

玩转 大数据

(图解版)

IDKW图解中心 编著

这是一本人人能读懂的大数据入门书

采用图解形式，解读“大数据”知识
用简单的图形解答复杂的问题

一口气读完，不让思维“断片”
利用碎片化时间，掌握关键信息

双色内文，让阅读不再枯燥
提供最接地气的案例，抓住大数据时代的商机



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

玩转 (图解版) 大数据

IDKW图解中心 编著

人民邮电出版社
北京

图书在版编目 (CIP) 数据

玩转大数据：图解版 / IDKW图解中心编著. — 北京：人民邮电出版社，2015.5
ISBN 978-7-115-39141-4

I. ①玩… II. ①I… III. ①企业管理—数据管理—图解 IV. ①F270.7-64

中国版本图书馆CIP数据核字(2015)第084828号

内 容 提 要

本书采用图解形式对“大数据”这一概念进行了全新解读，将大数据从一个 IT 术语变成了人人能读懂的大众词汇。本书从大数据的“前世今生”讲起，通过列举各行各业应用大数据的案例，全面解读了大数据和个人的关系、如何利用大数据创业、大数据中隐藏了多少价值、大数据给我们带来了哪些冲击和变革、如何利用大数据进行精准营销等读者十分关心的热点问题。本书图文并茂，读者利用零散时间，就能迅速了解大数据知识。

本书适合企业管理者、创业者、投资者、高校相关专业师生等想运用大数据改变未来生活的人士阅读。

◆编 著 IDKW 图解中心

责任编辑 李宝琳

执行编辑 包华楠

责任印制 焦志炜

◆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>

北京天宇星印刷厂印刷

◆开本：700×1000 1/16

印张：13

2015 年 5 月第 1 版

字数：100 千字

2015 年 5 月北京第 1 次印刷

定 价：39.00 元

读者服务热线：(010)81055656 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京崇工商广字第 0021 号



大数据，开启一个伟大时代的关键词

在过去的几年里，数据已经进入了爆炸式的增长阶段。数据更多、更杂，也更个性化、更智能化。在大数据时代，任何一切都变得有可能。

无论你有没有意识到，大数据早已对人们的生产、工作和生活产生了巨大的影响。正如马云所说：“我们还没搞懂 PC 互联网的时候，移动互联网来了，我们还没搞懂移动互联网的时候，大数据来了。”

作为如此火爆的一个流行词，“大数据”到底意味着什么？

身处不同行业、不同岗位的人可能会给出不同的答案：海量的信息、重要的资产、预测的基础、决策的依据、精准的营销、高端的技术、更多的机会、被“扒光”的隐私……

然而，数据的挖掘和分析只是整个变革过程中的一个技术手段，而远非变革的全部。数据产生于各行各业，这场变革也必将影响到各行各业。

无论如何，大数据开启了一个新的伟大时代。它不仅是一个时髦的概念，更是一个无可逆转的趋势。正如《连线》杂志创始主编凯文·凯利所说：“不管你现在做什么行业，未来都是数据生意。”

大数据并不是高高在上、遥不可及的东西，它与每个人休戚相关。对于不同的受众群体，大数据也不会厚此薄彼，每个人都身处大数据的浪潮中，都要经受大数据时代带给我们的诸多冲击。如果你不能与时俱进，那么就会被时代所淘汰。

大数据时代，知道却不理解“大数据”这个关键词，你就落伍了。本书采用图解形式对“大数据”这一概念进行了全新解读，将大数据拉下“神坛”，从一个IT术语变成了人人能读懂的大众词汇。

本书具有以下特点。

1. 模块清晰，专业务实

全书本着以应用为中心、以有用为标准、以使用为落脚点的基本原则，将全部内容划分成8个模块：大数据到底是什么、大数据和你的关系、大数据催生大创业、大数据产生大价值、大数据引发大变革、大数据演绎大未来、大数据驱动大营销、大数据带来大应用。

本书既传授知识，又引导思考，还指点行动。

2. 图解新颖，寓意于形

本书采用新颖的图解形式，寓意于形。如果你想一口气读完，那么轻松的图解形式会让你一直兴趣盎然；如果你只想利用碎片化时间，那么模块化的呈现一定不会让你的思维“断片”。无论你是为了追赶潮流希望了解大数据知识，还是为了增长知识探索大数据的奥秘，或者是想利用大数据开拓一片属于自己的新天地，本书都是你的首选。

3. 轻松呈现，精准表述

你不必担心，本书绝不是一部高深晦涩、让人一头雾水的“天书”，也不是连篇累牍、赘述理论的 IT 典籍，更不是杂七杂八、东拼西凑的大数据相关知识汇总。“专业知识 + 精准图解 + 现实案例 + 明快语言”，带你迅速融入大数据的世界，既使阅读变得更轻松，又使理解变得更深刻。

