来自Oracle专家的见解

Broadview
www.broadview.com.cn

深入理解 Oracle Exadata

Expert Oracle Exadata

实施Oracle在存储层进行SQL处理的
突破性解决方案

[美] Kerry Osborne
Randy Johnson
Tanel Põder 著

ACOUG
All China Oracle User Group
中国 Oracle 用户组

黄凯耀
张乐奕
张瑞 译

由Oracle Exadata研发部门前性能架构师Kevin Closson担任本书技术审校

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

Apress

深入理解 Oracle Exadata

Expert Oracle Exadata

[美] **Kerry Osborne , Randy Johnson , Tanel Põder** 著

黄凯耀　张乐奕　张瑞　译

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

本书深入地诠释了 Exadata 的各项特性，如智能扫描、混合列式存储、存储索引、智能闪存、IO 资源管理；系统地介绍了如何安装、配置和管理 Exadata；完美地阐述了 Exadata 的等待事件、性能监控和调优方法；详细地剖析了计算节点和存储节点的内部原理；全面地分享了作者们在实际项目中所获得的宝贵经验，如怎样进行大数据的高效移植、Exadata 上的一些常见误区、数据库资源管理，等等。本书是实践经验的总结和升华，可读性极强，不仅有对 Exadata 深入的研究，还有对它们优雅的展现，它将带领读者进入 Exadata 的殿堂。

版权贸易合同登记号图字：01-2011-7103

图书在版编目（CIP）数据

深入理解 Oracle Exadata /（美）奥斯本（Osborne, K.）等著；黄凯耀，张乐奕，张瑞译. —北京：电子工业出版社，2012.7
书名原文：Expert Oracle Exadata
ISBN 978-7-121-17489-6

Ⅰ. ①深… Ⅱ. ①奥… ②黄… ③张… ④张… Ⅲ. ①关系数据库系统－数据库管理系统 Ⅳ. ①TP311.138

中国版本图书馆 CIP 数据核字（2012）第 143244 号

策划编辑：张春雨
责任编辑：贾 莉
印 刷：中国电影出版社印刷厂
装 订：三河市皇庄路通装订厂
出版发行：电子工业出版社
北京市海淀区万寿路 173 信箱 邮编 100036
开 本：787×980 1/16 印张：37.5 字数：900 千字
印 次：2012 年 11 月第 3 次印刷
印 数：5001～6500 册 定价：99.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

原作者中文版序

First of all we'd like to say hello to our Chinese readers out there. A year ago at Oracle Open world 2011 we were asked if we would be interested in helping the Chinese translation teamtranslate this book into the Chinese language. Today we're very excited to see it go to press. One of the unexpected benefits of this process was getting yet one more technical review of its contents. In a work of this size and complexity there are bound to be a few mistakes that somehow make it through the editing process. The Chinese translation team provided outstanding feedback and helped us correct and clarify where needed.

It has been a year since the English version of this book went to press and over two years since Oracle began shipping Exadata V2. We've been amazed (but not too surprised) by the speed at which Exadata has become a world-wide phenomenon. One of the unique challenges in writing a book about such new technology was dealing with rapid changes to the product itself. Fortunately Oracle appears to have spent this time stabilizing and refining the Exadata platform rather than expanding on its feature set. There have been surprisingly few visible changes to the feature set and today this book continues to be the definitive resource for learning Exadata. The examples and labs illustrated in these pages still work today and provide valuable insight to the reader. We hope you find this book helpful as you come to understand the inner workings of intelligent storage and why it is such a leap forward in database technology.

—— Kerry, Randy, Tanel

首先我们要向这本书的中国读者问好。在一年前的 Oracle Open World 2011 上，我们被问及是否愿意帮助中国的翻译小组将这本书翻译成中文，而今天，我们非常兴奋地看到中文译本即将出版。在这个过程中，我们的意外收获是这次翻译又再次为本书的内容做了一次技术审校，就本书的内容及复杂度而言，在写作过程中出现一些错漏在所难免，而中文翻译小组的出色反馈帮助我们纠正和澄清了这些错误。

至今，这本书的英文版出版已有一年，而离 Oracle 发布 Exadata V2 也已超过两年。我们被 Exadata 在全球走红的速度所震惊（但是并不太惊讶）。写作一本如此崭新技术的书籍的一大挑战就是要面对产品本身的快速变化。幸运的是，Oracle 将时间花在了 Exadata 平台的稳定性及精致性建

