# 第四章

# 让虚拟基础架构变得高可用

*让基础架构上运行的虚拟机及其应用变得高可用，这能保护生产工作负载。*

- ☆何种高可用性（HA）策略最佳？
- ☆选择高可用性方法
- ☆案例学习
- ☆为高可用性虚拟架构做好准备
- ☆客户机NLB集群

# 实现虚拟基础架构高可用性

***让基础架构上运行的虚拟机及其应用变得高可用，这能保护生产工作负载。***

服务器虚拟化技术给企业提供了一个使数据中心变得更加“绿色”的机会，通过逐步减少物理服务器来达到此目的，通常整合的比率为 20:1 或者更多。服务器

虚拟化之后，企业可以将虚拟化扩大到数据中心内的其他方面，包括桌面电脑和应用程序。然而，一旦你实施了服务器虚拟化技术，你得转变依赖一个单一的解决方案，因为每个物理服务器将运行多个企业应用。

这样，企业将不能再失去这台服务器，它必需一直运行。但如果情况不是这样呢，那就必须能够将此台服务器的工作量转移到一台独立的主机上，使生产在任何时候都可用。

因此，研究并尝试不同的企业应用保护方案是重要的事情，以确保应用总是可用。有三种不同的方式可以使虚拟机上的应用实现高可用性：

1、 创建主机集群。

2、 创建客户机故障转移群集。

3、 创建客户机或网络负载平衡（NLB）群集。

这些都可以给虚拟机以及上面跑的应用提供一个特定级别的高可用性解决方案。通过在你的环境里使用高可用性（HA）群集配置，可以解决一些潜在的问题。

**你可以让系统远离主机故障**。当主机发生故障时，两件事情可能发生，一是正在运行着的虚拟机（VM）将被迁移至其他主机上面，二是此虚拟机内的应用将被

迁移至其他虚拟机。至于这两种情况中的哪个会发生，将取决于你采用三个高可用（HA）策略中的哪一个来保护主机。

**你必须对主机服务器进行维护**。生产应用仍将继续运行，这是因为其他虚拟机（VM）将支持此应用，或将此虚拟机（VM）转移到主机群集上的另外一个节点。

在数据中心里，你**可以动态地将应用负载从一台物理服务器迁移至另外一台**。这可以确保应用程序运行在最佳性能状态下，即使是在高峰负载期间。

在实施虚拟化后，这些高可用模式都是可以使用的。但是，究竟要选择哪一种模式，这将取决于每个应用程序它本身的重要程度。在大多数的企业中，几乎所有的用户都需要访问财务系统或电子邮件系统，但在一个很短的时间内，很少有用户会因为服务器网络维护而丢失什么。考虑每一个应用程序的类型，以确定最佳的高可用性（HA）模式。

## 哪一种高可用性（HA）策略最适合你的虚拟机（VM）环境？

对于虚拟机（VM）容错，每一种高可用性（HA）策略都提供了行之有效方法。但是，要觉得使用哪种方法并不总是很明显。表—第 4 页，“选择一个虚拟机的高可用性策略，”介绍了在给生产

***考虑每个应用的类型以确定最佳高可用性 HA 方法。***

环境中的虚拟机（VM）选择和实施一种高可用方案时的参考要点。

这些准则可以帮助你，但是，你至少应该创建主机故障转移群集。

每个主机上运行着多个生产的虚拟机。如果该主机出现故障，同时又没有实施高可用性（HA）方案，那么此主机上的每个虚拟机（VM）都将发生故障。根据应用的不同重要程度来决定高可用性（HA）的方案。当你在一台单独的物理机上运行着一个应用负载。在这种情况下，就没有理由为什么不能在运行一个基于主机级别集群的时候，同时再实施一个基于客户系统的高可用性（HA）解决方案，例如故障转移群集或者网络负载平衡（NLB）群集。

使用可以满足企业现有服务需要的方案，这可以确定你需要为每台虚拟机（VM）配置哪种高可用性（HA）方案。同时，你还必须考虑运行在虚拟机（VM）中的应用程序的服务政策。

