

Oracle R Enterprise: Harnessing the Power of R in Oracle Database

精通Oracle R

Enterprise大数据

在Oracle数据库中充分利用R的强大功能

将企业的大数据转化为有价值的资产！

and an Tierney (ACE Director)

著
译



清华大学出版社

精通 Oracle R Enterprise 大数据

在 Oracle 数据库中充分利用

R 的强大功能

[美] Brendan Tierney 著

张骏温 译

清华大学出版社

北 京

Brendan Tierney
Oracle R Enterprise: Harnessing the Power of R in Oracle Database
EISBN 978-1-25-958516-6
Copyright © 2017 by McGraw-Hill Education.

All Rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including without limitation photocopying, recording, taping, or any database, information or retrieval system, without the prior written permission of the publisher.

This authorized Chinese translation edition is jointly published by McGraw-Hill Education and Tsinghua University Press Limited. This edition is authorized for sale in the People's Republic of China only, excluding Hong Kong, Macao SAR and Taiwan.

Translation copyright © 2018 by McGraw-Hill Education and Tsinghua University Press Limited.
版权所有。未经出版人事先书面许可，对本出版物的任何部分不得以任何方式或途径复制或传播，包括但不限于复印、录制、录音，或通过任何数据库、信息或可检索的系统。
本授权中文简体字翻译版由麦格劳-希尔(亚洲)教育出版公司和清华大学出版社有限公司合作出版。此版本经授权仅限在中国大陆区域销售，不能销往中国香港、澳门特别行政区和中国台湾地区。
版权©2018 由麦格劳-希尔(亚洲)教育出版公司与清华大学出版社有限公司所有。

北京市版权局著作权合同登记号 图字：01-2017-5690
本书封面贴有 McGraw-Hill Education 公司防伪标签，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

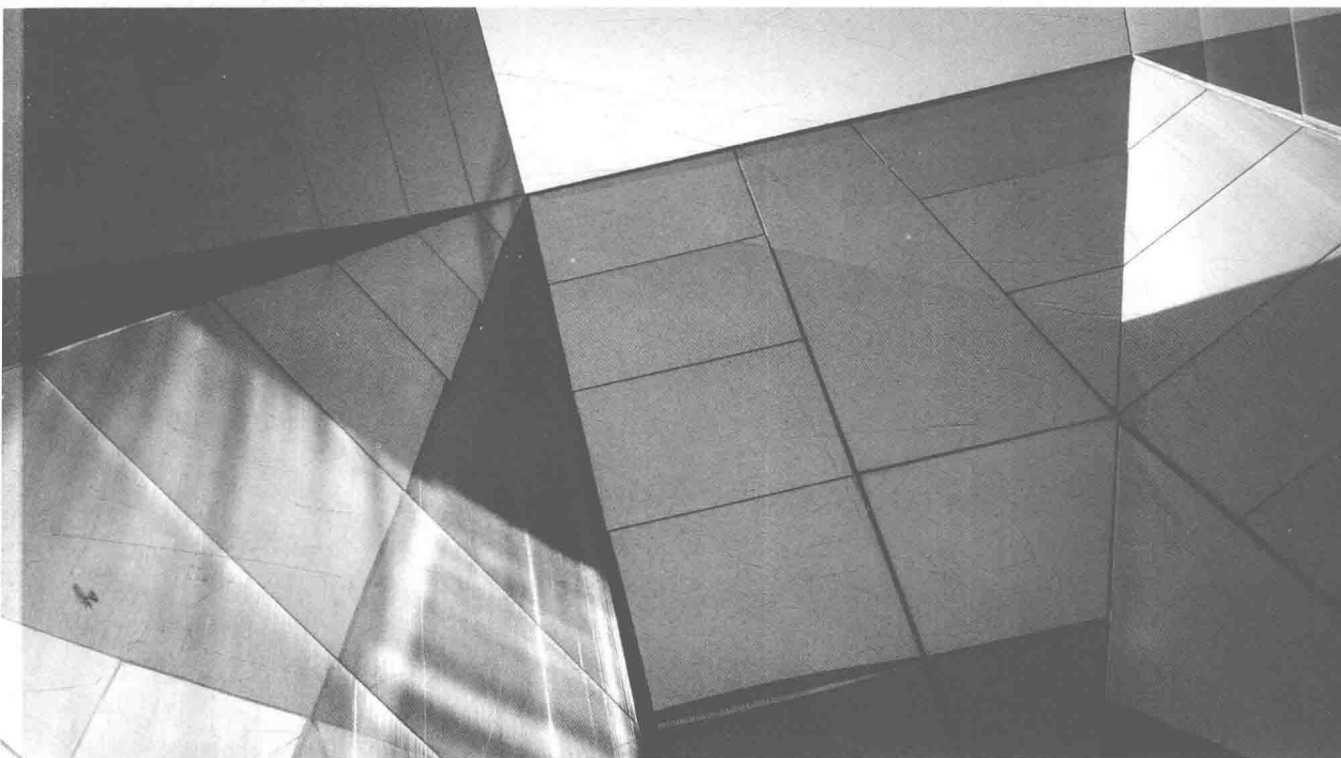
图书在版编目(CIP)数据

精通 Oracle R Enterprise 大数据 在 Oracle 数据库中充分利用 R 的强大功能 / (美)布兰登·蒂尔尼(Brendan Tierney)著；张骏温 译. —北京：清华大学出版社，2018
书名原文：Oracle R Enterprise: Harnessing the Power of R in Oracle Database
ISBN 978-7-302-49116-3
I. ①精… II. ①布… ②张… III. ①关系数据库系统 IV. ①TP311.138
中国版本图书馆 CIP 数据核字(2017)第 311925 号

责任编辑：王 军 韩宏志
封面设计：牛艳敏
责任校对：曹 阳
责任印制：杨 艳

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 装 者：北京密云胶印厂
经 销：全国新华书店
开 本：185mm×260mm 印 张：13 字 数：316 千字
版 次：2018 年 1 月第 1 版 印 次：2018 年 1 月第 1 次印刷
印 数：1 ~ 2500
定 价：59.80 元



译者序

在翻译本书的过程中，译者由衷地发出这样的感慨：技术一直沿着蜿蜒的道路发展和前行！R 语言已诞生 20 多年，但一直在专业圈子里使用，圈子以外的人对 R 语言并不熟悉。但现在情况不同了，因为大数据时代来了。“大数据”成为当今最时髦的词汇之一，不仅是 IT 专业人士，就连心灵鸡汤的制作者和传播者，也都出言必称大数据。总之，大数据风头正劲！R 技术也顺势起飞了，或者说有了更大的用武之地。

