

一本帮您快速了解大数据来龙去脉的书

BIG DATA TRAVEL

跟随大数据 旅行

杨池然 编著

大数据的前世今生，无处不在的大数据
大数据下的商业智能与平台架构
理解大数据分析，大数据的价值，大数据未来展望



机械工业出版社
CHINA MACHINE PRESS



跟随大数据旅行

杨池然 编著



机械工业出版社

图书在版编目(CIP)数据

跟随大数据旅行 / 杨池然编著. —北京: 机械工业出版社, 2013.10

ISBN 978-7-111-44579-1

I. ①随… II. ①杨… III. ①数据处理—研究 IV. ①TP274

中国版本图书馆 CIP 数据核字 (2013) 第 253666 号

机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码 100037)

策划编辑: 丁 诚

责任编辑: 丁 诚 孙晶晶

责任印制: 杨 曦

北京双青印刷厂印刷

2014 年 1 月第 1 版 · 第 1 次印刷

148mm × 210mm · 5.875 印张 · 81 千字

0 001 — 4 000 册

标准书号: ISBN 978-7-111-44579-1

定价: 29.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

电话服务

网络服务

社服务中心: (010) 88361066

教材网: <http://www.cmpedu.com>

销售一部: (010) 68326294

机工官网: <http://www.cmpbook.com>

销售二部: (010) 88379649

机工官博: <http://weibo.com/cmp1952>

读者购书热线: (010) 88379203

封面无防伪标均为盗版

这是一本短小而精悍的书，不需要花费太多时间就能够让读者弄清楚大数据到底是什么，还能帮助读者了解大数据的来龙去脉以及未来大数据对各行各业带来的影响与作用。大数据将带来新的科技革命浪潮并推动管理变革、IT 科技变革与业务变革、生态链变革以及分析变革。

本书第 1 章讲述大数据的来龙去脉。第 2 章从应用的角度讲述大数据实际的业界应用。第 3 章从大数据的技术角度讲述大数据对当今和未来技术的影响。第 4 章从数据分析的角度讲述大数据分析，而大数据分析是大数据的核心，内容从数据到信息，从分析流程到算法应用、数据挖掘、模型应用、可视化展现以及分析团队建设。第 5 章从大数据的价值讲述大数据的商业价值、分析价值、安全价值以及未来价值。第 6 章讲述大数据的未来，对大数据未来给予一个前瞻性预测，如未来 IT 科技、未来管理变革等。

本书适合 IT 科技人员、管理人员、数据分析人员、数据挖掘人员、统计分析人员和决策者以及想了解大数据的普通读者阅读。

序 言

未来世界什么最重要，数据！大数据技术将引领未来科技。大数据的出现正如互联网的出现一样，将改变人们生活的方方面面。

本书通过对大数据从诞生到实际应用的详细讲解，真正让读者从朦胧的状态过渡到理解大数据的实质到底是什么？理论与实践相结合的讲解，让读者真正体会到大数据的价值，客观而实际，摒弃虚无缥缈。

本书将从以下几个方面展开讲解：

◆ 第1章详细论述了什么是大数据以及大数据诞生的过程与未来发展方向和实际应用前景。

◆ 第2章详细讲解了大数据在各行各业的应用，针对社交网站上的大数据、电子商务中的大数据、医疗行业的大数据、移动互联网上的大数据、金融行业的大数据和物联网上的大数据进行详细的案例式讲解。

◆ 第3章从大数据商业智能与平台架构方面讲解大数据在业界的应用，如商业智能、MPP架构、云平台以及多平台共存等。

◆ 第4章讲解大数据在分析领域的应用价值，如大数据分析工具，大数据挖掘、模型，大数据可视化展现与可视化分析等。

◆ 第5章讲解大数据的实用价值，如商业价值、分析价值、安全价值以及未来价值。

◆ 第6章以未来展望的形式讲解大数据的未来，如未来智能科技、未来管理变革、未来商业智能以及未来中国大数据。

本书从大数据普及的视角讲解，无论是普通读者还是科技读者，都能从生动的案例中领略大数据的内涵。

感谢同事和家人在本书的编写过程中的支持。由于作者水平有限，且时间仓促，书中难免存在不妥与疏漏之处，欢迎广大读者批评指正。

杨池然

目 录

序言

第 1 章 大数据的前世今生	1
1.1 何谓大数据	1
1.1.1 关于大数据的三个疑问	4
1.1.2 谁在影响你	11
1.1.3 大数据——影响着谁	13
1.2 横空出世的大数据	16
1.3 大数据将走向何方	18
1.3.1 发现未来——大数据	18
1.3.2 各行各业——大数据应用	22
第 2 章 无处不在的大数据	26
2.1 SNS 社交网站中的大数据	27
2.2 电子商务中的大数据	30
2.2.1 购物行为与销量预测分析	31
2.2.2 商品关联分析	34