## 配置单点或者多点集群

物理服务器支持单点和多点集群技术。单点集群技术是基于各种形式的共同存储。 拿 VMware 举例，主机集群采用两项关键技术：高可用性（HA）和虚拟机文件系统（VMFS），它是一个共享式的文件系统，可让多个主机服务器连接到同一存储容器。VMFS 通常需要某种形式的 SAN，网络附加存储（NAS）或 iSCSI

存储容器。 VMware 还可以支持网络文件系统（NFS），这可以使小企业对主机服务器实施高可用性（HA）方案。VMware 的 HA 组件可以管理潜在的主机服务器故障。VMware 主机集群可以支持最多 32 个节点。

思杰 Xen 服务也可以依靠共享存储来为主机服务器提供高可用性（HA），存储通常为 NFS、NAS、SAN 或者 iSCSI 容器的形式。每台思杰主机服务器环境，管理员可以通过创建主机资源池的方式来配置高可用性。虽然其他虚拟化技术依靠管理数据库来控制多台主机配置信息，但是每一台思杰 Xen 的服务器主机存储着它自己资源池配置数据信息的副本。这消除了资源池配置的一个潜在的单点故障。思杰资源池也可以包含多达 32 个主机节点。

选择一个虚拟机的高可用性方案

| 虚拟机规格参数表 | 主机服务器集群 | 客户机故障转移群集 | 客户机网络负载平衡（NLB）群集 |
|---|---|---|---|
| 操作系统版本（例如，Windows服务器版本） | Web版，标准版企业版，或者数据中心版 | 企业版或者数据中心版 | Web版，标准版，企业版，或者数据中心版 |
| 客户机节点个数 | 单个节点 | 通常两个节点，最多支持16个节点 | 最多支持32个节点 |
| 虚拟机（VM）需要的资源 | 至少一块虚拟网络适配器 | iSCSI磁盘连接器和最少三块虚拟网络适配器：集群公共网，私有网和iSCSI | 最少需要两个虚拟网络适配器：集群公共网和私有网 |
| 服务器角色 | 多角色 | 状态应用服务器，文件和打印服务器，协作服务器的存储组件，网络基础设施组件如动态主机配置协议服务器 | 无状态的应用服务器，Web专用服务器，前端协作服务器，前端终端服务器 |
| 虚拟机内的应用 | 多应用 | SQL或数据库服务器，Exchange服务器，消息队列服务器，文件服务器，打印服务器 | Web园，Exchange客户端访问服务器，网络安全和加速服务器（ISA），VPN服务器，流媒体服务器，统一通信服务器 |

微软的 Hyper-V 依赖于 Windows Server 2008 的故障转移群集技术来创建主机集群。单个 Hyper - V 主机集群需要共享存储，存储只支持 SAN 或 iSCSI 容器，其他存储形式不被支持。 Hyper - V 的单点集群可以支持最多 16 个主机节点。

Hyper-V 的也可以支持多点集群，它跨越多个点来支持可能影响整个集群的故障。正因为如此，在 Hyper-V 的多点集群中不再需要共享存储，而是依靠速度更快的直连方式存储（DAS）。

