Sun StorageTek™ Availability Suite 4.0 软件安装和配置指南

Sun Microsystems, Inc.
www.sun.com

文件号码 819-6360-10
2006 年 6 月，修订版 A

请将有关本文档的意见和建议提交至： http://www.sun.com/hwdocs/feedback

目录

前言

《Sun StorageTek Availability Suite 4.0 软件安装和配置指南》介绍了有关如何有效地安装、设置和使用本软件的信息。

本书的结构

本书包含以下各章：

- 第 1 章提供了有关升级、安装和卸载 Availability Suite 软件的信息。
- 第 2 章介绍了初次使用 Sun StorageTek™ Availability Suite 软件之前需要执行的初始配置过程。
- 第 3 章讨论了 Remote Mirror 软件的配置问题。
- 词汇表定义了本书中使用的术语。

使用 UNIX 命令

本文档不会介绍基本的 UNIX® 命令和操作过程，如关闭系统、启动系统和配置设备等。欲获知此类信息，请参阅以下文档：

- 系统附带的软件文档
- Solaris™ 操作系统 (Solaris Operating System, Solaris OS) 的有关文档，其 URL 如下：

  `http://docs.sun.com`

Shell 提示符

Shell	提示符
C shell	*machine-name*%
C shell 超级用户	*machine-name*#
Bourne shell 和 Korn shell	$
Bourne shell 和 Korn shell 超级用户	#

印刷约定

字体*	含义	示例
AaBbCc123	命令、文件和目录的名称；计算机屏幕输出	编辑 .login 文件。 使用 ls -a 列出所有文件。 % You have mail.
AaBbCc123	用户键入的内容，与计算机屏幕输出的显示不同	% **su** Password:
AaBbCc123	保留未译的新词或术语以及要强调的词。要使用实名或值替换的命令行变量。	这些称为 *class* 选项。 要删除文件，请键入 **rm** *filename*。
新词术语强调	新词或术语以及要强调的词。	您**必须**成为超级用户才能执行此操作。
《书名》	书名	阅读 《用户指南》 的第 6 章。

* 浏览器的设置可能会与这些设置有所不同。

相关文档

应用	书名	文件号码

手册页	`sndradm`
	`iiadm`
	`dsstat`
	`kstat`
	`svadm`
	`dscfgadm`
		无

系统管理	《Sun StorageTek Availability Suite 4.0 Point-in-Time Copy 软件管理指南》	819-6370
	《Sun StorageTek Availability Suite 4.0 Remote Mirror 软件管理指南》	819-6365
集成	《Sun Cluster 和 Sun StorageTek Availability Suite 4.0 软件集成指南》	819-6375
故障排除	《Sun StorageTek Availability Suite 4.0 软件故障排除指南》	819-6380
发行说明	《Sun StorageTek Availability Suite 4.0 软件发行说明》	819-6385

访问 Sun 文档

您可以查看、打印或购买内容广泛的 Sun 文档，包括本地化版本，其网址如下：

`http://www.sun.com/documentation`

第三方 Web 站点

Sun 对本文档中提到的第三方 Web 站点的可用性不承担任何责任。对于此类站点或资源中的（或通过它们获得的）任何内容、广告、产品或其他资料， Sun 并不表示认可，也不承担任何责任。对于因使用或依靠此类站点或资源中的（或通过它们获得的）任何内容、产品或服务而造成的或连带产生的实际或名义损坏或损失， Sun 概不负责，也不承担任何责任。

联系 Sun 技术支持

如果您遇到通过本文档无法解决的技术问题，请访问以下网址：

`http://www.sun.com/service/contacting`

Sun 欢迎您提出意见

Sun 致力于提高其文档的质量，并十分乐意收到您的意见和建议。您可以通过以下网址提交您的意见和建议：

`http://www.sun.com/hwdocs/feedback`

请在您的反馈信息中包含文档的书名和文件号码：

《Sun StorageTek Availability Suite 4.0 软件安装和配置指南》，文件号码 819-6360-10

第 1 章

升级、安装和卸载 Availability Suite 软件

本章提供了在 Sun Solaris 10 操作环境及其后续更新发行版中升级、安装和卸载 Availability Suite (AVS) 软件的信息。

注 – 本章仅适用于未绑定在 Solaris 操作环境 (Operating Environment, OE) 中的 Availability Suite 版本。

本章讨论以下主题：

升级 Availability Suite 软件

AVS 4.0 仅支持从运行于 Solaris 8 或 9 的 AVS 3.2 进行升级。由于 AVS 3.2 不会运行于 Solaris 10 上，而 AVS 4.0 仅运行于 Solaris 10 上，因此从 AVS 3.2 到 AVS 4.0 的系统升级要求您首先将 Solaris 操作环境升级到 Solaris 10。

如果成功地将 Solaris OE 升级到 Solaris 10，则在 Solaris 10 安装了新 AVS 软件包之后，将会自动启用所有以前在 AVS 3.2 控制下的卷。

▼ 从 AVS 3.2 升级

要从 AVS 3.2 升级，请执行以下步骤：

1. **作为预防措施，请将 `dscfg` 数据库中包含的信息保存至远程位置中的某个文件：**

```
# dscfg -l > remote-node:/backup/database-file
```

此备份文件包含在 AVS 3.2 控制下的卷的列表，如果 Solaris OE 升级（下面的步骤 3）失败，则可以使用这些卷手动重构 AVS 集。例如，如果控制器编号在升级 Solaris 之后发生改变，则可能需要使用更新后的控制器编号来重新配置使用原始分片的 AVS 集。

2. **根据第 3 页 “卸载 Availability Suite 软件”中的指导，删除 AVS 3.2 软件包。**

3. **将操作环境升级到 Solaris 10 OE 或更高版本。**

运行 Solaris 8 的系统可以直接升级到 Solaris 10，而不必执行到 Solaris 9 的中间升级。请注意，在系统上全新安装 Solaris 10 OE 或更高版本不会被视为升级。

4. **根据第 2 页 “安装 Availability Suite 软件”中的指导，安装新的 AVS 软件包。**

安装 Availability Suite 软件

本节提供了安装 AVS 4.0 软件的信息。

▼ 安装 AVS 4.0

要安装 AVS 4.0，请执行以下步骤：

1. **如果计划在 Sun Cluster OE 中运行 AVS，则建议您在安装 AVS 之前首先安装 Sun Cluster OE。如果选择在已安装了 AVS 的系统上安装 Sun Cluster OE，则无需卸载 AVS。**

2. **安装 Sun Cluster OE 之后，运行 `dscfgadm` 来选择特定于 Sun Cluster 的配置位置。**

3. 按照以下顺序使用 `pkgadd(1M)` 来安装新 **AVS** 软件包：

   ```
   SUNWscmr
   SUNWscmu
   SUNWspsvr
   SUNWspsvu
   SUNWiir
   SUNWiiu
   SUNWrdcr
   SUNWrdcu
   ```

4. 首次使用 **AVS** 之前，请先执行第 5 页 “初始配置步骤”（第 2 章）中的步骤。

注 – 安装 Availability Suite 软件时会在根目录中创建文件 `reconfiguration`；然而，无需对 Solaris 执行旨在重新配置的重新引导便可以使用 Availability Suite 软件。

卸载 Availability Suite 软件

本节提供了卸载 AVS 3.2 软件的信息。

▼ 卸载 AVS 3.2

要卸载 AVS 3.2，请执行以下步骤：

1. 对于通过 **AVS** 启用的卷，请停止所有应用程序对这些卷执行的写操作。
2. 按照以下顺序使用 `pkgrm(1M)` 来卸载 **AVS** 软件包：

   ```
   SUNWrdcu
   SUNWrdcr
   SUNWiiu
   SUNWiir
   SUNWspsvu
   SUNWspsvr
   SUNWscmu
   SUNWscmr
   ```

第 2 章

初始配置步骤

在安装 Sun StorageTek Availability Suite 软件之后和首次使用它**之前**，您必须先对 Point-in-Time Copy 软件和 Remote Mirror 软件的某些文件进行配置。本章介绍必需的初始配置步骤：

