

让一切变得更简单!

Big Data

FOR

达人迷<sup>®</sup> DUMMIES<sup>®</sup>

# 写给大家看的 大数据

## 学习如何：

1 了解大数据工具及基础架构

2 探索大数据如何改变你的业务

3 将结构化数据和非结构化数据整合到大数据环境中

4 用预测分析进行更优化的决策

◎【美】Judith Hurwitz Alan Nugent  
Fern Halper Marcia Kaufman 著  
◎ 麦秆创智 译

 人民邮电出版社  
POSTS & TELECOM PRESS



***Big Data***

FOR  
**DUMMIES**

达人迷

**写给大家看的  
大数据**

◎【美】Judith Hurwitz Alan Nugent  
Fern Halper Marcia Kaufman 著  
◎ 麦秆创智 译

人民邮电出版社  
北京

## 图书在版编目 (C I P) 数据

写给大家看的大数据 / (美) 赫尔维茨  
(Hurwitz, J.) 等著; 麦秆创智译. — 北京: 人民邮电  
出版社, 2014. 10  
(达人迷)  
ISBN 978-7-115-35613-0

I. ①写… II. ①赫… ②麦… III. ①数据处理  
IV. ①TP274

中国版本图书馆CIP数据核字(2014)第097624号

## 版 权 声 明

Original English language edition Copyright © 2013 by Wiley Publishing, Inc.. All rights reserved including the right of reproduction in whole or in part in any form. This translation published by arrangement with Wiley Publishing, Inc.  
本书原英文版本版权© 2013 归 Wiley Publishing, Inc.所有。未经许可不得以任何形式全部或部分复制作品。本书中文简体字版是经过与 Wiley Publishing, Inc.协商出版的。

## 商 标 声 明

Wiley, the Wiley Publishing Logo, For Dummies, the Dummies Man and related trade dress are trademarks or registered trademarks of John Wiley and Sons, Inc. and/or its affiliates in the United States and/or other countries. Used under license.  
Wiley、Wiley Publishing 徽标、For Dummies、the Dummies Man 以及相关的商业特殊标志均为 John Wiley and Sons, Inc.及/或其子公司在美国和/或其他国家的商标或注册商标。未经许可不得使用。



- 
- ◆ 著 [美] Judith Hurwitz Alan Nugent  
Fern Halper Marcia Kaufman
  - 译 麦秆创智
  - 责任编辑 刘 洋
  - 责任印制 杨林杰
  - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号  
邮编 100164 电子邮件 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: 800×1000 1/16  
印张: 17.75  
字数: 314 千字 2014 年 10 月第 1 版  
印数: 1-3 000 册 2014 年 10 月北京第 1 次印刷  
著作权合同登记号 图字: 01-2013-6937 号

定价: 59.00 元

读者服务热线: (010)81055488 印装质量热线: (010)81055316  
反盗版热线: (010)81055315

## 内容提要

大数据是当前信息科技领域最为炙手可热的话题之一。本书简单而系统地介绍了大数据体系涉及的各项知识，涵盖大数据的基本概念、大数据的技术基础、大数据管理、大数据分析、大数据在现实工作中如何实现和实施等关键内容，涉及大数据基础架构、大数据使用的数据库和分布式技术、对大数据进行基础分析和高级分析的特点及异同，以及企业如何应用大数据转变其商业运作模式等内容，能够对想要了解大数据全貌，或是想要使用大数据的企业和个人提供全面的知识内容和学习借鉴。

