

IT 名家讲堂

大话 Oracle RAC

集群 高可用性 备份与恢复

张晓明 编著

- 【 RAC初体验 】
- 【 什么是RAC集群 】
- 【 Oracle Clusterware 】
- 【 RAC原理 】
- 【 存储方案——ASM初步 】
- 【 RAC维护工具集 】
- 【 HA与LB 】
- 【 备份 】
- 【 恢复 】
- 【 Flashback家族 】
- 【 RAC和Data Guard 】
- 【 RAC和Stream Replication 】
- 【 深入ASM 】
- 【 性能与RAC 】

 人民邮电出版社
POSTS & TELECOM PRESS

IT 名 人 堂

大话 Oracle RAC

集群 高可用性 备份与恢复

张晓明 编著

人 民 邮 电 出 版 社
北 京

图书在版编目 (C I P) 数据

大话Oracle RAC: 集群、高可用性、备份与恢复 / 张
晓明编著. —北京: 人民邮电出版社, 2009. 4
ISBN 978-7-115-20415-8

I. 大… II. 张… III. 关系数据库—数据库管理系统,
Oracle RAC IV. TP311.138

中国版本图书馆CIP数据核字 (2009) 第020831号

内 容 介 绍

本书以 Oracle 10g 为基础, 对 Oracle RAC 进行了全面的介绍和分析。全书分为两个部分, 共 14 章, 第一部分是集群理论篇, 这部分从集群基础知识入手, 通过分析集群环境和单机环境的不同, 介绍了集群环境的各个组件及其作用, 以及集群环境的一些专有技术, 包括 Oracle Clusterware、Oracle Database、ASM、Cache Fusion 等。第二部分是实践篇, 每一章都针对 RAC 的一个知识点展开讲解, 包括 Oracle Clusterware 的维护、HA 与 LB、备份、恢复、Flashback 家族、RAC 和 Data Guard 的结合使用、RAC 和 Stream 的结合使用, 最后对 ASM 进行深入介绍, 并给出性能调整的指导思想。

本书按照“发现问题→解决问题→实践与理论相结合”的方式进行介绍, 首先对现实问题进行分析, 然后提供合适的解决方案, 最后自然地引出 Oracle 中的理论知识点, 这种讲解方法能够有效地降低阅读难度, 帮助读者更好地掌握相关技能。

本书可以作为数据库开发人员、数据库管理员、数据库初学者及其他数据库从业人员的工作参考手册, 也可以作为大中专院校相关专业师生的参考用书和相关培训机构的培训教材。

大话 Oracle RAC——集群 高可用性 备份与恢复

◆ 编 著 张晓明

责任编辑 杜 洁

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号

邮编 100061 电子函件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

北京鑫正大印刷有限公司印刷

◆ 开本: 787×1092 1/16

印张: 30.5

字数: 746 千字

印数: 1—3 500 册

2009 年 4 月第 1 版

2009 年 4 月北京第 1 次印刷

ISBN 978-7-115-20415-8/TP

定价: 65.00 元

读者服务热线: (010)67132692 印装质量热线: (010)67129223

反盗版热线: (010)67171154



序

16 年前，当我第一次接触 Oracle 数据库时，Oracle 数据库对我来说就像一位带着面纱的美女，神秘而又憧憬。看着我们的 DBA 花了两三天的时间，在键盘上输入几百条甚至上千条命令，经过几次失败，才成功地在一台 HP G30 小型机上安装成功时，我就想也许还是做一个数据库开发员更适合我。那时的 Oracle 还没有图形化安装界面，所有操作都是通过命令行完成的。作为开发人员的我们不敢也不允许对数据库做数据增、删、改、查之外的其他操作。数据库的安装、配置、建表、建索引、增加用户等也只有 DBA 才能完成。

但在今天，情况已经完全不一样了，以前只有 DBA 能完成的工作现在大部分都可以由开发人员或其他非 DBA 人员来完成。大部分的开发人员都有在自己的 PC 机上安装 Oracle 数据库的经历，运行 Oracle Universal Installer 和运行其他普通的图形界面应用程序已经没什么区别，用户甚至可以不用提供任何参数，只是点击几下鼠标即可完成，整个过程可以不超过 30 分钟。但这并不意味着我们不需要 DBA，就像波音 747 飞机上的驾驶员，虽然 747 的自动导航系统可以控制飞机从起飞、巡航到降落，但我相信没有一个乘客会说不需要飞行驾驶员了，相反我们会提出更高的要求，因为我们需要更安全、更舒适的飞行。同样，DBA 也是如此。

