红帽高可用性配置，管理和维护
—最强版

原作者：史应生

背景介绍：

随着人们对业务需求和用户满意度期望值的不断提升，很多生产系统（尤其是金融，基金，证券，保险行业和电信）需要提供 7X24 小时的不间断服务。这就需要一个集群的环境，保证当正在提供服务的机器宕机后，集群中的其它机器可以在短时间内接管服务。并且保证能够把出现问题的机器自动启动，使其恢复到初始状态。而且在整个服务切换过程中，不需要任何的人为干预。这就是高可用性的解决方案。

在本文章中，作者针对红帽公司的集群套件，通过监控一个 web 服务器的应用，对集群套件的配置，管理和维护做了详细的说明。

适用读者：

1. 中/高级 Linux 系统管理员
2. 系统集成商
3. 解决方案构架师
4. 所有从事开源的兄弟姐妹

目录

第1章 红帽高可用软件介绍

红帽高可用软件最初建立在由 Mission Critical Linux, Inc. 开发的开源 Kimberlite http://oss.missioncriticallinux.com/kimberlite/ 群集工程。

基于 Kimberlite 建立的版本开始之后，红帽的开发者们在其上进行了大量的增进和修改。以下不全面的列表突出显示了一些此类增进。

- 打包并集成到红帽安装程序以便简化终端用户对其的使用。
- 增加了对多个群集成员的支持。
- 增加了对高可用性 NFS 服务的支持。
- 增加了对高可用性 Samba 服务的支持。
- 增加了图形化监控工具“群集配置工具”。
- 增加了图形化监控管理工具“群集状态工具”。
- 增加了对失效域的支持。
- 增加了对使用监视计时器来保障数据完好性的支持。
- 增加了将会自动重启失效程序的服务监控。
- 重新编写了服务管理器来实现额外的群集全局的操作。
- 一组各类错误修正。

红帽群集管理器软件吸收了来自 Linux-HA 工程的 STONITH 兼容电源开关模块，参见 http://www.linux-ha.org/stonith/。

红帽群集管理器是一组技术集合。它们被综合在一起来提供数据完好性和在失效情况下保持程序可用性的能力。通过使用冗余的硬件、共享的磁盘贮存区、电源管理、以及强健的群集管理和应用程序失效转移机制，群集能够满足企业市场的需要。

群集特别适合于数据库应用程序、网络文件服务器、以及带有动态内容的万维网服务器，它还可以和 Piranha 负载均衡群集软件（基于 Linux 虚拟服务器，LVS 计划）一起用来部署高可用性的电子商务站点。这类站点除了负荷平衡能力之外还具备提供完全的数据完好性和应用程序可用性的能力

1.1 群集总览

要设置群集，管理员必须把成员系统（member systems，通常被简称为成员，member）连

接到群集硬件，并把成员配置入群集环境。群集的基础是高级主机成员算式。该算式使用以下节点间的通讯方法来保证群集在任何时刻都会保持完全的数据完好性：

- 群集系统之间用于探测心跳（heartbeat）的网络连接
- 包含群集状态的共享磁盘贮存区上的共享状态（Shared state）

要使应用程序和数据在群集中具有高可用性，你必须把服务（service，如应用程序和共享磁盘贮存区）配置成一组分离的、被命名的属性和资源。你可以给它们分配 IP 地址来提供透明客户访问。譬如，管理员可以设置一个服务来为客户提供高可用性的数据库应用程序数据。你可以关联服务和失效域（failover domain）。失效域是有资格运行该服务的群集成员子集。一般来说，任何有资格的成员都可以运行服务，并存取共享磁盘贮存区上的服务数据。然而，为了保持数据的完好性，每个服务在某一时刻只能在一个群集成员上运行。你可以指定失效域中的成员是否是有序的，你还可以指定服务是否被限定只能在和它相关联的失效域中的成员上运行。（当和一个无限制的失效域相关联时，在没有可用的失效域成员的情况下，服务可以在任何群集成员上被启动。）

你可以设置一个活跃--活跃配置（active-active configuration），成员运行各自不同的服务；或者热备份配置（hot-standby configuration），主成员运行所有的服务，备份群集成员只有在主系统失效时才接管这些服务。

图 1－1 显示一个使用“活跃--活跃配置”的群集实例。

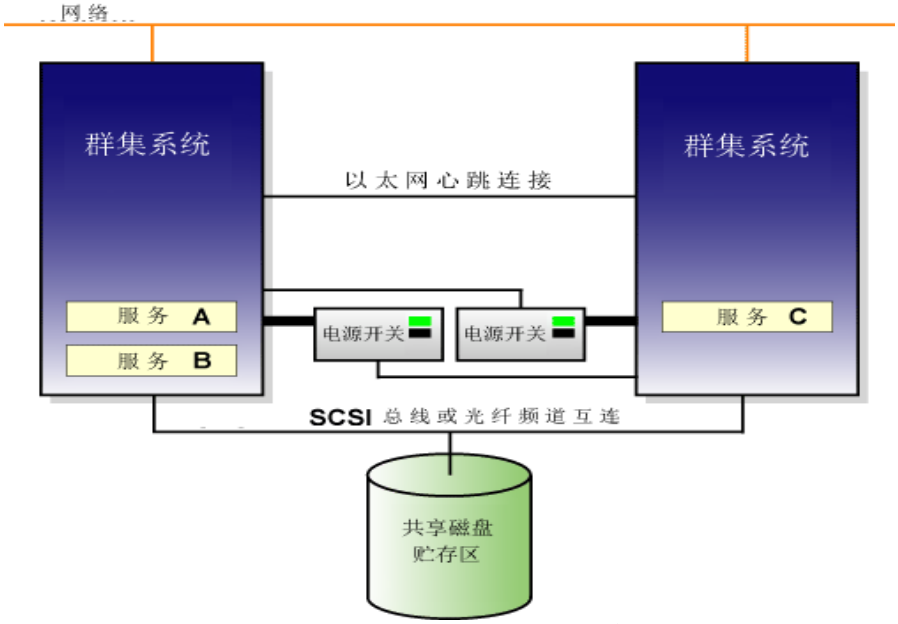

图 1－1 活跃-活跃配置的群集实例

如果某硬件或软件出现了故障，群集会自动在正常运行的成员上重新启动失效成员上的服务。服务失效转移（service failover）能力在保证不丢失数据的同时，还给用户带来了极少

的干扰。当失效的成员被恢复后，群集可以在成员间重新平衡其服务。

除此之外，你还可以完整地停止在一个群集成员上运行的服务，然后在另一个成员上重新运行它们。服务重新定位（service relocation）能力使你在群集系统需要维护时仍能够保持应用程序和数据的可用性。

1.2 群集特性

群集包括下列特性：

- 无单一失效点硬件配置

群集可以包含双控制器的 RAID 阵列、多个网络频道、以及冗余的不间断电源（UPS）系统来确保不会出现导致程序停运或数据丢失的单一失效情况。

另外，你还可以设置低费用的群集来提供比“无单一失效点”群集稍低的可用性。譬如，你可以设置带有单控制器的 RAID 阵列和只有单个以太网频道的群集。

- 服务配置体制

群集使你能够轻松地配置个别服务来使数据和应用程序具有高可用性。要创建一项服务，你需要指定服务所使用的资源以及服务的属性，包括服务名称、应用程序的启动、停止和状态脚本、磁盘分区、挂载点、以及你优选的要在其上运行该服务的群集成员。添加了一项服务后，群集管理软件就会把这些信息贮存在共享贮存区上的群集配置文件中，这样，所有群集成员都可以存取这些配置信息。

群集为数据库应用程序提供了使用简便的体制。譬如，数据库服务（database service）给数据库应用程序提供高可用性数据。在某个群集成员上运行的应用程序，如万维网服务器，给数据库客户系统提供网络存取能力。如果服务失效转移到另一个群集成员，应用程序仍旧能够存取共享的数据库数据。一个可网络存取的数据库服务通常被指派了一个 IP 地址，这个地址以及服务都会被失效转移以便保持客户的透明存取能力。

群集服务体制还可以被简便地扩展到其它应用程序。

- 失效域

通过把服务分配给限制的失效域（restricted failover domain），你可以限定在失效转移情况下有资格运行服务的成员。(被分配给限制的失效域的服务不能在没有包括在失效域中的成员上启动。）你可以通过首选项来给失效域中的成员排序，从而确定某个特定成员会运行这个服务（只要该成员处于活跃状态)。如果服务被分配给无限制的失效域（unrestricted failover domain），服务就可以在任何可用的群集成员上被启动（如果失效域中的所有成员都不可用的话)。

- 数据完好性保证

要保证数据完好性，在某一时刻只有一个群集系统能够运行服务或存取服务数据。在群集硬件配置中使用电源开关能够使某个成员在失效转移情况下重新启动另一个成员上的服务前重开它的电源。这会防止两个系统同时存取同一数据从而损坏它们。虽然这并不是必要的部件，但我们仍推荐你使用电源开关来确保所有失效情况下的数据完好性。监视定时器是另一种确保服务失效转移的正确操作的电源控制方法。

- 群集管理用户界面

群集管理界面提供了进行以下管理任务的设施：创建、启动、和停止服务；把服务从一个成员重新定位到另一个成员；修改群集配置（添加或删除服务或资源）；以及监视群集成员和服务。

- 以太网频道接合

要监视其它成员的健康状况，每个成员都监视远程电源开关的健康状况，并且通过网络发出心跳试通。由于以太网频道接合，多个以太网接口被配置成仿佛是一个一样，从而减少了典型的切换型以太网系统连接中的单一失效点所带来的威胁。

- 用于仲裁信息的共享贮存区

共享的状态信息包括群集是否活跃。服务状态信息包括服务是否在运行以及哪个成员正在运行该服务。每个成员都检查这些信息来保证其它成员处于最新状态。

在一个只有两个成员的群集中，每个成员都定期把一个时间戳和群集状态信息写入位于共享磁盘贮存区的两个共享群集分区上。要保证正确的群集操作，如果某成员无法在启动时写入主共享群集分区和屏蔽共享群集分区，它将不会被允许加入群集。此外，如果某群集成员不更新其时间戳，或者到系统的“ heartbeats ”（心跳）失败了，该成员就会从群集中删除。

图 1－2 显示了系统在群集配置中的通信方式。注意，用来通过串口进入系统控制台的终端服务器不是必需的群集部件。

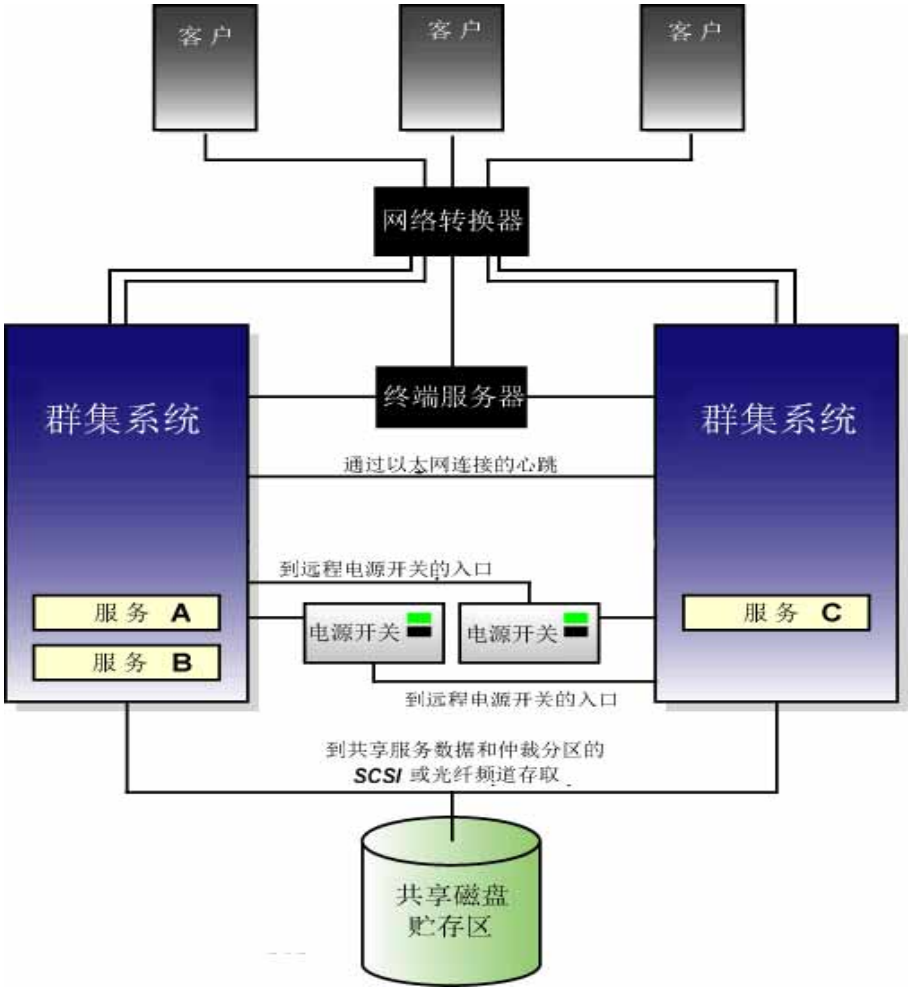

图 1—2．群集通信机制

- 服务失效转移能力

如果硬件或软件失效情况出现了，群集会采取恰当的行动来保持应用程序的可用性和数据完好性。例如，如果某成员失效了，另一个成员（在相关的失效转移域中的，若使用了的话；或在群集中的的另一个成员）会重新启动那些服务。已经在该成员上运行的服务将不会受到影响。

