

WILEY

“十二五”
国家重点图书出版规划项目

O'Reilly
媒体 总经理
Jim Stogdill
倾情作序推荐



新信息时代商业经济与管理译丛

大数据分析 决胜互联网金融时代

BIG DATA BIG ANALYTICS

EMERGING BUSINESS INTELLIGENCE AND
ANALYTIC TRENDS FOR TODAY'S BUSINESSES

【美】Michael Minelli 【美】Michele Chambers 【美】Ambiga Dhiraj◎著
阿里巴巴集团商家业务事业部◎译

IBM研究院首席科学家 Jeff Jonas

美国国家得宝公司主席及CEO Frank Blake

美国凯悦酒店首席信息官 Mike Blake

美国职业棒球大联盟媒体公司首席技术官 Joe Choti

Kaggle公司创始人及CEO Anthony Goldbloom

佩斯大学客户智能硕士工程创始人之一 Tony Branda

联袂推荐



人民邮电出版社
POSTS & TELECOM PRESS

专家赞誉

“完美的细节在某种程度上提供了一种洞察方法，它以真实世界的应用来洞察这个千变万化的商业世界，与你的经验背景无关。”

——弗兰克·布莱克 (Frank Blake)，美国家得宝公司 (The Home Depot) 主席及CEO

“大数据就像一堆巨大的拼图碎片。一旦从背景出发整体考虑，渐渐展开的画面将指导着更为精准的行动，最终得到更好的商业成果。这本书很及时，写得也很好，它可以帮助商界和IT界的管理人士在大数据的世界加速前进。”

——杰夫·乔纳斯 (Jeff Jonas)，IBM研究院首席科学家

“大数据影响了所有人，而本书是初学者必备的一本与应用有关的好书。迈克尔 (Michael) 和他的团队通过简单、易消化的风格抓住了大数据的本质，它值得与你的整个管理团队一起分享。”

——迈克·布莱克 (Mike Blake)，美国凯悦酒店首席信息官 (CIO)

“每天企业从各个部门产生大量的数据——网站、销售团队、社交媒体、数字图片及多媒体、交易记录等。对于商界及IT界的管理人员来说，本书是他们拥抱‘大数据世界’的必备好书，它能帮助他们理解大数据将如何改变商业竞争。”

——乔·焦蒂 (Joe Choti)，美国职业棒球大联盟媒体公司首席技术官 (CTO)

“大数据正在改变现代商业社会。它决定了银行如何授权贷款，对冲基金如何交易……本书向读者清晰地解释了他们需要了解的所有知识，从大数据对于技术的意义到它如何被应用。”

——安东尼·高德布卢姆 (Anthony Goldbloom)，Kaggle公司创始人及CEO

“作为客户智能方面的教授，本书将帮助我的学生理解大数据的多学科本质，以及如何解决销售问题，如何为客户更好地服务，如何使他们的公司经营地更好。极力推荐阅读此书！”

——托尼·布兰达 (Tony Branda)，佩斯大学 (Pace University) 客户智能硕士工程创始人

WILEY



Copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal.