2.3	医疗行业中的大数据	36
2.4	移动互联网上的大数据	40
2.5	金融行业中的大数据	45
2.5.1	信用评分	45
2.5.2	客户价值分群	49
2.5.3	客户挽留	51
2.5.4	客户精准营销	56
2.6	物联网上的大数据	58
第3章	大数据下的商业智能与平台架构	62
3.1	商业智能——在大数据 时代下的新发展	63
3.2	MPP 架构——在大数据 时代下的新架构	73
3.3	云平台——大数据与云平台的关系	76
3.4	其他架构——在大数据 时代下的其他创新	82
3.5	多平台共存——在大数据 时代下的必然选择	89
第4章	理解大数据分析	95
4.1	数据的价值转换过程	95

4.2	数据分析流程	98
4.3	数据分析人员的要求	105
4.4	数据分析的团队建设——大数据 分析部门组建	108
4.5	数据分析工具——大数据 分析的利器	116
4.6	算法应用——大数据分析 时代的基石	119
4.7	数据挖掘、模型与相关分析——大数据 淘金	124
4.8	可视化分析——大数据分析的 最高层级	140
4.9	报表与数据分析的关系——大数据 分析的深层关系	144
4.10	带点怀疑精神——不要完全相信 数据分析	147
第 5 章	大数据的价值	150
5.1	商业价值	152
5.2	分析价值	160

5.3 安全价值	162
5.4 未来价值	164
第 6 章 大数据未来展望	166
6.1 未来智能科技	166
6.2 未来管理变革	171
6.3 未来商业智能	174
6.4 未来中国大数据	176

第 1 章 大数据的前世今生

1.1 | 何谓大数据

未来世界什么最重要？“数据！”大数据时代汹涌来袭，就像当年互联网的出现一样，大数据技术将加速企业创新，引领社会变革，促进行政管理模式和商业模式的变革，推动政府执政能力的提升，并创造出无限商机，而数据将成为未来社会发展必需的“石油”和“金矿”。

当前，由互联网、移动互联网、物联网等行业的发展所产生的海量数据把我们淹没在数据的海洋之中。据 ICT research 研究预测，2013 年年底我国的数据中心基础设施产品市场的整体规模将达 157.85 亿元，未来每年将保持 6.5%~9.2% 的增长。大数据可视化分析、数据模型应用、数据安全和数据存储服务等将作为未来重要的发展方向。数据科学正成为一个

新兴的学科和领域，数据的重要性在我们的生活中日益重要。无论是个人、企业还是国家，谁能更好地抓住数据、理解数据、分析数据、应用数据，谁就能在未来的竞争中占据先机。可以肯定的是，在可以预见的未来，大数据将驱动科技、管理和商业模式的大变革，并由此创造出无限的商机，为人类社会带来“大知识”“大科技”“大商机”和“大发展”。

看了上边这些话，有些读者恐怕还是无法直观地认识到大数据的重要性，总是感觉大数据这样的尖端技术离我们的工作、生活还很远，是虚无缥缈的空中楼阁。事实上，眼下大数据技术已经渗透进我们社会生活中的方方面面，正在通过各种各样的方式影响着我们每一个人，而且这种影响力正在随着技术的发展和应用与日俱增。下面，先通过几个例子，帮助读者对大数据有一个初步的直观感受。

- 2008 年，阿里巴巴公司利用大数据技术通过查询用户的点击次数和购买次数，精准地预测到了 2008 年的全球金融危机，从而使自己的企业提前作好了防御金融危机的准备。

- 金融行业中的信用卡业务已经广泛利用大数

据分析技术进行风险预警、用户行为分析、客户挽留等商业行为分析。

- 谷歌公司曾经运用大数据技术分析流感人群，并跟踪预测到了某个地区流感的爆发，为流感防控提前作出了预警。

- 国家统计局通过对人口数据的分析得出中国人口老龄化的时代已经到来。

- 京东商城正在通过收集到的销售数据以及客户购买行为来向客户推销相关产品。

再讲一个更贴近我们生活的真实故事。

有一天，美国零售企业好市多（Costco）接到了一位母亲的投诉：“我的女儿才16岁，你们却向她推销孕妇类的产品。”但是，这一次投诉尴尬的却是这位母亲自己，因为她16岁的女儿确实怀孕了。为什么卖货的商场，却能比与女儿朝夕相处的母亲，更早地知道这件事？恰恰是因为这个女孩发现自己怀孕之后，不敢告诉母亲，只能时常在网上浏览一些和怀孕相关的网站，搜寻一些孕妇类的产品。这些浏览行为被当做数据记录下来后，好市多的数据分析团队，据此进行精确的商品促销定位，利用海量的数据进行精