然而，为了提供高虚拟机（VM）的高可用性， DAS 存储库必须使用第三方同步工具来实时同步。

无论你使用哪种虚拟化技术，最好的方式是让创建的主机集群尽可能为服务的连续性提供两个不同程度的保护：

**主机集群为虚拟机（VM）提供持续的运作**。如果一个主机发生故障，或者表明它已经不可用，那么此台主机上所有运行的虚拟机（VM）将自动转移到群集的另一个节点上。

**在主机例行维护时间，主机集群为虚拟机（VM）提供持续的运作**。取个例子，如果你需要在群集中的一个节点上安装系统更新，那么你可以在操作过程中将虚拟机迁移至其他节点，等操作完毕后再将虚拟机迁移回来。

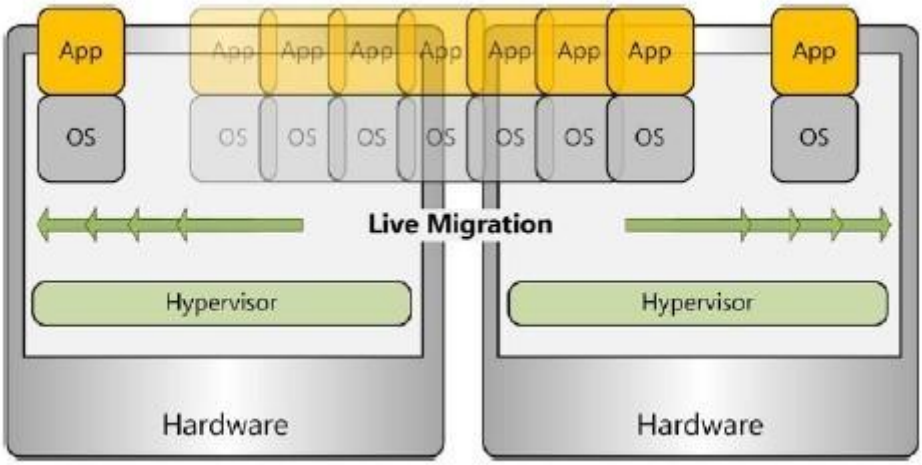
## 热迁移与高可用性对比

有许多关于热迁移的描述，它能够将正在运行的虚拟机（VM）从一台主机迁移到另一台上面。热迁通常用于两种情况：

1、 在执行例行管理时采用热迁移，这样管理员就可对主机服务器进行维护。

2、 动态热迁移，如果一个资源监视器检测到一个虚拟机（VM）没有可用资源，并且当前虚拟机所在的主机也没有足够的备用资源，来应对这一资源高峰需求。在这种情况下，资源监视器扫描所以的主机，发现有足够备用资源的所有主机。然后将此虚拟机迁移到这个新的主机上。

在所有情况下，虚拟机（VMs）的迁移是将此虚拟机的内存信息从一台主机迁移到另一台主机上。主机服务器必须使用同型号的处理器。否则，迁移之后的内存信息将不能在目标主机上工作。




虽然热迁移对动态数据中心来说是相当有用的，但是当主机发生故障时它就会变的完全没有使用价值。这是因为根本没有可能从发生故障的主机内存中读取任何信息。在主机故障转移期间，其实虚拟机（VMs）都处于停止状态，因为主机已经发生故障。当主机发生故障后，它上面的所有虚拟机（VMs）会被在另一台有备用资源的主机上重新启动。因此，高可用性（HA）不等同于实时迁移。

如果群集中的其他节点也需要维护，最好的方式是让创建的主机集群尽可能为服务的连续性提供两个不同程度的保护。

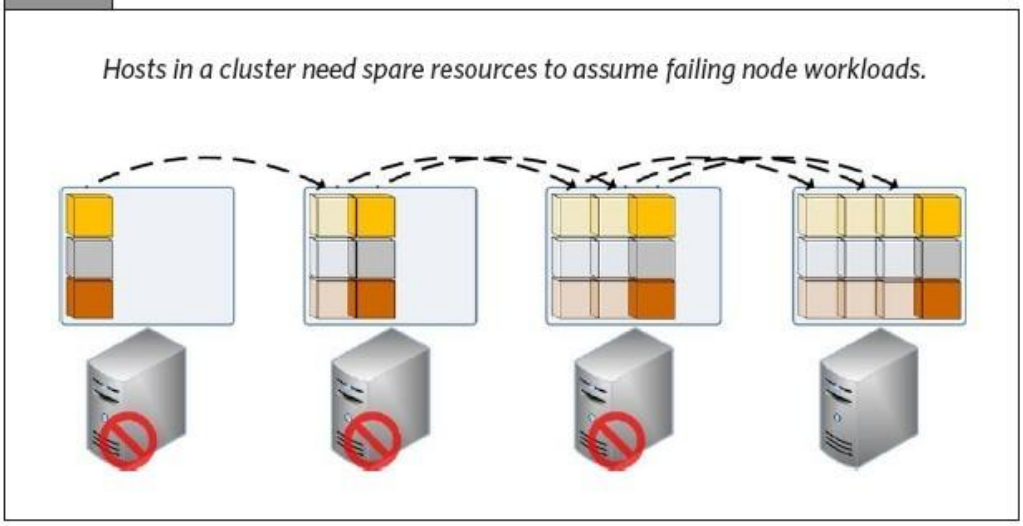
在这两种情况下，在虚拟机迁移过程中服务将被中断。当检测到群集中的一个节点出现故障，群集服务会促使虚拟机从故障节点转移到另一个节点。在这种情况下，它会使用一个迁移进程，将虚拟机从一个节点移动到另一个。依据你使用的虚拟化技术的不同，这可能会引起服务的中断。以热迁移为例，VMware 和 Citrix 的产品，支持迁移正在运行着的虚拟机（VM）。Hyper - V 的第一个版本不能执行热迁移。然而，今年下半年的晚些时候发布的 Hyper - VR2 版，将支持此功能。

