有任何选定值因可用 CPU 和内存总量限制而无效，则将显示黄色三角形。

创建资源池后，即可向其中添加主机。虚拟机的份额与同一父资源池内的其他虚拟机（或资源池）相关。

资源池属性

可以使用资源分配设置来管理资源池。

表 4–1 概述了可为资源池指定的属性。

表 4–1。 资源池属性

字段	描述
名称	新资源池的名称。
份额	资源池拥有的、相对于父级总数的 CPU 或内存份额值。同级资源池根据由其预留和限制限定的相对份额值共享资源。可以选择**低**、**正常**或**高**，也可以选择**自定义**来指定表示份额值的数字。
预留	保证为该资源池分配的 CPU 或内存量。非零预留将从父级（主机或资源池）的未预留资源中减去。这些资源被认为是预留资源，无论虚拟机是否与该资源池相关联也是如此。默认值为“0”。
可扩展预留	表示在接入控制期间是否考虑可扩展预留。当该复选框处于选中状态（默认设置）时，如果在该资源池中启动一个虚拟机，并且虚拟机的总预留大于该资源池的预留，则该资源池可以使用父级或祖先的资源。
限制	主机为该资源池提供的 CPU 或内存量的限制。默认设置为**无限**。要指定限制，请取消选中**无限**复选框。

资源池创建示例

此过程示例演示了如何使用 ESX/ESXi 主机作为父资源来创建资源池。

假定有一台 ESX/ESXi 主机，提供 6 GHz 的 CPU 和 3 GB 的内存，这些 CPU 和内存必须在营销部门和 QA 部门间进行共享。还需要不均等地共享资源，并授予一个部门 (QA) 更高的优先级。通过为每个部门创建一个资源池并使用**份额**属性区分资源分配优先级，可完成此任务。

此过程示例演示了如何使用 ESX/ESXi 主机作为父资源来创建资源池。

步骤

1 在“创建资源池”对话框中，键入 QA 部门的资源池的名称（例如 RP-QA）。

2 将 RP-QA 的 CPU 和内存资源**份额**指定为**高**。

3 创建第二个资源池 RP-Marketing。

将 CPU 和内存的“份额”保留为**正常**。

4 单击**确定**退出。

如果存在资源冲突，则 RP-QA 接收 4GHz 和 2GB 的内存，RP-Marketing 接收 2GHz 和 1GB 的内存。否则，它们可以接收超过此分配的量。这些资源随后即可供各自资源池内的虚拟机使用。

更改资源池属性

在创建资源池之后，可以更改其属性。

步骤

1 在 vSphere Client“清单”面板中选择资源池。

2 在**摘要**选项卡的“命令”面板中，选择**编辑设置**。

3 在“编辑设置”对话框中，可以更改选定资源池的全部属性。

将虚拟机添加到资源池

创建虚拟机时，可以通过新建虚拟机向导在创建过程中指定资源池位置。也可以将现有的虚拟机添加到资源池。

将虚拟机移至新的资源池时：

- 该虚拟机的预留和限制不会发生变化。
- 如果该虚拟机的份额为高、中或低，份额百分比会有所调整以反映新资源池中使用的份额总数。
- 如果已为该虚拟机指定了自定义份额，该份额值将保持不变。

  注意 由于份额分配是相对于资源池的，因此，当您将虚拟机移入资源池时可能必须手动更改虚拟机的份额，以便虚拟机的份额与新资源池中的相对值保持一致。如果虚拟机所占总份额的比例过大（或过小），将显示警告。

- “资源分配”选项卡中显示的有关资源池的预留和未预留 CPU 和内存资源的信息将发生变化，以反映与该虚拟机关联的预留（如果有）。

  注意 如果虚拟机已关闭或挂起，可以移动该虚拟机，但资源池的可用资源总量（例如预留和未预留的 CPU 和内存资源）不受影响。

步骤

1 从清单中的任意位置选择已存在的虚拟机。

  该虚拟机可以与独立主机、群集或另一个资源池关联。

2 将该虚拟机（或多个虚拟机）拖至所需的资源池对象。

如果某个虚拟机已启动，且目标资源池的 CPU 或内存不足以保证该虚拟机的预留，移动操作将会失败，因为接入控制不允许该操作。此时将显示一个错误对话框，解释这种情况。该错误对话框会将可用资源与所请求的资源进行比较，以便您考虑可否通过调整来解决此问题。

从资源池中移除虚拟机

通过将虚拟机移动到另一个资源池或将其删除，可以从资源池中移除虚拟机。

将虚拟机移至其他资源池

可以将虚拟机拖放到另一资源池。如果只需移动虚拟机，则无需将其关闭。

从某个资源池中移除虚拟机时，与该资源池相关联的份额总数将减少，从而使每个剩余的份额代表更多资源。例如，假定您有一个有权使用 6 GHz 的资源池，其中包含三台份额设置为**正常**的虚拟机。假定虚拟机受 CPU 限制，每个虚拟机获得 2 GHz 的相等分配额。如果将其中一个虚拟机移至其他资源池，剩余的两个虚拟机将各获得 3GHz 的相等分配额。

从清单中移除虚拟机或将其从磁盘中删除

右键单击虚拟机，并单击**从清单中移除**或**从磁盘删除**。

您需要关闭虚拟机才能将其完全移除。

资源池接入控制

在资源池内启动虚拟机时，或尝试创建子资源池时，系统会执行其他接入控制以确保不违反资源池的限制。

启动虚拟机或创建资源池之前，请在资源池的**资源分配**选项卡中检查**未预留的** CPU 和**未预留的**内存字段，以确定是否有足够的资源可用。

如何计算**未预留**的 CPU 和内存以及是否执行操作取决于**预留类型**。

表 4–2。 预留类型

预留类型	描述
固定的	系统检查所选资源池是否有足够的未预留资源。如果有，则可以执行操作。否则将显示一条消息，而且无法执行操作。
可扩展的（默认）	系统考虑所选资源池及其直接父资源池中的可用资源。如果对于父资源池也选中了**可扩展预留**选项，它还可以从其父资源池中借用资源。只要选中了**可扩展预留**选项，就会以递归方式向当前资源池的祖先借用资源。将该选项保持选中状态可提供更高的灵活性，但提供的保护将会同时减少。子资源池所有者预留的资源可能大于您的预期值。

系统不允许违反预先配置的**预留**或**限制**设置。每次重新配置资源池或启动虚拟机时，系统都会验证所有参数以确保仍能实现各服务级别保证。

可扩展预留示例 1

此示例显示了具有可扩展预留的资源池的工作方式。

假定某个管理员负责管理资源池 P，并定义了两个子资源池 S1 和 S2，分别用于两个不同的用户（或组）。

该管理员知道用户将要启动具有预留的虚拟机，但不知道每个用户需要预留多少资源。为 S1 和 S2 设置可扩展预留可使管理员更加灵活地共享和继承资源池 P 的公用预留。

如果不使用可扩展预留，管理员需要向 S1 和 S2 明确分配具体的资源量。这种具体的分配可能欠缺灵活性，特别是在较深的资源池层次结构中，并且可能使资源池层次结构中的预留设置操作复杂化。

可扩展预留会造成缺少严格的隔离。S1 可使用 P 的全部预留启动，致使 S2 无法直接使用任何 CPU 或内存资源。

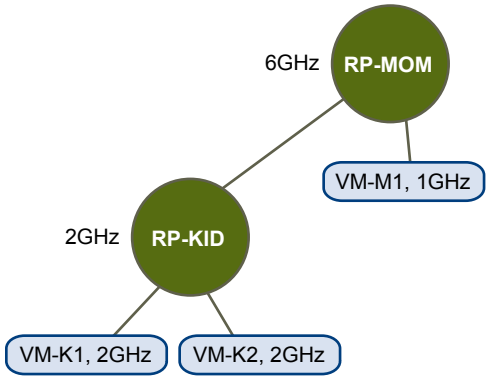
可扩展预留示例 2

此示例显示了具有可扩展预留的资源池的工作方式。

假设以下应用场景（如图 4–3 中所示）。

- 父资源池 RP-MOM 具有 6 GHz 的预留及一台预留了 1 GHz 的运行中的虚拟机。
- 您创建了一个具有 2 GHz 预留的子资源池 RP-KID，并选中**可扩展预留**。
- 您向子资源池添加两个各具有 2 GHz 预留的虚拟机（即 VM-K1 和 VM-K2），并尝试启动它们。
- VM-K1 可直接从 RP-KID（具有 2 GHz）预留资源。
- VM-K2 没有本地资源可用，因此它将从父资源池 RP-MOM 中借用资源。RP-MOM 现有资源为 6 GHz 减去 1 GHz（由虚拟机预留）再减去 2 GHz（由 RP-KID 预留），剩下 3 GHz 的未预留资源。利用 3 GHz 的可用资源，您可以启动这个 2 GHz 虚拟机。

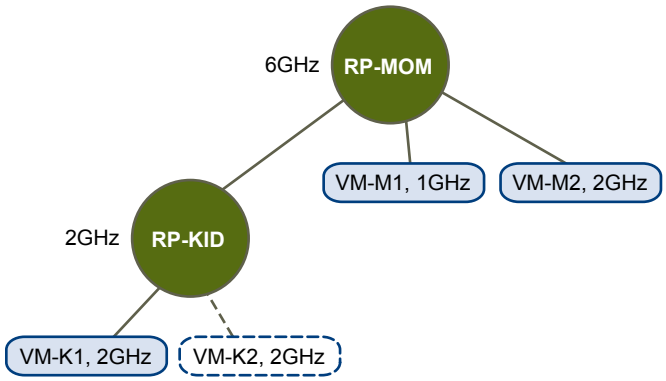
图 4–3。 可扩展资源池的接入控制：成功启动



现在假设另一个包含 VM-M1 和 VM-M2 的应用场景（如图 4–4 中所示）：

- 启动 RP-MOM 中总预留为 3 GHz 的两个虚拟机。
- 您依然可启动 RP-KID 中的 VM-K1，因为本地有 2 GHz 可用。
- 当您尝试启动 VM-K2 时，RP-KID 已无未预留的 CPU 容量，因此会检查其父级。RP-MOM 只有 1 GHz 的未预留容量可用（RP-MOM 的 5 GHz 已被占用 — 3 GHz 由本地虚拟机预留，2 GHz 由 RP-KID 预留）。因此，您无法启动需要 2 GHz 预留的 VM-K2。

图 4–4。 可扩展资源池的接入控制：无法启动



创建 DRS 群集 5

DRS 群集是一组具有共享资源和共享管理界面的 ESX/ESXi 主机和相关虚拟机。必须创建 DRS 群集，才能从群集级别资源管理中获益。

将主机添加到 DRS 群集时，主机的资源将成为群集资源的一部分。除了这种资源聚合外，您还可以使用 DRS 群集支持群集范围内的资源池并强制执行群集级别的资源分配策略。还提供下面的群集级别的资源管理功能。

- 负载平衡 — 将持续监控群集内所有主机和虚拟机的 CPU 和内存资源的分布情况和使用情况。在给出群集内资源池和虚拟机的属性、当前需求以及不平衡目标的情况下，DRS 会将这些衡量指标与理想状态下的资源利用率进行比较。然后，它会相应地执行虚拟机迁移（或提供迁移建议）。请参见第 41 页，“虚拟机迁移”。当您在群集中首次启动虚拟机时，DRS 将尝试通过在相应主机上放置该虚拟机或提出建议来保持适当的负载平衡。请参见第 40 页，“接入控制和初始放置位置”。
- 电源管理 - VMware 分布式电源管理功能处于启用状态时，DRS 会将群集和主机级容量与群集的虚拟机需求（包括近期历史需求）进行比较。如果找到足够的额外容量，它会将主机置于（或建议置于）待机电源模式，或者如果需要容量，则建议启动主机。根据提出的主机电源状况建议，可能需要将虚拟机迁移到主机并从主机迁移虚拟机。请参见第 55 页，“管理电源资源”。
- DRS 规则—您可以通过分配 DRS（关联性或反关联性）规则控制群集内主机上的虚拟机的放置位置。请参见第 47 页，“使用 DRS 规则”。

本章讨论了以下主题：

接入控制和初始放置位置

尝试在已启用 DRS 的群集内启动一个或一组虚拟机时，vCenter Server 会执行接入控制。它会检查群集内是否有足够的资源来支持虚拟机。

如果群集没有足够的资源来启动单个虚拟机，或在组启动尝试中无法启动任何虚拟机，将会显示一条消息。否则，对于每台虚拟机，DRS 将生成要在其上运行虚拟机的主机的建议，并执行以下操作之一

- 自动执行放置位置建议。
- 显示用户随后可以选择接受或覆盖的放置位置建议。

注意 对于独立主机或非 DRS 群集内的虚拟机，不提出任何初始放置位置建议。这些虚拟机将会在启动时被置于当前所驻留的主机上。

有关 DRS 建议及其应用的详细信息，请参见第 60 页，“DRS 建议页面”。

单个虚拟机启动

在 DRS 群集中，可以启动单个虚拟机，并接受初始放置位置建议。

启动单个虚拟机时，有两种类型的初始放置位置建议：

- 启动单个虚拟机，不需要任何必备条件步骤。

  用户将拥有虚拟机的初始放置位置建议列表，这些建议是互斥的。您只能选择一种建议。

- 启动单个虚拟机，但需要执行必备条件操作。

  这些操作包括在待机模式下启动主机或在主机间迁移其他虚拟机。在这种情况下，提供的建议具有多行，显示每个必备条件操作。用户可以接受整个建议，也可以取消启动虚拟机。

组启动

可以尝试同时启动多个虚拟机（组启动）。

选定进行组启动尝试的虚拟机不必位于同一个 DRS 群集内。可以在群集间选择虚拟机，但它们必须属于同一数据中心。也可以包括位于非 DRS 群集或独立主机上的虚拟机。这些虚拟机会自动启动并且不包括在任何初始放置位置建议中。

每个群集均提供组启动尝试的初始放置位置建议。如果组启动尝试的与放置位置相关的所有操作都处于自动模式，虚拟机将启动，而不提出任何初始放置位置建议。如果任何虚拟机的与放置位置相关的操作处于手动模式，则所有虚拟机（包括处于自动模式的虚拟机）都将手动启动，并且包括在初始放置位置建议中。

对于已启动的虚拟机所属的每个 DRS 群集，均会有一个包含所有必备条件的建议（或没有建议）。所有特定于此类群集的建议都显示在“启动建议”选项卡下。

如果进行了非自动组启动尝试，且包括了不受限于初始放置位置建议的虚拟机（即独立主机或非 DRS 群集上的虚拟机），vCenter Server 会尝试自动启动这些虚拟机。如果这些虚拟机自动启动成功，则会在“已开始启动”选项卡下列出。那些无法启动的虚拟机则在“失败的启动”选项卡下列出。

组启动示例

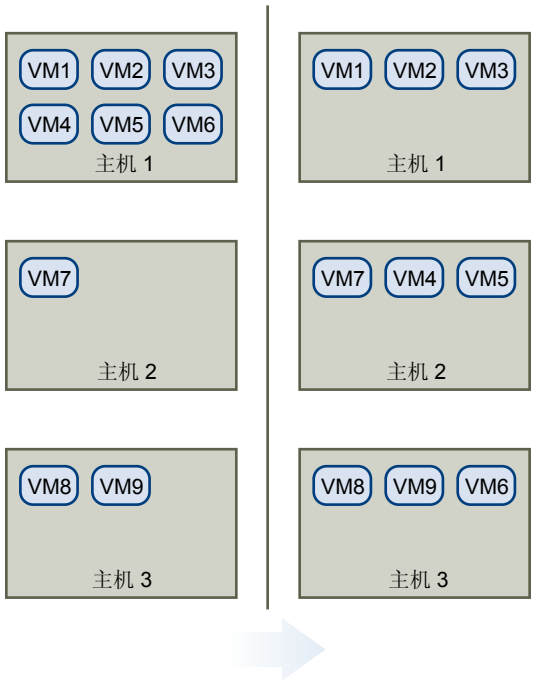
用户选择同一数据中心中的三个虚拟机进行组启动尝试。前两个虚拟机（VM1 和 VM2）在同一 DRS 群集 (Cluster1) 中，而第三个虚拟机 (VM3) 则在一台独立主机上。VM1 处于自动模式，而 VM2 处于手动模式。在此方案中，用户将获得 Cluster1 的初始放置位置建议（位于“启动建议”选项卡下），其中包含启动 VM1 和 VM2 的操作。将尝试自动启动 VM3，如果成功，则会在“已开始启动”选项卡下列出 VM3。如果此尝试失败，则会在“失败的启动”选项卡下列出 VM3。

虚拟机迁移

尽管 DRS 执行初始放置位置以便跨群集平衡负载，但是虚拟机负载和资源可用性中的更改可能会导致群集失衡。要更正此失衡情况，DRS 将生成迁移建议。

如果在群集上启用了 DRS，则可以更均匀地分配负载，从而降低不平衡程度。例如，请参见图 5–1。该图左侧的三台主机不平衡。假定主机 1、主机 2 和主机 3 具有相同的容量，且所有虚拟机的配置和负载（包括预留，如果已设置）均相同。但是，由于主机 1 有六个虚拟机，其资源可能被过度利用，而主机 2 和主机 3 有丰富的资源，因此 DRS 会将虚拟机从主机 1 迁移到主机 2 和主机 3（或提出迁移建议）。该图右侧显示了适当平衡负载之后所得到的主机配置。