设上，而并没有急于增加新功能，因此时至今日，Exadata 并没有很多显著的新增特性，而本书也仍然是学习 Exadata 的权威资料。书中展示的例子和实验仍然有效，并为读者带来了宝贵的见解。我们希望你能从本书中获益，希望本书可以帮助你了解智能存储的内部工作机理，也可以帮助你理解为什么称其为数据库技术的一个飞跃。

——Kerry，Randy，Tanel

本书序

2008 年 9 月，Oracle CEO Larry Ellison 在甲骨文全球用户大会（OOW）上宣布了软件及硬件集成一体化的数据库机——Oracle Exadata Database Machine（以下简称 Exadata）。Exadata 的推出不但震撼了业界、吸引了全球数据库专家的关注，也引起了 Oracle 数据库“粉丝”们对其技术的探究和追逐，而且也使得 Exadata 成为了网络搜索的热点 IT 词汇。

也正是由于 Exadata 在技术架构上的自我创新、功能上的丰富增强和性能上的极大优化，使得 Exadata 在市场推出 3 年多的时间里，得到了全球用户的广泛认可。目前，Exadata 的全球部署已经超过了 1000 台、用户遍及 67 个国家的 23 个行业。Exadata 数据库机已成为甲骨文 30 多年发展史中最成功的新产品。在甲骨文公司的云计算解决方案中，Exadata 作为数据库云服务器是 Oracle PaaS（Platform-as-a-Service，平台即服务）平台的基础构件，成为企业搭建云环境、构建云支撑平台的基石。

随着 Exadata 产品和技术的不断更新和广泛使用，无论是 Exadata 技术爱好者还是我们的广大用户，都非常希望有一本深入介绍 Exadata 的技术书籍，让读者不仅能从理论概念上更能从实际应用上来更好地理解和把握 Exadata 的技术机理，循序渐进地探索其内部的技术细节。

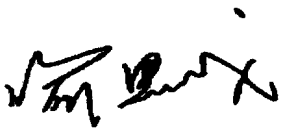
而 *Expert Oracle Exadata*（written by Kerry Osborne, Randy Johnson and Tanel Põder）一书的问世无疑是“雪中送炭”，本书的三位作者 Kerry、Randy 和 Tanel 都是大家熟知的 Oracle 技术领域内的大师，他们不但有自己的 Oracle 技术博客，还为全球客户实施和部署 Oracle Exadata 产品，积累了丰厚的实战经验，对 Exadata 技术的精髓有切身的体验和理解。本书一经出版便成为 Amazon 网站上受人关注的书籍。

我相信作为国内读者，更希望看到本书的中文版译著。机缘巧合的是，我们国内的三位译者，也是国内 Oracle 数据库“粉丝”们熟知的技术专家——甲骨文公司的黄凯耀（Kaya）、阿里巴巴的张瑞（Jacky）、云和恩墨的张乐奕（Kamus）。

三位译者出于对 Exadata 技术的热爱和把控，以及对读者的尊重，在翻译的过程中，不是简单地对照原文完成语句的翻译，而是在译文中仔细斟酌每一句话的含义，按照中文阅读习惯加以解释，增加了阅读的流畅性、可理解性，避免了生涩的直译。特别值得一提的是，对于一些原文较为晦涩的地方，三位译者根据自己的理解增加了“译者注”，我相信这在目前大多数技术书籍的译文中并不常见。

也正如三位译者在各自的译者序中所写，从2011年8月份开始着手翻译起，三位译者之间及与原著作者之间关于本书翻译的邮件沟通，来来回回将近500封，同时三位译者对原文中的一些错误之处也进行了一并纠正，真可谓为本译作“锦上添花”。

我相信这本汇集了原作者、译者及其他多位大师的技术、经验和点评的专业书籍，一定能让国内有此技术爱好的读者沉浸于书中、感同身受，有所收获。我也相信读完本书，能让你感受到“欲穷千里目，更上一层楼”的境界。



Oracle 全球副总裁　　喻思成

译者序 1

2011 年 7 月，我曾给中国 Oracle 用户组（ACOUG）做过一次有关 Exadata 的演讲，Exadata 的高性能在国内的 Oracle 社区中引起了不小的反响。其实，Exadata 在 2008 年底已经推出，自那时起，我们组（Oracle Real World Performance Group） 就一直从事着 Exadata 上的性能测试与项目开发工作。在我的博客上也有一些相关的文章涉及 Exadata，如 2010 年 5 月前后发表的系列文章 Exadata V2 架构分析，但那都是些零碎的片断。毋庸置疑，Exadata 在国内还是陌生的，但作为 Oracle 数据库的未来发展方向，让更多的人熟悉 Exadata 无疑是一件很有意义的事情，甚至很多兄弟团队都鼓励我们组写作一本关于 Exadata 及其性能调优的书。

