

MySQL最佳拍档!

DRBD 权威指南

基于Corosync+Heartbeat
技术构建网络RAID

陶利军 编著

使用开源软件构建高性能、高可用的数据存储

基于网络的RAID 1，充分保证数据及服务的正常运行

使用Heartbeat及Corosync作为底层通讯支持

使用LVM及GFS文件系统作为底层存储

DRBD基础详解到DRBD上线应用案例剖析

图形化的LVM/C管理界面

DRBD能够为你节约大量的存储设备预算



清华大学出版社

DRBD 权威指南

基于Corosync+Heartbeat
技术构建网络RAID

陶利军
编著

清华大学出版社

内 容 简 介

在 Linux 操作系统下,对于数据的安全可以通过 DRBD 来实现,通过它可以在不专业的硬件上实现专业的存储。DRBD 是一种网络 RAID 1,它能够将两台独立机器的存储作为 RAID1 来使用,两台机器来管理 RAID1。

本书分为 6 个部分共 16 章,内容包括:DRBD 理论基础、DRBD 技术基础、DRBD 技术特点、优化 DRBD 性能、安装和配置 DRBD、管理 DRBD、常见的管理任务、故障排除和错误恢复、DRBD 与其他应用程序的集成、Pacemaker 集群和 DRBD 的集成、DRBD 使用 LVM、GFS 与 DRBD、DRBD 与 Heartbeat 的集成、运维案例(Pacemaker+DRBD+其他 / Heartbeat+DRBD+NFS)、图形化管理工具 LCMC。

本书适合存储工程师、网络管理员、网络工程师作为参考书使用。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

DRBD 权威指南——基于 Corosync+Heartbeat 技术构建网络 RAID / 陶利军编著. —北京:清华大学出版社, 2013

ISBN 978-7-302-32461-4

I. ①D… II. ①陶… III. ①计算机网络—信息存储—研究 IV. ①TP393.0

中国版本图书馆 CIP 数据核字(2013)第 106324 号

责任编辑:栾大成

装帧设计:杨玉兰

责任校对:胡伟民

责任印制:沈 露

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京世知印务有限公司

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:188mm×260mm 印 张:26.75 插 页:1 字 数:670 千字

版 次:2013 年 9 月第 1 版

印 次:2013 年 9 月第 1 次印刷

印 数:1~3000

定 价:59.00 元

产品编号:053484-01

前言

在信息时代，无论是互联网提供的页面还是企业的数据都是生产力产生的结果，因此对于通过生产工具来保护生产力产生的结果是一件刻不容缓的事情，而且也是对生产力的尊重，更是对自己的尊重。在 Linux 操作系统下，对于数据的安全我们可以通过 DRBD 来实现，通过它可以在不专业的硬件上实现专业的存储效果。

说明：需要说明的是这里提供的任何方案都是基于中小规模、勤俭节约的企业或者是单位，DRBD 的宗旨就是使用最廉价的生产工具来实现最大的效益，如果你所在的企业或者是单位财大气粗，那么可以使用其他昂贵的架构。

DRBD 是一种网络 RAID 1，它能够将两台独立主机的存储设备作为 RAID1 来使用，两台机器来管理 RAID1，而这个 RAID1 的两个底层存储分别放置在这两台主机上，因此任何一台主机或者是底层存储损坏都不会影响企业或者是单位生产环境中的使用，等到损坏的主机或底层存储被修复或者是替换后将数据同步过去就可以了。

DRBD 的管理和控制不是特别复杂，只要认真学习本书中提到的方方面面即可，必要时可以访问 DRBD 的官方网站以便进行更深层次的学习。

本书中的内容包括了 6 个部分共 16 章的内容，讲述了 DRBD 的理论和生产环境的部署及运维，以及和其他程序的结合应用，包括：LVM、GFS 等基础底层软件，和 Heartbeat、Corosync、Pacemaker 等高可用软件，实现基于 DRBD 服务器之上的服务高可用。