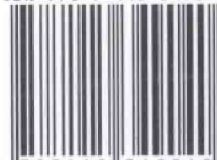
分类建议：经济管理

人民邮电出版社网址：www.ptpress.com.cn

sas



ISBN 978-7-115-34904-0



9 787115 349040 >

ISBN 978-7-115-34904-0

定价：49.00 元

WILEY

“十二五”
国家重点图书出版规划项目

O'Reilly
媒体 总经理
Jim Stogdill
倾情作序推荐



新信息时代商业经济与管理译丛

大数据分析 决胜互联网金融时代

BIG DATA BIG ANALYTICS

EMERGING BUSINESS INTELLIGENCE AND
ANALYTIC TRENDS FOR TODAY'S BUSINESSES

【美】Michael Minelli 【美】Michele Chambers 【美】Ambiga Dhiraj◎著
阿里巴巴集团商家业务事业部◎译

IBM研究院首席科学家 Jeff Jonas

美国国家得宝公司主席及CEO Frank Blake

美国凯悦酒店首席信息官 Mike Blake

美国职业棒球大联盟媒体公司首席技术官 Joe Choti

Kaggle公司创始人及CEO Anthony Goldbloom

佩斯大学客户智能硕士工程创始人之一 Tony Branda

联袂
推荐

人民邮电出版社

图书在版编目 (C I P) 数据

大数据分析：决胜互联网金融时代 / (美) 梅内里 (Minelli, M.), (美) 钱伯斯 (Chambers, M.), (美) 帝拉吉 (Dhiraj, A.) 著 ; 阿里巴巴集团商家业务事业部译 . -- 北京 : 人民邮电出版社, 2014.8 (2014.8 重印)
(新信息时代商业经济与管理译丛)
ISBN 978-7-115-34904-0

I. ①大… II. ①梅… ②钱… ③帝… ④阿… III. ①统计数据—统计分析 IV. ①0212.1

中国版本图书馆CIP数据核字(2014)第045953号

版权声明

Michael Minelli, Michele Chambers, Ambiga Dhiraj.
Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses.
Copyright © 2013 by John Wiley & Sons Ltd.
All rights reserved. This translation published under license.
Authorized translation from the English language edition published by Wiley Publishing, Inc..
本书中文简体字版由 John Wiley & Sons Ltd 公司授权人民邮电出版社出版, 专有版权属于人民邮电出版社。

-
- ◆ 著 [美]Michael Minelli [美]Michele Chambers
[美]Ambiga Dhiraj
译 阿里巴巴集团商家业务事业部
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>
大厂聚鑫印刷有限责任公司印刷
- ◆ 开本: 700×1000 1/16
印张: 12.5
字数: 175千字 2014年8月第1版
印数: 3 501—4 500册 2014年8月河北第2次印刷

著作权合同登记号 图字: 01-2013-6015号

定价: 49.00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

本书第 1 章主要介绍了大数据分析的定义和由来、大数据的重要性，以及大数据时代的历史背景和发展趋势。第 2 章通过对成功企业领导者的采访，详细介绍了大数据分析在信息、金融、传媒、医疗等领域的使用案例，并通过这些案例帮助读者了解如何将大数据分析方法应用到商业活动中。第 3 章详细介绍了大数据分析采用的一系列技术手段及案例，并指出大数据技术应用到实际生活中的一些关键问题。第 4 章介绍了承载大数据应用的计算、存储等底层基础设施的关键技术，并讨论了大数据计算的瓶颈及近年来涌现出的新技术。第 5 章详细介绍了在大数据时代商业分析的方法及涉及的关键技能，以及如何对分析结果进行归纳总结，并最终对企业的决策产生影响。第 6 章主要介绍了数据科学家的定义，以及数据科学和决策科学人才培养的方法论。第 7 章讨论了涉及数据隐私及数据安全的相关问题。

本书可供企业经理和管理人员阅读，通过本书可以使他们学习如何最有效地利用信息资源。大数据领域的相关从业者通过本书可以了解如何在大数据方向进行成功的实践。由于本书在艰深的战略理论和通俗易懂的示例之间做到了平衡，对技术性很强的话题，通过故事、比喻和类比的方式使读者更容易理解和接受，因此对大数据领域感兴趣的读者也可进行阅读。

献词

献给我的妻子 Jenny 和我们的 3 个不可思议的孩子, Jack、Madeline 和 Max; 同时献给一直陪伴着我的父母。

—— 迈克尔·梅内里 (Michael Minelli)

献给我的儿子 Cole, 他是我的生命之光, 教会了我换位思考; 也献给支持着我的家人们, Lisa Patrick、Pei Yee Cheng 和 Patrick Thean; 最后, 献给给予我慷慨帮助的同事们, Bill Zannine、Brian Hess、Jon Niess、Matt Rollender、Kevin Kostuik、Krishnan Parasuraman、Mario Inchiosa、Thomas Baeck、Thomas Dinsmore 以及 Usama Fayyad。

