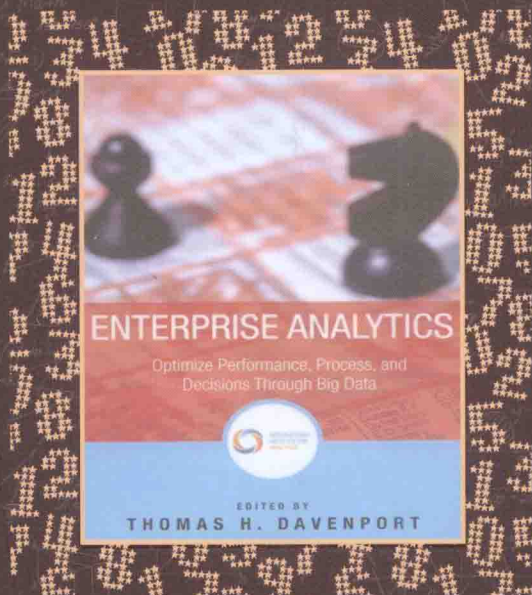


大数据分析

数据驱动的企业绩效优化、过程管理
和运营决策

[美] Thomas H. Davenport 编
吴峻申 译



ENTERPRISE ANALYTICS

Optimize Performance, Process, and Decisions Through Big Data



机械工业出版社
China Machine Press

ENTERPRISE ANALYTICS

Optimize Performance, Process, and Decisions Through Big Data

大数据分析

数据驱动的企业绩效优化、过程管理
和运营决策

[美] Thomas H. Davenport 编
吴峻申 译



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

大数据分析: 数据驱动的企业绩效优化、过程管理和运营决策 / (美) 达文波特 (Davenport, T. H.) 编; 吴峻申译. —北京: 机械工业出版社, 2015.3

(数据科学与工程丛书)

书名原文: Enterprise Analytics: Optimize Performance, Process, and Decisions Through Big Data

ISBN 978-7-111-49184-2

I. 大… II. ①达… ②吴… III. 企业管理 - 数据管理 IV. F270.7

中国版本图书馆 CIP 数据核字 (2015) 第 015569 号

本书版权登记号: 图字: 01-2013-0211

Authorized translation from the English language edition, entitled *Enterprise Analytics: Optimize Performance, Process, and Decisions Through Big Data*, 9780133039436 by Thomas H. Davenport, published by Pearson Education, Inc., Copyright © 2013 by International Institute for Analytics.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

Chinese simplified language edition published by Pearson Education Asia Ltd., and China Machine Press Copyright © 2015.

本书中文简体字版由 Pearson Education (培生教育出版集团) 授权机械工业出版社在中华人民共和国境内 (不包括中国台湾地区和香港、澳门特别行政区) 独家出版发行。未经出版者书面许可, 不得以任何方式抄袭、复制或节录本书中的任何部分。

本书封底贴有 Pearson Education (培生教育出版集团) 激光防伪标签, 无标签者不得销售。

大数据分析

数据驱动的企业绩效优化、过程管理和运营决策

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: 关敏

印刷: 三河市宏图印务有限公司

版次: 2015 年 3 月第 1 版第 1 次印刷

开本: 186mm × 240mm 1/16

印张: 12

书号: ISBN 978-7-111-49184-2

定价: 59.00 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) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

HZBOOKS | 华章IT | Information Technology



The Translator's Words 译者序

牛冰天

字念心 笔名冰天

最近几年，“大数据”已变成一个非常流行的术语。很多知名的从业人员已对此撰写了大量经典书籍，而这本书就是这样的书籍之一。本书由美国国际数据分析研究所的一群达人根据他们自己所写的讲义、方案和用例故事汇聚而成。书中到处显现了他们对数据分析的心得、观点和经验之谈。

本书内容主要分成五大部分。第一部分是对数据分析的一个综述。第二部分介绍如何在企业内部使用数据分析，并展开全部细节提醒人们需要注意哪些地方。第三部分则介绍一些应用数据分析的技术和方法论。通过叙述这些内容，消除人们对数据分析的陌生感，让人们更有信心使用数据分析。我相信读者如果能在实际中经常运用这些技术，必定会事半功倍。而第四部分是译者个人认为最精华的部分。众所周知，不管何种概念和理念要真正落到实处，必定要人来做。而人是一种复杂的生物，并不像工具和资源那样进行简单的操作就能达到我们想要的目的。因此，我认为读者应该仔细阅读这一部分的内容，看看这些专家学者是如何让人来更好地进行数据分析工作的。最后一部分有很多实例，我可以提前“剧透”一下，该部分的作者大多数都是客户公司的主管或负责人，他们将自己在工作中运用数据分析的一些感想和经验完完全全地展示出来了，让没有使用过数据分析的人不用再摸着石头过河，而是根据前辈的经验勇敢前行。

