

★ ★ ★
彩色图解版
★ ★ ★

一本书读懂 大数据

★
得数据者得天下，大数据时代，你也可以
成为大赢家！

吕长青 | 编著

大数据时代没有旁观者 每个人都看得懂的大数据入门书

【知识精简】

深入浅出地剖析大数据，
让你轻松阅读

【流程图解】

以易读好学的方式，
增加阅读趣味

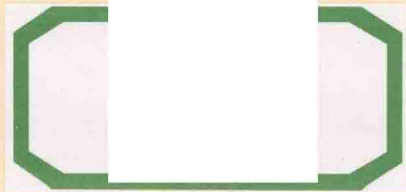
【案例翔实】

各行业经典有趣的案例，
大数据全揭秘

【数据掘金】

企业个人数据掘金方案，
领跑数据时代

北京工业大学出版社



一本书读懂 大数据

吕长青 | 编著



RFID

北京工业大学出版社

图书在版编目 (CIP) 数据

一本书读懂大数据 / 吕长青编著. —北京: 北京工业大学出版社, 2016. 11

ISBN 978-7-5639-4934-2

I. ①一… II. ①吕… III. ①数据处理 IV.
①TP274

中国版本图书馆CIP数据核字 (2016) 第236093号

一本书读懂大数据

编 著: 吕长青

责任编辑: 翟雅薇

封面设计: 书心瞬息

出版发行: 北京工业大学出版社

(北京市朝阳区平乐园 100 号 邮编: 100124)

010-67391722 (传真) bgdcbs@sina.com

出 版 人: 郝 勇

经销单位: 全国各地新华书店

承印单位: 三河市兴国印务有限公司

开 本: 710 毫米 × 1000 毫米 1/16

印 张: 16.25

字 数: 218 千字

版 次: 2016 年 11 月第 1 版

印 次: 2016 年 11 月第 1 次印刷

标准书号: ISBN 978-7-5639-4934-2

定 价: 38.00 元

版权所有 翻印必究

(如发现印装质量问题, 请寄回本社发行部调换 010-67391106)

序言

好像一转眼间，大数据就如浪潮般席卷而来，正在以前所未有的方式深层次、多角度地改变着社会，包括商业模式、卫生领域、人们的生活和工作等。大数据展现了其恢宏磅礴的气势和无法阻挡的霸气，各种大数据运用层出不穷，大数据创造的奇迹一个接着一个在世界范围内传播。

事实上，信息爆炸必然会导致数据飞速增长，尤其是互联网、智能设备、物联网、车联网，以及遍布全球的传感器，皆是数据来源或者承载的方式。然而尽管大数据如此火爆，人们目前对大数据的利用率却非常低。众所周知，人类的大脑只开发了 20%，如果将剩余的 80% 也开发出来，会发生怎样的转变呢？如果大数据也能够将全部数据价值挖掘出来，会产生什么样的变化呢？没人知道，但无人可以否认，大数据已然有着广泛而光明的前景，其未来价值不可估量。

但大数据火热的同时，也带来一些问题，如数据安全和数据隐私问题。随处都有人们留下的数据，如果这些数据不能合理运用，则会造成严重的后果。

大数据时代既给人们带来前所未有的机遇，也带来了严峻的挑战。如果能顺利乘上大数据时代的快车，那么则能迎来新时期，对企业、

个人等可谓是一种新生；但如果落伍，无法顺应时代趋势，将面临非常严峻的挑战。那么大数据时代，路在何方呢？

本书内容翔实，覆盖面广，从点到面，从问题到根源，详细全面地阐述了大数据所带来的影响、大数据应用、大数据解决方案等问题，同时也指出在这种局势下，个人和企业该如何应对，从而把握时代趋势，从容赢得先机。

书中还搭配了丰富翔实的表格和图片，便于读者理解。相比文字，图片和表格往往更让人印象深刻。本书在行文上全面抛弃了“假、大、空”的套话，采用通俗易懂的语言，打破了人们对大数据的畏惧感。读过本书，人们会感慨：原来这就是大数据，我要这么应对，很简单啊。这正是编者所希望的。