大数据时代是一个人人都想抓住的机遇，衷心希望读者在阅读本书之后能有所收获。

最后，感谢王淑燕、刘伟、程富建、姚小凤、李作学、程淑丽、毕春月对本书的框架和体系给出了具体的思路和修改意见，感谢王春霞、高春燕、徐滕、张天骄、陈里、史册、王琴、贾晶晶提供了部分素材，感谢齐艳霞、王胜会、姜娣、蔚星星、金成哲、黄成日、杨茜审阅了部分内容，感谢贾月、孙立宏、罗章秀、刘井学为本书做了图形和版式设计。

CHAPTER 第1章 1

大数据到底是什么

1.1 定义大数据 // 2

大数据的
“前世今生”

3

大数据的
三个来源

7

大数据的
五个特征

8

1.2 理解大数据 // 11

“大数据”不等于
“数据大”

11

大数据与传统
数据的区别

13

结构化、非结构化
和半结构化的数据

15

CHAPTER 第2章 2

大数据和你的关系

2.1 为什么要读大数据 // 20

大数据时代
已经来临

20

得数据者得天下

22

2.2 大数据能做什么 // 27

大数据记录了
一切

27

大数据改变工作
与生活

28

影响甚至改变
未来

30

2.3 大数据和你有什么关系 // 32

所有人都生活在
数据中

32

“预言家”和
“英式管家”

34

CHAPTER 第3章 3

大数据催生大创业

3.1 大数据创业前景广阔 // 40

数据生产公司：
将大数据商品化

42

数据服务公司：
专注于解决方案

46

项目数据分析师事务所：
致力于数据分析

47

3.2 大数据创业的难点 // 49

创业公司面临的
三个问题

49

大数据专业人才的
缺乏

51

CHAPTER 第4章 4

大数据产生大价值

4.1 数据预测：放大数据价值 // 54

大数据的核心
就是预测

54

一切均可以
数据化

62

4.2 数据挖掘：发现潜在价值 // 67

数据是一种
资产

67

化“数”为
“据”是关键

72

4.3 数据运营：创造新的价值 // 77

大数据的价值
转化

79

大数据的潜在
价值和 innovation

82

CHAPTER 第5章 5

大数据引发大变革

5.1 思维变革：信息风暴颠覆思维惯例 // 94

从“样本”到
“总体”

94

从“精确性”
到“混杂性”

97

从“因果关系”到
“相关关系”

99

5.2 商业变革：大数据改写商业规则 // 101

商业模式的
创新

101

跨界的融合

104

大数据颠覆传统
媒体行业

106

大数据冲击
金融行业

107

大数据决定
企业竞争力

109

5.3 管理变革：用大数据说话 // 111

信息扁平化

111

基于大数据
的决策模式

114

5.4 技术变革：再造互联网 // 116

数据可视化

116

大数据与云
计算

121

5.5 教育变革：让教育更智能 // 124

在线教育的
浪潮

124

微学校和个
性化教育

130

未来的教育趋势

133

CHAPTER 第6章 6

大数据演绎大未来

6.1 大数据时代的三大发展趋势 // 138

数据成为资产

138

产业垂直整合

143

泛互联网化

146

6.2 大数据统治世界 // 152

数据垄断的困境
与隐忧

152

数据开放与隐私
的权衡

156

6.3 更多的改变正蓄势待发 // 160

大数据引领数据
智能时代

160

大数据驱动新的
工业革命

162

CHAPTER 第7章 7

大数据驱动大营销

7.1 从数据库营销到大数据营销 // 166

数据制造市场

166

从“产品至上”
到“用户至上”

168

7.2 大数据驱动精准营销 // 171

挖掘大数据
背后的需求

171

用大数据定
位客户

173

CHAPTER 第8章 8

大数据带来大应用

8.1 大数据浪潮席卷各行各业 // 176

大数据的政府
应用
176

大数据的行业
应用
182

大数据的个人
应用
189

8.2 大数据技术落地困难重重 // 191

大数据思维
尚未形成
191

数据收集来
源良莠不齐
193

数据分析的局限
194



Chapter 1

第 1 章 大数据到底是什么

“大数据”无疑是目前最受关注的热词之一，它不只是一个词汇，更是一门技术，而且开启了一个新时代。

1.1 定义大数据

2014年12月15日，由中国计算机学会（CCF）主办、CCF大数据专家委员会承办、中国科学院计算技术研究所与CSDN共同协办的“2014中国大数据技术大会”在北京圆满落幕。

从2008年仅60人参加的技术沙龙，到2014年数千人参加的技术盛宴，中国大数据技术大会见证了整个大数据生态圈技术的发展与演变。可以说，大数据引发了新一轮“工业革命”，几乎每个行业都面临着大数据问题。越来越多的人开始关注大数据，越来越多的企业开始参与到大数据的竞争中来。