Oracle 在软件领域是当之无愧的领导者，在全球各国各大行业、企业和政府部门的关键应用中，Oracle 的数据库产品得到广泛使用，因而，能让 Oracle 公司将 R 技术融入其产品中，也足见 R 技术在当今的热度。

R 语言或者说 R 技术，最适用的领域是数据分析，其中有大量的算法、模型和图形图表工具。通常的使用模式是，数据分析人员在自己的 PC 机或笔记本电脑上安装 R 语言并使用。这种使用模式对于数据量不大的情况，是非常适用的，但当数据量很大时，这种模式就不适用了。受到 PC 机或笔记本电脑的硬件资源的限制，在对大量数据进行分析时，会非常耗时。而 Oracle R Enterprise 就能解决这个问题。它能使数据分析人员在 Oracle 数据库中使用 R，即原来在 PC 机或笔记本电脑上完成的任务现在可在数据库服务器上完成了。这是一个革命性的变化，这意味着可利用数据库服务器的硬件资源来进行大数据量的分析，从而大大提高了数据分析的速度。

本书作者 Brendan Tierney 是业内知名专家，不仅精通 Oracle 数据库和 R 语言，在相关理论和实践方面，也有高深的造诣和丰富的经验。本书就是其多年经验的结晶。

本书开篇介绍如何在 Oracle 数据库上安装 R 语言，然后沿着安装-使用-卸载路径展开，将全书分为 14 章。

第 1 章是对 Oracle R Enterprise 的综述。

第 2 章详细介绍 Oracle R Enterprise 的安装。包括安装的先决条件、在 Oracle R Enterprise 数据库的服务器端和客户端机器上的安装、验证安装是否成功。

第 3 章介绍 Oracle R Enterprise 的最基本用法。

第 4 章介绍 Oracle R Enterprise 的透明层，这是能在 Oracle R Enterprise 数据库中使用 R 的关键所在。

第 5 章介绍 Oracle R Enterprise 的各种软件包。

第 6 章~第 14 章介绍 Oracle R Enterprise 的各种用法。

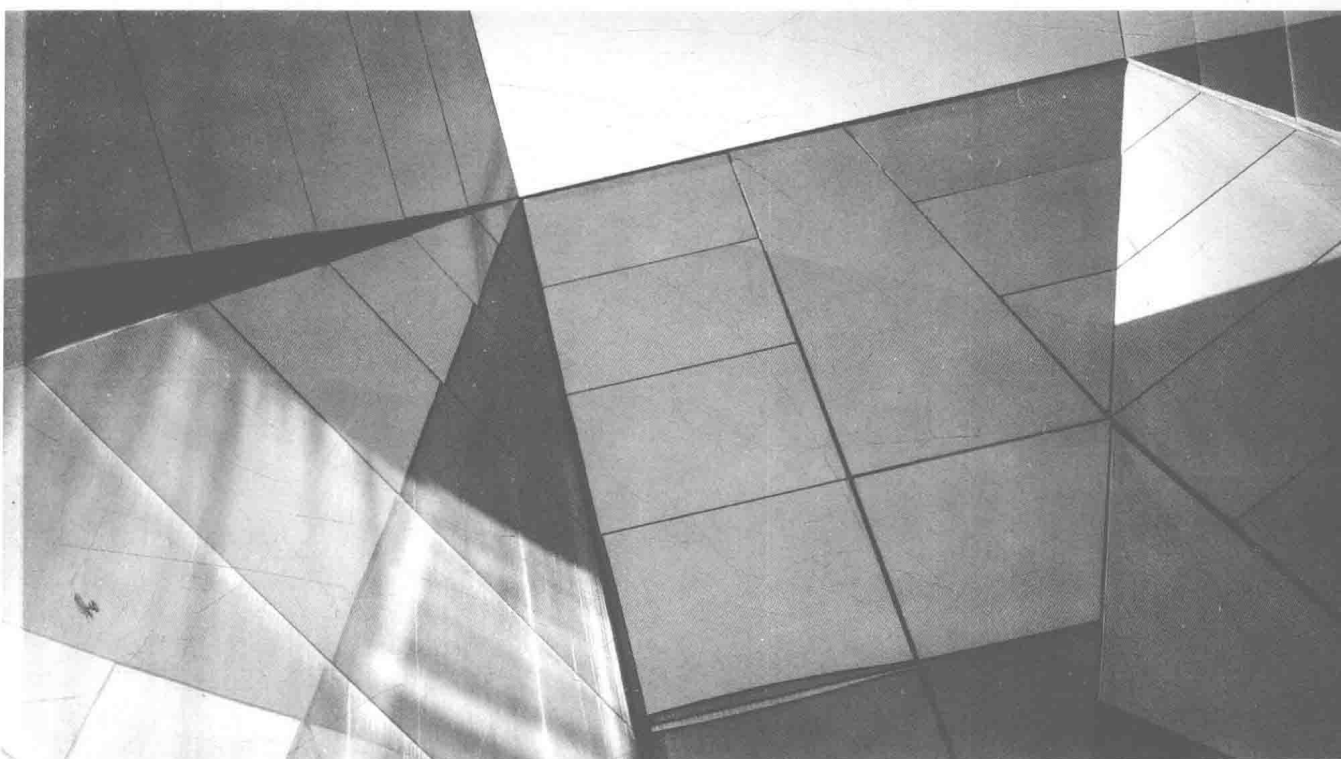
这是一本针对性很强的专业书籍，书中列举大量示例，实用性很强，对于在大数据环境工作的 DBA、应用软件开发人员及数据分析人员有较强的指导作用。但也要指出，由于作者长于实践，在叙述时难免有些口语化，对同一事物常用多个词汇来表述，如：书中常将 package、function 和 task 混用，还喜欢用一些比较怪的词汇，这样一来，对于不熟悉 Oracle 数据库的初学者，可能造成一些困惑，也给翻译工作带来一些困难。译者在翻译过程中，尽可能地将这些问题进行了更正。

不过，瑕不掩瑜，总的说来，这是一本很好、很贴近实用的书。很荣幸能将这本书翻译成中文版奉献给国内广大读者。这里要感谢清华大学出版社的编辑们，他们为本书的翻译投入了巨大热情并付出了很多心血，没有他们的帮助和鼓励，本书不可能顺利付梓。