由于编者时间仓促和经验上有不足之处，书中难免有疏漏和不成熟的地方，敬请读者予以批评指正。

目 录

第一章

新起航：

互联网+大数据，产业展翅

- 1.1 新趋势：大数据时代真的来临了002
- 1.2 新概念：你该这么认识大数据005
- 1.3 四大特征：帮你把脉大数据011
- 1.4 五大事实：大数据不是传统数据014
- 1.5 六大误解：拨开迷雾认识真正的大数据018
- 1.6 时效性：昨天的大数据已不是今天的大数据024
- 1.7 发展历程：看大数据如何横空出世026
- 1.8 大影响：大数据带来的四大变革030

第二章

全方位：

360 度全面了解大数据

- 2.1 起源：第三次浪潮的华彩乐章 036
 - 2.1.1 量累积：为有源头数据来 036

2.1.2	颠覆性：摩尔定律现象	039
2.1.3	量质变：由数据到大数据	042
2.1.4	新篇章：一切皆可数据化	045
2.2	现状：数字时代的“石油”与“黄金”	049
2.2.1	新引擎：互联网+的“发动机”	049
2.2.2	新宝藏：“互联网+大数据”掘金新方式	051
2.2.3	揭幕战：百度、谷歌、微软一较高下	054
2.2.4	云数据：存在“云”里的大数据	057
2.2.5	新势力：物联网的快速崛起	060
2.3	发展趋势：站在风口的大数据	066
2.3.1	新岗位：催生千万大数据相关岗位	066
2.3.2	新资产：未来最大的交易商品	070
2.3.3	私定制：量身定制服务时代	073
2.3.4	数共享：越来越开放的数据	075
2.3.5	重安全：数据王国安全为根基	079
2.3.6	看未来：大数据引领智能化社会	083

第三章

掘金红海：

大数据的行业解决方案

3.1	电商：乘上大数据翅膀展翅高飞	088
3.1.1	“双十一”背后的大数据	088
3.1.2	销售分析：察觉数字背后的奥秘	092

3.1.3	大数据如何为电商企业增加利润	098
3.1.4	大数据动态分析定制专属广告	102
3.1.5	以人为本的移动互联网营销	104
3.2	传媒影视：什么样的内容才是观众最喜爱的	108
3.2.1	大数据催生《纸牌屋》	108
3.2.2	社交媒体传播成就了《致青春》	113
3.2.3	大数据告诉你《芈月传》有多火	115
3.2.4	IP 是什么，为何这么重要	118
3.2.5	大数据或将决定影视内容生产	119
3.3	个人生活：大数据离我们的生活并不遥远	122
3.3.1	教育改革，世界大课堂出现与个性化教育	122
3.3.2	医疗，健康信息隐藏在数据里	125
3.3.3	可穿戴设备，随时上传数据	127
3.3.4	信用生活，生活更智能	129
3.3.5	大数据对人际关系的影响	131
3.4	互联网金融：大数据理财时代	134
3.4.1	互联网金融的异军突起	134
3.4.2	钱该放卡里，还是网络里	135
3.4.3	社交化的互联网金融	136
3.4.4	传统金融行业该如何应对大数据的兵临城下	138
3.5	大数据预测：新时代的“预言家”	142
3.5.1	世界杯“预言帝”	142
3.5.2	奥斯卡金像奖提前揭晓“结果”	144

3.5.3	预测准不准谁说了算	147
3.5.4	快数据，由预测到现测	150
3.5.5	数据预测与人的直觉	153

第四章

火眼金睛：

找出隐藏在数据背后的奥秘

4.1	浪里淘沙：穷尽大数据海域终见沙	158
4.1.1	用最少数据获取最多信息	158
4.1.2	利用所有的数据	160
4.1.3	快速、多源、复杂化的数据	161
4.1.4	数据挖掘流程全解	163
4.1.5	为什么要挖掘数据	167
4.2	沙里淘金：数据分析助你一臂之力	169
4.2.1	四种入门分析方法，你 get 了吗	169
4.2.2	分类：道不同不相为谋	175
4.2.3	大数据分析的五个基本面	177
4.2.4	并不绝对准确，允许出错	179
4.3	可视化：让数据自己说话，让观众听到结果	182
4.3.1	图片比数字更让人印象深刻	182
4.3.2	将数据隐藏的信息图片化	186
4.3.3	常用的数据可视化	187
4.3.4	数据可视化的发展历程	192