图 5–1。 负载平衡



当群集不平衡时，DRS 将根据默认的自动化级别，提出建议或迁移虚拟机：

- 如果所涉及的群集或任何虚拟机为手动或半自动，则 vCenter Server 不执行自动操作来平衡资源。“摘要”页面会指示有迁移建议，“DRS 建议”页面会显示最有效地利用群集内资源的更改建议。
- 如果所涉及的群集或虚拟机均为全自动，则 vCenter Server 将根据需要在主机间迁移正在运行的虚拟机，以确保高效利用群集资源。

注意 即使是在自动迁移设置中，用户也可以显式迁移单个虚拟机，但 vCenter Server 可能会将这些虚拟机迁移到其他主机，以优化群集资源。

默认情况下，自动化级别是为整个群集指定的。也可以为单个虚拟机指定自定义的自动化级别。

DRS 迁移阈值

DRS 迁移阈值允许您指定要生成并应用的建议（如果建议中所涉及的虚拟机处于全自动模式）或要显示的建议（如果处于手动模式）。此阈值还用来度量主机（CPU 和内存）负载之间可以接受的群集不平衡程度。

可以移动阈值滑块以使用从“保守”到“激进”这五个设置中的一个。这五种迁移设置将根据其所分配的优先级生成建议。每次将滑块向右移动一个设置，将会允许包含下一较低优先级的建议。“保守”设置仅生成优先级 1 的建议（强制性建议），向右的下一级别则生成优先级 2 的建议以及更高级别的建议，然后依次类推，直至“激进”级别，该级别生成优先级 5 的建议和更高级别的建议（即，所有建议）。

每个迁移建议的优先级是使用群集的负载不平衡衡量指标进行计算的。该衡量指标在 vSphere Client 群集的“摘要”选项卡中显示为“当前主机负载标准偏差”。负载越不平衡，所生成迁移建议的优先级会越高。有关该指标以及如何计算建议优先级的详细信息，请参见 VMware 知识库文章“计算 VMware DRS 迁移建议的优先级”。

在建议收到优先级后，会将该级别与您所设置的迁移阈值进行比较。如果优先级低于或等于阈值设置，则会应用该建议（如果相关虚拟机均处于全自动模式），或向用户显示该建议以进行确认（如果处于手动或半自动模式）。

迁移建议

如果创建带有默认模式（手动或半自动）的群集，则 vCenter Server 将在“DRS 建议”页面上显示迁移建议。

系统将提供足够的建议，以强制实施规则并平衡群集的资源。每条建议均包含要移动的虚拟机、当前（源）主机和目标主机，以及提出建议的原因。原因可能为以下之一：

- 平衡平均 CPU 负载或预留。
- 平衡平均内存负载或预留。
- 满足资源池预留。
- 满足 DRS（关联性或反关联性）规则。
- 主机正在进入维护模式或待机模式。

注意 如果使用 VMware 分布式电源管理功能，那么，除了迁移建议外，DRS 还会提供主机电源状况建议。

DRS 群集必备条件

任何添加到 DRS 群集的主机都必须满足某些必备条件才能成功使用群集功能。

共享存储器

确保受管主机使用共享存储器。共享存储器通常位于存储区域网络 (SAN) 上，但也可以借助于 NAS 共享存储器来实现。

有关 SAN 的详细信息，请参见《iSCSI SAN 配置指南》和《光纤通道 SAN 配置指南》；有关其他共享存储器的信息，请参见《ESX 配置指南》或《ESXi 配置指南》。

共享 VMFS 卷

配置所有受管主机以使用共享 VMFS 卷。

- 将所有虚拟机的磁盘置于可通过源主机和目标主机访问的 VMFS 卷上。
- 将共享 VMFS 的访问模式设置为公用。
- 确保 VMFS 卷足够大，可以存储虚拟机的所有虚拟磁盘。
- 确保源主机及目标主机上的所有 VMFS 卷都使用卷名称，并且所有虚拟机都使用这些卷名称来指定虚拟磁盘。

注意 虚拟机交换文件还需要放在源主机和目标主机均可以访问的 VMFS 上（就像 `.vmdk` 虚拟磁盘文件一样）。如果所有的源主机及目标主机都是 ESX Server 3.5 或更高版本并且使用主机-本地交换，则此要求将不再适用。这种情况下，支持将带有交换文件的 VMotion 置于非共享存储器上。默认情况下，交换文件置于 VMFS 上，但管理员可以使用高级虚拟机配置选项替代此文件位置。

处理器兼容性

为了避免限制 DRS 的功能，应当将群集内源和目标主机的处理器兼容性最大化。

VMotion 在基础 ESX/ESXi 主机之间传输虚拟机的运行架构状况。VMotion 兼容性是指目标主机的处理器必须能够使用等效指令，从源主机的处理器在挂起时的状态继续执行。处理器时钟速度和缓存大小可能不同，但处理器必须属于相同的供应商类别（Intel 与 AMD）和相同的处理器系列，以便达到通过 VMotion 迁移所需的兼容性。

处理器系列（如 Xeon MP 和 Opteron）是由处理器供应商定义的。可以通过比较处理器的型号、步进级别和扩展功能来区分同一系列中的不同处理器版本。

在某些情况下，处理器供应商在同一处理器系列中引入了重大的架构更改（例如 64 位扩展及 SSE3）。如果不能保证通过 VMotion 成功迁移，VMware 会识别这些异常情况。

vCenter Server 提供了一些有助于确保通过 VMotion 迁移的虚拟机满足处理器兼容性要求的功能。这些功能包括：

- 增强型 VMotion 兼容性 (EVC) - 可以使用 EVC 帮助确保群集内主机的 VMotion 兼容性。EVC 可以确保群集内的所有主机向虚拟机提供相同的 CPU 功能集，即使这些主机上的实际 CPU 不同也是如此。可以避免因 CPU 不兼容而导致通过 VMotion 进行的迁移失败。

  可以在“群集设置”对话框中配置 EVC。为了使群集能够使用 EVC，群集内的主机必须满足某些要求。有关 EVC 和 EVC 要求的更多信息，请参见《基本系统管理》。

- CPU 兼容性掩码 - vCenter Server 会将虚拟机可用的 CPU 功能与目标主机的 CPU 功能进行比较，以确定是允许还是禁止通过 VMotion 迁移。通过将 CPU 兼容性掩码应用到单个虚拟机，可以向虚拟机隐藏某些 CPU 功能，从而防止由于 CPU 不兼容而造成的 VMotion 迁移失败。

VMotion 要求

要启用 DRS 迁移建议的使用，群集内的主机必须是 VMotion 网络的一部分。如果主机不在 VMotion 网络中，DRS 仍可提供初始放置位置建议。

要为 VMotion 进行配置，群集内的每台主机必须满足下列要求：

- ESX/ESXi 主机的虚拟机配置文件必须驻留在 VMware 虚拟机文件系统 (VMFS) 上。
- VMotion 不支持裸磁盘，也不支持对借助于 Microsoft 群集服务 (MSCS) 群集的应用程序进行迁移。
- VMotion 要求在所有启用了 VMotion 的受管主机之间设置专用的千兆以太网迁移网络。在受管主机上启用 VMotion 后，需要为受管主机配置唯一的网络标识对象并将其连接到专用迁移网络。

创建 DRS 群集

可使用 vSphere Client 中的新建群集向导创建 DRS 群集。

前提条件

可以在没有特殊许可证的情况下创建群集，但必须要有许可证才能为 DRS（或 VMware HA）启用群集。

步骤

1 在 vSphere Client 中，右键单击数据中心或文件夹，然后选择**新建群集**。

2 在**名称**字段中为群集命名。

  该名称显示在 vSphere Client“清单”面板中。

3 通过单击 **VMware DRS** 框来启用 DRS 功能。

  还可以通过单击 **VMware HA** 来启用 VMware HA 功能。

4 单击**下一步**。

5 选择 DRS 的默认的自动化级别。

	初始放置位置	迁移
手动	显示推荐的主机。	显示迁移建议。
半自动	自动放置。	显示迁移建议。
全自动	自动放置。	自动执行迁移建议。

6 设置 DRS 的迁移阈值。

7 单击**下一步**。

8 指定该群集的默认电源管理设置。

如果启用电源管理，则选择 DPM 阈值设置。

9 单击**下一步**。

10 如果适用，请启用增强型 VMotion 兼容性 (EVC)，并选择它应以何种模式运行。

11 单击**下一步**。

12 选择虚拟机的交换文件位置。

可以将交换文件与虚拟机本身存储在同一目录中，或者将交换文件存储在主机指定的数据存储中（主机-本地交换）。

13 单击**下一步**。

14 查看列出所选选项的摘要页。

15 单击**完成**以完成群集的创建，或单击**上一步**返回并对群集设置进行修改。

新群集不包括任何主机或虚拟机。

要将主机和虚拟机添加到群集，请参见第 49 页，“将主机添加到群集”和第 51 页，“从群集内移除虚拟机”。

设置虚拟机的自定义自动化级别

创建 DRS 群集后，可以为各个虚拟机自定义自动化级别，以替代群集的默认自动化级别。

步骤

1 在 vSphere Client 清单中选择群集。

2 右键单击并选择**编辑设置**。

3 在“群集设置”对话框的 **VMware DRS** 下，选择**虚拟机选项**。

4 选中**启用单个虚拟机自动化级别**复选框。

5 选择单个虚拟机，或者选择多个虚拟机。

6 右键单击并选择自动化模式。

7 单击**确定**。

注意 其他 VMware 产品或功能（如 VMware vApp 和 VMware 容错）可能会替代 DRS 群集内虚拟机的自动化级别。有关详细信息，请参见特定于产品的文档。

禁用 DRS

可以关闭群集的 DRS。

禁用 DRS 后，群集的资源池层次结构和 DRS 规则（请参见第 47 页，“使用 DRS 规则”）不会在您再次打开 DRS 时重新建立。因此，如果禁用 DRS，将从群集内移除资源池。为了避免丢失资源池，应该将 DRS 自动化级别更改为手动（并禁用所有虚拟机替代项），从而将 DRS 挂起，而不是禁用它。这样便可在阻止自动 DRS 操作的同时保留资源池层次结构。

步骤

1 在 vSphere Client 清单中选择群集。

2 右键单击并选择**编辑设置**。

3 在左侧面板中选择**常规**，并取消选中**打开 VMware DRS** 复选框。

4 单击**确定**，关闭 DRS。

6 使用 DRS 群集管理资源

创建 DRS 群集后，可以对其进行自定义，并使用它来管理资源。

要自定义 DRS 群集及其包含的资源，可以配置 DRS 规则，并添加和移除主机和虚拟机。在定义群集的设置和资源后，应当确保它是并保持为有效群集。还可以使用有效 DRS 群集管理电源资源，并与 VMware HA 进行交互操作。

本章讨论了以下主题：

使用 DRS 规则

可以通过使用 DRS 关联性和反关联性规则控制群集内主机上的虚拟机的放置位置。关联性规则指定将两个或多个虚拟机放置在同一主机上。反关联性规则限于两个虚拟机，并且要求不能将这两个虚拟机放置在同一主机上。

如果这两个规则冲突，则优先使用老的规则，并禁用新的规则。DRS 仅尝试满足已启用的规则，即使它们发生冲突。将忽略禁用的规则。与关联性规则的冲突相比，DRS 将优先阻止反关联性规则的冲突。

要检查是否有违反任何已启用的 DRS 规则的情况，请在 vSphere Client 的清单面板中选择该群集，再选择 DRS 选项卡，然后单击**错误**。如果违反了某规则，则在此页面中将会显示与之相对应的错误。请阅读该错误以确定为什么 DRS 不能满足特定规则。

注意 DRS 规则与单个主机的 CPU 关联性规则不同。

创建 DRS 规则

可以创建 DRS 规则以指定虚拟机关联性或反关联性。

步骤

1 在 vSphere Client 清单中选择群集。

2 右键单击并选择**编辑设置**。

3 在左面板的 **VMware DRS** 下选择**规则**。

4 单击**添加**。

5 在“虚拟机规则”对话框中，命名该规则。

6 在弹出菜单中选择一个选项：

- **聚集虚拟机**

  一个虚拟机不能设置多条这样的规则。

- **单独的虚拟机**

  这种类型的规则不能包含两个以上（不含两个）的虚拟机。

7 单击**添加**，然后单击**确定**。

此时会创建该规则。

编辑 DRS 规则

可以编辑 DRS 规则。

步骤

1 显示清单中的群集。

2 右键单击群集并选择**编辑设置**。

此时将显示群集的“设置”对话框。

3 在左窗格的 **VMware DRS** 下，选择**规则**。

4 在右窗格中选择规则，然后单击**编辑**。

5 在对话框中进行更改，然后单击**确定**。

禁用 DRS 规则

可以禁用 DRS 规则。

步骤

1 在 vSphere Client 清单中选择群集。

2 从右键单击菜单中选择**编辑设置**。

3 在左面板中选择**规则**（在 **VMware DRS** 下）。

4 取消选中规则左侧的复选框，然后单击**确定**。

下一步

稍后可以通过重新选中该复选框启用该规则。

删除 DRS 规则

可以删除 DRS 规则。

步骤

1 在 vSphere Client 清单中选择群集。

2 从右键单击菜单中选择**编辑设置**。

3 在左面板中选择**规则**（在 **VMware DRS** 下）。

4 选择要移除的规则并单击**移除**。

此时会删除该规则。

将主机添加到群集

对于由同一 vCenter Server 管理的主机（受管主机）和未由该服务器管理的主机，将主机添加到群集的步骤有所不同。

添加某个主机之后，部署到该主机的虚拟机将变为群集的一部分，而且 DRS 会建议将某些虚拟机迁移到群集内的其他主机。

将受管主机添加到群集

当将 vCenter Server 正在管理的独立主机添加到 DRS 群集时，该主机的资源将与群集相关联。

可以决定是要将现有的虚拟机和资源池与群集的根资源池相关联，还是移植资源池层次结构。

注意 如果主机没有子资源池或虚拟机，其资源将添加到群集，但不会创建带有顶层资源池的资源池层次结构。

步骤

1 从清单或列表视图中选择主机。

2 将主机拖至目标群集对象。

3 选择要对主机的虚拟机和资源池执行的操作。

- **将此主机的虚拟机放入群集的根资源池中**

  vCenter Server 会移除主机上所有现有的资源池，而该主机层次结构中的虚拟机都将被附加到根。因为份额分配是相对于资源池的，而上述操作破坏了资源池层次结构，所以在选择此选项后可能必须手动更改虚拟机的份额。

- **为此主机的虚拟机和资源池创建资源池**

  vCenter Server 创建将成为群集的直接子级的顶层资源池并将主机的所有子级添加到新资源池。您可以命名这个新的顶层资源池。默认为**已从 <主机名> 移植**。

此时主机即会添加到群集。

将非受管主机添加到群集

可将非受管主机添加到群集。该主机当前并未由群集所在的 vCenter Server 系统管理，而且在 vSphere Client 中不可见。

步骤

1 选择要添加主机的群集，然后在右键单击菜单中选择**添加主机**。

2 输入主机名、用户名和密码，然后单击**下一步**。

3 查看摘要信息并单击**下一步**。

4 选择要对主机的虚拟机和资源池执行的操作。

- **将此主机的虚拟机放入群集的根资源池中**

  vCenter Server 会移除主机上所有现有的资源池，而该主机层次结构中的虚拟机都将被附加到根。因为份额分配是相对于资源池的，而上述操作破坏了资源池层次结构，所以在选择此选项后可能必须手动更改虚拟机的份额。

- **为此主机的虚拟机和资源池创建资源池**

  vCenter Server 创建将成为群集的直接子级的顶层资源池并将主机的所有子级添加到新资源池。您可以命名这个新的顶层资源池。默认为**已从 <主机名> 移植**。

此时主机即会添加到群集。

将虚拟机添加到群集

可通过以下三种方式将虚拟机添加到群集。

- 如果将某个主机添加到一个群集，则该主机上的所有虚拟机均会添加到此群集。
- 当创建虚拟机时，创建新的虚拟机向导会提示您选择放置虚拟机的位置。可以选择独立主机或群集并选择主机或群集内的任意资源池。
- 可以使用迁移虚拟机向导将虚拟机从一台独立主机迁移到一个群集或者从一个群集迁移到另一个群集。要启动此向导，请将虚拟机对象拖到群集对象上或右键单击虚拟机名称，然后选择**迁移**。

  注意 可以直接将虚拟机拖到群集内的资源池。在这种情况下，迁移虚拟机向导会启动，但是资源池选择页不会显示。因为资源池控制资源，所以不允许直接向群集内的主机迁移。

从群集内移除主机

可以从群集内移除主机。

前提条件

从 DRS 群集内移除主机之前，请考虑将涉及到的问题。

- 资源池层次结构 – 即使在将某个主机添加到群集时使用了 DRS 群集并决定移植主机资源池，在将该主机从群集内移除后，其上也只保留根资源池。在这种情况下，层次结构将随群集保留。可以创建一个特定于主机的资源池层次结构。

  注意 必须先将主机置于维护模式，才能将其从群集内移除。相反，如果先断开主机的连接，然后再将其从群集内移除，则主机会保留反映群集层次结构的资源池。

- 虚拟机 – 主机必须处于维护模式才能从群集中移除，而且对于要进入维护模式的主机，必须将所有已启动的虚拟机迁移出该主机。当请求主机进入维护模式时，会询问您是否要将该主机上所有已关闭的虚拟机迁移到群集内的其他主机上。
- 无效群集 – 当从群集内移除主机时，可供群集使用的资源会减少。如果群集有足够的资源用于满足群集内所有虚拟机和资源池的预留需要，则群集会调整资源的分配以反映减少的资源量。如果群集没有足够的资源满足所有资源池的预留需要，但是有足够的资源满足所有虚拟机的预留需要，就会出现警报，而且该群集会被标记为黄色。DRS 继续运行。