2011 年 8 月，一个机缘巧合的机会，博文视点的张春雨老师联系上我，希望我可以参与到这本书的翻译工作中。初看之下，这本书的内容非常丰富，是对 Exadata 的一个全面系统的介绍。于是我们一拍即合，这本书的翻译工作就此掀开序幕。另外两位译者是阿里巴巴的张瑞（HelloDBA）与云和恩墨的张乐奕（Kamus）。张瑞是阿里巴巴的架构师，负责数据库性能优化与应用架构改进，研究软硬件结合的数据库解决方案。张乐奕是云和恩墨的技术总监，Oracle ACE Director, 也是国内知名的 Oracle 技术专家。于是，我们的中文翻译小组正式成立。

2011 年 8 月中旬，我们第一次接触到了这本书的电子版。开始的日子是忙碌的，每天工作之余，翻译上几页，这是一个锻炼人耐力的过程。出差途中，不管是在飞机上还是火车上，翻译这件事儿也帮助我打发了一些无聊时光。我学会了一个道理，积少成多，贵在坚持。不过也有难熬的时候，特别对于晦涩的章节，但最终理清作者的思路时，也会感到欢欣鼓舞，即使几小时已经倏忽间流走了。而印象最深的，是与张瑞和张乐奕对里面技术点的讨论，几百封邮件的往来帮助我们一起澄清了对原著诸多晦涩段落的理解。这是一个通力合作的过程，它是痛苦的，又是快乐的，我们只有尽情享受其中。

三位作者的文风其实各有特点。Tanel 是全球 Oracle 社区的著名人物，他的 Oracle Session Snapper 出名已久，性能优化的相关章节主要由他执笔，这两章充满睿智，里面完美地体现了 Tanel 对性能调优的理解，既有对全局的系统方法论的阐述，也有对每个性能指标含义的具体说明，还有对 SQL、存储节点的调优监控思路。翻译的时候，我总有一种心有戚戚焉的感觉。这些论述即使在非 Exadata 平台也有非常大的借鉴意义。

Randy 则具体关注 Exadata 的管理，包括了对数据库资源管理器、配置和恢复、存储节点与计算节点的详细剖析。对资源管理的详细阐述必将会刮起一阵清新之风，在消退 Oracle 资源管理器神秘感的同时，让读者也掌握了如何构建合适的资源管理模型。这些章节是对 Exadata 整体架构的

高可用性与高可配置性的极佳体现。

Tanel 和 Randy 还合作了关于如何移植的主题，这也是充满实践性的真知灼见的一章，里面提到的方案在实施中会有很好的借鉴意义。

Kerry 则专注于对 Exadata 相关特性的描述。通过大量的例子深入浅出地介绍了 Exadata 的主要特性，同时还大量地挖掘了各特性后面的细节。

由三位各有所长的作者联合执笔，终于成就了这本恢宏巨作。

我负责翻译的章节主要包括第 7、8、9、11、12 章，以及前言部分。第 7、8、9 章由 Randy 执笔，是关于系统资源管理、系统配置和恢复的内容。第 11 和 12 章由 Tanel 执笔，是关于性能优化的章节。

对翻译工作，不得不提的一点是对名词术语的翻译。坦白讲，以前阅读译著时，一个令人难受的地方就是原来很熟悉的英文术语与译者的翻译联系不起来。这每每让我有阅读原著的冲动。我们在本书的翻译过程中在尽量避免这个问题，对于大家所熟知的英文术语，我们尽量不做翻译。当不翻译真的很影响阅读的流畅性时，我们才会进行翻译，如“Grid Disk”，这个词汇在第 14 章就出现了 140 多次，不翻译会很影响阅读效果，所以虽然我们平时都直称“grid disk”，本书中我们还是把它翻译成了“网格盘”（或许“网格盘”一词也会因此流行起来呢）。当然，我们会在前面的术语翻译约定中列举出来。

