HZ BOOKS
华章IT

数据库
技术丛书

Oracle RAC 核心技术详解

高斌 著

Oracle RAC Core Technology

- Oracle RAC领域的扛鼎之作，一线资深技术专家多年工作经验结晶
- 深入解析Oracle RAC的内部实现原理和核心技术，并通过大量实际案例揭示Oracle RAC的最佳实践

机械工业出版社
China Machine Press

数据库
技术丛书

Oracle RAC
核心技术详解

Oracle RAC Core Technology

高斌 著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

Oracle RAC 核心技术详解 / 高斌著 . —北京：机械工业出版社，2015.10
（数据库技术丛书）

ISBN 978-7-111-51830-3

I. O… II. 高… III. 关系数据库系统 IV. TP311.138

中国版本图书馆 CIP 数据核字（2015）第 240846 号

本书由 Oracle RAC 资深专家结合其多年的研究和工作经验撰写而成，不仅系统介绍了 RAC 的原理，同时也包含了一些测试和实际案例来帮助读者更好地理解 RAC 的工作原理，以及如何在实际工作中运用这些知识解决问题。作者以通俗易懂的语言，从基本原理着手，并用简单的测试和实例来验证这些原理，再通过具体的实际案例来详细介绍如何用这些理论知识解决实际问题。本书不仅可以帮助你更好地理解 Oracle RAC 这套复杂的系统，也能够了解 Oracle 的原厂工程师是如何分析和处理问题的。

本书分为两部分，共 13 章。第一部分（第 1 章 ~ 第 9 章）对集群管理软件进行了详细介绍，涵盖关于 10g CRS 和 11g GI 的核心技术的详细介绍，包含集群的核心组件 CSS、CRS 以及 11gR2 新增的 OHAS 组件和守护进程，同时还包含诊断集群问题常用的诊断工具介绍。此外，由于从 11gR2 版本开始，ASM 已经变成了集群管理软件的一个组件，作者在这部分还介绍了和集群、数据库相关的 ASM 知识。第二部分（第 10 章 ~ 第 13 章）讲解 RAC 数据库的核心技术——内存融合（Cache Fusion），详细介绍了内存融合技术的核心组件以及工作原理，并且还对 RAC 相关的性能调优知识进行了介绍，同时作者对 RAC 中的连接管理和工作负载管理也做了详细解释。

Oracle RAC 核心技术详解

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码：100037）
责任编辑：张梦玲　　责任校对：董纪丽
印　　刷：北京瑞德印刷有限公司　　版　　次：2015 年 10 月第 1 版第 1 次印刷
开　　本：186mm × 240mm　1/16　　印　　张：31.75
书　　号：ISBN 978-7-111-51830-3　　定　　价：99.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88379426　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　88379649　68995259　　读者信箱：hzit@hzbook.com

Preface 序 言 一

2012 年，我和 Allen[⊖]在红木城第一次见面，当时他来总部参加 RAC Pack 培训，同时也和我们一起处理了美国客户的一些问题。虽然只有短短的两个星期，但我能感觉到 Allen 是一名出色的工程师，他对 RAC 的理解给我留下了深刻的印象。在之后的时间里，我仍能接触到 Allen 处理过的一些关键问题，这进一步印证了我对他的看法。本书覆盖了 RAC 的核心技术，而且也讲解得非常准确和精彩，的确是一本难得的好书。相信读者在阅读了本书之后，不仅能够对 RAC 的核心技术和原理有所了解，在技术上有所收获，而且能够了解 Oracle 的技术部门是如何分析解决问题的，并在具体的工作中实践、应用本书提到的技术和分析方法。

杨依岩

Oracle RAC/Clusterware 全球技术领导小组组长

⊖ Allen 即为本书作者高斌。

序 言 二 Preface

行者

Oracle 全球技术支持中心在大连成立到现在，8 年多的时间过去了。从最初的质疑：大连这个地方有做技术的吗？到后来的肯定：大连不错呀，能够找到懂技术的人。这 8 年来，有人离开，有人坚守，有人离开又回来。无论怎样，每个人都在历练中不断成长，亲历的人知道，这一切与起点无关，与行动有关。这 8 年来，好几个人说要写本关于 Oracle 数据库的书，Allen 是将此想法付诸行动的人，而我钦佩有执行力的人。

人们常常羡慕自己所不能的。自己技术不行，就会很羡慕技术能力强的人，喜欢与他们相处，沾沾灵气，所以当 Allen 邀请我为他的新书写序时，我诚惶诚恐。作为一个没有机会成为技术高手的人，缺乏专业的评判力，无法为这本书摇旗呐喊，但有幸与做技术的人相处久了，也会有些心得，现写下来，与大家分享。