步骤

1 选择主机，然后在右键单击菜单中选择**进入维护模式**。

2 主机处于维护模式后，可以将其拖到其他清单位置，该位置可以是顶层数据中心或者其他群集。

  移动主机时，主机的资源会从群集内移除。如果将主机的资源池层次结构移植到群集上，则该层次结构将随群集保留。

移动主机后，可以：

- 将主机从 vCenter Server 中移除。（在右键单击菜单中选择**移除**。）
- 在 vCenter Server 下将主机作为独立主机运行。（在右键单击菜单中选择**退出维护模式**。）
- 将主机移至另一个群集。

使用维护模式

当需要维护主机时（例如，要安装更多内存），请将主机置于维护模式。主机仅会因用户要求而进入或离开维护模式。

如果主机将进入维护模式，则需将其上正在运行的虚拟机迁移到其他主机。此时主机将处于**进入维护模式**这一状况，直到关闭所有正在运行的虚拟机或将虚拟机迁移到其他主机为止。如果主机正在进入维护模式，则无法启动其上的虚拟机，也无法将虚拟机迁移到该主机。

当主机上不再有正在运行的虚拟机时，该主机的图标将发生变化，并新增显示**维护模式**，并且该主机的“摘要”面板会指示新的状况。在维护模式下，主机不允许您部署虚拟机，也不允许您启动虚拟机。

注意 如果主机进入所请求的模式后会违反 VMware HA 故障切换级别，则 DRS 不会建议将任何虚拟机从进入维护或待机模式的主机中迁出（在全自动模式下，则不执行这样的迁移）。

使用待机模式

将主机置于待机模式时，会将其关闭。

通常，主机由 VMware DPM 功能置于待机模式以优化电源使用情况。还可以手动将主机置于待机模式。但是，DRS 可能会在其下次运行时撤消（或建议撤消）更改。要强制主机保持关闭状态，请将其置于维护模式并将其关闭。

从群集内移除虚拟机

可以从群集内移除虚拟机。

可通过以下两种方式从群集内移除虚拟机：

- 当从群集内移除主机时，所有未迁移到其他主机的已关闭的虚拟机也会被移除。主机只有在维护模式或断开的情况下才可以被移除。如果从 DRS 群集内移除主机，群集可能会因群集过载而变成黄色。
- 可以使用迁移虚拟机向导将虚拟机从一个群集迁移到一台独立主机或者从一个群集迁移到另一个群集。要启动此向导，请将虚拟机对象拖到群集对象上或右键单击虚拟机名称，然后选择**迁移**。

  如果虚拟机属于 DRS 群集规则组，则 vCenter Server 会在允许迁移之前显示警告。该警告提示从属的虚拟机没有自动迁移。必须在执行迁移操作之前确认该警告。

DRS 群集有效性

vSphere Client 会指示 DRS 群集是有效、过载（黄色）还是无效（红色）。

DRS 群集由于多个原因而变得过载或无效。

- 群集可能由于一台主机发生故障而过载。
- 如果 vCenter Server 不可用，并且使用与 ESX/ESXi 主机直接相连的 vSphere Client 启动虚拟机，则 DRS 群集将变为无效。
- 如果用户在虚拟机进行故障切换时减少父资源池上的预留，则群集将变为无效。
- 如果在 vCenter Server 不可用时使用与 ESX/ESXi 主机相连的 vSphere Client 对主机或虚拟机进行更改，则这些更改将生效。但是，当 vCenter Server 再次可用时，您可能会发现群集由于不再满足群集要求而变为红色或黄色。

当考虑群集有效性情况时，应当了解以下术语。

预留	保证分配给资源池的固定量，由用户输入。
使用的预留	预留总量或每个子级资源池所使用的预留量（以较大者为准），以递归方式相加。
未预留	这个非负数会根据资源池类型不同而有所不同。

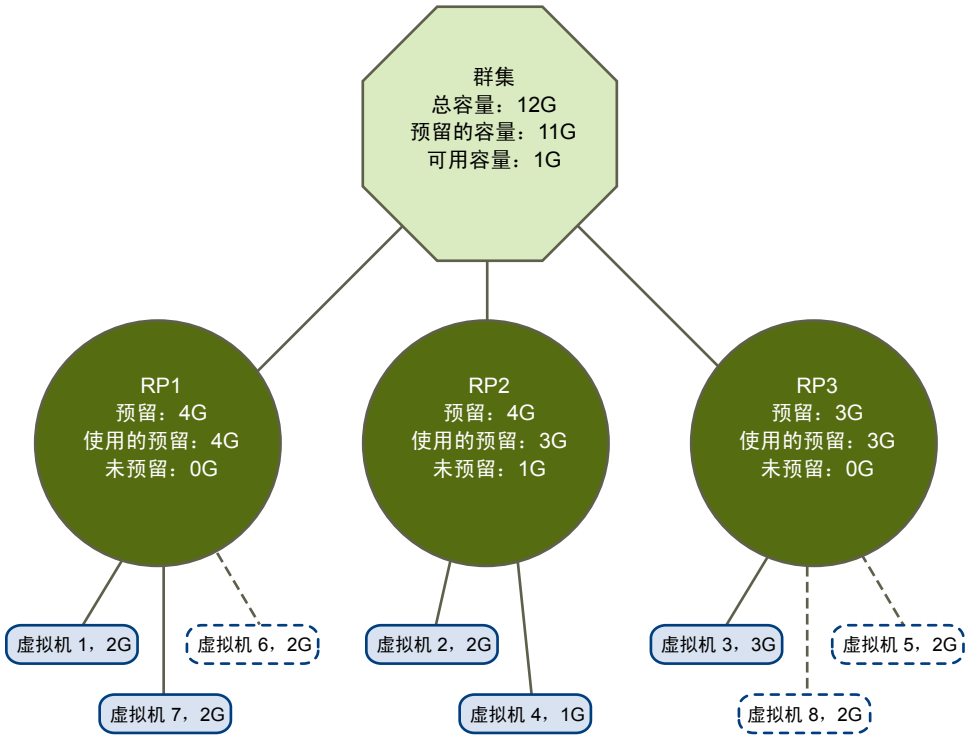
不可扩展的资源池	预留减去已使用的预留。
可扩展的资源池	（预留减去已使用的预留）加上任何可从祖先资源池借来的未预留资源。

有效 DRS 群集

有效群集拥有足够资源来满足所有预留以及支持所有正在运行的虚拟机。

图 6–1 显示具有固定资源池的有效群集的示例以及如何计算其 CPU 和内存资源。

图 6–1。 具有固定资源池的有效群集

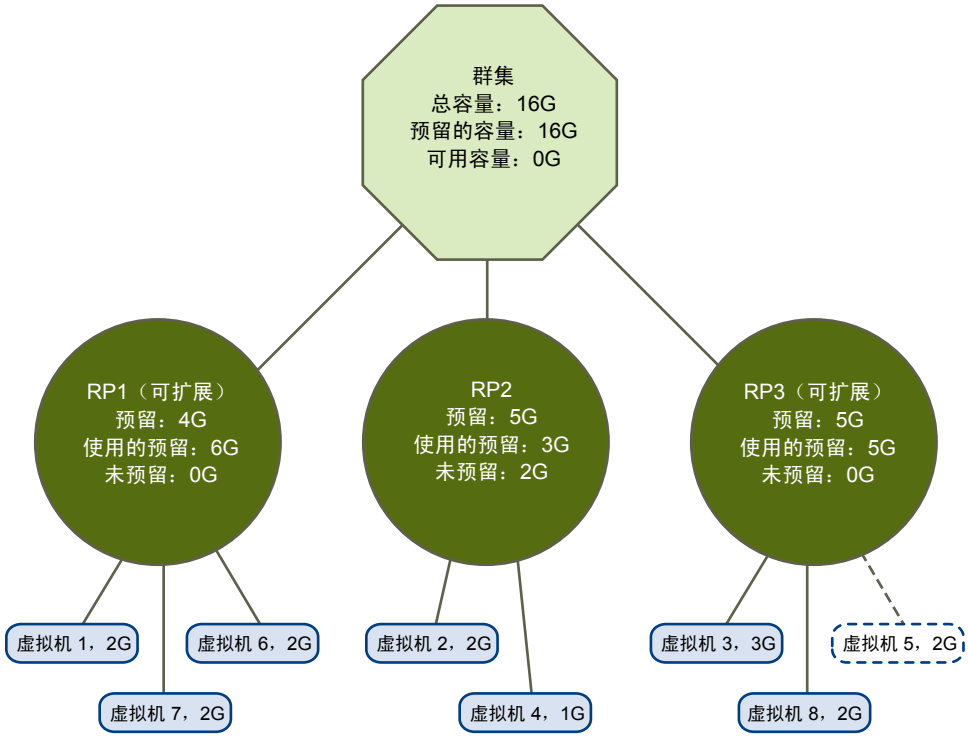


该群集具有以下特性：

- 总资源为 12GHz 的群集。
- 三个类型均为**固定**（未选择**可扩展预留**）的资源池。
- 三个资源池合起来的总预留为 11GHz (4+4+3GHz)。总数显示在群集的**预留的容量**字段中。
- RP1 是使用 4GHz 预留量创建的。两个虚拟机。启动了 VM1 和 VM7，分别各占 2GHz（**使用的预留**：4GHz）。未剩下资源用于启动额外的虚拟机。VM6 显示为未启动。它不消耗任何预留。
- RP2 是使用 4GHz 预留量创建的。启动了两个虚拟机，分别各占 1GHz 和 2GHz（**使用的预留**：3GHz）。还剩 1GHz 未预留。
- RP3 是使用 3GHz 预留量创建的。启动了一个占用 3GHz 的虚拟机。没有资源可于启动额外的虚拟机。

图 6–2 举例说明具有某些资源池（RP1 和 RP3）的有效群集，这些资源池的预留类型为**可扩展**。

图 6–2。 具有可扩展资源池的有效群集



可按如下方式配置有效群集：

- 总资源为 16GHz 的群集。
- RP1 和 RP3 的类型为**可扩展**，RP2 的类型为“固定”。
- 这三个资源池合起来所使用的总预留是 16GHz（其中 RP1 占 6GHz，RP2 占 5GHz，RP3 占 5GHz）。16GHz 显示为顶层群集的**预留的容量**。
- RP1 是使用 4GHz 预留量创建的。启动了三个虚拟机，分别各占用 2GHz。这些虚拟机中的两个（例如，VM1 和 VM7）可以使用 RP1 的预留，第三个虚拟机 (VM6) 可以使用群集资源池中的预留。（如果此资源池的类型为**固定**，则无法启动额外的虚拟机。）
- RP2 是使用 5GHz 预留量创建的。启动了两个虚拟机，分别各占 1GHz 和 2GHz（**使用的预留**：3GHz）。还剩 2 GHz 未预留。

  RP3 是使用 5GHz 预留量创建的。启动了两个虚拟机，分别各占 3GHz 和 2GHz。即使此资源池的类型为**可扩展**，也无法启动额外的 2GHz 虚拟机，因为父资源池的额外资源已被 RP1 占用。

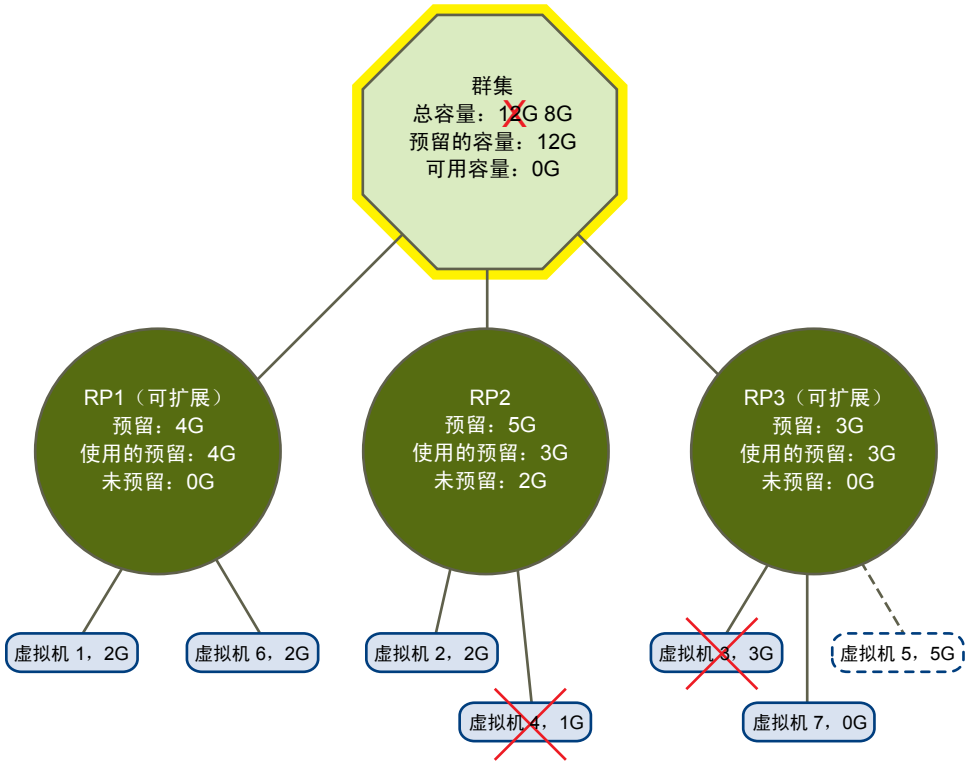
过载的 DRS 群集

当资源池和虚拟机的树在内部是一致的，但群集内没有足够容量来支持子资源池所预留的所有资源时，群集将会变为过载（黄色）。

始终会有足够的资源来支持所有正在运行的虚拟机，因为当主机不可用时，其所有的虚拟机也不可用。当群集容量突然减少时（例如，群集内的一台主机不可用时），群集通常会变为黄色。VMware 建议留足额外的群集资源，以避免群集变为黄色。

请考虑以下示例，如图 6–3 中所示。

图 6–3。 黄色群集



在此示例中：

- 总资源为 12GHz（分别来自三台各有 4GHz 资源的主机）的群集。
- 预留了总共 12GHz 资源的三个资源池。
- 三个资源池合起来所使用的总预留为 12GHz (4+5+3GHz)。该数值显示为群集内**预留的容量**。
- 由于其中一个 4GHz 主机不可用，因此总资源减少至 8GHz。
- 同时，故障主机上运行的 VM4 (1GHz) 和 VM3 (3GHz) 都不再运行。
- 该群集现在正在运行的虚拟机总共需要 6GHz 资源。该群集仍有 8GHz 的资源可用，足够满足虚拟机需求。

  由于不再能达到 12GHz 的资源池预留，因此群集会被标记成黄色。

无效 DRS 群集

当树内部不再一致，即未遵守资源限制时，已启用 DRS 的群集会变为无效（红色）。

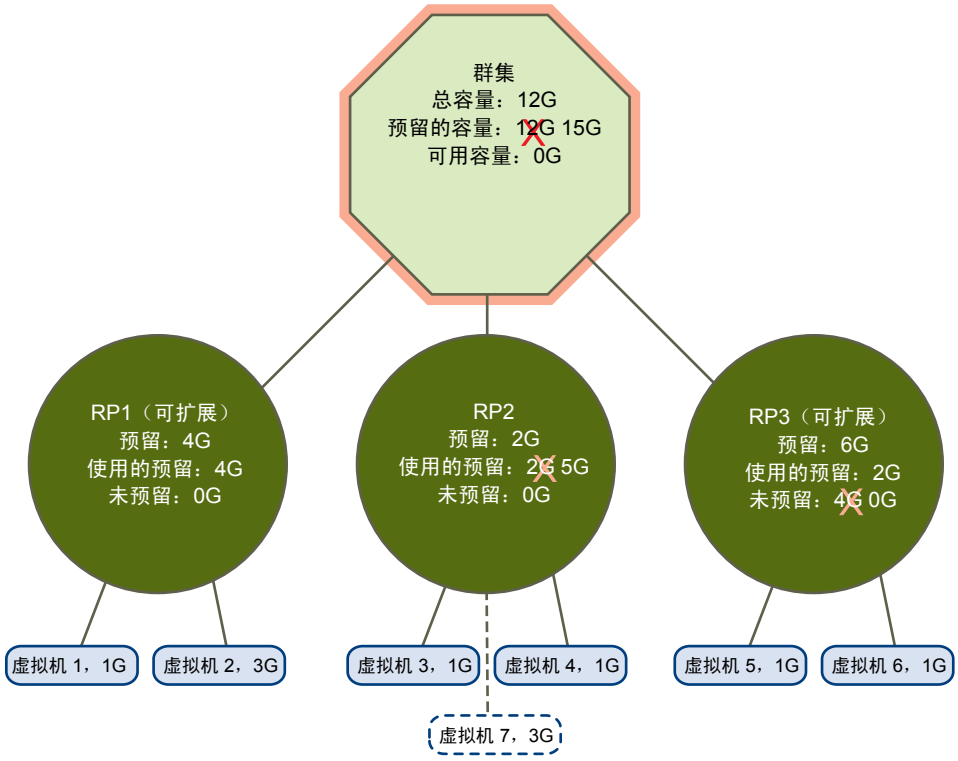
群集内资源的总数量与该群集是否为红色并无直接关联。如果在子级别中存在不一致，即使在根级别中存在足够的资源，群集也可能为红色。

可通过关闭一个或多个虚拟机、将虚拟机移至树中有足够资源的部分或者编辑红色部分的资源池设置，来解决红色 DRS 群集问题。添加资源通常仅在处于黄色状况时才有用。

如果在虚拟机正在进行故障切换时重新配置资源池，则群集也可能会变为红色。正在进行故障切换的虚拟机会断开连接，并且不会算入父资源池所使用的预留。在故障切换完成前，可减少父资源池的预留。故障切换完成后，会再次将虚拟机资源纳入父资源池计算中。如果池的使用量大于新的预留，则该群集将变为红色。