2011 年 2 月，我们开始进行本书的翻译校对工作。这又是三位译者头脑碰撞的日子，我们努力清除原文理解上的每一个障碍，并积极与三位原作者就我们所发现的众多瑕疵进行了邮件讨论并加以改正，同时对于里面的重点和难点加入了独具特色的译者注。相信本书将是 Expert Oracle Exadata 全球的最新译本，同时也是独具特色的中文版本。如果读者在阅读原版时产生了疑惑，而会想起参考此中文版本，那将是我们莫大的荣幸。

看着眼前厚厚的著作，思绪万千，开始翻译的日子似很遥远，又历历在目。在这里，要感谢 Oracle Real World Performance Group，感谢杨中对翻译工作的支持，感谢帮助我做了认真细致校对工作的李昕、曲卓、董志平、陈长青、孙笑盈，感谢 Oracle 的众多同事提供的帮助和指导，特别是来自 COE（Center Of Excellence）的许向东。感谢喻总在百忙之中为本书做序。最后，还要谢谢我的妻子，她对我加班加点的翻译工作不仅毫无怨言，还从一个学习者的角度，校对了其中的一些章节。谢谢你们！

黄凯耀于深圳

2012-4-9

译者序 2

这本书的翻译计划是从 2011 年 8 月份开始的，据我所知，最早是博文视点的编辑“侠少”找到阿里巴巴的张瑞（Jacky）和甲骨文的黄凯耀（Kaya），然后 Jacky 再找到我。

实际上，我个人开始想要翻译这本 Exadata 技术书籍倒是从更早的时候就开始了，这本书在 Amazon 上的发行日期是 2011 年 8 月 9 日，其实早在 2011 年 2 月份就已经有另外一本关于 Exadata 性能的书籍（Achieving Extreme Performance with Oracle Exadata，作者全部是 Oracle 公司员工），但是论作者的知名度，仍然是本书更受人关注。最早知道这本书是从本书联合作者 Tanel Poder 的个人技术 Blog 中，那是 2011 年 3 月份，Tanel 发文说已经可以在 Apress 网站上购买新书 *Expert Oracle Exadata* 的 Alpha 版本，Tanel 是全球最受人尊重的 Oracle 技术专家之一，而一本技术书籍可以预先购买 Alpha 版本也是很稀奇的事情，再加上 Exadata 正是当今 IT 界的“当红炸子鸡”，理所当然这本书非常值得期待。在 2011 年 4 月份，我个人跟某出版社联系过，表达了如果该书可以引进中国，那么我很愿意组织人手进行翻译的工作，对方的回复是正在谈版权，之后没有消息。然后，Tanel 在 6 月份发文说，本书已经即将定稿，再之后，就是 8 月份，该书正式发售。而在正式发售的当月，博文视点就开始寻找中文版本的译者，可以说是非常迅速。而对于版权的猜测，那一定是博文视点拿到了版权，而某出版社失利了。:-D

以上的情况，让我在收到 Jacky 的邀请以后，毫不犹豫地接受了工作，无论工作如何繁忙，我都愿意让这本书的中文译者里有我的名字，这对于我而言可以说是一种荣幸。2011 年 8 月 17 日收到这本书的 PDF 电子版（当然后来又收到纸质版），从 8 月份开始，Kaya、Jacky 和我都迅速地投入了翻译的工作，在整个过程中，通过不断地沟通，我们按照每个人的经验和对各个章节的熟悉程度以及感兴趣程度，大致是均分了各个章节。我负责翻译的章节是第 1、2、4、6、13、16 章，原本我给自己定下的计划是每两周翻译一章，那么最快可以在两个月内完成翻译，再加上校稿，本来计划在 3 个月内可以完成所有的翻译，也就是如果一切顺利，这本书的中文译本应该在 2011 年年底的时候就跟大家见面了。但是，计划永远是赶不上变化的，除了工作的繁忙和个人的懒惰，我们几个译者还都在其他方面出现了这样或那样的意外情况，导致整个翻译工作整体滞后。所幸，还不算太迟，我想在你们看到本书的时候，这个世界上应该还没有更新的 Exadata 书籍可以参考。所以，这本书仍然是迄今为止想要了解 Exadata、想要使用 Exadata、想要监控调整 Exadata 的最佳参考书籍。