还要感谢甲骨文(中国)公司的技术产品售前总监许向东和她的团队，在百忙之中审阅了译稿的初稿，并提出很多修改意见。

本书全部章节由张骏温翻译，参与本次翻译活动的还有赵爱华、薛群群、王欢、张彩霞、张越、杨萌、刘东、张若楠、江周娴、张艳，在此一并谢过！

由于水平有限，翻译工作中可能会有不准确的内容，如果读者在阅读过程中发现失误和遗漏之处，欢迎批评指正。

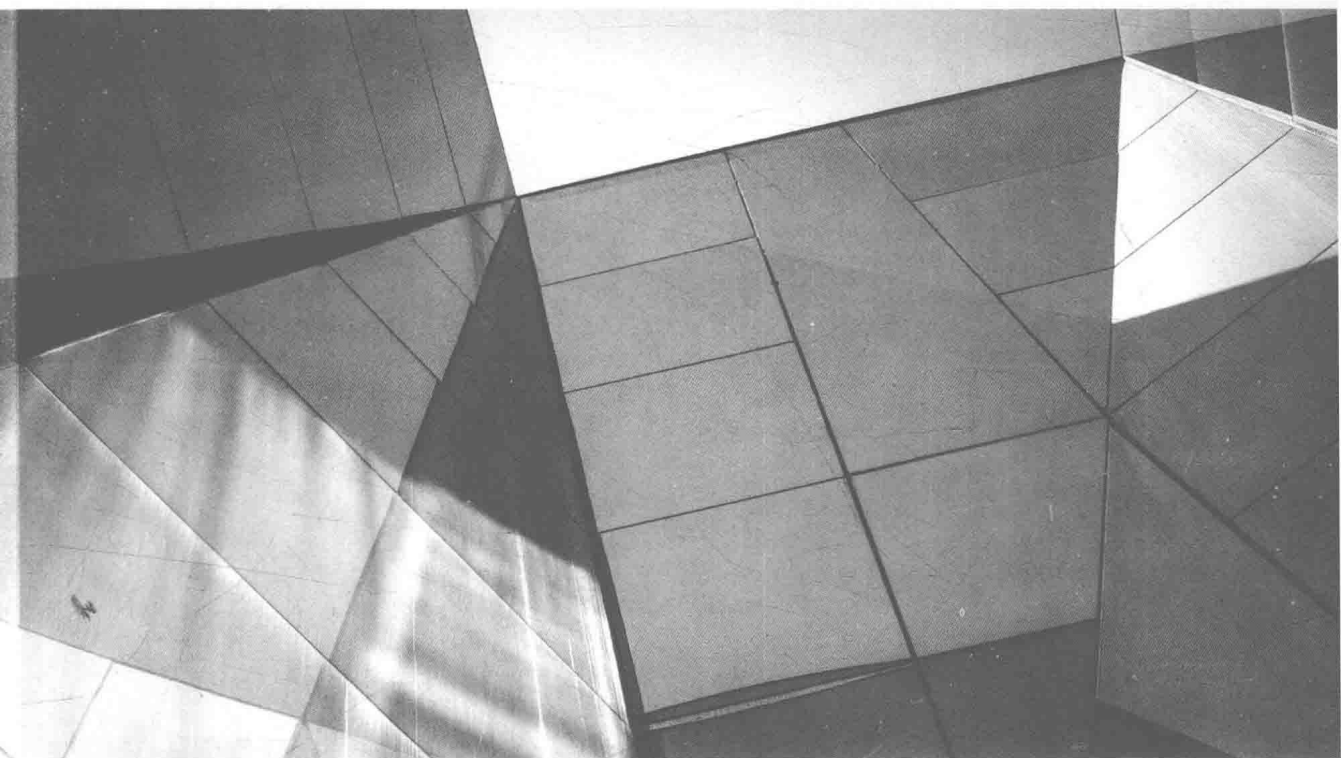


作者简介

Brendan Tierney 是 Oracle ACE 总监，是都柏林理工学院的数据科学、数据库和大数据讲师，也是一位独立咨询师(Oralytics)。Brendan 曾在爱尔兰、英国、比利时、荷兰、挪威、西班牙、加拿大和美国等多个国家的项目中工作，在数据挖掘、数据科学、大数据和数据仓库领域拥有逾 24 年的工作经验，是公认的数据科学和大数据专家。Brendan 是 Oracle User Group 社区的活跃分子，是 OUG 在爱尔兰的领导者之一。Brendan 是 *UKOUG Oracle Scene* 杂志的编辑，定期在全球技术会议上发表演讲；也是一位活跃的博客写手，曾为 OTN、Oracle Scene、IOUG SELECT Journal、ODTUG Technical Journal 和 ToadWorld 撰写文章。他还是位于爱尔兰的 DAMA 的董事会成员。Brendan 已撰写 *Predictive Analytics Using Oracle Data Miner* 和 *Real World SQL and PL/SQL: Advice from the Experts* 两本书籍。