也正因如此，译者在翻译过程中对这些闪耀着思维“火花”的见解深为叹服，因此也更加感到压力巨大，因为如果不能原汁原味地将这些内容展现给读者，就会让译者怀有“暴殄天物”的负罪感。因此，译者在翻译过程中对很多词句做了大量斟酌和修饰工作，就是为了让读者阅读起来能更加顺畅并更易于理解各位作者的思想。老实说，翻译是门“力气活”，鉴于译者个人翻译能力的局限，书中肯定也有不尽如人意之处，希望读者谅解。我也欢迎大家随时提出修改和批评意见。那么现在，各位就可通过阅读本书来

看看译者翻译的成果了。谢谢各位！

最后，感谢父母亲在本人翻译此书期间，在生活起居上给予的无微不至的关怀，以及无暇理会其他事情时的理解和支持。好了，译者的废话就到此为止了。请欣赏各位达人的佳作吧！

吴峻申
于外高桥办公室

Foreword 序

收集于此书中的研究成果体现了一批人所做的贡献，正是他们让美国国际数据分析研究所（International Institute for Analytics, IIA）在今天能如此成功。本书是三分勤奋、两分持续不懈和少许缘分的成果，也正是这一成果让我们这个羽翼未丰的公司开始“起步”。

首先说勤奋。这里我们显然要感谢 Tom Davenport 把最初收集的 IIA 研究成果编辑和整理成册。关于 Tom 所整理的这些原材料，我们也感谢在 IIA 成立头两年中贡献这些独到的研究成果的教职人员，特别是 Bill Franks、Jeanne Harris、Bob Morison、James Taylor、Eric Peterson 和 Keri Pearlson。当我们在扩展医疗行业数据分析覆盖范围时，Marcia Testa（哈佛公共卫生学院）和 Dwight McNeil 也发挥了关键的作用。在最初的零售行业数据分析研究过程中，Ananth Raman（哈佛商学院）和 Marshall Fisher（沃顿商学院）也对我们进行了指导，给了我们不少帮助。我们期待在这两个领域里还会有其他类似于本书的书籍出现。当然，也要感谢所有愿为我们的诸多研究主题花费时间参与实践的企业。

我们也感谢整个 SAS 团队对 IIA 持续的信任，他们早已验证了我们的使命和方向，并从那时起就展示出对我们的信任。特别要感谢 Scott Van Valkenburgh（所有在白板前进行的交流讨论）、Deb Orton、Mike Bright、Anne Milley 和 Adele Sweetwood。我们也感谢 IIA 其他供应商对我们的支持，包括埃森哲、戴尔、英特尔、SAP 和 Teradata 公司。

本书的出版也要归功于 IIA 两位牛人的坚持不懈。Katherine Busey 是 IIA 在波士顿的第一位员工，正是此人让在 Pearson 公司的 Jeanne Glasser 相信 IIA 的研究成果不仅仅对我们的研究客户来说是值得阅读的东西。还要感谢 Callie Youssi，是她协调了 IIA 所有的教学研究活动，而这并不是一个轻松简单的活。

我们很难想象 Tom 没有妻子和管家 Jodi 助他一臂之力会如何。谢谢你们俩把全部“赌注”压在了身为企业家的我身上，特别是在我的企业处于充满挑战的第一年时。

还有少许缘分，Tom 和我们都承蒙 Eric McNulty 的照顾，是他富有远见地让我们走在一起成为 IIA 的喉舌，并帮助我们设定了本书早期的出版和研究标准。

Eric McNulty

——Jack Phillips

美国国际数据分析研究所首席执行官

本书的出版离不开许多人的帮助。首先，我要感谢我的家人，正是他们让我有机会从事这项工作。其次，我要感谢我的同事，正是他们让我有机会从事这项工作。最后，我要感谢我的读者，正是他们让我有机会从事这项工作。

前言：企业数据分析“新世界”

Thomas H. Davenport

数据分析的兴起

数据分析并不是个新东西——我已经发现过 1954 年有关企业数据分析团队内容的参考资料——但比起以前，它们在商务和组织运营中看起来更加重要了。用于决策制定和管理的数据分析方法的兴起缘于下列几个因素：

