

BIG 大数据 DATA

从概念到运营

【美】比尔·施玛泽 (Bill Schmarzo) 著 钱峰 译

UNDERSTANDING
HOW DATA POWERS
BIG BUSINESS



中信出版集团 · CHINACITICPRESS

BIG DATA

大数据 从概念到运营

UNDERSTANDING HOW DATA POWERS BIG BUSINESS

【美】比尔·施玛泽 (Bill Schmarzo) 著 钱峰 译

图书在版编目 (CIP) 数据

大数据: 从概念到运营 / (美) 施玛泽著; 钱峰译

北京: 中信出版社, 2016.8

书名原文: Big Data: Understanding How Data

Powers Big Business

ISBN 978-7-5086-6487-3

I. ①大… II. ①施… ②钱… III. ①企业管理-研究 IV. ①F270

中国版本图书馆 CIP 数据核字 (2016) 第 167496 号

Big Data: Understanding How Data Powers Big Business by Bill Schmarzo, ISBN:9781118858202

Published by John Wiley & Sons, Inc.

Copyright © 2013 by John Wiley & Sons, Inc., Indianapolis, Indiana. All Rights Reserved. This translation published under license. Authorized translation from the English language edition, Published by John Wiley & Sons. No part of this book may be reproduced in any form without the written permission of the original copyrights holder.

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

Simplified Chinese translation Copyright ©2016 by CITIC Press Corporation

大数据: 从概念到运营

著者: [美] 比尔·施玛泽

译者: 钱峰

策划推广: 中信出版社 (China CITIC Press)

出版发行: 中信出版集团股份有限公司

(北京市朝阳区惠新东街甲4号富盛大厦2座 邮编 100029)

(CITIC Publishing Group)

承印者: 北京楠萍印刷有限公司

开本: 880mm × 1230mm 1/32

印张: 10.5 字数: 149千字

版次: 2016年8月第1版

印次: 2016年8月第1次印刷

京权图字: 01-2013-8339

广告经营许可证: 京朝工商广字第 8087 号

书号: ISBN 978-7-5086-6487-3

定价: 49.00元

版权所有·侵权必究

凡购本社图书, 如有缺页、倒页、脱页, 由销售部门负责退换。

服务热线: 400-600-8099

投稿邮箱: author@citicpub.com

BIG DATA | 序言

Understanding How Data Powers Big Business

换一种方式来思考。

你的竞争者已经开始利用大数据来获益，而你那传统的信息技术设备却不足以应付大数据的管理、分析和处理。

换一种方式来思考。

你应该重视大数据。大数据对组织产生的最重要的影响就在于它能升级现有的业务流程并且发现新的赢利机遇。一个组织对于影响业务的关键因素（如客户、产品、活动以及运营）的分析再多也不为过。大数据能够以较低的粒度，用更加及时可行的方法进

行分析。大数据能支持新的业务应用，比如个性化营销、基于位置信息的服务、预见性维护的属性分析和机器行为分析。大数据有望重塑组织的价值创造过程并创建全新的、更具吸引力的、更赚钱的客户互动模式。大数据能够使你的企业从回顾型、批量处理的业务监管中解放出来，转而获得更多具有预见性的、实时的业务优化方面的深刻见解。

换一种方式来思考。

大数据使你能够接受数据丰富性，并且在实时、低粒度的情况下掌握分析组织内部和外部数据的能力。例如，商业智能交叉分析模型能很好地处理 GB 级的数据，但是在如今这个有着 PB 级数据、数以千计的横向扩展的处理节点以及数据库内分析的时代，它已经像马鞭和马车一样落伍了。^① 标准的关系数据库技术也无法表达大数据分析所基于的复杂分支和迭代逻辑。你需要一个升级的现代基础设施以充分利用大数据。

换一种方式来思考。

虽然关于大数据的讨论大都集中于 Hadoop^② 和其他大数据技术创新，但是真正驱动技术和业务的是大数据经济学，它结合了开源数据管理和高级分析。使用建立在横向扩展的商品架构上的软件的费用是使用数据仓库架构的费用的 1/20。这一巨大经济变

① 1GB=1 024M, 1TB=1 024G, 1PB=1 024T。——编者注

② Hadoop 是一个开发和运行处理大规模数据的软件平台。——编者注

化使你不得不重新考虑许多传统的数据和分析模型。三年前不可能实现的数据转换和数据富集现在已经方便且廉价地实现了，无论企业是大是小，都能够运用云技术在海量指标中挖掘 PB 级数据。

