

大数据应用与技术丛书

Big Data MBA: Driving Business Strategies with Data Science

大数据MBA

通过大数据实现与分析驱动
企业决策与转型

Economics

[美] Bill Schmarzo

Metamorphosis

著

于楠

译

Think
Differently

Business Drivers

- ① All time data
- ② Unstructured data
- ③ Real-time analysis
- ④ Predictive analytics

Key Business
Processes

#4

Monetization

#3

Optimization

#2

Insights

#1 Monitoring

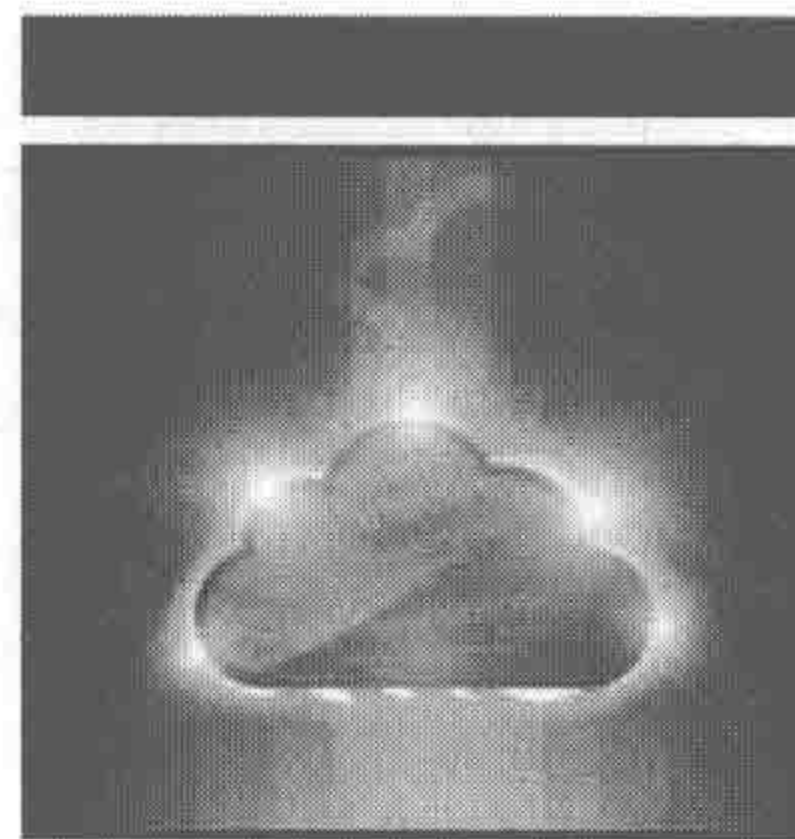
Prescriptive analytics
(Recommendations)

WILEY



清华大学出版社

大数据应用与技术丛书



大数据 MBA

通过大数据实现与 分析驱动企业决策与转型

[美] Bill Schmarzo 著

于楠 译

清华大学出版社

北 京

Bill Schmarzo

Big Data MBA: Driving Business Strategies with Data Science

EISBN: 978-1-119-18111-8

Copyright © 2016 by Bill Schmarzo

All Rights Reserved. This translation published under license.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc., and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

北京市版权局著作权合同登记号 图字: 01-2016-5740

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

本书封面贴有 Wiley 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

大数据 MBA: 通过大数据实现与分析驱动企业决策与转型/(美)比尔·施玛泽(Bill Schmarzo) 著; 于楠 译.
—北京: 清华大学出版社, 2017

(大数据应用与技术丛书)