4.3.5 让图表“动”起来	194
4.3.6 新闻界，可视化的力量	196

第五章

以人为本：

方今之急在于人才而已

5.1 数据分析专家：大数据海域航行靠舵手	199
5.1.1 为什么要成为数据分析专家	199
5.1.2 优秀数据分析专家的与众不同之处	201
5.1.3 自主学习是数据分析专家的特质	204
5.1.4 企业如何招聘数据分析人才	207
5.1.5 教你快速成为数据分析专家	208
5.2 数据分析团队：航行在大数据海域的航空母舰	211
5.2.1 为什么要打造分析团队	211
5.2.2 团队组织结构	213
5.2.3 团队管理的几大原则	214
5.2.4 如何优化团队	217
5.2.5 数据高级分析工作应由谁来做	219
5.2.6 优秀数据分析团队的特征	220
5.2.7 团队建设的五个阶段	222
5.2.8 处理好团队成员间的冲突	225

第六章

领跑者：

寻求大数据时代的出路

6.1 企业：企业的大数据怎么做	229
6.1.1 战略：构建大数据战略	229
6.1.2 系统：搭建数据化运营系统	231
6.1.3 产品：善于使用 BAT 等大数据	233
6.1.4 营销：全方位分析客户，实现精准营销	234
6.2 个人：如何搭上大数据时代的顺风车	238
6.2.1 取势：善战者谋其势而利导之	238
6.2.2 明道：物有本末、事有终始，知所先后，则近道矣	240
6.2.3 优术：工欲善其事，必先利其器	241
6.2.4 践行：大道至简，知行合一	244
6.2.5 合众：独学而无友，则孤陋而寡闻	246

第一章 新起航：

互联网 + 大数据，产业展翅

信息爆炸由来已久，但最近几年人们的感受更为深刻，这一方面是因为网民数量的大幅增加，另一方面是互联网、云计算等纷纷兴起，导致各种设备数量增速极快，每天都有无数的数据因此而产生；大数据已成为战略性资源，影响越来越大，因此人们感觉到大数据已然就在身边，且与自己的生活息息相关，而大数据所席卷而来的变革具有颠覆性、全面性，正深入各个层面。大数据时代真的来临了。



新趋势： 大数据时代真的来临了

从古人使用计量单位来计算农作物，或者计算打猎收获开始，数据便出现了，从那以后，数据便成为常用的方式，有时会统计下人口数量，有时会统计作物收成，有时会根据数据进行作物或猎物分配，有时也会根据数据进行分析，比如今年作物的产量相对去年是增产还是减产，猎物数量相比去年同时期是否有所增加。数据收集和数据分析便成为促进社会进步的重要方式。

17、18 世纪，微积分、统计学以及概率论的提出，为数据收集和数据分析提供了理论基础，同时科学家还发明了一些新的工具，可以对以往无法进行统计的数据进行收集和分析，如犯罪率、农作物收成预测、结婚率等，数据收集和数据分析由此得以迈入新的阶段。

后来，数据对社会生活的影响越来越重，19 世纪时，约翰·斯诺曾使用数据科学绘制了“霍乱群聚”图。那个时期，伦敦一直为霍乱所困扰，科学家进行了大量调查，最后认为导致霍乱的根源在于有毒空气，斯诺博士对此有所怀疑，他怀疑霍乱跟水源有关，于是他将霍乱爆发频繁、严重的地区当作调查对象，详细调查了这些地区被污染的水源，从而找到了导致霍乱产生的根源。

随后数据科学得以运用在工业中，极大提升了工作效率，增加了工厂的产量。不久后，人们在其他领域使用数据科学，发现一个很重要的事实：数据波及哪个领域，哪个领域就会发生一场变革。而且人们如能合理利用数据，则能提升该领域的价值，节省相关费用。

人类社会步入互联网时代后，数据迎来了新的发展机遇，从而正式迈入大

数据时期。大数据对人类社会的影响是颠覆性的，从小作坊到大型企业、到各类组织、政府，无一不深受大数据的影响。大数据对个人也产生了深远影响，使生活、职场、思维都发生了颠覆性的变革。

大数据颠覆传统的决策方式，以海量数据为基础，从数据中挖掘出有价值的信息，并以此来指导决策，甚至主导决策、决定决策，这种方式颠覆了以往的以直觉、阅历、经验等为主的决策方式。人们发现数据量越是丰富，其分析结果就越准确，因此随着数据增长速度加快，人们收集数据的能力逐步提升，在未来，数据将主导决策。这对企业、个人的影响很大。