本章还介绍了以下主题，供您参阅：

初始配置步骤概述

表 2-1 概述了必需的和可选的初始配置任务。

表 2-1 Availability Suite 软件的初始配置摘要

任务	说明
1. 配置以下文件： • `/etc/hosts` • IP 堆栈 （IPv4 和 IPv6） • （可选） `/etc/services` • `/etc/nsswitch.conf` • （可选)`/usr/kernel/drv/rdc.conf`	第 6 页 “配置系统文件”
2.（可选）调整为软件配置使用的默认卷数。	第 11 页 “修改设置”
3.（可选）调整异步队列。	《Sun StorageTek Availability Suite 4.0 Remote Mirror 软件管理指南》
4. 选择位图卷。	第 14 页 “使用位图卷”
5.（可选）建立一个可选的远程镜像卷配置文件。	第 16 页 “使用卷集文件”

配置系统文件

本节介绍如何编辑和检查以下系统文件，以保证 Sun StorageTek Remote Mirror 软件正常运行：

- 第 6 页 “编辑 `/etc/hosts` 文件”
- 第 7 页 “设置 IPv6 地址”
- 第 10 页 “编辑 `/etc/services` 文件”
- 第 10 页 “编辑 `/etc/nsswitch.conf` 文件”
- 第 17 页 “备份配置信息”

▼ 编辑 `/etc/hosts` 文件

此步骤可确保运行 Remote Mirror 软件的计算机能够读取和识别 `/etc/hosts` 文件中的主机名。

- **将即将使用 Remote Mirror 软件的所有计算机的名称和 IP 地址添加到 `/etc/hosts` 文件。**

  在每台要安装和运行 Remote Mirror 软件的计算机上编辑此文件。

配置 IP 堆栈 （IPv4 和 IPv6）

如果使用 Internet 协议版本 6 (Internet Protocol version 6, IPv6) 传输协议进行复制，请在使用 Remote Mirror 软件的主机上为接口同时配置 IPv4 和 IPv6 堆栈。 IPv6 协议提供了更强的可寻址功能。对于 Solaris 10 OS，请参见 《System Administration Guide: IP Services》以获取有关 IPv6 的更多信息。

要使用 IPv6 协议，请将 IPv4 和 IPv6 接口定义为相同的名称。必须对主和辅助主机进行定义，以使两台计算机使用相同的传输协议。

▼ 设置 IPv6 地址

此示例过程显示了如何设置网络接口以使用 IPv6 地址。使用此过程可测试远程镜像主机的连接情况。以下过程假定您使用如下配置信息：

网络接口	`hme1`
主主机接口名	`sndrpri`
辅助主机接口名	`sndrsec`

1. **使用文本编辑器在主主机和辅助主机上创建 `/etc/hostname6.hme1` 文件。在主主机上，将接口名称 `sndrpri` 添加到该文件。在辅助主机上，将接口名称 `sndrsec` 添加到该文件。保存并关闭这两个文件。**

```
primary-host# more /etc/hostname6.hme1
sndrpri
secondary-host# more /etc/hostname6.hme1
sndrsec
```

2. **关机并重新启动主主机和辅助主机以激活 IPv6。**

```
# /etc/shutdown -y -i 6 -g 0
```

3. **两台计算机重新引导后，获取 `hme1` 接口地址的 IPv6 `inet` 地址。**

4. 以下的示例中，此地址为 fe80::a00:20ff:febd:c33f/128。

```
# ifconfig -a
lo0: flags=1000849<UP,LOOPBACK,RUNNING,MULTICAST,IPv4> mtu 8232 index 2
        inet 127.0.0.1 netmask ff000000
hme0: flags=1000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4> mtu 1500 index 3
        inet 192.9.200.125 netmask ffffff00 broadcast 192.9.200.255
        ether 8:0:20:ae:85:fa
lo0: flags=2000849<UP,LOOPBACK,RUNNING,MULTICAST,IPv6> mtu 8252 index 2
        inet6 ::1/128
hme0: flags=2000841<UP,RUNNING,MULTICAST,IPv6> mtu 1500 index 3
        ether 8:0:20:ae:85:fa
        inet6 fe80::a00:20ff:feae:85fa/10
hme1: flags=2000841<UP,RUNNING,MULTICAST,IPv6> mtu 1500 index 4
              ether 8:0:20:bd:c3:3f
        inet6 fe80::a00:20ff:febd:c33f/128
```

5. 编辑 /etc/inet/ipnodes 文件，添加在步骤 4 中获得的 **inet** 地址，将主主机地址指定给 sndrpri，将辅助主机地址指定给 sndrsec。请勿包含地址中的 /128 部分。

注 – 确保所有运行 Remote Mirror 软件的系统上的 /etc/inet/ipnodes 文件均包含每个系统的 IPv6 inet 地址和名称。

---

6. 保存并关闭文件，然后检查文件内容。

   以下示例中， sndrsec 是辅助主机接口名称。

```
primary-host# more /etc/inet/ipnodes
#
# Internet host table
#
::1             localhost
127.0.0.1       localhost
fe80::a00:20ff:febd:c33f                    sndrpri
fe80::a00:20ff:fee1:195e                    sndrsec
```

7. 编辑 /etc/nsswitch.conf 文件以确保 ipnodes: 指向 files。

   在此文件中查找以下文本并确保 ipnodes: 行不是注释行。

```
# consult /etc "files" only if nis is down.
hosts: files nis [NOTFOUND=return] files
ipnodes: files
```

8. **对即将使用 Remote Mirror 软件的所有计算机，将其主机名和 IPv6 `inet` 主地址添加到每台计算机的 `/etc/hosts` 文件。**

   在每台要安装和运行 Remote Mirror 软件的计算机上编辑此文件。

注 – 如果未完成此步骤（如第 6 页 “编辑 `/etc/hosts` 文件”所述），则在启用 Remote Mirror 软件时会显示以下错误消息：

```
sndradm: Error: neither sndrpri nor sndrsec is local
```

---

9. **确保可以从一个系统 ping 到另一个系统，并且这些系统使用的均是 IPv6 协议。**

   要从主主机发出 ping 指令，请输入以下内容：

```
# ping -s sndrsec
PING sndrsec: 56 data bytes
64 bytes from sndrsec (fe80::a00:20ff:fee1:195e): icmp_seq=0. time=0. ms
64 bytes from sndrsec (fe80::a00:20ff:fee1:195e): icmp_seq=1. time=0. ms
64 bytes from sndrsec (fe80::a00:20ff:fee1:195e): icmp_seq=2. time=0. ms
```

   要从辅助主机发出 ping 指令，请输入以下内容：

```
# ping -s sndrpri
PING sndrpri: 56 data bytes
64 bytes from sndrpri (fe80::a00:20ff:febd:c33f): icmp_seq=0. time=0. ms
64 bytes from sndrpri (fe80::a00:20ff:febd:c33f): icmp_seq=1. time=0. ms
64 bytes from sndrpri (fe80::a00:20ff:febd:c33f): icmp_seq=2. time=0. ms
```

10. **使用 `netstat`(1M) 命令检验接口是否具有正确的 IPv6 地址和 IPv6 名称。**

    在 `sndrpri` 和 `sndrsec` 主机上都使用此命令。例如：

```
# netstat -in
Name  Mtu  Net/Dest      Address        Ipkts  Ierrs Opkts  Oerrs Collis Queue
lo0   8232 127.0.0.0     127.0.0.1      3844   0     3844   0     0      0
hme0  1500 192.0.0.0                       192.9.200.225 22007  0    1054   0
0      0

Name  Mtu  Net/Dest      Address        Ipkts  Ierrs Opkts  Oerrs Collis Queue
lo0   8252 ::1                         ::1                         3844   0
3844   0     0
hme1  1500 fe80::a00:20ff:febd:c33f   fe80::a00:20ff:febd:c33f    43     0 65 0 0
```

```
# netstat -i
Name  Mtu  Net/Dest      Address        Ipkts  Ierrs Opkts  Oerrs Collis Queue
lo0   8232 loopback      localhost      3844   0     3844   0     0      0
hme0  1500 arpanet                         rick1                  22038  0
1067   0     0       0

Name  Mtu  Net/Dest                 Address                      Ipkts  Ierrs
Opkts  Oerrs Collis
lo0   8252 localhost localhost                    3844   0     3844   0     0
hme1  1500 sndrpri                   sndrpri                   43     0     65
0     0
```