我从事 IT 工作已近 20 年，还从来没像今天这样关心数据库的安全和可靠性，因为数据库已经成为我们一切业务的核心。我们公司电信企业提供业务服务，这就要求我们的数据库一年 365 天不间断地可靠运行，任何一种中断都是不能接受的。每天数据库要处理上亿次的交易，每天数据量接近 8GB，并还在不断增长中。为提高数据库的可用性、扩展性和可靠性，我们的 DBA 团队尝试了不同的技术，最后采用了 Oracle RAC。经过整个团队的努力，现在我们已经成功地在我们的生产环境部署了 Oracle RAC。张晓明先生，既是本书的作者，也是带领我们 DBA 团队成功实施 Oracle RAC 的技术领头人。我相信不久，随着 RAC 技术越来越成熟，会有越来越多的公司选择 RAC。作为 DBA，就像作者在书中写的那样：“对于还从没接触过 RAC 的 DBA 来说，现在该是未雨绸缪的时候了。”

在书中，作者也遵循了著名的 Oracle 大师，Oracle 公司副总裁 Thomas Kyte 提倡的 “*Test thing yourself and base decision on proven examples*”。他从一个简单的实施实例出发，循序渐进、由浅入深地讲解了 Oracle RAC 的实施细节、RAC 的工作原理和实施、维护过程中可

能碰到的问题和解决技巧。我相信所有的 DBA 都可受益匪浅。

IBM 创始人 Thomas J. Watson 曾说过: “*Though have been father of every advance since time began, ‘I didn’t think’ has cost the world millions of dollars*”, 我想这本书能带给你的不仅是知识和技术, 更是一种不同的思维和观点。我希望每位读者能从书中得到更多的思考, 并为这个世界节省更多的 “millions of dollars”。



林镇武 副总裁
北京无限新锐网络科技有限公司

前言

从 2008 年 2 月开始动笔到定稿出版，这本书的编写几乎花费了我 10 个月的时间，大大超出了最初 3 个月的计划。这个过程虽然不能说寝食难安，但其中的酸甜苦辣也是如人饮水，冷暖自知，如果一开始就知道要投入这么大精力，我真怀疑自己是否有勇气开始动笔。

回顾自己学习 RAC 的经历，我对它又爱又恨。因为对技术本身的痴迷加上高“薪”技术的诱惑，所以爱；因为没有能力掌控，所以恨。时至今日，我仍然不敢说自己已完全掌握了 RAC，RAC 的学习过程是对我所有学习过的 IT 技术的大复习，包括 OS、Network、Storage、Database 等，甚至可能还有很多领域我从未涉足。

相信很多对 Oracle 有一定经验的 DBA 和我有同样的感觉，RAC 比普通的 Oracle 更难入门。不仅因为比比皆是的晦涩艰深的术语，也不仅因为它覆盖的技术领域太广，更主要的是可用的参考资料太少。我翻遍了所有能够获得的书籍资料，但是这些资料都是从很基本、很琐碎的细节开始，从下到上（Bottom-Up）的学习方式是虽然看到了“树”，但是看不到“林”。虽然无树不成林，技术细节总是必要的，但过于琐碎的分解给我造成更大的迷惑，到底 RAC 的总线是什么？GCS、GES、GRD、PCM 这些东西是如何堆砌在一起的？

这段痛苦经历促使我决定以自己的方式讲述 RAC 的故事，如果大家能在这个旅程中豁然开朗，对我而言将是一件很有趣又备感欣慰的事。

本书特点

本书最大的特点是侧重于理论讲述和实战演练，尤其是对理论的剖析有一定的深度，并通过大量完整的案例来论证这些理论。DBA 和开发人员特点不同，对于开发人员来说，从某种角度上，强调创造力决定高度。但是 DBA 要求的是稳定压倒一切，DBA 必须去适应自己所使用的产品，而不能期待产品去适应自己，更没有机会去改变产品，因此了解这个产品的工作原理是非常必要的。

现在网络发达，遇到问题时，有论坛、BBS 各种渠道可以利用。相信大家都有类似这样的体会：“哇，这家伙怎么能想到这样做？怎么敢这样做？”。其实这些高手的所谓“特技”都不是偶然，也不是贸然下手，只是他对产品有比你更深刻的认识，知道问题的前因，知道行动的后果，所以看起来他可以“天马行空、为所欲为”，而你只能“亦步亦趋、如履薄冰”。因此，我个人更强调对基础的掌握，而不要迷信有什么技巧或捷径。请相信只要能在一个领域持续上数年，你就在这个领域里能成为专家。