当失效的系统重新引导并能够写入共享群集分区后，它就能够重新加入群集来运行服务。依据服务的配置情况而定，群集可以重新在成员间平衡服务。

- 手工重新定位服务的能力

除了自动转移失效服务外，群集还使你能够在一个群集系统上完整地停止服务，然后在另一个系统上启动它们。你可以在按计划进行对某个成员系统的维护任务的同时，仍旧提供应用程序和数据可用性。

- 事件记录设施

为确保问题在影响服务可用性之前被检测到并被解决，群集守护进程使用传统的 Linux syslog 子系统来记录消息。你可以定制所记录消息的严重性级别。

应用程序监视

群集服务体制还可以监视应用程序的状态和健康情况。这样，如果出现了应用程序特有的失效情况，群集会自动重新启动该应用程序。在应用程序失效情况下，它会试图在最初运行它的成员上重新启动；如果这不奏效，它会试图在另一个群集成员上启动。你可以通过给服务分配失效域来指定哪些成员具备运行该服务的资格。

1.3 系统所需最小配置

软件准备：

- **操作系统**：红帽企业版 Linux AS4 Update3 即 RHEL4U3
- **红帽集群套件**：Red Hat Cluster Suite (RHCS)

  注意：RHCS 的版本必须和操作系统对应。比如操作系统是 RHEL4U3，RHCS 也必须是 4.0Update3 的版本

硬件准备：

- **服务器**：两台（红帽官方文档称支持的最多节点数是 16 个）
- **存　储**：支持 ISCSI,SAN,NAS。

  如果在做 Fail-over 时,两台服务器所使用的资源不需要共享,就不需要存储。
- **Fence 设备**：Fence 设备的作用时在一个节点出现问题时，另一个节点通过 fence 设备把出现问题的节点重新启动，这样做到了非人工的干预和防止出现问题的节点访问共享存储，造成文件系统的冲突。

  关于 Fence 设备，有外置的比如 APC 的电源管理器。很多服务器都是内置的，只不过不同厂家的叫法不同而已。比如 HP 的称为 iLo,IBM 的称为 BMC,Dell 的称为 DRAC。在本文章中，我们使用 HP 的服务器，所以 Fence 设备为 iLo.
- **网络交换机**：一个

应用程序：

在配置 HA 前，保证所要监控的应用在任何一个节点上工作正常。

监控脚本：

如果集群软件监控的服务是操作系统自带的服务，那么监控脚本位于/etc/rc.d/init.d/下面。

如果是自己开发的应用，需要按照 LSB 的标准编写集群的监控脚本。您可以按照/etc/rc.d/init.d/这个目录下的脚本，结合自己的应用，编写自己的监控脚本。监控脚本必须具备对服务的启动，停止和状态监控三个功能。

网络规划：

每个服务器的 IP 地址和主机名。

Fence 设备的 IP 地址和主机名。

集群对外提供服务的虚拟 IP 地址 VIP 和虚拟主机名。

第2章 系统配置

2.1 系统主机信息和网络配置

设置了基本的群集硬件后，下一步是在每个成员上安装红帽企业 Linux，并保证所有系统都能够识别连接了的设备。其步骤如下：

1. 在所有群集成员上安装红帽企业 Linux。

此外，在安装红帽企业 Linux 时，强烈推荐你执行以下步骤：

- 在安装红帽企业 Linux 前收集成员和接合以太网端口的 IP 地址。注意，接合以太网端口的 IP 地址可能是专用 IP 地址（如 10.x.x.x）。
- 不要把本地文件系统（如 /、/etc、/tmp、和/var）放在共享磁盘或和共享磁盘相同的 SCSI 总线上。这有助于防止其它群集成员不小心地挂载这些文件系统，还可以保留用于群集磁盘的总线上有限的 SCSI 标记号码。
- 把 /tmp 和 /var 放在不同的文件系统中。这能够提高成员的性能。
- 当成员引导时，请确定成员检测磁盘设备的顺序和红帽企业 Linux 安装时的检测顺序相同。如果设备没有按同一顺序被检测到，成员可能不能够引导。
- 在使用配置了大于零的逻辑单元号码（LUN）的 RAID 贮存区时，有必要通过在 /etc/modprobe.conf 添加以下内容来启用 LUN 支持：

```
options scsi_mod max_scsi_luns=255
```

修改了 modprobe.conf 文件后，有必要使用 mkinitrd 来重新建构初始内存磁盘。关于使用 mkinitrd 来创建 ramdisk 的信息。

2. 重新引导成员。
3. 在使用终端服务器时，配置红帽企业 Linux 来把控制台消息发送到控制台端口。
4. 编辑每个群集成员上的 /etc/hosts 文件，包括在群集中使用的 IP 地址或保证其地址在 DNS 中。
5. 降低交替内核引导超时的限度来缩短成员的引导时间。
6. 请确保连接远程电源开关的串口上没有关联任何登录程序（或 getty）。要执行这项任务，编辑 /etc/inittab 文件，使用一个散列符号（#）来把所有和用在远程电源开关上的串口相对应的项目都变成注释。然后使用 init q 命令。
7. 校验是否所有系统都检测到了全部已安装硬件：
   - 使用 dmesg 命令来显示控制台启动消息。

- 使用 cat /proc/devices 命令来显示内核中配置了的设备。

8. 通过使用 ping 命令来在成员间发送测试包来校验各成员是否能够通过所有的网络接口通信。

9. 如果打算配置 Samba 服务，请校验是否安装了 Samba 服务所需的 RPM 软件包。

2.1.1 主机信息配置

/etc/hosts 文件包含 IP 地址到主机名的转换表。每个成员上的 /etc/hosts 文件都必须包含以下项目：

- 所有群集成员的 IP 地址和相关联的主机名
- 用于点对点以太网心跳连接的 IP 地址（可以是专用地址）

除了使用 /etc/hosts 文件外，你还可以使用 DNS 或 NIS 这类命名系统来定义群集所用的主机名。不过，要限制依赖关系的数量或优化可用性，强烈推荐你使用 /etc/hosts 文件来为群集网络接口定义 IP 地址。

以下是成员上的 /etc/hosts 文件示范：

```
127.0.0.1       localhost.localdomain   localhost
192.168.100.11      redhat_ha1
192.168.100.13      redhat_ha2
192.168.100.29      redhat_ha

192.168.100.15      ilo_ha1
192.168.100.17      ilo_ha2
```

注：192.168.100.29 和 redhat_ha 分别为对外虚拟 IP 和虚拟主机名（对外提供服务的 IP 地址和主机名）

2.1.2 修改主机名和网关信息

修改/etc/sysconfig/network

```
NETWORKING=yes
HOSTNAME= redhat_ha1
GATEWAY=192.168.100.254
```

注：按照相应的配置修改 redhat_ha2 的机器

2.1.3 设置网卡绑定功能

修改/etc/sysconfig/network-scripts/ifcfg-bond0

```
DEVICE=bond0
IPADDR=192.168.100.11
NETMASK=255.255.255.0
NETWORK=192.168.100.0
BROADCAST=192.168.100.255
ONBOOT=yes
BOOTPROTO=none
USERCTL=no
```

修改/etc/sysconfig/network-scripts/ifcfg-eth0

```
DEVICE=eth0
USERCTL=no
ONBOOT=yes
MASTER=bond0
SLAVE=yes
BOOTPROTO=none
```

修改/etc/sysconfig/network-scripts/ifcfg-eth1

```
DEVICE=eth1
USERCTL=no
ONBOOT=yes
MASTER=bond0
SLAVE=yes
BOOTPROTO=none
```

添加如下信息到/etc/modprobe.conf

```
alias bond0 bonding
options bond0 mode=1 miimon=100 use_carrier=0
```

注：按照相应的配置修改 redhat_ha2 的机器

2.1.4 缩短内核引导超时限度

有可能通过缩短内核引导超时限度来缩短成员的引导时间。在红帽企业 Linux 的引导过程中，引导装载程序允许你指定要引导的另一个内核。指定内核的默认超时时间是十秒钟。

修改成员的内核引导超时限度，按照以下方式来编辑恰当的文件：

在使用 GRUB 引导装载程序时，你应该修改 /boot/grub/grub.conf 中的超时参数来指定恰当的 timeout 秒数。 要把间隔设为 3 秒钟，按照以下方式来编辑参数：

timeout = 3

在使用 LILO 或 ELILO 引导装载程序时，编辑 /etc/lilo.conf 文件（x86 系统）或 elilo.conf 文件（Itanium 系统），并指定想要的 timeout 数值（以十分之一秒为单位）。以下的例子把超时限度设为三秒钟：

timeout = 30

要应用 /etc/lilo.conf 文件中的改变，使用 /sbin/lilo 命令。

在 Itanium 系统上，要应用 /boot/efi/efi/redhat/elilo.conf 文件中的改变，使用 /sbin/elilo 命令。

2.2 关闭不需要的系统服务

目的：提高系统的启动速度。

```
# chkconfig kudzu off
# chkconfig sendmail off
# chkconfig nfs off
# chkconfig smartd off
# chkconfig cups off
# chkconfig rhnsd off
# chkconfig iptables off
# chkconfig autofs off
# chkconfig acpid off
# chkconfig apmd off
```

注：在两个机器上都要操作

2.3 为系统打补丁

对于集群软件的监控脚本，如果需要调用/etc/rc.d/init.d/functions 这个文件中的函数，就需要打补丁。目的是使监控脚本符合 Linux LSB 的规范。

```
# cp functions.patch /etc/rc.d/init.d
# cd /etc/rc.d/init.d
# cp functions functions.orig
# patch functions functions.patch
```

注：在两个机器上都要操作

2.4 安装高可用的 HA 监控脚本

```
# cp redhat_script /etc/rc.d/init.d
```

注：其中，redhat_script 是集群软件的监控脚本。

如果监控的是系统自带的服务，就不需要这一步。比如：在本配置中，对于 web 服务器，它的监控脚本就是/etc/rc.d/init.d/httpd。这个脚本是系统自带的。

2.5 共享存储配置

在本配置中，我们使用了共享存储的其中一个 LUN，用于存放 web 服务器的动态和静态页面，即 ”DocumentRoot”。

下面的操作只需要在一个主机上操作：

```
# fdisk /dev/sdb
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF disklabel
Building a new DOS disklabel. Changes will remain in memory only,
until you decide to write them. After that, of course, the previous
content won't be recoverable.
```

```
The number of cylinders for this disk is set to 78325.
There is nothing wrong with that, but this is larger than 1024,
and could in certain setups cause problems with:
1) software that runs at boot time (e.g., old versions of LILO)
2) booting and partitioning software from other OSs
   (e.g., DOS FDISK, OS/2 FDISK)
Warning: invalid flag 0x0000 of partition table 4 will be corrected by w(rite)

Command (m for help): p
         Device Boot      Start         End      Blocks   Id  System

Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-78325, default 1):
Using default value 1
Last cylinder or +size or +sizeM or +sizeK (1-78325, default 78325): +10G

Command (m for help): w

The partition table has been altered!

Calling ioctl() to re-read partition table.
Syncing disks.
```

```
# mke2fs -j /dev/sdb1
```