当一个节点出现故障，群集服务会通过在另外一个节点重启此虚拟机（VM）来达到迁移的目的。在这种情况下，虚拟机（VM）的停机时间在增加，因为故障节点上的所有虚拟机都会被关闭。当你需要对一个节点执行维护时，使用迁移工具将虚拟机（VM）从一个主机节点迁移到另一个节点。请记住，你必须在群集中的每台主机服务器上留有备用资源，以便支持迁移操作（图 1）。理想情况下，每台主机服务器将拥有足够的备用资源，以支持群集中至少一个其他节点故障。

## 配置客户机故障转移群集

当虚拟机（VM）作为一个应用加入到主机集群中后，虚拟机的高可用就已被配置。

FIGURE 1




当虚拟机（VM）作为一个应用加入到主机集群中后，虚拟机的高可用性就已被配置。但是，虚拟机不像传统的应用程序。即便虚拟机将会始终在运行——或者尽可能在运行——在集群中的一台主机上，这个模型并不适用于你的生产网中的每一个应用负载。这是因为集群主机服务器不会对虚拟机中的应用起作用。

## 子机故障切换集群工作环境

所有的虚拟机都可以作为主机集群系统内的一个应用来对待，从而实现高可用。然而，虚拟机和传统的应用程序毕竟存在差别。虽然对于虚拟机来讲，我们也需要保持它始终处于运行状态（或者是在尽可能多的时间段内保持运行），但主机集群的这种操作方式并不是通用于生产环境中的所有工作负载的。原因在于对主机所做的集群不会直接作用于运行于在虚拟机中的应用程序。因此这些应用程序并不能感受到主机所具有的高可用性，而在这一点上，子机故障切换集群中就完全不同，这

## 虚拟化环境必须考虑高可用性

对于那些通过部署高可用软件或集群系统来保护他们所拥有的虚拟机的用户而言，已经非常了解其优点，然而，我们不得不注意这些方案中所带有的一些潜在风险。

如果您在一台物理主机上运行了 20 台虚拟机，不希望在发生主机硬件故障后，所有的虚拟机都无法工作。正如 IDC Enterprise Platform Group 的副总裁 John Humphreys 先生所认为的那样，这种环境中，部署一定形式的高可用系统是非常有必要的。

“在虚拟化架构中，我们把 5 台、6 台、甚至 10 台运行关键业务的服务器放到了同一台物理主机上，”Humphreys 说，“如果我们这样做了，接下来就需要一种可以保障高可用性的服务。用户可能会花费多一点以保障他们的虚拟机可以始终处于运行中。”

起码来说，高可用软件可以把虚拟机从一台发生了部件故障的“残疾”虚拟机上。自动切换到指定的故障切换目标服务器上，以避免长的宕机时间。通过多种不同的高可用软件或传统的集群模式，可以实现不同级别的高可用性。

虽然高可用软件安装和运行都非常地便捷，但是它也有一些缺点，一家中立机构——San Diego Data Processing Corp.公司的系统管理员 Rick J. Scherer 这么认为。这家 IT 企业在他们位于两个不同地点的生产和灾难恢复数据中心里都采用了 VMware 的虚拟化软件。Scherer 为了实现高可用，同时部署了 VMware 的高可用软件和 Microsoft Cluster 软件。

他说，在生产环境里有 29 台拥有足够 CPU、内存和 I/O 资源的主机，可以支持多台虚拟机。在四月份时，Scherer 在 29 台服务器上运行了总计 400 台虚拟机。而他计划在今年末，再部署额外的 100 台虚拟机。