因此，从理论到实践，再到生产环境下的使用；从部署安装到版本升级，再到故障解决；从日常管理到优化，再到感性化的图形界面，本书中都清楚地渗透到每一个环节。

第 1 部分 DRBD 理论基础

本部分内容包括以下 3 章。

- DRBD 技术基础
- DRBD 技术特点
- 优化 DRBD 性能

第 2 部分 安装和配置 DRBD

本部分内容包括以下 3 章。

- 通过预编译二进制安装 DRBD
- 从源代码编译安装 DRBD
- 配置 DRBD

第 3 部分 管理 DRBD

本部分内容包括以下 2 章。

- 常见的管理任务
- 故障排除和错误恢复

第 4 部分 DRBD 与其他应用程序的集成

本部分内容包括以下 4 章。

- Pacemaker 集群和 DRBD 的集成
- DRBD 使用 LVM
- GFS 与 DRBD
- DRBD 与 Heartbeat 的集成

第 5 部分 运维案例

本部分内容包括以下 2 章。

- Pacemaker + DRBD+其他
- Heartbeat+DRBD+NFS

第 6 部分 图形化管理工具 LCMC

本部分内容包括以下 2 章。

- 认识 Linux 集群管理器 LCMC
- 使用 Linux 集群管理器 LCMC

内 容 声 明

关于本书内容的说明，如果你在哪里看到了与本书雷同的内容，你需要确定一下它的内容是否来自于相应软件的官方网站、man 文档、howto、README、Changelog、INSTALL、LICENSE、*.conf、*.cfg 等，这些原创，在我看来，什么是原创？只有这些才是原创（我个人的观点，别拿砖头拍我！），我们只不过是它们的衍生和应用。本书中的内容就是这样，这是我个人的一个学习方法，对于每一个新使用的软件，我都会看它提供的相关文档和其官方网站，配置文件绝对是软件的精华所在，因此在本书中讲述了大量的配置文件，没办法，Linux 下的服务不就是命令加上配置吗？

由于这些官方网站、man 文档、howto、README、Changelog、INSTALL、LICENSE、*.conf、*cfg 等都是英文，因此对于我们的认识和阅读很不方便，事实上也正是缺乏这些文档的知识才导致我们一直徘徊在技术的门口，因此本人便基于这个基础来编写本书，将这些最基础也是最权威的文档通过理解来实现汉语化，以方便更多人阅读，以个人的感觉这些东西实际上是我们最需要的，它是认知的第一步，毕竟我们的官方语言是汉语。

书中的内容是我在工作中的一个总结，我没有去刻意地改变一个说法，相反只要是官方文档中要的，我就尽可能地使用它们的提法、说法及方法。

使用对象

- 广大的 Linux 爱好者；
- 具有一定 Linux 基础的系统管理员；
- Linux 下的系统工程师；
- 存储服务器管理员；
- 培训中心；
- 运维人员；
- 一切应该了解和使用 DRBD 的用户。

关于读者

全书分为 6 个部分共 16 章的内容，如果你是初学者，那么不要从第 1 部分开始，而要从第 2 部分开始，因为第 1 部分属于 DRBD 的理论基础部分，对于初学者来说看起来会很累，首先可以从搭建起一个运行 DRBD 的服务器开始。

接下来学习第 3 部分。在第 2、第 3 部分学懂之后，就可以进一步学习第 4 部分，然后再回头看第 1 部分，但熟悉 DRBD 及 Heartbeat、Pacemaker 的用户可以直接阅读第 5 部分。

本书的最后一部分是针对 DRBD 的图形化管理界面所写的，对于不熟悉命令行的用户可以使用体验一下，而对于熟悉了命令行的用户也不妨使用图形管理界面体验一下它的管理控制。