▼ 编辑 /etc/services 文件

端口 121 是供 Remote Mirror rdc 守护进程使用的默认端口。

```
# cat /etc/services
...
rdc 121/tcp # SNDR server daemon
...
```

如果您更改了该端口号，则必须在此配置集内的所有远程镜像主机（即，主主机和辅助主机以及一对多、多对一和多中继配置中的所有主机）上进行同样的更改。

1. **在每台运行 Remote Mirror 软件的计算机上编辑 /etc/services 文件。**
2. **关闭并重新启动所有主机，以使新的端口号生效。**

▼ 编辑 /etc/nsswitch.conf 文件

如果此文件包括 hosts: 和 services: 条目，请检验 files 是否置于 nis、nisplus、ldap、dns 或计算机使用的其他任何服务之前。例如，对于使用 NIS 命名服务的系统，文件应包含以下行：

```
hosts: files nis
services: files nis
```

- **如果主机和服务条目不正确，请编辑该文件并将其保存。**

  如果您使用的是 IPv6 协议，请参见第 7 页 “配置 IP 堆栈（IPv4 和 IPv6）”以了解对此文件进行的更改。

修改设置

以下各节介绍了如何修改 Remote Mirror 软件的设置。

注 – 编辑此节中的文件后，为使更改生效，请使用 `shutdown` 命令关闭服务器并重新启动。如果要编辑 `rdc.conf` 文件以使用 64 个以上的卷集，请确保您具有足够的系统资源 （如很大的交换空间）。

设置位图操作模式

根据 `/usr/kernel/drv/rdc.conf` 中的 `rdc_bitmap_mode` 设置，磁盘上存储的位图在系统崩溃后仍可保留下来。默认设置为 `1` （强制位图记录每一次的写操作内容）。

- **打开 `rdc.conf` 文件并定位到以下部分。编辑位图模式的值，保存并关闭文件。**

```
#
# rdc_bitmap_mode
# - Sets the mode of the RDC bitmap operation, acceptable values are:
#   0 - autodetect bitmap mode depending on the state of SDBC (default).
#   1 - force bitmap writes for every write operation, so an update resync
#       can be performed after a crash or reboot.
#   2 - only write the bitmap on shutdown, so a full resync is
#       required after a crash, but an update resync is required after
#       a reboot.
#
rdc_bitmap_mode=1;
```

增加卷集的数目

默认的已配置卷集的数目为 64。要配置 64 个以上的卷集，请在每台运行 Remote Mirror 软件的计算机上编辑 `/usr/kernel/drv/rdc.conf` 文件中的 `rdc_max_sets` 字段。

- **打开 `rdc.conf` 文件并定位到以下部分。编辑卷集的值，保存并关闭文件。**

例如，要使用 128 个卷集，请按照以下所示更改此文件：

```
#
# rdc_max_sets
# - Configure the maximum number of RDC sets that can be enabled on
# this host.The actual maximum number of sets that can be enabled
#   will be the minimum of this value and nsc_max_devices (see
#   nsctl.conf) at the time the rdc kernel module is loaded.
#
rdc_max_sets=128;
```

增加存储卷设备的限额

Availability Suite 软件的默认存储卷限额为 4096。存储卷驱动器设备 （即卷）的默认值由 `nsctl.conf` 文件中的 `nsc_max_devices` 值设置。

Remote Mirror 软件和 Point-in-Time Copy 软件分摊使用这些卷数。例如，如果仅使用 Point-in-Time Copy 软件，则可以拥有 1365 个卷集，每个卷集均包含主卷、阴影卷和位图卷。如果同时使用 Remote Mirror 和 Point-in-Time Copy 软件包，则卷集数目将在这两个软件包之间分摊。

修改此限额或许会对某些安装有益。必要时，有足够内存的站点可以增加此限额以启用更多的存储器卷。而对于可用内存有限的站点，则可以降低此限额以释放系统资源。

- **打开 `nsctl.conf` 文件，然后定位到 `nsc_max_devices` 字段。编辑该字段的值，保存并关闭文件。**

使用 dscfgadm 初始化配置数据库和启动服务

Availability Suite 软件的启动和关闭是通过 Service Management Facility (SMF) 服务实现的，可使用 `dscfgadm` 实用程序管理该服务。

```
# svcs |  grep nws_
online         Mar_14   svc:/system/nws_scm:default
online         Mar_14   svc:/system/nws_sv:default
online         Mar_14   svc:/system/nws_ii:default
online         Mar_14   svc:/system/nws_rdc:default
online         Mar_14   svc:/system/nws_rdcsyncd:default
```

dscfgadm 实用程序

dscfgadm 提供的功能可用来设置配置位置以及启用和禁用 Availability Suite 服务，从而控制 Availability Suite 配置服务。

```
# dscfgadm [-x]
```

用法

```
dscfgadm [-x ]
dscfgadm [-x ] -i
dscfgadm [-x ] -e [-r] [-p]
dscfgadm [-x ] -d [-r]
```

选项

-i 用于显示有关 Availability Suite 服务的信息

-e 用于启用 Availability Suite SMF 服务 （默认情况下启用所有服务）

-d 用于禁用 Availability Suite SMF 服务 （默认情况下禁用所有服务）

-r 用于启用/禁用 Remote Mirror 软件

-p 用于启用 Point-in-Time Copy 软件

-x 用于显示详细的调试信息

初始化配置数据库和启动服务

默认情况下，未启动 Availability Suite 服务，且系统中不存在 Availability Suite 配置数据库。在执行 dscfgadm 时，如果不带任何选项 （或仅带 -x 选项），则它将在交互模式下运行。在此模式下，您可以初始化 Availability Suite 软件所需的本地配置数据库，还可以选择是否在此时启动 Availability Suite 服务。

如果您选择在数据库初始化期间不启动 Availability Suite SMF 服务，则可在稍后使用 dscfgadm -e 命令来启动 SMF 服务。

```
# dscfgadm -e
```

注 – 只有启动了 Availability Suite 服务，才能使用 Availability Suite 软件。

启用或禁用服务

以后要启用或禁用服务，请使用带 `-e` 和 `-d` 选项的 `dscfgadm` 命令。

```
# dscfgadm -e
```

```
# dscfgadm -d
```

默认的操作可作用于所有的服务，但 `-r` 和 `-p` 选项只能分别禁用或启用 Remote Mirror 或 Point-in-Time Copy 服务。

注 – 在禁用某一服务之前，请确保停止应用程序对该服务所使用的卷的所有写入操作。

注 – 这些设置在系统引导后仍保持不变。

使用位图卷

Point-in-Time Copy 软件和 Remote Mirror 软件都使用原始卷来存储位图。不支持位图文件。

建议的位图卷位置

对于 Point-In-Time Copy 软件，请将位图原始卷存储在不包含其主卷和阴影卷的其他磁盘上；而对于 Remote Mirror 软件，请将位图原始卷存储在不包含复制卷的其他磁盘上。为这些位图卷配置 RAID（例如镜像分区），并确保没有将镜像成员与主卷和阴影卷或复制卷存储在同一个磁盘上。

在群集环境下使用 Point-In-Time Copy 软件时，位图卷与其对应的主卷或阴影卷必须位于同一磁盘组或群集资源组。

位图卷的大小要求

位图卷的大小基于主卷的大小和所创建卷集的类型 （独立卷集、从属卷集或压缩从属卷集）。

- 独立或从属阴影卷集要求：

  每 1 GB 主卷 （四舍五入到最接近的整 GB 数）需 8 KB 的位图卷，另加用于系统开销的 24 KB。

  例如，要对一个 3 GB 主卷进行阴影操作，位图卷大小必须为 (3 × 8 KB) + 24 KB，即 48 KB。 50 GB 的主卷需要 424 KB 的位图卷。

- 压缩从属阴影卷集要求：

  每 1 GB 主卷 （四舍五入到最接近的整 GB 数）需 264 KB 位图卷，另加用于系统开销的 24 KB。

  例如，要对一个 3 GB 主卷进行阴影操作，位图大小必须为 (3 × 264 KB + 24 KB，即 816 KB。压缩从属阴影卷集中 50 GB 的主卷需要 13224 KB 的位图卷。