—— 米歇尔·钱伯斯 (Michele Chambers)

献给来自世界各地、对决策科学怀有无比激情的 Mu Sigma 公司的同事们。

—— 安碧嘉·帝拉吉 (Ambiga Dhiraj)

作者介绍

迈克尔·梅内里 (Michael Minelli)

是一个市场和销售专家,在商业分析解决方案领域有 16 年的经验。现在,他是销售和全球联盟的副总裁、万事达卡公司的信息服务顾问,他负责带领销售和战略联盟团队为万事达卡的数据资产赚钱。万事达卡数据资产包含 18 亿张银行卡,代表了全球 210 个国家和地区的 3 400 万商家。在成为万事达卡顾问之前,梅内里负责革命分析 (Revolution Analytics) 的销售团队,他负责新商业开发和战略联盟团队,为公司的软件和服务提供商提供开源的 R 工程的支持。

在加入革命分析 (Revolution Analytics) 之前,梅内里作为 SAS 公司的销售总监和全球客户经理,创建了他自己的基金会并且专长于市场和销售分析解决方案。在 SAS 的 11 年中,他担任过很多职责,创造了一个涵盖客户智能、风险管理、供应链和财务的大型销售分析项目的成功记录。

在加入 SAS 之前,他是 MMS 公司的一名销售人员,一家专注于 ERP 和商业智能的创业公司 (MMS 公司在 1999 年被 AMOS 公司收购)。

本书是梅内里的第二本著作。他之前是《Partnering with the CIO》的作者之一。

迈克尔拥有佩斯大学的市场营销学士学位。

米歇尔·钱伯斯 (Michele Chambers)

拥有诺瓦东南大学的计算机工程学士学位和杜克大学的 MBA 学位。她是一位有着 25 年技术经验的创业高管。钱伯斯女士曾经是负责 IBM 公司大数据分析的副总裁，她的团队负责充分利用 IBM 的包含 IBM Netezza 并行计算平台在内的大数据平台去和客户打交道，钱伯斯女士同时负责为 IBM Netezza 产品规划其愿景、战略、销售以及市场策略。她致力于帮助公司拓展新的领域，应用分析优化的方法获取高商业价值并创造可持续的差异化市场。

钱伯斯女士对结果和成长性分外关注，她成功推行了几项业务，包含 Netezza 公司的高级分析解决方案，这项业务仅仅在 6 个月内就创造了 8 000 万美元的税收。另外，她还成功创立了一系列 SAP（企业管理系列软件）解决方案，这套方案在第一年就创造了超过 1 000 万美元的税收，并引发了早期的供应链软件分析业务。

在她业余时间，作为单身妈妈的钱伯斯女士往往显示出通过她的数学才能解决真实世界难题方面的超前能力。

安碧嘉·帝拉吉 (Ambiga Dhiraj)

是 Mu Sigma 公司客户交付部的经理，Mu Sigma 是决策科学和分析解决方案的首席供应商，帮助企业制度化地做出数据驱动的决策。在 Mu Sigma 公司的职业生涯中，安碧嘉带领她的客户交付团队，解决在多个垂直行业市场的企业市场营销、供应链和风险分析领域的商业问题。利用在计算机工程、研究和咨询领域的丰富经验，她用一种整合数据、商业和科技的独特方法帮助 Mu Sigma 公司在决策科学领域发展创新的解决方案。

中文版序

托夫勒在《第三次浪潮》一书中说过，如果说 IBM 的主机拉开了信息化革命的大幕，那么“大数据”才是第三次浪潮的华彩乐章。

《大数据分析：决胜互联网金融时代》正是这样一本书，他虽然不能做为工程师具体实现大数据的工程文档，但却是实现大数据的技术方向参考，也是商家、数据分析师、数据科学家学习方向的参考。

《大数据分析：决胜互联网金融时代》主要讲述了什么是大数据，它为什么如此重要？廓清了大数据的定义以及在现代社会中的重要性。采用 3V（海量、速度、多样性）描述对大数据最基本特征的归纳。