作者声明

本书的内容是我在工作中的一个总结，在生产环境中都使用过，并非纸上谈兵，但是书中的例子，我尽可能地不使用生产环境中的，一是怕对你造成误导；二是不想说什么权威。

我在前面提到了文稿内容的来源。对于文稿的构成，一部分是对员工培训的文稿；一部分是在培训中心的文案；还有一部分是我在学习中的笔记，由这三部分融合而成，而非简单的拼凑。

另外，毕竟我们都是做互联网的，每天面对着无数个页面，我所要说的是，如果读者在学习本书的过程中发现有和网络上相似的内容，那么确定一下是否是两者（即笔者和您看到网络上文章的作者）参考了同一个官方的资料，本人绝对没有有意抄袭其他作者的内容，这是第一；第二，如果真的是我写的内容确实和您的内容有相同之处，那么及时和我联系（绝对是缘分！）；第三，互联网给了我们发展，也给了我们交流，如果您在看本书的过程中发现有个别说法、提法和您的相同，那么请您海涵，往往是一个说法、提法用久了就觉得是自己的说法了（我相信谁都会犯这种错！）；第四，由于本人是做运维（系统管理和网络管理），因此在写作风格上也是按照自己的认知过程所写，既没有受过专业的训练也没有模仿某个作者或者某个作品的写作风格，如果真的和您的写作风格相同，那么绝对是巧合（这个就不要计较了！）；第五，本套书中引用了互联网的一些内容，由于同一个内容被转来转去，确实很难找到原出处，因此在应用的内容处只指明了来源于互联网。

由于本人才疏学浅，因此，对于本书难免会有疏漏和不足，因此，如果广大读者如果有什么建议和意见可以给本人发邮件：drbd_service@126.com。

关于本书

作者简介

写给读者

敬献给所有在外漂泊的男人们——有感于居住证。

小时候听过一首《水手》的歌，懂得了什么是男人，长大后拼着一腔热血闯世界，渐渐地将自己容纳为这个陌生城市的一部分。

可是突然有一天醒了，这个城市并不属于我，在我们尽义务的时候必须是这个城市的一部分，个人所得税，这是每一个公民依法缴纳的义务，不缴就是违法，比如捐款；又比如养老保险，这个我不想说得太明白，你懂得！

在你享受权利的时候，你必须不是这个城市的一部分，比如“北京的外乡人连续救人却因是外地人不给帐篷住”，北车营村党支部书记翟瑞生的解释是“前期救灾物资比较紧缺，需要先紧着本村人口”。又如在孩子读书上学的时候，“北京公布异地高考方案，外地人只能上职校不能上高中”，相信这句话不知道有多少人看到“记者今天上午从市教委获悉，2013年符合相关条件的随迁子女可以参加中等职业学校考试录取；2014年符合相关条件的随迁子女可以参加高等职业学校考试录取。”也就是说，外地人只能上技工学校，生气吗？你懂得！

“自2014年起，凡进城务工人员持有有效北京市居住证明，有合法稳定的住所，合法稳定职业已满6年，在京连续缴纳社会保险已满6年，其随迁子女具有本市学籍且已在北京连续就读高中阶段教育3年学习年限的，可以在北京参加高等职业学校的考试录取。学生从高等职业学校高职毕业后，可参加优秀应届毕业生升入本科段学习的推荐与考试录取。”你懂得吗？

小时候学过“少小离家老大回”的诗句，临近春节的时候突然接到一个哥们的电话，电话中说“哥，我回去了，过完年我不想回来了，想回去找点事情做”，是的，也许早点回去比晚点回去更好。