采用这种方式还有一个原因，RAC 覆盖知识面广、资料难觅也是众所周知的。时至今日，

我不敢妄谈精通，只是尽可能地把所知中规中矩地记录下来，不敢随意发挥，希望能够抛砖引玉，并对大家有所启发和帮助。

本书适用读者

大话 Oracle RAC，书如其名，本书不是一本 Oracle 数据库的基础入门书籍，而是关于 Oracle RAC 的入门指导。

这本书适合于初、中级数据库管理员和数据库开发人员，但是本书不会特别讲述什么是 SGA，什么是数据文件，什么是字典视图，什么是日志，也不会专门介绍如何创建表空间、用户等。如果你对上面这些名词、操作都非常陌生，那么这本书不适合你，请先夯实单实例的基础（可以参考下面的“学习资源推荐”）再来看这本书。

本书假定读者已经脱离了入门阶段。如果你对 Oracle 单实例数据库有一定基础，并做过类似 DBA 工作，但是由于条件所限，一直对 RAC 敬仰如滔滔江水，渴望拥有却不知如何下手，那么这本书就非常适合你。如果你需要在最短时间内充电 RAC 相关知识，并且敢在简历上写下“有一定 RAC 经验”，那么这本书就是为你而准备的。

可能你已经有了有一定的 Oracle 使用经验，比如做过些安装、管理、备份操作，对 Oracle 数据库有了一定感性认识，但是并没有完整的理论框架，那是最好的，你可以通过这本书学习到 Oracle 最核心的理论基础，并看到这些理论在单实例和 RAC 环境下的各种应用，不过还是建议你找一本基础入门的书籍放在手边，方便速查。

学习资源推荐

- http://www.oracle.com/pls/db102/portal.all_books，Oracle 10.2 官方文档，最权威的文档，适合系统学习。
- <http://www.oracle.com/technology/index.html>，OTN，资料新、覆盖面广，尤其适合快速入门、图文并茂；缺点是分类方式不太适合中国用户的习惯。
- <https://metalink.oracle.com>，Metalink，各种深度的文档一应俱全，也是所有补丁的发源地，不过用户需要申请账号。
- <http://www.itpub.net>，这是很多 Oracle DBA 开始成长的地方。
- <http://www.eygle.com>，这是 Eygle 的个人网站，Eygle（Oracle ACE Director）是国内 Oracle 数据库领域中的知名资深人士。
- <http://asktom.oracle.com>，大名鼎鼎的 AskTom。
- <http://www.databasejournal.com> 和 <http://www.oracle-base.com> 这两个站点程度类似，经常会有针对某专题的系列文章，适合初学者。

本书结构与内容

本书共 14 章，分为集群理论篇、实战篇两大部分。

工欲善其事，必先利其器，学习 RAC，环境是最重要的。所以第 1 章“RAC 初体验”

将引导读者在一台 PC 机上搭建出一个 2 节点的 RAC 环境，并在建好之后快速感受 RAC 两大亮点——负载均衡、高可用性。这一章内容虽然是介绍安装，但并不是简单的单击“上一步”、“下一步”按钮，其间渗透了许多作者的体会和心得，所以即使有安装经验的读者，也不妨做一次快速翻阅，相信也会有所收获。

从第 2 章至第 5 章主要是基础理论介绍，第 2 章首先概述了集群的概念、分类，并在其基础上对 RAC 架构做一个快速浏览。接下来的 3 章，每一章都讨论了 RAC 的一个组件。

如果对 RAC 环境做一个解剖，可以发现集群软件架构可以分为 3 个组件：Clusterware Manager（集群件）、Distribute Lock Management（DLM，分布式锁管理器）、Oracle 实例。这 3 个组件在 Oracle 的发展历史各不相同。Oracle 在版本 9 之前一直是依赖厂商的集群件产品，比如 Sun 的 Clusterware。直到 Oracle 9 才开始提供自己的集群件产品 Cluster Manager，但是只支持部分平台。到了版本 10，这个产品已经演变成一个功能完整、支持所有平台的独立产品 Oracle Clusterware，这个产品不仅支持 RAC，而且还能为其他非 Oracle 软件提供高可用支持。第 3 章专门介绍 Oracle Clusterware。而第 7 章会演示如何利用这个产品搭建一个 Web 集群。

无论 Oracle 还是 IBM、微软的数据库产品，无论单实例数据库还是集群数据库，它们核心功能都是一样的——事务处理，并在保证数据完整的前提下提供最大的并发支持。大家都知道，数据库系统是通过“锁机制”来实现的，而在集群环境下，使用的是“分布式锁机制”。其实不仅仅是数据库，所有支持多用户并发的系统都需要某种“锁机制”，因此无论“锁”还是“分布式锁”都不是 Oracle 所特有的，它是所有这些系统的共性。