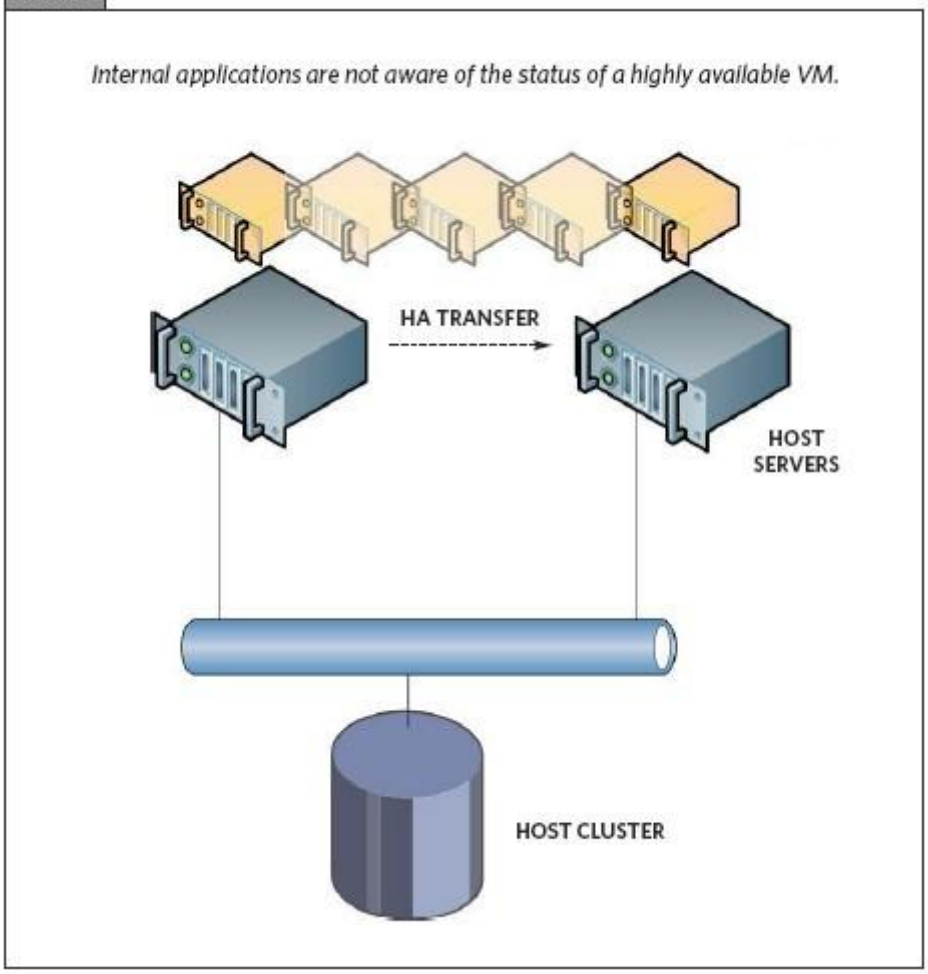
作为最低级别的高可用，Scherer 在 ESX 集群内运行了 VMware HA 软件。

然而，在这种高可用解决方案中，需要管理员去监控主机服务器的资源，以保证故障切换后有足够的资源可以支持所有虚拟机的运行。“我们试图保证在系统中留有 10%到 20%的余量，这样才可以知道拥有了足够的可用资源。”Scherer 这样说。

无论是采用高可用软件产品还是集群的办法，最重要的是一定要做到可以预防虚拟环境中所存在单点故障问题。——Bridge Botelho

种模式中应用程序被直接作为集群中的一部分而安装。在主机集群模式中，并不能保证当主机故障后，虚拟机仍然可以保持运行（如图 2）。尽管事实上多数应用程序不能感知，故障发生后从一个主机节点到另一个主机节点的切换过程，但是这种主机高可用模式还是可以满足大多数应用程序的需求。

FIGURE 2




很多状态敏感型（state-sensitive）应用程序，如 Microsoft Exchange，并不能很好地工作于主机集群模式下。在发生切换时，甚至可能会导致数据的丢失。

交易型应用程序，尤其是那些支持超高速交易模式的应用，也无法和主机集群模式良好地结合，因为从设计上讲，这些应用已经定义了在故障切换时的特殊处理方法。因此，当虚拟机需要做故障切换时，这些应用不能按照集群系统所规划的方式去正常工作。