准营销分析，找到有价值的客户并推销商品。这就是运用大数据技术的一个典型案例。

看完上面几个案例，读者应该能真实地感受到大数据就在我们身边。它将直接或间接地影响我们每一个人的工作和生活，它将改变我们传统的商业模式、消费行为与沟通方式，开启 IT 科技的新纪元。总而言之，生活在信息化的大数据时代，每秒钟都将发生新的故事。

1.1.1 关于大数据的三个疑问

在详细介绍大数据技术之前，需要先弄清楚几个读者最关心的同时也是最基本的问题。

1. 大数据到底是什么

伴随着互联网、物联网、移动互联网、电子商务、社交网等网络信息化的推广和应用，各行各业在几十年的信息化过程中，积累了大量的历史数据，而这些新的数据还在不断增长。与此同时，数据的形态也发生了巨大的变化，以往的格式化数据正在过渡到格式化和非格式化数据共存的时代（如

视频、图像、文本数据)。各类数据源每天都在生成大量的数据并传输到云计算平台,以供前端数据挖掘、模型建设、数据分析以及数据可视化展现等使用,为各业务层提供服务,真正地实现了让数据说话。大数据与其他行业之间的关系如图1-1所示。

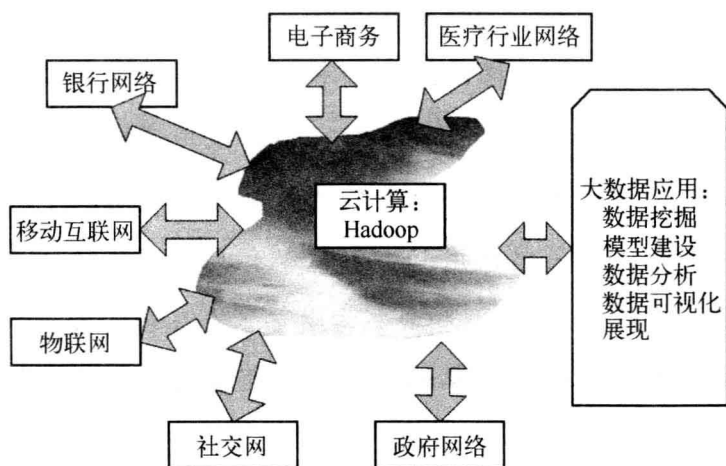


图 1-1 大数据与其他行业之间的关系

● 云计算。云计算简称为“云”，是一种通过互联网按需交付计算资源（从应用到数据中心都属于计算资源）和按使用付费的基础架构。打个比方，我们目前所用的电就是按使用的电量收费，如果不用电，也就不需要付费。云计算就相当于国家电

网，属于供应商，向用户提供电，用户用多少就收多少费用，不用资源就收回，可以供其他人使用。云计算分为公共云（公共云由一些公司运营和拥有，这些公司使用这种云为其他组织和个人提供服务，按合理的价格计算资源并收费。使用公共云的好处，在于用户无需购买硬件、软件或支持基础架构，这些都是由提供商拥有并管理的）、私有云（私有云由单个公司运营和拥有，资源使用不对外）和混合云（混合云以私有云作为基础，同时结合了公共云服务的策略使用。事实上私有云是不会独立于公司其他的 IT 资源和公共云而单独存在的。大多数使用私有云的公司都将发展为管理跨数据中的工作负载、私有云和公共云创建了混合云）。云服务商通过云计算平台提供基础设施即服务（IaaS）、应用平台即服务（PaaS）、软件即服务（SaaS）。

● 物联网、移动互联网、银行网络、电子商务、医疗行业网络、社交网以及政府网络属于应用层，可以利用云计算实现不同的服务需求。

● 大数据应用层：对于物联网、移动互联网、银行网络、电子商务、医疗行业网络、社交网以及政

府网络等应用层产生的数据，通过大数据应用层进行数据的分析，如数据挖掘、模型建设、数据分析、数据可视化展现等。

● 云计算与大数据的关系。云计算与大数据之间的关系正如一枚硬币的正反两面，有了云计算提供软件和硬件相结合的计算平台，大数据才能有用武之地，才能满足大数据需要的计算与分析需求。可以说，有了云计算才能玩转大数据，有了大数据所以需要云计算。

回到前面的问题——大数据是什么？大数据就是指难以用传统的软件工具存储、管理和分析的大容量数据。而大数据技术则是指通过对大数据的提取、交互、整合和分析，为政府、企业或其他机构的管理层提供决策信息，发现隐藏在数据背后的信息，挖掘数据信息的价值。

大数据技术具有四项典型特征：海量数据、数据的多样性、交互响应的高速性和数据的易变性。

（1）海量数据

海量数据是指数据量规模增长到 PB 级以上，而不仅是几个 TB 数据量。