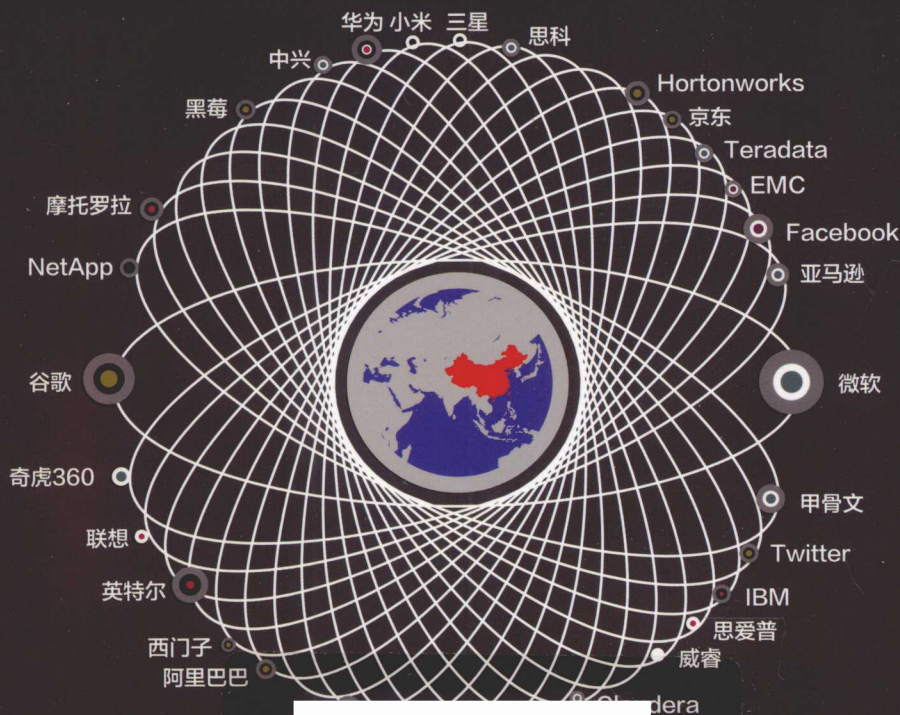


大数据席卷中国，一场你无法置身事外的革命！

大数据在中国

赵伟 著



江苏文艺出版社
JIANGSU LITERATURE AND ART
PUBLISHING HOUSE

大数据在中国

BIG DATA IN CHINA



江苏文艺出版社
JIANGSU LITERATURE AND ART
PUBLISHING HOUSE

图书在版编目 (CIP) 数据

大数据在中国 / 赵伟著. — 南京 : 江苏文艺出版社, 2014

ISBN 978-7-5399-7375-3

I. ①大… II. ①赵… III. ①互联网络-应用-企业管理-研究-中国 IV. ①F279.23-39

中国版本图书馆CIP数据核字(2014)第088379号

书 名	大数据在中国
著 者	赵 伟
责任编辑	孙金荣
策划编辑	一 航
文字校对	郭慧红
封面设计	罗久才
出版发行	凤凰出版传媒股份有限公司 江苏文艺出版社
出版社地址	南京市中央路165号, 邮编: 210009
出版社网址	http://www.jswenyi.com
经 销	凤凰出版传媒股份有限公司
印 刷	北京兆成印刷有限责任公司
开 本	700毫米×1000毫米 1/16
印 张	19.5
字 数	209千字
版 次	2014年6月第1版 2014年6月第1次印刷
标准书号	ISBN 978-7-5399-7375-3
定 价	35.00元

(江苏文艺版图书凡印刷、装订错误可随时向承印厂调换)

导读
INTRODUCTION 大数据正在改变中国

大数据给了中国企业与个人哪些新机遇？一场关于思维、工作与生活的自我变革！本书为《给你一个团队，你能怎么管？》作者的突破之作！

这也许是迄今为止最易懂、最实用的大数据类图书！因为除了本书，再没有另外一本书能让你如此接近中国大数据时代的现在与未来。

全球知名咨询公司麦肯锡最先提出“大数据”概念。随后大数据一词越来越多地被提及，它上过《纽约时报》《华尔街日报》的专栏头条，进入过美国白宫官网的新闻。比尔·盖茨、巴菲特、谷歌公司创始人拉里·佩奇、甲骨文公司创始人拉里·埃利森等商业领袖，都第一时间开始关注大数据。