多少人拼搏更变自己的命运来改变后代的命运，考职称、读博士……

为什么不去推动体制的变更呢？老牌英国为什么会落后了？

目 录

第 1 部分 DRBD 理论基础

第 1 章 DRBD 技术基础.....2	2.9 脑裂通知和自动恢复.....10
1.1 内核模块.....2	2.10 支持磁盘刷新.....10
1.2 用户空间工具.....3	2.11 磁盘错误处理策略.....11
1.3 资源.....3	2.12 处理过期数据的策略.....11
1.4 资源角色.....4	2.13 三路复制.....12
第 2 章 DRBD 技术特点.....5	2.14 使用 DRBD 代理实现远 距离复制.....13
2.1 单主模式.....5	2.15 基于“运送”的复制.....14
2.2 双主模式.....5	2.16 浮动对点.....14
2.3 复制数据传输模式.....5	第 3 章 优化 DRBD 性能.....15
2.4 多种传输复制数据的协议.....6	3.1 优化 DRBD.....15
2.5 有效地同步策略.....7	3.1.1 优化 DRBD 吞吐量.....15
2.5.1 可变的同步速率.....7	3.1.2 优化 DRBD 延时.....17
2.5.2 锁定同步速率.....7	3.2 关于调优的建议.....18
2.5.3 基于校验和的复制.....8	3.2.1 为 DRBD 设置 CPU 掩码.....18
2.6 延时同步复制数据.....8	3.2.2 修改网络的 MTU.....18
2.7 在线设备验证.....8	3.2.3 使用 deadline I/O 调度方式.....19
2.8 复制流量的完整性检测.....9	

第 2 部分 安装和配置 DRBD

第 4 章 通过预编译二进制	4.2.4 Ubuntu Linux.....29
安装 DRBD.....23	第 5 章 从源代码编译安装 DRBD.....30
4.1 由 LINBIT 提供的安装包.....23	5.1 下载 DRBD 源代码.....30
4.2 由发行商提供的软件包.....24	5.2 从 DRBD 仓库牵出 DRBD 源程序.....31
4.2.1 SUSE Linux Enterprise Server.....24	5.3 从源代码构建特定平台的 DRBD 二进制安装包.....34
4.2.2 Debian GNU/Linux.....25	5.3.1 检测构建 DRBD 的依赖性.....34
4.2.3 CentOS.....25	

5.3.2 准备内核源码	35	6.3 配置资源	56
5.3.3 准备 DRBD 的编译树	36	6.3.1 配置举例	57
5.3.4 构建 DRBD 用户空间工具	37	6.3.2 配置 global 部分	59
5.3.5 编译 DRBD 作为内核模块	37	6.3.3 配置 common 部分	60
5.4 构建 DRBD 的 RPM 安装包	44	6.3.4 配置 resource 部分	60
5.5 构建 DRBD 的 Debian 安装包	53	6.4 第一次启用 DRBD 资源	61
第 6 章 配置 DRBD	54	6.5 初始化设备同步	64
6.1 准备底层存储设备	54	6.6 使用基于“运送”的复制	65
6.2 准备网络配置	54		

第 3 部分 管理 DRBD

第 7 章 常见的管理任务	68	7.8 启用双主模式	84
7.1 检测 DRBD 状态	68	7.8.1 永久双主模式	84
7.1.1 通过 drbd-overview		7.8.2 临时双主模式	85
命令获取 DRBD 的状态	68	7.8.3 系统启动时自动提升	
7.1.2 通过伪文件系统/proc/drbd		双主模式	85
了解 DRBD 的状态信息	68	7.9 使用在线设备验证	85
7.1.3 连接状态	70	7.9.1 启用在线设备验证	86
7.1.4 资源角色	72	7.9.2 执行在线验证	86
7.1.5 磁盘状态	73	7.9.3 自动在线验证	86
7.1.6 I/O 状态标志	74	7.10 配置同步的速率	87
7.1.7 性能指标	74	7.10.1 在配置中锁定同步速率	89
7.2 启用和禁用资源	76	7.10.2 临时锁定同步速率的配置	91
7.2.1 启用资源	76	7.10.3 可变同步速率的配置	91
7.2.2 禁用资源	76	7.11 配置基于校验和的同步	92
7.3 重新配置资源	76	7.12 配置阻塞策略和暂停复制	92
7.4 资源的升级和降级	77	7.13 配置处理 I/O 错误策略	93
7.5 基本的手动故障转移	77	7.14 配置复制流量完整性检测	94
7.6 更新 DRBD 版本	78	7.15 调整资源大小	95
7.6.1 更新仓库	78	7.15.1 在线增大资源大小	95
7.6.2 更新安装包	78	7.15.2 离线增大资源大小	95
7.6.3 配置迁移	81	7.15.3 在线缩小资源大小	96
7.7 将 DRBD 8.4 降级到 8.3	84	7.15.4 离线缩小资源大小	97