如果启用一个位图卷过大的阴影卷集，则即使浪费空间也会创建此阴影卷集。如果启用一个位图过小的阴影卷集，则启用命令会失败并返回一条错误消息。

Availability Suite 软件提供的 `dsbitmap` 实用程序可以计算 Point-in-Time Copy 阴影卷集或 Remote Mirror 卷集所需的位图卷大小。

1. **要获取 Point-in-Time Copy 位图卷的大小，请使用此命令：**

```
dsbitmap -p data-volume [bitmap-volume]
```

2. **要获取 Remote Mirror 位图卷的大小，请使用此命令：**

```
dsbitmap -r data-volume [bitmap-volume]
```

有关 `dsbitmap` 实用程序的更多信息，请参阅 `dsbitmap(1SCM)` 手册页。

使用卷集文件

启用 Remote Mirror 软件时，您可以指定一个可选的**卷集文件**，它包含有关该卷集的信息：卷、主主机和辅助主机、位图卷、操作模式等等。使用卷集文件时，可使用 `sndradm -f` *volset-file* 选项。

也可从命令行输入关于每个卷集的信息；但如果有多个卷集时，将这些信息放置在一个文件中会比较方便。另一个优点是您可以针对特定的卷集进行操作，而将其他的卷集排除在外。与将卷集添加到 I/O 组不同，您可以在一个卷集文件中混合使用不同的复制模式。指定卷集文件的字段如下所示：

phost pdev pbitmap shost sdev sbitmap `ip {sync|async} [g` *io-groupname*`] [C` *tag*`] [q` *qdev*`]`

表 2-2 对这些字段进行了介绍。有关卷集文件格式的更多信息，请参见 `rdc.cf` 手册页。

以下是文件条目的示例：

```
atm10 /dev/vx/rdsk/oracle816/oratest /dev/vx/rdsk/oracle816/oratest_bm \
atm20 /dev/vx/rdsk/oracle816/oratest /dev/vx/rdsk/oracle816/oratest_bm \
ip sync g oragroup
```

表 2-2 卷集文件的字段

字段	含义	说明
phost	主主机	主卷所在的服务器。
pdev	主设备	主卷分区。只能指定完整路径名 （例如， `/dev/rdsk/c0t1d0s4`）。
pbitmap	主位图	卷分区 （存储主分区的位图的分区）。必须指定完整路径名。
shost	辅助主机	辅助卷所在的服务器。
sdev	辅助设备	辅助卷分区。必须指定完整路径名。
sbitmap	辅助位图	卷分区 （存储辅助分区的位图的分区）。必须指定完整路径名。
`ip`	网络传输协议	指定 `ip`。
`sync` | `async`	操作模式	• `sync` （同步）模式下，当远程卷更新结束后，才确认 I/O 操作已完成。 • `async` （异步）模式下，在更新远程卷之前即确认主主机的 I/O 操作已完成。

表 **2-2** 卷集文件的字段 （续）

g *io-groupname*	I/O 组名	可使用字符 g 指定的 I/O 组名。在本例中，组名为 oragroup。
C *tag*	群集标记	该标记可将操作限定到属于该群集资源组的远程镜像集。
q *qdev*	磁盘队列卷	在异步集或组内作为基于磁盘的 I/O 队列使用的卷。您必须指定完整的路径名称。例如：/dev/rdsk/c1t2d0s6。

备份配置信息

必须定期备份 Sun StorageTek、 VERITAS Volume Manager 和 Solaris Volume Manager 的配置信息。要进行任何与卷集相关的更改，请使用 /usr/sbin/iiadm 命令 （如 《Sun StorageTek Availability Suite 4.0 Point-in-Time Copy 软件管理指南》中所述），并考虑：

- 将这些备份命令放置到一个 shell 脚本中，并将此脚本作为每日执行的 cron(1M) 作业的一部分运行。
- 将命令输出保存到一个可定期备份到磁带的位置。

您可能希望建立一个 cron(1M) 作业，以便定期对配置数据库进行自动备份。另外，请在每次更改配置 （如添加或删除卷）之后备份配置信息。

▼ 备份配置信息

- **将配置数据库 (/etc/dscfg_local) 复制到一个安全位置。**

```
# cp /etc/dscfg_local /var/backups/dscfg_db
```

第 3 章

配置 Remote Mirror 软件

Sun StorageTek Availability Suite Remote Mirror 软件是用于 Solaris 10 （Update 1 及更高版本）操作系统的卷级**复制**工具。 Remote Mirror 软件在物理上相互独立的**主**和**辅助**站点间实时地复制磁盘卷的写操作。Remote Mirror 软件可以与任何支持 TCP/IP 的 Sun 网络适配器和网络链路一起使用。

由于此软件是基于卷的，因此它与存储器无关，并且支持 Sun 及第三方产品的原始卷或任何卷管理器。此外，该产品还支持那些通过单台主机 （运行 Solaris OS）来写入数据的所有应用程序或数据库。对于配置为允许多台运行 Solaris OS 的主机向一个共享卷写入数据的数据库、应用程序或文件系统 （例如：Oracle® 9iRAC、 Oracle® Parallel Server），此产品不提供支持。

作为灾难恢复和业务连续性规划的一部分， Remote Mirror 软件可以在远程站点保存重要数据的最新副本。通过 Remote Mirror 软件，您可以对业务连续性规划进行预演或测试。对于高可用性解决方案，可以配置 Sun StorageTek Availability Suite 软件以在 Sun™ Cluster 3.*x* 环境中进行故障转移。

当应用程序访问数据卷、连续向远程站点复制数据或记录更改 （以便用于日后快速重新同步）时， Remote Mirror 软件处于活动状态。

Remote Mirror 软件既允许从主站点到辅助站点 （通常称为**正向同步**），也允许从辅助站点到主站点 （通常称为**反向同步**）手动初始化重新同步。

Remote Mirror 软件中的复制和配置是基于卷集来完成的。远程镜像集包括主站点和辅助站点 （用于跟踪和记录更改以实现快速重新同步）上的主卷、辅助卷和位图卷，以及用于**异步复制**模式的可选的**异步队列**卷。建议将主卷和辅助卷设为相同大小。您可以使用 `dsbitmap` 工具来确定所需的位图卷大小。有关配置远程镜像集或 `dsbitmap` 工具的更多信息，请参见 《Sun StorageTek Availability Suite 4.0 Remote Mirror 软件管理指南》。

复制

复制既可以同步进行，也可以异步进行。在同步模式下，只有主主机和辅助主机都确认了应用程序的写操作，此写操作才会得到确认。在异步模式下，只要应用程序的写操作得到本地存储的确认并写入异步队列，写操作即得到确认。异步队列将以异步方式将写操作推至辅助站点。

同步复制

同步操作的数据流如下：

1. 在位图卷中设置记录位。
2. 并行初始化本地写操作和网络写操作。
3. 两项写操作完成后，清除记录位 （**惰性清除**）。
4. 写操作得到应用程序的确认。

同步复制的优点在于主站点和辅助站点始终是同步的。此复制类型只有当链路的等待时间很少，并且链路能够满足应用程序所需带宽时才可行。这些限制通常会将同步解决方案局限于校园内或大城市中。

在这种情况下，一个写入操作的平均服务时间为：

位图写操作 + MAX （本地数据写操作，网络传输往返时间 + 远程数据写操作）

在校园内和大城市中，网络传输往返时间可以忽略，因此服务时间大约是未安装 Remote Mirror 软件时所观测到的时间的两倍。

假定一个写操作需要 5 毫秒，则：

5 毫秒 + MAX （5 毫秒， 1 毫秒 + 5 毫秒） = 11 毫秒

注 – 在轻负荷的系统上， 5 毫秒是一个合理的假定值。在更符合实际情况的负荷系统上，排队等待累积会使该值增大。

不过，如果网络传输往返时间达到大约 50 毫秒 （这对于远距离复制来说很平常），则网络等待时间会使同步复制解决方案变得不切实际，如下例所示：

5 毫秒 + MAX （5 毫秒， 50 毫秒 + 5 毫秒） = 60 毫秒

异步复制

异步复制将远程写操作与应用程序写操作分开。此模式下，网络写操作在添加到异步队列时便得以确认。这意味着辅助站点与主站点可能不同步，直到所有写操作均发送至辅助站点。在此模式下，数据流如下：