大数据有三大特点：第一是大，海量的数据；第二是快，可以通过公共数据库快速地获取；第三是我们不再热衷于寻找因果关系，而更加关注于相关关系。

大数据决定着企业的未来发展，正如《纽约时报》2012年2月的一篇专栏文章所称，大数据时代已经来临，在商业、经济及其他领域中，决策

将日益基于数据和分析作出，而并非基于经验和直觉。

哈佛大学社会学教授加里·金说：“这是一场革命，庞大的数据资源使得各个领域开始了量化进程，无论学术界、商界还是政府，所有领域都将开始这种进程。”

大数据最核心的作用是可以“预见未来”。华尔街用它准确预测出股票的涨跌，投资机构搜集并分析上市企业声明，从中寻找其破产的蛛丝马迹；美国疾病控制和预防中心依据它，分析全球范围内流感等疫病的传播状况；美国总统奥巴马通过它获取更多的政治选票。

如今，大数据已经到了中国，它将改变中国，改变一切传统企业，改变整个市场的格局。我们的工作、生活、社交都将与它息息相关。从移动、电信、阿里巴巴、微博、微信、百度到互联网的每一个角落，整个互联网就相当于一个大数据库。

了解大数据，学会其中的原理与运用的技巧，也许一个百度搜索热词就能给你一个商机，你的想法创意很快就能通过各种公共数据库得到验证。退一步来说，就算你开一个网店，起点也比他人要高出许多。高端论坛中经常会提到的“互联网思维”如果离开大数据的支持，就会如无源之水，无本之木。

本书用最精简的文字讲述、分析了大数据的特点、原理以及在当下中国的各个行业和领域的运用。同时也给我们的企业与个人提供了新的参考：大数据时代将对中国企业转型提供哪些支持？我们的优势和劣势在哪里？我们该如何搭上大数据的快车，实现商业价值与个人理想呢？

如果说您已经通过许多同类书籍知道了大数据是什么，那么本书则更注重为您解惑和提供实用指导：面对大数据，我们应该怎么做？书中不仅展示了谷歌、微软、亚马逊、Facebook、Twitter 等大数据先锋公司最具代表意义的应用案例，也展示了阿里巴巴、腾讯、百度、奇虎 360、小米等首批搭上大数据快车的企业的中国式突围。

这本书要解决的第一个问题是，我们应该如何理解“大数据”？

大数据的英文表达是 Big Data，意思是“海量数据”。数据的规模大到了已经无法用当前的技术和工具来处理，那就必须突破瓶颈，从而产生数据革命。对数据的处理包括很多方面，有收集、整理、分类、存储、分析、预测和输送等等。

大数据就这么简单吗？当然不是。数据如同人体的血液，大数据则是整个人体系统与血液有关的部分。最早涉及这个概念的是天文学和基因学领域，因为这两个学科非常依赖对数据的分析方法，尤其是对“海量数据”的分析。它也是电脑和互联网结合的产物，因为电脑实现了数据的“数字化”，让它们像数字一样容易储存，互联网则实现了数据的“网络化”，让它们通过网络可以自由快速地传输。

从此，大数据才真正拥有了无穷的生命力。互联网的技术不断发展，渗透进我们的工作和生活，加上移动网络、物联网与其他各种联网设备的

出现与普及，一个必然产生的现象就是数据的迅速增长。有90%的数据是互联网出现以后才产生的，它以指数级的速度在我们的生活中不断增加，从海量至于无穷大，世界正被数据淹没。

我们需要更加关注的，是数据从量变开始质变，并且体现在多个方面，触发蝴蝶效应，推动其他领域的变化。

第一，催发数据性思维。

一种全新的思维模式，这首先是大数据时代的到来带给我们的。思维方式的改变，推动产业的根本性变革。这种变革甚至是颠覆性的，因为思维是最大的生产力，也是社会文明进步的决定性力量。

历来的商业变革，其发端都不是某种技术，而是思维方式的转变。思维产生需求，需求推动技术。人们有了新的思考，有了新的视野，才开始低头审视旧的经济体制和传统的商业理念，去思索和创新商业逻辑，研究更实用、先进和高效的模式。反之，如果人们不能与时俱进，吸收并且创造顺应潮流的新思维，再通过新思维去重新组织资源，建立架构和制定策略，那么貌似强大的体量反而会变成继续前进的累赘——就像19世纪的清帝国。