Oracle Exadata 的举世瞩目，对整个数据库硬件/软件市场的震撼，在全球或者仅仅是中国国内的引人瞩目，乃至热销，都已经无须赘言。作为数据库从业者，也许你没有听过 Netezza，也许你没有听过 Twinfin，也许你没有听过 Hana，但是你一定听过 Exadata，这绝不仅仅是由于 Oracle 公司一贯的好战、勇于进攻、大力宣传的风格，而是 Exadata 确实具有独步天下的功能。也许我们不能说在经过最精细的调整以后，Exadata 在数据仓库领域与其他竞争对手相比一定具有绝对的优势，但是，不要忘记，在现在这个世界里，又有多少是纯粹的数据仓库系统呢？又有多少用户愿意让 OLTP 用一套系统，而数据仓库又用另外一套系统呢？这其中的数据传输开销和系统设计复杂性的开销，如果能够消减甚至避免，那么又何乐而不为呢？Exadata 正是这样的一套软硬件一体的平台，同时支持 OLTP 类型负载和数据仓库类型负载，通过 Oracle Database 11gR2 中的资源管理器来更加精细地调控硬件资源，让两种类型的负载都能获得各自需要的资源，并顺畅执行。

如果我们抛却 Exadata 在存储节点中的软件特性，它使用的各个硬件组件并不是划时代的，无论是 Infiniband 还是 Flashcache/SSD，都已经出现很久了，在企业级市场中也已被很多用户使用，但是将这些组件放在一起，并且预先调整为一个平衡的系统（没有任何一处明显的性能瓶颈），这是划时代的。Oracle 将软硬一体机的概念推广到了开放性平台上，极大地挑战了 Teradata 的市场，用开放性的硬件+开放性的操作系统+开放性的数据库软件，构造出了一个平衡的、性能超强的平台，这同样是划时代的。

好吧，前面我们提到了“抛却 Exadata 在存储节点中的软件特性”是吗？这就好比说，把皇冠上最闪亮的那颗宝石先摘下来，别闪花了我们的眼睛。现在，我们要把这颗宝石放回去了，智能扫描（Smart Scan）、存储索引（Storage Index）、混合列式压缩（Hybrid Columnar Compression），无论哪一项软件特性都足以震撼数据处理市场，而当它们结合在一起，配合上 Oracle Database 原本就具有的高性能，再配合前面说的这个平衡的硬件架构，我们就得到了足以颠覆一切固有理念的惊人性能。在 Exadata 的 POC 现场，有客户因为实在无法接受 Exadata 展示出来的飞一般的速度而怀疑 Oracle 的技术人员在造假。这在无奈的同时，无疑也是一种自豪吧。

Exadata 的出现，颠覆了一些我们既有的数据库管理理念，但是无论如何，Exadata 中运行的是 Oracle Enterprise Linux（当然也有 Solaris，不过是 x86-64 版本，至少到目前为止，Oracle 还没有计划显示会出现 SPARC 平台上的 Exadata），Linux 上运行的是 Oracle Database 11gR2，对于所有数据库技术从业者来说，之前积累的操作系统管理知识，Oracle 数据库/RAC 管理知识都仍然适用。我们需要的只是与时俱进，将 Exadata 的特有知识点加入我们以前的知识体系中。本书是最佳的入手点，因为本书中不但有 Exadata 的特性阐述，也同样有使用经验和最佳实践。要知道本书的作者都是真正的 Exadata 使用者，而本书的技术审校者（Kevin）更是 Exadata 的性能架构师（不过，Kevin 现在已经离开 Oracle 公司，加盟 EMC，去玩 Greenplum 了）。

我唯一希望的是，大家在阅读这本中文译本的时候，不至于产生去重新阅读原著的冲动（虽然，我仍然建议大家去阅读原著），因为如果那样，那只能表示我们的翻译实在是很不适合中文读者的

理解。如果你觉得本书优秀，那么基本上可以说这是原作者的功劳，当然，我也希望你们看到我们三位译者的努力。我们在翻译完各自的章节以后，又互相审阅了其他人的章节，我们尽量斟酌每一句话的翻译，希望读起来是符合中文阅读习惯的，对于一些比较难于理解的片段（比如 Kevin 说的某些话），我们通过邮件跟作者进行了沟通以确保译文是正确体现了作者意图的，对于一些原文较为晦涩的地方，我们也根据自己的理解增加了“译者注”，我相信这也是目前大多数技术书籍的译文中并不常见的，我们甚至在想，如果译者注足够多，那么就可以出一本批注版的书籍了（:-D）。这其中，由于 Kaya 在 Exadata 中的实战经验尤为丰富，更是付出了格外的精力。你们现在看到的这本 *Expert Oracle Exadata* 的中文版，应该是全球的最新版本，因为在我们的翻译过程中，不但将本书对应英文版出版以后提交给作者的错误修订全部都更正到了本书中，而且我们还在翻译过程中发现了更多的错误，Kaya 通过邮件直接跟三位作者沟通并一一确认，最终对于确实是错误的描述也都全部做了更正。实际上，这也是本书推迟到现在才出版的原因之一。