本书语言生动，内容覆盖面广，理论结合实例，非常适合对大数据感兴趣的广大读者。对于从事与大数据相关工作的人员，本书也有很高的参考价值。

## 献词

朱迪丝将此书献给她的丈夫沃伦，她的孩子萨拉和大卫，以及她的母亲伊莱恩。她还想以此书纪念她的父亲大卫。

艾伦将此书献给他的妻子简，感谢她的爱与支持；以及他的3个孩子，克里斯、杰夫和格雷格；并纪念他的父母，感谢他们帮助他在这条道路上启航。

费恩将此书献给她的丈夫克雷，女儿凯蒂和琳赛，以及她的妹妹阿德里安娜。

玛西亚将此书献给她的丈夫马修，她的孩子萨拉和艾米丽，以及她的父母格洛丽亚和拉里。

## 作者致谢

我们衷心地感谢Wiley的朋友们，尤其是我们的编辑Nicole Sholly。此外，我们还感谢我们的技术编辑Brenda Michelson，感谢她细致入微的贡献。

几位作者想对以下几位技术领域的思想领袖为本书所做的贡献表示感谢，谢谢他们慷慨地花时间将这些关于混合云的技术和其他一些广泛的知识与大家分享。他们提供的帮助体现在几个方面，包括技术简介、分享研究成果、案例学习，以及内容审阅。我们感谢下面的个人和机构提供的宝贵帮助：

Context Relevant: Forrest Carman

Dell: Matt Walken

Epsilon: Bob Zurek

IBM: Rick Clements, David Corrigan, Phil Francisco, Stephen Gold, Glen Hintze, Jeff Jones, Nancy Kop, Dave Lindquist, Angel Luis Diaz, Bill Mathews, Kim Minor, Tracey Mustacchio, Bob Palmer, Craig Rhinehart, Jan Shauer, Brian Vile, Glen Zimmerman

Kognitio: Michael Hiskey, Steve Millard

Opera Solutions: Jacob Spoelstra

RainStor: Ramon Chen, Deidre Mahon

SAS Institute: Malcom Alexander, Michael Ames

VMware: Chris Keene

Xtremedata: Michael Lamble



## 作者介绍

朱迪丝·S.赫尔维茨 (Judith S.Hurwitz) 是Hurwitz & Associates的主席兼CEO, Hurwitz & Associates是一家专注于新兴科技的调研和咨询公司, 涵盖领域包括云计算、大数据、数据分析、软件开发、服务和管理。朱迪丝是一名技术策略师、资深管理者以及优秀的作者。作为参与技术创新和应用的先驱, 朱迪丝多年来以受信顾问的身份服务了许多行业中的领军企业。她帮助这些企业针对平台融合的商业价值完成了新型商业模式的转型。她是Huiwitz集团的创始人。她还曾在多家大型公司任职, 包括阿波罗计算机公司和约翰·汉考克。她撰写过大量涵盖分布式软件各方面知识的文章。2011年她主笔编写了《Smart or Lucky? How Technology Leaders Turn Chance into Success》一书(2011年由巴斯出版社出版)。朱迪丝是For Dummies系列中多本书的联合作者, 包括《Hybrid Cloud For Dummies》(John Wiley & Sons出版社于2012年出版), 《Cloud Computing For Dummies》(John Wiley & Sons出版社于2010年出版), 《Service Management For Dummies》以及《Service Oriented Architecture For Dummies》第二版(均由John Wiley & Sons出版社于2009年出版)。同时, 她还是多本For Dummies系列丛书定制版本的联合作者, 包括《Platform as a Service For Dummies——CloudBees特别版》(John Wiley & Sons出版社于2012年出版), 《Cloud For Dummies——IBM中型企业限定版》(John Wiley & Sons出版社于2011年出版), 《Private Cloud For Dummies——IBM限定版》(2011年)以及《Information on Demand For Dummies——IBM限定版》(2008年)(均由John Wiley & Sons出版社出版)。

朱迪丝拥有波士顿大学的学士和硕士学位, 曾在若干家新兴公司的咨询部门工作, 并被波士顿大学艺术与科学学院于2005年评为优秀校友。她曾在波士顿大学校友理事会任职。除此之外, 她还获得过2005年的马萨诸塞州技术领导委员会奖。