这种通过新思维来颠覆旧秩序的案例，不仅发生在国家的竞争之间，还在信息技术领域频繁上演。它总是最先由信息与技术的新思维起步，然后渗透到传统领域，继而蔓延到全社会，改变每一个行业和每一个人。

相关案例到处都是，比如黑莓、摩托罗拉、诺基亚、柯达、雅虎等公司，比如华为、联想、中兴、百度、360、中国移动以及它们的竞争对手——谷歌、思科、微软、三星、西门子等。

数据性思维表现在：

1. 对全部数据进行分析，而不是随机抽样；
2. 并不过于追求精确性，而是重视数据的复杂性；
3. 更多挖掘数据的相关性，而不是因果关系。

虽然对于后两条在不同的领域存在争议，在不同的需求面前也存在侧重点的区别，但就数据性思维而言，这三条特征足以让我们的世界天翻地覆。当然，许多超级巨头的没落也并不是因为它们没有数据性思维——原因总是很复杂，但它们归根结底总是倒在自己的思维上。它们被新的思维超越，并且彻底击倒；如果不奋起直追，就会被淘汰，退出历史舞台。

昔日巨人在感慨自己未及早行一步，追赶者在叹息何处寻觅超级思维人才。而在今天，数据性思维就是最新的智慧，虽然它的威力还没有进化到可以导致巨头轰然倒下的地步。不过，需要警告的是，如果你今天不给予它足够的重视，在下一拨没落帝国的名单中，你的名字将会赫然在列，成为失败者中的一员。哪怕你现在风光无限，在被超越和打败的时候，也会无能为力。

第二，产出“数据资产”。

思维充分展示了它的力量，而在大数据时代，另一个变化是数据成为一种资产并被产出。这是因为，当我们需要更加全面的数据来提高预测分析精度时，就必须制造更多的便宜、方便和自动的数据工具，收集、加工或生产数据。就像你住的房间越大，需要的佣人就越多，他们的工作量就越大，而你的需求是佣人最好不犯错误，干活越快越好，越多越好，给你提供完美的生活。

这些工具包括什么呢？上网时我们使用的浏览器、软件、窗口和后门工具，它们会记下你的各种个人信息数据；生活中我们使用的智能手机、GPS 导航器、智能手表或其他数码产品，时刻都在产生数据；还有路由器、智能玩具甚至电视机、冰箱等，如果它们联入网络，在为你服务的同时，也在产出无数的数据。

当你和女朋友出去购物时，你需要刷卡、移动上网，那么超市收银的刷卡器、路由器、无线 WLAN 和 3G 网络、银行 ATM 以及监控摄像头，都在收集和产生与你有关的数据。

除非你躺在家里什么都不干——扔掉手机、电脑、固定电话和所有的银行卡，也不出门购物，变成彻头彻尾的隐居者，从这个世界消失，否则你时刻都是一个产生数据的源头，并被接入一个无限大的数据互联网。全世界几十亿人共同组成了一个庞大的数据生产工厂，提供了几乎无法计量的数据资产。

为什么我说这是一种资产呢？因为数据在互联网领域意味着流量，而流量就代表着财富。流量拿来变现，可以是广告，可以是游戏，也可以应用于电子商务。拥有这些数据的人或企业，就能由此获得巨大的利益。这时，商机就产生了。

因此，在大数据时代开启之前，意识到数据的资产价值的公司就已开始布局。它们早就盯上了数据生产的源头——消费者，通过出售数据工具来把人们接入互联网，获得人们的数据流量。

第三，数据资产可以变现。

虽然在事实上，任何资产都是可以变现的，但没有哪种资产的价值可

以超过“数据”。我们拥有了数据资产，就必须通过分析来挖掘它有多少价值，通过预测来确定价值的方向，通过整理来划分价值的类型，最后再“变现”为用户价值、股东价值乃至全社会的价值。

在这个过程中，核心环节就是预测——这也是大数据分析的目的。数据是海量的，也是复杂的；数据无孔不入，无所不至，怎样采取实用技术来进行划分？如何通过数学建模来从数据中整理价值走向，建立价值链？最关键的是，我们要预测事态发生的可能性，然后制订相应计划，采取合理措施。