书名原文: Big Data MBA: Driving Business Strategies with Data Science

ISBN 978-7-302-47736-5

I. ①大... II. ①比... ②于... III. ①数据处理—应用—企业管理 IV. ①TP274②F272

中国版本图书馆 CIP 数据核字(2017)第 167826 号

责任编辑: 王 军 于 平

装帧设计: 孔祥峰

责任校对: 牛艳敏

责任印制: 沈 露

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: 北京嘉实印刷有限公司

经 销: 全国新华书店

开 本: 185mm×230mm 印 张: 15.25 字 数: 315 千字

版 次: 2017 年 8 月第 1 版 印 次: 2017 年 8 月第 1 次印刷

印 数: 1~4000

定 价: 49.80 元

产品编号: 069436-01

译者序

有人做过这样一个统计，一分钟内，微博、推特上新发的数据量超过 10 万；社交网络“脸书”的浏览量超过 600 万……这些数据意味着什么？对普通人来说，这些数据无非代表闲来无事的一条评论、一张图片或一些感慨，而对数据挖掘高手来说，这些数据背后可能蕴藏着巨大财富。比如华尔街一家投资公司的老板曾通过分析社交媒体账户留言判断民众情绪，抛售股票大赚。数据的作用不止于此，购物网站用户留言可以显示企业某一产品的销售状况；求职网站的岗位空缺可以推测当前就业率，甚至美国疾控中心根据网民搜索，分析出全球范围内流感等病疫的传播状况等。

有人说大数据是第四生产要素，也有人预言大数据将来的价值会超过石油，成为未来新的“能源”。那么，大数据究竟是什么？我们只能说，数据的价值日益不可估量。越来越多的政府、企业、机构开始意识到数据正在成为最重要的资产，数据分析能力正在成为组织的核心竞争力，但是，又有多少企业能够真正利用好数据和分析，成功获得有效运营和业绩增长呢？

本书作者 Bill • Schmarzo 泽是大数据行业领军企业 Dell EMC 的首席技术官。在本书中，他毫无保留地分享了在这一领域 20 多年的专业经验，通过引入“大数据商业模式成熟度指数”概念，分析具体企业的大数据运营案例，指导企业像“数据科学家一样思考”，化繁为简，教会企业如何利用大数据和数据科学构建合理的关键商业计划，获得竞争优势，挖掘隐藏在大数据中的巨大财富。

作为译者，本人非常愿意推荐《大数据 MBA 通过大数据实现与分析驱动企业决策与转型》一书。本书不需要太多的大数据技术背景，无论你是企业用户、管理人员、有志从事这一行业的学生或其他人士，本书都称得上是一本使用数据和分析解决商业问题的绝好

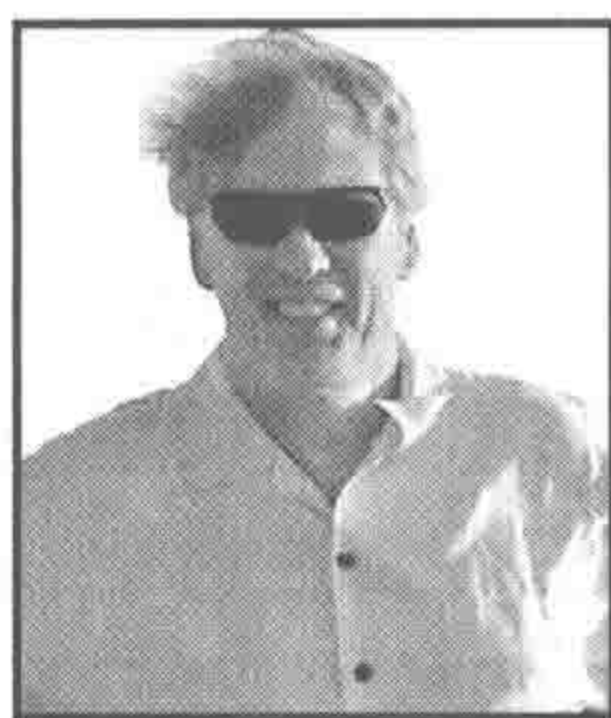
实用指南。本书内容通俗易懂，语言清晰简单。对于普通读者，本书也是一本了解大数据策略和数据科学的好书。

在撰写译者序期间，正值我在美国密苏里新闻学院访学，亲身感受到数据技术对新闻业的影响同样深刻。通过大数据，媒体可以对生活的各个方面进行预测性报道，宏观如经济指数预测，微观如交通拥堵、流行疾病等预测。根据大数据分析结果，为用户提供定制化的信息推送服务，也是媒体未来发展的趋势之一。这一点和利用数据分析、增强企业商业策略竞争力也有着异曲同工之妙。这些都体现了大数据的影响并不仅限于信息通信产业，而是对各个领域都开始了量化进程。