## 为高可用的虚拟化架构做准备

尽管有三种不同的虚拟机高可用架构，我们仍然需要为创建主机集群做准备，而所需的投入要取决于所选择的管理程序（hypervisor）。VMware 和 Citrix 的管理程序在创建主机集群时，都需要 NFS 存储系统的支持。而 Microsoft 的 Hyper-V，则需要使用 SAN 或 iSCSI 目标存储。SAN 存储对于中小规模企业而言过于昂贵，iSCSI 是更为经济的选择。例如，我们可以借助于一台标准服务器（带有多块做了 RAID 的硬盘），来快速地创建可以支持各种管理程序的共享存储子系统。通过在服务器上安装 iSCSI 目标端软件程序，我们可以把这台存储服务器转变为一台低成本的共享存储，以支持主机故障切换集群的创建。

| 步骤 | 思杰 XenServer | 微软 Hyper-V | VMware ESXi |
|---|---|---|---|
| 1 | 安装 XenServer | 安装 Windows Server 2008 Server Core 并启用 Hyper-V | 安装 ESXi Server |
| 2 | 把主机连接到共享存储 | 把主机连接到共享存储 | 把主机连接到共享存储 |
| 3 | 创建资源池 | 配置和启用故障切换集群功能 | 创建高可用集群 |
| 4 | 创建虚拟机的高可用 | 创建虚拟机的高可用 | 创建虚拟机的高可用 |

当共享存储设备准备完成后，请遵循上表中不同管理程序：Citrix XenServer，Microsoft Hyper-V 以及 VMware ESXi，下创建主机集群的步骤。而虚拟机的创建则可以在主机集群创建之前或之后。但一定要注意的一点是组成虚拟机的文件必须要保存在共享存储设备上，这样才可以保证集群中的所有主机都可以访问。

由于以上的这些原因，我们应该考虑创建高可用的虚拟机（也就是说在主机内部的虚拟机层面建设集群系统），实现和应用程序相关的集群系统。这样的集群系统，可以确保当我们把应用程序迁移到资源池内的虚拟机层上时，依然具有持续地高可用性和稳定性。故障切换集群仅支持状态型的工作负载（stateful workloads）,或者是那些可以从用户会话进程中记录数据的工作负载。

***考虑建立高可用性VMS（在虚拟机层理创建集群）实现应用感知集群。***

对于无状态的工作负载（stateless workloads），或者是仅提供只读服务的工作负载，则可以依赖 NLB 实现高可用。和故障切换集群相似，NLB 是一个可以在虚拟化层获得完全支持的高可用解决方案。

为了确保虚拟工作负载内，状态型应用程序的高可用性，很多公司倾向选择构建于单个站点上的集群系统。这样的集群，通常来说在虚拟架构下是很容易创建的，而且也不需要数据复制软件的支持（数据复制软件通常需要从第三方供应商购买获得）。当我们在创建单站点的子机集群系统时，需要考虑如下这些因素：