7.16	禁用后台驱动器缓存	98
7.17	配置脑裂行为	98
7.17.1	脑裂通知	99
7.17.2	脑裂自动恢复策略	100
7.18	创建三个节点的设置	101
7.18.1	设备堆叠的注意事项	101
7.18.2	配置堆叠资源	102
7.18.3	启用堆叠资源	103
7.19	使用 DRBD 代理	103
7.19.1	DRBD 代理部署注意事项	103
7.19.2	安装	104
7.19.3	许可文件	105
7.19.4	配置代理	105
7.19.5	控制 DRBD 代理	106
7.19.6	DRBD 的代理插件	107
7.19.7	故障查找	108

第 4 部分 DRBD 与其他应用程序的集成

第 9 章 Pacemaker 集群和 DRBD

的集成	115
9.1 认识 Pacemaker	115
9.1.1 Pacemaker 的功能	115
9.1.2 Pacemaker 的结构	116
9.1.3 Pacemaker 的内部组成	117
9.1.4 Pacemaker 的集群类型	118
9.2 在 Pacemaker 中添加 DRBD	
后台服务	119
9.3 在集群中使用基于资源级别的	
的栅功能	120
9.3.1 使用资源级别的栅功能	
——dopd	120
9.3.2 使用基于资源级别的	
栅功能	122

第 8 章 故障排除和错误恢复

8.1 处理硬盘驱动器错误	109
8.1.1 从 DRBD 手动分离硬盘	
驱动器	109
8.1.2 出现 I/O 错误自动分离硬盘	
驱动器	109
8.1.3 在使用内部元数据的情况下	
替代失败磁盘	110
8.1.4 在使用外部元数据的情况下	
替代失败磁盘	110
8.2 处理节点错误	111
8.2.1 处理临时 Secondary	
节点错误	111
8.2.2 处理临时 Primary 节点错误	111
8.2.3 处理永久节点错误	111
8.3 手动脑裂恢复	112

9.4 在 Pacemaker 集群中使用堆叠的

DRBD 资源	122
9.4.1 向 Pacemaker 集群中添加	
异地灾备	122
9.4.2 在 Pacemaker 集群中使用	
堆叠资源实现 4 路冗余	124
9.5 两个基于 SAN 的 Pacemaker	
集群之间 DRBD 数据复制	127
9.5.1 DRBD 资源配置	127
9.5.2 Pacemaker 资源配置	128
9.5.3 站点的故障转移	129