如图 6–4 中的示例所示，如果用户（以不支持的方式）能够启动一个使用资源池 2 下 3 GHz 预留的虚拟机，则该群集会变为红色。

图 6–4。 红色群集



管理电源资源

通过 VMware 分布式电源管理 (DPM) 功能，DRS 群集可以根据群集资源利用率来启动和关闭主机，从而减少其功耗。

VMware DPM 监控内存和 CPU 资源的群集中所有虚拟机的累积需求，并将其与群集中所有主机的总可用资源量进行比较。如果找到足够的额外容量，则 VMware DPM 会将一台或多台主机置于待机模式，并将其虚拟机迁移到其他主机，然后将其关闭。相反，当认为容量不够时，DRS 会使这些主机退出待机模式（将它们启动），并使用 VMotion 将虚拟机迁移到这些主机上。当进行这些计算时，VMware DPM 不仅考虑当前需求，而且还接受任何用户指定的虚拟机资源预留。

注意 ESX/ESXi 主机不能自动退出待机模式，除非它们正在 vCenter Server 管理的群集中运行。

VMware DPM 可以使用三个电源管理协议之一使主机退出待机模式：智能平台管理接口 (IPMI)、Hewlett-Packard Integrated Lights-Out (iLO) 或 Wake-On-LAN (WOL)。每个协议均需要其各自的硬件支持和配置。如果主机不支持以上任何协议，则无法通过 VMware DPM 将其置于待机模式。如果主机支持多个协议，则将按以下顺序使用协议：IPMI、iLO、WOL。

注意 不要在待机模式中断开主机，或在未启动它的情况下将其从 DRS 群集中移出，否则 vCenter Server 将不能再次启动主机。

为 VMware DPM 配置 IPMI 或 iLO 设置

IPMI 是硬件级别规范，而 Hewlett-Packard iLO 是嵌入式服务器管理技术。它们均介绍并提供用于远程监控和控制计算机的接口。

必须在每台主机上执行以下过程。

前提条件

IPMI 和 iLO 需要硬件底板管理控制器 (BMC) 提供用于访问硬件控制功能的网关，并允许使用串行或 LAN 连接从远程系统访问该接口。即使主机自身已关闭，BMC 仍是启动的。如果已正确启用，则 BMC 可响应远程启动命令。

如果计划将 IPMI 或 iLO 用作唤醒协议，则必须配置 BMC。BMC 配置步骤根据型号而异。有关详细信息，请参见供应商的文档。使用 IPMI，还必须确保 BMC LAN 通道已配置为始终可用且允许操作员特权命令。在某些 IPMI 系统上，当启用“LAN 上的 IPMI”时，必须在 BIOS 中对其进行配置并指定特定的 IPMI 帐户。

仅使用 IPMI 的 VMware DPM 支持基于 MD5 和纯文本的身份验证，但不支持基于 MD2 的身份验证。如果主机的 BMC 报告操作员角色支持并启用了 MD2 的身份验证，则 vCenter Server 使用 MD5。否则，如果 BMC 报告支持和启用了基于纯文本的身份验证，则使用基于纯文本的身份验证。如果既未启用 MD5 身份验证，也未启用纯文本身份验证，则 IPMI 无法与主机配合使用，并且 vCenter Server 将尝试使用 LAN 唤醒。

步骤

1 在 vSphere Client 清单中选择主机。

2 单击**配置**选项卡。

3 单击**电源管理**。

4 单击**属性**。

5 输入以下信息。

- BMC 帐户的用户名和密码。（该用户名必须能够远程启动主机。）
- 与 BMC 关联的网卡的 IP 地址，不同于主机的 IP 地址。该 IP 地址应是具有无限租期的静态或 DHCP 地址。
- 与 BMC 关联的网卡的 MAC 地址。

6 单击**确定**。

测试 VMware DPM 的 LAN 唤醒

如果根据 VMware 准则配置用于 VMware DPM 功能的 LAN 唤醒并成功对其进行测试，系统将完全支持对 WOL 的使用。为群集首次启用 VMware DPM 之前，或在要添加到正在使用 VMware DPM 的群集的任何主机上，必须执行以下步骤。

前提条件

在测试 WOL 之前，请确保群集满足先决条件。

- 群集必须至少包含两个 ESX 3.5（或 ESX 3i 版本 3.5）或更高版本的主机。
- 每台主机的 VMotion 网络链路必须工作正常。VMotion 网络还应当是单个 IP 子网，而不是由路由器分隔的多个子网。
- 每台主机上的 VMotion 网卡都必须支持 WOL。要检查 WOL 支持，请首先通过在 vSphere Client 的“清单”面板中选择主机，再选择**配置**选项卡，然后单击**网络**以确定对应于 VMkernel 端口的物理网络适配器的名称。获取此信息后，单击**网络适配器**，并查找对应于网络适配器的条目。相关适配器应在**支持 LAN 唤醒**列中显示“是”。
- 要显示主机上每个网卡的 WOL 兼容状态，请在 vSphere Client 的“清单”面板中选择主机，再选择**配置**选项卡，然后单击**网络适配器**。网卡应在**支持 LAN 唤醒**列中显示“是”。
- 每个支持 WOL 的 VMotion 网卡所插入到交换机端口应设置为自动协商链路速度，而不是设置为固定速度（例如，1000 Mb/s）。当主机关闭时，许多网卡仅在可切换到 100 Mb/s 或更慢速度时，才支持 WOL。

在验证这些必备条件之后，测试每个将使用 WOL 支持 VMware DPM 的 ESX/ESXi 主机。在测试这些主机时，请确保已针对该群集禁用了 VMware DPM 功能。

小心 确保对添加到 VMware DPM 群集（将 WOL 用作唤醒协议）的任何主机进行测试，如果测试失败，则禁止其使用电源管理。如果未完成此操作，则 VMware DPM 可能会关闭随后无法再次启动的主机。

步骤

1 在 vSphere Client 中主机的**摘要**选项卡上单击**进入待机模式**命令。

此操作将关闭主机。

2 通过在主机的**摘要**选项卡上单击**启动**命令，尝试使主机退出待机模式。

3 观察主机是否再次成功启动。

4 对于未能成功退出待机模式的任何主机，在“群集设置”对话框的**主机选项**页面中选择主机并将其**电源管理**设置更改为“已禁用”。

执行此操作后，VMware DPM 不会将该主机作为要关闭的候选主机。

为 DRS 群集启用 VMware DPM

在执行了每台主机上正使用的唤醒协议所需的任何配置或测试步骤后，可以启用 VMware DPM。

要执行此操作，请配置电源管理自动化级别、阈值和主机级替代项。这些设置在群集的设置对话框中的**电源管理**下进行配置。

自动化级别

是否自动执行由 VMware DPM 生成的主机电源状况和迁移建议取决于为该功能选择的电源管理自动化级别。

该自动化级别在群集的设置对话框中的**电源管理**下进行配置。可用选项包括：

- 关闭 - 禁用该功能且不提供建议。
- 手动 - 提供主机电源操作和相关虚拟机迁移建议，但不自动执行。这些建议显示在 vSphere Client 中群集的 **DRS** 选项卡上。
- 自动 - 如果可以自动执行虚拟机迁移，则将自动执行相关的主机电源操作。

注意 电源管理自动化级别与 DRS 自动化级别不同。

VMware DPM 阈值

由 VMware DPM 功能生成的电源状况（主机启动或关闭）建议按优先级进行分配，建议范围为 — 从优先级 1 到优先级 5。

这些优先级分类的基础为：DRS 群集内资源的过度利用率或利用率不足，以及预期对主机电源状况的改善。。优先级 1 的建议是强制性的，而优先级 5 的建议仅带来轻微改善。

该阈值在群集的设置对话框中的**电源管理**下进行配置。每次将 VMware DPM 阈值滑块向右移动一个级别后，都会使自动执行的一组建议的优先级或显示为要手动执行的建议的优先级下降一个级别。在“保守”设置中，VMware DPM 仅生成优先级 1 的建议，向右的下一级别则生成优先级 2 以及更高级别的建议，然后依次类推，直至“激进”级别，该级别生成优先级 5 的建议以及更高级别的建议（即，生成所有建议）。

注意 DRS 阈值和 VMware DPM 阈值本质上是相互独立的。您可以区分它们分别提供的迁移和主机电源状况建议的激进程度。

主机级替代项

在 DRS 群集内启用 VMware DPM 时，默认情况下，群集内的所有主机都将继承其 VMware DPM 自动化级别。

通过选择群集的设置对话框的**主机选项**页面并单击其**电源管理**设置，可以替代单个主机的此默认值。可以将此设置更改为以下选项：

- 已禁用
- 手动
- 自动

注意 如果由于退出待机模式测试失败而将主机的电源管理设置为禁用，请不要更改其设置。

在启用和运行 VMware DPM 之后，通过查看每台主机的**上次退出待机模式的时间**信息（显示在“群集设置”对话框的**主机选项**页面以及每个群集的**主机**选项卡上），可以验证它是否正常工作。此字段会显示时间戳，以及 vCenter Server 上次尝试使主机退出待机模式的结果是**成功**还是**失败**。如果未曾进行此类尝试，则字段会显示**从不**。

注意 **上次退出待机模式的时间**字段的时间派生自 vCenter Server 事件日志。如果清除此日志，则时间将重置为**从不**。

监控 VMware DPM

可以在 vCenter Server 中使用基于事件的警报来监控 VMware DPM。

使用 VMware DPM 时可能出现的最严重的错误是：主机在 DRS 群集需要其容量时无法退出待机模式。可以使用在 vCenter Server 中预先配置的**退出待机错误**警报来监控出现此错误时的情况。如果 VMware DPM 无法使主机退出待机模式（vCenter Server 事件 `DrsExitStandbyModeFailedEvent`），可以将此警报配置为向管理员发送警示电子邮件或者使用 SNMP 陷阱发送通知。默认情况下，在 vCenter Server 能够成功连接到该主机之后，将清除此警报。

要监控 VMware DPM 活动，还可以为以下 vCenter Server 事件创建警报。

表 6–1。 vCenter Server 事件

事件类型	事件名称
正在进入待机模式（即将关闭主机）	`DrsEnteringStandbyModeEvent`
已成功进入待机模式（主机已成功关闭）	`DrsEnteredStandbyModeEvent`
正在退出待机模式（即将启动主机）	`DrsExitingStandbyModeEvent`
已成功退出待机模式（已成功启动）	`DrsExitedStandbyModeEvent`

有关创建和编辑警报的详细信息，请参见《基本系统管理》指南。

如果是使用监控软件而不是 vCenter Server，并且该软件会在意外关闭物理主机时触发警报，那么，当 VMware DPM 使主机进入待机模式时可能会出现生成“无效”警报的情况。如果不希望接收这些警报，请配合供应商部署一个与 vCenter Server 集成的监控软件版本。还可以使用 vCenter Server 本身作为监控解决方案，因为从 vSphere 4.x 开始，它本身能够识别 VMware DPM 且不会触发这些无效警报。

查看 DRS 群集信息 7

通过使用 vSphere Client 中群集**摘要**和 **DRS** 选项卡，可以查看有关 DRS 群集的信息。还可以应用显示在 **DRS** 选项卡中的 DRS 建议。

本章讨论了以下主题：

查看群集摘要选项卡

可以从 vSphere Client 的“清单”面板访问群集的“摘要”选项卡。

此选项卡的“常规”、“VMware DRS”和“VMware DRS 资源分发”部分显示有关群集的配置和操作的有用信息。以下各节将介绍这些部分中出现的字段。

群集摘要选项卡常规区域

群集的“摘要”选项卡的“常规”区域提供有关群集的常规信息。

表 7–1。 常规区域

字段	描述
VMware DRS	表示 VMware DRS 是否已启用。
VMware HA	表示 VMware HA 是否已启用。
VMware EVC 模式	表示增强型 VMotion 兼容性是否已启用。
总 CPU 资源	分配给该群集的 CPU 资源总量。
总内存	分配给该群集的内存资源总量。
主机数	该群集上主机的数目。
处理器总数	该群集上所有主机中的处理器数目。
虚拟机数目	该群集上虚拟机的数目。
使用 VMotion 的总迁移数	群集中所执行的迁移总数。

群集摘要选项卡 VMware DRS 区域

仅当启用了 VMware DRS 时，VMware DRS 区域才会出现在群集“摘要”选项卡中。

表 7–2。 VMware DRS 区域

字段	描述
迁移自动化级别	手动、半自动和全自动。
电源管理自动化级别	关闭、手动和自动。
DRS 建议	正在等待用户确认的 DRS 迁移建议的数量。如果该值非零，则打开群集的 **DRS** 选项卡的“建议”页面。
DRS 故障	当前未完成的 DRS 故障的数量。如果该值非零，则打开群集的 **DRS** 选项卡的“故障”页面。
迁移阈值	表示要应用或生成的迁移建议的优先级。
目标主机负载标准偏差	从迁移阈值设置派生的值，低于此值，将处于负载不平衡状态。
当前主机负载标准偏差	指示群集内当前负载不平衡的值。此值应当小于目标主机负载标准偏差，除非有未应用的 DRS 建议或限制阻止达到该级别。
查看资源分发图表	打开可提供 CPU 和内存使用情况信息的资源分发图表。

VMware DRS 资源分发图表

VMware DRS 资源分发图表显示 CPU 和内存的使用情况信息。

单击 VMware DRS 群集的**摘要**选项卡上的“查看资源分发图表”链接，可打开此图表。

CPU 利用率

CPU 利用率按单个虚拟机显示，并按主机分组。图表用彩色框显示每个虚拟机的信息，不同颜色表示已向其递送的授权资源的百分比（由 DRS 计算）。如果虚拟机正在接受其可用量，则此框应为绿色。如果较长时间处于非绿色，您可能需要调查造成此情况的原因（例如，未应用的建议）。

如果将指针放在与虚拟机相对应的框上，则将显示其使用情况信息（已消耗与可用量）。

通过单击“%”或“MHz”按钮，可以在相应模式间切换显示 CPU 资源信息。

内存利用率

内存利用率按单个虚拟机显示，并按主机分组。

如果将指针放在与虚拟机相对应的框上，则将显示其使用情况信息（已消耗与可用量）。

通过单击“%”或“MHz”按钮，可以在相应模式间切换显示内存资源信息。

使用 DRS 选项卡

DRS 选项卡在从 vSphere Client 的“清单”面板中选择 DRS 群集对象时可用。

此选项卡显示有关为群集提供的 DRS 建议、在应用此类建议时发生的故障和 DRS 操作历史记录的信息。可以从该选项卡访问三个页面。这些页面的名称为“建议”、“故障”和“历史记录”。

DRS 建议页面

可通过单击 DRS 选项卡上的**建议**按钮访问此页面。

DRS 选项卡的“建议”页面显示以下群集属性。

表 7–3。 DRS 建议页面

字段	描述
迁移自动化级别	DRS 虚拟机迁移建议的自动化级别。**全自动**、**半自动**或**手动**。
电源管理自动化级别	VMware DPM 建议的自动化级别。**关闭**、**手动**或**自动**。
迁移阈值	要应用的 DRS 建议的优先级（或更高级别）。
电源管理阈值	要应用的 VMware DPM 建议的优先级（或更高级别）。

另外，该页的“DRS 建议”部分还会显示为了通过迁移或电源管理优化群集内的资源利用率而生成的一组最新建议。此列表上仅显示等待用户确认的手动建议。

可从该页面执行的操作：

- 要刷新建议，请单击**运行 DRS**，建议即得到更新。此命令会显示在全部三个 DRS 页面上。
- 要应用所有建议，请单击**应用建议**。
- 要应用建议的子集，请选中**替代 DRS 建议**复选框。这将激活每个建议旁边的**应用**复选框。选中每个所需建议旁边的复选框，然后单击**应用建议**。

表 7–4 显示了 DRS 为每个建议提供的信息。

表 7–4。 DRS 建议信息

列	描述
优先级	所提供建议的优先级 (1-5)。优先级 1（最高级别）表示因主机正进入维护或待机模式或违反 DRS 规则而需要强制移动。其他级别表示建议能在多大程度上提高群集性能，从优先级 2（显著提高）到优先级 5（轻微提高）。在 ESX/ESXi 4.0 的先前版本中，建议采取星级（1 到 5 星）而不是优先级。星级越高，越需要移动。有关优先级计算的信息，请参见 VMware 知识库文章，网址为 http://kb.vmware.com/kb/1007485。
建议	由 DRS 建议的操作。本列显示的内容取决于所提供建议的类型。 ■ 对于虚拟机迁移：要迁移的虚拟机的名称、（虚拟机当前正在其上运行的）源主机以及（虚拟机要迁移到其上的）目标主机。 ■ 对于主机电源状况更改：要启动或关闭的主机的名称。
原因	提供建议的原因。DRS 之所以建议迁移虚拟机或转换主机电源状况，原因可能与以下任一情况有关。 ■ 平衡平均 CPU 或内存负载。 ■ 满足 DRS（关联性或反关联性）规则。 ■ 主机正在进入维护模式。 ■ 降低功耗。 ■ 关闭特定主机。 ■ 增加群集容量。 ■ 平衡 CPU 或内存预留。 ■ 保持未预留的容量。

DRS 建议只能使用 vCenter Server 来配置。如果 vSphere Client 与 ESX/ESXi 主机直接相连，则迁移功能不可用。要使用迁移功能，请通过 vCenter Server 管理主机。

DRS 故障页面

DRS 选项卡的**故障**页面显示阻止提出 DRS 操作建议（处于手动模式）或应用 DRS 建议（处于自动模式）的故障。

可通过单击 DRS 选项卡上的**故障**按钮访问此页面。

可以使用“包含”文本框自定义问题的显示。从文本框旁的下拉框中选择搜索条件（时间、问题和目标），然后输入相关的文本字符串。

可单击问题以显示有关该问题的其他详细信息，包括特定故障及其阻止的建议。如果单击故障名称，则《DRS 故障排除指南》将提供该故障的详细说明。还可通过单击**查看《DRS 故障排除指南》**从**故障**页面访问此指南。

对于每个故障，DRS 均将提供表 7–5 中显示的信息。

表 7–5。 DRS 故障页面

字段	描述
时间	故障发生时的时间戳。
问题	对阻止提出或应用建议的条件的描述。选择此字段后，将在“问题详细信息”框中显示与其关联的故障的详细信息。
目标	所需操作的目标。

DRS 历史记录页面

DRS 选项卡的“历史记录”页面显示最近根据 DRS 建议采取的操作。

可通过单击 DRS 选项卡上的**历史记录**按钮访问此页面。

对于每个操作，DRS 均将提供表 7–6 中显示的信息。

表 7–6。 DRS 历史记录页面

字段	描述
DRS 操作	所采取操作的详细信息。
时间	操作发生时的时间戳。

默认情况下，此页面上的信息将保留 4 个小时，并且会跨会话保留（可以注销会话，而当您再次登录时，该信息仍然可用）。

可以使用“包含”文本框自定义近期操作的显示。从文本框旁的下拉框中选择搜索条件（“DRS 操作”和“时间”），然后输入相关的文本字符串。

配合使用 NUMA 系统和 ESX/ESXi 8

在支持 NUMA（非一致性内存访问）的服务器架构中，ESX/ESXi 支持对 Intel 和 AMD Opteron 处理器的内存访问进行优化。

在了解如何执行 ESX/ESXi NUMA 调度以及 VMware NUMA 算法如何工作之后，可以指定 NUMA 控件以优化虚拟机的性能。

本章讨论了以下主题：

什么是 NUMA?

