

A BOOK TO READ

章沛轩◎著

大数据，是生活管理、工作管理、
商业管理的准确、有效的利器！



一本书读懂

BIG DATA

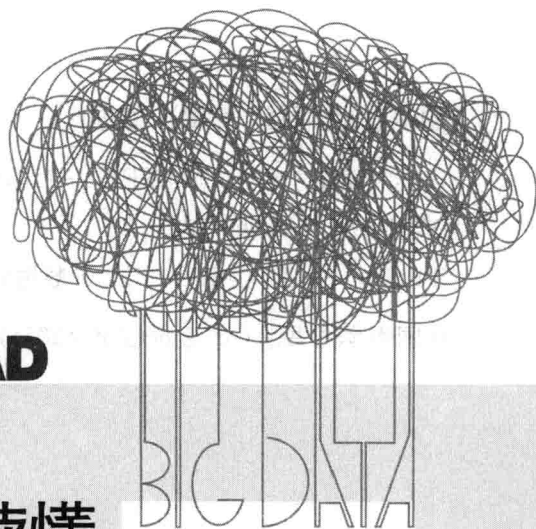
大数据

理解大数据，革新思维
掌握大数据，提高能力
运用大数据，赚钱轻松

BIG DATA

中国商业出版社

A BOOK TO READ



一本书读懂

大数据

章沛轩◎著

BIG DATA

中国商业出版社

图书在版编目 (CIP) 数据

一本书读懂大数据 / 章沛轩著. —北京 : 中国商业出版社, 2015.1

ISBN 978-7-5044-8823-7

I. ①一… II. ①章… III. ①数据处理 IV. ① TP274

中国版本图书馆 CIP 数据核字 (2015) 第 011353 号

责任编辑 : 朱丽丽

中国商业出版社出版发行

010-63180647 www.c-cbook.com

(100053 北京广安门内报国寺 1 号)

新华书店总店北京发行所经销

北京市京东印刷厂印刷

*

710 × 1000 毫米 16 开 14 印张 200 千字

2015 年 7 月第 1 版 2015 年 7 月第 1 次印刷

定价 : 32.00 元

* * * *

(如有印装质量问题可更换)



前言

preface

在商场购物付款的时候，你付给对方的可能不是花花绿绿的纸钞，而是在一个机器上刷下你的银行卡，输入一个相应的数字就可以了；每个月领的薪水，老板不会直接给你一沓钞票，而是让会计定期给你的银行卡上增加一些数额；电话响起，一连串的数字立马让你明白对方是谁以及电话重要与否；在公司，老板对你的评价可能是以数字的形式呈现的……数字，存在于我们工作、生活的方方面面，这让我们对自己方方面面的关注都集中在数字的多少和变化上。人们突然发现，不仅是生活和工作，做生意、进行企业管理乃至整个时代都被数据“控制”，我们已经离不开数据了。“大数据”也成了人们竞相议论的热词，但很少有人真正了解大数据。

关于大数据，目前业界并没有一个统一的定义，只有几个一致的观点。有两份资料很好地诠释了大数据的本质。第一份来自于高德·纳咨询公司的梅尔夫·阿德里安（Merv Adrian）2011年第一季度刊登在《天睿杂志》（Teradata Magazine）上的一篇文章。他说：“大数据超出了常用硬件环境和软件工具在可接受的时间内为其用户收集、管理和处理数据的能力。”



另一份来自于麦肯锡全球数据分析研究所（Mckinsey Global Institute）在2011年5月发表的一篇论文：“大数据是指大小超出了典型数据库软件工具收集、存储、管理和分析能力的数据集。”这些资料暗示着大数据的界定会随着技术的进步而变化。

在大数据时代，只有让数据成为自己的利器才能决胜。

对于企业来说，首先，需要拥有一套具有商业敏感度的数据决策框架，可以使企业“看”得更准，并能够对近期做了什么是对的、什么是错的进行判断。这样一来，快速的数据反馈可以让每个决策的误差得到适时修正。其次，让数据真正从“看”到“用”，让用数据成为构建企业生产力的重要部分。