在每个主机上建立挂装目录:

```
# mkdir /share_fs
```

第3章群集配置

在进行的了系统的配置后，就可以安装群集系统软件和群集配置软件了。

3.1安装红帽群集管理器软件包

3.1.1使用软件包管理工具来安装

在光盘驱动器中插入红帽群集套件光盘。如果你使用了图形化桌面，该光盘会自动运行软件包管理工具。点击「前进」来继续，见图 3-1。

图 3-1. 软件包管理工具

选择红帽群集套件的复选箱，然后点击「细节」来查看软件包描述，见图 3-2。

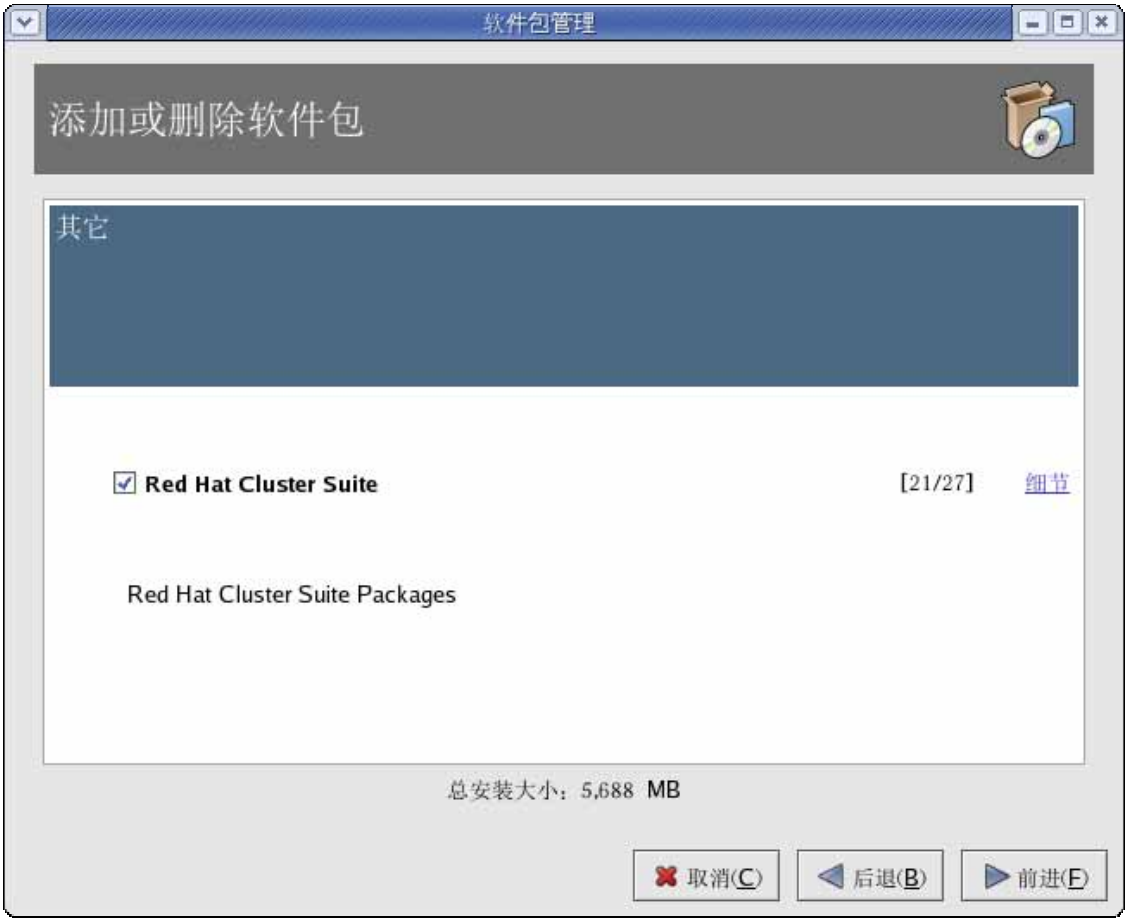

图 3-2. 选择软件包组

软件包管理工具显示了要安装的软件包的总览。点击「前进」来安装软件包。 安装完成后，点击「结束」来退出软件包管理工具

3.1.2 使用 rpm 来安装

如果你没有使用图形化桌面环境，你可以在 shell 提示下使用 rpm 工具来手工安装软件包。把红帽群集套件光盘插入光盘驱动器。登录 shell 提示，转换到该光盘的 RedHat/RPMS/ 目录下，然后以根用户身份键入以下命令。（把 <version> 和 <arch> 替换成要安装的软件包的版本和体系）：

```
# cd  /media/cdrom/RedHat/RPMS
#rpm –ivh  ccs-<version>.<arch>.rpm
           cman-<version>.<arch>.rpm
           cman-kernel-<version>.<arch>.rpm
           dlm-<version>.<arch>.rpm
```

dlm-kernel-smp-<version>.<arch>.rpm
fence-<version>.<arch>.rpm
gulm-<version>.<arch>.rpm
iddev-<version>.<arch>.rpm
ipvsadm-<version>.<arch>.rpm
magma-<version>.<arch>..rpm
magma-plugins-<version>.<arch>.rpm
perl-Net-Telnet-<version>.<arch>.rpm
piranha-<version>.<arch>.rpm
rgmanager-<version>.<arch>.rpm
system-config-cluster-<version>.<arch>.noarch.rpm

3.2群集配置工具

使用下面的方法来进入群集配置工具：

在 shell 提示下，键入 system-config-cluster 命令

程序首次启动时，群集配置工具会被显示。完成了群集配置后，这个命令就会默认启动群集状态工具。

3.2.1 创建一个新的群集配置

选择 Create New Configuration 按钮 (见图 3-3)

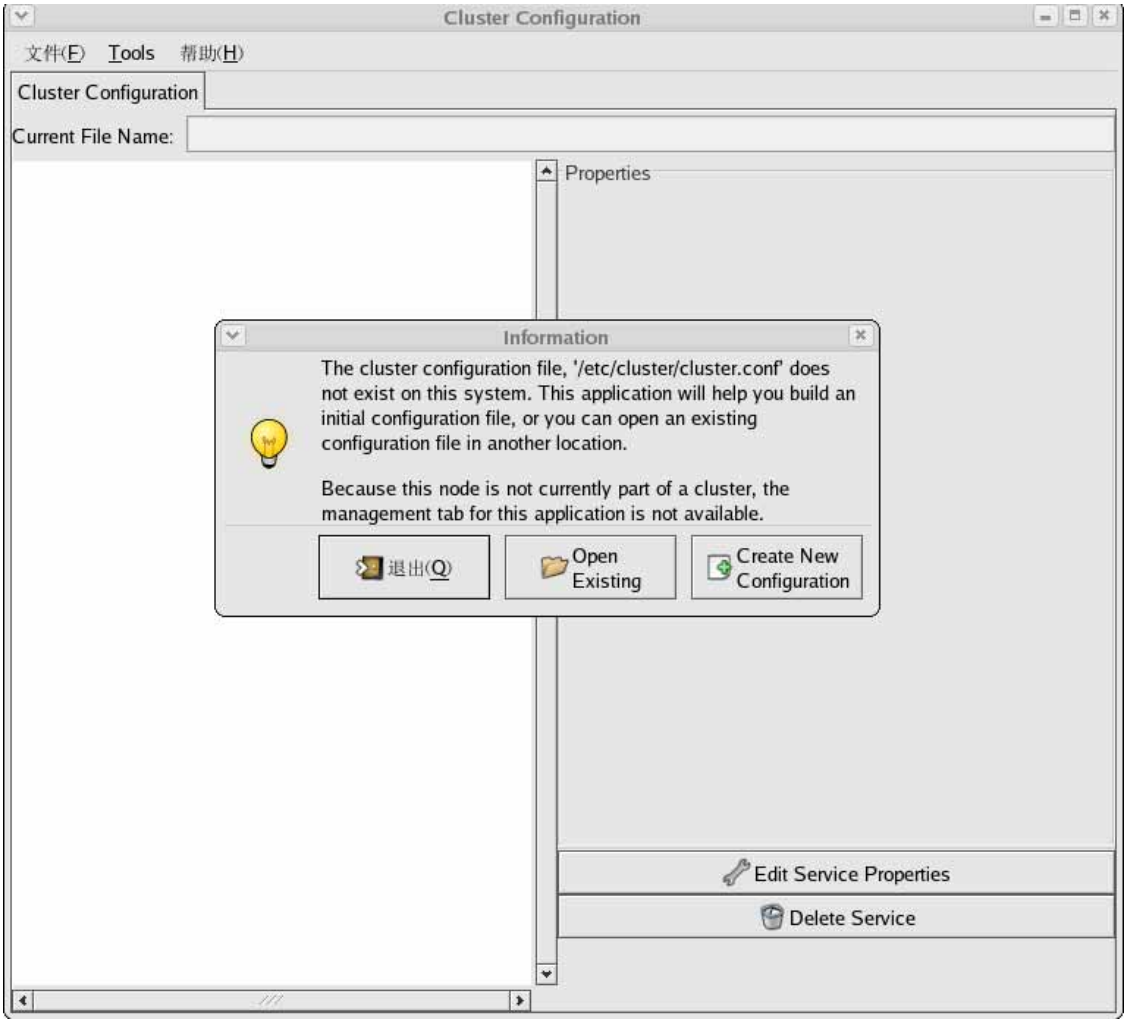

图 3-3

3.2.2选择锁机制

选择 Distributed Lock Manager (DLM)，然后点击确定按钮(见图 3-4)

图 3-4

3.2.3添加群集成员节点

点击 Cluster->Cluster Nodes->Add a Cluster Node(见图 3—5)

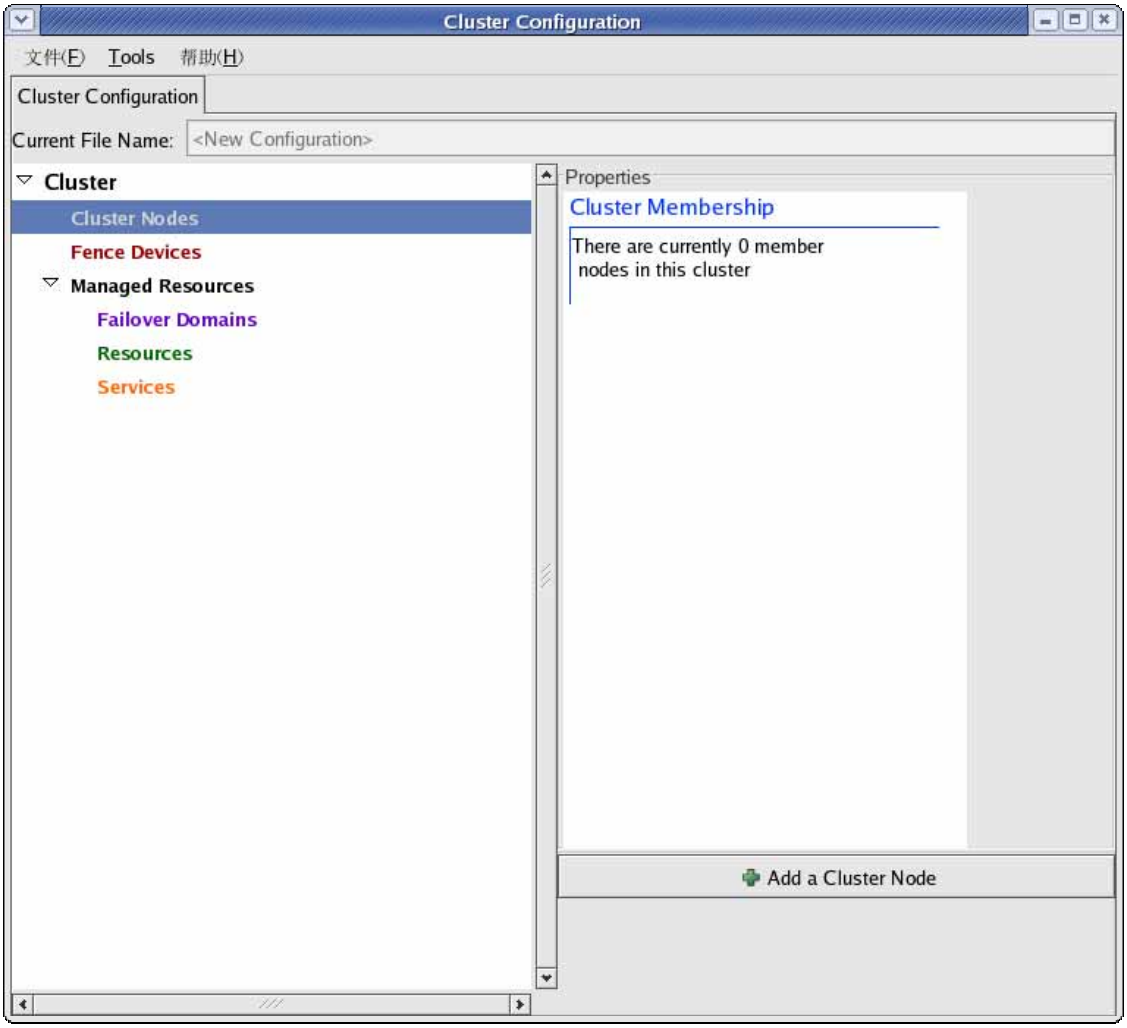

图 3-5

3.2.3.1输入第一个成员节点的信息

在 cluster Node Name 中输入第一个机器的主机名，redhat_ha1，Quorum Votes：不填，然后点击确定．(见图 3－6)

图 3-6

3.2.3.2 输入第二个成员节点的信息

在 cluster Node Name 中输入第二个机器的主机名，redhat_ha2，Quorum Votes：不填，然后点击确定．(见图 3-7)

图 3-7

3.2.4 添加 Fence 设备

要确保数据完好性，在某一时间内只有一个成员能够运行服务和存取服务数据。在群集硬件配置中使用电源控制器（又称电源开关）就使成员能够在失效转移情况下，重新启动另一个成员上的服务之前重开它的电源。这会防止两个系统同时存取同一数据从而损害它。虽然这并不是必需的，我们推荐你使用电源控制器来保证在所有失效情况下的数据完好性。监视计时器是另一种电源控制，它可以确保服务失效转移操作的正确运行。

3.2.4.1增加第一个 Fence 设备

输入第一个 Fence 设备的信息,点击 Cluster->Fence Devices->Add a Fence Device 来加入第一个 iLO 设备：(见图 3-8) 选择 HP ILO Device

Name: 为 ilo 的名字

Login: Administrator:为 ilo 的登录名

Password: 为 ilo 的登录密码

Hostname:为 ilo 的主机名

图 3-8

3.2.4.2增加第二个 Fence 设备

输入第二个 Fence 设备的信息,点击 Cluster->Fence Devices->Add a Fence Device 来加入第一个 iLO 设备：(见图 3-9) 选择 HP ILO Device