NUMA 系统是具有多个系统总线的高级服务器平台。可以在单个系统映像中利用大量处理器，具有极高的性价比。

在过去的十年中，处理器时钟速度获得了巨大的提升。但是，几 GHz 的 CPU 需要具备大量的内存带宽，才能有效利用其处理能力。即使是运行占用大量内存的工作负载（例如科学计算应用程序）的单个 CPU，也会受到内存带宽的限制。

在对称多处理 (Symmetric MultiProcessing, SMP) 系统上，这个问题会变得更加严重，因为许多处理器必须竞争同一系统总线上的带宽。一些高端系统通常通过构建高速数据总线来尝试解决这个问题。但是这种解决方案价格昂贵而且可扩展性也受到限制。

NUMA 是一种替代方法，它使用高性能连接将多个具有成本效益的小型节点连接起来。每个节点均包含处理器和内存，很像一个小型 SMP 系统。但是，高级内存控制器允许节点使用所有其他节点上的内存，从而创建了单个系统映像。当处理器访问不在自己节点内的内存（远程内存）时，数据必须通过 NUMA 连接来传输，这种传输的速度比访问本地内存的速度慢。顾名思义，这种技术的内存访问时间是不一致的，而且取决于内存的位置和通过其访问内存的节点。

对操作系统的挑战

因为 NUMA 架构提供单个系统映像，所以通常可以运行没有经过专门优化的操作系统。例如，IBM x440 完全支持 Windows 2000，尽管 Windows 2000 并未针对与 NUMA 配合使用而设计。

在 NUMA 平台上使用这种操作系统有许多缺点。远程内存访问的滞后时间较长，会使处理器得不到充分利用，经常要等待数据传输到本地节点，而且 NUMA 连接会成为具有高内存带宽需求的应用程序的瓶颈。

而且，这种系统上的性能会有很大变化。例如，如果应用程序在一次基准运行时将内存放置在本地，但后来的一次运行碰巧将所有的这些内存放在远程节点上，此时性能就会发生变化。此现象会让容量规划变得困难。最后，多个节点之间的处理器时钟可能会不同步，因此直接读取时钟的应用程序可能会出现错误的行为。

一些高端 UNIX 系统支持在编译器和编程库中进行 NUMA 优化。此支持需要软件开发人员调整和重新编译他们的程序才能获得最佳的性能。针对一个系统进行的优化不能保证在下一代相同的系统上也能正常发挥作用。其他系统允许管理员明确决定运行应用程序的节点。对于要求其所有内存均必须是本地内存的某些应用程序，可能接受这种做法，不过当工作负载变化时会造成管理负担并且会导致节点之间不平衡。

理想情况下，系统软件提供了透明的 NUMA 支持，因此应用程序可以立即受益，无需进行修改。该系统应充分利用本地内存并且智能调度程序，不需要管理员经常干预。最后，该系统必须在不影响公平性或性能的情况下，对不断变化的状况作出良好的响应。

ESX/ESXi NUMA 调度的工作方式

ESX/ESXi 使用复杂的 NUMA 调度程序来动态平衡处理器负载、内存局部性或处理器负载。

1 由 NUMA 调度程序管理的每个虚拟机均分配有主节点。主节点是系统的 NUMA 节点之一，其中包含处理器和本地内存，如系统资源分配表 (SRAT) 所示。

2 将内存分配给虚拟机时，ESX/ESXi 主机优先从主节点分配内存。

3 NUMA 调度程序可以动态更改虚拟机的主节点以响应系统负载的变化。该调度程序可能会将虚拟机迁移到新的主节点，以减少处理器负载的不平衡。因为这可能会导致使用更多远程内存，所以调度程序可能会将虚拟机的内存动态迁移到新的主节点，以改善内存局部性。在改善总体内存局部性的同时，NUMA 调度程序还可能在节点之间交换虚拟机。

一些虚拟机不受 ESX/ESXi NUMA 调度程序管理。例如，如果为虚拟机手动设置了处理器关联性，NUMA 调度程序可能无法管理该虚拟机。如果虚拟机上的虚拟处理器数量超过单个硬件节点上可用的物理处理器内核数，则无法自动管理该虚拟机。未受 NUMA 调度程序管理的虚拟机仍然可以正确运行。但是，这些虚拟机不能从 ESX/ESXi 的 NUMA 优化中受益。

ESX/ESXi 中的 NUMA 调度和内存放置策略可以透明地管理所有虚拟机，因此管理员不需要明确处理在节点之间平衡虚拟机这一复杂事情。

无论客户机操作系统的类型如何，优化措施都可以顺利发挥作用。ESX/ESXi 甚至为不支持 NUMA 硬件的虚拟机（例如 Windows NT 4.0）也提供了 NUMA 支持。因此，即使是使用旧版操作系统，也可以利用新的硬件。

VMware NUMA 优化算法和设置

本节介绍了 ESX/ESXi 在维持资源保证量的同时，用来充分提高应用程序性能的算法和设置。

主节点和初始放置位置

当启动虚拟机时，ESX/ESXi 会向其分配主节点。虚拟机仅在其主节点内的处理器上运行，而且新分配的内存也来自该主节点。

除非虚拟机的主节点更改，否则虚拟机仅使用本地内存，从而避免了与其他 NUMA 节点的远程内存访问相关联的性能损失。

新的虚拟机最初以循环方式分配到主节点，第一个虚拟机分配到第一个节点，第二个虚拟机分配到第二个节点，以此类推。此策略确保在系统的所有节点上均匀地使用内存。

诸如 Windows Server 2003 之类的一些操作系统提供了这一级别的 NUMA 支持（称为初始放置位置）。对于仅运行单个工作负载（例如基准配置，它不会在系统的正常运行时间过程中发生变化）的系统，这可能够用了。但是，初始放置位置还不够完善，不能保证预期支持工作负载变化的数据中心级系统的良好性能和公平性。

要了解仅采用初始放置位置的系统的缺点，请考虑以下示例：管理员启动四个虚拟机，系统将其中两个虚拟机置于第一个节点上，将剩下的两个虚拟机置于第二个节点上。如果第二个节点上的两个虚拟机均停止，或者这两个虚拟机均闲置，则系统将完全不平衡，全部负载都会置于第一个节点上。即使系统允许剩余的虚拟机中可以有一个虚拟机远程运行在第二个节点上，它也会由于所有的内存都保留在原始节点上而遭受严重的性能损失。

动态负载平衡和页面迁移

ESX/ESXi 结合了传统的初始放置位置方法和动态再平衡算法。系统定期（默认情况下每两秒一次）检查各个节点的负载，并且确定是否应通过将虚拟机从一个节点移至另一个节点来再平衡负载。

此计算考虑了虚拟机和资源池的资源设置，以便在不违反公平性或资源可用量的情况下改善性能。

再平衡器选择合适的虚拟机，并将其主节点更改为负载最少的节点。如果可以的话，再平衡器会移动目标节点上已经有一些内存的虚拟机。从此之后（除非再次移动），虚拟机将在新的主节点上分配内存，并且仅在新主节点内的处理器上运行。

再平衡是维持公平性和确保完全使用所有节点的有效解决方案。再平衡器可能需要将虚拟机移至已经分配少量内存或没有分配内存的节点上。这种情况下，虚拟机会遭受与大量远程内存访问相关联的性能损失。ESX/ESXi 通过将内存从虚拟机的原始节点以透明的方式迁移到新的主节点，可以消除该损失：

1 系统选择原始节点上的页（4 KB 连续内存），并将其数据复制到目标节点中的页上。

2 系统使用虚拟机监控层和处理器的内存管理硬件来无缝地重新映射虚拟机的内存视图，因此系统将目标节点上的页用于后续的所有引用，从而消除了远程访问内存所带来的损失。

当虚拟机移至新的节点时，ESX/ESXi 主机立即开始按此方式迁移其内存。主机会管理迁移速率，以避免让系统负担过重，特别是在虚拟机剩下很少的远程内存或目标节点的可用内存很少时。如果虚拟机只是短时间内移至新的节点，则内存迁移算法还可以确保 ESX/ESXi 主机不会进行不必要的内存移动。

当初始放置位置、动态再平衡和智能内存迁移配合使用时，即使工作负载出现变化，也能确保 NUMA 系统的良好内存性能。当主要工作负载出现变化时（例如启动新的虚拟机时），系统需要一些时间来重新调整，将虚拟机和内存迁移到新的位置。经过很短的时间之后（通常是几秒钟或几分钟），系统就可以完成重新调整并达到稳定状况。

针对 NUMA 优化的透明页共享

许多 ESX/ESXi 工作负载存在跨虚拟机共享内存的机会。

例如，几个虚拟机可能正在运行同一客户机操作系统的多个实例，加载了相同的应用程序或组件，或包含公用数据。这些情况下，ESX/ESXi 系统使用专用的透明页共享技术安全地消除了内存页的冗余副本。采用内存共享，在虚拟机中运行的工作负载消耗的内存通常要少于其在物理机上运行时所需的内存。因此，可以高效地支持更高级别的过载。

ESX/ESXi 系统的透明页共享也针对在 NUMA 系统上的使用而经过了优化。在 NUMA 系统上，页按照节点进行共享，因此对于频繁共享的页面，每个 NUMA 节点都有自己的本地副本。当虚拟机使用共享页时，它们不需要访问远程内存。

跨 NUMA 节点和在 NUMA 节点内共享内存页

“VMkernel.Boot.sharePerNode”选项控制内存页是否仅可以在单个 NUMA 节点内共享（删除重复数据），还是可以跨多个 NUMA 节点共享。

“VMkernel.Boot.sharePerNode”默认情况下处于打开状态，并且仅在同一 NUMA 节点内共享相同页。这可改善内存局部性，因为对共享页的所有访问均需使用本地内存。

注意 此默认行为在 ESX 的所有先前版本中亦然如此。

关闭“VMkernel.Boot.sharePerNode”选项后，可以跨不同 NUMA 节点共享相同页。这增加了共享和删除重复数据的数量，从而以内存局部性为代价降低了总体内存消耗。在内存受限的环境（如 VMware View 部署）中，可能需要删除许多相似虚拟机上的重复数据，因此跨 NUMA 节点共享页面可能非常有益。

NUMA 架构中的资源管理

可以使用不同类型的 NUMA 架构进行资源管理。能够提供 NUMA 平台以支持业界标准操作系统的系统包括：基于 AMD CPU 或 IBM 企业 X 型架构的系统。

IBM 企业 X 型架构

IBM 企业 X 型架构是支持 NUMA 的架构之一。

IBM 企业 X 型架构支持最多具有四个节点的服务器（在 IBM 术语中也称为 CEC 或 SMP 扩展联合）。每个节点最多可以包含四个 Intel Xeon MP 处理器，总共 16 个 CPU。下一代 IBM eServer x445 使用增强版本的企业 X 型架构，并扩展为八个节点，每个节点最多四个 Xeon MP 处理器，总共 32 个 CPU。第三代 IBM eServer x460 提供了类似的可扩展性，但另外还支持 64 位 Xeon MP 处理器。所有这些系统的高可扩展性均源于企业 X 型架构的 NUMA 设计；基于 POWER4 的 IBM 高端 pSeries 服务器也采用了该设计。

基于 AMD Opteron 的系统

基于 AMD Opteron 的系统（如 HP ProLiant DL585 Server）也提供了 NUMA 支持。

BIOS 节点交叉设置决定了系统行为更像 NUMA 系统还是更像统一内存架构 (UMA) 系统。请参见《HP Proliant DL585 Server》中的技术摘要。另请参见 HP 网站上的《基于 HP ROM 的安装实用程序用户向导》。

默认情况下，禁用节点交叉，因此每个处理器都有自己的内存。BIOS 生成系统资源分配表 (SRAT)，因此 ESX/ESXi 主机将系统作为 NUMA 来进行检测并应用 NUMA 优化。如果启用节点交叉（也称为交叉内存），则 BIOS 不生成 SRAT，因此 ESX/ESXi 主机不会将系统作为 NUMA 来进行检测。

目前提供的 Opteron 处理器的每个插槽最多有四个内核。当节点内存处于启用状态时，会划分 Opteron 处理器上的内存，以便每个插槽有一些本地内存，但其他插槽的内存则是远程的。单内核 Opteron 系统的每个 NUMA 节点有单个处理器，而双内核 Opteron 系统的每个 NUMA 节点有两个处理器。

SMP 虚拟机（有两个虚拟处理器）无法驻留在具有单个内核的 NUMA 节点内，例如单内核 Opteron 处理器。这也意味着 ESX/ESXi NUMA 调度程序无法管理这些虚拟机。未受 NUMA 调度程序管理的虚拟机仍然可以正确运行。但是，这些虚拟机不会从 ESX/ESXi NUMA 优化中受益。单处理器虚拟机（具有单个虚拟处理器）可以驻留在单个 NUMA 节点内，并且由 ESX/ESXi NUMA 调度程序进行管理。

注意 现在，对于小型 Opteron 系统会在默认情况下禁用 NUMA 再平衡，以确保调度的公平性。可以使用“Numa.RebalanceCoresTotal”和“Numa.RebalanceCoresNode”选项更改此行为。

指定 NUMA 控制

如果您有一些占用大量内存的应用程序或者有少量的虚拟机，可能要通过明确指定虚拟机 CPU 和内存放置位置来优化性能。

如果虚拟机运行占用大量内存的工作负载（例如内存中的数据库或具有大型数据集的科学计算应用程序），这样做很有用。如果已知系统工作负载很简单而且不会变化，您可能还想手动优化 NUMA 放置位置。例如，对于一个由运行 8 个虚拟机而且具有类似工作负载的 8 个处理器组成的系统，很容易进行明确地优化。

注意 大多数情况下，ESX/ESXi 主机的自动 NUMA 优化会产生良好的性能。

ESX/ESXi 为 NUMA 放置位置提供了两组控制，因此管理员可以控制虚拟机的内存和处理器位置。

vSphere Client 允许您指定两个选项。

CPU 关联性 虚拟机应仅使用给定节点上的处理器。

内存关联性 服务器应仅在指定的节点上分配内存。

如果在虚拟机启动前设置了这两个选项，则虚拟机仅在选定的节点上运行，而且其所有的内存均在本地分配。

虚拟机已经开始运行后，管理员还可以手动将虚拟机移至另一个节点。这种情况下，必须手动设置虚拟机的页迁移速率，以便虚拟机前一个节点中的内存可以移至新的节点。

手动设置 NUMA 的放置位置可能会干扰 ESX/ESXi 资源管理算法，这种算法尝试向每个虚拟机赋予公平份额的系统处理器资源。例如，如果将十个虚拟机（具有占用大量处理器的工作负载）手动置于一个节点，并且仅将两个虚拟机手动置于另一个节点，则系统不可能为所有这十二个虚拟机赋予相等份额的系统资源。