这已经是我和清华大学出版社王军老师合作的第二本大数据系列书籍了，在此，我非常感谢王军老师在整个编辑过程中对我的理解和包容，以及其他各位老师、朋友为本书付出的心血。你们的专业与严谨令我非常佩服，特此表示衷心的感谢。同样感谢我的家人对我翻译工作的支持和鼓励。没有你们，本书就不可能顺利出版。由于本书内容包罗万象，跨度很大，涉及企业管理、经济、历史等方方面面，为本书的翻译增加了不少难度。尽管一直努力保证翻译的准确性，但学识有限，译文难免会有疏忽之处，恳请广大读者批评指正。对于本书的任何想法和意见都欢迎发送邮件至 celene1122@gmail.com。

译者

作者简介



作者 Bill Schmarzo 是 Dell EMC 公司首席技术官(CTO)。作为首席技术官，他负责制定 Dell EMC 全球服务的大数据策略、指导大数据服务产品和服务功能。他还直接与企业合作，帮助他们明确在何处以及如何开始他们的大数据之旅。Bill 是 *Big Data: Understanding How Data Powers Big Business* 一书的作者，他还撰写白皮书，是个狂热的博主，经常发表演讲，介绍如何利用大数据和数据科学影响企业的关键商业计划。他是旧金山大学管理学院的一名研究员，教授“大数据 MBA”课程。

Bill 在数据仓库、商业智能和分析方面具有三十年以上的经验。他撰写了“EMC 大数据愿景研讨会”方案，并与 Ralph Kimball 共同编写了一系列分析应用程序类文章。Bill 曾在数据仓库研究所工作，担任分析应用课程的负责人。此前，他是雅虎(Yahoo!)分析副总裁，负责 Business Objects 的分析应用程序商业部门，职责包括其行业定义的分析应用程序的开发、市场营销和销售。

Bill 拥有柯伊学院数学、计算机科学和工商管理理学学士学位，以及爱荷华大学工商管理硕士学位。他最近的博文链接是 http://infocus.emc.com/author/william_schmarzo/。



技术编辑简介

Jeffrey Abbott 领导 Dell EMC 公司围绕大数据开展全球服务营销实践，帮助客户了解如何通过战略商业计划识别和抓住机会，同时提高对提升客户价值的服务产品组合的认识。作为内容开发人员和项目负责人，Jeffrey 认为基于角色的广告系列要传达清晰、简洁的讯息。在加入 Dell EMC 之前，他帮助构建和推广了 CA Technologies 公司的云计算生态系统。该系统结合了在线社区、云开发平台和云服务电子商务网站。他还在该公司的“思想领导小组”工作了几年，围绕信息技术领域中的主要革命性趋势，制定和推广了可执行的信息发送和社交媒体项目。他在 Dell EMC、Citrix 和 Ardenne 等公司担任过各种产品营销职位，并在十年时间中，在众多精品营销公司运行客户端账户。他曾在佛蒙特大学学习小企业管理，现在和妻子、两个儿子共同居住在马萨诸塞州的萨德伯里地区。Jeffrey 喜欢滑雪、徒步旅行、摄影和驾驶老爷车。

致 谢

致谢是件危险的事，虽然不像与鳄鱼或者不开心的芝加哥小熊队球迷摔跤一样危险，但在此我需要向很多人表示感谢。如何防止致谢部分比这本书还长呢？本书中总结了许多内容，包括讨论、辩论、演示、参与以及在深夜里与许多同事和客户喝啤酒、吃比萨。感谢所有与我一起经历大数据之旅的人。

我可能会在致谢部分忘记感谢许多人。下面，致谢开始……

Jeff Abbott 对这本书的贡献，真是说也说不完。他不仅是本书的技术编辑，而且还承担了一项吃力不讨好的任务——编辑我所有的博客。Jeff 很耐心地包容了我的写作风格，巧妙地将这些材料编辑为可读性强且易于理解的文档。再怎么感谢他的耐心、他的指导和我们之间的友情都不足为过。