对于个人来说，大数据让每个人的思维方式出现了重大改变。很多时候，我们欠缺的不是解决问题的方式，而是定位问题的能力。我们应该思考：有没有数据可以改善我们的盲点；我们该如何学会用“假定数据是可获取的”来重新思考周围的一切。当这种数据化思考成为我们的“直觉”时，就能够把数据的力量和思考的力量融合在一起，产生出新的无与伦比的合力了。到那时我们会发现，周围的一切都将因此而焕发出新的生机。

本书拒绝从纯专业技术的角度去阐述什么是大数据，而是从应用的角度，从数据运用的普遍性和实用性方面揭示大数据的作用。让更多的人了解什么是大数据和如何运用大数据；让更多的人将大数据这个工具运用到实践中。

目录

contents

第一篇 认识大数据的魔力

第一章 认识大数据的基本常识

- 大数据时代悄然而至 / 3
- 大数据产生的背景 / 6
- 一切都可以数据化 / 9
- 统计学：解析大数据的工具 / 13
- 数据大小怎么算 / 16

第二章 数据的收集管理和使用

- 收集数据不是最要紧的 / 19
- “活”做数据收集 / 21
- “活”看数据指标 / 24
- 数据的存放和管理 / 26
- 注意数据分类的维度 / 29
- 如何应用存储的数据 / 34

第三章 看看大数据的价值在哪里

- 数据中蕴含着宝藏 / 38

- 数据的分类估值 / 40
- 认清数据的五大价值 / 43
- 大数据不一定有大价值 / 48
- 大数据的价值遗憾 / 50
- 旧数据也会有新用途 / 54

第二篇 大数据时代下的企业运营

第一章 大数据时代的管理者

- 顶尖企业的数据管理模式 / 59
- 学会用客观数据说话 / 63
- 管理者必备的数据运用能力 / 67

第二章 大数据的管理工具

- 数据下的PDCA管理循环 / 71
- 用数据预测未来并规避风险 / 75
- 在绩效管理中引入商务智能 / 78
- 如何对员工进行量化评估 / 82

第三章 大数据下的商业运行

- 将信息变成一种企业竞争优势 / 87
- 响应潮流，企业构建大数据战略 / 89
- 养企业一定要学会“养数据” / 91

第四章 商业典范的大数据应用

- 数据改写引擎巨头商业模式 / 94
- 腾讯QQ的成功：一家独大 / 98
- Linux数据技术的商业支援 / 102

通过数据建立创新体系 / 106

第三篇 走进大数据营销时代

第一章 了解数据库营销

数据库营销的特点 / 113

数据库营销的作用 / 115

数据库营销的实施步骤 / 118

营销量化数据的四种类型 / 122

第二章 大数据改变营销模式

大数据是如何改变传统营销的 / 125

大数据下的自动化营销 / 129

大数据可以全程辅助营销 / 131

营销面对新的机遇与挑战 / 133

第三章 大数据精准的营销推广

大数据让广告智能化 / 137

大数据下的品牌代言 / 141

大数据下用户体验的威力 / 144

大数据下的预测性营销推广 / 148

第四篇 无处不在的大数据运用

第一章 有趣的大数据运用

大数据下的地震预测 / 153

是什么方便了我们的出行 / 156



- 大数据帮了奥巴马 / 159
- 大数据让罪犯无处可逃 / 161
- 大数据造就美味机器 / 164

第二章 我们离不开的数据生活

- 我们的健康信息藏在数据里 / 169
- 大数据带来的医疗新突破 / 173
- 智能时代让生活更健康 / 176

第三章 名副其实的“情网”

- 大数据下的人际关系网 / 181
- 网络约会成功的秘密 / 184
- 网络中的人际交流 / 188
- 大数据下的浪漫情感 / 191

第四章 大数据预见的行业未来

- 从小教室到世界大课堂 / 194
- 大数据下的传媒影视 / 196
- 制造业的大数据之路 / 199
- 餐饮旅游业的未来 / 202
- 被大数据颠覆的传统农业 / 204

第五章 大数据时代的个人数据管理

- 做好个人大数据管理 / 207
- 用标签管理个人大数据 / 209
- 让自己成为数据分析师 / 211
- 自己的大数据是可以卖的 / 214

第一篇 认识大数据的魔力



第一章 认识大数据的基本常识

大数据时代悄然而至

我们每天坐在电脑前在网上搜索、购物、发信息、发图片，用手机打电话、发彩信……很多人都在不经意中制造和使用着数据。因此，产生的数据量正在以惊人的速度增长——数据已经渗透到生活和工作的方方面面，成为时代的特征。