1. 设置记录位。
2. 并行执行本地写操作和异步队列写操作。
3. 写操作得到应用程序的确认。
4. 清理线程读取异步队列项并执行网络写操作。
5. 清除记录位 （惰性清除）。

服务时间为以下操作所需的时间：

位图写操作 + MAX （本地写操作，异步队列项数据）

用 5 毫秒作为一个写操作所需的服务时间值，则异步写操作所需的服务时间大约为：

5 毫秒 + MAX （5 毫秒， 5 毫秒） = 10 毫秒

如果在较长的一段时间内，卷或**一致性组**的写入速率超过网络排出速率，则异步队列将被填满。因此，设置适当的大小非常重要。本文档稍后将会讨论估计适当卷大小的方法。

以下两种模式可控制 Remote Mirror 软件在异步磁盘队列填满时的操作。

- 阻塞模式 (Blocking mode)

  在阻塞模式 （默认设置）下， Remote Mirror 软件会阻止并等待异步磁盘队列排出到一定程度，然后再向异步队列添加写操作。这将影响应用程序的写操作，但是能够维护链路上写操作的顺序。

- 非阻塞模式 (Non-blocking mode)

  在非阻塞模式 （不适用于基于内存的队列）下， Remote Mirror 软件在异步磁盘队列填满时并不阻止，但会进入**记录**模式并记录写操作。随后的**更新式同步**将从 0 位向前读取，并且不保存写操作的顺序。如果使用这种模式，当异步磁盘队列填满而写操作顺序丢失时，则相关联的卷或一致性组会不再一致。

注 – 强烈建议在启动更新式同步 （例如，使用 `autosync` 守护进程）之前，在辅助站点上执行即时复制操作。

一致性组

在同步模式下，跨多个卷的应用程序的写操作排序是确定的。因为在需要排序时，应用程序会等待一个操作完成后才发出另一个 I/O 操作，并且只有写操作同时在主和辅助站点完成后， Remote Mirror 软件才会发出完成信号。

在默认的异步模式下，每个卷的队列都由一个或多个独立线程进行排出操作。由于此操作独立于应用程序，因此不会保留写入多个卷时的写操作顺序。

如果应用程序需要对写操作排序，则 Remote Mirror 软件提供了一致性组功能。每个一致性组都有单一的网络队列，并且尽管允许并行执行多个写操作，写操作顺序仍可通过序列号保留下来。

规划远程复制

在规划远程复制时，需要考虑业务需求、应用程序写负荷以及网络特性。

业务需求

决定复制业务数据时，您需要考虑最长延迟时间。对于辅助站点上的数据，您能允许的最长过期时间是多久？这决定了复制模式和快照安排。另外，务必要了解正在复制的应用程序是否要求以正确的顺序将写操作复制到辅助卷。

应用程序写负荷

了解写负荷的平均值和峰值对于决定主站点和辅助站点之间的网络连接类型十分重要。要确定配置，请收集以下信息：

- 数据写操作的平均速率和大小

  平均速率为应用程序在一般负荷情况下的数据写操作量。应用程序读操作对于准备和规划远程复制并不重要。

- 数据写操作的峰值速率和大小

  峰值速率是应用程序在一段测量持续时间内写入的最大数据量。

- 峰值写操作速率的持续时间和频率

  持续时间为峰值写操作速率持续的时间长短，频率为这种情况发生的频繁程度。

如果不了解这些应用程序特性，则可在应用程序运行时使用工具 （如 `iostat` 或 `sar`）测量写操作流量来测量它们。

网络特性

了解应用程序写负荷后就可以确定网络链路的需求。需要考虑的最重要的网络特性是网络带宽及主站点和辅助站点间的网络等待时间。如果在安装 Sun StorageTek Availability Suite 软件之前网络链路已存在，则可使用工具 （如 `ping`）来帮助您确定站点间链路的特性。

要使用同步复制，网络等待时间必须足够低，这样应用程序响应时间便不会因每次写操作的网络传输往返时间而受到较大的影响。而且，网络带宽必须足以处理应用程序峰值写操作期间产生的写操作流量。如果网络无法随时处理写操作流量，则应用程序响应时间将受到影响。

要使用异步复制，网络链路的带宽必须足以处理应用程序的平均写操作流量。在应用程序峰值写操作阶段，过量的写操作将写入本地异步队列，然后在以后网络流量允许时写入辅助站点。只要设置了适当的异步队列大小，在突发的写操作量超过网络限制时，仍然可以使应用程序响应时间减到最低。

请参见本文档的第 23 页 “配置异步队列”一节。您选择的 Remote Mirror 异步选项模式 （阻塞模式或非阻塞模式）决定了当异步队列填满时，软件的处理方式。

配置异步队列

如果您使用异步复制，则本节中介绍了有关配置设置的规划。这些设置基于远程镜像集或一致性组。

磁盘或内存队列

在其 3.2 版中， Remote Mirror 软件添加了对基于磁盘的异步队列的支持。为了便于从以前的版本升级，将仍支持基于内存的队列，但新的基于磁盘的队列提供了创建更大更高效队列的能力。更大的队列允许更大的突发写操作，而不会影响应用程序的响应时间。而且，基于磁盘的队列比基于内存的队列对系统资源的影响小。

异步队列的大小必须足以处理应用程序峰值写操作期间有关的突发写操作流量。大的队列能够处理长时间的突发写操作，但同时会进一步扩大辅助站点和主站点不同步的可能性。请使用峰值写操作速率、峰值写操作持续时间、写操作大小和网络链路特性来确定队列的大小。请参见第 27 页 “设置正确的基于磁盘的异步队列大小”。

您选择的异步队列选项 （阻塞模式或非阻塞模式）决定了当异步队列填满时，软件的处理方式。请使用 dsstat 工具确定异步队列的统计信息，包括高水位标志 (high-water mark, hwm)，该标志表示使用过的最大队列空间。要将异步队列添加到远程镜像集或一致性组，请使用 -q 选项运行 sndradm 命令：sndradm -q a

队列大小

可使用 dsstat(1SCM) 命令监视异步队列以检查**高水位标志** (hwm)。如果由于应用程序写入的数据超出了队列的处理能力而导致 hwm 经常达到队列总大小的 80% 到 85%，请增加队列大小。此原则同时适用于基于磁盘的队列和基于内存的队列。但是，重新调整不同类型队列大小的步骤是不同的。

基于内存的队列

- 队列中默认的最大写操作数量 （可调整）是 4096。可使用 sndradm -W 命令更改此值。
- 512 字节数据块 （默认队列大小）的默认最大数量 （可调整）是 16384，即大约 8 MB 的数据。可使用 sndradm -F 命令更改此值。

基于磁盘的队列

磁盘队列的有效大小为磁盘队列卷的大小。只能通过将磁盘队列卷替换为不同大小的卷来重新调整其大小。例如，如果队列大小为 16384 个数据块，请确保 hwm 未超过 13000 到 14000 个数据块。如果超过此数量，请使用以下步骤重新调整队列大小。

▼ 重新调整队列大小

1. **使用 sndradm -l 命令将卷置于记录模式 (logging mode)。**
2. **重新调整队列大小。**
   - 对于基于内存的队列：使用 sndradm -F 命令。
   - 对于基于磁盘的队列：使用 sndradm -q 命令将现有的磁盘队列卷替换为更大的卷。
3. **使用 sndradm -u 命令执行更新式同步。**

▼ 显示当前队列的大小、长度和 hwm

1. **键入以下命令显示队列大小：**

- 对于基于内存的队列：

```
# sndradm -P
/dev/vx/rdsk/data_t3_dg/vol0 -> priv-2-
230:/dev/vx/rdsk/data_t3_dg/vol0
autosync: off, max q writes: 4096, max q fbas: 16384, async
threads: 8, mode: async, state: replicating
```

max q fbas 指定的队列大小以数据块为单位 （此示例中为 16384 个数据块）。队列中操作项的最大值由 max q writes 指定 （此示例中为 4096）。此示例中的值表示该队列中每个操作项的平均大小为 2K。