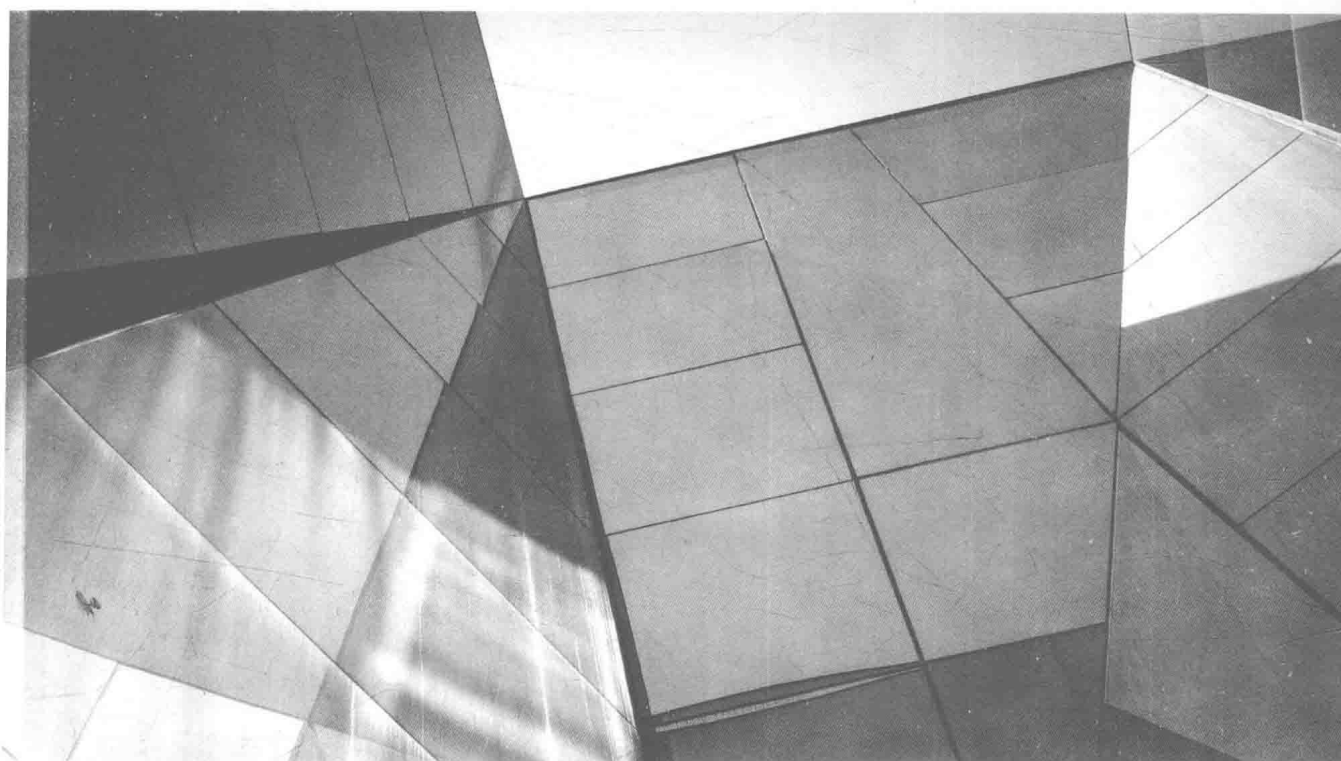
Web 和 blog: www.oralytics.com

Twitter: @brendantierney



技术编辑简介

Mark Hornick 是 Oracle Advanced Analytics Product Management 的主任，目前致力于 Oracle 的 R 技术。他和内部及外部的客户一道从事将 R 应用于 Oracle Database、Oracle Exadata 和 Oracle Big Data Appliance 中可变高级分析中的工作——无论是本地的，还是在云中的。Mark 是 *Using R to Unlock the Value of Big Data*、*Oracle Big Data Handbook* 和 *Java Data Mining: Strategy, Standard, and Practice* 三本书的合著者，并在 blogs.oracle.com/R 上撰写博客。Mark 于 1999 年随着 Thinking Machines Corp. 的并购而加入 Oracle 的 Data Mining Technologies 组，并于 2010 年转向 R 技术。在那之前，Mark 致力于如下领域的研究和开发工作：分布式对象管理、扩展事务模型、工作流管理系统和 GTE Laboratories(现在的 Verizon)的电信网络监控对象模型。Mark 是 IOUG 的 Business Intelligence Warehousing and Analytics (BIWA) SIG 的创始者和 Oracle Advisor，担任 BIWA Summit 活动的 Content Selection Committee Chair。他也是 R Consortium 的 Oracle 代表。Mark 从 Rutgers University 获得计算机专业学士学位，从 Brown University 获得计算机专业硕士学位。



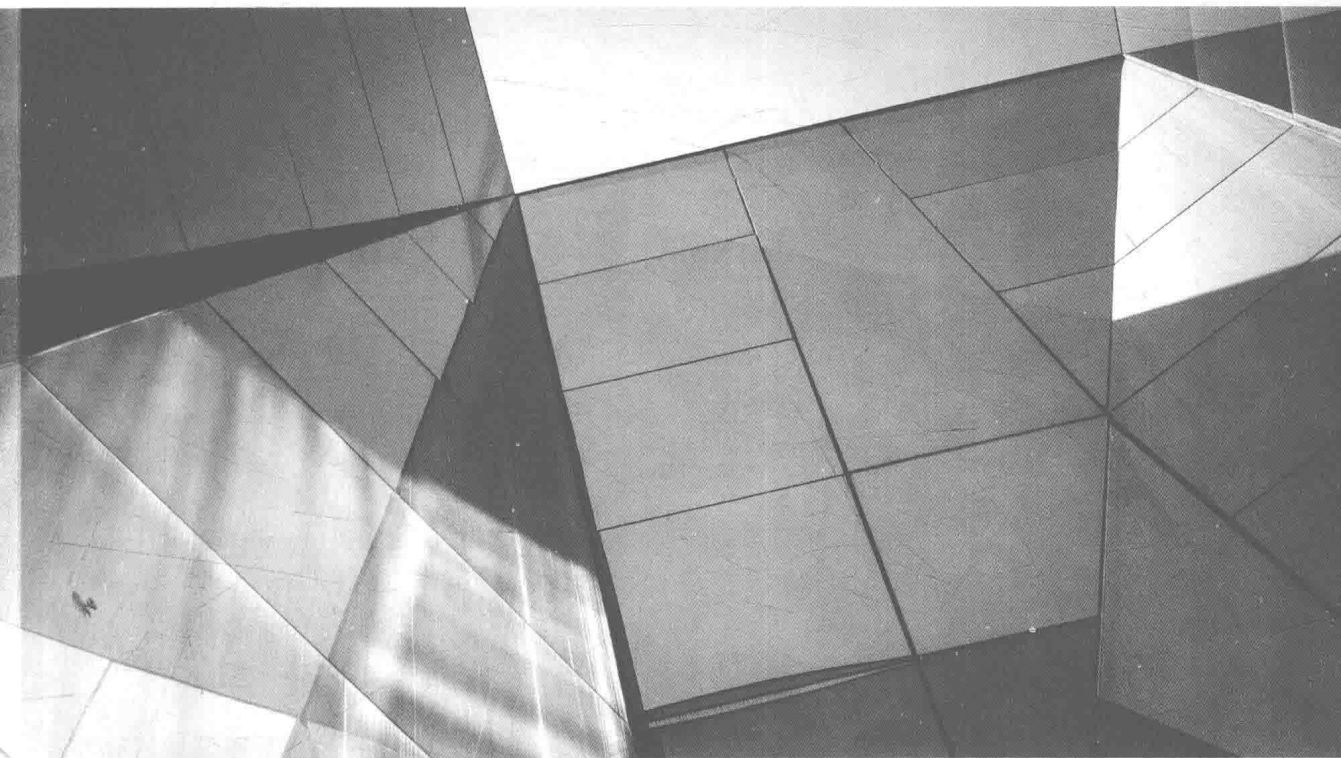
致 谢

感谢 Grace、Daniel 和 Eleanor，感谢你们在我同时撰写本书和另一本书时所给予的一贯支持和鼓励。的确，同时写两本书不是件轻松的事，没有你们的支持，这是不可能的。