仅仅从多样性、体量和增长速度上把握，无法说清普通数据与大数据的本质区别。现代企业中的大数据分为两种：一种是大的交易数据（企业、消费者、供应链等）；一种是大的交互数据（SNS、互联网、物联网、移动终端）。也就是说个人数据、物质世界数据、社会数据三维联合分析描述了事物。

作者在第 2 章中讲述了大数据在各行业中的发展以及应用。例如对市场营销（数字营销，数据库营销）、反欺诈、风险控制、医药及医疗健康领域、计算广告学等领域予以深入阐述，对大数据的应用以及未来方向实例化。

大数据的发展是在大规模计算技术（分布式计算、存储、管理和分析技术）的基础上发展起来的，第 3 章中对这些内容进行了详尽论述。

数据存储、数据处理这两个系统是大数据的基础，而并行计算是满足速度、规模的唯一平台，在第 4 章信息管理中重点进行了阐述。同时进一步

阐述了大数据计算的局限性（主要体现在磁盘上限、I/O 上限、内存上限、CPU 上限）。

大数据的新兴技术包括 SSD、GPU 图形处理运算单元的应用。当然我们看到很多公司包括阿里巴巴已经在开始使用这些新技术。

在第 5 章中作者还从商业分析角度，介绍了大数据分析的分类，并着重介绍了数据分析的“最后一公里”（即如何将分析应用到生产实践），为在企业内部应用数据分析提供了框架和步骤。沟通周期中的每一步都是企业分析综合的关键所在，应用数据分析的步骤为：沟通、执行、评估、协调激励机制。

在第 6 章中，作者从个人发展方面阐述了大数据的各种角色及拥有的技能，以及数据科学家的发展方向。

在隐私极度敏感的今天，如何冲破个人隐私、商业利益、行政垄断的限制，实现数据的低成本、高效率、大规模的聚集和整合？作者在第 7 章中着重阐述了保护隐私的七条原则，以及对信息隐私性的定义（个人信息、敏感信息、其他信息）并举例说明了各类信息的例子。

这是一本好书，它从各个角度对大数据进行了阐述。它廓清了大数据的基本概念、技术、商业实践、信息管理中存储和处理，以及数据安全和隐私方面的世界通行做法。

感谢我的同事（张中、征宇、圣香、依韵）在繁忙的工作之余翻译了这本书，是他们辛勤的劳动，换来了《大数据分析：决胜互联网金融时代》与大家见面的机会。

阿里巴巴集团
高级数据挖掘专家
张雁冰（纶巾）
2014 年 6 月

序言：大数据与企业进化

当我的朋友迈克尔·梅内里（Mike Minelli）邀请我来写这个序言的时候，一开始我都不确定我应该在纸面上写什么。序言通常是由一部分本书总结和一部分此领域概述组成。但是当我读完 Mike 寄给我的草稿，我意识到这真是一本好书，以至于它完全不需要前述序言的任何一部分。不需要任何来自我的额外帮助，本书就能让你深刻理解现在正在发生的事情以及它发生的原因，并且能让你在这个以数据为中心的时代变革中发现你所在产业的机会。不仅如此，这本书还充满了该怎么去做的实用性建议。但是也许这里有建立一个全面理解的机会。探索大数据意味着要横跨广博的科技进步轨迹，因此与其用无论如何你都将会阅读到的本书总结来困扰你，我还不如尝试在大数据究竟意味着什么的蓝图上做一些涂鸦。

此序言是基于大数据不仅仅是另一项技术的论点展开的。它也不仅仅是另一个经过 Gartner 技术趋势周期传送带而送往全球系统集成商途中的礼品盒。我相信大数据将继电子计算机、互联网之后，成为信息时代的第三个新纪元，在这种情况下它也将从根本上改变企业进化的轨迹。企业将经历一场类似于人类自我认知突破的变化。