- 对于基于磁盘的队列：

```
# sndradm -P
/dev/vx/rdsk/data_t3_dg/vol0 -> priv-
230:/dev/vx/rdsk/data_t3_dg/vol0
autosync: off, max q writes: 4096, max q fbas: 16384, async
threads: 1, mode: async, blocking diskqueue:
/dev/vx/rdsk/data_t3_dg/dq_single, state: replicating
```

显示的是磁盘队列卷 (/dev/vx/rdsk/data_t3_dg/dq_single)。可通过检查卷的大小来确定队列大小。

2. 键入以下命令以显示队列的当前长度及其 hwm：

```
# dsstat -m sndr -d q
name              q role    qi      qk  qhwi    qhwk
data_a5k_dg/vol0  D  net     4      13     5     118
```

其中：

- qi 为队列中的当前操作项数
- qk 为队列中的当前数据总大小 （以 KB 为单位）
- qhwi 为队列中在任何时刻曾经出现过的最大操作项数。
- qhwk 为队列中在任何时刻曾经出现过的数据最大值 （以 KB 为单位）。

3. 要显示流摘要和磁盘队列信息，请键入：

```
# dsstat -m sndr -r bn -d sq 2
```

4. 要显示更多信息，请使用其他显示选项运行 dsstat(1SCM)。

大小设置正确的队列的 dsstat 输出范例

注 – 此示例仅显示了本节所需的命令输出的一部分；实际上 dsstat 命令可显示更多信息。

以下 dsstat(1SCM) 内核统计信息的输出显示了有关异步队列的信息。在这些示例中，队列设置为正确的大小，并且队列当前未满。此示例显示以下设置和统计信息：

```
# dsstat -m sndr -r n -d sq -s priv-2-230:/dev/vx/rdsk/data_t3_dg/vol67
name              q role    qi      qk  qhwi   qhwk    kps   tps  svt
data_t3_dg/vol67  D  net    48     384   240   1944     10     1   54
```

其中：

- qi 条目表明总共已有 48 个写事务放入队列中
- qk 条目表明已有 384 KB 放入队列中
- qhwi 条目表示队列中操作项的 hwm 为 240 项；当前尚未达到
- qhwk 条目表示队列中数据 （以 KB 为单位）的 hwm 为 1944；当前尚未达到

假定磁盘队列的卷大小为 1 GB（或 2097152 个磁盘数据块），则 1944 个数据块的 hwm 远远低于 80% 的最高点。针对该写负荷，磁盘队列的大小是正确的。

大小设置不正确的磁盘队列的 dsstat 输出范例

以下 dsstat(1SCM) 内核统计信息的输出显示了有关异步队列的信息，此队列的大小设置不正确：

```
# sndradm -P
/dev/vx/rdsk/data_a5k_dg/vol0 -> priv-230:/dev/vx/rdsk/data_a5k_dg/vol0
autosync:off, max q writes:4096, max q fbas:16384, async threads:2, mode:async,
state:replicating

# dsstat -m sndr -d sq
name              q role    qi      qk  qhwi   qhwk    kps   tps  svt
data_a5k_dg/vol0  M  net  3609    8060  3613   8184     87    34   57
k/bitmap_dg/vol0     bmp     -       -     -      -      0     0    0
```

此示例显示了默认的队列设置，但应用程序写入的数据超出了队列的处理能力。 8184 KB 的 qhwk 值与 16384 个数据块 (8192 KB) 的 max q fbas 之间的差异表明应用程序正在逐渐接近允许的 512 字节块的最大限制。接下来的几个 I/O 操作很有可能无法进入队列。

这种情况下，增大队列是一种可行的解决方案。不过，请考虑改善网络链路 （例如使用更大带宽的接口）以满足长期效益。还可以考虑使用即时卷副本并复制阴影卷。请参见 《Sun StorageTek Availability Suite 4.0 Point-in-Time Copy 软件管理指南》。

总结：

- 如果填充速率小于或等于排出速率，则默认的队列大小已足够。
- 如果排出速率小于填充速率，则增加队列大小可提供临时的解决方案。但是，如果写操作持续较长一段时间，则队列最终仍会填满。

设置正确的基于磁盘的异步队列大小

请考虑以下示例。此示例中，每小时运行一次 `iostat` 以记录将要复制的 I/O 负荷。此示例中，假定使用 DS3 （45 MB/秒）链路。同时假定此应用程序使用单个一致性组，因此涉及单个队列。

假定应用程序在收集数据期间处于普通状况，则当收集了 24 小时的数据后，便可以确定平均写操作速率、异步队列的适当大小、远程站点在一天之后的过期情况以及选择的网络带宽是否合适此应用程序。

表 3-1 为基于磁盘的队列确定正确大小的示例

时间	kwr/s	wr/s	网络吞吐量	队列增长	队列大小
	A	B	C	A/1000 - C)*3600	
6 am	0	0	4 MBps[1]		
7 am	1000	400	4 MBps		
8 am	2000	1000	4 MBps		
9 am	2000	1000	4 MBps		
10 am	4000	1800	4 MBps		
11 am	5000	2400	4 MBps	3.6 GB	3.6 GB
12 pm	1000	400	4 MBps	-10 GB	
1 pm	1200	600	4 MBps		
2 pm	1000	500	4 MBps		
3 pm	1200	400	4 MBps		
4 pm	2000	600	4 MBps		
5 pm	1000		4 MBps		
6 pm	800		4 MBps		
7 pm	800		4 MBps		

表 3-1 为基于磁盘的队列确定正确大小的示例 （续）

时间	kwr/s	wr/s	网络吞吐量	队列增长	队列大小
8 pm	3200	1000	4 MBps		
9 pm	8000	2500	4 MBps	14 GB	14 GB
10 pm	8000	2500	4 MBps	14 GB	28 GB
11 pm	1000	400	4 MBps	-10	18
12 pm	0		4 MBps	-14	4
1 am	0		4 MBps	-14	
2 am	0		4 MBps		
3 am	0		4 MBps		
4 am	0		4 MBps		
5 am	0		4 MBps		
平均带宽	1.8 MBps				

1 兆字节/秒

填写好上表并计算队列的增长和大小后，很明显 30 GB 的队列已足够。尽管队列会增大，并且辅助站点会因此逐渐脱离同步，但在夜间运行的批处理作业能够保证队列在翌日的正常工作时间之前已为空，而且两个站点同步。

此试验还证明网络带宽适合应用程序产生的写负荷。

配置异步队列清理线程

Sun StorageTek Availability Suite 软件提供了设置清理异步队列的线程数的功能。更改此数值可允许网络上的每个卷或一致性组同时存在多重 I/O。辅助节点上的 Remote Mirror 软件可使用序列号处理 I/O 的写操作顺序。

确定对于复制配置最有效的队列清理线程数时必须考虑许多变量。这些变量包括集或一致性组的数量、可用的系统资源、网络特性，以及是否存在文件系统。如果集或一致性组的数量较少，则较多的清理线程数可能更高效。建议您进行一些基本的测试或以稍有不同的值与此变量原型加以比较，以确定对配置最有效的设置。

配置知识、网络特性及 Remote Mirror 软件的操作可以指导您选择合适的网络线程数。Remote Mirror 软件使用 Solaris RPC 作为传输机制，这些 RPC 是同步的。对于每个网络线程，独立的线程可达到的最大吞吐量为 I/O 大小/传输往返时间。考虑工作负荷主要为 2 KB 的 I/O，传输往返时间为
60 毫秒的情况。每个网络线程的吞吐量为：

2 KB/0.060 秒 = 33 KB/秒

在单个一致性组中包含单个卷或多个卷的情况下，默认的两个网络线程会将网络复制限制在 66 KB/秒。建议增加此数字。如果将复制网络设置为 4 MB/秒，则理论上 2KB 工作负荷的最佳网络线程数为：

(4096 KB/秒) / (2 KB/0.060 IO/秒) = 123 个线程

这里假定的是线性的可调节性。而实际观察表明，添加的网络线程超过 64 个后将不再受益。考虑在没有一致性组的情况下， 30 个卷在 4 MB/秒的链路上以 8 KB I/O 进行复制。默认的每卷 2 个网络线程会产生 60 个网络线程，如果工作负荷平均分布在这些卷上，则理论上带宽为：