2003年，刚进大学的小强还没有自己的电脑，他省吃俭用花600元买了一个不知名品牌的MP3播放器，容量只有128M。拿到MP3播放器后他非常欣喜，因为这个MP3播放器能存储大约50首普通压缩率的MP3歌曲，还能当软盘用。而他之前一直使用3.5寸的软盘来存储数据，一张软盘的容量仅为1.44M。之前他用来听歌的设备是一台索尼随身听，要听新歌只能花钱买磁带，每盘磁带大约30元，只能存储10首歌，而且没法自己挑选想要听的歌。

10年后的2013年，小强已经参加工作多年，他平时使用智能



手机听歌、上网，使用平板电脑玩游戏、购物、看电影，家里的笔记本电脑反而用得越来越少了。最近他想买一个移动硬盘来存储高清电影，他在网上花600元买了一个2T的移动硬盘。这个2T的移动硬盘大约能存储1000部高清电影，如果用来存储普通压缩率的MP3歌曲，大约能存储80万首。

不考虑货币购买力变化及产品功能等问题，只考虑数据容量，同样是600元，2013年购买到的容量是2003年的1600倍。可是，小强还是觉得容量不够用，这10年里到底是哪里出了问题呢？是什么促进了存储介质的发展呢？

不仅是小强，几乎所有的个人、单位都已经觉得原来购买的存储设备容量不够用。随着互联网的逐渐成熟和发展，移动带宽迅速提升，云计算、互联网应用更加丰富。更多的传感设备、移动终端接入网络，由此产生的数据及数据增长速度迅速攀升。数据已经渗透到世界每一个角落、每一个行业和每一个业务职能领域，逐渐成为重要的生产因素和时代的标志；而人们对于海量数据的运用将预示着新一波生产率增长和消费者盈余浪潮的到来。

大数据时代已经悄然来临。在这样的背景下，人们逐渐认识到大数据给自己带来的作用，大数据的概念也逐渐在人们的脑海中凸显。这就让人们需要一个巨大容量的存储设备。所以，大数据产生的一个显著标志就是：为了满足市场的要求，数据存储设备的容量越来越大，处理数据的速度越来越快。

以“脸谱”为例。2012年10月，“脸谱”声称其用户量已突破10亿人——近乎世界总人口的1/6。仅仅是用户量就是一个巨大的数字，这让“脸谱”接待这些客户有些吃力，“脸谱”不得不

开发各种各样的新技术以跟上用户暴涨的脚步，而这些都需要数据去支撑。

数据暴涨的不仅是人数，更有因人数剧增而产生的一些技术性数据。时代的发展让“脸谱”以惊人的速度处理这些数据，据统计，“脸谱”每天需要处理近30亿个“赞”、25亿次内容分享以及3亿张上传照片。也就是说，这家公司每天存储的数据超过100PB，将这些数据分析处理后，得到超过500PB的新数据。这就相当于2000台苹果电脑硬盘驱动器存储的数据量。你可能还不了解“PB”是什么样的数据单位，这个单位究竟有多大，接下来的内容会让你知道这是个多么庞大的数字量。

权威机构的一项调查发现，90%的企业的数据量在迅速上涨，其中近20%的企业的数据量每年增长一半甚至更多。

国际数据公司2011年6月的报告显示，全球数据量在2011年已达到1.8ZB，在5年里增加了5倍。1.8ZB有多大呢？如果把所有这些数据都刻录存入普通DVD光盘里，光盘的高度将等同于从地球到月球的一个半来回，也就是大约72万英里，相当于每位美国人每分钟写3条推特，要不停地写2.6976万年。IDC预测全球数据量大约每两年翻一番，2015年全球数据量将达到近8ZB，到2020年，全球将达到35ZB。我们可以想象，这是一个多么庞大的数据。当然，这也让我们发现，大数据时代已经悄然而至。事实上，今天的全球互联网巨头都已意识到大数据时代数据的重要意义。包括易安信、惠普、IBM、微软在内的全球IT巨头纷纷通过收购大数据相关厂商来实现技术整合，这足以看出它们对大数据的重视。