如今，大数据已经不算一个新概念，从专业术语到大众热词，大数据在短短几年时间内迅速走红，成为了家喻户晓的流行语。虽然很多人常常把“大数据”一词挂在嘴边，但是却“知其然而不知其所以然”，大

数据究竟是怎么一回事谁也不能说清楚。

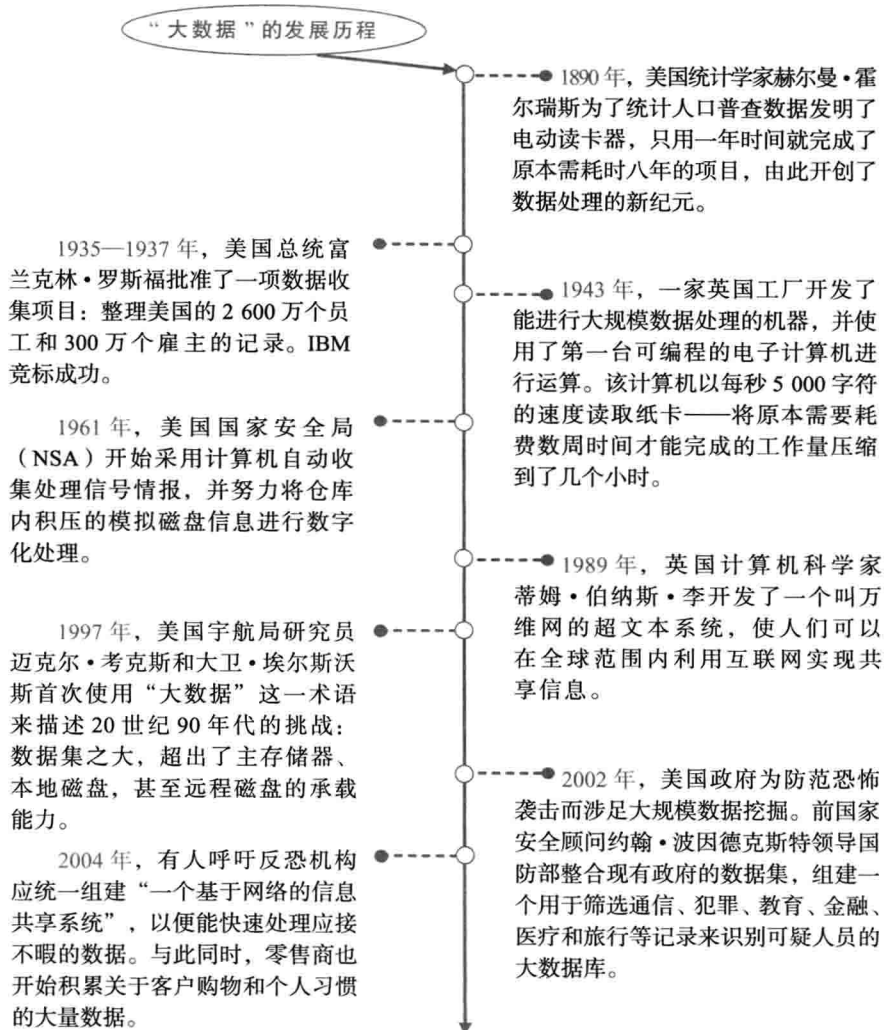
那么，大数据到底是什么？大数据是怎么“火”起来的？大数据又为什么这么“火”呢？

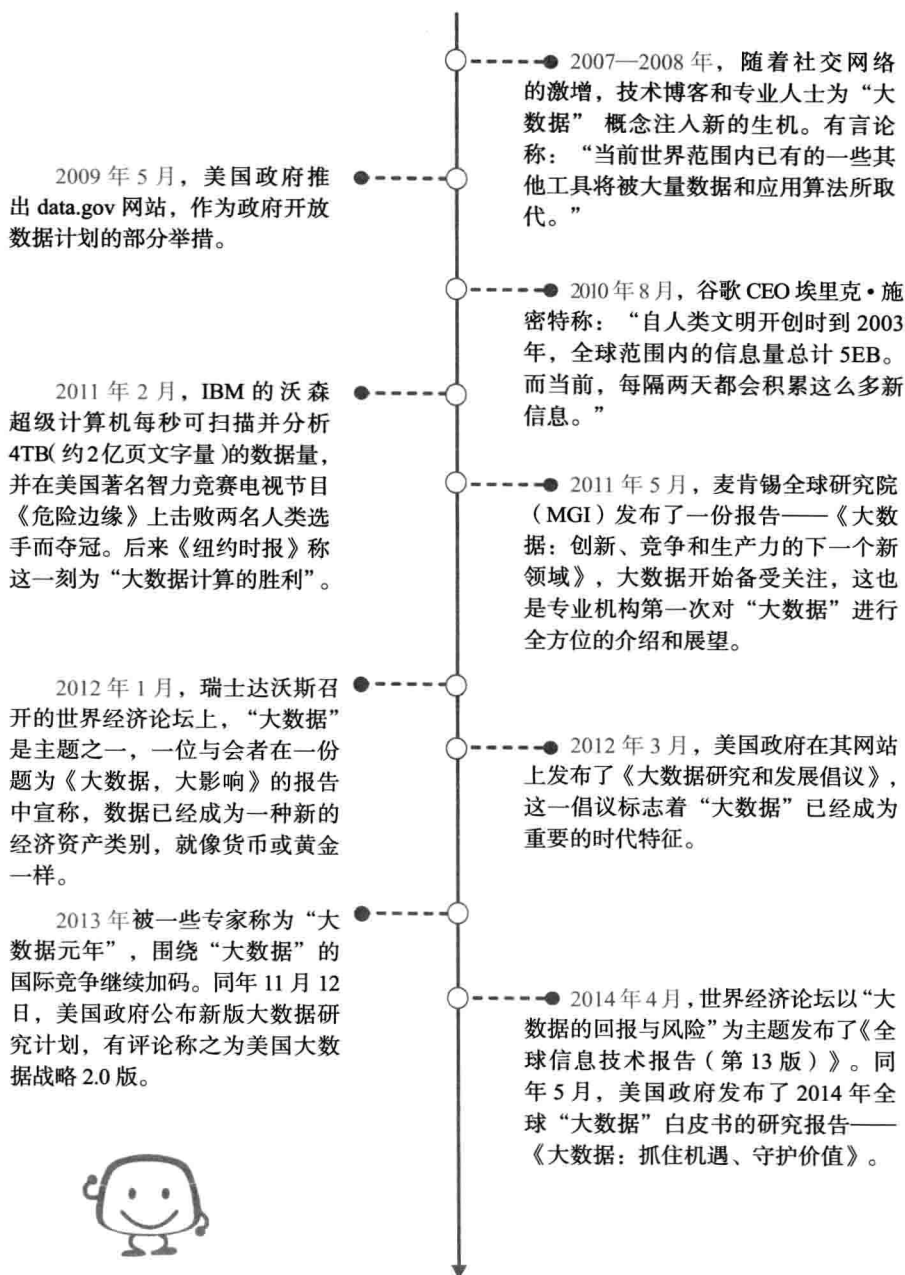
大数据的“前世今生”

处理不断增长的数据是一项巨大的挑战。大数据是在数据爆发式增长、社会化趋势下产生的，各行各业对大规模信息的处理需求从根本上推动了大数据相关技术的发展，其发展历程最早可追溯到 19 世纪末。

大规模的数据信息价值的挖掘主要源于国家安全与军事方面的需求，例如，第二次世界大战期间，英国的科学家研发了能进行大规模数据处理的机器；“9·11 事件”后，美国政府组建了用于筛选通信、犯罪、教育、金融、医疗和旅行等记录来识别可疑人员的大数据库等。随着大数据存储和处理技术的革新，以及大数据分析算法的研发，大数据在教育、金融、医疗等多个领域有着广泛应用。