Oracle 单实例的锁机制包括两种：Lock 和 Latch，Oracle RAC 同样也离不开这两种锁机制，但是 Oracle RAC 又多了一种新的锁——分布式锁，这种锁的管理机制叫作分布式锁管理器（Distributed Lock Manager，DLM）。Oracle 集群最初使用的是操作系统的 DLM，但是操作系统的 DLM 仅适用于数量很少的资源的分布式管理，比如对文件，对于 Oracle 数据库动辄 G 级的数据量，这种 DLM 就非常不适合。所以 Oracle 开发了自己的 DLM 机制，在 RAC 之前的 OPS 中，这种技术叫作 PCM（Parallel Cache Management），到了 RAC，这个技术改名为 Cache Fusion。可以说，RAC 的学习、管理、优化都是围绕着 Cache Fusion 进行的。不过与 Clusterware 不同，Oracle 并没有把 DLM 独立为一个产品，而是把 DLM 功能集成到数据库内核中，如果安装过程中检测到 Clusterware 的存在，就会激活内核的 DLM 功能。

第 4 章围绕着“锁”在单实例和 RAC 的使用，详细讲述了 RAC 原理。这一章是全书最重要的部分，也是作者花费心血最多的章节，希望读者详细研读。

以“锁”为总线来学习 RAC，这个灵感来自于我在雅虎工作期间主持的一次内部培训，也正是这个灵感造就了本书。

当时雅虎的一个系统要从 SQL Server 转移到 Oracle，大家希望能对 SQL Server 和 Oracle 的区别有所了解，于是我就主持了一次内部培训。当时对培训主题的选择是煞费苦心，如果从两个产品的媒体发布数据、销售数量做比较，对于技术人员显然没有任何意义；如果单从产品功能、用户友好性来比较，也没有任何说服力。在否定了若干方案后，我开始思考“数据库最本质的功能是什么？”，思索的结果就是最终选定“并发和锁”作为比较的主题。通过介绍两种产品的“锁机制”是如何实现的，以及各自的实现方法对并发能力的影响，从而达到比较的目的。这个讲义（PPT 格式）可以在我的 Blog（<http://space.itpub.net/75321>）找到。

需要说明的是，这个讲义是以 Oracle 9i 和 SQL Server 2000 为基础，其中关于 Oracle 9i 的部分对 Oracle 10g 仍然有效，但是 SQL Server 2000 部分对其他版本是否适用，就需要根据具体情况来定。

这次内部培训之后，借助这个灵感，我对 RAC 知识重新梳理，终于找到了学习 RAC 最有效的方法（至少我个人这样认为），也最终产生这本书。

第 5 章带领读者认识 ASM，这是 Oracle 10 推出的自动存储管理系统，这个系统同时综合了集群文件系统和卷管理器两方面功能，也是 Oracle 10 RAC 的首选存储方案。

本书的第 6 章到第 12 章是实战部分，每一章都是一个独立的讨论主题。

Oracle 10g 中的管理工具被大大加强，提供了基于 Web 界面的管理方式。这些工具虽然简单好用，但对于初学者来说恰恰也是缺点，首先这些工具隐藏了背后的运作机制，其次对于这些工具本身的学习也会分散读者的精力，何况在某些现场根本没有图形界面可以使用。所以，我更推荐读者在学习阶段还是使用“笨”方法：手工输入命令+观察输出结果。在有一定的经验以后，再去使用各种图形工具，更何况 Oracle 中需要记忆的命令数量也是非常有限的，差不多一章的内容就可以全部覆盖。

因此，第 6 章把 RAC 环境中可能用到的命令进行整理，并按照一种便于记忆的方式进行归类。但这一章并不是命令手册，很多命令会有些较深入的分析。当然在以后的章节学习中，读者可能还会需要反复查阅这一章。

第 7 章介绍 RAC 最主要的亮点“HA 和 LB”，第 8 章介绍 RAC 的备份，第 9 章介绍恢复技术中的完全恢复、不完全恢复，第 10 章介绍其他恢复技术，包括数据块恢复、Flashback 功能家族介绍。

Oracle 的高可用家族除了 RAC 还有两个兄弟：Data Guard 和 Stream Replication，相对于 RAC，这两种方案都比较容易上手，也是目前企业常用的 HA 方案。但是简单好学并不是代表功能会打折扣。这三兄弟每个既能独当一面，也可以互相配合使用，提供更强大的功能。但是，这三个兄弟针对的问题并不一样，适用场合也不尽相同。因此学习 Oracle 的高可用方案时，必须搞清楚这些方案的区别，最贵的并不一定是最适合的。因此，第 11 章比较了 RAC 和 Data Guard 的区别，以及二者如何搭配使用。第 12 章解释了 RAC 和 Stream Replication 的区别和合作。