因此让我们从头开始。工业时代是一个社会巨大变化的时代。我们先后以蒸汽、电力作为原动力，在生产率上释放了惊人的增长。结果就是人类历史上财富的第一次持续增长。

那些早期的工业企业需要大量并逐步专业化的工人。为了协调所有这些人的工作，就发展出了一系列的规则和职权体系来管理。一个大规模的企业不再是拥有者个人意愿的直接行使，它成为某种有机体。

企业体制诞生自法国的拿破仑政府官僚体系，它强调专业技能、固定规则和森严的等级制度。Bureaucracy（官僚）中的词根“bureau”字面上的含义就是办公桌，而文件就是它们之间的存储介质和信号机制。

官僚体系就是能够处理大范围刺激并且协调众多参与者的组织形式，但是至今它仍严格地限制了它的进化过程。官僚体系就是人类工业企业的线形虫。

拥有超过 24 000 个物种的线形虫是一种数量众多、适应性强的蛔虫，它的神经系统通常由 302 个神经单元构成。除了 20 个神经元组成它的咽神经系统外，剩下的部分充当了它初级的大脑。但它也能够保持体内平衡，引导运动，检测它周围环境信息，做出复杂反应，甚至管理某些基础学习行为。所以，它是对官僚机构的良好近似。

尽管它表现出了复杂行为，但却是线形虫完全无意识做出的。它的反应，就像官僚体系一样，是被动和事务性的。虫子撞上物体，然后受到刺激，神经元被触发，最后虫子做出反应。它的反应也许是离开或者是吃掉它碰上的物体。就像货架一旦售罄，就生成订单，办公桌间传递文件，货车抵达，货架再次被装满。

虫子和企业都是复杂事件处理引擎，但是它们大部分是确定性的。但是企业在进化，它变得更加注意周边环境并且更快速地进行反应。信息时代，或者说第二次工业革命时代，它的主要部分就是这个。

1954 年安达信会计事务所的 Joe Glickauf 在 UNIVAC 1 数字计算机上为通用电气实施了电子薪资系统。他因此把美国企业带进了信息时代的计算机阶段（顺带着，创建了 IT 咨询产业）。贯穿整个 20 世纪 50 年代，其他企业快速地采纳了类似的系统来处理一系列的企业流程。企业依旧是一个线形虫，但是我们正在解构它，并积极地数字化它的神经系统。

它依旧是一条一模一样的虫子。虽然基本上做出同样的反应，但是毫无疑问它变得更加高效，并且能够更快地做出反应。因为我们通过使用计算机模拟表单来让那些已有的系统自动化起来，通过将文件柜的内容转储到数据库中，发货单、账务、客户管理文档都轻松地迁移进入计算机

了。我们正在解构虫子，但是我们还没有重构它。

因此企业依旧是一个官僚体系，仅仅是一个更高效、更快速响应和可扩展的官僚体系。但是这是企业和信息技术共生进化的开始，成为企业进一步进化史的起点。数字化成为即将发生的进化过程的基石。

大概 30 年前，Leonard Kleinrock、Lawrence Roberts、Robert Kahn 和 Vint Cerf 发明了互联网并迎来了信息时代的第二个阶段——网络时代。

突然间我们的小虫子以前所未有的方式和它的同伴以及周边生态联系起来。企业间的信息传递变得像办公桌间信息传递一样自然，意识到这个迫在眉睫的变革后，杰克·韦尔奇和其他人一起推动传统的信件最终让位给数字化的传递。纸质的采购订单、发货单时代最终消亡了。数字化时代的最初 35 年一直关注在内部流程，而现在重心更多地关注到了和外部世界的交互上面。（我需要解释一下，尽管 EDI（电子数据交换协议）已经存在一段时间，但成为互联网的基本结构后它才真正蓬勃发展起来。）虫子就像进化出了第六感，它能够看得更远，对未来做更深入的预测和更快速的反应。