艾伦·F.纽金特 (Alan F.Nugent) 是Hurwitz & Associates的首席顾问。艾伦是一名极富经验的技术指导和有着30多年行业经验的资深老手。最近, 他在Mzinga公司担任首席执行官和首席技术官, Mzinga公司关注基于云的分布式大数据、智能社交和社区管理解决方案开发及部署。在Mzinga之前, 艾伦曾任CA公司的执行副总裁和首席技术官, 负责制订公司的战略技术方向。他是作

为CA企业系统管理（ESM）部门的高级副总裁和总经理加入CA公司的，负责管理基础架构产品名录和数据。在2005年4月加入CA之前，艾伦是Novell的高级副总裁和CTO，在此期间，他推动了整个公司向开源产品和身份驱动解决方案上转移。在此之前，作为BellSouth的顾问级CTO，艾伦领导公司将BellSouth许多不相关的客户和工作数据整合和转变成为单一的数据实例。

艾伦是Provo，UT自适应计算董事会的独立成员，是位于西雅图，WA的SpaceCurve的主席和董事顾问，是位于Waltham，MA的N-of-one的成员和顾问。同时他还是一名高产的作者，主要关注领域包括商业和技术，许多年来，他一直在持续地与人们分享他在各个领域的想法和经验。

费恩·哈博博士（Fern Halper, PhD）是Hurwitz & Associates的成员，同时也是TDWI高级研究分析部总监。她在数据分析、商业分析和战略规划方面有超过20年的经验，曾经发表过大量关于数据分析和高级分析的文章。她在预测分析和文本分析方面进行过广泛而深入的研究。费恩维护着一个技术博客。她还在AT&T贝尔实验室和Lucent科技担任要职，负责研发新型数据分析系统，同时为互联网业务设计开发战略和进行产品线规划。费恩在若干个大学中任教信息技术相关的课程。她从Colgate University获得了学士学位之后在Texas A&M大学攻读博士学位。

费恩是4本For Dummies系列书籍（公共版）的联合作者，包括《Hybrid Cloud For Dummies》（John Wiley & Sons出版社出版，2012年），《Cloud Computing For Dummies》（John Wiley & Sons出版社出版，2010年），《Service Oriented Architecture For Dummies》第二版和《Service Management For Dummies》（均为John Wiley & Sons出版社出版，2009年）。她还是若干定制版For Dummies系列书籍的联合作者，包括《Cloud For Dummies——IBM中型企业限定版》（John Wiley & Sons出版社出版，2011年），《Platform as a Service For Dummies——CloudBees特别版》（John Wiley & Sons出版社出版，2012年），以及《Information on Demand For Dummies——IBM限定版》（John Wiley & Sons出版社出版，2008年）。

玛西亚·A.考夫曼（Marcia A. Kaufman）是Hurwitz & Associates的联合创始人和COO，Hurwitz & Associates是一家专注于新兴科技的调研和咨询公司，涵盖领域包括云计算、大数据、数据分析、软件开发、服务和管理。她在虚拟化和云计算的商业价值方面发表过大量作品，强调云基础模型、商业模式、数据



加密和终端安全，以及云环境中的在线交易处理。玛西亚在商业策略、行业研究、分布式软件、软件品质、信息管理和分析方面有超过20年的研究经验，涉及行业包括金融服务业、制造业和服务业。在任职Data Resources.Inc. (DRI) 期间，她构建了复杂的行业模型和预测机制。她获得了Connecticut大学数学与经济学院的文学学士学位和波士顿大学的MBA学位。