这方面的应用十分广泛，像流行病、股价、经济走势的预测。当预测可行并成为一种机制时，人们就有了更进一步的可能——通过适当的干预，来引导事情向着期望的方向发展。

例如，针对用户的个人喜好和对消费能力的统计，电商网站向他们推荐不同的商品，引导他们多消费。游戏公司会在游戏运行过程中对玩家的行为数据进行分析，及时调整针对不同玩家的计费点和关卡难度，也对玩家的消费欲望和能力进行分析，最大化地从玩家身上获得收入。现在所有的网络公司都已拥有了“精准营销”的能力，通过恰当的技术手段向不同的用户展现不同的广告（个性化广告），以此来提高广告收入，也帮助自己的广告用户提高产品销量。

这些例子都很好地证明了数据资产的价值，以及价值的可塑性。它也是一个巨大而完善的产业链，它的分析、预测和控制技术，亦广泛地应用于其他行业，对我们个体的生活、工作和思维的提高都有不可估量的帮助。

本书抛出的第二个问题是：大数据非常热，但如何才能把它变成一个广泛实用的利器而不是阶段性的泡沫呢？

我们都有这样的感觉：大数据正在逼近和包围我们，因为现在人人都 在讨论，企业在探讨，政府也在研究，新闻铺天盖地，理论遍地开花。数据也在不停地产生，个人、企业、政府机构、互联网，都在产生数据。各式各样的数据正在冲击我们，颇有“乱花渐欲迷人眼”和“身在庐山”的感觉，难免感到迷茫与困惑。

我们究竟应该从大数据中挖掘到什么？

企业和组织怎样通过大数据完成蜕变和转型？

大数据技术应该怎样实用化、有效化，让它适用于每一个人？

这些是人们普遍关注的问题，要想实现这些具体的目标，我们必须先破除障碍。障碍既是技术上的，又是心态和认知上的。技术障碍最容易攻克，心态和认识上的障碍往往很难在短时间内消除。

从国内这几年大数据发展的现状来看，人们从数据中已经发现了新的价值，在技术层面也已经做足了准备，但之所以大数据市场还未能大规模地启动（除少数几家公司外），原因在于心态、认知的滞后和旧观念的羁绊。

上海一家IT公司的经理朱先生拿他亲身经历的一件事举例。不久前，朱先生开车在路上行驶，被一辆摩托车刮蹭了，由于肇事者拒不认账，朱先生只好请求有关的交管机构调取马路上的监控视频，希望对这场事故责任有准确的认定。这当然是一个好的方法，但他得到的答复是：“你的请求是可以的，不过需要将你的汽车押在交管部门15天，相关数据也要在15

天后才能调取出来。”

朱先生很苦恼：“明明一件很容易、很迅速就可办到的事，为何弄成这样？”

这就是观念或心态的问题。我们知道，即便在中国一个较小的县城，马路上和关键的区域也都装上了摄像头；在大的城市比如北、上、广、深，传感器早已经星罗棋布，到处都是。收集和记录视频是一件信手拈来的易事，这些数据也具备了为社会、为民众提供更多优质服务的可行性。同时，现在的大数据技术也已经完全可以高效处理和分析这些数据。可事实往往相反，人们依然无法从中获益，有的机构最在乎的是自己拥有多少权限，而不是提供多少服务。当然，相关的机构也对此有怨言，他们也有自己的理由和无奈之处。原因就在于人们服务意识的缺乏，以及旧的思维和心态的顽固影响，这便构成了一个牢固的阻力场，让每个人深陷其中，都想出来，但又都拔不动腿。

大数据不仅应用空间广泛，也有巨大的需求空间。不论在医疗、交通、航空、电力、电信、金融、城市管理等公共服务领域，还是在制造业、创新行业等各式各样的企业当中，都存在从数据中发现新价值，提供更个性化、更有价值的服务的需求。也就是说，我们已经完全准备好了技术，但还没有准备好观念和意识。现在中国正大力推进智能电网建设，虽然智能电表安装很快，但后端的数据整合远远跟不上，导致用户虽已装上了智能电表，但得不到应有的服务。比如当电表只剩下最后几十度电时，都无法获得短信提醒服务。这和朱先生的遭遇如出一辙，不是技术上做不到，而