就在今天，我重新审阅完了自己翻译的第 6 章，回顾了一下从 2011 年 8 月份开始，我们三位译者和博文视点的侠少关于翻译本书的邮件沟通，来来回回将近 300 封邮件，我相信在本书中文版最终定稿的时候，沟通邮件量一定会超过 300 封（实际上最终的沟通邮件将近 500 封）。我们扪心自问，已经尽了自己最大的努力，但是一定还会有这样或那样的不足，还望读者海涵。

最后，我要感谢我的妻子和可爱的儿子，在我工作之余的很多个深夜，我仍然在翻译此书，是我的妻子极大地包容了我，没有她的支持，没有她承担几乎全部家务和对我们年仅 1 岁多的儿子的照料，也许我的翻译进度还会拖后。谢谢你，我爱你们。感谢 Kaya、Jacky，还有博文视点的侠少，与你们关于本书翻译讨论的 500 封邮件是宝贵的财富。感谢我的大学师妹——董楠，她是《老美国志异》、《此地无人生还》、《满是镜子的房间》三本畅销书籍的译者，喜欢摇滚的朋友应该热爱这几本书籍，本书某些段落的措辞有得到她的指教。另外，我同样要感谢我所在的公司——云和恩墨的多位同事，是你们帮我承担了由于翻译工作而落下的本应属于我的工作，感谢杨廷琨（老杨同时帮助审阅了本书的第 1 章），感谢盖国强，还有帮助我审阅中文译稿的同事们——仇实、刘洋、余广宏、董禹、宋春风，译稿里面也有你们的功劳，谢谢你们。

张乐奕（Kamus）于上岛咖啡，北京

2012 年 2 月 29 日

译者序 3

2008 年，Oracle 在 OOW 上发布了与 HP 合作开发的 Exadata V1。当时，我就对 Exadata 充满了好奇，很快我就在自己的博客上写了第一篇介绍 Exadata 的文章 *Oracle Database Machine*。现在看来，文章中很多观点都是错误和可笑的，但当时介绍 Exadata 的技术资料非常少，很多观点只能来自于猜测。从那时开始，我一直保持着对 Exadata 的关注，并在一次 Oracle 介绍 Exadata 新技术的会议上，认识了 Kaya（本书的另外一位译者），这也为我们共同翻译本书埋下了伏笔。

2009 年，Oracle 发布了 Exadata V2，它不仅采用了 SUN 的硬件，更是革命性地引入了 Flash 存储，并采用智能闪存（Exadata Smart FlashCache）技术，让 Exadata 同时支持 DW 和 OLTP 应用，成为真正全能型的数据库软硬件一体机。2010 年 2 月 1 日，我在博客上发表了第二篇关于 Exadata 的技术文章《Oracle Exadata 技术浅析》，引起了大家的热烈讨论，并被广泛转载。文章中介绍了 Exadata 的新特性，并且在没有资料提及的情况下，提出 SmartScan 应该只能在特殊访问路径下（直接路径扫描）才能启用的观点。从此以后，我对 Exadata 着了迷，无奈除了一些官方文档以外，几乎没有其他任何资料，而我又没机会亲身操作 Exadata，只能通过各种途径了解 Exadata 的最新信息。

2010 年，我参加了在北京举办的 OOW 大会，第一次近距离看到了 Exadata 的真面目，也更加深刻体会到“Hardware and Software Engineered to Work Together”这句话的真正含义。从此，我开始致力于推动 Flash 存储技术在数据库领域的应用，研究软硬件结合的数据库解决方案。2011 年 10 月举办的 OTN China Tour 活动上，受 ACOUG 的邀请，我做了《软硬件结合的数据库解决方案》主题演讲，介绍了我们在软硬件结合方面的一些尝试，并且解读了 Exadata 和 Oracle 刚推出的 ODA（Oracle Database Appliance）的架构，引起了非常大的反响。