Name: 为 ilo 的名字

Login: Administrator:为 ilo 的登录名

Password: 为 ilo 的登录密码

Hostname:为 ilo 的主机名

图 3-9

3.2.5 建立 Fence 设备和节点联系

在添加了 fence 设备后，需要建立 fence 设备和每个节点的对应关系，使每个节点可以通过 Fence 设备对节点的开机，关机和重启进行管理或者对节点的状态进行查询。

3.2.5.1建立第一个节点和 Fence 设备的关系

点击 Cluster Nodes->redhat_ha1->Manage Fencing For This Node 按钮（图 3－10)。进入图 3－11 所示界面，点击 Add a New Fence Level 按钮，进入 3－12 所示界面，点击 Fence-Level-1,进入 3-13 所示界面，点击 Add a New Fence to this Level，进入 3-14 所示界面，选择 ilo-ha1,然后点击确定按钮，进入 3-15 所示界面，最后点击关闭按钮。

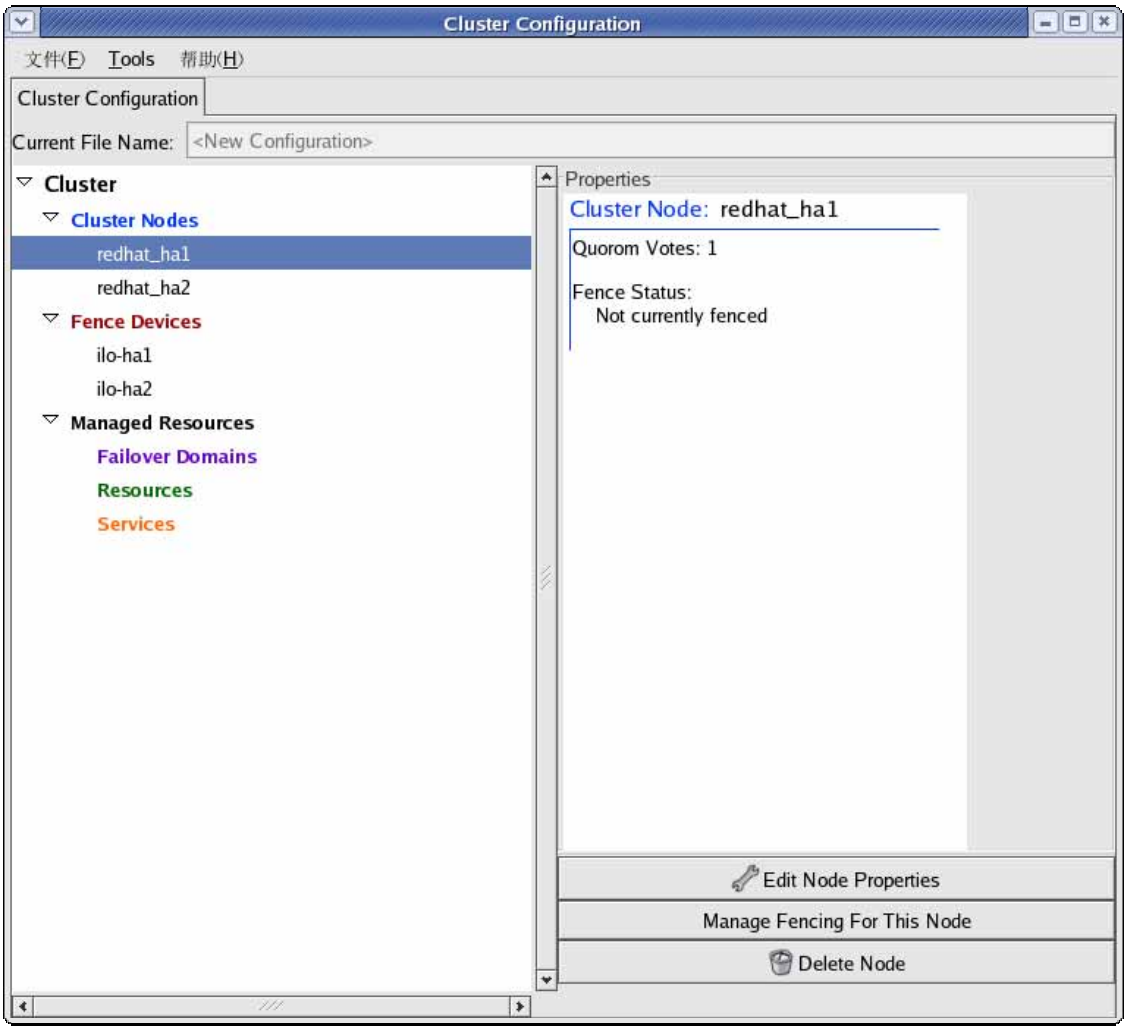

图 3-10

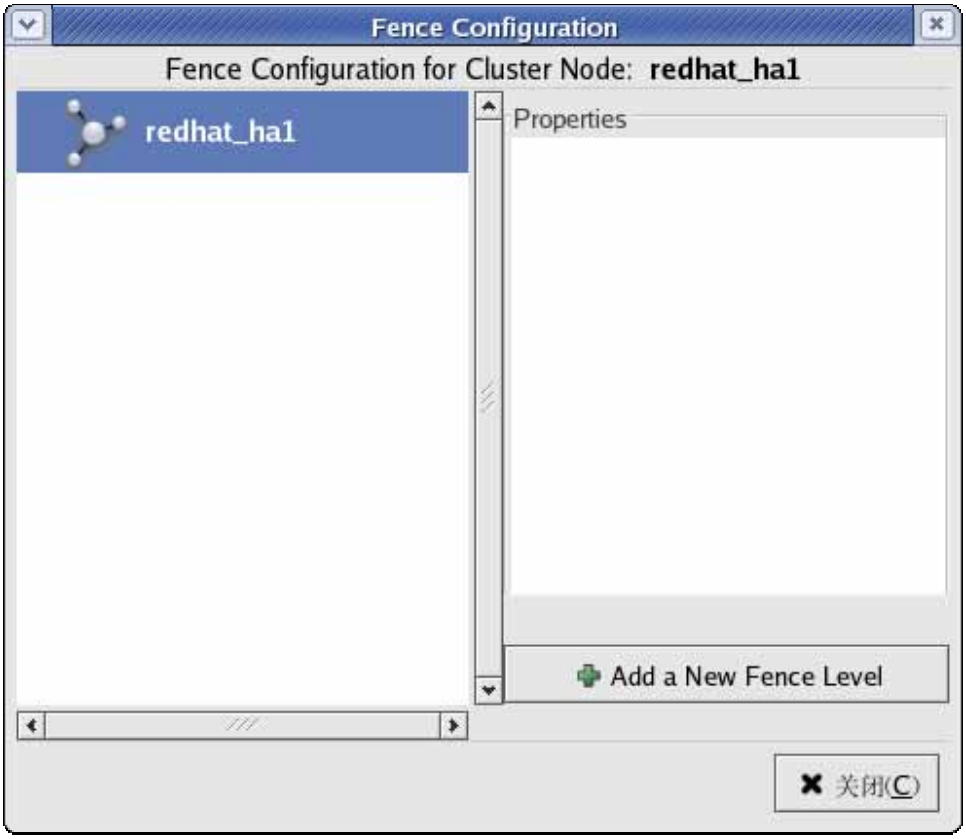
Fence Configuration
Fence Configuration for Cluster Node: redhat_ha1
redhat_ha1
Properties
Add a New Fence Level
关闭(C)

图 3-11

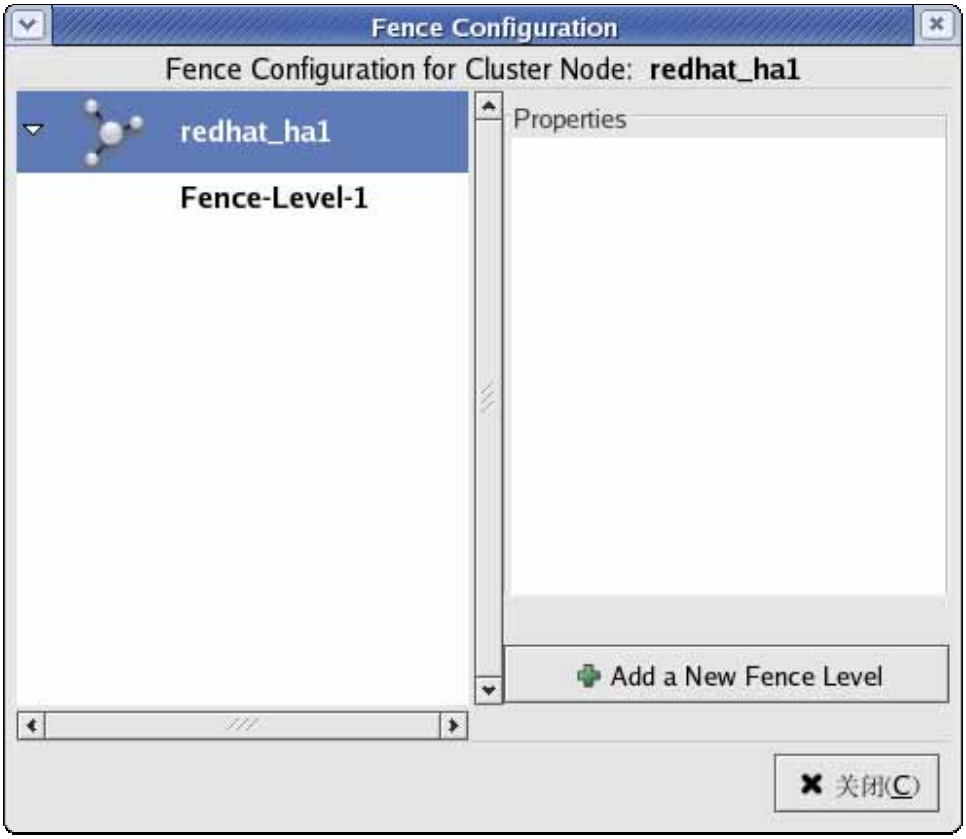
Fence Configuration
Fence Configuration for Cluster Node: redhat_ha1
redhat_ha1
Fence-Level-1
Properties
Add a New Fence Level
关闭(C)

图 3-12

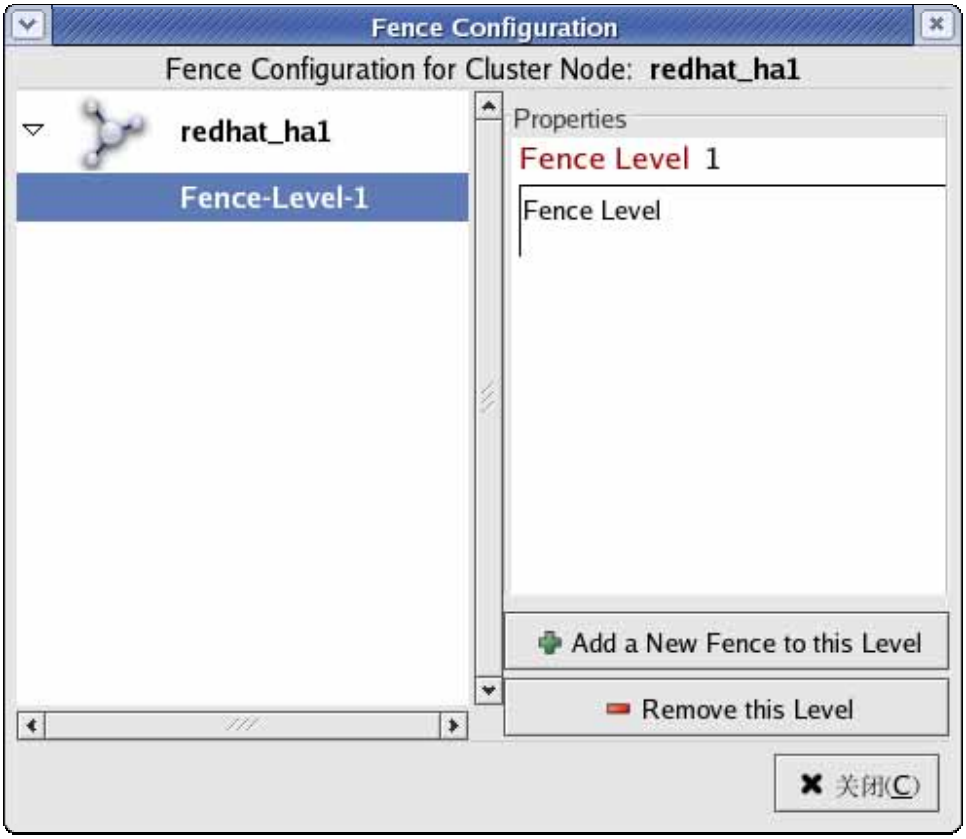
Fence Configuration
Fence Configuration for Cluster Node: redhat_ha1
redhat_ha1
Fence-Level-1
Properties
Fence Level 1
Fence Level
Add a New Fence to this Level
Remove this Level
关闭(C)

图 3-13


Fence Properties
Add a New Fence
ilo-ha1
Fence Device Type:
HP ILO Device
RSA Fence Option
(No Parameters)
ILO Fencing
(No parameters)
取消(C)
确定(O)