第 10 章 DRBD 使用 LVM

10.1 LVM 简介	130
10.2 使用逻辑卷作为 DRBD	
后端设备	133

10.3 在 DRBD 的同步中自动使用 LVM 快照	144	12.1 安装 Heartbeat	161
10.4 配置 DRBD 资源作为物理卷	150	12.1.1 下载安装 glue	161
10.5 在现有卷组中添加新的 DRBD 卷	153	12.1.2 下载安装 heartbeat	163
10.6 DRBD 与 LVM 的嵌套配置	154	12.1.3 下载安装 agents	165
10.7 使用 Pacemaker 实现高可用	156	12.2 配置 Heartbeat	167
第 11 章 GFS 与 DRBD	157	12.2.1 ha.cf 文件	169
11.1 GFS primer	157	12.2.2 haresources 文件	175
11.2 创建适合于 GFS 的 DRBD 资源	157	12.2.3 authkeys 文件	175
11.3 配置 LVM 识别 DRBD 资源	158	12.3 启动 Heartbeat	176
11.4 配置集群支持 GFS	158	12.3.1 环境部署	176
11.5 创建 GFS 文件系统	159	12.3.2 启动主 Heartbeat	177
11.6 使用 GFS 文件系统	159	12.3.3 启动备用 Heartbeat	180
第 12 章 DRBD 与 Heartbeat 的集成	161	12.4 测试 Heartbeat	182
		12.4.1 正常情况下访问	183
		12.4.2 宕掉主节点	183
		12.4.3 重新启动主节点	186
		12.4.4 命令参考	233
		13.5 pacemaker-cli 安装包	273
		13.5.1 命令参考	274
		13.5.2 配置 Corosync	289
		13.5.3 启动 Corosync	290
		13.6 校验安装	292
		13.6.1 校验 Corosync 安装	292
		13.6.2 校验 Pacemaker 安装	293
		13.7 配置服务	293
		13.7.1 添加虚拟 IP 地址	294
		13.7.2 添加 Web 服务	299
		13.8 添加其他配置	300
		13.8.1 根据需要设置 STONITH 设备	300
		13.8.2 禁用故障恢复使用节点	300
第 13 章 Pacemaker+DRBD+其他	191		
13.1 安装集群软件	191		
13.1.1 在 CentOS 5.6 系统上	191		
13.1.2 在 CentOS 6.3 系统上	194		
13.1.3 安装集群管理软件	199		
13.1.4 其他设置	200		
13.2 Corosync 安装包	201		
13.2.1 配置文件	202		
13.2.2 命令参考	206		
13.3 Pacemaker 安装包	214		
13.3.1 配置文件	216		
13.3.2 命令参考	218		
13.4 Crm 工具	223		
13.4.1 crm 命令行选项	224		
13.4.2 用户接口使用介绍	225		

第 5 部分 运维案例

第 13 章 Pacemaker+DRBD+其他	191	13.4.3 命令参考	233
13.1 安装集群软件	191	13.5 pacemaker-cli 安装包	273
13.1.1 在 CentOS 5.6 系统上	191	13.5.1 命令参考	274
13.1.2 在 CentOS 6.3 系统上	194	13.5.2 配置 Corosync	289
13.1.3 安装集群管理软件	199	13.5.3 启动 Corosync	290
13.1.4 其他设置	200	13.6 校验安装	292
13.2 Corosync 安装包	201	13.6.1 校验 Corosync 安装	292
13.2.1 配置文件	202	13.6.2 校验 Pacemaker 安装	293
13.2.2 命令参考	206	13.7 配置服务	293
13.3 Pacemaker 安装包	214	13.7.1 添加虚拟 IP 地址	294
13.3.1 配置文件	216	13.7.2 添加 Web 服务	299
13.3.2 命令参考	218	13.8 添加其他配置	300
13.4 Crm 工具	223	13.8.1 根据需要设置 STONITH 设备	300
13.4.1 crm 命令行选项	224	13.8.2 禁用故障恢复使用节点	300
13.4.2 用户接口使用介绍	225		