非常感谢 Mark Hornick 担任本书的技术编辑。他几乎是对本书的每一行、每章中的例子都给出了详细的意见与建议。

还有几位也要在此提及，他们为本书的某些章节提供了支持与帮助。首先，我要感谢 Roel Hartman 在 APEX 方面所给予的帮助。另外，感谢 Oracle Data Mining 团队的 Charlie Berger、Mark Kelly 和 Pedro Rata 以及 Oracle R Enterprise 团队的 Sherry LaMonica。

更多的感谢给予 McGraw-Hill Professional 的 Wendy、Brandi、Claire 和 Rachel，他们从本书还处于创意阶段就和我一起工作。还要感谢 Bart Reed 对本书所做的文字编辑工作。



前 言

高级分析和 Big Data 的世界正处于不断发展中——不断发展出新产品来帮助我们管理各种形式和规模的数据。公司要面对的一个挑战是处理它们日益增长的数据量(volumes)并提取有意义和有用的信息。

R 语言是一种开源语言，已经诞生 20 多年了。我们已经看到，R 语言在全世界几乎每个行业中都得到广泛传播，与之相随的每年数以千计的毕业生进入市场。这使得组织可培养数据分析师并让他们从事分析数据的工作。尽管 R 语言有许多优点，但它仍有大量的局限。局限主要集中在这种语言随当今大部分组织所面对的典型数据量而调整的能力方面。

而在 Oracle R Enterprise 中，Oracle 已经克服了 R 语言的这些局限。Oracle 已经获得了 R 语言并将它集成到 Oracle Database 中。这样一来便解决了任何可扩展性问题和性能问题。Oracle R Enterprise 能使 R 脚本并行执行，并将数据库服务器用作计算引擎以便允许极大量的数据在给定时间内得到处理。通过将 R 语言集成到 Oracle Database 中，你便可以使用 SQL 和 PL/SQL(数据库的主要编程语言)来执行脚本并处理结果。这种集成，再结合 SQL 语言，使你可以很容易地把分析和图片包含到产品环境或前端应用中。任何能够调用和处理 SQL 的编程语言现在都可以运行脚本并处理结果。

本书的目标读者有三类。第一类是数据科学家，他们使用 R 语言进行分析和高级分

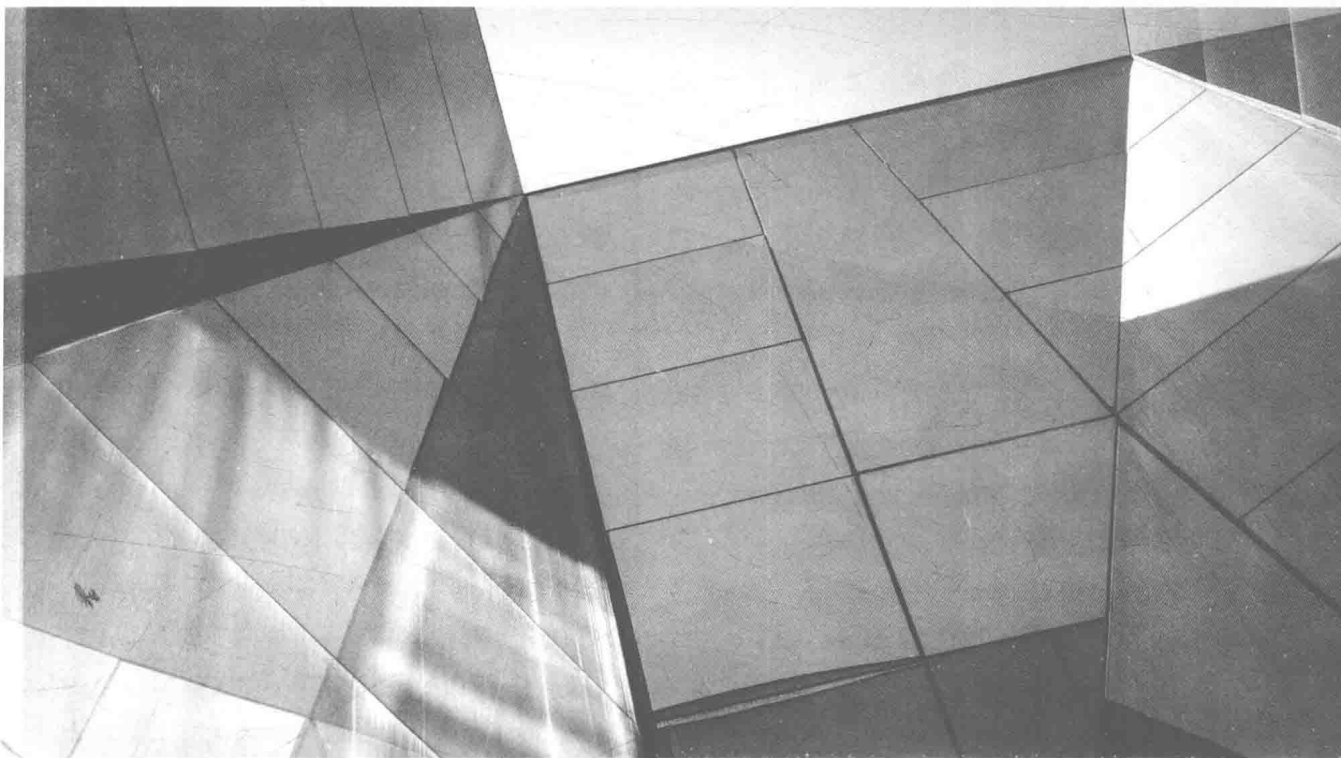
析工作。随着数据集的增长,他们将 Oracle R Enterprise 用作 Oracle Database 和数据库服务器的接口。本书将帮助他们理解如何使用 Oracle R Enterprise 及如何结合其他产品来使用 Oracle R Enterprise。第二类包括 SQL 和 Oracle Business Intelligence 的开发者。这类用户有一种不断增长的需求:将使用 R 得到的分析结果并集成到他们的应用和分析仪表盘(analytic dashboard)中。这些开发者可使用 Oracle R Enterprise 所携带的 SQL API 函数(function)来很容易地把用户定义的 R 脚本集成到工作流、应用程序和仪表盘中。第三类用户包括那些使用分析、Big Data 和高级分析等的多重角色人员。这些用户利用 Oracle Database、Hadoop、R 语言、SQL、PL/SQL、APEX 以及各种高级分析工具进行工作。本书包含 Oracle R Enterprise 如何被这三类用户中的每一类使用的细节。