图 3-14

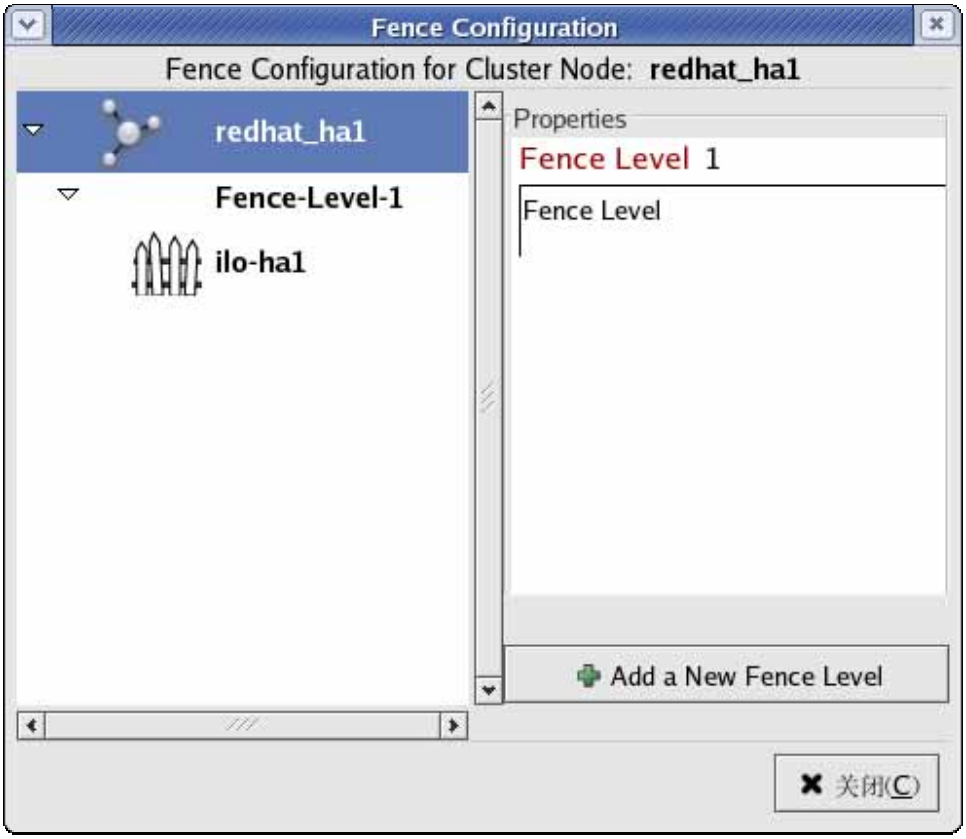

图 3-15

3.2.5.2建立第二个节点和 Fence 设备的关系

点击 Cluster Nodes->redhat_ha2->Manage Fencing For This Node 按钮（图 3－16)。进入图 3－17 所示界面，点击 Add a New Fence Level 按钮，进入 3－18 所示界面，点击 Fence-Level-1，进入 3-19 所示界面，点击 Add a New Fence to this Level，进入 3-20 所示界面，选择 ilo-ha2，然后点击确定按钮，进入 3-21 所示界面，最后点击关闭按钮。

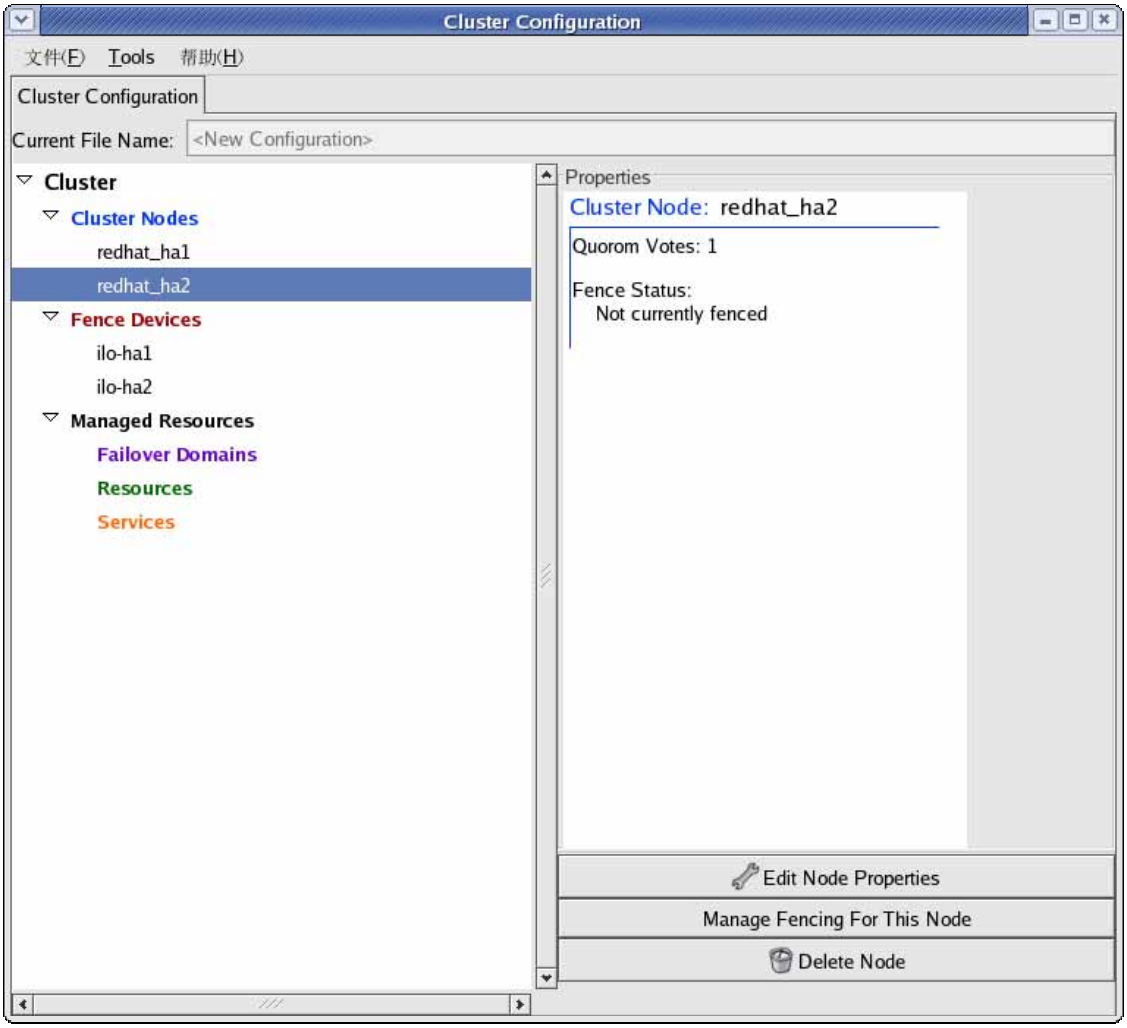
Cluster Configuration
文件(F) Tools 帮助(H)
Cluster Configuration
Current File Name: <New Configuration>
Cluster
Cluster Nodes
redhat_ha1
redhat_ha2
Fence Devices
ilo-ha1
ilo-ha2
Managed Resources
Failover Domains
Resources
Services
Properties
Cluster Node: redhat_ha2
Quorom Votes: 1
Fence Status:
Not currently fenced
Edit Node Properties
Manage Fencing For This Node
Delete Node

图 3-16

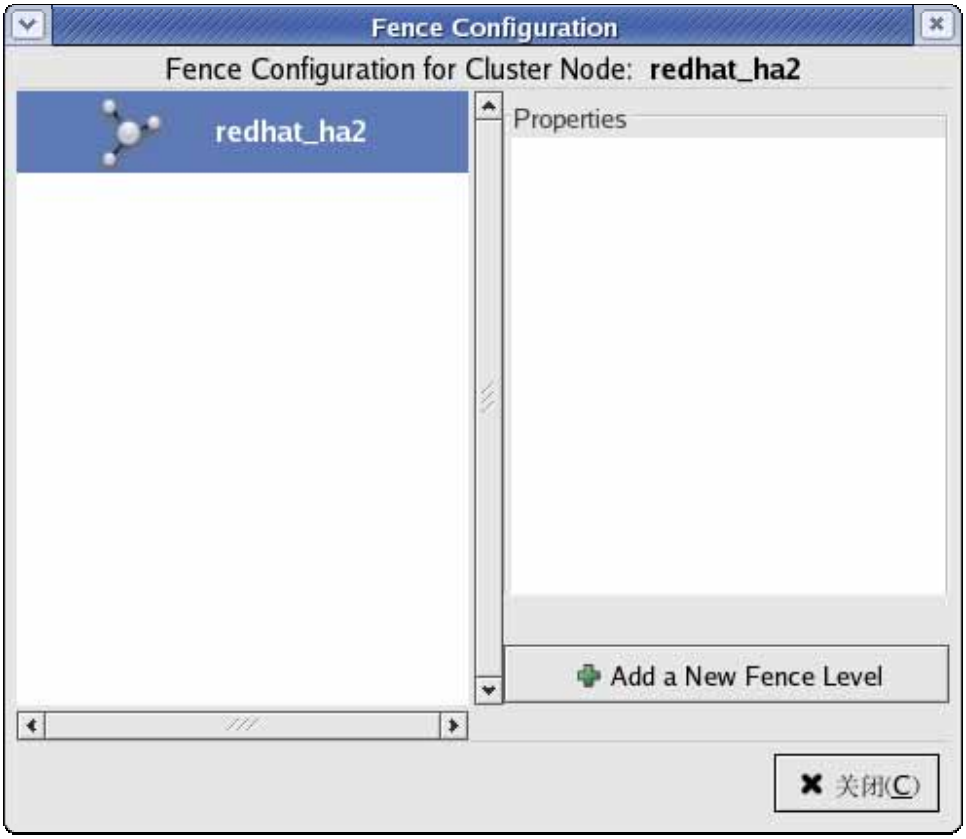
Fence Configuration
Fence Configuration for Cluster Node: redhat_ha2
redhat_ha2
Properties
Add a New Fence Level
关闭(C)

图 3-17

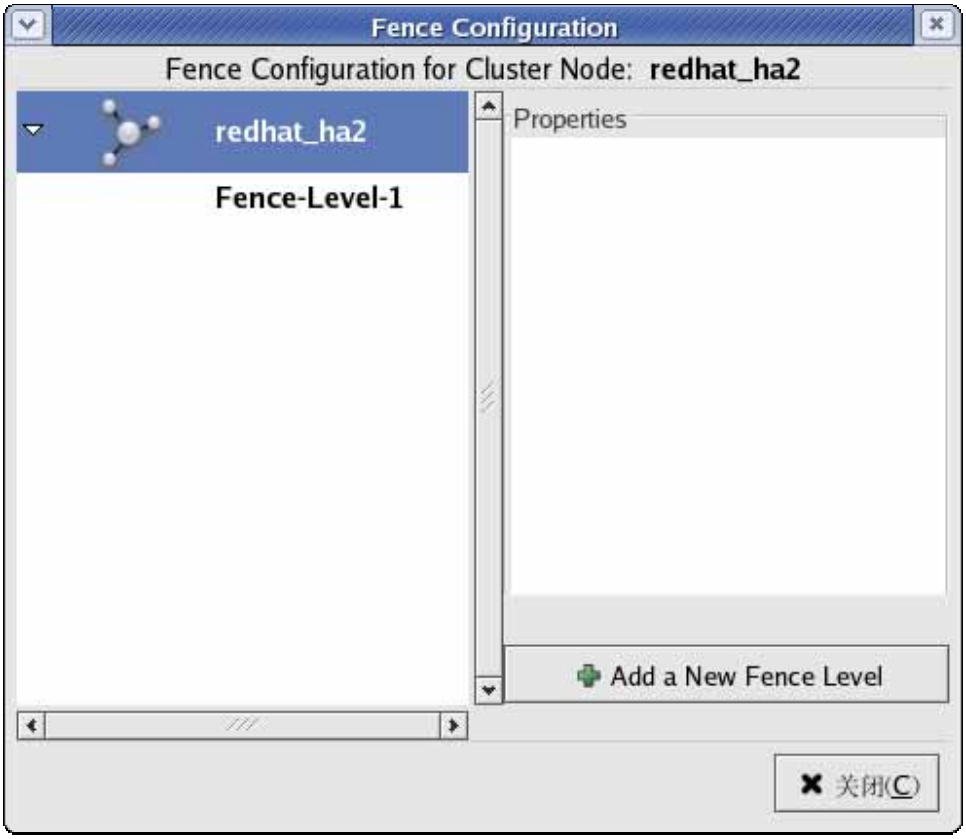
Fence Configuration
Fence Configuration for Cluster Node: redhat_ha2
redhat_ha2
Fence-Level-1
Properties
Add a New Fence Level
关闭(C)