虽然 Jen Sorenson 只是这本书的公关编辑，但她所做的远超过其职责范围。由于采纳了珍的建议，本书的许多章节(如第 6 章中使用童话主题公园的例子)变得更加有趣。事实上，第 6 章可能是我最喜欢的章节，因我很难驾驭数据科学算法。而珍完成了一项颇有难度的课题(数据科学算法)，并将它赋予生命，这真是太不可思议了。

谈到数据科学，Pedro DeSouza 与 Wei Lin 是我见过的两位最优秀的数据科学家。我更感激能和他们成为朋友。在过去几年中，他们一直帮助我学习有关数据科学的内容，这反映在本书的许多章节中(最明显的如第 5 章和第 6 章)。但更重要的是，他们教会我非常有价值的人生一课，即谦卑是最好的学习方式。我无法用言语表达我对他们本人及其专业度的钦佩之情。

Joe Dossantos 和 Josh Siegel 两位可能会惊讶地发现他们的名字出现在我的致谢名单中，但事实上他们不应该这样。他们与我共同度过了大数据之旅中的许多阶段，帮助我进

一步了解了大数据如何影响商业世界。他们的影响体现在整本书中。

Wiley 出版社的两位编辑，Adaobi Obi Tulton 和 Chris Haviland，绝对不简单！他们尽心尽力，使整个编辑过程尽可能地顺畅。他们非常理解我的想法，所以我接受了他们提出的 99% 以上的建议。他俩也是我第一本书的编辑，我猜想在他们同意做我第二本书的编辑时，已经记不起我之前给他们带来了多少麻烦。虽然我们从来没有见过面，但我能够感觉与他们非常亲近。感谢你们的耐心和指导以及幽默感！

我特别要感谢 Mouwafac Sidaoui 教授，我与他在旧金山大学的管理学院共同教授大数据工商管理硕士。除了他之外，我找不出更好的合作者。他机智、谦逊、严厉、幽默、老练，具备一个朋友所能拥有的一切品质，让任何人都想与其成为朋友。在他的帮助下，我成为旧金山大学的管理学院研究员，他还为我安排了下一个职业——教书。

我也要感谢 Elizabeth Davis 院长，以及旧金山大学工商管理学院的硕士生，他们愿意作为测试本书概念和技术的小白鼠。他们帮助我确定哪些想法能够实施，如何解决那些不奏效的方法。

另外，特别感谢 Dell EMC 公司，它使我在大数据商业转型潜力前沿进行研究工作。它让我自由地追求新想法、新概念和新产品。许多情况下，助我走在大数据发展的前沿。我找不到像 Dell EMC 一样更好的雇主和合作伙伴了。

感谢名单中还应该包括 Dell EMC 里优秀、富有创造力的员工们，我与他们定期互动。但是这个列表太长了，在此我仅主要感谢 Ed、Jeff、Jason、Paul、Dan、Josh、Matt、Joe、Scott、Brandon、Aidan、Neville、Bart、Billy、Mike、Clark、Jeeva、Sean、Shriya、Srini、Ken、Mitch、Cindy、Charles、Chuck、Peter、Aaron、Bethany、Susan、Barb、Jen、Rick、Steve、David 等许多人。

我要感谢我的家人，在我创作期间他们一直支持包容我。在我面临截稿压力时，我的妻子 Carolyn 还为我制作了美味的墨西哥辣肉卷。我的儿子 Alec、Max、女儿 Amelia 在整本书的写作过程中都对我给予了很大的支持。这个了不起的家庭一直在保佑我(但是别再偷拿我放在冰箱里的墨西哥辣肉卷了！)

最重要的是，我想要感谢我有幸合作过的 Dell EMC 的客户。客户一直在大数据转型的一线阵地，和客户合作能够更好地了解哪些大数据策略起作用而哪些并不奏效，继而在一阵阵地与 Dell EMC 最优秀的客户携手并进。诚然，我工作中最好的一部分就是有机会与我们的客户合作。嗨！为了和他们合作，我都愿意忍受飞来飞去的航空旅行了！