**使用交叉原则。**假设我们是在主机集群系统上创建了一个双节点的虚拟机，必须要确保这两台虚拟机节点不是位于主机集群系统中的同一个主机节点上。如果虚拟机集群安装在了同一个主机节点上，那么在该主机节点故障的情况下，整个虚拟

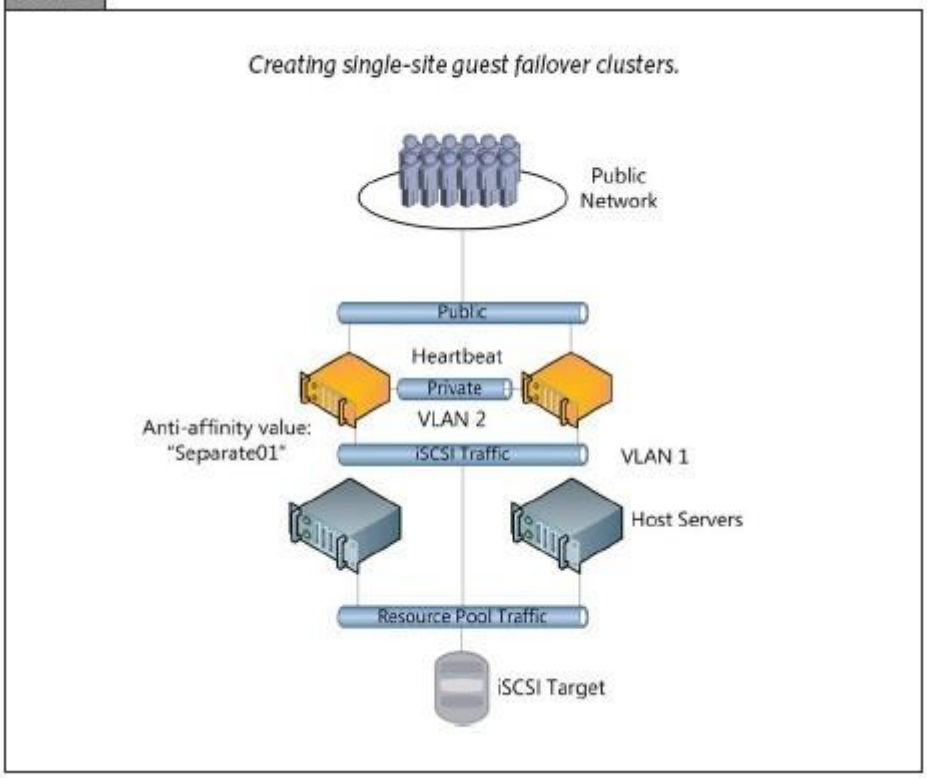
机集群就会停止运行。这样虚拟机集群也就失去了其存在的意义。所以我们要控制虚拟机集群中的节点位于不同的主机节点上，或者是不同的主机集群上。在 Windows 操作系统中，可以通过 Cluster.exe 命令来满足交叉原则，其他的管理程序（hypervisor）也可以通过各自的方法实现。

**对 VLAN 的依赖**。依赖于管理程序的子 VLAN 功能，可以实现子机集群的内部流量和其他的网络流量之间的隔离。我们可以为每台虚拟机内的虚拟网卡设置不同的 VLAN。

**对 iSCSI 存储系统的依赖**。依赖于 iSCSI 存储系统，可以实现子机集群中的共享存储设置。在 iSCSI 协议支持下，我们可以通过网卡接口实现到共享存储设备的访问。所有的虚拟机都可以轻松地访问 iSCSI 存储资源设备，因为它们之间的连接只需要通过网卡就可以实现。

借助于以上三个基本原则，我们就可以创建单站点子机故障切换集群系统，并使其运行在资源池的虚拟化层中（图 3）。正如我们需要私有集群系统创建独立的网络流量一样，iSCSI 存储系统也需要相对独立的网络。而且用于外部最终用户到集群内虚拟机访问的流量也需要独立的网络支持，因此我们需要在虚拟机内和主机上创建多个虚拟机网卡。

FIGURE 3




当子机故障切换集群中的一个节点所运行的主机发生故障时，另一个节点会自动检测到子虚拟机节点失效。然后自动地把该虚拟机上的应用迁移到子机集群中的两外一台虚拟机上。而最终用户在迁移过程中不会感觉到系统停机。

当故障切换集群中的应用程序从一个节点到另一个节点切换时，最终用户可能会感觉到一定程度的响应延迟。然而，这种延迟仅仅会持续几微秒的时间（具体时间取决于应用的不同），因此通常也不会被用户感知。