换一种方式来思考。

企业在面对大数据时最大的隐患是什么？是无所作为。是被动地等待技术供应商为你解决问题，等待技术转换的沙砾沉淀下来。你将 Hadoop 引进企业，加载了一些数据，并让一些人来研究它。但这不是做科学实验的时候，这是一项严肃的技术，它依靠海量的实时数据和高级分析创造了新的业务模式，它的价值已经在许多行业得到了证实——零售业、金融服务业、电信业、制造业、能源业、运输业以及酒店业。

换一种方式来思考。

人们需要做什么？作为商业和信息技术行业的领导者，你们应该消除隔阂，互相了解，确认组织中最重要业务流程，然后去审视大数据，特别是具体的交易数据（暗数据）、非结构化数据、实时数据访问和预测性分析，思考大数据将如何发掘关于你的客户、产品、活动和运营的可操作性见解。利用大数据更快更频繁地做出更好的决策，并在过程中发现新的赢利机会，借助大数据开启“让我赚更多钱”的模式。行动起来，大胆一点！不要怕犯错误，如果你失败了，立刻从失败中站起来继续前进，吸取教训。

换一种方式来思考。

BIG DATA 引言

Understanding How Data Powers Big Business

大数据是如今技术领域的热点话题。这样的话题每过四五年就会出现一批，并且会成为企业必须掌握的技术，带领企业抵达一片“乐土”，在那里所有技术缺陷和困难都能得到解决。企业努力从供应商和分析人员提供的信息中分辨什么是技术能做的，什么是不能做的。在一些案例中，它们成功地将技术融入了企业的技术版图，这些技术包括关系数据库、企业资源计划（ERP）、客户端—服务器架构、客户关系管理（CRM）、数据仓库技术、电子商务、商业智能（BI）和

开源软件。

大数据则不同。可能因为大数据在本质上更多地关注企业转型问题，而不是技术问题，大数据将企业从一个回顾型、批量处理、有数据约束的监督型商业环境，转为预测型、实时、渴求数据、追求最优化的商业环境。使用大数据并不是为了均分业务或追赶潮流。大数据利用收集到的关于你的客户、产品和运营的独特的、可操作的见解，重塑你的价值创造过程、优化你的关键业务计划并发掘新的赢利机会。大数据是关于赚钱的，这也是这本书要阐述的问题——如何利用那些关于你的客户、产品和运营的独特的、可操作的见解为企业赚钱？

本书从一个实用主义的、可亲身实践的角度来阐述大数据带来的商业机遇。书中没有太多的理论，而是有很多切合实际的建议、技巧、方法和可以参考的工作表，以及这些年来我通过与一些世界领导企业合作积攒的许多案例。当你翻阅这本书时，你会学习到以下内容：

- 帮助你的企业学习大数据的一般定义，利用大数据业务模型成熟度索引和你的企业进行交流，探讨大数据可以在哪些具体业务领域中实现有意义的业务价值（第1章）。
- 回顾从以往大数据事件中得到的历史教训，决定哪

些部分你可以运用在目前或者将来的大数据的机遇中（第2章）。

- 学习一套可以利用你目前的业务流程来识别正确地聚焦大数据项目的度量指标，从而推动企业走向成功的新流程（第3章）。

- 创建一个高效的组织结构来支持你的大数据计划，包括一些新角色的融合，将诸如数据科学与用户体验小组、新的首席数据官、首席分析官这样的角色融入你的数据和分析组织中（第4章）。

- 回顾一些常见的人类决策困境和不足，思考“质疑的终结”带来的影响，知道如何将能抵抗这些人类决策弱点的深刻见解付诸实践（第5章）。

- 学习一种打破式方法论，或者从功能性角度被称为“分解”的方法，将组织的商业策略和关键业务计划分解为关键业务价值驱动因子、关键成功因素，以及支持性的数据、分析和技术要求（第6章）。

- 通过将大数据商业价值驱动因子（失衡交易数据、新的非结构化数据源、实时数据连接和预测性分析）与诸如迈克尔·波特的五力分析和价值链分析等价值创造模型进行参照对比，从而深入探究大数据的工商管理硕士课程内容，展望大数据将在哪些领域，通过什么方式优化你的

核心业务流程并开发新的赢利机会（第7章）。

- 了解为什么从客户行为和产品使用数据的新来源获得的有关用户和产品的深入见解，再加上高水平分析，能够支持更引人注目、更具相关性且赢利更多的用户体验（第8章）。

- 促使业务和信息技术领域的利益相关者建立合作，畅想他们能够借助大数据实现什么，发掘关于大数据如何影响核心业务流程的案例，确保对大数据的最终期望的状态和关键成功因素达成一致（第9章）。

- 围绕识别、构建和传达大数据支撑的商业解决方案和应用过程，学习融合了所有技术、方法、工具以及工作表的处理方法（第10章）。

- 回顾那些使新数据管理和先进分析方法得以实现的关键大数据技术（Hadoop、MapReduce、Hive等）和分析方法（R、Mahout、MADlib等）的发展，探索这些技术对你现有的数据仓库和商业智能环境能够产生的影响（第11章）。