图 3-18

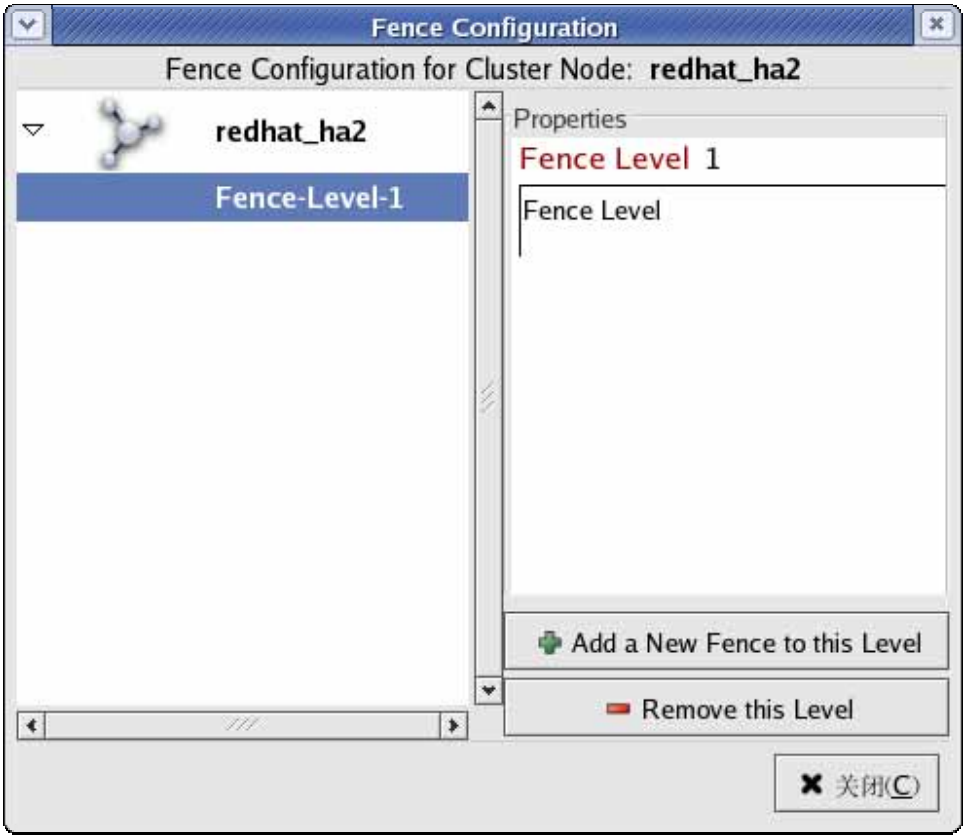
Fence Configuration
Fence Configuration for Cluster Node: redhat_ha2
redhat_ha2
Fence-Level-1
Properties
Fence Level 1
Fence Level
Add a New Fence to this Level
Remove this Level
关闭(C)

图 3-19


Fence Properties
Add a New Fence
ilo-ha2
Fence Device Type:
HP ILO Device
RSA Fence Option
(No Parameters)
ILO Fencing
(No parameters)
取消(C)
确定(O)

图 3-20

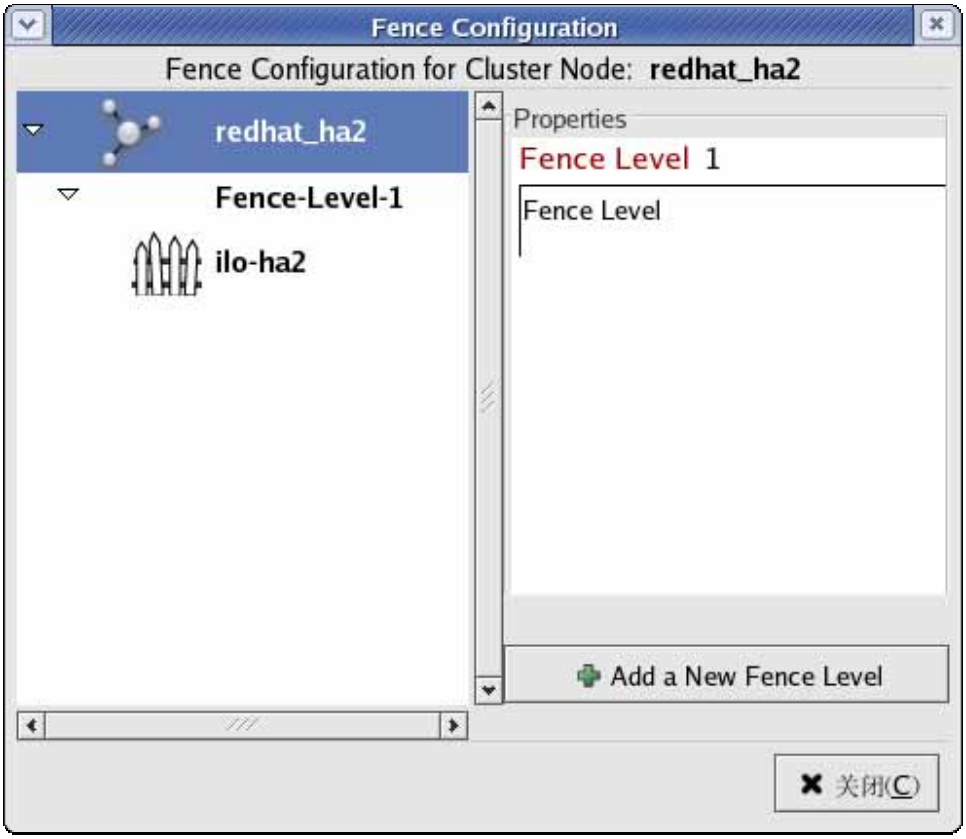

图 3-21

3.2.6创建一个失效域

失效转移域是在系统失效时具备运行服务资格的群集成员的一个带有名称的子集。失效转移域的特点如下：

无限制 — 允许你指定要优选的成员子集，但是被分派到这个域的服务可以在任何可用的成员上运行。

有限制 — 允许你限制能够运行某个特定服务的成员。如果在限制的失效转移域中没有一个可用的成员，服务就无法被启动（手工启动或被群集软件启动）。

无序 — 当服务被分派给一个无序的失效转移域，运行服务的成员就会从失效转移域成员中不按优先顺序被选择。

有序 — 允许你在失效转移域成员中指定一个优选顺序。在列表最前面的是最优先的，跟着是次一级的，依此类推。

按照默认设置，失效转移域是无限制和无序的。

在有好几个成员的群集中，使用有限制的失效转移域能够在最大程度上减少设置群集来运

行服务（如 httpd，它要求你在运行这个服务的所有成员上设置一模一样的配置）的工作量。与其设置整个群集来运行这个服务，你必须只在和服务相关的有限制的失效转移域中的成员上设置。

选择 Cluster->Managed Resources->Failover Domains->Create a Failover Domain. 在 name for new Failover Domain 中输入:redhat_fd，点击确定按钮. (见图 3-22)

图 3-22

注：Failover Domain 的名字要符合总行的规定和本支票影像系统的统一命名规范。

点击 Available Cluster Nodes,分别选中 redhatha1 和 redhatha2,然后点击关闭按钮. (见图 3—23，图 3—24)

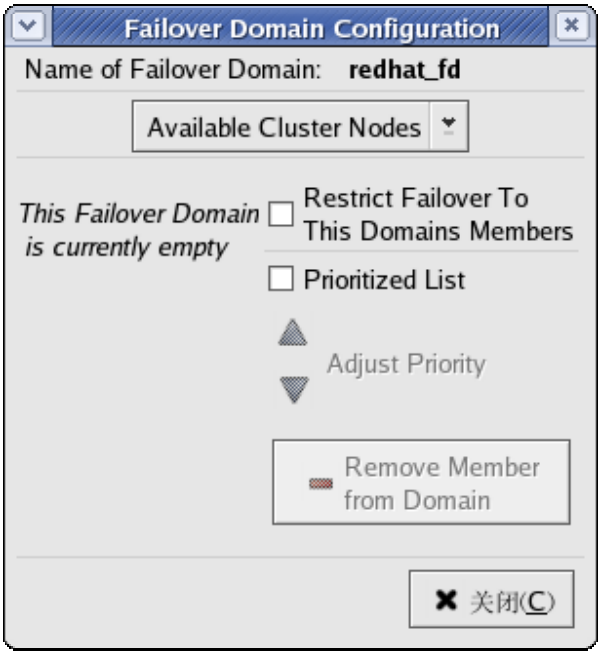

图 3-23

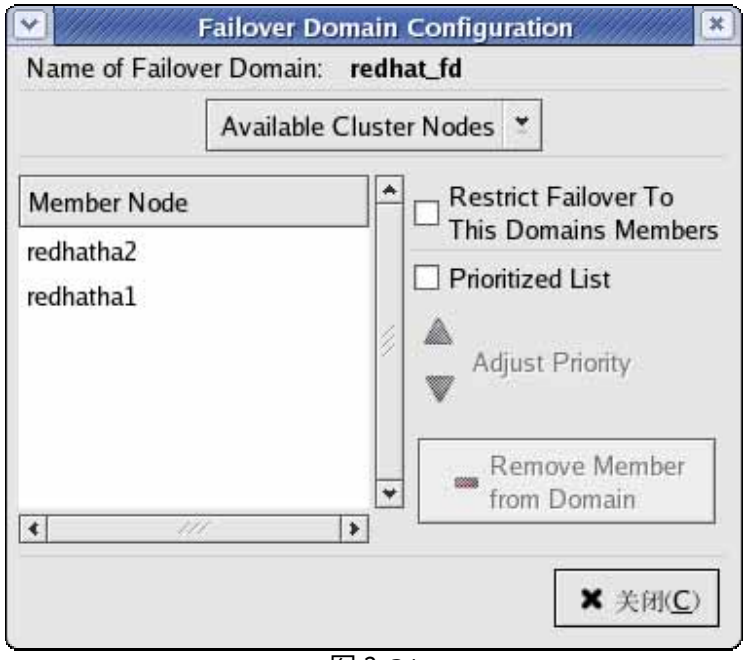

图 3-24

3.2.7创建群集资源

红帽群集资源包括下面七种资源类型：

- GFS 文件系统
- 非 GFS 文件系统(ext2,ext3)
- IP 地址
- NFS 加载
- NFS 客户端
- NFS 输出
- 服务脚本

点击 Cluster->Managed Resources->Resources->Create a Resource(见图 3—25)

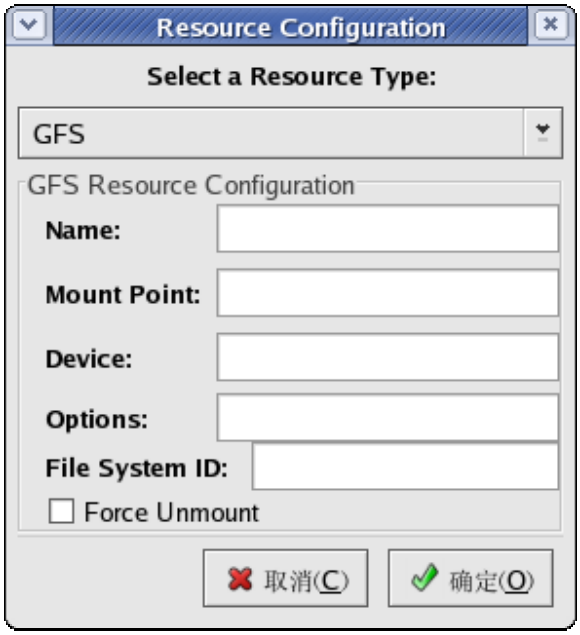

图 3-25

3.2.7.1添加一个文件系统资源

在 select a Resource Type: File System(见图 3—26)

Name:share_fs

File System Type: ext3

Mount Point:/share_fs

Device:/dev/sdb1

选中:Force unmount 点击确定按钮

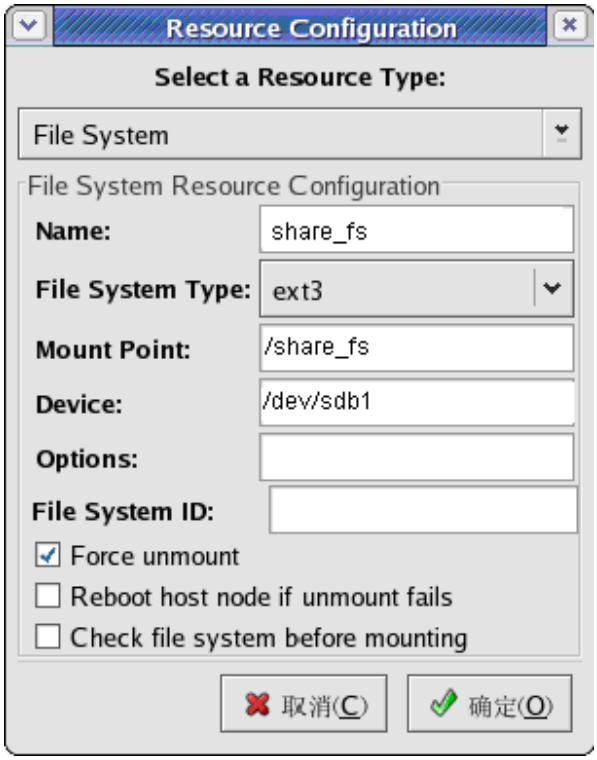

图 3-26

3.2.7.2添加一个服务 IP 地址资源

点击 Cluster->Managed Resources->Resources->Create a Resource(见图 3—27)