前言

我从来没有计划写第二本书，因我本以为写一本书足以涵盖与这个话题有关的所有内容。但自从我完成第一本书之后，这个领域就发生了太大的变化，我深深地感到有必要继续帮助企业机构探索这个领域，帮助他们利用数据和分析方法改变其商业模式。我也并不仅仅探讨商业过程中“让我赚更多的钱”这一部分。大数据能够在教育、扶贫、假释复原、卫生保健、安全和降低犯罪等诸多方面显著地“提高生活质量”。

我的第一本书面向信息技术领域的读者。但是，我很快意识到，抢占大数据领域最大的赢家当属商业。因此，本书主要面向商业领域的读者，并基于以下几个关键前提：

- 组织机构不需要大数据策略。相比之下，他们需要整合了大数据的商业策略。
- 商业领袖将分析方法转化为信息技术的时代已经结束，未来的企业领导者必须将分析方法视为与会计、财务、管理科学和营销一样的商业规则。
- 数据货币化和商业转型的关键在于释放机构的创造性思维，必须让企业用户能够“像数据科学家一样思考”。
- 最后，大数据的商业潜力只受到企业用户创造性思维的限制。

在我完成第一本书之后，我开始在旧金山大学管理学院讲授“大数据 MBA”课程。由于表现出色，我成为该校的第一名管理学院研究员。与学校里那些优秀、富有创造力的学生们，以及和 Mouwafac Sidaoui 教授共事的经历，促使我着手编写第二本书。这本书主要面向这些学生以及未来的商业领袖。

我希望本书中最为引人入胜的话题是数据科学的力量。目前已经有许多关于数据科学的书籍，它们大都旨在帮助人们成为数据科学家。但是我却认为这些书籍中缺失了一些内容：我们需要帮助未来商业领袖像数据科学家一样思考，而不是去创造一个充满数据科学

家的世界。

因此，本书的重点即，帮助未来的商业领袖将数据和分析方法整合到他们的商业模式中，并通过帮助企业“像数据科学家一样思考”，释放机构的创意，引领文化的转型。

本书概览

商业利益相关者放弃对信息技术数据和分析方法控制的日子已经结束了。商业利益相关者在对机构进行数据收集和分析工作提供支持和货币化方面必须处于前沿及核心的位置。商业领导者需要了解在何处以及如何使用大数据，利用客户、产品和运营数据的新来源之间的冲突，再加上数据科学，优化关键业务流程，发现新的货币化机会，创造竞争优势新来源。然而，将商业用户改造成数据科学家是不现实的。最重要的是，我们应当教会商业用户像数据科学家一样思考，这样他们便可以利用信息技术，与数据科学家合作，进行用例识别、需求定义、业务估值以及最终的分析操作。

本书提供了一个商业化框架，辅以相应的支持方法和实践练习。这些方法和练习不仅可以帮助商业用户明确在何处以及如何利用大数据获得商业优势，而且还为操作分析方法、建立正确的组织结构，以及将机构用户体验的分析观察结果推向客户和一线员工等多方面提供了相应的指导。

本书的组织架构

本书分为四个部分：

- 第I部分“大数据的商业潜力”包括第1章至第4章，为本书奠定了以商业论述为中心的基调。这里，我介绍了大数据商业模式成熟度指数，并围绕“机构并不需要大数据策略，而需要大数据的应用策略”这一角度，对大数据进行了讨论。
- 第II部分“数据科学”包括第5章至第7章，介绍了数据科学背后的原理。这些章节介绍了一些数据科学的基础知识，并探讨了商业智能和数据科学的互补性，以及这两个学科在解决问题上的异同。
- 第III部分“商业利益相关者的数据科学”包括第8章至第12章，旨在教会商业用户和商业领袖“像数据科学家一样思考”。这部分介绍的方法和练习用于强化数据科学思维。其中包含很多实际的练习。
- 第IV部分“构建跨企业间的支持”包括第13章至第15章，讨论了机构面临的挑战。这部分主要探讨“构想”，这一话题可能是本书中最重要的话题，因为大数据的商业潜力只受到商业用户的创造性思维的限制。