- 在各种不同的业务信息系统中被用来分析的数据量已经有了惊人的增长。
- 强大且廉价的分析所有这些数据的计算机和软件。
- 在企业组织内部，大量训练有素的经理被升迁至具有相应职责的职位上。
- 人们需要区分产品和报价之间的区别，优化价格和库存资产并理解是什么驱动了业务绩效的各个方面。

也正因此，很多因素向人们表明，有关数据分析的计划、工作和组织正在在全世界各地涌现。比如，根据 LinkedIn 的数据显示，在 1990 ~ 2010 年间，从事数据分析和数据科学工作的人数已增长了 10 倍。几乎每家大型咨询公司都已提出了一种数据分析实践。根据 Google Trends (趋势) 显示，在 2005 ~ 2012 年间，使用术语“数据分析”来进行搜索的数量已增长了 20 多倍；在 2010 年，使用术语“大数据”（不知道什么时候开始有这个定义的）来进行搜索的数量呈现了一个更加惊人的上升趋势。我们这个时代已经被称为“数据分析时代”、“算法时代”，在将数据分析应用到职业棒球运动的小说和电影出现后，又被称为“点球成金的时代”。^①

① 原文这里是 Money-ball，这是来源于一部名为《点球成金》、根据同名小说改编而成的电影，影片主要讲述了一位美国职业棒球队经理和他的朋友采用数学建模的方式挖掘棒球人才的故事。——译者注

企业数据分析

在数据分析中，正在逐渐引起重视的某个重要属性是，它已成为一项“企业”资源，至少对许多企业来说是如此。也就是说，人们不再将数据分析分离成几小块内容，如市场调研、精算或质量管理，而是意识到数据分析能力能让整个企业受益。人们正在集中管理不同的团队，或者说就算没有集中管理，至少彼此之间也正在进行协调和沟通。在整个企业中，人们正在储备和评估那些数据分析方面的人才。企业级别的团队正在决策他们自己的计划、举措和各部分内容之间的优先次序。其目标是最大化对企业的影响。

因此本书的主题和很多章节都有关于以企业级别的方式管理数据分析的理论和实践。对于首席数据分析官或正出现的一些能承担这个角色的人来说，即使人数依旧很少，如果说存在着有一套能指导他们工作的方针，那么这本书能提供很多方针中的内容。虽然我们还没有让数据分析成为一套被人们广泛认可的业务功能，但我们正清晰地朝着这一方向迈进。

“大数据”的兴起

人们对数据分析的热情因大数据的流行而倍增。这个概念指的是那些太冗长又不太结构化，且很难通过传统手段来管理和分析的数据。但这种定义显然是相对的，会随着时间的流逝而变化。目前，“太冗长”通常意为流转在 1PB 容量中 (1000TB)[⊖] 的数据库或数据；比如，Google 公司每天要处理大约 24PB 的数据。而“不太结构化”通常意为那些不易于放入常规数据库传统行列里的数据。

大数据的例子包括众多大量的在线信息，包括来自于网络社交媒体内容（推特、博客、论坛帖子）的“点击流数据”。此外，大数据还包含来自零售行业和犯罪情报收集环境中的视频数据以及表演给人们看的娱乐节目视频。它也包括呼叫中心和情报监测中的音频数据。在生命科学领域中，它还包括生物研究与医药中的基因和蛋白质数据。

很多 IT 供应商和解决方案提供商以及他们的某些客户都只把这个术语当做数据分析中另外一个热门的词汇，或者是当做为了更好地理解业务而管理和分析的数据。但它不仅仅是供应商的炒作；如果能以一个稳定的基础分析大数据，那么还会产生相当多的业务效益。

那些在大数据领域比较卓越的公司都已开始使用其他诸如泛在传感器和“物联网”这样的新技术。事实上，每一台机械或电子设备都会留下一条描绘其性能、所处方位或状态的痕迹。这些设备和使用它们的人都可以通过导向其他大量数据源的互联网来进行沟通。当所有

⊖ 1TB=10³GB, 1PB=10⁶GB。——译者注

这些比特数据能和有线或无线电话、电缆、人造卫星等媒质结合在一起的时候，未来的数据就会变得更大。

采用这些工具的公司最终会以最细粒度的级别理解他们所处的业务环境并快速适应它。他们将会通过监控和分析使用模式来区分商业产品与服务。并且在生命科学领域，有效地使用大数据也能理所当然地帮助人们治愈那些最危险的疾病。