玛西亚是4本For Dummies系列书籍（公共版）的联合作者，包括《Hybrid Cloud For Dummies》（John Wiley & Sons出版社出版，2012年），《Cloud Computing For Dummies》（John Wiley & Sons出版社出版，2010年），《Service Oriented Architecture For Dummies》第二版以及《Service Management For Dummies》（均为 John Wiley & Sons出版社出版，2009年）。她还是若干定制版For Dummies系列书籍的联合作者，包括《Cloud For Dummies——IBM 中型企业限定版》（John Wiley & Sons出版社出版，2011年），《Platform as a Service For Dummies——CloudBees特别版》（John Wiley & Sons出版社出版，2012年），《Private Cloud For Dummies——IBM限定版》（2011年）以及《Information on Demand For Dummies》（2008年）（均为John Wiley & Sons出版社出版）。

## 译者序

如果有人问我现在科技领域最热门的话题是什么，我一定会毫不犹豫地回答大数据；如果有人问我下一个十年能够对科技领域产生最深远影响的话题是什么，我一定会毫不犹豫地回答大数据；如果有人问我一个企业想要从已有的业务中深入挖掘更多价值，洞悉市场内涵，需要什么技术，我一定会毫不犹豫地回答大数据。所以我想，你一定也对大数据这个话题不再陌生了，而是和我一样，颇有摩拳擦掌，跃跃欲试之感。

这样的话，我当一次扫兴的人，提前为你剧透一下本书译毕之后的感想吧：大数据，果然是个博大精深的话题！跟随本书，我仿佛把近十年来的技术和理念发展重新经历和梳理了一遍，而后又把未来十年可能发生的变化和进步提前预览了一遍，合上书才发现，要想用好大数据，我要恶补的东西真是太多了。

好吧，你也许和读这本书之前的我一样，对技术和商业都有些粗浅的了解，也许和我一样正运营着一家公司，每天满脑子想的都是怎么把公司做成百年老店，看到亚马逊、阿里巴巴等大公司的成功，就像看到别人家考上哈佛的孩子一样，回来看看自己家的，非常希望能给他多补充点营养，打好底子。抑或是，你有一个和我一样想法的上级，他把自己的心肝宝贝拜托给你，让你出山打探地形寻点儿良药，回来告诉他如果能把大数据和现有业务结合起来，为公司的未来注入无限活力和无限可能，需要怎么做。恭喜你，你遇到了一本好书。

这本书定不能解决你所有的疑惑，毕竟大数据是太错综复杂的话题，它结合了从业务到技术、从底层架构到实际应用、从原始数据采集到复杂数据分析的方方面面。但通过这本书，你能够很好地一探大数据世界的全景，对你需要了解各个方面有基本的认知，下一步才好有的放矢地深入研究。对于架构师而言，也许是虚拟化和分布式；对研发人员而言，也许是Hadoop和MapReduce；对于数据库工程师而言，也许是NoSQL和数据集市；对于产品经理而言，也许是融合实时数据采集与现有业务流程……对于CEO而言，也许是如何配置一个能够支持大数据的团队，早早地搭上大数据这班快车。

大数据一定还做些什么需要你的智慧，欢迎进入大数据的世界！

由于译者水平有限，匆忙之中难免有疏漏，欢迎各位读者批评斧正。反馈意见和建议可发送至本书编辑的电子邮箱：[liuyang@ptpress.com.cn](mailto:liuyang@ptpress.com.cn)。

# 前 言

---

欢迎来到《写给大家看的大数据》的世界！大数据已经成为非常重要的技术趋势，它最有潜力能够颠覆式地改变世界上任何组织使用信息的方式以增强体验以及改变组织的商业模式。一家企业应该如何正确地将大数据运用到工作中去？将海量数据转化为知识是什么意思？在本书中，我们将向你展开一幅新画卷，看看软件、硬件和传输方式上的技术变革是如何改变数据的使用方式的。