2011 年底，我终于等到了亲身体验 Exadata 的机会，Oracle 提供了一个 Exadata V2 Quarter Rack 供我们测试。我们根据在线交易网站的特点，专门设计了一个测试模型，并且先在高端的小型机和存储上进行测试，然后再在 Exadata 上运行相同的测试，以此来评估 Exadata 和传统主机存储之间的性能差异。Exadata 的表现让我们非常惊讶，性能完全超过了小型机和存储。Exadata 使用 FlashCache 技术，在读多写少的应用场景下，表现超乎想像，不仅预热速度快，而且性能与将数据全部放在 Flash 上相差无几，提供了非常好的性能价格比。不仅如此，我们还特别模拟了存储节点宕机的情况，当某个存储节点宕机时，Exadata 表示毫无影响，而且性能下降非常小。通过一周的测试，我们对 Exadata 高性能、高可用和高度灵活的特性有了更深的认识。

作为一名 Exadata 技术爱好者，一直苦于找不到一本深入介绍 Exadata 的书籍，直到偶然一次机会我发现了 *Expert Oracle Exadata* 这本书，本书的几位作者都是 Oracle 技术领域的大师，之前我

一直订阅他们的博客，了解最新的 Oracle 前沿技术，冲着它是几位大师的合著，再看看内容简介，我知道它就是我期盼已久的那本书。当时这本书还没有完成，我找到出版社希望能在本书出版后第一时间拿到，没想到侠少不仅帮我拿到了书，同时也拿到了本书的版权，并且希望我能翻译这本书，我毫不犹豫地答应了，因为我已经期待了太长时间。但是当我读了一点又开始犹豫了，因为对于 Exadata 来说我只是个初学者，而且本书的技术含量非常高，凭我一己之力很难完成这项工作，我马上想到了两个人，一位是 Kaya（黄凯耀），另一位是 Kamus（张乐奕），Kaya 来自于 Oracle RealWorld Database Performance Group，专门研究 Oracle 和 Exadata 的性能优化，我相信国内没有人比他更熟悉 Exadata。Kamus 是国内知名的 Oracle 技术专家，Oracle ACE Director，他也是 Exadata 技术爱好者。事实证明，他们就是最合适的人。

虽然我以前也经常写文章，但是翻译书籍还是第一次，深深体会到翻译工作的艰辛。因为 Exadata 是业界最新的技术，翻译的过程也是学习的过程，很多技术都是第一次碰到，需要理解并准确表达出来，这对翻译者来说是个严峻的考验。我们三个人一起翻译和讨论，不知不觉就过去了半年，来来回回发了几百封邮件，往往一个句子甚至一个词，都要讨论很多遍才能确定下来，正是在这个讨论的过程中，很多技术问题都被我们搞清楚了。我们坚信翻译不是简单的文字转换，一定要自己先搞清楚，才能翻译出来给大家。正因为如此，我们在翻译的过程中，也发现了原书中的很多错误，并且 Kaya 把我们发现的错误都整理出来发给了原书的作者，所以大家读到的应该是最新版本。

如果要说本书最难翻译的部分是什么，当属“Kevin 说”，因为 Kevin 说的要不是“技术哲学”，就是“技术原理”，有时候我们读过几遍之后，还不知道他在说什么，我们已经尽可能翻译得通俗易懂，但是仍然有很多不完美的地方，请大家谅解。另外，因为 Exadata 的技术非常新，我们在难以理解的地方都加上了译者注，希望可以帮助大家。本书中出现的术语，我们都采用中英文对照的方式，也有一些常用的 Oracle 术语我们选择不翻译，因为英文比中文更容易理解。

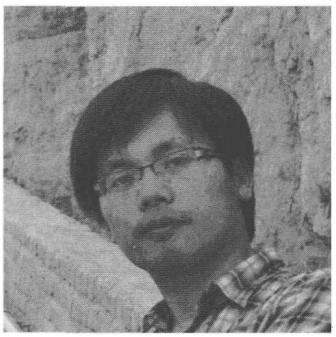
最后，我要感谢家人对我的支持，尤其是我的爱人 Emily 和宝贝 Michael，每天工作到深夜，回家后看到你们熟睡的脸庞，觉得一切付出都是值得的。还要感谢和我一起战斗过的 Kaya 和 Kamus，你们给了我太多的帮助。感谢原书的作者，正是因为你们写了这样一本伟大的著作，才有现在的中文版。最后感谢读者的支持和理解，希望你们喜欢这本书。