在下一个十年，几乎每个行业和业务职能都有希望被以此为基础的大数据分析而改变。早早地开始使用大数据的企业能获取重大的竞争优势。“小数据”时代那些早期的数据分析竞争者（包括美国第一资本银行、美国前进保险公司和万豪酒店）已经脱颖而出，取得了相当大的竞争优势；对于那些想把握住大数据机遇的公司来说，现在正是一个最佳时机。

所有这种数据的可用性意味着几乎所有业务或企业活动都可被看作大数据问题或计划。在多数机器都已有一个或多个微处理器的制造业中，一种大数据趋势已经形成。拥有无数客户接触点和点击流数据的消费者市场也已成为一个大数据问题。各国政府也已开始意识到他们坐拥等待被分析的庞大数据集合。Google 甚至把汽车租赁也描述为一个大数据问题。

本书主要以小数据分析为基础，偶尔也指大数据、数据科学家和其他相关主题。当然，很多来自传统数据分析理论的想法也和大数据分析高度相关。

IIA 和为本书所做的研究

在过去的 15 年左右，我一直在对数据分析进行研究。2010 年，一位信息行业创业者 Jack Phillips 和我一起创立了 IIA。这家依旧很年轻的组织的创建初衷是对数据分析和数据分析技术的用户及供应商进行研究并提供咨询服务。在这之前，我已带领组建了一个有关数据分析赞助研究计划的团队。但我知道这距离产生相应的研究成果内容，还有很长一段路要走。

对研究所最早的支持来自业内领先的数据分析供应商 SAS。我们也和 SAS 的核心合作伙伴一起工作过，这包括 Intel、埃森哲和 Teradata 公司。稍后，其他的核心供应商，包括 SAP 和 Dell 公司，也成为我们 IIA 的赞助商。IIA 的赞助商不仅为研究提供了财务支持，还提供了担任 IIA 教员职位的数据分析理论研究者及其思想领袖。

在招募了其他具有教学或独立咨询背景的教员后，我们开始有了研究成果。你会在本书中看到好几个研究产出实例。IIA 产生了三种类型的产出：研究摘要（关于某特定数据分析主题的文档，通常为 3 ~ 5 页）；领导力实践摘要（有关公司领导力或典型数据分析议题的研究案例）；关于网络会议、网络广播和音频会议的报道。虽然在某些情况下，我们要把不止一

篇摘要或文档合并在一起凑成本书中的一章，但重点在于这些文档是简短和易于理解的。

在手边有了某些原始研究成果后，我们开始招募能参与 IIA 工作的公司或组织。我们最初的方案是关注那些常见的“企业”主题——系统化数据分析的方式、用于数据分析的技术架构、设计方式等。我们发现人们对这些主题反响热烈，所以其中有很多主题都涵盖在本书中。开始有大量的公司和个人成员参与了 IIA。

然而，对我们的想法最强烈的反应还是在特定行业研究方面。虽然看上去大多数公司都对那些有关数据分析最佳实践的常用材料相当有兴趣，但它们更感兴趣的还是如何对医疗和零售（我们首先进行特定行业研究计划）的行业运用数据分析。但这项研究不在本书的特色范围之内——我们会为有关特定行业数据分析的内容撰写几本其他的书籍——但我们会把来自这些行业的领先实践摘要包括进来作为本书中的几章。

本书结构

本书中的所有章节内容都产自或来源于 IIA 的项目。所有作者（或每一章中至少有一位作者）都是 IIA 的教员。有一些主题以类似（但不完全相同）的形式出现在某些杂志文章或书籍中，但大多数还没有在 IIA 外部发表过。本书描述了几大主题。第一部分是对数据分析及其价值的综述。第二部分讨论了数据分析应用。第三部分涵盖了一些数据分析技术。第四部分描述了数据分析中与“人”相关的内容。第五部分由一些在企业内进行数据分析活动的案例研究组成。

About the Author 作者简介

Thomas H. Davenport 是 IIA 的创始人之一，现任研究总监，也是哈佛商学院客座教授、巴布森学院特聘教授和德勤会计事务所高级顾问；被《*Optimize*》杂志评为业内第三领先的商务策略分析师（仅次于 Peter Drucker 和 Tom Friedman）。Davenport 也是一位世界著名的思想领袖，他帮助成百上千家公司重振了他们的管理实践。他的“竞争力分析”理念最近刚刚被《哈佛商业评论》杂志评为过去 10 年最重要的管理理念之一。相关的文章也被评为《HBR》（*Harvard Business Review*，哈佛商业评论）75 年历史中十大必读文献之一。Davenport 在 2010 年 2 月出版了相关书籍《*Analytics at Work: Smarter Decisions, Better Results*》，这也被《*CIO Insight*》杂志评为 15 本必读书籍之一。