注意 可以在 `resxtop`（或 `esxtop`）实用程序的“内存”面板中查看 NUMA 配置信息。

使用 CPU 关联性将虚拟机与单个 NUMA 节点相关联

通过将虚拟机与单个 NUMA 节点上的 CPU 编号相关联（手动 CPU 关联性），可能会改善虚拟机上应用程序的性能。

步骤

1 使用 vSphere Client，右键单击虚拟机并选择**编辑设置**。

2 在“虚拟机属性”对话框中，选择**资源**选项卡并选择**高级 CPU**。

3 在“调度关联性”面板中，为不同的 NUMA 节点设置 CPU 关联性。

注意 必须为 NUMA 节点中的所有处理器手动选择这些框。CPU 关联性是按照处理器指定的，而不是按照节点指定的。

使用内存关联性将内存分配与 NUMA 节点相关联

可以指定虚拟机上所有的后续内存分配使用与单个 NUMA 节点关联的页（也称为手动内存关联性）。当虚拟机使用本地内存时，该虚拟机上的性能会得到改善。

注意 只有在指定了 CPU 关联性时，才能指定要用于以后内存分配的节点。如果仅对内存关联性设置进行了手动更改，则自动 NUMA 再平衡功能将无法正常工作。

步骤

1 使用 vSphere Client，右键单击虚拟机并选择**编辑设置**。

2 在“虚拟机属性”对话框中，选择**资源**选项卡并选择**内存**。

3 在“NUMA 内存关联性”面板中，设置内存关联性。

示例 8–1。 将虚拟机绑定到单个 NUMA 节点

以下示例说明了将最后四个物理 CPU 手动绑定到 8 路服务器上双路虚拟机的单个 NUMA 节点。

CPU（例如 4、5、6 和 7）是物理 CPU 编号。

1 在 vSphere Client“清单”面板中，选择该虚拟机并选择**编辑设置**。

2 选择**选项**并单击**高级**。

3 单击**配置参数**按钮。

4 在 vSphere Client 中，为处理器 4、5、6 和 7 打开 CPU 关联性。

接着，您希望此虚拟机仅在节点 1 上运行。

1 在 vSphere Client“清单”面板中，选择该虚拟机并选择**编辑设置**。

2 选择**选项**并单击**高级**。

3 单击**配置参数**按钮。

4 在 vSphere Client 中，将 NUMA 节点的内存关联性设置为 1。

完成这两个任务可以确保虚拟机仅在 NUMA 节点 1 上运行，并在可能的情况下从同一个节点分配内存。

性能监控实用程序：resxtop 和 esxtop A

通过 resxtop 和 esxtop 命令行实用程序，您可以实时详细查看 ESX/ESXi 使用资源的情况。可以按以下三种模式之一启动任一实用程序：交互（默认）、批处理或重放。

resxtop 和 esxtop 的基本区别在于：resxtop 可以远程（或本地）使用，而 esxtop 只能通过本地 ESX 主机的服务控制台来启动。

本附录讨论了以下主题：

使用 esxtop 实用程序

esxtop 实用程序仅在 ESX 主机的服务控制台上运行，而且使用它必须拥有根用户特权。

使用所需选项键入该命令：

```
esxtop [-] [h] [v] [b] [s] [a] [c filename] [R vm-support_dir_path]  [d delay] [n iter]
```

esxtop 实用程序从 .esxtop4rc 读取其默认配置。该配置文件由八行组成。

前七行包含小写字母和大写字母，指定在 CPU、内存、存储适配器、存储设备、虚拟机存储器、网络和中断面板上以什么顺序显示哪些字段。这些字母对应于各个 esxtop 面板的“字段”或“顺序”面板中的字母。

第八行包含有关其他选项的信息。最重要的是，如果以安全模式保存了配置，那么，不从 .esxtop4rc 文件的第七行移除 s，就不会获得不安全的 esxtop。用一个数字指定更新之间的延迟时间。与交互模式相同，键入 c、m、d、u、v、n 或 I 将确定 esxtop 启动的面板。

注意 不要编辑 .esxtop4rc 文件。请在运行中的 esxtop 进程中选择这些字段和顺序，进行更改，并使用 W 交互命令保存该文件。

使用 resxtop 实用程序

resxtop 实用程序是 vSphere CLI 命令。

必须先下载和安装 vSphere CLI 包，或将 vSphere Management Assistant (vMA) 部署到 ESX/ESXi 主机或 vCenter Server 系统，才可以使用任何 vSphere CLI 命令。

在安装完成之后，从命令行启动 resxtop。对于远程连接，可以直接连接到 ESX/ESXi 主机或通过 vCenter Server 进行连接。

命令行选项与 esxtop（除 R 选项外）相同，但具有附加连接选项。

注意 resxtop 不使用由其他 vSphere CLI 命令共享的所有选项。

表 A–1。 resxtop 命令行选项

选项	描述
[server]	要连接到的远程主机的名称（必需）。如果直接连接到 ESX/ESXi 主机，则使用该主机的名称。如果间接连接到 ESX/ESXi 主机（即通过 vCenter Server 进行连接），则在该选项中使用 vCenter Server 系统的名称。
[vihost]	如果采用间接连接方式（通过 vCenter Server），则此选项应当包含您连接到的 ESX/ESXi 主机的名称。如果直接连接到 ESX/ESXi 主机，则不使用此选项。
[portnumber]	要连接到的远程服务器端口号。默认端口为 443，除非在服务器上更改了这一端口，否则不需要此选项。
[username]	在连接到远程主机时要进行身份验证的用户名。远程服务器会提示输入密码。

也可以通过在命令行上省略 server 选项，使该命令默认为 localhost，以在本地 ESX/ESXi 主机上使用 resxtop。

在交互模式中使用 esxtop 或 resxtop

默认情况下，resxtop 和 esxtop 以交互模式运行。交互模式在不同的面板中显示统计信息。

对于每个面板都提供帮助菜单。

交互模式命令行选项

可以在交互模式中将各种命令行选项与 esxtop 和 resxtop 配合使用。

表 A–2 列出了在交互模式中可用的命令行选项。

表 A–2。 交互模式命令行选项

选项	描述
h	显示 resxtop（或 esxtop）命令行选项的帮助。
v	显示 resxtop（或 esxtop）版本号。
s	以安全模式调用 resxtop（或 esxtop）。在安全模式中，禁用了指定更新之间延迟的 –d 命令。
d	指定更新之间的延迟。默认值为 5 秒。最小值为 2 秒。可以使用交互命令 s 更改此命令。如果指定的延迟少于 2 秒，延迟将设置为 2 秒。
n	迭代次数。对显示执行 n 次更新，然后退出。
server	要连接的远程服务器主机的名称（仅 resxtop 需要）。
portnumber	要连接到的远程服务器上的端口号。默认端口为 443，除非在服务器上更改了这一端口，否则不需要此选项。(仅限 resxtop)
username	连接到远程主机时要进行身份验证的用户名。远程服务器也会提示输入密码（仅限 resxtop）。
a	显示所有统计信息。该选项会替代配置文件设置并显示所有统计信息。配置文件可以是默认的 ~/.esxtop4rc 配置文件或用户定义的配置文件。
c<filename>	加载用户定义的配置文件。如果未使用 -c 选项，则默认配置文件名为 ~/.esxtop4rc。使用 W 单键交互命令创建自己的配置文件，同时指定其他文件名。

公共统计信息描述

当 resxtop（或 esxtop）以交互模式运行时，不同的面板上会显示一些统计信息。以下统计信息是所有四个面板的公共信息。

四个 resxtop（或 esxtop）面板的顶部所显示的“正常运行时间”行显示了当前时间、自上一次重新引导以来所经过的时间、当前运行的环境数量和平均负载。环境是 ESX/ESXi VMkernel 可调度的实体，类似于其他操作系统中的进程或线程。

其下显示的是过去 1 分钟、5 分钟和 15 分钟内的平均负载。平均负载同时考虑了正在运行和准备运行的环境。平均负载为 1.00 表示完全利用了所有物理 CPU。平均负载为 2.00 表示 ESX/ESXi 系统可能需要当前可用数目两倍的物理 CPU。同样，平均负载为 0.50 表示 ESX/ESXi 系统上的物理 CPU 有一半得到了利用。

统计信息列和顺序页

可以定义在交互模式中字段的显示顺序。

如果按下 f、F、o 或 O，系统会显示一个页面，该页面在最上面的一行指定字段顺序和字段内容的简短描述。如果对应于字段的字段字符串中的字母为大写，则显示该字段。字段描述前面的星号表示是否显示字段。

这些字段的顺序对应于字符串中字母的顺序。

从“字段选择”面板中，您可以：

- 通过按下对应的字母，切换字段的显示。
- 通过按下对应的大写字母，向左移动字段。
- 通过按下对应的小写字母，向右移动字段。

交互模式单键命令

以交互模式运行时，resxtop（或 esxtop）可识别几个单键命令。

所有交互模式面板都可以识别表 A–3 中列出的命令。如果已经在命令行上提供 s 选项，则用来指定更新之间延迟的命令会处于禁用状态。所有的交互排序命令按降序排序。

表 A–3。 交互模式单键命令

键	描述
h 或 ?	显示当前面板的帮助菜单，给出命令的简短摘要以及安全模式的状态。
空格	立即更新当前面板。
^L	擦除和重绘当前面板。
f 或 F	显示将统计信息列（字段）添加到当前面板或从当前面板移除统计信息列（字段）的面板。
o 或 O	显示用来更改当前面板上统计信息列顺序的面板。
#	提示您输入要显示的统计信息的行数。只要值大于 0，就会替代根据窗口大小测量自动确定要显示的行数。如果在一个 resxtop（或 esxtop）面板中更改该数值，此更改会影响所有四个面板。
s	提示您输入更新之间的延迟，以秒为单位。小数值可以识别到微秒。默认值为 5 秒。最小值为 2 秒。此命令在安全模式中不可用。
W	将当前设置写入 esxtop（或 resxtop）配置文件。这是写入配置文件的推荐方式。默认文件名是通过 -c 选项指定的文件名，如果不使用 -c 选项，则为 ~/.esxtop4rc。还可以在该 W 命令生成的提示中指定其他文件名。
q	退出交互模式。
c	切换到 CPU 资源利用率面板。
m	切换到内存资源利用率面板。
d	切换到存储（磁盘）适配器资源利用率面板。
u	切换到存储（磁盘）设备资源利用率屏幕。
v	切换到存储（磁盘）虚拟机资源利用率屏幕。

表 A–3。 交互模式单键命令（续）

键	描述
n	切换到网络资源利用率面板。
I	切换到中断面板。

CPU 面板

CPU 面板显示了服务器范围的统计信息以及单个环境、资源池和虚拟机 CPU 利用率的统计信息。

资源池、正在运行的虚拟机或其他环境有时会称为组。对于属于虚拟机的环境，显示正在运行的虚拟机的统计信息。所有其他环境按逻辑方式聚合到包含这些环境的资源池中。

表 A–4 论述了此面板中显示的统计信息。

表 A–4。 CPU 面板统计信息

行	描述
PCPU USED(%)	PCPU 指的是物理硬件执行上下文。如果超线程不可用或已禁用，则它可以是物理 CPU 内核；如果超线程已启用，则可以是逻辑 CPU（LCPU 或 SMT 线程）。 PCPU USED(%) 显示： ■ 每个 PCPU 的 CPU 使用情况百分比 ■ 所有 PCPU 的平均 CPU 使用情况百分比 CPU 使用情况 (%USED) 是自上次屏幕更新以来所使用的 PCPU 名义频率的百分比。它等于在此 PCPU 上运行的环境的 %USED 的总和。 **注意** 如果 PCPU 的运行频率高于其名义（额定）频率，则 PCPU USED(%) 可能大于 100%。
PCPU UTIL(%)	PCPU 指的是物理硬件执行上下文。如果超线程不可用或已禁用，则它可以是物理 CPU 内核；如果超线程已启用，则可以是逻辑 CPU（LCPU 或 SMT 线程）。 PCPU UTIL(%) 表示 PCPU 处于非闲置状态的实际时间百分比（原始 PCPU 利用率），它显示每个 PCPU 的 CPU 利用率百分比和所有 PCPU 的平均 CPU 利用率百分比。 **注意** PCPU UTIL(%) 可能由于电源管理技术或超线程而与 PCPU USED(%) 不同。
CCPU(%)	ESX 服务控制台报告的 CPU 总时间的百分比。如果使用的是 ESXi，则此字段不显示。 ■ `us` — 用户时间百分比。 ■ `sy` — 系统时间百分比。 ■ `id` — 闲置时间百分比。 ■ `wa` — 等待时间百分比。 ■ `cs/sec` — 服务控制台记录的每秒上下文切换次数。
ID	运行中环境内资源池或虚拟机的资源池 ID 或虚拟机 ID，或运行中环境的环境 ID。
GID	运行中环境内资源池或虚拟机的资源池 ID 或虚拟机 ID。
NAME	运行中环境内资源池或虚拟机的名称，或运行中环境的名称。
NWLD	运行中环境内资源池或虚拟机中的成员数量。如果使用交互命令 `e`（请参见交互命令）对组进行扩展，则所生成的全部环境的 NWLD 为 1（一些资源池（如控制台资源池）只有一个成员）。
%STATE TIMES	由以下百分比构成的 CPU 统计信息集合。对于环境，百分比是一个物理 CPU 内核的百分比。
%USED	由资源池、虚拟机或环境使用的物理 CPU 内核周期百分比。%USED 可能取决于 CPU 内核的运行频率。当以较低的 CPU 内核频率运行时，%USED 可能小于 %RUN。在支持涡轮加速模式的 CPU 上，CPU 频率也可能高于名义（额定）频率，并且 %USED 可能大于 %RUN。
%SYS	代表资源池、虚拟机或环境在 ESX/ESXi VMkernel 中处理中断和执行其他系统活动所用的时间百分比。该时间是用于计算“%USED”的时间的一部分。
%WAIT	资源池、虚拟机或环境在阻止或遇忙等待状况所占的时间百分比。该百分比包括资源池、虚拟机或环境闲置的时间百分比。

表 A–4。 CPU 面板统计信息（续）

行	描述
%IDLE	资源池、虚拟机或环境闲置的时间百分比。从“%WAIT”中减去该百分比，可得出资源池、虚拟机或环境等待某个事件所用的时间百分比。VCPU 环境的“%WAIT - %IDLE”之差可用来估计客户机 I/O 等待时间。要查找 VCPU 环境，请使用单键命令 e 展开虚拟机，并搜索以“vcpu”开头的环境 NAME（名称）。（请注意，VCPU 环境可能还会等待除 I/O 事件之外的其他事件，因此，此测量值只是估计。）
%RDY	资源池、虚拟机或环境准备运行的时间百分比，但不是所提供的、要在其上执行的 CPU 资源的时间百分比。
%MLMTD（最大限制）	ESX/ESXi VMkernel 故意未运行资源池、虚拟机或环境的时间百分比，因为如果运行的话，会违反资源池、虚拟机或环境的限制设置。由于资源池、虚拟机或环境在被阻止以此方式运行时准备运行，“%MLMTD”（最大限制）时间也包括在“%RDY”时间内。
%SWPWT	资源池或环境等待 ESX/ESXi VMkernel 交换内存所用的时间百分比。“%SWPWT”（交换等待）时间包括在“%WAIT”时间内。
EVENT COUNTS/s	由每秒事件速率构成的 CPU 统计信息集合。这些统计信息仅供 VMware 内部使用。
CPU ALLOC	由以下 CPU 分配配置参数构成的 CPU 统计信息集合。
AMIN	资源池、虚拟机或环境属性“预留”。
AMAX	资源池、虚拟机或环境属性“限制”。-1 值表示无限制。
ASHRS	资源池、虚拟机或环境属性“份额”。
SUMMARY STATS	由以下 CPU 配置参数和统计信息构成的 CPU 统计信息集合。这些统计信息仅适用于环境，而不适用于虚拟机或资源池。
AFFINITY BIT MASK	显示环境的当前调度关联性的位掩码。
HTSHARING	当前超线程配置。
CPU	当 `resxtop`（或 `esxtop`）获得该信息时，正在运行的环境的物理或逻辑处理器。
HTQ	表示环境当前是否已隔离。“N”表示否，“Y”表示是。
TIMER/s	该环境的定时器速率。
%OVRLP	调度资源池、虚拟机或环境时，代表不同资源池、虚拟机或环境在调度资源池、虚拟机或环境期间所用系统时间的百分比。该时间不包括在“%SYS”中。例如，如果当前正在调度虚拟机 A，而且虚拟机 B 的网络数据包已由 ESX/ESXi VMkernel 处理，则虚拟机 A 所用的时间显示为“%OVRLP”，而虚拟机 B 所用的时间显示为“%SYS”。
%RUN	调度的总时间百分比。该时间不算超线程和系统时间。在支持超线程的服务器上，%RUN 可以是“%USED”大小的两倍。
%CSTP	资源池在就绪、共同取消调度状况中所用的时间百分比。 （注意：您可能会看到该统计信息显示出来，但其仅供 VMware 使用。）

可以如表 A–5 中所述使用单键命令来更改该显示。

表 A–5。 CPU 面板单键命令

命令	描述
e	在展开显示 CPU 统计信息和不展开显示 CPU 统计信息之间切换。 展开显示中包括按属于资源池或虚拟机的各个环境细分的 CPU 资源利用率统计信息。各个环境的所有百分比是单个物理 CPU 的百分比。 考虑以下示例： ■ 如果在 2 路服务器上按资源池细分的“%Used”为 30%，则说明该资源池正在利用两个物理 CPU 30% 的资源。 ■ 如果在 2 路服务器上按属于资源池的环境细分的“%Used”为 30%，则说明该环境正在利用一个物理 CPU 30% 的资源。
U	按资源池或虚拟机的“%Used”列对资源池、虚拟机和环境进行排序。

表 A–5。 CPU 面板单键命令 （续）

命令	描述
R	按资源池或虚拟机的“%RDY”列对资源池、虚拟机和环境进行排序。
N	按 GID 列对资源池、虚拟机和环境进行排序。这是默认的排序顺序。
V	仅显示虚拟机实例。
L	更改“NAME”列的显示长度。

内存面板

内存面板显示了服务器范围和组的内存利用率统计信息。与 CPU 面板类似，组对应于资源池、正在运行的虚拟机或正在消耗内存的其他环境。

内存面板顶部第一行显示了当前时间、自上一次重新引导以来所经过的时间、当前运行的环境数量和内存过载平均值。显示过去 1 分钟、5 分钟和 15 分钟内内存过载的平均值。内存过载为 1.00 表示内存 100% 过载。请参见第 24 页，“内存过载”。