以下是本书中每章节的具体内容：

- 第1章：“大数据商业任务”。该章围绕大数据展开讨论，重点介绍了大数据如何应用于商业转型，以及大数据经济学，而非大数据技术本身。
- 第2章：“大数据商业模式成熟度指数”。该章介绍了整本书的基础——大数据商业模式成熟度指数(Big Data Business Model Maturity Index, BDBM)，阐述 BDBM 的五个阶段，BDBM 提供衡量机构将数据和分析方法整合到商业模型中有效性的方法。
- 第3章：“大数据策略文档”。该章介绍了 CXO 级别的文档和过程，帮助机构从商业角度确定从何处以及如何开始大数据之旅。
- 第4章：“用户体验的重要性”。这是我最喜欢的话题之一。该章通过更为简单、直接的方法，挑战传统的商业智能报告和控制面板，为关键商业利益相关者，包括一线员工、渠道伙伴以及终端客户提供可行的见解。
- 第5章：“商业智能和数据科学之间的差异”。该章探讨了商业智能和数据科学的不同之处，突出了各自的差异和互补性。
- 第6章：“数据科学 101”。该章(我最喜欢的一章)回顾了我的数据科学团队通常使用的 14 种不同的分析技术，以及应该考虑在何种商业情况下使用这些技术。该章还使用了童话主题公园的奇妙案例进行研究。
- 第7章：“数据湖泊”。该章介绍了数据湖泊的概念，解释了数据湖泊如何摒弃昂贵的数据仓库资源，释放数据科学团队的创造性。
- 第8章：“像数据科学家一样思考”。这是本书的核心部分。该章涵盖“像数据科学家一样思考”的六个步骤。该章内容相当深刻，因此在阅读时需要动笔(也可能用到橡皮)在纸上记录、练习。
- 第9章：““By”分析技术”。该章深入探讨“像数据科学家一样思考”这个重要的概念，即“By”分析技术。
- 第10章：“评分开发技术”。该章介绍了评分如何推动企业用户和数据科学家之间的协作，创建用于指导机构关键商业决策的可行评分。
- 第11章：“货币化训练”。该章为机构提供了一种方法。这些机构虽然有大量的客户、产品和运营数据，但却不知道如何通过该数据获利。该章能够开拓眼界！
- 第12章：“商业蜕变训练”。该章提供了有趣的即学即用的练习。由于考虑了大数据商业模式成熟度指数的蜕变阶段，它探讨了潜在的数据和分析方法对机构的影响。
- 第13章：“构想的力量”。该章开始解决你可能遇到的一些机构和文化上的挑战。特别是该章中介绍了一些构想技术，可以帮助释放机构的创造性思维。

- 第 14 章：“释放企业机构影响力”。该章更为详细地介绍了大数据对机构带来的影响，尤其是首席数据(货币化)执行官的作用。
- 第 15 章：“故事”。本书包含了一些非传统意义上的案例研究。利用该章介绍的方法，可以创建与自身机构相关的大数据“故事”。任何人都可以找到大数据案例研究，但这并不意味着每个人都可以创建一个大数据故事。

本书读者对象

本书面向企业用户和企业管理人员。我撰写本书旨在将其作为我的大数据 MBA 课程教材，因此其中包含了学生的实践练习与模板。通过这些练习，学生能够顺利获得大数据 MBA 毕业证书。

我认为读者阅读我的第一本书 *Big Data: Understanding How Data Powers Big Business* 也会有所收获。这本书的受众对象是信息技术领域的读者。虽然这两本书之间有部分内容重叠(10%~15%)，但第一本书中做了铺垫，并介绍了一些大数据概念，而本书则对这些概念进行了更详细的探讨。

所需工具

除了铅笔、橡皮擦、几张纸和你的创意之外，并不需要特殊的工具。拿一杯印度茶拿铁，一些 Chipotle 快餐，享受学习的乐趣吧！

网站提供的资源

你可以从本书的网站(www.wiley.com/go/bigdatamba)下载“像数据科学家一样思考”的工作簿，也可以扫描封底的二维码下载。你会发现意外的惊喜！

这对你意味着什么

旧金山大学“大数据 MBA”课程的学生告诉我，这本教材让他们有机会周密地策划，推动跨机构间合作，解决问题，接受挑战，再利用数据和分析方法将提出的想法付诸实践。什么样的雇主会不想要一个知道如何实现这些过程的未来领导者呢？