第 13 章是对 ASM 的深度介绍，其中 ASM Alias 部分很重要，对于用户可能遇到的问题给出了具体的解决办法。

第 14 章主要介绍了调优方面的内容，但只介绍了 SQL 调整策略以及 RAC 环境下要考虑的特殊内容。之所以只选择这些内容，是因为 Oracle 性能调整是一个非常大的主题，经典话题包括主机、内存、网络调整，SQL 语句调整、等待事件调整等，新话题可以讨论 Oracle 10g 新添的若干工具 AWR、ASH、ADDM 等。如果在一章中把这些内容都一网打尽，只能是蜻蜓点水、点到为止，与其这样不痛不痒地堆砌，还不如有针对性地深入讨论几个重点。需要更多了解性能优化内容的读者，可以参考其他资料和文档。

本书声明

(1) 除了特别声明外，本书实例中使用的操作系统是 Red Hat AS 4.0 Update 4，使用的数据库是 Oracle 10 R2 版。

(2) 部分工具软件的下载地址。

Oracle Database、Oracle Clusterware、ASM Lib 驱动

<http://www.oracle.com/technology/global/cn/software/index.html>

Red Hat Advanced Server

<https://www.redhat.com/apps/download>

(3) 作者在编写本书过程中，以“够用就好”为原则，尽量覆盖到 RAC 所有知识点。但所有观点都出自作者的个人见解，疏漏、错误之处在所难免，欢迎大家指正。读者如果有好的建议或者学习本书过程中遇到问题，欢迎到作者的博客（<http://space.itpub.net/75321>）留言进行探讨，或者发送邮件到 mingyan926@hotmail.com，希望能够与大家一起交流和进步。

感 谢

在此，感谢芮玉奎先生，他在内容结构的编排方面给我很多帮助，感谢 eygle，正是他的推荐，才促成我与人民邮电出版社的这次合作。在本书编写过程中，还得到了很多朋友的支持和帮助，包括清华大学的焦丽静、IT168 的贝贝、阿里巴巴的李红星、邹大鹏，Oracle 公司的 rebecca、bryanxu，爱康网的 cindy，还有公司领导林镇武，同事史季强、邝俊彪等，在此一并感谢。

在看得见的地方学习知识，在看不到的地方学习智慧。同时也祝愿大家在 Oracle RAC 的学习道路上顺风顺水。

张晓明
2009 年 2 月

目 录

第1部分 集群理论篇

第1章 RAC 初体验	2	2.3 RAC 集群	78
1.1 本书使用环境	3	2.3.1 存储层	79
1.1.1 硬件环境	3	2.3.2 网络层	79
1.1.2 软件环境	4	2.3.3 集群件层	80
1.1.3 本书使用的环境	6	2.3.4 应用层	81
1.2 如何在 PC 机上搭建 RAC 环境	8	2.4 小结	81
1.2.1 需要下载的软件	8	第3章 Oracle Clusterware	82
1.2.2 安装过程	9	3.1 Clusterware 和 RAC 的关系	82
1.3 任务列表	28	3.2 Oracle Clusterware 组成	83
1.4 规划阶段	28	3.2.1 磁盘文件	83
1.4.1 确认主机名和 3 个 IP 地址	29	3.2.2 Clusterware 后台进程	89
1.4.2 存储方案选型	29	3.2.3 网络组件	91
1.5 实施阶段	30	3.3 Clusterware 的日志体系	95
1.5.1 主机配置	31	3.4 小结	95
1.5.2 安装 Oracle Clusterware	40	第4章 RAC 原理	96
1.5.3 安装 Oracle Database	51	4.1 数据库基本原理	96
1.5.4 配置 Listener	55	4.1.1 并发访问和数据一致性	97
1.5.5 创建 ASM	59	4.1.2 事务和隔离级别	98
1.5.6 创建数据库	63	4.1.3 Oracle 支持的隔离级别	99
1.6 客户端测试	71	4.2 Oracle 单实例的并发控制机制	100
1.6.1 客户端配置	71	4.2.1 Lock	100
1.6.2 体验 Failover	72	4.2.2 数据记录的行级锁	101
1.6.3 体验 LoadBalance	73	4.2.3 Latch	105
1.6.4 修改归档模式	74	4.2.4 Latch 和 Lock 对比	106
第2章 什么是 RAC 集群	76	4.2.5 进一步理解	106
2.1 集群分类	76	4.3 RAC 下的并发控制	109
2.2 集群环境的特殊问题	76	4.3.1 DLM 中资源和锁	110
2.2.1 并发控制	77	4.3.2 Non-Cache Fusion 资源	111
2.2.2 健忘症 (Amnesia)	77	4.3.3 Cache Fusion 资源	112
2.2.3 脑裂 (Split Brain)	77	4.3.4 GRD (Global Resource Directory)	114
2.2.4 IO 隔离 (IO Fencing)	78		