在 select a Resource Type: IP Address 输入：192.168.100.29，点击确定按钮。

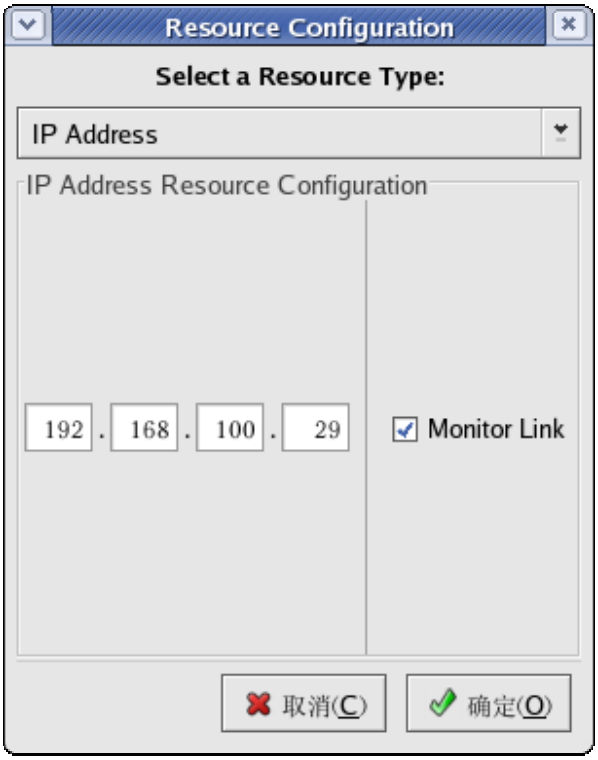

图 3-27

3.2.7.3添加一个服务控制脚本资源

点击 Cluster->Managed Resources->Resources->Creat a Resource(见图 3－28) 在 select a Resource Type: Script Name: redhat_init_script　File (with path): /etc/rc.d/init.d/httpd，　点击确定按钮。

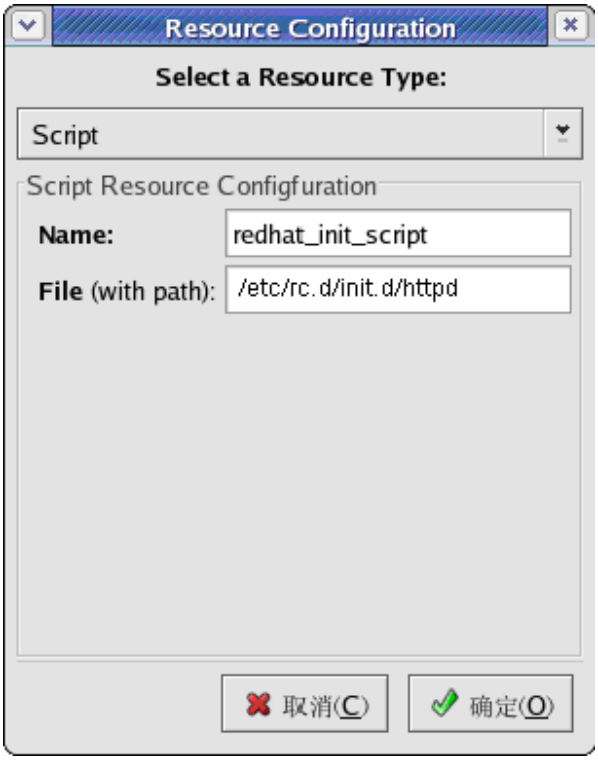

图 3-28

3.2.8 创建群集服务

在配置服务前，收集所有关于服务资源和属性的可用信息。在某些情况下，还可能为一个服务指定多个资源（例如，多个 IP 地址和磁盘设备）。

表 3-1 中描述了你可以使用群集配置工具来指定的服务属性和资源。

属性	描述
服务名称	每个服务必须有一个独特的名称。服务名称可以包含 1 到 63 个字符，必须包含字母（大小写）、数字、下划线、点和短横线（连线）的组合。服务名称必须以字母或下划线开头。
失效转移域	通过把服务关联到某个现存的失效转移域来识别要在其上运行服务的成员。当启用了有序的失效转移后，如果运行服务所在的成员失效的话，该服务就会被自动重新安置到有序成员列表中的下一个成员。（优先顺序是通过失效域列表的成员名称顺序来建立的）。

属性	描述
检查间隔	指定成员检查和服务相关的应用程序的健康状况的频繁程度（以秒为单位）。例如，当你为 NFS 或 Samba 服务指定了非零的检查间隔时，红帽群集管理器会校验必要的 NFS 或 Samba 服务是否在运行。对于其它类型的服务，红帽群集管理器会在调用了服务脚本的 status 从句之后检查其返回状态。按照默认设置，检查间隔为零就表明服务监视被禁用。
用户脚本	若适用，指定用于启动和停止该服务的脚本的完整路径名。
IP 地址	一个或多个互联网协议（IP）地址可能被分派给某个服务。这个 IP 地址（有时称为“浮动”IP 地址）和与成员的主机名以太网接口相关的 IP 地址不同，因为它在失效转移发生时会和服务一起被自动重新安置。如果客户使用这个 IP 地址来使用服务，它们就不知道哪个成员在运行该服务，失效转移是对所有客户透明的。 注意，群集成员必须在用于服务的每个 IP 地址中的 IP 子网里配置了一个网络接口卡。 每个 IP 地址的子网掩码和广播地址也可以被指定；如果没有指定，群集就会使用互连子网的网络所使用的子网掩码和广播地址。
设备特殊文件	指定用在服务中的每个共享磁盘分区。
文件系统和共享选项	如果服务使用一个文件系统，请指定文件系统的类型、挂载点、和其它挂载选项。你可以指定在 mount(8) 说明书页中描述的任何标准文件系统挂载选项。你没有必要为原始设备（若用于服务）提供挂载信息。ext2 和 ext3 文件系统在群集中被支持。 指定是否要强制卸载文件系统。强制卸载允许群集服务管理体系在重新安置或失效转移之前卸载文件系统，即使文件系统正被使用。这是通过中止任何正在存取文件系统的应用程序来做到的。 你还可以指定是否要通过 NFS 设置的存取权限来导出这个文件系统。 通过提供 Samba 共享名称来指定是否要使文件系统能够通过 Samba 而被

属性	描述
	SMB 客户存取。

3.2.8.1创建一个群集服务

点击 Cluster->Managed Resources->Services->Create a Service(见图 3－29) Name: redhat-service，然后点击确定按钮。在接下来的对话框 Failover Domain 中，选择 redhat_fd(见图 3－30)

注：Service Name 的名字要符合总行的规定和本支票影像系统的统一命名规范。

图 3-29

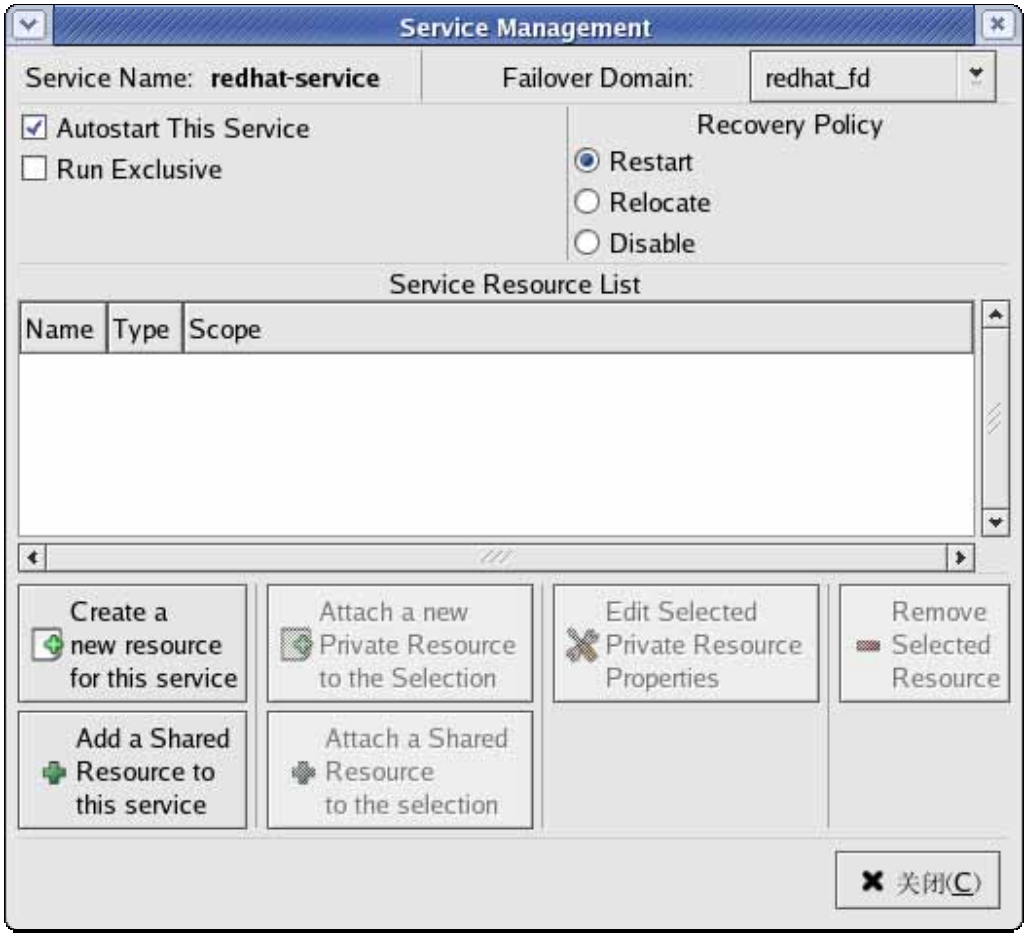

图 3-30

3.2.8.2为新创建的群集服务加入建立的资源

点击：Add a Shared Resource to this service 按钮，选择 share_fs ,然后点击确定按钮.

点击：Add a Shared Resource to this service 按钮，选择192.168.100.29,然后点击确定按钮

点击：Add a Shared Resource to this service 按钮，选择redhat_init_script ,然后点击确定按钮

然后点击关闭按钮.（见图3－31，图3－32）

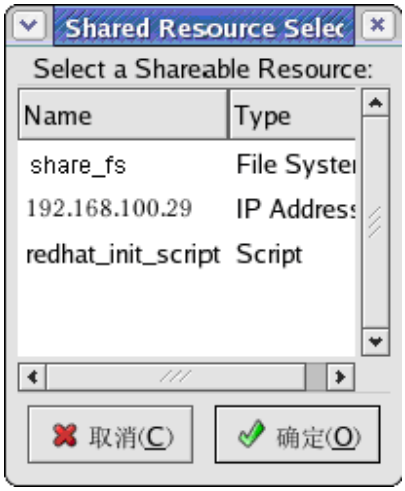
Shared Resource Selec
Select a Shareable Resource:
Name
Type
share_fs
File Syste
192.168.100.29
IP Address
redhat_init_script
Script
取消(C)
确定(O)

图3-31

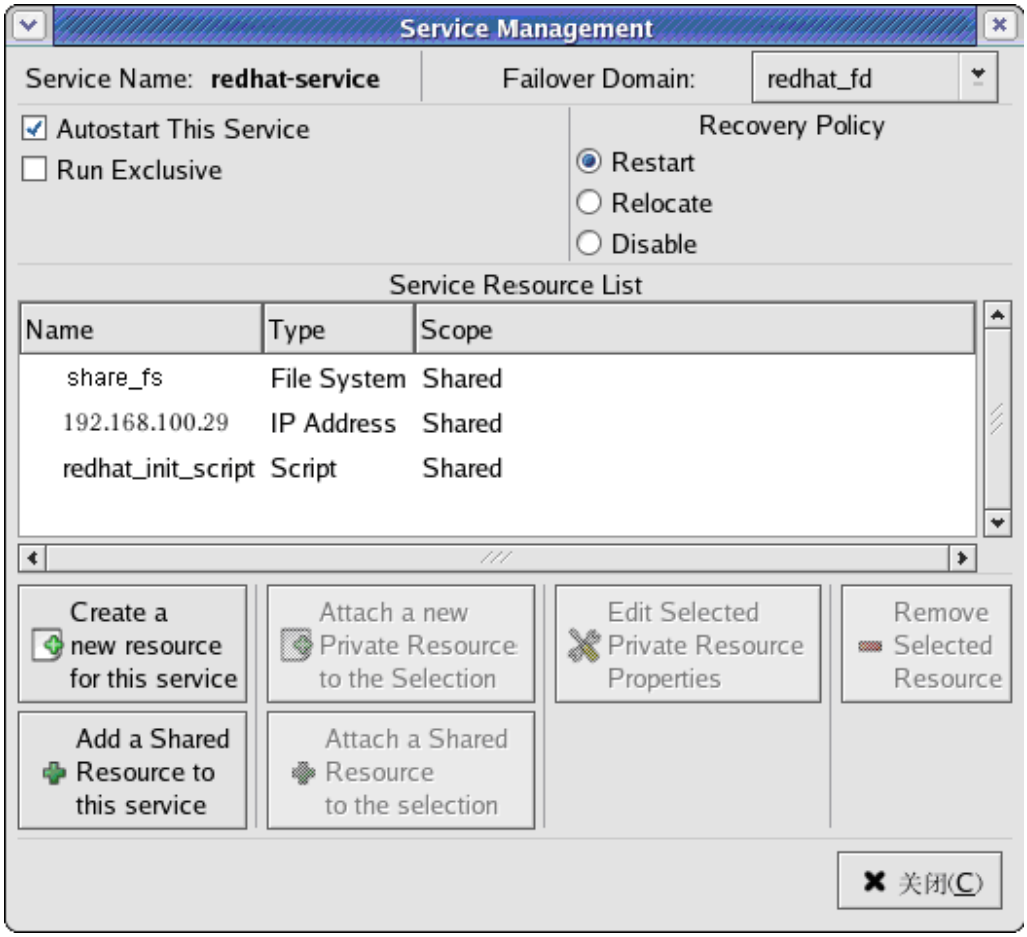
Service Management
Service Name: redhat-service
Failover Domain:
redhat_fd
Autostart This Service
Run Exclusive
Recovery Policy
Restart
Relocate
Disable
Service Resource List
Name
Type
Scope
share_fs
File System
Shared
192.168.100.29
IP Address
Shared
redhat_init_script
Script
Shared
Create a new resource for this service
Attach a new Private Resource to the Selection
Edit Selected Private Resource Properties
Remove Selected Resource
Add a Shared Resource to this service
Attach a Shared Resource to the selection
关闭(C)