聪明与勤奋

勤奋的人对知识如饥似渴，学习起来不惜气力。对初学者来说，这是个优秀的品质，当你分不清什么是重要的、什么是无关紧要的时候，每一个细节都需要关注，要用能吃苦的精神换取量的积累。但对熟练者来说，用力过猛会是瓶颈，忙着积累知识，就会少了思考的时间。因此熟练者要恰到好处地用力，将知识转化为领悟，最终养出处理问题的直觉。

聪明的人对任何新知识都可以很快上手。因为擅长少花力气多办事，常能找到捷径，快速地完成手头的工作，这类人的思维能力强，但耐不住反复实践的寂寞，他们敬佩内功深厚的大师，却偏偏对奇思巧技没有免疫力，学得快却学不深。

用力过猛的问题不难解决，常走捷径却是顽疾。聪明的人需要审视自己的本能反应，反应太快不好，需耐下性子放慢动作，一帧一帧地审视自己，找到症结并加以改变。好在聪明伶俐的人只要过了这一关，前途则不可限量，因此这样的努力也值得了。

专家说的？依据是什么？

在 Oracle 公司工作的好处是有机会接触各路专家。最初，觉得专家很厉害，什么都知道。渐渐地，接触的专家多了，虽然对他们敬佩依旧，但是对专家的盲目崇拜却被彻底地根除了。因为专家也会犯错误，专家在解决问题的时候也没有什么独门绝技。在遇到未知的问题时（这是常有的事），他们都是利用已有的知识，通过系统的分析，一步一步地定位问题。

我们习惯寻找答案，懒于思考为什么。我们迷恋知识，畏于亲身实践。一个优秀的工程师不能仅仅满足于见过多少问题，知道多少答案，还要有能力解决未知的问题。学会在已知和未知的边界做有效的思考，要常常问自己，写这篇文档的作者是怎样想到这个解决方案的？这一解决方案的依据是什么？文档中的方案为什么能解决用户的问题？有没有其他的替代方案？这个设计是如何工作的？它为什么要这么设计呢？等等。要学会区分事实和观点，提出观点时要以事实为依据，无论是专家的建议，还是自己的猜测，都要经得起验证。

别想太多，去做就好

常常有工程师对培训寄予厚望，幻想着能获得一本武功秘籍，以成为绝世高手，可惜技术的积累不是速成的。参观台北故宫博物院时，最震撼我的藏品是镂空同心象牙球，它每层的雕花各不相同，每个空心球都可以转动。这个精品是三代人用几十年的时间成就的，这和技术的积累过程是同一个道理。

在某个领域有很高造诣的人均由特殊材料造就，而组成这种特殊材料的基本元素平白无奇，无非是大量的学习和实践，就像钻石不过是由碳元素组成的单质晶体，而在高压高温的环境下成为了金刚不坏之身。

专家不是教出来的。专家需要在实践中成长，需要承担有挑战的工作，拥有反复试错的勇气，通过有效的反馈不断地纠错。在成为专家的过程中，你也许会一路高歌，也许会进两步退一步，也许会在某段时间内停滞不前，无论你把自己想象得过于聪明还是笨拙，时间都会不徐不疾地前行，打磨着浮躁的心，就像我在 Allen 身上看到的一样，他在持续的自我进化中，越来越优秀。

王兰

Oracle 全球软件服务（数据库和集成系统）总监

前　言 Preface

实时应用集群（Realtime Application Cluster，RAC）——Oracle 服务器技术的核心产品之一。该产品能够提供优秀的高可用性、灵活性、可扩展性、可测量性。十多年来，已经有成千上万的企业使用 RAC 作为企业核心数据库，并将关键应用部署到 RAC 数据库中，而且这种趋势已经不可逆转。

想一想，自己已经和 Oracle 打了十几年交道，尤其是近几年，每天都要处理 RAC 问题，也算积累了一些心得。于是就有了写本专门介绍 RAC 核心技术的书的想法，接着，这本书就在几个月之后出现了。

作为 DBA，尤其是负责维护 RAC 系统的 DBA，如果希望了解 RAC 的工作原理、各个特性的功能以及管理方法、常见问题的诊断方法，那么本书无疑是很好的选择。如果你是 Oracle 数据库技术爱好者，本书也可以帮助你很好地了解 Oracle 的集群技术和“神奇”的内存融合技术（Cache Fusion）。当然，本书对应用程序开发人员或者系统设计人员也会有所帮助，因为 Oracle 的很多设计理念非常先进，很多地方是值得借鉴的。由于本书着重介绍 Oracle RAC 的核心技术，其中包含比较多的原理和工作机制描述，如果读者具有一定的 Oracle 数据库基础知识和性能调优知识，会更容易理解本书的内容。当然，作者也尽量使用朴实的语言来解释相关的原理和工作机制，希望能够满足更多读者的需求。