4.3.5	PCM Lock	114
4.3.6	Cache Fusion	116
4.3.7	RAC 并发控制总结	121
4.4	RAC 架构	121
4.4.1	SGA 的变化	122
4.4.2	后台进程的变化	122
4.4.3	文件	124
4.4.4	SCN	126
4.4.5	Cache Fusion、GCS、GES	127
4.5	RAC 和 Clusterware 的交互	127
4.5.1	Clusterware 层	128
4.5.2	RAC 层	129
4.6	小结	132
第 5 章	存储方案——ASM 初步	133
5.1	Shared-Disk 和 Shared-Nothing 架构	133
5.2	Oracle 10g RAC 的存储方案介绍	134
5.3	ASM 架构	134
5.3.1	ASM 实例	135
5.3.2	文件	136
5.4	配置 ASM	136
5.4.1	安装位置	136
5.4.2	创建 ASM 磁盘	136
5.4.3	配置 ASM 实例	144
5.4.4	创建磁盘组	146
5.4.5	在数据库中使用 ASM 的 磁盘组	147
5.5	ASM 实例：将数据库迁移到 ASM	147
5.5.1	试验说明	147
5.5.2	操作步骤	148
5.5.3	最后验证	155
5.6	小结	156

第 2 部分 实 战 篇

第 6 章	RAC 维护工具集	158
6.1	Oracle Clusterware 工具集	158
6.2	节点层	158
6.3	网络层	159
6.4	集群层	161
6.4.1	crsctl	162
6.4.2	OCR 命令系列	167
6.5	应用层	176
6.5.1	crs_stat	176
6.5.2	onsctl	180
6.5.3	srvctl	186
6.5.4	恢复	192
6.6	小结	196
第 7 章	HA 和 LB	197
7.1	什么是高可用性	197
7.2	Failover	198
7.2.1	Client-Side Connect Time Failover	198
7.2.2	TAF (Transparent Application Failover)	198
7.2.3	Client-Side Failover 和 TAF 的 对照试验	199
7.2.4	Server-Side TAF	202
7.3	Oracle Clusterware HA 框架	210
7.3.1	术语介绍	210
7.3.2	配置命令	213
7.3.3	完整实例	219
7.4	LoadBalance	223
7.4.1	Connection Balancing	223
7.4.2	利用 Service 分散负载	227
7.5	测试 LoadBalance	228
7.5.1	通过 Listener 日志区分路 由来源	228
7.5.2	测试方法	229
7.5.3	测试过程	230
7.6	小结	234
第 8 章	备份	235
8.1	概述	235
8.1.1	备份	235
8.1.2	恢复	236
8.1.3	RMAN 工具	238
8.2	使用 Flash Recovery Area	238
8.2.1	配置 Flash Recovery Area	239
8.2.2	Flash Recovery Area 的空间 监控	239
8.3	使用 RMAN	240
8.3.1	RMAN 工具的使用方法	240

8.3.2 RMAN 配置.....	242	10.7.4 命令和视图	332
8.4 完全备份和增量备份	248	10.8 Flashback Drop	333
8.4.1 完全备份	249	10.8.1 Tablespace Recycle Bin	333
8.4.2 增量备份	251	10.8.2 Flashback Drop 操作	335
8.4.3 其他 RMAN 命令	254	10.8.3 Recycle Bin 的维护	336
8.5 RAC 的备份实例	261	10.9 Flashback Query 和 Flashback Table	337
8.6 小结	264	10.9.1 Oracle 9i 的 Flashback Query	337
第 9 章 恢复	265	10.9.2 Flashback Version Query	339
9.1 为什么 Oracle 数据库能够恢复	265	10.9.3 Flashback Transaction Query	354
9.2 日志	267	10.9.4 Flashback Table	356
9.2.1 日志线程 (Redo Thread)	268	10.9.5 UNDO Retntion	359
9.2.2 RAC 的联机日志	268	10.10 小结	360
9.3 恢复种类	269	第 11 章 RAC 和 Data Guard	361
9.3.1 Instance Recovery	269	11.1 Data Guard 介绍	361
9.3.2 Media Recovery	269	11.1.1 Data Guard 架构	362
9.3.3 Crash Recovery	269	11.1.2 日志发送 (Redo Send)	362
9.3.4 Online Block Recovery	274	11.1.3 日志接收 (Redo Receive)	365
9.4 介质恢复	274	11.1.4 日志应用 (Redo Apply)	367
9.4.1 完全恢复	275	11.1.5 数据保护模式	367
9.4.2 不完全恢复	280	11.1.6 自动裂隙检测和解决	368
9.4.3 恢复到单实例	294	11.2 RAC 和 Standby 配置实例	369
9.5 小结	302	11.2.1 RAC Primary 和 Single Standby	370
第 10 章 其他恢复技术	303	11.2.2 RAC Primary 和 RAC Standby	381
10.1 数据块恢复	303	11.3 小结	394
10.2 如何检查数据块一致性	304	第 12 章 RAC 和 Stream Replication	395
10.2.1 使用初始化参数	304	12.1 Stream 的工作原理	395
10.2.2 DBV 工具	305	12.2 Data Guard 和 Stream 对比	396
10.2.3 Analyze 命令	307	12.3 RAC Stream 配置实例	397
10.2.4 RMAN 工具	307	12.3.1 Single (Source) 和 RAC (Target)	398
10.2.5 dbms_repair 包	308	12.3.2 RAC 对 RAC 的复制	404
10.3 模拟数据块不一致	309	12.3.3 STRMMON 工具	425
10.4 数据块不一致处理办法	313	12.4 小结	427
10.4.1 收集信息	313	第 13 章 深入 ASM	428
10.4.2 设计恢复方法	314	13.1 非 ASM 的存储结构	428
10.4.3 恢复操作	316	13.2 ASM 存储结构	429
10.5 数据块恢复实例	317		
10.6 Flashback 家族介绍	326		
10.7 Flashback Database	326		
10.7.1 Flashback Database 架构	326		
10.7.2 启用 Flashback Database	327		
10.7.3 Flashback Database	329		