大数据不是一个单一的市场。事实上，这个概念融合了已经有一段时间历史的数据管理技术。大数据技术能够帮助组织在正确的时间、以正确的方式储存、管理和操作海量的数据，以得到这些海量数据最有意义的内涵。理解大数据这一概念的关键在于，数据必须要经过管理，才能给出为满足业务需求而设计的解决方案。大部分公司在大数据之路上还只是迈出了第一步。大部分公司只不过在尝试着通过一些技术手段将大量数据收集起来，在其中寻找可能存在的隐藏着模式，而这些看似杂乱无章的数据，其背后有可能隐藏着一次重大变革的指引。有些数据可能指示了顾客在购买模式正在发生改变，或是指示着业务中的一些新元素，需要在它产生不可逆转的影响前引起人们的注意。

随着企业逐渐开始重视新的大数据解决方案，大量新的机会也随之开始涌现。例如，制造商可以检测到设备传感器采集的数据，以此决定在某种灾难性事件发生前处理器应该如何应对。零售商也可以实时地监控销售数据，并在执行某项交易时及时根据客户情况调整产品数量。大数据解决方案还可用于医疗行业，帮助医生确定某种病情的病因，并对治疗方案提供指导信息。

与此同时，大数据并非一种孤立的解决方案。实现大数据解决方案需要必要的基础设施提供灵活性以及对数据的分发和管理。所以，要想发挥出这项技术趋势的最大威力，将业务和技术策略提升至同等地位是非常有必要的。

出于以上种种原因，我们觉得帮助读者理解大数据背后的技术以及理解企业使

用新兴技术的方式非常重要，例如如何通过Hadoop、MapReduce和新型数据库引擎来将企业的数据转化为价值。本书为审视大数据提供了新的视角，什么是大数据？它如何能够让企业比以往更有效地利用数据？看完本书之后，相信你能够做出更为睿智的决策。

## 关于本书

大数据对于很多人来说是一个全新的概念，它要求人们对技术和业务需求有更为深入的调查和理解。不同的人需要了解大数据的不同方面，有些人想深入钻研技术细节，有些人想要了解应用大数据所带来的经济影响。打造大数据的生态圈需要整体架构，需要商业技巧，同时还需要大量的规划。

不管你阅读本书的目的是什么，在理解了以下问题之后，你会更了解大数据以及大数据可能会对你的业务产生的影响。

- ✔ 大数据的结构是怎样的？如何在不打断数据中心正常运行的前提下管理海量数据？
- ✔ 何时应当将大数据的分析结果与已有的数据仓库整合到一起？
- ✔ 在使用大数据时如何进行安全保障？怎么保证公司的信息安全？
- ✔ 不同的数据技术有些什么不同的价值？何时应当将它们纳入大数据策略中的一部分？
- ✔ 在大数据分析的过程中，哪些类型的数据源能产生帮助？如何将不同类型的分析应用到不同的业务问题中去？

## 面向新手

虽然我们试图为所有人提供帮助，但落实到编写这本书的过程中，我们还是挑选了一些我们认为会对《写给大家看的大数据》最感兴趣的人的特质。我们觉得你应该是这样的：

- ✔ 你很聪明。你不是傻瓜（译者注：本书原名《Big Data For Dummies》），但是可能本书的书名让你产生了一种不舒服的感觉，所以你都不敢在这本书旁边转悠，如果你可以做决定，你大概会把封面的标题换掉。
- ✔ 你是商界人士，很少甚至不想了解技术方面的知识。但是作为一个活在21世纪的人，技术是大家都没法回避的一道坎儿。人们都在说：“这一切都是大数据！”所以你想找个好办法，快速让自己了解其他人都在说什么。

✔ 你是一个IT精英，对很多技术有深入了解。但事实是，你在大数据领域是个新手。所有人都说大数据和以前的任何一种技术不一样。这个时候，你想要的是一张广角蓝图。

不管你是哪一种人，欢迎选购本书。我们就在这里，随时准备好为你提供帮助。

## 本书如何组织

为了便于阅读，我们将本书分成了七个部分。你可以直接阅读自己感兴趣的部分。

### 第一部分：大数据入门

在这一部分中，我们从技术和商业两个角度解释了所有你了解大数据所需要的基本概念；同时还分析了其中一些重要的概念和元素，这样你就可以在任何一场关于大数据的讨论中站稳脚跟了。