Elizabeth Craig 是埃森哲卓越绩效研究院波士顿分院的研究人员。她和 Peter Cheese 以及 Robert J. Thomas 合著了《*Talent-Powered Organization*》（Kogan Page, 2007）。

Jeanne G. Harris 是埃森哲卓越绩效研究院芝加哥分院的高级主管研究人员。她和 Thomas H. Davenport 以及 Robert Morison 合著了《*Analytics at Work: Smarter Decisions, Better Results*》（哈佛商业出版社，2010）。她还与人合著了 2007 年出版的书籍《*Competing on Analytics: The New Science of Winning*》，也由哈佛商业出版社出版。

Robert Morison 是 IIA 企业研究成果订阅的教学带头人。他是一位成就非凡的商业研究者、作者、决策领导者和管理咨询顾问。他也是《*Analytics at Work: Smarter Decisions, Better Results*》（哈佛商业出版社，2010）、《*Workforce Crisis: How to Beat the Coming Shortage of Skills and Talent*》（哈佛商业出版社，2006）和《哈佛商业评论》杂志上的 3 篇文章的合著者，其中一篇还获得了 2004 年 McKinsey 最佳文献奖。他还为大量企业、行业和政府组织做过演讲，并在美国 PBS 电视台的《晚间商业新闻报告》节目中担任过劳动力问题评论员。他最近成为 nGenera 公司的执行副总监和研究主管，之前他在 Concours 集团公司、CSC Index 公司

和美国通用电气资讯服务公司担任过管理职位。

Keri E. Pearlson 博士是信息管理和使用领域的专家。她曾经在一些世界最大型的公司里和 CIO 以及执行主管一起工作过。她善于帮助执行主管们起草企业战略,使其成为具有 Web 2.0 能力的企业,也善于设计和传授高级领导力课程,并精于管理一些有关信息系统高级主管感兴趣的问题的多客户端程序,使其能满足高级主管对信息系统的要求。她还特别专长于帮助 IT 主管们和他们的主管同事一起参与企业战略制定。她也是 IIA 的一位教员,是 CIO 咨询服务公司 KP Partners 的创业合伙人和总裁。

Bill Franks 是 IIA 的一位教员,也是 Teradata 公司全球联盟计划的首席分析总监。他还负责过商业数据分析创新中心(由 Teradata 和 SAS 公司联合赞助),该中心专注于帮助客户追求创新性的数据分析成果。除此之外, Bill 还在 Teradata 公司的高级数据分析领域部门工作过,帮助公司确定企业战略和定位。他也是书籍《*Taming the Big Data Tidal Wave*》(John Wiley & Sons 出版社,2012 年 4 月, www.tamingthebigdatatidalwave.com)的作者。

Eric T. Peterson 是 IIA 的一位教员。他也是 Web Analytics Demystified 公司的创始人。他以从业人员、咨询顾问和分析师的身份在 Web 数据分析领域工作了 10 多年。他还是 3 本畅销的 Web 数据分析书籍的作者:《*Web Analytics Demystified*》、《*Web Site Measurement Hacks*》和《*The Big Book of Key Performance Indicators*》。在 www.webanalyticsdemystified.com, 他是读者最多的 Web 数据分析作家之一。

John Lucker 是德勤咨询公司主要领导人,在那里他负责德勤的高级数据分析和建模实践,这在专业性的服务行业中是处于领先地位的一个数据分析组织。他在高级数据分析、预测建模、数据挖掘、计分和规则引擎以及大量其他的高级数据分析业务解决方案领域里具有极其丰富的工作经验。

James Taylor 是 IIA 的一位教员以及 Decision Management Solutions 公司的 CEO。决策管理系统运用了业务规则、预测分析和优化技术来解决企业今日所面对的最棘手问题,并改变企业开展业务的方式。他在为客户开发软件和解决方案方面具有 20 多年的工作经验。他已带领 Decision Management Solutions 公司帮助在保险、银行、医疗管理和电信行业中具有领先地位的公司取得了成功。