60 * (8 KB / 0.060 IO/秒) = 8 MB/秒

这超过了网络带宽。不需要进行调整。

异步队列清理线程数的默认设置为 2。要更改此设置，可使用 `sndradm` 命令行界面与 `-A` 选项。对 `-A` 选项的描述是：在异步模式下复制某个集时， `sndradm -A` 用来指定可创建的用于处理异步队列的最大线程数 （默认值为 2）。

要确定当前配置的服务于异步队列的清理线程数，可使用 `sndradm -P` 命令。例如，您可以看到下面的集具有 2 个异步清理线程。

```
# sndradm -P
/dev/md/rdsk/d52 -> lh1:/dev/md/sdsdg/rdsk/d102
autosync: off, max q writes: 4096, max q fbas: 16384, async threads: 2, mode:
async, group: butch, blocking diskqueue: /dev/md/rdsk/d100, state: replicating
```

以下示例显示了如何使用 `sndradm -A` 选项将异步队列清理线程数更改为 3：

```
# sndradm -A 3 lh1:/dev/md/sdsdg/rdsk/d102
```

网络调整

Remote Mirror 软件将自身直接插入到系统的 I/O 路径中，监视所有流量，以确定其目标是否为远程镜像卷。系统将会跟踪目标为远程镜像卷的 I/O 命令，并管理这些写操作的复制过程。由于 Remote Mirror 软件直接插入到系统的 I/O 路径中，因此会对系统产生某些性能方面的影响。网络复制所需的额外 TCP/IP 处理也会消耗主机 CPU 资源。在主和辅助远程镜像主机上执行本节所述的操作过程。

TCP 缓冲区大小

TCP **缓冲区**大小是指传输控制协议在等待确认前允许传输的字节数。要获得最大吞吐量，请务必对正在使用的链路使用最佳的 TCP 发送和接收套接字缓冲区大小。如果缓冲区太小，则 TCP 拥塞窗口将永远无法完全打开。如果接收端缓冲区太大，则 TCP 流控制会中断，且发送端超过接收端，从而导致 TCP 窗口关闭。如果发送主机比接收主机快，则可能发生这种情况。只要仍有多余的内存，发送端的窗口过大不会造成问题。

注 – 在共享的网络上将缓冲区大小增加到过高的值可能会影响网络性能。有关调整大小的信息，请参见 《Solaris System Administrator Collection》。

表 3-2 显示了 100BASE-T 网络可能的最大吞吐量。

表 3-2 网络吞吐量和缓冲区大小

等待时间	缓冲区大小 = 24KB	缓冲区大小 = 256KB
10 毫秒	18.75 MBps[1]	100 MBps
20 毫秒	9.38 MBps	100 MBps
50 毫秒	3.75 MBps	40 MBps
100 毫秒	1.88 MBps	20 MBps
200 毫秒	0.94 MBps	10 MBps

1 兆字节/秒

查看和调整 TCP 缓冲区大小

您可以通过使用 `/usr/bin/netstat`(1M) 和 `/usr/sbin/ndd`(1M) 命令来查看和调整 TCP 缓冲区的大小。调整时需要考虑的 TCP 参数包括：

- `tcp_max_buf`
- `tcp_cwnd_max`
- `tcp_xmit_hiwat`
- `tcp_recv_hiwat`

更改其中一个参数后，请使用 `shutdown` 命令重新启动 Remote Mirror 软件，以允许该软件使用新的缓冲区大小。但是关闭并重新启动服务器后， TCP 缓冲区又恢复为默认大小。为了保存更改，需要在启动脚本中设置这些值，如本节后面的部分所述。

调整网络以查看 TCP 缓冲区和值

下面介绍了用于查看 TCP 缓冲区及其值的步骤。

▼ 查看所有 TCP 缓冲区

- **键入以下命令查看所有 TCP 缓冲区：**

```
# /usr/sbin/ndd /dev/tcp ? | more
```

▼ 按缓冲区名称查看设置

- **键入以下命令按缓冲区名称查看设置：**

```
# /usr/sbin/ndd /dev/tcp tcp_max_buf
1073741824
```

此命令显示值 1073741824。

▼ 查看套接字的缓冲区大小

- **可使用 /usr/bin/netstat(1M) 命令来查看特定网络套接字的缓冲区大小。**
  例如，查看端口 121 （默认的 Remote Mirror 端口）的大小：

```
# netstat -na |grep "121 "
*.121 *.* 0 0 262144 0 LISTEN
192.168.112.2.1009 192.168.111.2.121 263536 0 263536 0 ESTABLISHED
192.168.112.2.121 192.168.111.2.1008 263536 0 263536 0 ESTABLISHED

# netstat -na |grep rdc
*.rdc *.* 0 0 262144 0 LISTEN
ip229.1009 ip230.rdc 263536 0 263536 0 ESTABLISHED
ip229.rdc ip230.ufsd 263536 0 263536 0 ESTABLISHED
```

此示例显示的值 263536 为 256 KB 的缓冲区大小。在主主机和辅助主机上的设置必须是相同的。

▼ 在启动脚本中设置和检验缓冲区大小

注 – 在主主机和辅助主机上创建此脚本。

---

1. 使用以下值在文本编辑器中创建脚本文件：

```
#!/bin/sh
ndd -set /dev/tcp tcp_max_buf 16777216
ndd -set /dev/tcp tcp_cwnd_max 16777216

# increase DEFAULT tcp window size
ndd -set /dev/tcp tcp_xmit_hiwat 262144
ndd -set /dev/tcp tcp_recv_hiwat 262144
```

2. 将文件另存为 `/etc/rc2.d/S68ndd`，然后退出该文件。

3. 设置 `/etc/rc2.d/S68ndd` 文件的权限和所有权。

```
# /usr/bin/chmod 744 /etc/rc2.d/S68ndd
# /usr/bin/chown root /etc/rc2.d/S68ndd
```

4. 关闭并重新启动服务器。

```
# /usr/sbin/shutdown -y g0 -i6
```

5. 按照第 31 页 “查看套接字的缓冲区大小”中的介绍检验其大小。

Remote Mirror 软件如何使用 TCP/IP 端口

主节点和辅助节点上的 Remote Mirror 软件会侦听 `/etc/services` 中指定的一个公认的端口（端口 121）。Remote Mirror 通过套接字（在主站点上为任意指定的地址；在辅助站点上为公认的地址）写入从主站点到辅助站点的流量。而运行状况监视心跳则在另一个连接上进行传输（在辅助站点上为任意指定的地址；在主站点上为公认的地址）。Remote Mirror 协议在这些连接上使用 SUN RPC。

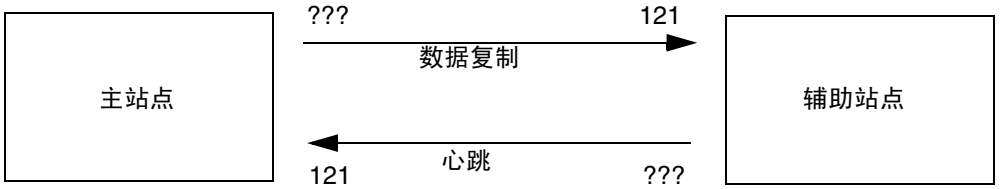

图 3-1 Remote Mirror 如何使用 TCP 端口地址

默认的 TCP 侦听端口

端口 121 是供 Remote Mirror `sndrd` 守护进程使用的默认 TCP 端口。要更改端口号，请使用文本编辑器编辑 `/etc/services` 文件。

如果您更改了该端口号，则必须在配置集内的所有远程镜像主机 （即，主主机和辅助主机以及一对多、多对一和多中继配置中的所有主机）上进行相同的更改。另外，您还必须关闭和重新启动所有受影响的主机，以使端口号的更改生效。

将 Remote Mirror 与防火墙一起使用

由于 RPC 需要确认，因此必须打开**防火墙**，以允许数据包的源或目的字段中有公认的端口地址。如果该选项可用，请确保同时配置防火墙以允许 RPC 流量。

在写入复制流量时，目标为辅助站点的数据包的目标字段包含公认的端口号，这些 RPC 的确认将在源字段包含公认的端口号。

对于运行状况监视，来自辅助站点的心跳在目标字段中带有公认的端口号，其确认将在源字段中包含此地址。

Remote Mirror 软件与 Point-in-Time Copy 软件

为了确保正常操作期间在两个站点上具有最高级别的数据完整性和系统性能，建议将 Sun StorageTek Availability Suite 4.0 Point-in-Time Copy 软件与 Remote Mirror 软件结合使用。