通过本书,你可以学到并理解:

- 如何安装 Oracle R Enterprise。
- 如何设置和配置 Oracle Database schemas 以便使用 Oracle R Enterprise。
- 透明层是什么以及如何能利用它来无缝地使用数据库中的数据。
- 如何利用 Oracle Database 中的数据来访问、使用、转换和抽样数据。
- Oracle R Enterprise 的不同成分和这个产品所携带的附加的高级分析算法。
- 如何利用 R 来使用数据库中的 Oracle Data Mining 算法。
- 如何在 Hadoop 上使用 Oracle R Advanced Analytics 来进行分析以及如何使用 Spark-enabled 算法。
- 如何利用 Oracle Data Miner GUI 工具来使用 Oracle R Enterprise 和用户定义的 R 脚本。
- 如何把用户定义的 R 添加到诸如 Oracle APEX 和 OBIEE 之类的应用程序中。
- 为了支持 Oracle R Enterprise,数据库管理员需要执行什么任务。

本书大部分章节都列举了丰富的示例。为节省时间以及避免输入错误,本书提供了一组代码文件,每章的代码都放在一个单独文件中。

可从 McGraw-Hill Professional 网站 www.mhprofessional.com 下载压缩文件。只需要输入书名或 ISBN (1259585166),然后单击本书主页的 Downloads & Resources 选项卡即可。另外,扫描本书封底的二维码也可获得代码文件。



目 录

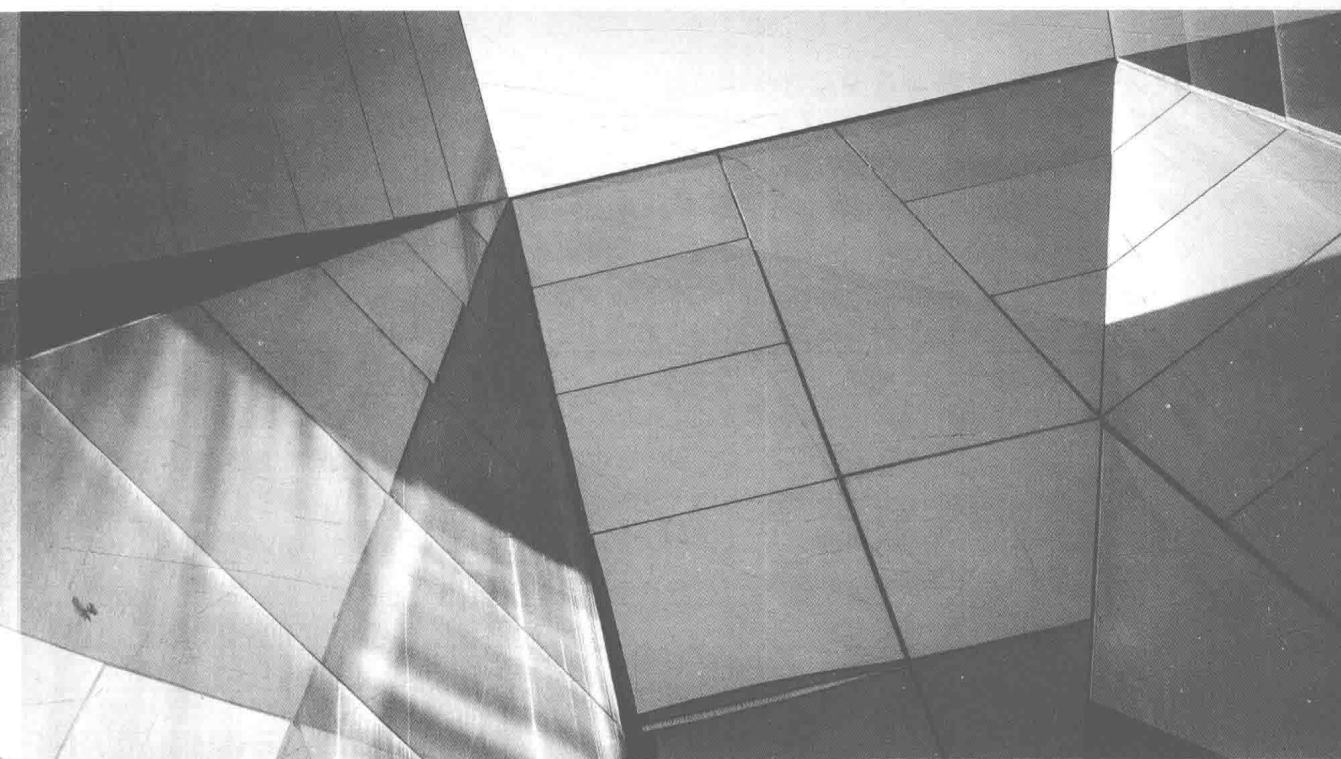
第 1 章 Oracle R Enterprise 简介..... 1	
1.1 本书的目标 3	
1.2 Oracle Advanced Analytics 选项 4	
1.3 Oracle R Enterprise(Oracle R 企业版)..... 5	
1.4 利用 Oracle R Enterprise 易于 部署 R 7	
1.5 Oracle 的 R 技术 9	
1.6 客户如何使用 Oracle R Enterprise 和 Oracle Advanced Analytics 10	
1.7 小结 11	
第 2 章 安装 Oracle R Enterprise 13	
2.1 安装的先决条件 13	
2.2 设置 Oracle Database 14	
2.3 安装 Oracle R Enterprise 15	
2.3.1 Oracle R Enterprise 的 软件包 15	
2.3.2 安装前的要求 16	
2.3.3 在 Oracle Database Server 上的安装 17	