但这些全新的网络不仅仅影响了我们的企业和外部世界交互的方式，它们甚至开始侵蚀官僚体系的基石：层级体系。

在工业时代官僚体系的严格层级制度尽管曾经是劳动者的力量倍增器，但实际上也意味着企业不可能比它最高领导层中最聪明的人聪明。受限于层级体系、僵化制度和专有化职能，企业的总体智慧远逊于它的参与者拥有的总体智慧。

伴随着全球化、复杂的关系网络、更快速的市场循环周期，企业环境的复杂度在快速地增长，很久以来就已经超过了一个独立个体能够理解的复杂度。毕竟世界上仅有一个乔布斯。改变必然发生。

因此企业已经开启了（缓慢的）面向更敏捷、网络化、学习型组织的旅程，它能够众包自身内部不同层级和来自客户群内部的智慧。它们对于学习的反映就是开始出现局部应急行为。这就是 Facebook “快速前进，

打破陈规”的企业箴言背后的东西。换句话说就是自主性的局部化以及决策者不能掌控所有的事情。

当然在网络时代的企业依然拥有组织结构图。但是它们已经不是故事的全部。如今，我们需要分析电子邮件模式、电话记录、即时通信内容和其他人类联系的实际证据来判断真正的组织模型，这种模型就像官方组织架构图内部浮现的晶格。

因此企业的进化不再是仅仅沿着效率和生产率渐进式改进方向前进。依靠技术和复杂现实与周期的驱使，企业的形态正在变化。

企业正在发展出外部感官和必要的神经元来处理它所发现的东西。它也从被动、僵化转变成为复杂和主动，目的就是更好地适应后工业时代的要求。

我们已经站在了信息时代大数据新纪元的门口。企业已经利用计算机和互联网时代有了显著的进化，从而能够适应更加复杂的世界，但是更大的变革还在前方。

本书将带你领略整个大数据的来历，所以在此我不会详细解释大数据的意义。我将仅仅详细描述企业进化到下一个阶段的必要条件。而这个关键点是：大数据不是更大数据的商业智能。

我们不再局限于那个最近统治了企业信息化技术 55 年的结构化的事务世界。大数据代表了存储和分析方式两者的全然转变，而不仅仅只是和大小相关。

你所在企业的“商业智能”今天处理的绝大部分数据是内部生成的高度结构化的交易数据。它就像一个被激发的神经元的记录。商业智能分析师的角色经常被总结为企业的“动觉”（即本体感觉），产出报告来告知层级首脑躯干正在做什么或做了什么。

但是大数据拥有做到不同的潜力。举一个例子，数据经常有不同的来源，并且是它最初的非结构化形式。再举一个例子，我们将不仅仅分析我们已经做了什么；使用含义丰富并且深入细节的原始采集数据，我们将分

析在我们周围的世界将要发生什么。

现在我们可以考虑包括网页日志、视频剪辑、呼叫中心记录、每个网站的页面、社交数据、政府公开数据、合作伙伴数据集等做我们分析的全集。分析将是更加深入细节的外在感知过程，而不再局限于纯粹的自省。我们的消费者正在做什么？他们都认识谁？他们被呼叫的时候是否愤怒或者愉快？他们的网友喜欢什么，他们在什么时候和多大程度上会被网友影响？我们的哪些顾客是最相似的？对于我们的竞争对手他们会说什么？他们从我们的竞争对手那里会买什么？我们竞争对手的停车场是否停满车辆？等等。

也许更重要的是，这些大量的数据如何能够被直接转化成产品，或者至少是我们产品的某个属性？我们是否能完成这样的闭环：先从我们所在环境感知内容，然后理解内容，再到需要做什么？

数据科学这个词语阐述的观念是我们正在使用数据来将科学方法应用到我们的商业中。我们创建（或者发现）假设，进行实验，观察我们的消费者的反应是否符合预期，然后基于这些结果生产新的产品或交互。具有前瞻性的企业正在完成这个闭环以至于整个过程不需要人工干预但产品却基于消费者行为或者其他数据进行实时更新。