### 第二部分：大数据的技术基础

第二部分主要提供给技术或商业领域的专业人士，我们分析了几种不同类型的大数据元素，同时强调了几个重要的支持大数据的技术概念。在这一部分中，你将会了解到让大数据能够实际可行的基础架构类型。

### 第三部分：大数据管理

第三部分针对的同样是技术或商业领域的专业人士，但话题更为深入，我们将会深入分析不同的数据库选择，以及几种新型技术，例如MapReduce和Hadoop。了解这些关键技术能够帮助你掌握在大数据这一趋势背后更重要的核心是什么。

### 第四部分：数据分析与大数据

你会如何去分析和处理你手上大量的非结构化数据？在本书的第四部分中，我们更深入地讲解了不同类型的分析手段，它们能够很好地帮助你从你的数据中获取真实意义。这个部分将会引导你思考如何将大数据应用到你的实际业务中。

### 第五部分：大数据实现

这一部分深入到实际管理数据到底意味着什么，包括如何开始操作数据以及如

何保护这些数据的安全性和私密性。这一部分会让你在这个关键领域产生许多思考。

## 第六部分：现实中的大数据解决方案

在这一部分中，你将会学到企业是如何应用大数据转变其商业运转模式的。如果想要一窥未来你在大数据方面能做的工作，这一部分非常适合你。

## 第七部分：十项注意

如果你是第一次接触For Dummies系列丛书，你可能不太熟悉这一标志性的“十项注意”部分。在这一部分中，Wiley出版社的编辑要求For Dummies的作者将书中内容总结成一个个能够让读者接受的“十项”清单。一开始我们对这个部分感到非常抓狂，但结束后，我们深深地觉得这一设计非常有意义。在你已经阅读完关于大数据的最佳实践，以及我们在这一部分中精炼出的“要做”和“不要做”清单之后，相信你也会和我们有同样的感觉。

### 专用术语

我们整理了一些人们在讨论大数据时常常用到的术语。虽然在正文中介绍每一个术语时都会详细地解释它，但我们仍然认为应该将其整理成术语集，这对于读者而言将会是非常有用的资源。

## 书中的图标



注意其中的内容，你能为自己省掉许多麻烦。



如果把这些小知识点忘记了，你会觉得很遗憾的。



用这个图标标识出来的内容，是我们想要你特别留意的内容。





在这里你会找到一些技术性的知识点。

## 从哪里开始

我们已经大致向你描绘了大数据的轮廓，还介绍了其中的关键组成部分。我们建议你先阅读本书的前4章，你将能够了解大数据到底是什么、实现大数据都需要哪些关键技术的帮助。接下来的两章介绍了大数据所需的基础框架，这是理解大数据非常关键的环节。在随后的8章中，我们详细介绍了构成大数据的各种不同的数据结构。

你可以按照章节顺序从头到尾阅读本书，但如果你不是这么循规蹈矩的人，我们努力保持“*For Dummies*”系列的风格，尽量保持每个章节的独立性，这样的话，你就可以直接前往你最感兴趣的话题开始阅读了。不管你从哪里开始，我们都祝你读得开心。

很多小节本身都可以扩充成一本完整的书。大数据本身以及它融合的新兴技术在Hurwitz & Associates公司是我们非常关注的话题，我们真诚邀请你访问我们的网站，上面有很多我们的博文和见解：[www.hurwitz.com](http://www.hurwitz.com)。

有时，John Wiley & Sons公司会更新它发行的技术书籍。如果本书在技术上有任何更新，这些更新将被发布在[www.dummies.com/go/bigdatafdupdates](http://www.dummies.com/go/bigdatafdupdates)上。