表 A–6。 内存面板统计信息

字段	描述	
PMEM (MB)	显示服务器的计算机内存统计信息。所有数字都以兆字节为单位。	
	total	服务器中计算机内存总量。
	cos	分配给 ESX 服务控制台的计算机内存量。
	vmk	正由 ESX/ESXi VMkernel 使用的计算机内存量。
	other	除 ESX 服务控制台和 ESX/ESXi VMkernel 之外其他各项正在使用的计算机内存量。
	free	可用的计算机内存量。
VMKMEM (MB)	显示 ESX/ESXi VMkernel 的计算机内存统计信息。所有数字都以兆字节为单位。	
	managed	由 ESX/ESXi VMkernel 管理的计算机内存总量。
	min free	ESX/ESXi VMkernel 旨在保持可用的计算机内存最小量。
	rsvd	当前由资源池预留的计算机内存总量。
	ursvd	当前未预留的计算机内存总量。
	state	计算机内存的当前可用性状况。可能的值为 high、soft、hard 和 low。high 表示计算机内存没有任何压力，low 表示有压力。
COSMEM (MB)	显示 ESX 服务控制台报告的内存统计信息。所有数字都以兆字节为单位。如果使用的是 ESXi，则此字段不显示。	
	free	闲置的内存量。
	swap_t	配置的总交换量。
	swap_f	可用的交换量。
	r/s is	从磁盘换入内存的速率。
	w/s	内存交换到磁盘的速率。
NUMA (MB)	显示 ESX/ESXi NUMA 统计信息。只有当 ESX/ESXi 主机正运行在 NUMA 服务器上时，才会显示该行。所有数字都以兆字节为单位。 对于服务器中的每个 NUMA 节点，显示两个统计信息： ■ NUMA 节点中由 ESX/ESXi 管理的计算机内存总量。 ■ 该节点中当前可用的计算机内存量（在圆括号中）。	

表 A–6。 内存面板统计信息（续）

字段	描述	
PSHARE (MB)	显示 ESX/ESXi 页共享统计信息。所有数字都以兆字节为单位。	
	shared	正共享的物理内存量。
	common	环境之间共用的计算机内存量。
	saving	由于页共享而节省的计算机内存量。
SWAP (MB)	显示 ESX/ESXi 交换使用量统计信息。所有数字都以兆字节为单位。	
	curr	当前的交换使用量。
	目标	ESX/ESXi 希望交换使用量所处的位置。
	r/s	由 ESX/ESXi 系统从磁盘换入内存的速率。
	w/s	由 ESX/ESXi 系统将内存交换到磁盘的速率。
MEMCTL (MB)	显示内存虚拟增长统计信息。所有数字都以兆字节为单位。	
	curr	使用 `vmmemctl` 模块回收的物理内存总量。
	target	ESX/ESXi 主机尝试使用 `vmmemctl` 模块回收的物理内存总量。
	max	ESX/ESXi 主机可以使用 `vmmemctl` 模块回收的最大物理内存量。
AMIN	该资源池或虚拟机的内存预留。	
AMAX	该资源池或虚拟机的内存限制。-1 值表示无限制。	
ASHRS	该资源池或虚拟机的内存份额。	
NHN	资源池或虚拟机的当前主节点。该统计信息仅适用于 NUMA 系统。如果虚拟机没有主节点，则显示短划线 (-)。	
NRMEM (MB)	分配到虚拟机或资源池的当前远程内存量。该统计信息仅适用于 NUMA 系统。	
N% L	分配到虚拟机或资源池的当前本地内存百分比。	
MEMSZ (MB)	分配到资源池或虚拟机的物理内存量。	
GRANT (MB)	映射到资源池或虚拟机的客户机物理内存量。消耗的主机内存等于 GRANT - SHRDSVD。	
SZTGT (MB)	ESX/ESXi VMkernel 想要分配到资源池或虚拟机的计算机内存量。	
TCHD (MB)	资源池或虚拟机的工作集估计。	
%ACTV	正由客户机引用的客户机物理内存的百分比。这是瞬时值。	
%ACTVS	正由客户机引用的客户机物理内存的百分比。这是慢速移动平均值。	
%ACTVF	正由客户机引用的客户机物理内存的百分比。这是快速移动平均值。	
%ACTVN	正由客户机引用的客户机物理内存的百分比。这是估计值。（您可能会看到该统计信息显示出来，但其仅供 VMware 使用。）	
MCTL?	是否已安装内存虚拟增长驱动程序。**N** 表示否，**Y** 表示是。	
MCTLSZ (MB)	通过虚拟增长从资源池回收的物理内存量。	
MCTLTGT (MB)	ESX/ESXi 系统尝试通过虚拟增长从资源池或虚拟机回收的物理内存量。	
MCTLMAX (MB)	ESX/ESXi 系统可以通过虚拟增长从资源池或虚拟机回收的最大物理内存量。该最大值取决于客户机操作系统类型。	
SWCUR (MB)	该资源池或虚拟机当前使用的交换量。	
SWTGT (MB)	ESX/ESXi 主机所希望的资源池或虚拟机的交换使用量目标。	
SWR/s (MB)	ESX/ESXi 主机为资源池或虚拟机从磁盘换入内存的速率。	

表 A–6。 内存面板统计信息（续）

字段	描述
SWW/s (MB)	ESX/ESXi 主机将资源池或虚拟机内存交换到磁盘的速率。
CPTRD (MB)	从检查点文件中读取的数据量。
CPTTGT (MB)	检查点文件大小。
ZERO (MB)	置零的资源池或虚拟机物理页。
SHRD (MB)	共享的资源池或虚拟机物理页。
SHRDSVD (MB)	由于资源池或虚拟机共享页而节省的计算机页。
OVHD (MB)	资源池的当前空间开销。
OVHDMAX (MB)	可能由资源池或虚拟机造成的最大空间开销。
OVHDUW (MB)	用户环境的当前空间开销。（您可能会看到该统计信息显示出来，但其仅供 VMware 使用。）
GST_NDx (MB)	为 NUMA 节点 x 上的资源池分配的客户机内存。该统计信息仅适用于 NUMA 系统。
OVD_NDx (MB)	为 NUMA 节点 x 上的资源池分配的 VMM 开销内存。该统计信息仅适用于 NUMA 系统。

表 A–7 显示了可以在内存面板中使用的交互命令。

表 A–7。 内存面板交互命令

命令	描述
M	按“映射的组”列对资源池或虚拟机排序。
B	按“组 Memctl”列对资源池或虚拟机排序。
N	按“GID”列对资源池或虚拟机排序。这是默认的排序顺序。
V	仅显示虚拟机实例。
L	更改“NAME”列的显示长度。

存储适配器面板

默认情况下，按照存储适配器来汇总存储适配器面板中的统计信息。还可以按存储通道、目标或 LUN 查看统计信息。

存储适配器面板显示了表 A–8 中所示的信息。

表 A–8。 存储适配器面板统计信息

列	描述
ADAPTR	存储适配器的名称。
CID	存储适配器通道 ID。只有展开对应的适配器时，该 ID 才可见。请参见下面的交互命令 `e`。
TID	存储适配器通道目标 ID。只有展开对应的适配器和通道时，该 ID 才可见。请参见下面的交互命令 `e` 和 `a`。
LID	存储适配器通道目标 LUN ID。只有展开对应的适配器、通道和目标时，该 ID 才可见。请参见下面的交互命令 `e`、`a` 和 `t`。
NCHNS	通道数量。
NTGTS	目标数量。
NLUNS	LUN 数量。
NWDS	环境数量。
BLKSZ	以字节为单位的块大小。该统计信息仅适用于 LUN。
AQLEN	存储适配器队列深度。适配器驱动程序被配置为能够支持的 ESX/ESXi VMkernel 活动命令的最大数目。

表 A–8。 存储适配器面板统计信息 （续）

列	描述
LQLEN	LUN 队列深度。允许 LUN 具有的 ESX/ESXi VMkernel 活动命令的最大数目。
%USD	ESX/ESXi VMkernel 活动命令使用的队列深度（适配器、LUN 或环境）百分比。
LOAD	ESX/ESXI VMkernel 活动命令加上 ESX/ESXi VMkernel 排队命令与队列深度（适配器、LUN 或环境）的比率。
ACTV	ESX/ESXi VMkernel 中当前处于活动状态的命令数目。
QUED	ESX/ESXi VMkernel 中当前排队的命令数目。
CMDS/s	每秒发出的命令数目。
READS/s	每秒发出的读取命令数目。
WRITES/s	每秒发出的写入命令数目。
MBREAD/s	每秒读取的兆字节数。
MBWRTN/s	每秒写入的兆字节数。
DAVG/cmd	每条命令的平均设备滞后时间，以毫秒为单位。
KAVG/cmd	每条命令的平均 ESX/ESXi VMkernel 滞后时间，以毫秒为单位。
GAVG/cmd	每条命令的平均虚拟机操作系统滞后时间，以毫秒为单位。
DAVG/rd	每个读取操作的平均设备读取滞后时间，以毫秒为单位。
KAVG/rd	每个读取操作的平均 ESX/ESXi VMkernel 读取滞后时间，以毫秒为单位。
GAVG/rd	每个读取操作的平均客户机操作系统读取滞后时间，以毫秒为单位。
DAVG/wr	每个写入操作的平均设备写入滞后时间，以毫秒为单位。
KAVG/wr	每个写入操作的平均 ESX/ESXi VMkernel 写入滞后时间，以毫秒为单位。
GAVG/wr	每个写入操作的平均客户机操作系统写入滞后时间，以毫秒为单位。
QAVG/cmd	每条命令的平均队列滞后时间，以毫秒为单位。
QAVG/rd	每个读取操作的平均队列滞后时间，以毫秒为单位。
QAVG/wr	每个写入操作的平均队列滞后时间，以毫秒为单位。
ABRTS/s	每秒中止的命令数目。
RESETS/s	每秒重置的命令数目。
PAECMD/s	每秒的 PAE（物理地址扩展）命令数目。
PAECP/s	每秒的 PAE 副本数。
SPLTCMD/s	每秒的拆分命令数目。
SPLTCP/s	每秒的拆分副本数。

表 A–9 显示了可以在存储适配器面板中使用的交互命令。

表 A–9。 存储适配器面板交互命令

命令	描述
e	在展开显示存储适配器统计信息和不展开显示存储适配器统计信息之间切换。允许查看按属于已展开存储适配器的各个通道细分的存储资源利用率统计信息。系统会提示您输入适配器名称。
P	在展开显示存储适配器统计信息和不展开显示存储适配器统计信息之间切换。允许查看按属于已展开存储适配器的路径细分的存储资源利用率统计信息。请勿汇总到适配器统计信息。系统会提示您输入适配器名称。
a	在展开显示存储通道统计信息和不展开显示存储通道统计信息之间切换。允许查看按属于已展开存储通道的各个目标细分的存储资源利用率统计信息。系统会提示您输入适配器名称和通道 ID。展开通道本身之前，需要先展开通道适配器。

表 A–9。 存储适配器面板交互命令（续）

命令	描述
t	在展开显示存储目标统计信息和不展开显示存储目标统计信息之间切换。允许查看按属于已展开存储目标的各个路径细分的存储资源利用率统计信息。系统会提示您输入适配器名称、通道 ID 和目标 ID。展开目标本身之前，必须先展开目标通道和适配器。
r	按“READS/s”列排序。
w	按“WRITES/s”列排序。
R	按“MBREAD/s read”列排序。
T	按“MBWRTN/s written”列排序。
N	首先按“ADAPTR”列排序，然后依次按每个“ADAPTR”内的“CID”列、每个“CID”内的“TID”列、每个“TID”内的“LID”列、每个“LID”内的“WID”列排序。这是默认的排序顺序。

存储设备面板

存储设备面板显示了服务器范围的存储利用率统计信息。

默认情况下，该信息按存储设备分组。还可以按照路径、环境或分区对统计信息分组。

表 A–10。 存储设备面板统计信息

列	描述
DEVICE	存储设备的名称。
PATH	路径名称。只有对应的设备展开到路径时，该名称才可见。请参见下面的交互命令 p。
WORLD	环境 ID。只有对应的设备展开到环境时，该 ID 才可见。请参见下面的交互命令 e。环境统计信息按环境和设备显示。
PARTITION	分区 ID。只有对应的设备展开到分区时，该 ID 才可见。请参见下面的交互命令 t。
NPH	路径数量。
NWD	环境数量。
NPN	分区数量。
SHARES	份额数量。该统计信息仅适用于环境。
BLKSZ	以字节为单位的块大小。
NUMBLKS	设备的块数。
DQLEN	存储设备队列深度。这是设备被配置为能够支持的 ESX/ESXi VMkernel 活动命令的最大数目。
WQLEN	环境队列深度。这是允许环境具有的 ESX/ESXi VMkernel 活动命令的最大数目。这是对于环境而言每个设备的最大值。只有对应的设备展开到环境时，此列才有效。
ACTV	ESX/ESXi VMkernel 中当前处于活动状态的命令数目。该统计信息仅适用于环境和设备。
QUED	ESX/ESXi VMkernel 中当前排队的命令数目。该统计信息仅适用于环境和设备。
%USD	由 ESX/ESXi VMkernel 活动命令使用的队列深度百分比。该统计信息仅适用于环境和设备。
LOAD	ESX/ESXi VMkernel 活动命令加上 ESX/ESXi VMkernel 排队命令与队列深度的比率。该统计信息仅适用于环境和设备。
CMDS/s	每秒发出的命令数目。
READS/s	每秒发出的读取命令数目。
WRITES/s	每秒发出的写入命令数目。
MBREAD/s	每秒读取的兆字节数。
MBWRTN/s	每秒写入的兆字节数。
DAVG/cmd	每条命令的平均设备滞后时间，以毫秒为单位。

表 A–10。 存储设备面板统计信息（续）

列	描述
KAVG/cmd	每条命令的平均 ESX/ESXi VMkernel 滞后时间，以毫秒为单位。
GAVG/cmd	每条命令的平均客户机操作系统滞后时间，以毫秒为单位。
QAVG/cmd	每条命令的平均队列滞后时间，以毫秒为单位。
DAVG/rd	每个读取操作的平均设备读取滞后时间，以毫秒为单位。
KAVG/rd	每个读取操作的平均 ESX/ESXi VMkernel 读取滞后时间，以毫秒为单位。
GAVG/rd	每个读取操作的平均客户机操作系统读取滞后时间，以毫秒为单位。
QAVG/rd	每个读取操作的平均队列读取滞后时间，以毫秒为单位。
DAVG/wr	每个写入操作的平均设备写入滞后时间，以毫秒为单位。
KAVG/wr	每个写入操作的平均 ESX/ESXi VMkernel 写入滞后时间，以毫秒为单位。
GAVG/wr	每个写入操作的平均客户机操作系统写入滞后时间，以毫秒为单位。
QAVG/wr	每个写入操作的平均队列写入滞后时间，以毫秒为单位。
ABRTS/s	每秒中止的命令数目。
RESETS/s	每秒重置的命令数目。
PAECMD/s	每秒的 PAE 命令数目。该统计信息仅适用于路径。
PAECP/s	每秒的 PAE 副本数。该统计信息仅适用于路径。
SPLTCMD/s	每秒的拆分命令数目。该统计信息仅适用于路径。
SPLTCP/s	每秒的拆分副本数。该统计信息仅适用于路径。

表 A–11 显示了可以在存储设备面板中使用的交互命令。

表 A–11。 存储设备面板交互命令

命令	描述
e	展开或汇总存储环境统计信息。该命令允许查看由属于已展开存储设备的各个环境分隔的存储资源利用率统计信息。系统会提示您输入设备名称。统计信息按环境和设备显示。
p	展开或汇总存储路径统计信息。该命令允许查看由属于已展开存储设备的各个路径分隔的存储资源利用率统计信息。系统会提示您输入设备名称。
t	展开或汇总存储器分区统计信息。该命令允许查看按属于已展开存储设备的各个分区分隔的存储资源利用率统计信息。系统会提示您输入设备名称。
r	按“READS/s”列排序。
w	按“WRITES/s”列排序。
R	按“MBREAD/s”列排序。
T	按“MBWRTN”列排序。
N	先按“DEVICE”列排序，再依次按“PATH”、“WORLD”和“PARTITION”列排序。这是默认的排序顺序。
L	更改“DEVICE”列的显示长度。