大数据产生的背景

为什么大数据似乎在一夜之间悄然而至，并迅速走红呢？为什么人们会把2012年称为“大数据的跨界年度”？大数据之所以会在2012年进入主流大众的视野，我们分析得知，缘于三种趋势的合力。

第一，随着互联网的发展，许多高端消费公司为了提供更先进的、更完美的服务，加大了对大数据的应用。

比如“脸谱”就使用大数据来追踪用户，然后通过“搜索和识别你所熟知的人”，“脸谱”可以给出好友推荐建议。用户的好友数目越多，他对“脸谱”的信任度就越高。好友越多同时也就意味着用户分享的照片越多、发布的状态更新越频繁、玩的游戏也越多样化。后文会提到，“脸谱”因此在和同行的竞争中占得先机。

商业社交网站领英则使用大数据为求职者和招聘单位建立关联。有了领英，猎头公司就不再需要对潜在人才进行繁琐的识别和访问。只需一个简单的搜索，他们就可以找到潜在人才，并与他们进行联系。同样，求职者也可以通过联系网站上的其他人，将自己推销给潜在的人力资源负责人，入职自己中意的公司。

杰夫·韦纳是领英的首席执行官，他在谈到该网站的未来发展时谈到一个经济图表，这是一个能实时识别“经济机会趋势”

的全球经济数字图表，他说，实现该图表及其预测能力时所面临的挑战就是一个大数据问题。

可以看出，大家都在利用大数据产生利益，反过来，利用大数据的人就成了催生大数据时代到来的力量之一。

第二，人们在无形中纷纷为大数据投资。

还是以实际的公司为案例。“脸谱”与领英两家公司都是在2012年上市的。“脸谱”在纳斯达克上市，领英在纽约证券交易所上市。从表面上来看，谷歌和这两家公司都是消费品公司，而实质上，它们都是利用大数据吃饭的企业。除了这两家公司以外，一家为大中型企业提供运营智能的大数据企业Splunk也在2012年完成了上市。这些企业的公开上市使华尔街对大数据业务的兴趣非常浓厚。因此，硅谷的一些风险投资家们开始前赴后继地为大数据企业提供资金，这给大数据的发展提供了前所未有的良机。大数据将引发下一波重大转变，在这场转变中，硅谷有望在未来几年取代华尔街。

作为“脸谱”的早期投资者，加速合伙公司在2011年年底宣布为大数据提供一笔不小的投资，2012年年初，加速合伙公司支出了第一笔投资。著名的风险投资公司格雷洛克合伙公司也针对这一领域进行了大量的投资。

第三，商业用户和其他以数据为核心的消费产品，也开始期待以一种同样便捷的方式来获得大数据的使用体验。

我们在网上看电影、买产品——这些已经成为现实。既然互联网零售商可以为用户推荐一些阅读书目、电影和产品，为什么这些产品所在的企业却做不到呢？举个例子说，为什么房屋租赁



公司不能明智地决定将哪一栋房屋提供给租房人呢？毕竟，该公司拥有客户的租房历史和现有可用租房屋库存记录。随着新技术的出现，公司不仅能够了解到特定市场的公开信息，还能了解到有关会议、重大事项及其他可能会影响市场需求的信息。通过将内部供应链与外部市场数据相结合，公司可以更加精确地预测出可租的房屋类型和可用时间。

类似地，通过将这些内部数据和外部数据相结合，零售商每天都可以利用这种混合式数据确定产品价格和摆放位置。通过考虑从产品供应到消费者的购物习惯这一系列事件的数据（包括哪种产品卖得比较好），零售商就可以提升消费者的平均购买量，从而获得更高的利润。

所以，商业用户也成为了推动大数据发展的动力之一。

从我们所举的例子看，好像是少数的几家公司推动大数据的产生，的确是这样。但总的来说，大数据的产生既是时代发展的结果，也是利益驱使的结果。当然，那些小公司的发展，乃至个人的服务需求也在为大数据的产生添砖加瓦，只是单个个体的效果不明显，但反映在整个大数据产业中依然是巨大的，其中的道理不再多说了。