大数据几乎改变了每个行业，从医疗保险到天空探险、从传统银行到互联网金融、从小微企业到跨国公司、从路边小饭馆到五星级餐饮等，几乎所有的行业数据增长速度都越来越快。如今一家小企业所拥有的数据量都有可能超过图书馆所拥有的数据量。根据相关预测，大数据在未来很长一段时间里，将保持高速增长的状态，将会对每一个行业带来深远的影响。

在过去，创新多依靠人类自身的突发奇想，而在大数据时代，创新将会变得简单，只要找到数据间的相关关系就能促使创新出现。人们以海量数据为基础，深入挖掘用户需求，挖掘产品或者技术上的不足之处，从而促进技术升级和换代。随着数据应用范围越来越广，数据所爆发出的能量将更为惊人。

今天，数据比以往更深入且全面地与社会、集体、个人等交织在一起，整个地球都在数据的包围之中，人们自身也成为数据的生产者、利用者。尤其是随着云计算、云空间的出现，计算机处理能力的提升，存储成本的降低，物联网的兴起，移动互联网的火热等，大数据已经涉及人类生活的层层面面，成为社会进步、经济发展、人类文明进步的必经之路。

人类从大数据中获益匪浅：银行能够了解到信用风险；政府可以根据数据适时调整政策，以惠万民；医疗运用数据监测我们的健康情况，使一些较为顽固

的疾病可以及早发现……总之，大数据正以燎原之势强势袭来，且将在未来社会（如智能社会）中发挥更为重要的作用。

事实上，信息爆炸由来已久，但最近几年人们的感受更为深刻，这一方面是因为网民数量的大幅增加，另一方面是物联网、云计算等兴起，导致各种设备数量增速极快。另外就是互联网企业在入口和应用场景的竞争越来越激烈，这些竞争也是数据资源的竞争，大数据已成为战略性资源，影响越来越大，因此人们才感觉到大数据已然就在身边，且与自己的生活息息相关，而大数据所席卷而来的变革具有颠覆性、全面性，正深入各个层面，基于此，人们难免会想到，大数据时代真的来临了。

1.2

新概念： 你该这么认识大数据

无论你是青睐或是讨厌，我们都生活在处处是“数据”的时代，数据包括数字、文字、图像、视频、声音等，我们每天打电话、上网等行为都在不断地增加数据总量。大数据已是一种事实存在，我们唯一能做的就是认识大数据，接受大数据的存在，同时做好应对大数据的准备。

大数据无疑是最近最热门的话题之一，和互联网+、云计算一样，影响深远且意义深刻。随着互联网的发展，各类信息每天以爆炸姿态快速增长，大量数据环绕在我们身边，不过尽管人们对大数据并不陌生，却很少有人能说清大数据的概念，即使是对新技术、新趋势抱以热忱之心的IT人，也很难说清大数据的概念。

下面三个案例，会让你对大数据有初步的了解。

案例分享

大数据对社会的影响

Google 发现流感病毒感染地区

2009年，美国境内出现了一种新型流感病毒，病毒具有禽流感病毒特征，传播快速，有蔓延全境的危险。疾控中心不能快速应对，最快也需一周的时间才能确定流感爆发的位置。流感病毒传染性强，一周的时间，足以传染到更多地区，后果严重。而此时Google（谷歌）通过分析搜索引擎上数千万美国人搜索的词汇，与疾控中心在2003年和2005年的流感病毒传染期的数据进行对比，通过搜索记录，如“哪些药治疗发热和咳嗽最有效”等，可以判断此地区是否受流感病毒感染。最终Google发现了搜索记录和病毒的相关联系，成功找出美国和世界其他一些受新型流感病毒感染的地区。



▲ 受流感病毒感染的地区

案例分享

大数据对企业的影响

啤酒与尿布经典营销案例

通常人们很难将啤酒和尿布联系起来，然而全球零售业巨头沃尔玛却出人意料地将两者摆在一起，推销给消费者。原来沃尔玛在分析消费者购物行为数据

交易号	产品
T01	啤酒
T01	尿布
T02	啤酒
T02	尿布
T03	尿布