换句话说，企业的OODA循环[Observe（观察），Orient（定位），Decide（决策），Act（行动）。美国陆军上校Boyd发明的OODA循环描述了在面对不确定性条件下的决策模型]至少实现在企业的一个战术时间范围内的生产中。人设计算法，除此之外他们的介入是不必要的。与传统重点关注OODA循环中的OO过程的商业智能不同，现代企业需要直接整合决策和行动阶段从而跟上现代市场的动态变化。做更多的分析是不够的，将来的企业需要更多的能够实时行动的产品和组织敏捷能力。

打个比方，我们人类通过感知器官自己绘制的映射来体验世界，然后我们把这些映射保存下来作为记忆。这些映射像脚手架一样，起到对我们心智和自我展现过程的支撑作用。它们是我们由处置到推理的进化入口，与此同时我们从被动反应的虫子进化为具备推理能力的人类。

通过存储复杂而又丰富的相互作用，企业也开始创建并存储类似映射的结构。我们不再将复杂的相互作用精简成卡通般的汇总事务，我们开始存储整个全景，包括每一个感知器和触碰点收集到的每一个原始数据。通过网络和我们现在收集到的关系数据，企业的记忆开始看起来类似人脑的关联模型。企业不可能变成一个人，但是它也绝不仅仅如一只虫子般简单（我知道最高法院肯定不同意我的论述），因为它开始拥有智能。

大数据时代将是重大转变之一。在过去的 55 年，信息技术的关注点在于为了自动化、效率和生产率进行数字化。现在我认为我们将看到信息技术升级为对企业智能的支撑。

直到现在，我们评估一个项目绝大部分靠 ROI。但是我们将很快考虑使用智能化（意思是让某些东西更聪明）来评估。我们在商业和信息技术上的目标将是应用数据与分析来提升企业的智能。就像这个公式： $IQ_{\text{企业}} = f(\text{数据}, \text{算法})$ 。对技术来说这是一个完全不同的目标框架，这将意味着全新的方式来组织和概念化它的投资和交付。

我们收集的数据和我们开发的代码是如何提升我们组织的智能的？我们能否开始考虑某些类似智商的术语在我们的企业上——某种感官知觉、回忆、推理和响应能力的组合？我们将会从关注投资回报率转向关注智能的获取？不管怎样，我们将构建比运行它的人类更聪明、更快速反应的企业。

当然这不是事务性 IT 的终结。企业还是会保留“退化的 IT”，就像我们的大脑依然保留了来自我们倾向性进化过程的区域。毕竟，我们依然会不需思考地把手从热火炉旁边移开，企业也会继续自动地把货架再次补充上。但是一个智能的企业将拥有一个通过理智推导行动的完整大脑。企业将比它的管理层中的任何人更智能，它将不需要人工干预地运行 OODA 循环。

大数据是信息时代的新纪元，我们为之工作的企业将更加智能。

Jim Stogdill
O'Reilly 媒体 总经理

致谢

我们要特别感谢在整个写作和出版过程中一直为我们提供帮助的扩展团队：斯托克斯·亚当斯（Stokes Adams）、迈克·巴洛（Mike Barlow）、西克·曹（Sheck Cho）、斯泰西·里维拉（Stacey Rivera）、宝拉·桑顿（Paula Thorton）。

我们还要鸣谢对这本书做出了有益贡献的人们和他们所属的组织。

Chuck Alvarez	摩根士丹利
Tasso Argyros	Teradata
Amr Awadallah	Cloudera
Ravi Bandaru	诺基亚
Mike Barlow	Cumulus Partners
Randall Beard	尼尔森
David Botkin	Playdom
Nate Burns	纽约州立大学布法罗分校
David Champagne	革命分析
Drew Conway	IA Ventures
Joe Cunningham	Visa
Yves de Montcheiul	Talend
Anthony Deighton	QlikTech
Deepinder Dhingra	Mu Sigma
Zubin Dowlaty	Mu Sigma
Shaun Doyle	Cognitive Box
Michael Driscoll	Dataspota