目 录

第1部分 大数据的商业潜力

第1章 大数据商业任务	3
1.1 大数据 MBA 介绍	3
1.2 关注大数据的驱动竞争差异	5
1.2.1 利用技术推动竞争差异	6
1.2.2 论经济驱动的商业转型 经验	6
1.3 “不同思考方式”的重要性	8
1.3.1 别想着大数据技术, 想想 商业转型	8
1.3.2 别想着商业智能, 想想 数据科学	9
1.3.3 别想着数据仓库, 想想 数据湖泊	9
1.3.4 别想着“发生了什么”, 想想“会发生什么”	10
1.3.5 别想着最高收入人群, 想想合作	11
1.4 本章小结	12

1.5 家庭作业	12
第2章 大数据商业模式成熟度指数	15
2.1 介绍大数据商业模式 成熟度指数	16
2.1.1 阶段1: 商业监测	18
2.1.2 阶段2: 商业观察	18
2.1.3 阶段3: 商业优化	21
2.1.4 阶段4: 数据货币化	22
2.1.5 阶段5: 商业蜕变	24
2.2 大数据商业模式成熟度 指数经验心得	25
2.2.1 经验1: 专注原始 大数据价值	25
2.2.2 经验2: 充分利用见解, 创建新的盈利机会	26
2.2.3 经验3: 为企业机构 变革做准备	26
2.3 本章小结	27
2.4 家庭作业	28

第 3 章 大数据策略文档	29
3.1 建立通用商业术语	30
3.2 介绍大数据策略文档	31
3.2.1 确定机构的关键商业计划	32
3.2.2 Chipotle 餐馆中最重要的 是什么	33
3.2.3 确定关键商业实体和 关键决策	34
3.2.4 明确经济助力因素(用例)	37
3.2.5 识别和优化数据源	39
3.3 介绍优先级矩阵	42
3.4 使用大数据策略文档, 赢得 世界职业棒球大赛	43
3.5 本章小结	47
3.6 家庭作业	48
第 4 章 用户体验的重要性	51
4.1 “无知的”用户体验	52
4.2 消费者案例分析: 提高 客户参与	54
4.3 商业案例研究: 启用 一线员工	56
4.3.1 门店经理仪表盘	56
4.3.2 示例用例: 竞争分析	58
4.3.3 其他用例	59
4.4 B2B 案例研究: 使渠道 更有效	60
4.4.1 顾问是你的合作伙伴 ——助他们成功	61
4.4.2 理财顾问案例研究	61
4.4.3 理财顾问仪表盘的 信息部分	63
4.4.4 理财顾问仪表盘的 建议部分	65

4.5 本章小结	68
4.6 家庭作业	69

第 II 部分 数据科学

第 5 章 商业智能和数据科学 之间的差异	73
5.1 什么是数据科学	74
5.1.1 商业智能与数据科学: 问题是不同的	75
5.1.2 商业智能问题	76
5.1.3 数据科学的问题	76
5.2 分析师各具特点	77
5.3 分析方法不同	78
5.3.1 商业智能分析师的参与 过程	78
5.3.2 数据科学家的参与过程	80
5.4 数据模型不同	82
5.4.1 商业智能的数据模型	82
5.4.2 数据科学的数据建模	83
5.5 商业角度的不同	85
5.6 本章小结	88
5.7 家庭作业	88
第 6 章 数据科学 101	89
6.1 数据科学案例研究设置	89
6.2 基础的探索性分析	91
6.2.1 趋势分析	91
6.2.2 箱形图	94
6.2.3 地理(空间)分析	95
6.2.4 配对图	95
6.2.5 时间序列分解	96
6.3 分析算法与模型	97
6.3.1 聚类分析	98
6.3.2 正态曲线当量(NCE)分析	99