2.3.4 安装客户端	22	4.2 探寻 ORE 透明层背后的真相	50
2.4 验证 ORE 的安装	25	4.3 小结	53
2.5 安装 RStudio	26	第 5 章 Oracle R Enterprise 程序包	55
2.6 使用 Oracle 的预制应用	27	5.1 Oracle R Enterprise 程序包	55
2.6.1 预制应用 Oracle Database Developer	27	5.2 探索 ORE 包函数和程序包版本	57
2.6.2 预制应用 Oracle OBIEE Sample App	28	5.3 ORE 设置和选项	59
2.6.3 预制应用 Oracle Big Data Lite	29	5.4 获得 ORE 的帮助	61
2.7 小结	29	5.5 ORE 演示脚本	62
第 3 章 Oracle R Enterprise 入门	31	5.6 小结	66
3.1 创建和管理到数据库的 ORE 连接	31	第 6 章 探索数据	67
3.1.1 使用 ore.connect 创建 ORE 连接	32	6.1 探索性数据分析(EDA)	68
3.1.2 使用 ore.sync 同步数据	33	6.2 数据抽样	73
3.1.3 使用 ore.attach 将对象加入搜索空间	35	6.3 数据聚合	78
3.2 执行 SQL 命令	36	6.4 数据转换	80
3.3 在 Oracle Database 中处理数据	36	6.4.1 派生属性	80
3.4 在数据库中存储数据	40	6.4.2 分块属性	81
3.4.1 使用 ore.creat 函数建表	40	6.4.3 组合数据	82
3.4.2 使用 ore.drop 函数删除表	40	6.5 数据排序	82
3.4.3 ore.create 函数和 ore.drop 函数组合使用的示例	41	6.6 小结	83
3.5 在数据库自带的 R Datastore 中存储 ORE 对象	41	第 7 章 使用 ODM 算法建立模型	85
3.6 断开与数据库的连接	46	7.1 Oracle Data Mining(Oracle 数据挖掘)	86
3.7 小结	46	7.1.1 ORE 中可用的 ODM 算法	87
第 4 章 透明层	47	7.1.2 利用 OREdm 软件包在 Oracle 中进行自动数据准备	87
4.1 透明层概述	47	7.2 使用 OREdm 软件包建立模型和对数据进行评分	88
		7.2.1 属性重要性	88
		7.2.2 关联规则分析	89
		7.2.3 决策树	92
		7.2.4 支持向量机	94

7.2.5	朴素贝叶斯	98
7.2.6	广义线性模型	99
7.2.7	聚类	100
7.3	保存数据挖掘模型	101
7.4	小结	103
第 8 章	利用 ORE 和其他算法建立模型	105
8.1	什么算法是可用的?	106
8.2	为建模而准备数据	106
8.3	使用 ORE 算法建立模型	107
8.3.1	广义线性模型	108
8.3.2	线性和逐步回归模型	109
8.3.3	神经网络	111
8.3.4	随机森林	112
8.4	使用 R 软件包和算法构建模型	113
8.5	使用 ore.predict 优化数据库自带的评分过程	115
8.6	小结	117
第 9 章	在用户定义的 R 脚本库中生成 R 脚本	119
9.1	使用 R 脚本库	120
9.2	创建和删除 R 脚本	120
9.3	授予和撤消用户定义的 R 脚本的特权	122
9.4	管理 R Script Repository	123
9.5	使用 SQL API 创建脚本	123
9.5.1	创建一个 R 脚本	124
9.5.2	删除一个脚本	126
9.5.3	授予和撤消访问权限	126
9.5.4	用户定义的 R 脚本的数据字典视图	127
9.6	小结	127

第 10 章	嵌入式 R 执行	129
10.1	通过 R 接口使用嵌入式 R 执行	129
10.1.1	如何使用 ore.doEval 函数	130
10.1.2	如何使用 ore.tableApply 函数	132
10.1.3	如何使用 ore.groupApply 函数	134
10.1.4	如何使用 ore.rowApply 函数	135
10.1.5	如何使用 ore.indexApply 函数	136
10.1.6	并行执行嵌入式 R 函数	137
10.2	通过 SQL 接口使用嵌入式 R 执行	138
10.2.1	如何使用 rqEval SQL 函数	138
10.2.2	如何使用 SQL 函数 rqTableEval	140
10.2.3	如何使用 SQL 函数 “rqGroupEval”	144
10.2.4	如何使用 SQL 函数 rqRowEval	146
10.3	小结	147
第 11 章	针对 Hadoop 的 Oracle R Advanced Analytics	149
11.1	连接到 Apache Hive 上并处理数据	152
11.2	使用 ORCH 管理 Map-Reduce 作业	155
11.3	通过 ORAAH 使用 Spark	157

11.4 小结	158	13.3 小结	182
第 12 章 通过 Oracle Data Mining		第 14 章 针对 Oracle DBA 的	
使用 ORE	159	ORE	183
12.1 Oracle Data Mining	160	14.1 在数据库中创建一个新的	
12.2 Oracle Data Miner	164	ORE 模式	183
12.3 通过 SQL Node 来包含 R		14.2 在 Oracle R Enterprise 中	
脚本	165	安装新的 R 软件包	184
12.4 使用 R 节点	167	14.2.1 在数据库服务器上安装新	
12.5 小结	168	的 R 软件包	185
第 13 章 在 APEX 和 OBIEE 中使用		14.2.2 在客户机上安装新	
ORE	171	软件包	186
13.1 Oracle APEX	171	14.3 ORE 的全局变量和选项	187
13.1.1 在 APEX 应用程序中		14.4 使用 ore.parallel 特性	189
包含 ORE 脚本	173	14.5 卸载 Oracle R Enterprise	189
13.1.2 向 APEX 应用程序中		14.5.1 从 Oracle Database 服务	
添加 R 图形	175	器中卸载 Oracle R	
13.2 Oracle Business Intelligence	177	Enterprise	190
13.2.1 设置 OBIEE 以便能		14.5.2 从客户端卸载 Oracle	
使用 ORE	178	R Enterprise	191
13.2.2 在 OBIEE RPD 中使用		14.6 小结	192
R 脚本	180		
13.2.3 在仪表板上呈现由 R			
脚本产生的结果	180		