Oracle RAC 是由集群管理软件和数据库软件两部分构成的：集群管理软件（在版本 10g 和 11g 的第 1 版中叫 CRS（Cluster Ready Service），从 11g 第 2 版开始叫 GI（Grid Infrastructure））负责提供集群的基础服务并管理所有的应用程序资源；数据库软件负责运行 Oracle RAC 数据库并为应用程序提供服务。本书内容也正是按照集群管理软件、数据库软件的顺序来编写的。在本书大部分章节中，作者首先会对涉及的原理部分进行介绍并辅以简单的测试和说明，之后通过一些实际的案例对所介绍的工作原理进行验证，并运用我们学到的知识解决实际问题。另外，虽然 11gR2 版本的集群管理软件和 10gR2 版本相比发生了很大的改变，但是很多核心的概念、架构仍然得到了延续，本书也是先从 10g 版本开始，之后介绍 11gR2 版本的性能提升和改变。

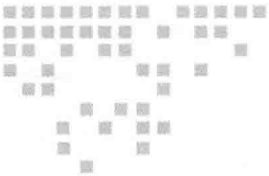
本书共分两部分，一共 13 章。第一部分：第 1 章～第 9 章，包含了对集群管理软件的详细介绍，读者可以在这部分找到关于 11g GI 和 10g CRS 核心技术的详细介绍，以及集群核心组件 CSS、CRS 和 11gR2 新增的 OHAS 组件与守护进程的描述，同时还有诊断集群问题的常用诊断工具的讲解。最后，由于从 11gR2 版本开始 ASM 已经变成了集群管理软件的一个组件，所以作者在这部分内容中还介绍了和集群、数据库相关的 ASM 知识。第二部分：第 10 章～第 13 章，包含了 RAC 数据库的核心技术——内存融合，作者用大量的篇幅介绍内存融合技术的核心组件以及工作原理，并且还对 RAC 相关的性能调优知识进行了介绍。最后，作者还讲解了 RAC 中的连接管理、工作负载管理知识。

目　录 Contents

第一部分 Part 1

集群管理软件

- 第 1 章　Oracle 集群技术介绍
- 第 2 章　安装 Oracle 集群
- 第 3 章　11gR2 集群新增组件
- 第 4 章　11gR2 集群新增的集群守护进程
- 第 5 章　CSS 部分
- 第 6 章　CRS 部分
- 第 7 章　Oracle 集群管理软件的启动顺序
- 第 8 章　集群诊断工具概述
- 第 9 章　ASM 基础



Chapter 1 第 1 章

Oracle 集群技术介绍

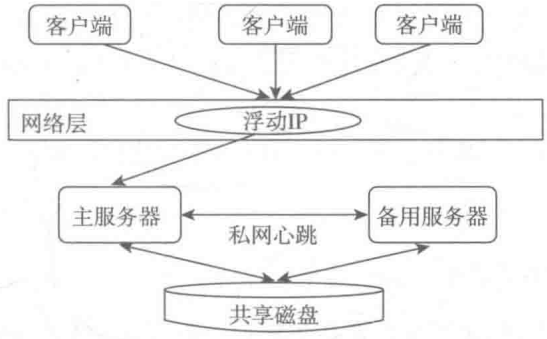
在本章，作者首先会对集群技术进行简单的介绍，之后对 Oracle 集群技术进行介绍，使读者对 Oracle 集群产品有一个基本的认识。

1.1 集群技术简介

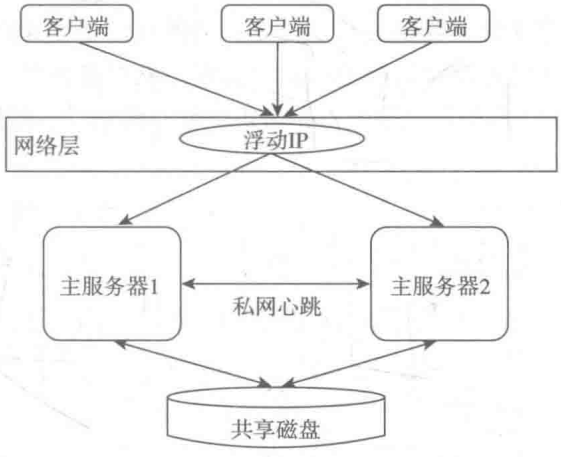
集群是一组相互独立的、通过高速网络互联的计算机，这些计算机（节点）通过单一系统的模式加以管理，并且向外提供统一的服务。从客户的角度看来，集群就像一台独立的服务器。根据集群的功能特点可以大概将其分为以下 3 类。