13.3 ASM 文件系统	430	13.7.3 ASM 和 TTS	453
13.3.1 ASM Metadata	431	13.8 ASM 的限制	456
13.3.2 Physical Metadata	431	13.9 小结	457
13.3.3 Virtual Metadata	433	第 14 章 性能与 RAC	458
13.3.4 Disk Recovery	434	14.1 RAC 的若干特点	459
13.4 条带化 (ASM Striping)	435	14.2 AWR	461
13.4.1 ASM 文件模板 (ASM File Template)	435	14.2.1 启用 AWR	462
13.4.2 ASM 别名 (ASM File Alias)	436	14.2.2 修改 AWR 配置	462
13.5 RDBMS 和 ASM 之间的交互	445	14.2.3 产生 AWR 报告	462
13.6 ASM 的实例恢复	446	14.3 ASH	464
13.7 ASM 和 OS Filesystem 之间交互	447	14.4 SQL 调整策略	466
13.7.1 dbms_file_transfer 包	447	14.4.1 SQL 语句的执行过程	466
13.7.2 RMAN 的 CONVERT 方法	453	14.4.2 SQL 性能调整的基本方法	467
		14.4.3 PL/SQL 和 SQL	471
		14.5 小结	473

1

第 1 部分 集群理论篇

第 1 章 RAC 初体验

第 2 章 什么是 RAC 集群

第 3 章 Oracle Clusterware

第 4 章 RAC 原理

第 5 章 存储方案——ASM 初步

第1章 RAC 初体验

长期以来，业界流行着这样一个“潜规则”：以是否有 RAC 经历来衡量 DBA 水平。虽然这种评价方法并不准确，但从某种程度上也反映了 RAC 在 Oracle 体系金字塔的塔尖地位。造成这种现象原因有很多，作为一种高端解决方案，RAC 的实施成本一度也是很高的，除了硬件成本、OS License、Oracle License 成本，还包括第三方的 Clusterware 软件成本、人力成本，所以 RAC 案例一直集中在电信、银行、制造等大企业，中小企业鲜有问津，绝大多数的 DBA 很难有机会接触到这种环境。所以 RAC 的神秘很大程度上要归功于“高墙锁深宫”，技术本身的因素反而是次要的。

不过最近几年，随着芯片技术的发展，硬件成本逐年降低，物美价廉的 PC Server 越来越深入人心。而在软件方面，Linux 已经不再是一款“教学”软件，越来越多的企业已经在核心业务上选择 Linux 作为首选平台，Oracle 尤其重视在 Linux 平台上的应用，它也是第一个支持 Linux 平台的商业数据库。从 Oracle 本身来说，随着 Oracle 10g 版本的推出，Oracle 开始提供 RAC 的“一站式”解决方案，包括针对存储的 ASM、针对集群的 Oracle Clusterware。并且这些软件都可以从 Oracle 网站免费下载。因此在 Oracle 10g 上部署 RAC 成本非常便宜——不需要昂贵的小型机和专有的操作系统，普通的 PC Server + Linux 就足以胜任；不需要采购第三方的集群软件、存储软件、不需要三方的人力资源，只需要一个 DBA 就可以完全搞定。

因此，近一两年来，PC Server + Linux + Oracle 10g RAC 这种方案在中小企业有逐渐流行的趋势。对于还从没接触过 RAC 的 DBA 来说，现在该是未雨绸缪的时候了，说不定哪天你的技术总监就会请你分析一下 RAC 的可行性。希望当你对着老板那双充满期待的眼睛时，能够底气十足地说出一二三来，而不是汗流浹背、支支吾吾。