13.9 执行 Active/Passive 测试	301	14.11.1 drbd	342
13.10 Pacemaker 与 DRBD 的结合	303	14.11.2 drbdsetup	348
13.10.1 安装 DRBD	303	14.11.3 drbdmeta	351
13.10.2 配置 DRBD	309	14.11.4 drbd-overview	352
13.11 主备切换测试	315	14.12 对 Primary 角色的切换	
13.11.1 维护切换测试	315	配置文件	353
13.11.2 故障切换测试	319	14.12.1 drbd.d	353
第 14 章 Heartbeat+DRBD+NFS	325	14.12.2 global_common.conf	353
14.1 安装 DRBD	325	14.12.3 *.res 文件	354
14.1.1 安装 DRBD 8.2.1	325	14.12.4 drbd.conf	355
14.1.2 启动	327	14.13 脑裂处理	365
14.1.3 设置为 Primary 端	328	14.13.1 拔掉主(Primary)节点的	
14.2 使用 DRBD 设备	329	所有网线	366
14.3 对 Primary 角色的切换	330	14.13.2 拔掉备用(Secondary)节点	
14.4 设置 NFS	331	的所有网线	367
14.5 与 Heartbeat 结合	331	14.13.3 主(Primary)宕机	369
14.6 配置 DRBD	332	14.13.4 另一种情况	372
14.7 客户端挂接 NFS	334	14.13.5 又一种情况	375
14.8 故障	334	14.14 添加资源	376
14.9 升级安装	334	14.14.1 添加磁盘	376
14.9.1 安装 DRBD 8.4.1	334	14.14.2 添加配置文件	378
14.9.2 编辑 DRBD 的配置文件	335	14.14.3 启用资源	380
14.10 目录结构分析	338	14.15 主备节点切换	383
14.10.1 rpm 安装包	338	14.15.1 运行时切换	383
14.10.2 tar 安装包	340	14.15.2 停止 DRDB 服务切换	385
14.11 用户命令	342		

第 6 部分 图形化管理工具 LCMC

第 15 章 认识 Linux 集群		第 16 章 使用 Linux 集群	
管理器 LCMC	387	管理器 LCMC	390
15.1 安装方式	387	16.1 添加机器和集群	390
15.2 下载并安装 LCMC	387	16.1.1 添加主机向导	392
15.3 运行 LCMC	389	16.1.2 添加集群向导	395

16.2 集群管理	399
16.2.1 连接和断开集群	399

16.2.2 添加和移除集群	400
16.2.3 集群资源管理	400

第 1 部分

DRBD 理论基础

本部分内容包括以下 3 章。

- DRBD 技术基础：

讲述 DRBD 的组成：内核模块和用户空间工具；DRBD 的资源及资源角色。

- DRBD 技术特点：

讲述了 DRBD 几乎所有的技术，正是 DRBD 有如此多的技术才保证了 DRBD 能够稳定地运行，也正是如此 DRBD 才被广泛地使用。

- 优化 DRBD 性能：

调优 DRBD，使其发挥更大的性能。

第 1 章 DRBD 技术基础

DRBD(Distributed Replicated Block Device), 翻译为中文就是“分布式复制块设备”, 解释说明一下就是基于软件实现的无共享, 基于块级别设备(硬盘、分区、逻辑卷等)的镜像内容复制解决方案。