图3-32

3.2.8.3 保存群集设置

点击文件->保存，使用默认的文件名和路径（见图3－33），然后点击文件－>退出。

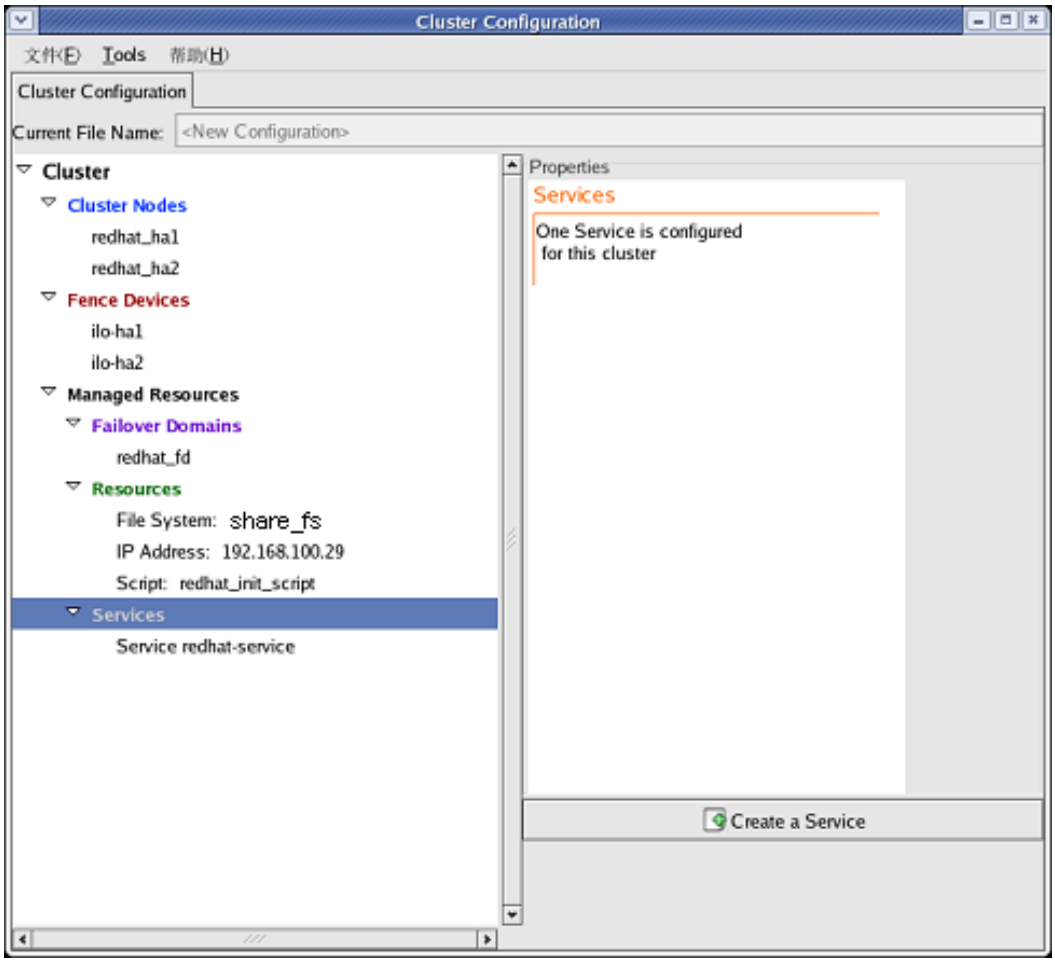

图3-33

3.2.9启动群集管理器

3.2.9.1同步节点间的群集配置信息

```
scp /etc/cluster/cluster.conf redhat_ha2:/etc/cluster/cluster.conf
```

3.2.9.2启动群集相应服务

在所有群集成员节点上分别依次启动下面的服务

```
service ccsd start
service cman start
service fenced start
service rgmanager start
```

第4章群集管理

本章描述在群集被安装和配置后所涉及的管理和维护任务。

4.1群集状态工具总览

群集状态工具显示了群集服务、群集成员、和应用程序服务的状态，以及和服务操作有关的统计数据。群集配置文件（由群集配置工具所维护）被用来决定如何管理成员、服务和群集守护进程。使用群集状态工具来启动和停止那个成员上的群集服务、重新启动应用程序服务、或把应用程序服务转移到另一个成员上。

4.2显示群集和服务状态

监视群集和应用程序服务状态能够帮助识别和解决群集环境中的问题。以下工具可以在显示群集状态方面提供帮助：

- clustat 命令
- 日志文件消息
- 群集监视 GUI

群集和服务状态包括以下信息：

- 群集成员系统状态
- 心跳频道状态
- 服务状态以及哪个群集系统在运行该服务或拥有该服务
- 监视群集系统的服务状态

使用群集状态工具来启动和停止那个成员上的群集服务、重新启动应用程序服务、或把应用程序服务转移到另一个成员上。当配置了群集服务，并相关的群集进程启动后，在 shell 提示符中，运行 system-config-cluster，点击 Cluster Management 标签，就会显示当前群集的服务状态。见图 4－1。

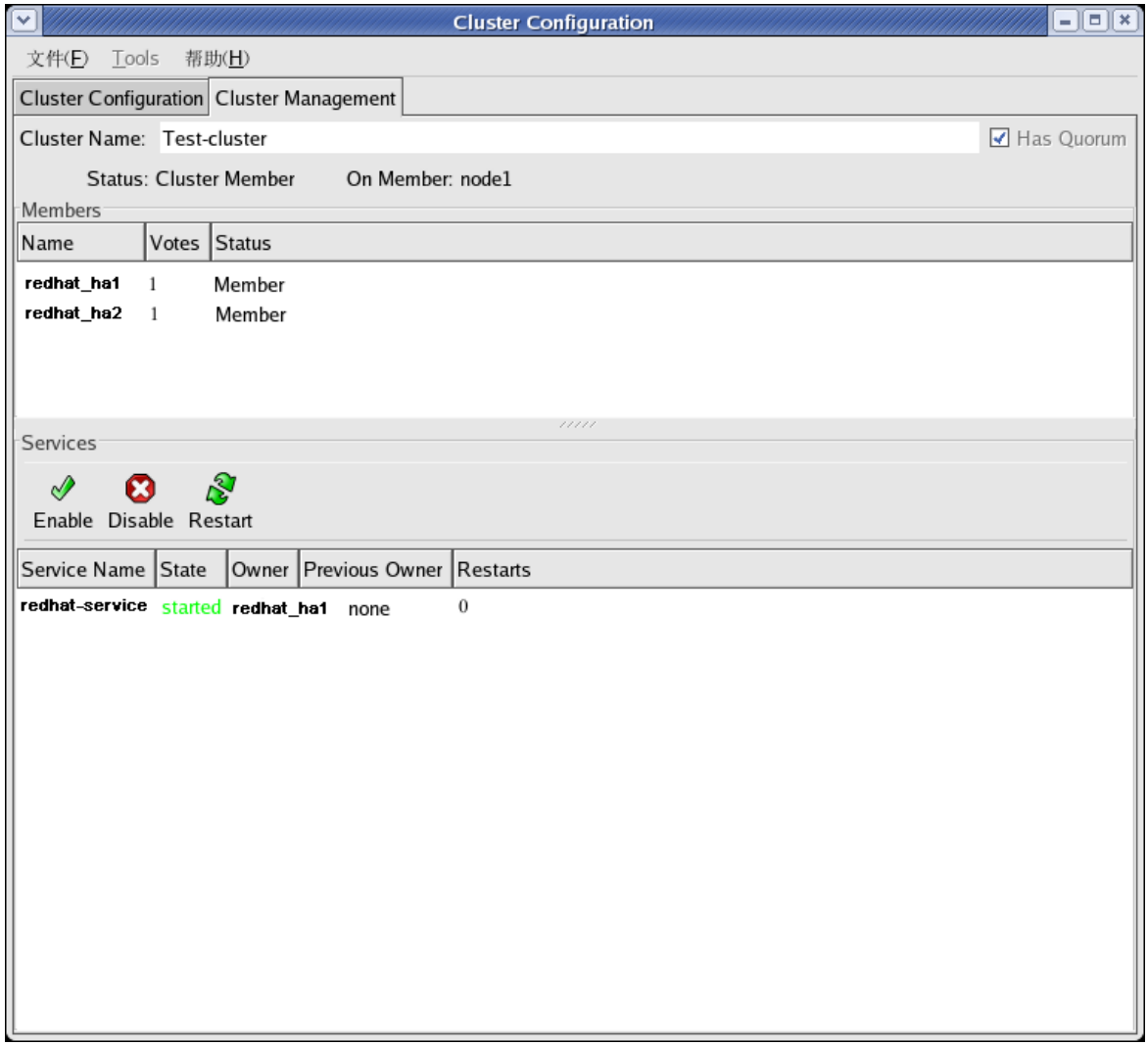

图 4－1

以下表格描述了如何分析群集状态工具和 clustat 命令中显示的状态信息。

成员状态	描述
「活跃」	成员系统在和另一个成员系统通信及存取仲裁分区。
「不活跃」	成员系统无法和另一个成员系统通信。

表 4-1. 成员状态

服务状态	描述
「运行」	服务资源在拥有它的群集系统上被配置了并可被利用。
「待用」	服务在成员上失效了，正在另一个服务上等待被启动。

服务状态	描述
「禁用」	服务被禁用了，没有被分派所有者。
「停止」	服务没有在运行；正在等待一个能够启动它的成员。
「失效」	服务没有被成功启动，而且群集无法成功的停止该服务。

表 4-2. 服务状态

要在 shell 提示下显示当前群集状态的快照，启用 clustat 工具。其示例输出如下：

```
Member Status: Quorate
Member Name Status
------ ---- ------
redhat_ha1 Online, Local ,rgmanager
redhat_ha2 Online,rgmanager
```

要从 shell 提示下在指定时间段内监视群集并显示其状态，启用带有 -i time 选项的 clustat 命令。这里的 time 指定状态快照所间隔的秒数。例如：

```
clustat -i 10
```

4.3启动和停止群集软件

4.3.1 clusvadm 工具使用

clusvcadm 工具提供了命令行用户界面，它使管理员能够监视和管理群集系统和服务。使用 clusvcadm 工具来执行以下任务：

禁用和启用服务

重新定位和重新启动群集服务

锁定和解锁服务状态

clusvcadm 命令行的选项如下：

-d service 禁用某服务。

-e service 启用某服务。

-e service -m member 启用指定成员上的某服务。

-l 锁定服务状态。

-r service -m member 把某服务重新定位到指定成员上。

-q 沉默操作。

-R 重新启动某服务。

-s 停止某服务。

-u 解锁服务状态。

-v 显示 clusvcadm 的当前版本信息。

详情请参阅 clusvcadm(8) 的说明书页。

你可以在 shell 提示下启动群集软件，只需键入：

clusvcadm -d redhat_service

如何启动服务在 redhat_ha1 上?

clusvcadm -e redhat-service -m redhat_ha1

如何把服务从 redhat_ha1 重新定向到 redhat_ha2 上?

clusvcadm -r redhat-service e -m redhat_ha2

4.3.2 通过 Service 命令进行操作

如果想启动群集服务，在所有群集成员节点上分别依次启动下面的服务。

service ccsd start

service cman start

service fenced start

servcie rgmanager start

如果想停止群集服务，在所有群集成员节点上分别依次启动下面的服务。

servcie rgmanager stop

service fenced stop

service cman stop

service ccsd stop

至此，我们完成了基于两个节点的红帽高可性的配置，管理和维护的讲解。

后记：

由于时间有限，利用出差的间隙，在飞机上完成了整个文档，难免有写笔误和不周之处。欢迎大家给出中肯的意见和批评指导。

意见和反馈：shiyingsheng@yahoo.com.cn