虚拟机存储面板

该面板显示了以虚拟机为中心的存储统计信息。

默认情况下，按照资源池聚合统计信息。一个虚拟机具有一个对应的资源池，因此该面板实际上按照虚拟机显示统计信息。还可以按照环境或按照环境和设备查看统计信息。

表 A–12。 虚拟机存储面板统计信息

列	描述
ID	运行中环境内资源池的资源池 ID 或运行中环境的环境 ID。
GID	运行中环境内资源池的资源池 ID。
NAME	运行中环境内资源池的名称或运行中环境的名称。
Device	存储设备名称。只有对应的环境展开到设备时，该名称才可见。请参见下面的交互命令 `i`。
NWD	环境数量。
NDV	设备数量。只有对应的资源池展开到环境时，此数量才有效
SHARES	份额数量。该统计信息仅适用于环境。只有对应的资源池展开到环境时，此列才有效。
BLKSZ	以字节为单位的块大小。只有对应的环境展开到设备时，此列才有效。
NUMBLKS	设备的块数。只有对应的环境展开到设备时，此列才有效。
DQLEN	存储设备队列深度。这是设备被配置为能够支持的 ESX/ESXi VMkernel 活动命令的最大数目。只有对应的环境展开到设备时，显示的数字才有效。
WQLEN	环境队列深度。该列显示了允许环境具有的 ESX/ESXi VMkernel 活动命令的最大数目。只有对应的环境展开到设备时，该数字才有效。这是对于环境而言每个设备的最大值。
ACTV	ESX/ESXi VMkernel 中当前处于活动状态的命令数目。该数字仅适用于环境和设备。
QUED	ESX/ESXi VMkernel 中当前排队的命令数目。该数字仅适用于环境和设备。
%USD	由 ESX/ESXi VMkernel 活动命令使用的队列深度百分比。该数字仅适用于环境和设备。
LOAD	ESX/ESXi VMkernel 活动命令加上 ESX/ESXi VMkernel 排队命令与队列深度的比率。该数字仅适用于环境和设备。
CMDS/s	每秒发出的命令数目。
READS/s	每秒发出的读取命令数目。
WRITES/s	每秒发出的写入命令数目。
MBREAD/s	每秒读取的兆字节数。
MBWRTN/s	每秒写入的兆字节数。
DAVG/cmd	每条命令的平均设备滞后时间，以毫秒为单位。
KAVG/cmd	每条命令的平均 ESX/ESXi VMkernel 滞后时间，以毫秒为单位。
GAVG/cmd	每条命令的平均客户机操作系统滞后时间，以毫秒为单位。
QAVG/cmd	每条命令的平均队列滞后时间，以毫秒为单位。
DAVG/rd	每个读取操作的平均设备读取滞后时间，以毫秒为单位。
KAVG/rd	每个读取操作的平均 ESX/ESXi VMkernel 读取滞后时间，以毫秒为单位。
GAVG/rd	每个读取操作的平均客户机操作系统读取滞后时间，以毫秒为单位。
QAVG/rd	每个读取操作的平均队列读取滞后时间，以毫秒为单位。
DAVG/wr	每个写入操作的平均设备写入滞后时间，以毫秒为单位。
KAVG/wr	每个写入操作的平均 ESX/ESXi VMkernel 写入滞后时间，以毫秒为单位。
GAVG/wr	每个写入操作的平均客户机操作系统写入滞后时间，以毫秒为单位。
QAVG/wr	每个写入操作的平均队列写入滞后时间，以毫秒为单位。
ABRTS/s	每秒中止的命令数目，以毫秒为单位。
RESETS/s	每秒重置的命令数目，以毫秒为单位。

表 A–13 显示了可以在虚拟机存储面板中使用的交互命令。

表 A–13。 虚拟机存储面板交互命令

命令	描述
e	展开或汇总存储环境统计信息。允许查看由属于某个组的各个环境分隔的存储资源利用率统计信息。系统会提示您输入组 ID。该统计信息按环境显示。
l	展开或汇总存储设备（即 LUN）统计信息。允许查看按属于已展开环境的各个设备分隔的存储资源利用率统计信息。系统会提示您输入环境 ID。
V	仅显示虚拟机实例。
r	按“READS/s”列排序。
w	按“WRITES/s”列排序。
R	按“MBREAD/s”列排序。
T	按“MBWRTN/s”列排序。
N	先按“虚拟机”列排序，然后按“环境”列排序。这是默认的排序顺序。
L	更改“NAME”列的显示长度。

网络面板

网络面板显示了服务器范围的网络利用率统计信息。

统计信息按照所配置的每个虚拟网络设备的端口进行排列。有关物理网络适配器统计信息，请参见对应于物理网络适配器所连接到端口的行。有关在特定虚拟机上配置的虚拟网络适配器的统计信息，请参见对应于虚拟网络适配器所连接到端口的行。

表 A–14。 网络面板统计信息

列	描述
PORT-ID	虚拟网络设备的端口 ID。
UPLINK	“Y”表示对应的端口是上行链路。“N”表示不是。
UP	“Y”表示对应的链路是上行链路。“N”表示不是。
SPEED	以兆位/秒为单位的链路速度。
FDUPLX	“Y”表示对应的链路以全双工方式运行。“N”表示不是。
USED-BY	虚拟网络设备端口用户。
DTYP	虚拟网络设备类型。“H”表示集线器，“S”表示交换机。
DNAME	虚拟网络设备名称。
PKTTX/s	每秒传输的数据包数。
PKTRX/s	每秒接收的数据包数。
MbTX/s	每秒传输的兆位数。
MbRX/s	每秒接收的兆位数。
%DRPTX	丢弃的传输数据包百分比。
%DRPRX	丢弃的接收数据包百分比。
TEAM-PNIC	用于绑定上行链路的物理网卡的名称。

表 A–15 显示了可以在网络面板中使用的交互命令。

表 A–15。 网络面板交互命令

命令	描述
T	按“Mb Tx”列排序。
R	按“Mb Rx”列排序。
t	按“Packets Tx”列排序。
r	按“Packets Rx”列排序。
N	按“PORT-ID”列排序。这是默认的排序顺序。
L	更改“DNAME”列的显示长度。

中断面板

中断面板显示有关中断向量的使用信息。

表 A–16。 中断面板统计信息

列	描述
VECTOR	中断向量 ID。
COUNT/s	每秒中断总数。此值是每个 CPU 的累积计数。
COUNT_x	在 CPU x 上的每秒中断数。
TIME/int	每个中断的平均处理时间（以微秒为单位）。
TIME_x	在 CPU x 上每个中断的平均处理时间（以微秒为单位）。
DEVICES	使用中断向量的设备。如果没有为设备启用中断向量，则其名称将包含在尖括号（< 和 >）中。

使用批处理模式

批处理模式允许您收集资源利用率统计信息并将其保存到文件中。

在准备好批处理模式之后，可以在此模式中使用 `esxtop` 或 `resxtop`。

准备批处理模式

要以批处理模式运行，必须先准备批处理模式。

步骤

1 以交互模式运行 `resxtop`（或 `esxtop`）。

2 在每个面板中，选择所需列。

3 使用 `W` 交互命令将该配置保存到文件（默认为 `~/.esxtop4rc`）中。

现在可以在批处理模式中使用 `resxtop`（或 `esxtop`）。

在批处理模式中使用 esxtop 或 resxtop

在准备好批处理模式后，可以在此模式中使用 esxtop 或 resxtop。

步骤

1 启动 resxtop（或 esxtop）将输出重定向到文件。

例如：

```
esxtop -b > my_file.csv
```

文件名必须具有 .csv 扩展名。该实用程序不强制要求这点，但后处理工具需要该扩展名。

2 使用诸如 Microsoft Excel 和 Perfmon 之类的工具处理在批处理模式中收集的统计信息。

在批处理模式中，resxtop（或 esxtop）不接受交互命令。在批处理模式中，该实用程序运行到产生所请求的迭代次数为止（有关详细信息，请参见下面的命令行选项 n），或运行到通过按 Ctrl+c 终止进程为止。

批处理模式命令行选项

可以将批处理模式与命令行选项配合使用。

在批处理模式中可以使用表 A–17 中的命令行选项。

表 A–17。 批处理模式中的命令行选项

选项	描述
a	显示所有统计信息。该选项会替代配置文件设置并显示所有统计信息。配置文件可以是默认的 ~/.esxtop4rc 配置文件或用户定义的配置文件。
b	以批处理模式运行 resxtop（或 esxtop）。
c <filename>	加载用户定义的配置文件。如果未使用 -c 选项，则默认配置文件名为 ~/.esxtop4rc。使用 W 单键交互命令创建自己的配置文件，同时指定其他文件名。
d	指定统计信息快照之间的延迟。默认值为 5 秒。最小值为 2 秒。如果指定的延迟少于 2 秒，延迟将设置为 2 秒。
n	迭代次数。resxtop（或 esxtop）对统计信息迭代执行此次数的收集和保存操作，然后退出。
server	要连接的远程服务器主机的名称（仅 resxtop 需要）。
portnumber	要连接到的远程服务器上的端口号。默认端口为 443，除非在服务器上更改了这一端口，否则不需要此选项。(仅限 resxtop)
username	连接到远程主机时要进行身份验证的用户名。远程服务器还会提示您输入密码（仅限 resxtop）。

使用重放模式

在重放模式中，esxtop 重放借助于 vm-support 收集的资源利用率统计信息。

在准备好重放模式之后，可以在此模式中使用 esxtop。请参见 vm-support 手册页。

在重放模式中，esxtop 接受与交互模式相同的交互命令集，并运行到不再有 vm-support 收集的快照要读取为止，或者运行到请求的迭代次数已完成为止。

准备重放模式

要以重放模式运行，必须先准备重放模式。

步骤

1 在 ESX 服务控制台上以快照模式运行 `vm-support`。

请使用以下命令。

```
vm-support -S -d duration -I interval
```

2 解压缩所生成的 tar 文件，以便 `esxtop` 可以在重放模式中使用该文件。

现在可以在重放模式中使用 `esxtop`。

在重放模式中使用 esxtop

可以在重放模式中使用 `esxtop`。

您不必在 ESX 服务控制台上运行重放模式。可以运行重放模式，按照与批处理模式相同的样式产生输出（请参见下面的命令行选项 `b`）。

步骤

◆ 要激活重放模式，请在命令行提示符处输入以下内容。

```
esxtop -R <vm-support_dir_path>
```

重放模式命令行选项

可以将重放模式与命令行选项配合使用。

表 A–18 列出了可用于 `esxtop` 重放模式的命令行选项。

表 A–18。 重放模式中的命令行选项

选项	描述
R	vm-support 收集的快照目录的路径。
a	显示所有统计信息。该选项会替代配置文件设置并显示所有统计信息。配置文件可以是默认的 `~/.esxtop4rc` 配置文件或用户定义的配置文件。
b	以批处理模式运行 `esxtop`。
c<filename>	加载用户定义的配置文件。如果未使用 `-c` 选项，则默认配置文件名为 ~/.esxtop4rc。使用 `W` 单键交互命令创建自己的配置文件，同时指定其他文件名。
d	指定面板更新之间的延迟。默认值为 5 秒。最小值为 2 秒。如果指定的延迟少于 2 秒，延迟将设置为 2 秒。
n	迭代次数。`esxtop` 对显示执行此次数的更新，然后退出。

高级属性

B

可以为主机或单个虚拟机设置高级属性以帮助自定义资源管理。

大多数情况下，调整基本资源分配设置（预留、限制和份额）或接受默认设置可以获得适当的资源分配结果。但是，可以使用高级属性为主机或特定虚拟机自定义资源管理。

本附录讨论了以下主题：

设置高级主机属性

可以为主机设置高级属性。

小心 VMware 仅建议高级用户设置高级主机属性。大多数情况下，使用默认设置即可获得最佳结果。

步骤

1 在 vSphere Client“清单”面板中，选择要自定义的主机。

2 单击**配置**选项卡。

3 在**软件**菜单中，单击**高级设置**。

4 在“高级设置”对话框中，选择适当的项目（例如**CPU** 或**内存**），并在右侧面板中滚动以查找和更改属性。

高级 CPU 属性

可以使用高级 CPU 属性自定义 CPU 资源使用情况。

表 B–1。 高级 CPU 属性

属性	描述
CPU.MachineClearThreshold	如果使用启用了超线程的主机，请将此值设置为 0 以禁用隔离。
Power.CpuPolicy	如果将此属性设置为默认值（静态），则 VMkernel 不会直接设置 CPU 电源管理状况，而且只响应来自 BIOS 的请求。如果启用此策略（设置为动态），则 VMkernel 将根据当前的使用情况动态选择相应的电源管理状况。这可以在省电的同时不降低性能。在不支持电源管理的系统上启用此选项将导致产生错误消息。

高级内存属性

可以使用高级内存属性自定义内存资源使用情况。

表 B–2。 高级内存属性

属性	描述	默认值
Mem.CtlMaxPercent	根据所配置内存大小的百分比，使用 `vmmemctl` 限制从任何虚拟机回收的最大内存量。指定 0 将禁止所有虚拟机使用 `vmmemctl` 进行回收。	65
Mem.ShareScanTime	指定要扫描整个虚拟机以寻找页面共享机会所用的时间，以分钟为单位。默认值为 60 分钟。	60
Mem.ShareScanGHz	指定每 1 GHz 可用主机 CPU 资源为寻找页面共享机会，每秒内可用于扫描的最大内存页面量。 默认值为每 1GHz 的速率为 4 MB/秒。	4
Mem.IdleTax	指定闲置内存消耗率，以百分比为单位。虚拟机对闲置内存的消耗量大于对正在使用的内存的消耗量。0 % 的消耗率定义的分配策略将忽略工作集并严格按照份额分配内存。较高的消耗率产生的分配策略允许要重新分配的闲置内存远离以非生产性方式累积闲置内存的虚拟机。	75
Mem.SamplePeriod	指定虚拟机执行时间的周期性时间间隔（以秒为度量单位），在该执行时间内监控内存活动来估计工作集大小。	60
Mem.BalancePeriod	指定自动内存重新分配的周期性时间间隔，以秒为单位。可用内存量的重大更改也会触发重新分配。	15
Mem.AllocGuestLargePage	将此选项设置为 1，将让主机大页作为客户机的备用大页。在使用客户机大页的服务器工作负载中减少 TLB 缺失并改善性能。0 = 禁用。	1
Mem.AllocUsePSharePool 和 Mem.AllocUseGuestPool	将这些选项设置为 1 可减少内存碎片。如果主机内存有碎片，则主机大页的可用性会降低。这些选项可以提高让主机大页作为客户机备用大页的可能性。0 = 禁用。	1
LPage.LPageDefragEnable	将此选项设置为 1 可启用大页碎片整理。0 = 禁用。	1
LPage.LPageDefragRateVM	每个虚拟机上每秒内最多可尝试的大页碎片整理次数。可接受的值在 1 到 1024 之间。	2
LPage.LPageDefragRateTotal	每秒内最多可尝试的大页碎片整理次数。可接受的值在 1 到 10240 之间。	8
LPage.LPageAlwaysTryForNPT	将此选项设置为 1 将允许始终尝试为嵌套页表 (NPT) 分配大页。0 = 禁用。 如果启用此选项，则所有客户机内存都受到使用嵌套页表的计算机（例如，AMD Barcelona）中的大页支持。如果 NPT 不可用，则只有部分客户机内存受到大页支持。	1

高级 NUMA 属性

可以使用高级 NUMA 属性自定义 NUMA 使用情况。

表 B–3。 高级 NUMA 属性

属性	描述	默认值
Numa.RebalanceEnable	将此选项设置为 0 可针对虚拟机禁用所有的 NUMA 再平衡和初始放置位置，从而有效地禁用 NUMA 调度系统。	1
Numa.PageMigEnable	如果将此选项设置为 0，则系统不会在节点间自动迁移页面以改善内存局部性。手动设置的页面迁移率仍然有效。	1
Numa.AutoMemAffinity	如果将此选项设置为 0，则系统不会自动使用 CPU 关联性集合来设置虚拟机的内存关联性。	1

表 B–3。 高级 NUMA 属性（续）

属性	描述	默认值
Numa.MigImbalanceThreshold	NUMA 再平衡器计算节点之间 CPU 的不平衡，考虑每个虚拟机的 CPU 时间可用量与其实际消耗量之间的差值。该选项控制节点之间触发虚拟机迁移所需的最小负载不平衡，以百分比为单位。	10
Numa.RebalancePeriod	控制再平衡周期的频率，以毫秒为单位指定。再平衡的频率越大，CPU 开销也越大，是在运行大量虚拟机的计算机上尤其如此。频繁的再平衡还可以提高公平性。	2000
Numa.RebalanceCoresTotal	指定主机上启用 NUMA 再平衡器所需的处理器内核的最小总数。	4
Numa.RebalanceCoresNode	指定每个节点上启用 NUMA 再平衡器所需的处理器内核的最小数量。 在小型 NUMA 配置（例如，2 路 Opteron 主机）中禁用 NUMA 再平衡时，此选项和 Numa.RebalanceCoresTotal 会非常有用，在这样的配置中，如果启用了 NUMA 再平衡功能，而且处理器总数或每个节点上的处理器较少，则会影响调度的公平性。	2
VMkernel.Boot.sharePerNode	控制内存页是只能在单个 NUMA 节点内共享（重复数据删除），还是可以跨多个 NUMA 节点共享。 与其他 NUMA 选项不同，该选项显示在“高级设置”对话框中的“VMkernel”下。这是因为，与此处显示的其他 NUMA 选项（可以在系统运行时进行更改）不同的是，VMkernel.Boot.sharePerNode 是一个引导时选项，只有在重新引导之后才会生效。	有效（已选）

设置高级虚拟机属性

可以为虚拟机设置高级属性。

步骤

1 在 vSphere Client“清单”面板中选择虚拟机，然后从右键单击菜单中选择**编辑设置**。

2 单击**选项**，然后单击**高级 > 常规**。

3 单击**配置参数**按钮。

4 在显示的对话框中，单击**添加行**以输入新参数及其值。

高级虚拟机属性

可以使用高级虚拟机属性自定义虚拟机配置。

表 B–4。 高级虚拟机属性

属性	描述
sched.mem.maxmemctl	通过虚拟增长而从选定虚拟机中回收的最大内存量，以兆字节 (MB) 为单位。如果 ESX/ESXi 主机需要回收更多内存，则会强制它进行交换。交换的优先级低于虚拟增长。
sched.mem.pshare.enable	为选定的虚拟机启用内存共享。 此布尔值默认为“有效”。如果将虚拟机的该属性设置为“无效”，则将关闭内存共享。
sched.swap.persist	指定关闭虚拟机时应保留还是删除虚拟机的交换文件。默认情况下，当虚拟机启动时系统为虚拟机创建交换文件，当虚拟机关闭时删除该交换文件。
sched.swap.dir	虚拟机交换文件的 VMFS 目录位置。默认为虚拟机的工作目录，即包含其配置文件的 VMFS 目录。此目录必须保留在虚拟机可访问的主机上。如果移动虚拟机（或从虚拟机创建的任何克隆），则可能需要重置此属性。

索引