是观念需要改变。

大数据方兴未艾，正在源源不断地释放出更多的能量，推动我们的产业升级、社会转型和企业转型。我们要想从中获益更多，就需要全民参与、全民破冰、全民转变观念。只有这样，数据才会成为我们的“新能源”，而非“烫手山芋”。只有首先破除这些障碍，大数据在我们的生活中才真正具有无限广阔的应用空间，成为推动社会发展的核心动力。

要想成功破冰，更新旧的观念，我们必须从内部数据和内部问题切入。只空喊口号、形而上地研究似是而非的问题是起不到多少实际作用的。在不同目标、不同状态下的企业、部门和个人，其实都可以开启自己的大数据之旅，从大数据中挖掘到更多价值，但一定要脚踏实地，拒绝空谈，从实际出发，从小处做起，从细节开始，打下坚实的基础。

只有从现有的数据入手、以拥抱客户为核心任务、以解决内部问题为基本的出发点，才能切实地推动我们的大数据发展，成功开启破冰之旅。比如公交系统，在整合了各种公交平台的信息体系后，将公交车、地铁和出租车的出行进行“三位一体”的融合，人们就可以用手机实时地查询正在等待的公交车还有几分钟到达，确认实时的路况和交通拥堵信息，灵活选择自己的出行工具和路线。

当前，大数据对于企业与机构的显著应用效应主要体现在两个维度：一是在前端发现更多的商机，对客户实现全方位的关照，挖掘和拓展更大的增值空间；二是管控风险、提高效率和降低成本。

有一位在电信公司任职的项目经理对我说：“现在电信之间的竞争日趋

白热化，市场也将饱和，怎样快速地掌控市场和用户的动向，制定个性化和针对性的营销方案，就成为在激烈的竞争中获胜的关键。我们在采用新的大数据方案之后，直接提升了服务速度和个性化的服务效率，数据中心的反馈时间从原来的两周缩短到现在的不到两天。这样一来，我们的客户经理就能够实时获取大量信息，更有效地分析客户的使用习惯，占领服务客户需求的先机。”

这就是在解决问题的基础上成长，在挖掘数据价值的前提下破除障碍。对大数据在现实中的具体应用，尤其对企业来说，首先必须以客户为中心，制定战略规划和自己的大数据蓝图。其次是根据自己的业务优先级，建立数据分析体系，逐步提升大数据的分析与应用能力。最后就是定制可量化的指标，获得投资回报率。

从目前的大数据实践来看，我们的应用主要体现在对现有数据的整合、挖掘，进而去提升客户端的应用价值和带动企业的效率。如果想让数据革命真正变成席卷全球的第三次工业革命，产生新的生产力，还有很大的创新空间。

对中国而言，一个伟大的时代才刚刚开始。

赵伟

2014年元月10日 北京

大数据在中国
BIG DATA IN CHINA

目 录

导读 大数据正在改变中国 / 001

序 大数据在中国 / 004

001

CHAPTER 1

大数据,你还不知道的部分

◎FB 数据单元——信息导航图 / 002

◎核心:整理、分析、预测、控制 / 006

◎大数据先行者 / 015

◎谨慎:不是所有人都需要 / 021

◎棱镜门背后的大数据革命 / 024

037

CHAPTER 2

大数据时代给世界的巨大转型机会

◎我们能做什么,我们要做什么 / 038

◎反馈经济——新的营销模式在兴起 / 050

◎数据大爆炸——信息过量 / 055

◎全球产业链正面临大调整 / 060

065 CHAPTER 3 中国如何搭上大数据快车？

- ◎中国的大数据现状 / 066
- ◎追赶者的中国——走到至关重要的十字路口 / 080
- ◎消除“数据割据”与“数据孤岛” / 087
- ◎工信部的规划：四项关键技术创新工程 / 090
- ◎建立大数据政府 / 092

097 CHAPTER 4 中国首批重视大数据的千亿公司

- ◎原动力：对信息共享的需求 / 098
- ◎云计算和大数据 / 100
- ◎在中国的发展 / 110
- ◎阿里巴巴：云帝国构想 / 118
- ◎腾讯：大社交战略 / 123
- ◎360：最大数据中心 / 127
- ◎百度：大数据时代的三层布局 / 134

139 CHAPTER 5 大数据与技术变革

- ◎告别小数据时代 / 140
- ◎数据服务产业链 / 146
- ◎技术支持与发展 / 151
- ◎“脚印追踪”——个性化的数据推荐系统 / 156