1.1.1 高可用集群

高可用集群（High Availability Cluster）简称 HA cluster。高可用的含义是最大限度地提供可用的服务。从集群的名字上可以看出，此类集群实现的功能是保障用户的应用程序持久、不间断地提供服务。而高可用集群还可分为 cold failover 和 hot failover。cold failover 就是我们经常提到的主 / 备模式，集群通常存在两个节点，其中一个作为主节点向外提供服务，备用节点随时准备在主节点出现问题时向外提供服务。主 / 备节点都同时访问相同的共享磁盘。当然，cold failover 的主 / 备节点彼此也进行心跳操作来维护集群的一致性；cold failover 同样也提供浮动 IP 功能来保证对应用程序的透明性。大家可以通过下图了解典型的 cold failover 系统的基本架构。



hot failover 一般是指集群的每个节点都处于活动状态，每个节点都能够分担应用程序负载，当某一个节点出现问题之后，其他节点可以分担问题节点的负载，而这些变化对应用程序来说全部都是透明的。读者可以通过下图了解典型的 hot failover 系统的基本架构。



Oracle RAC 就是非常典型的 hot failover 集群。

1.1.2　负载均衡集群

负载均衡集群也是由两台或者两台以上的服务器组成的，分为前端负载调度和后端服务两个部分。负载调度部分负责把客户端的请求按照不同的策略分配给后端服务节点，而后端服务节点是真正提供应用程序服务的部分。与 HA cluster 不同的是，在负载均衡集群中，所有的后端服务节点都处于活动动态，它们都对外提供服务，分摊系统的工作负载。Oracle RAC 也是很典型的负载应用集群。

1.1.3　高性能计算集群

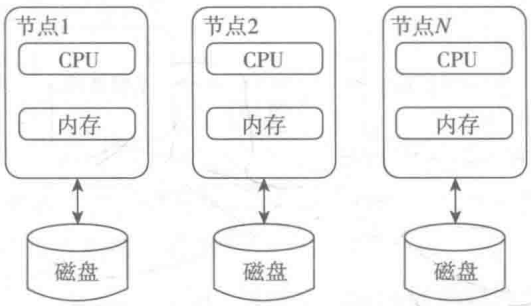
这类集群能够提供单个计算机所不能提供的强大计算能力，能进行大规模的数值计算和数据处理，并且倾向于追求综合性能。从某种程度来讲，可以认为高性能计算机集群类似于

一台虚拟的超级计算机，它能够进行庞大而且高速的计算工作，而这些工作是单一计算机所无法完成的。高性能计算机集群通常采用的方式就是并行技术。Oracle RAC 也提供了非常优秀的并行计算能力。

如果从数据共享的角度对集群分类的话，还可以将其分成 share-nothing 和 share-everything 两种结构。

1.1.4 share-nothing 结构

在这种结构下，集群中的每一个节点在物理上都是独立的，而且它们访问的磁盘也是相互独立的，而工作负载会分布到集群的不同节点上。这种结构的好处在于集群的结构很简单，节点和节点之间的交互和依赖很少，而且避免了动态锁管理（Dynamic Lock Management，DLM）、多节点并发控制所带来的性能瓶颈。但是，share-nothing 也有很大的缺点：由于节点和节点间相互比较独立，所以在进行集群规划和工作负载划分的时候需要非常小心。由于彼此访问的磁盘也是互相独立的，所以当某一个节点出现问题时，可能会引起某些数据无法被访问的情况，即使磁盘之间可以对数据进行镜像处理，成本也会非常大，这会严重影响系统的性能和可用性，而且此类集群的可测量性、可扩展性非常差。下图描述了比较典型的 share-nothing 集群的基本架构。



1.1.5 share-everything 结构

这种架构的集群和 share-nothing 结构的正相反，这类集群中的所有节点都会访问共享的磁盘，所以很多时候可以把这种架构称为“共享磁盘架构”，这种架构的最大特点就是：通过高速的存储局域网将多个节点连接在一起，实现对共享磁盘的并发读、写操作。share-everything 集群架构成功地避免了 share-nothing 集群的缺点。这种集群架构能够实现非常好的高可用性、负载均衡、可测量性和扩展性。但是，share-everything 也有一些问题：由于所有节点会向相同的磁盘进行并发读、写操作，所以需要一套控制机制（通常称之为 DLM）来对读、写进行一定程度的串行化（也就是需要加锁），从而解决并发读、写操作下的冲突并保证数据的一致性。Oracle 集群技术就采用了 share-everything 结构。下图是比较典型的 share-everything 集群的基本架构。