早在 1980 年，著名未来学家阿尔文·托夫勒便在《第三次浪潮》一书中，将大数据热情地赞颂为“第三次浪潮的华彩乐章”。2008 年 9 月，《自然》杂志推出了名为“大数据”的封面专栏。但 2009 年以后，“大数据”这一术语才成为互联网信息技术行业的流行词汇。





到底什么是大数据，目前并没有统一的定义。康奈尔大学的计算机科学家乔恩·克莱因伯格认为：“大数据这个术语本身很模糊，但在某些方面却又非常具体。大数据是可能改造一切的一种过程的标签。”

关于“大数据”的几种不同定义

麦肯锡公司的定义

大小超出常规的数据库工具获取、存储、管理和分析能力的数据集。

McKinsey&Company

麦肯锡公司同时强调：
并不是说一定要超过特定
TB 值的数据集才能算是大
数据。

国际数据公司的定义

海量的数据规模（Volume）、快速的数据
流转和动态的数据体系（Velocity）、多
样的数据类型（Variety）、巨大的数据价值
（Value）。



也就是所谓的大数
据的“4V”特征。

维基百科的定义

所涉及的数据量规模巨大到无法通过人
工，在合理时间内达到截取、管理、处理并
整理成为人类所能解读的信息。

WIKIPEDIA
The Free Encyclopedia

或称巨量数据、海量数
据、大资料。