Stacy Blanchard 是埃森哲公司中组织效能服务和人力资本分析部门的领导者,在为企业调整战略、文化和领导力方面具有超过 15 年的工作经验,曾在全球许多客户所处的情景和行业中工作过。她把现实世界中的经验和公认的用来指导与调整最高管理层(C-suite)管理思路的方案结合在一起驱动企业转型过程。在就职于埃森哲之前,她是 Hagberg 咨询集团公司的 CEO,这是一家专门针对企业战略、文化和领导力进行评估、调整与转型的咨询企业。

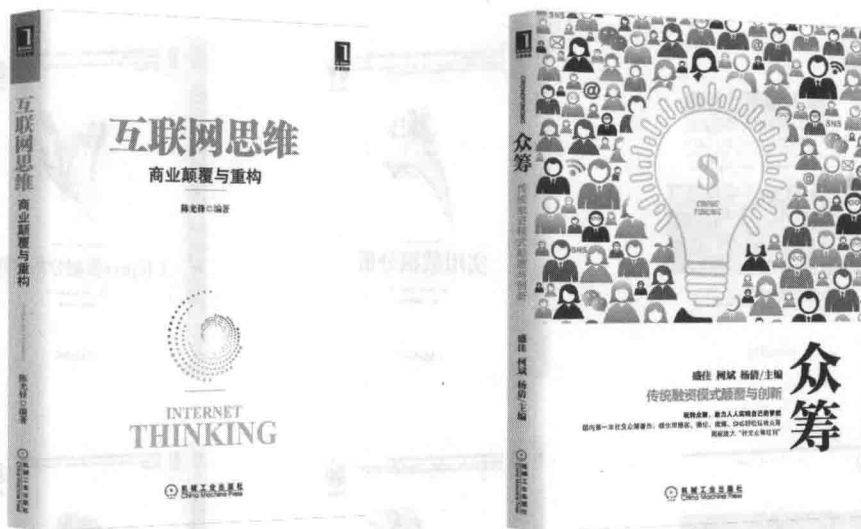
Carl Schleyer 是 Sears Holdings 公司（一家 IIA 赞助商）运营和数据分析主管，负责通过海量数据的收集和分析工作来支持人才与人力资本方面的企业战略及战术。作为这个岗位的一部分工作内容，Carl 创建了 Sears Holdings 公司的第一支数据分析团队，致力于追求纯粹的人力资本。他的工作激情在于通过有影响力的企业决策来揭示数据的价值。Carl 在零售行业具有 20 多年的丰富经验，并在人力资源领域中各种各样的岗位工作过。

Leandro DalleMule 是花旗银行集团公司全球数据分析高级主管。在此之前，他是德勤公司数据分析咨询实践部门的高级经理、通用电气公司的风险管理经理和美孚公司巴西分公司的品牌推广经理。

Callie Youssi 是 IIA 负责研究业务的副总裁。在 IIA 遍布世界各地的分支机构中的教员发现数据分析中那些最引人注目的实际应用时，她能对它们进行构建、管理和支持。她负责汇总和分析 IIA 客户最感兴趣的领域并确保有一支雄厚的师资后备力量来解决这些领域中的问题。

Katherine Busey 是 IIA 负责业务开发的副总裁，负责为 IIA 拓展全球性的商务发展机遇。她和 IIA 的供应商、商业伙伴和研究客户一起工作，发现数据分析领域中那些新的发展趋势并将这些内容汇总起来告诉供应商和从业人员。

推荐阅读



互联网思维：商业颠覆与重构

作者：陈光锋 ISBN：978-7-111-45551-6 定价：49.00元

国内首部互联网思维系统化专著
深度解析十大核心互联网思维，全面揭秘互联网七字诀

互联网市场上有成有败，BAT为何屹立不倒？小米为何能够在如此短的时间内创造神话？传统企业互联网化浪潮下的生存之道是什么？这其中的奥妙就在于互联网思维，意欲冲向互联网浪潮之巅，产品、营销、运营都离不开互联网思维。

众筹：传统融资模式颠覆与创新

作者：盛佳等 ISBN：978-7-111-46681-9 定价：49.00元

作为互联网融资的革新模式，众筹是搅局还是颠覆？
众多经典案例，揭秘如何利用社交网络玩转众筹

我们在众筹前加上定语“社交”，在社交前加上定语“网络”，我们明白，基于网络社交关系的众筹是一次革命，其基石正是本书力倡的信息对称的权利与投资民主的权利。众筹是企业发展的加速剂，而且这一切都能在社交网络上完成，众筹实在是一件神奇的事情。