- 用大数据故事图简要介绍大数据的最佳实践、方法和价值创造的技术，这张简单的图概括了实现用大数据优化价值创造过程并发掘新的赢利机会这一承诺的要点和方法（第12章）。

- 总结本书，回顾一系列将指导你和你的企业踏上大数据之旅的行动号召——从教育和认识，到识别大数据之旅开始的地方和启动方式，以及大数据支撑的商业解决方案和应用的全过程（第13章）。

身处数据和分析行业的魅力在于，只需要一项技术创新，我们就能迎来下一个大数据浪潮。零售终端、通话账单以及信用卡数据为消费品行业、零售业、金融服务业以及电信公司提供了利用大数据的先机，网站点击产生的数据对电子商务和数字媒体产业提供了帮助。社交媒体、手机应用和传感器产生的数据加剧了当今各行业的大数据狂潮，不仅有商业机构对消费者的电子商务（B2C），也有商家对商家的电子商务（B2B），而且有越来越多的行业加入进来。可穿戴设备、脸部识别、DNA（脱氧核糖核酸）图谱和虚拟现实产生的数据将释放新一轮由大数据驱动的价值创造机遇。

在这场数据巨变中存活下来并繁荣发展的，是那些将数据和分析作为核心能力的组织。这些组织对数据如饥似渴，将数据看作一项需要积累的资产，而不是需要避免的业务支出。这样的企业将分析方法当作知识产权来管理，认为它需要获取、培养，有时甚至需要进行法律保护。

本书就是献给这样的企业。本书提供了包括技术、工具以

及方法在内的一套指南，帮助组织满足对数据如饥似渴的需求，培养复杂的数据管理和分析能力，做出必要的组织的调整和投资，从而利用关于你的客户、产品及运营活动的洞见来优化关键业务流程和发掘新的赢利机遇。

BIG DATA 目录

Understanding How Data Powers Big Business

序 言 // VII

引 言 // XI

01

大数据的业务机遇

业务转型势在必行 // 006

沃尔玛的案例分析 // 007

大数据业务模型成熟度索引 // 010

大数据业务模型成熟度观察报告 // 027

02

大数据的历史教训

被通用产品代码颠覆的快速消费品行业和零售业 // 033

大数据运动中的经验教训 // 039

03 大数据的业务影响

大数据的颠覆性影响 // 047

用正确的标准进行管理 // 048

数据货币化机遇 // 053

04 大数据的组织影响

数据分析周期 // 066

数据科学家的角色和职责 // 070

大数据小组中的新角色 // 076

解放组织的创造力 // 081

05 理解决策理论

商业智能的挑战 // 090

疑问的终结 // 092

大数据对用户界面的影响 // 093

人们在决策时遭遇的挑战 // 096

06 建立大数据战略

大数据战略文件 // 110

将战略文件变为行动 // 115

星巴克的大数据战略文件案例 // 117

旧金山巨人队的大数据战略文件案例 // 121

07 理解你的价值创造过程

大数据价值创造的四大驱动因素 // 134

迈克尔·波特的价值创造模型 // 149

08 大数据对用户体验的影响

非智能的用户体验 // 170

了解确定用户体验的关键决定 // 172

利用大数据分析改善与客户的互动 // 174

挖掘并利用对客户的认知 // 177

大数据带来的全新用户体验 // 183

09 大数据愿景训练

大数据愿景训练流程 // 199

排序过程 // 219

利用用户体验模型推动愿景训练 // 226

10 大数据方案工程

方案工程流程 // 236

方案工程的业务方案 // 249

阅读年度报告 // 258

11 大数据对架构的影响

- 建立新的数据架构 // 270
- 新的大数据技术 // 272
- 将大数据带入传统的数据仓库世界 // 280

12 开始你的大数据之旅

- 爆炸性的数据增长推动业务机遇 // 299
- 传统的技术和方法难以满足需求 // 300
- 大数据业务模型成熟度索引 // 302
- 推动业务和信息技术方面的利益相关者合作 // 303
- 落实大数据的洞见 // 305
- 大数据推动价值创造过程 // 306

13 大数据行动清单

- 确认组织的关键业务计划 // 311
- 从业务和信息技术方面利益相关者的合作开始 // 312
- 规范展望流程 // 313
- 利用模型推动创造性过程 // 314
- 明确技术和架构选择 // 314
- 建立在现有的内部业务流程的基础上 // 315
- 挖掘新的货币化机遇 // 316
- 了解对组织的影响 // 317

Understanding How Data
Powers Big Business

01

大数据的业务机遇