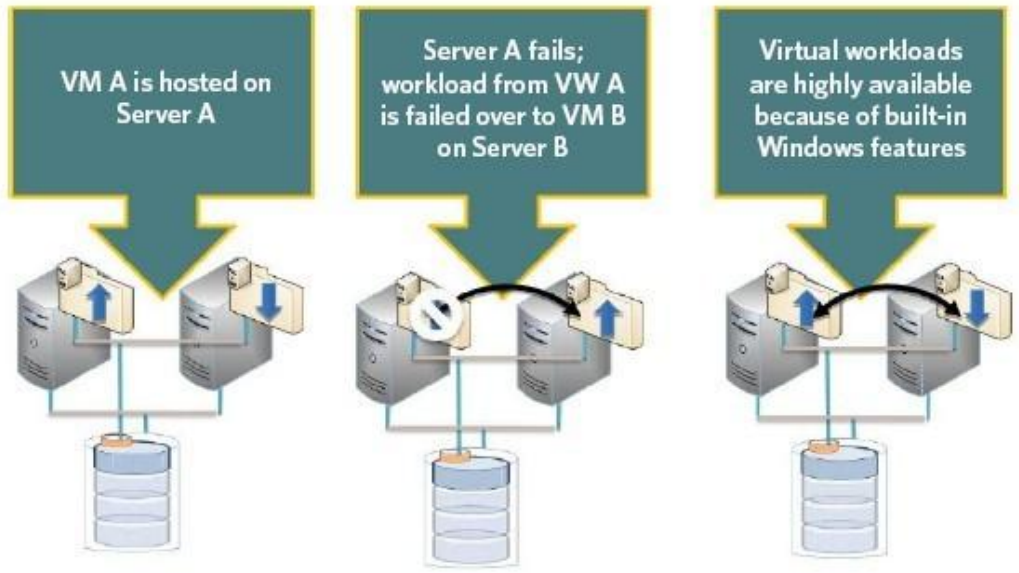
## 使用客户机 NLB 集群

虽然 NLB 是一种 HA 解决方案，但它与失效切换集群不同。在失效切换集群中，仅有一个节点运行指定服务。当节点失效时，另一个节点接管服务并成为该服务的主节点。以上表现由失效切换集群模式的结构所致。因为该模式，同一时刻，只能有一个节点访问给定的存储卷。因此，同一时刻，集群应用只能运行在单一节点上。

NLB 或者服务器负载均衡集群中，每个集群节点提供同样的服务。访问某个服务时，客户被定向到一个单一的 IP 地址。然后，NLB 服务会将用户重定向到集群中第一个可用的节点。因为集群中的每个节点都可以提供同样的服务，所以通常情况下，它们都处于只读模式，并且是无状态的。

FIGURE 4

*Guest application failover during a host failure.*



因为超级管理器（Hypervisor）网络层提供了大量的网络服务（比如，NLB 重定向），因此，完全可以在虚拟机里支持 NLB 集群。这意味着你可以创建一个多节点——最大 32 个 NLB 节点——为生产虚拟机中可用的无状态服务提供 HA 保护。不过，每个 NLB 集群中的节点必须有两个网卡，一个用于管理、另一个用于公网的访问。在虚拟机中，这可以通过添加另外一块网络适配器来实现。

当你在虚拟机中运行生产业务时，首先必须确认是否支持那样的配置。否则，如果出现什么问题，你还需要将虚拟机转换到物理机上。然后，你最好从厂商那里得到些支持。当准备虚拟机时，资源池管理员应当考虑这些配置。

受支持的配置可以运行在独立的（主机）失效集群，或者客户机级的 HA 配置中。当配置虚拟机的时候，请不要忘记还有产品授权的需求。阅读相关的技术文章以正确的配置你的网络。如果没有相关的文章，那就阅读下产品的配置文档。

# 我们的编辑团队

您若有何意见与建议，欢迎与我们的编辑联系。

诚挚感谢以下人员热情参与 TechTarget 中国《高级虚拟化系列手册》的内容编辑工作！


**关于作者**

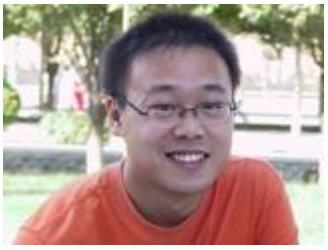
Danielle Ruest 与 Nelson Ruest 是无间断服务、可用性和基础架构优化领域的 IT 专家。他们合著了大量书籍，其中包括 Virtualization: A Beginner’ s Guide 以及 Windows Server 2008。联系方式：infos@reso-net.com。

**李哲贤**

TechTarget中国特邀技术编辑。六年存储行业从业经验。曾先后服务于国内外几家知名存储厂商，对存储虚拟化、容灾备份、数据中心建设等方面有较深入了解。现服务于某跨国企业，从事服务器存储销售支持工作。

李建军

TechTarget中国特邀技术编辑。目前供职于某国际知名存储厂商的全球技术支持中心，负责NAS 产品全球技术支持。自2003年起曾先后在五百强外企从事UNIX、IP、存储相关产品技术支持。



于富春

TechTarget中国虚拟化论坛版主。在大型网站搭建及管理领域具有丰富经验，长期专注于Redhat及开源linux系统、VMware虚拟化产品的学习研究。