作为一个“过来人”的经验而言，RAC 本身并不难，只是涉及的知识面太广。仅仅一个安装过程就需要 OS 配置、存储配置（如果没有专门的存储工程师）、网络配置，最后才是 Oracle 自己的配置，基本上 RAC 的学习过程就是对 IT 知识的全面复习，对不同的用户可能是学习不同的新知识。

“纸上得来终觉浅，绝知此事要躬行”，所有技术都不是通过读书就能学会的，必须要亲身实践，才能把抽象理论变成自己的才能。要想掌握 RAC，有一个练习环境是非常重要的。因此这一章我将带着你“从无到有”搭建一个 RAC 出来，然后给出一些简单的使用案例。这一章的定位是帮助读者对 RAC 有一个感性认识，而后面的每一章节都是针对一个具体的知识点。因此这一章中除了必要的解释外，更多的是给出其他相关章节的链接，比如“参见 1.1.1 小节”，读者可以自行决定阅读顺序。

1.1 本书使用环境

工欲善其事，必先利其器，学习 RAC，环境是最重要的。

本书使用的环境是一个两节点的 RAC，和在第 1.2 节搭建的虚拟环境不一样，这是一个真实的环境。图 1-1 描述了这个环境的拓扑结构，在这幅图中，涉及的硬件（Hardware）用 Hn 表示，涉及的软件（Software）用 Sn 来表示，其中 n 不是代表数量，而是代表种类。

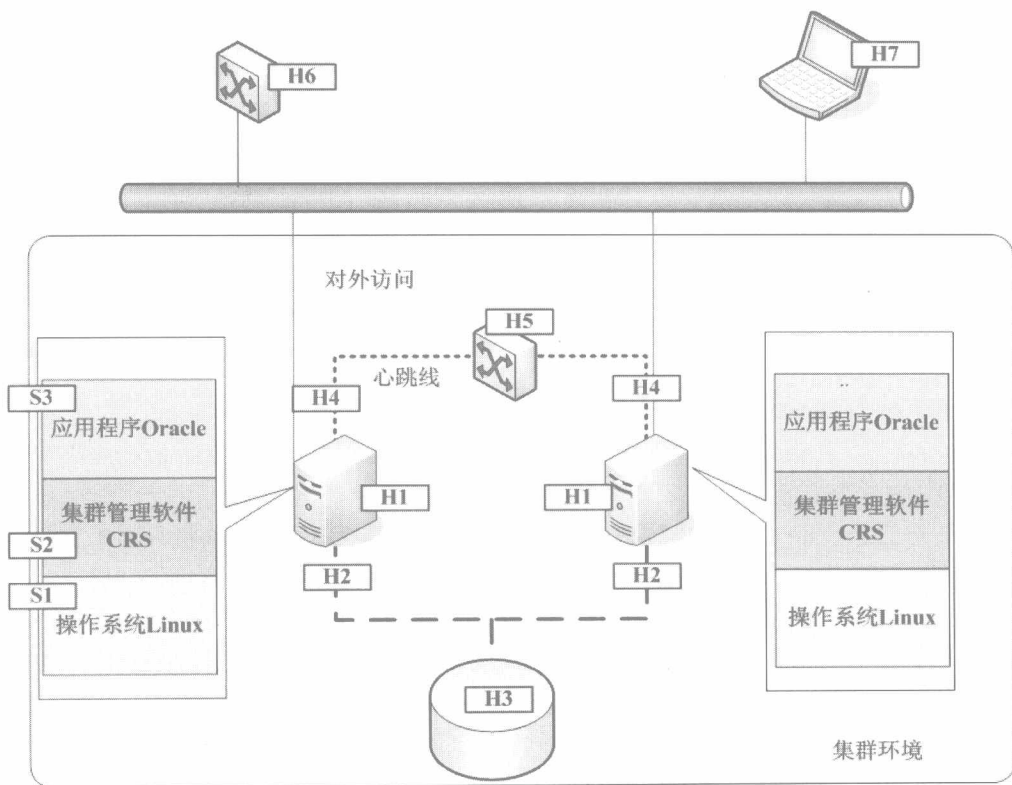


图 1-1 RAC 环境拓扑图

1.1.1 硬件环境

概括地说，整个 RAC 的硬件环境包括主机、共享存储、网络互联设备，下面我们详细讨论每个组件。

1. 节点（主机）

在这个双节点的 RAC 环境，共有两台主机，即图 1-1 中的两个 H1。在 RAC 术语中，主机通常叫作“节点”，二者是一个概念，后面章节有时会混用这两个词。

每个主机的硬件配置应该一样。每个主机至少需要两块物理网卡，图 1-1 中的 H4 就代表两块网卡，它所跨越的两条线代表两条网线。