我翻译了本书第 3、5、10、14、15 章以及附录部分。

张　瑞

2012 年 4 月 1 日于杭州

译者介绍



黄凯耀，2006年加入Oracle，在Real World Performance Group（隶属于Oracle公司总部数据库产品管理部门）工作，担任首席软件工程师。主要从事关键客户的现场性能测试、现实客户碰到的重大问题解决、Oracle数据库的质量保证、数据库间的竞争分析等工作。特别专注于大型数据库（VLDB）在OLTP与OLAP环境下的高性能与高可扩展性的最佳实践。目前工作重点在于Oracle Exadata的性能测试与实施。乐于技术的总结与分享，个人技术博客为www.os2ora.com。

张乐奕（Kamus）

云和恩墨（北京）信息技术有限公司技术总监

Oracle ACE Director

Itpub Oracle数据库管理版/高可用版版主

ACOUG（www.acoug.org）联合创始人

OESIG（www.oesig.org）创始人

个人Blog：www.dbform.com

张乐奕，云和恩墨的联合创始人之一，致力于通过不断的技术探索，帮助中国用户理解和接触新技术，推广数据库技术应用。曾先后任职于UT斯达康、电讯盈科、甲骨文等知名企业，担任DBA及技术顾问工作。现任职于云和恩墨（北京）信息技术有限公司。

具备丰富的行业经验与技术积累，对于数据库技术具有深刻的理解。热切关注Oracle技术和其他相关技术，对于Oracle数据库RAC以及高可用解决方案具有丰富的实践经验。长于数据库故障诊断，数据库性能调优。作为社区和网络的活跃者，在公开演讲和出版方面，多有建树。

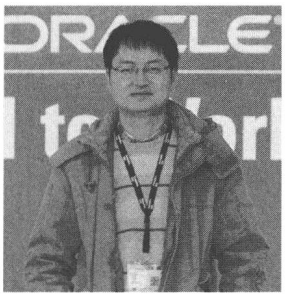
2004年2月，作为主要作者出版了《Oracle数据库DBA专题技术精粹》一书。

2005 年 6 月，作为主要作者出版了《Oracle 数据库性能优化》一书。

2007 年 12 月，获 ITPUB 论坛年度原创技术文章奖，同年 3 月，被 Oracle 公司授予 Oracle ACE 称号。

2010 年 3 月，与 Eygle 联合创立 ACOUG 用户组，目前 ACOUG 是中国最活跃的 Oracle 用户组，持续进行着技术分享。

2011 年 03 月，被 Oracle 公司授予 Oracle ACE Director 称号，同年创办中国 Exadata 特别用户组。



张瑞，网名 HelloDBA，Oracle ACE，2005 年加入阿里巴巴，数据库架构师，负责数据库性能优化与应用架构改进，主导推动了阿里巴巴数据库技术的变革。同时也是 Exadata 技术爱好者，致力于推动 Flash 存储技术在数据库领域的应用，研究软硬件结合的数据库解决方案。个人有技术博客 HelloDB.net，乐于分享数据库领域的最佳实践和研究成果，并创立了 AskHelloDBA.com 专业数据库问答社区，解答各种数据库技术问题，定期举办 AskHelloDBA 数据库技术论坛。

术语翻译约定

A

ASM 范围安全策略	ASM-Scoped Security

B

半机柜	half rack
并行子进程	slave process
部分重建	partitial reconstruction

C

CPU 量子	cpu quantum
CPU 资源短缺	CPU starvation
查询协调进程	query coordinator
出列	dequeue
串行直接路径读取	serial direct path reads
磁盘修复计时器	disk repair timer
磁盘组	disk group
存储节点	storage cell
存储节点卸载处理	cell offload processing
存储节点卸载效率	cell offload efficiency
存储节点应急过程	storage cell rescue procedure

D

DBM 配置器	DBM Configurator
单一客户端访问名字	Single Client Acess Name

F

访问方法	access method

G

高度冗余	high redundancy
告警提醒	alert notification
骨干交换机	spine switch
管理网络	management network
过量分配	over-provisioning

H

互连流量	interconnect traffic
混合列式压缩	hybrid columnar compression

J

集群感知	cluster-aware
集群软件	clusterware
计算节点	compute node
交错分配	interleaving
节点安全策略	Cell Security
节点盘	cell disk

K

客户端访问网络	client access network
块内链接	intra-block chaining

L

类别	category
类别 IORM	Category IORM