# 目 录

<b>第一部分 大数据入门</b> ..... 1	聚合大数据.....25
<b>第1章 大数据基础</b> ..... 3	管理不同类型的数据 .....25
数据管理的演化过程..... 4	将不同类型的数据整合到大数据环
理解数据管理的几个关键..... 5	境中 .....25
关键1: 创建可管理的数据结构 ... 5	<b>第3章 当老古董遇上新生代: 分布</b>
关键2: Web和内容管理 ..... 7	<b>式计算</b> ..... 27
关键3: 管理大数据 ..... 7	分布式计算简史.....27
大数据的定义..... 9	感谢DARPA .....27
构建成功的大数据管理架构.....10	可持续模型的价值 .....28
捕捉、组织、集成分析与模拟 .....10	了解分布式计算基础.....29
建立架构基础 .....11	为什么大数据需要分布式计算 .....29
性能问题 .....13	计算经济的改变 .....30
传统与高级分析 .....15	时延带来的问题 .....30
大数据之旅.....16	当需求遇上解决方案 .....31
<b>第2章 研究大数据类型</b> ..... 17	获取所需的性能.....31
定义结构化的数据.....18	<b>第二部分 大数据的技术基础</b> .....33
探索大结构化数据源 .....18	<b>第4章 深入大数据技术组件</b> ..... 35
理解关系型数据库在大数据中的	探索大数据栈.....36
角色 .....19	第0层: 带冗余的物理基础架构 .....37
定义非结构化数据.....21	物理冗余网络 .....38
探索非结构化数据源 .....21	管理硬件: 存储与服务器.....39
理解CMS在大数据管理中的角色 ...23	基础架构操作 .....39
理解实时需求和非实时需求.....23	

第1层：安全框架 .....	39	微软Azure .....	64
进/出应用程序和互联网的界面与 接口 .....	40	OpenStack .....	65
第2层：可操作数据库 .....	42	在使用云服务时需要注意什么 .....	65
第3层：组织数据服务与工具 .....	43	<b>第三部分 大数据管理</b> .....	<b>67</b>
第4层：可分析的数据仓库 .....	44	<b>第7章 操作型数据库</b> .....	<b>69</b>
大数据分析 .....	45	RDBMS在大数据领域的重要性 .....	71
大数据应用程序 .....	46	非关系型数据库 .....	72
<b>第5章 虚拟化及其如何支持分布式 计算</b> .....	<b>47</b>	Key-Value型数据库 .....	73
理解虚拟化的基本知识 .....	47	文档数据库 .....	75
在大数据中使用虚拟化的重要性 .....	48	MongoDB .....	76
服务器虚拟化 .....	50	CouchDB .....	77
应用程序虚拟化 .....	50	纵列数据库 .....	78
网络虚拟化 .....	51	图形数据库 .....	79
处理器和内存虚拟化 .....	51	空间数据库 .....	81
数据和存储虚拟化 .....	52	混合持久化 .....	83
使用Hypervisor管理虚拟化 .....	53	<b>第8章 MapReduce基础</b> .....	<b>85</b>
抽象化与虚拟化 .....	54	MapReduce溯源 .....	85
实现在大数据中的虚拟化 .....	54	理解Map函数 .....	86
<b>第6章 云和大数据</b> .....	<b>56</b>	添加Reduce函数 .....	88
大数据领域中的云 .....	56	结合Map和Reduce .....	89
理解云部署和分发模型 .....	57	优化MapReduce .....	91
云部署模型 .....	57	硬件/网络拓扑 .....	92
云分发模型 .....	59	同步 .....	92
大数据需要云 .....	60	文件系统 .....	92
在大数据中使用云 .....	61	<b>第9章 探索Hadoop的世界</b> .....	<b>94</b>
大数据云市场的服务提供商 .....	62	谈谈Hadoop .....	94
亚马逊公共弹性计算云（EC2） .....	63	理解Hadoop分布式文件系统 （HDFS） .....	95
谷歌的大数据服务 .....	64		