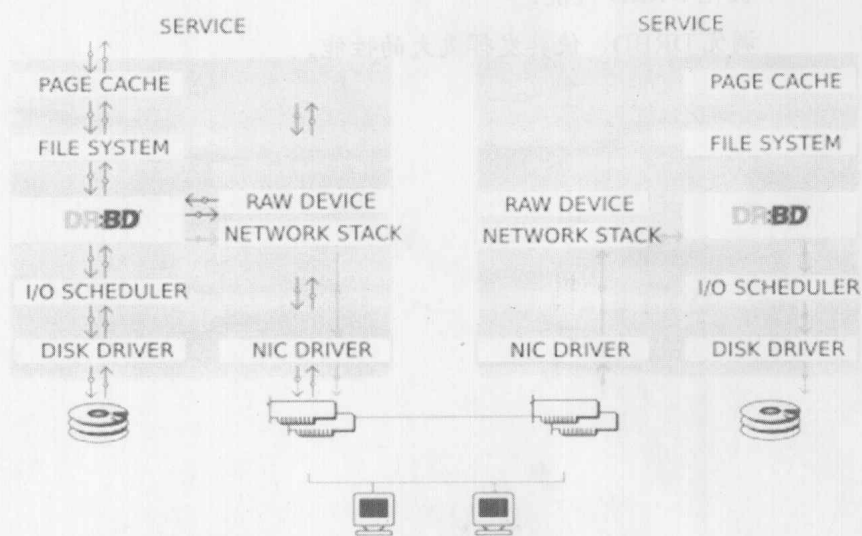
DRBD 镜像数据的特点如下。

- 实时性: 应用程序修改了设备上的数据之后, 复制数据将会持续地被复制到对点上。
- 透明性: 应用程序不需要知道数据被存储在多个节点主机上。
- 同步或者异步: 对于复制数据, 可以使用同步复制, 也可以使用异步复制。对于同步镜像, 只有在所有节点执行完成写操作之后, 应用程序才会被通知写完成; 而对于异步镜像, 当本地节点执行完成写操作之后, 在数据同步还没有完成之前就会通知应用程序写完成。

1.1 内核模块

DRBD 技术的核心功能是通过 Linux 内核模块的方式实现, 具体来说, DRBD 技术由虚拟块设备驱动程序构成了驱动设备, 所以 DRBD 技术位于系统的 I/O 堆栈的底部附近。正因为如此, DRBD 技术非常灵活多变, 这使得它的复制解决方案适合于任何应用添加高可用。

图解:



在上图中, DRBD 的位置处于 Linux 输入/输出堆栈内。

1.2 用户空间工具

DRBD 提供了一组管理工具,通过这些管理工具来和内核模块通信,以便配置和管理 DRBD 资源。

Drbdadm: 它是 DRBD 程序套件中高级管理工具,它能够从配置文件/etc/drbd.conf 中获取 DRBD 的所有配置参数,并且充当了 drbdsetup 和 drbdmeta 命令的前端(就是说 drbdadmin 实际上是调用了 drbdsetup 和 drbdmeta 命令)。

说明:需要说明的一点是,drbdadmin 有一个“演示模式”,也就是说通过“-d”选项能够演示 drbdadm 调用 drbdsetup 和 drbdmeta 命令而实际上并没有执行这些命令。

Drbdsetup: 配置 DRBD 模块载入到内核,所有 drbdsetup 的参数必须通过命令行传递,它和 drbdadm 在使用上的区别在于灵活性,大多数用户很少直接使用这个命令,或者是根本就不直接使用它。

Drbdmeta: 该命令用于创建(create),转储(dump),恢复(restore)和修改(modify)DRBD 元数据结构。它类似于 drbdsetup 命令,大多数用户很少直接使用 drbdmeta 命令。

1.3 资源

在 DRBD 中,resource 是集合的统称,它指定的是特定数据复制的各个方面,包括以下名称。

- **资源名字(resource name):** 在 DRBD 中,资源名字可以是任意的,它可以使用不包含空格以外的任何 US-ASCII 字符,但是在具体使用时有 26 个英文字母和 10 个数字就足够了,不要牵涉更多的字符。
 - **卷(volumes):** 任何资源都是一个包含了一个或者多个的复制组,卷就是用于存储数据的空间。DRBD 确保精确地写入资源中的所有卷(volume)。卷的标号从 0 开始,一个资源中最多包括 65 535 个卷。卷中的数据包括复制数据集和 DRBD 内部使用的元数据集。
- 从 drbdadm 层面来看,卷包含在资源内,可以通过<resource>/<volume>格式来表达资源名称和卷的关系。
- **DRBD 设备(device):** 这是一个由 DRBD 管理的虚拟块设备,它的主要设备号为 147,次要设备号从 0 开始,这是习惯。在资源中每一个 DRBD 设备相当于一个卷,相关的块设备通常被命名为“/dev/drbdX”,这里的 X 是设备的次要号。DRBD 也允许自定义块设备的名称,但是必须以“drbd_”开始。

较早的 DRBD 版本,使用了 NBD 设备的主设备号 43,DRBD 很早就不再使用这个设备号了,而是使用 LANANA 注册的 147,也就是现在 DRBD 设备的主设备号为 147。