6.3.3 关联分析.....	100	7.9 家庭作业.....	127
6.3.4 图形分析.....	101	第 III 部分 商业利益相关者的数据科学	
6.3.5 文本挖掘.....	102	第 8 章 像数据科学家一样思考.....	131
6.3.6 情感分析.....	103	8.1 像数据科学家一样思考的 过程.....	132
6.3.7 遍历模式分析.....	104	8.2 本章小结.....	142
6.3.8 决策树分类器分析.....	105	8.3 家庭作业.....	143
6.3.9 同期群分析.....	106	第 9 章 “By” 分析技术.....	145
6.4 本章小结.....	108	9.1 “By” 分析技术的简介.....	146
6.5 家庭作业.....	110	9.2 “By” 分析练习.....	148
第 7 章 数据湖泊.....	111	9.3 基于 “By” 分析的 Foot Locker 用例.....	152
7.1 数据湖泊简介.....	112	9.4 本章小结.....	154
7.2 支持商业运行的数据 湖泊特征.....	114	9.5 家庭作业.....	154
7.3 使用数据湖泊跨越分析 鸿沟.....	115	第 10 章 评分开发技术.....	157
7.4 数据和分析环境的现代化.....	117	10.1 分数的定义.....	158
7.4.1 行动 1: 创建基于 Hadoop 的数据湖泊.....	117	10.2 FICO 分数示例.....	158
7.4.2 行动 2: 分析沙箱的简介.....	118	10.3 其他行业评分的示例.....	161
7.4.3 行动 3: 摒弃数据 仓库中的 ETL 过程.....	119	10.4 继续讨论勒布朗·詹姆斯的 示例.....	162
7.5 分析辐射型分析架构.....	120	10.5 继续讨论 Foot Locker 的 示例.....	166
7.6 早期学习.....	121	10.6 本章小结.....	169
7.6.1 经验 1: 命名并不重要.....	122	10.7 家庭作业.....	169
7.6.2 经验 2: 它是数据湖泊, 而不是数据湖区.....	122	第 11 章 货币化训练.....	171
7.6.3 经验 3: 数据治理是 一个生命周期, 而不是 一个项目.....	123	11.1 健身跟踪器货币化示例.....	172
7.6.4 经验 4: 数据湖泊优于而 不是落后于数据仓库.....	124	11.1.1 步骤 1: 了解产品使用.....	172
7.7 未来是什么.....	125	11.1.2 步骤 2: 开发商业利益 相关者的人物角色.....	173
7.8 本章小结.....	126	11.1.3 步骤 3: 对潜在的建议 集思广益.....	174

11.1.4 步骤 4: 确定支持的
数据源 175

11.1.5 步骤 5: 对盈利机会
进行优先级排序 177

11.1.6 步骤 6: 开发货币化
计划 178

11.2 本章小结 179

11.3 家庭作业 179

第 12 章 商业蜕变训练 181

12.1 商业蜕变回顾 182

12.2 商业蜕变训练 183

12.2.1 明确商业蜕变构想 183

12.2.2 了解你的客户 184

12.2.3 明确价值主张 184

12.2.4 定义数据和分析需求 185

12.3 卫生保健中的商业蜕变 190

12.4 本章小结 193

12.5 家庭作业 194

第 IV 部分 构建跨企业间的支持

第 13 章 构想的力量 197

13.1 构想: 为创造性思维助力 198

13.1.1 大数据前景研讨会的
流程 198

13.1.2 前期研究 199

13.1.3 采访商业利益相关者 200

13.1.4 用数据科学进行探索 200

13.1.5 研讨会 202

13.1.6 设计研讨会 204

13.2 最优化矩阵 206

13.3 本章小结 208

13.4 家庭作业 208

第 14 章 释放企业机构影响力 209

14.1 首席数据货币化执行官 209

14.1.1 首席数据货币化执行官的
职责 210

14.1.2 首席数据货币化执行官
机构 210

14.1.3 分析卓越中心 211

14.1.4 首席数据货币化执行官的
领导力 212

14.2 隐私、信任度和决策治理 212

14.2.1 隐私问题=信任问题 213

14.2.2 决策管理 214

14.3 释放企业机构的创新能力 214

14.4 本章小结 216

14.5 家庭作业 217

第 15 章 故事 219

15.1 客户和员工的分析 221

15.2 产品和设备分析 224

15.3 网络和运营分析 225

15.4 一个好的商业故事的特点 227

15.5 本章小结 228

15.6 家庭作业 228