作为整体灾难恢复规划的一部分，即时副本可以复制到物理上的远程站点，提供卷的一致性副本。通常这种方式被称为批量复制，此操作的过程和优点如 “最佳操作指南”《Sun StorageTek Availability Suite Software–Improving Data Replication over a Highly Latent Link》中所述。

远程镜像辅助卷的即时副本可在从主站点 （主卷所在的站点）启动辅助卷的同步之前建立。开始重新同步之前，可在辅助站点上启用 Point-in-Time Copy 软件创建复制数据的即时副本，以防止双重故障。如果在重新同步的过程中产生了并发的故障，则即时副本可用作返回位置，且在并发故障问题解决后继续进行重新同步。一旦辅助站点与主站点完全同步后，便可以禁用 Point-in-Time Copy 软件卷集，或将其用于辅助站点的其他目的 （远程备份、远程数据分析或其他功能）。

在启用、复制或更新操作中内部执行的 Point-in-Time Copy 软件 I/O 操作可以更改阴影卷的内容，而不使任何新的 I/O 进入 I/O 堆栈。当发生这种情况时，I/O 不会在 `sv` 层被中断。如果该阴影卷同时也是远程镜像卷，则 Remote Mirror 软件也不会察觉到这些 I/O 操作。在这种情况下，I/O 操作修改的数据将不会被复制到目标远程镜像卷。

为支持这种复制，可将 Point-in-Time Copy 软件配置为向 Remote Mirror 软件提供已更改的位图。如果 Remote Mirror 软件处于记录模式，则它会接受位图，然后将 Point-in-Time Copy 软件位图与自身中该卷的位图进行 "OR" 比较，并将 Point-in-Time Copy 软件位图的变化添加到自身中要复制到远程节点的变化列表中。如果 Remote Mirror 软件处于卷的复制模式，则拒绝来自 Point-in-Time Copy 软件的位图。于是，启动、复制或更新操作将失败。一旦重新启用 Remote Mirror 记录模式，便可重新进行 Point-in-Time Copy 软件操作。

注 – 只有当远程镜像卷集处于记录模式时， Point-in-Time Copy 软件才能在远程镜像卷上成功地执行启用、复制、更新或复位操作。否则， Point-in-Time Copy 操作将失败， Remote Mirror 软件将报告操作被拒绝。

远程复制配置

Remote Mirror 软件可以创建一对多、多对一和多中继卷集。

- 一对多复制可用于将数据从主卷复制到驻留在一台或多台主机上的多个辅助卷。一个主站点卷和每个辅助站点卷分别组成一个单独的卷集。例如，对于一个主主机卷和三个辅助主机卷，您需要配置三个卷集：主主机卷 A 和辅助主机卷 B1、主主机卷 A 和辅助主机卷 B2，以及主主机卷 A 和辅助主机卷 B3。
- 多对一复制可用于通过多个网络连接、在两台以上的主机间复制卷。本软件支持将多台不同主机上的卷复制到单台主机上的卷中。此术语不同于卷到卷的一对多配置。
- 多中继复制是指一个卷集的辅助主机卷可以作为另一个卷集的主主机卷。在一个主主机卷 A 和一个辅助主机卷 B 的情况下，由辅助主机卷 B 充当辅助主机卷 B1 的主主机卷 A1。

Remote Mirror 软件还支持将上述几种配置结合使用。

词汇表

asynchronous queue（异步队列） 用于存储要复制到远程站点的写操作的本地磁盘或内存。写操作进入队列后由应用程序确认，然后在网络性能允许的情况下稍后转发到远程站点。

asynchronous replication（异步复制） 异步复制在远程映像更新前即向源主机确认主 I/O 事务已完成。即，当本地写操作已结束并且远程写操作已进入队列时，便向主机确认 I/O 事务已完成。推迟至辅助副本消除了 I/O 响应时间因长距离传播而产生的延时。

auto synchronization（自动同步） 在主主机上启用自动同步选项后，如果系统重新引导或者发生链路失败，则同步守护进程 (`autosyncd`) 会试图重新同步卷集。

blocking（阻塞模式） （异步队列）在阻塞模式下，如果异步队列已满，则之后的写操作都将延迟，直到队列腾出足够的空间允许进行写操作。阻塞模式是默认的异步运行选项，能够保证发送到辅助站点的数据包的写操作顺序。设置阻塞选项后，如果异步队列已满，则应用程序响应时间可能会受影响。

configuration location（配置位置） 一个位置， Sun StorageTek Availability Suite 软件在其中存储它使用的有关所有已启用卷的配置信息。

consistency group（一致性组） 一致性组是指共享同一个异步队列以维护写操作顺序的一组远程卷。

`dsstat` Sun StorageTek Availability Suite 工具集中的一个工具，可用于显示来自 Remote Mirror 和 Point-in-Time 快照产品的内核统计信息。

firewall（防火墙） 一台用作两网络间接口并控制这些网络间流量的计算机，目的是保护内部网络免受来自外部网络的电子攻击。

forward resynchronization（正向重新同步） 请参见更新式同步。

full synchronization
（整卷式同步） 整卷式同步执行卷对卷的完全复制，是最耗时的同步操作。大部分情况下，是使用主卷对辅助卷进行同步。然而，在恢复出现故障的主磁盘时，可能需要使用幸存的远程镜像作为源来执行反向同步。

hwm 请参见高水位标志

high water mark
（高水位标志） 高水位标志是指所使用的异步队列最大容量。

lazy clear
（惰性清除） 清除内核中的位的操作，但只有在另一个位被设置或内核中的副本被回收之后，才会向磁盘写回位图块。由于在系统发生故障之后只是重新传输更改，因此这种操作很安全。

logging
（记录） 一种跟踪模式，其中由位图来跟踪对磁盘的写入，而不是将每个 I/O 事件实时记录到日志中。当远程服务中断或受损时，此方法可以记录尚未复制到远程站点的已更新磁盘数据。每个源卷中不再与其远程副本匹配的数据块都被标记出来。通过使用此日志，本软件可以执行优化的更新式同步来重建远程镜像，而不必执行卷对卷的完全复制。

non-blocking
（非阻塞模式） （异步队列）在非阻塞模式下，如果异步队列已满，则 Remote Mirror 软件进入记录模式并放弃队列的内容。非阻塞模式不能保证发往辅助站点的数据包的写操作顺序，但是能够保证异步队列已满时不会影响应用程序响应时间。

primary or local: host or volume
（主或本地：主机或卷） 主机应用程序主要依赖的系统或卷。例如，产品数据库由此获取访问数据。本软件将把此数据复制到辅助站点。

replication
（复制） 完成卷集的初始同步之后，该软件将确保主卷和辅助卷始终含有相同的数据。复制是由用户层应用程序的写操作所触发的；复制是一个持续的进程。

reverse synchronization
（反向同步） 预演数据恢复时所使用的操作。日志将记录预演过程中在辅助系统上进行的更新测试。当主系统恢复时，在辅助系统上执行的更新测试将被来自主系统映像的数据块覆盖，从而使远程数据卷集保持一致。

secondary or remote: host or volume
（辅助或远程：主机或卷） 主组件的远程副本，在此对数据副本进行读写。远程副本在对等服务器上传送，无需主机介入。一台服务器可同时充作某些卷的主存储设备和其他卷的辅助 （远程）存储设备。

synchronization
（同步） 在目标磁盘上建立源磁盘的副本的过程，是软件镜像的前提条件。

synchronous replication
（同步复制） 由于 I/O 响应时间的传输延迟的不利影响，同步复制只限于较短的距离 （几十公里）。

TCP buffer
（TCP 缓冲区） TCP 缓冲区大小为传输控制协议在等待确认前允许传输的字节数。

update synchronization
（更新式同步） 更新式同步只复制那些由记录模式标记的磁盘数据块，减少了恢复远程镜像卷集的时间。

volume set file
（卷集文件） 一个含有特定卷集相关信息的文本文件。此文本文件与配置位置不同，配置位置包含 Remote Mirror 和 Point-in-Time Copy 软件使用的所有已配置卷集的有关信息。

索引

S

T

W

X

Y