第 1 章

Oracle R Enterprise 简介

高级分析和 **Big Data** 领域正处于不断扩展之中。**Big Data** 领域的扩展带来新的产品，帮助我们管理所有的各种不同形式和数量的数据。众多公司一直要面对的一个永恒挑战是处理数量不断增长的数据并提取有意义的和有用的信息。市场上有许多工具和供应商，可提供便于提取有意义信息的解决方案。我们所能得到的主要数据管理工具之一就是数据库，包括关系型数据库、NoSQL 数据库、Hadoop 上的数据库，以及更多。传统上，数据库用于抓取、存储和分析一个组织中的各种形式的数据。但近来我们看到了大量不同的帮助我们管理特定场景中数据的数据管理解决方案。在 **Big Data** 领域中，这一现象尤为突出，在 Hadoop 和 NoSQL 数据库的使用上，我们已经看到了一个重大的提升。类似地，在分析工具环境中，我们已经看到了能使处理数据和从数据中提取有意义信息变得容易的语言和工具的发展。一种最普遍适用的数据分析语言就是 **R**。

图 1-1 显示了分析的不同分类。通常，大部分组织都会关注定性分析，因为定性分析关注组织此地此刻的运行状况、当前正在发生什么和为什么某些事会发生。当一个组

织成熟起来时(有时是因为竞争性的原因)，它会转入预测分析领域。这一领域关注于理解可能存在于数据中的模式以及组织如何利用这些模式来预测可能发生的事情。有一些示例，包括预测客户的税收、预测传感器或设备何时失效、管理员工并预计谁可能会离职等。近来，正是这类预测分析吸引了大部分的注意力。这项技术和用于预测分析的机器学习技术已经出现了几十年了，但我们现在看到的是，这些技术正被用在比以往更广范围的工业和问题中。

规范性分析	物联网	物联网自动化
	不确定型决策	可变条件下决定最佳结果
	协同决策	决策中涉及的多个预测模型和其他技术
	自我优化	不断改进；寻找最佳结果
自动分析	后端应用	批处理、决策、预警等
	前端应用	决策、打分、告警等
	实时	在数据处理过程中实时预测
	自动更新	自动重建模型
	嵌入式分析	高级分析是……的一部分
	What-if 分析	如果数据变了，会发生什么？
预测分析	预测建模	接下来会发生什么？
	预报	如果这些趋势继续下去，会发生什么？
	模拟	如果……，有可能会发生什么？
	告警	需要行动
描述性分析	统计分析	高级统计技术、相关性等
	查询/深入探讨(drill down)	问题的实质是什么？
	特别报告	多少、多频繁、在哪里？
	标准报告	发生了什么？

图 1-1 分析的不同分类

R 语言是一种已经出现了二十多年的开源语言。我们已经见到了 R 语言在全世界以及几乎每个行业中被广泛采用，随之而来的是每年有数以千计的有 R 使用经验的毕业生进入市场。这就允许组织创建数据科学团队来分析其数据。R 语言可使你使用由世界各地专家们开发的各种 R 软件包。这些软件包建立在基本 R 语言的功能上，提供了大量的统计和机器学习技术。一些 R 软件包是专为某些类型的问题或专为某些行业所面对的特殊问题而开发的。

尽管 R 语言有许多优点，但它也有许多局限。这些局限主要集中在扩展自身以便能处理当今大部分组织所使用的典型数据量的能力上。Oracle 对 R 语言所做的正是使其能以一种结合方式和 Oracle Database 及数据库服务器一起工作。这样做的目的是要克服 R 语言的三个主要局限：

- **可扩展性(Scalability)** R 语言最初只能使用本地计算机的 CPU。无论你的计算机有多少个 CPU，数据处理都会受到限制。随着数据量和分析复杂性的增加，

很快就会达到机器的极限。有大量可用的 R 软件包能克服这个问题，但它们涉及多线程编程，这就超过了大多数编程者的编程能力，更不要说数据分析师了。

- **内存管理** R 语言受限于机器中可用的 RAM 的量。因为大多数分析都是在分析员的台式机或笔记本电脑上进行的，当处理组织的数据时，可用的 RAM 很快就会耗光。有关这种局限的一个常见例子是当试图打开一个 Excel 电子表格时，却打不开，并出现了出错信息。并行化处理数据可能是一个挑战。相同的情况再次出现，确实有可用的 R 软件包帮助你克服这个挑战，但这些软件包是有限的并且需要丰富的编程知识。
- **产品部署** 数据分析师和数据科学家的利用 R 语言所做的大部分开发工作都是在他们的本地电脑或小型服务器上进行的。这里的挑战是，如何把所有这些分析成果和预测模型带出分析实验室环境并把它们添加到进行批处理的后端上的生产架构中或一个前端应用程序中。把 R 语言与其他语言集成起来的能力是极为有限的，在很多情况下，需要利用其他语言，诸如 C、Java、SQL 等，重新开发分析和预测模型。

当想要把自己的高级分析工作带出实验室环境并植入一个企业架构及各种运营和商务解决方案中时，大部分组织都面临着必须处理这些局限的问题。

因为有了 Oracle R Enterprise，Oracle 已经克服了 R 语言的这些局限。Oracle 已经得到了 R 语言并把它集成到 Oracle Database 中。通过这么做，Oracle 已经克服了前面描述的可扩展性和性能问题。Oracle R Enterprise 使得 R 脚本能够并行执行并把数据库服务器用作计算引擎以便在给定时间内使大量数据得到处理。因为 R 语言与 Oracle Database 集成在一起，现在就可以使用 SQL 和 PL/SQL 等数据库方面的主要编程语言来执行 R 脚本并处理所得的结果。因为有了与 SQL 语言的这种集成，现在就可以很容易地把任何分析或 R 图形包含到产品环境或前端应用中。任何可以调用和处理 SQL 的编程语言现在都可以运行 R 脚本并处理所得的结果。

1.1 本书的目标

在此处就确立本书的目标是非常重要的，这样就可使你理解本书通篇涉及的材料类型以及读完本书的所有章节和例子后你能够做什么。本书并不讲述如何编写 R 代码，所以不是面对初学者的入门读物。本书面向那些具有利用 R 语言工作和使用 R 语言经验的人，以及具有在 Oracle Database 中利用数据工作的经验的人。本书的每一章都贯穿了 R 开发者在利用 Oracle Database 工作时所需要知道和用到的主题和领域。一些章节 SQL 开发者，这些人员与 R 语言有某种接触，或接受了把存储于 Oracle Database 中的 R 脚本添加到各种仪表盘、报告环境、商务应用程序、后端处理等的任务。每一章都提供了说明如何使用 Oracle R Enterprise 的各种特性，包括 R API 函数套件和 ORE SQL API 函数的例子。这些例子都很简单，以便你能够很容易地跟着它们学习并在自己的测试环境中重现它们。随着使用这些函数的经验的增长，你会发现自己会创建更复杂的例子了。