

一本书搞懂 企业大数据

ENTERPRISE
BIG DATA



董超◎编著

李克强总理：

大数据等新一代互联网技术深刻改变了世界，
也让各国站在科技革命的同一起跑线上

阿里巴巴集团创始人马云：

世界正在快速改变，很多人还不知道IT是什么，
今天IT已经在向DT（数字科技）时代快速跨越

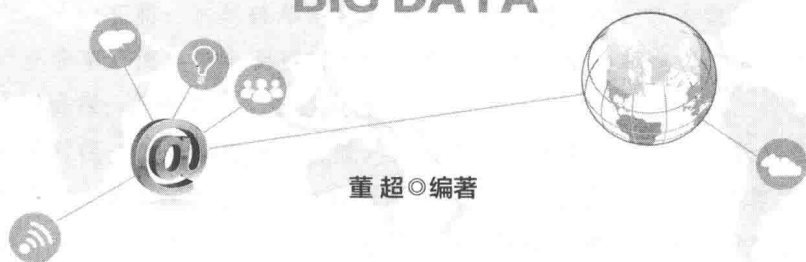
“十三五”规划指出：“实施国家大数据战略，推进数据资源开放共享。”



化学工业出版社

一本书搞懂 企业大数据

ENTERPRISE
BIG DATA



董超◎编著



化学工业出版社

·北京·

《一本书搞懂企业大数据》内容涵盖：认识大数据、企业迈入大数据时代、企业大数据的技术架构、企业大数据的运营、金融业与大数据、工业与大数据、零售业与大数据、大数据面临的问题及应对等，供读者参考。

本书有助于数据分析人员、企业管理者、战略决策者、技术供应商等，精准、快速地实施大规模数据分析，从而实现更多价值。

图书在版编目(CIP)数据

一本书搞懂企业大数据 / 董超编著. —北京：化学工业出版社，2017.1

ISBN 978-7-122-28523-2

I. ①一… II. ①董… III. ①企业管理-数据管理 IV. ①F272.7

中国版本图书馆CIP数据核字(2016)第277062号

责任编辑：陈 蕾 刘 丹
责任校对：宋 玮

装帧设计：尹琳琳

出版发行：化学工业出版社（北京市东城区青年湖南街13号 邮政编码100011）

印 装：三河市延风印装有限公司

710mm×1000mm 1/16 印张11½ 字数207千字 2017年1月北京第1版第1次印刷

购书咨询：010-64518888（传真：010-64519686）

售后服务：010-64518899

网 址：<http://www.cip.com.cn>

凡购买本书，如有缺损质量问题，本社销售中心负责调换。

定 价：48.00元

版权所有 违者必究

前言

PREFACE



“大数据”成为近几年的热词，大数据已经渗透到了全世界市场的各个领域，并逐渐成为重要的组成因素。对海量数据的运用将提升人们对市场以及未来走向的预知和把控。“让数据产生价值”，随着企业对大数据越来越重视，这个声音也正在变成企业的口号。

当前全球正处在新一轮科技革命和产业变革的关键时期，以大数据为代表的信息技术产业变革，使各个企业在信息技术发展上处于新的起跑线，如何充分挖掘利用大数据资源，是企业当前面临的重大课题。

在互联网时代，数据本身就是资产，而大数据则意味着这些资产正在变得庞大无比，在现实生活中，善于运用大数据的企业已经从数据中获得了丰厚的回报。

对一个企业来说，不断地为产品扩展销售渠道和提升服务质量是至关重要的。在进行大量数据分析的前提下，更好的分析数据，获取用户信息才能够更精准的获取市场动态，并且掌握市场对产品和服务的需求，从而优化产品及服务，提升产品的销售量。

大数据的战略意义，一方面在于掌握庞大的数据信息；另一方面则在于大数据与产业的结合，对海量数据进行专业化处理，实现数据的“增值”和应用的“价值”。随着资源整合和产业链拓展，大数据应用正在成为新的经济增长点，逐渐在智能家居、智能制造、智慧出行、智慧医疗、互联网金融等应用中找到用武之地。

2016年3月17日，《中华人民共和国国民经济和社会发展第十三个五年规划纲要》发布。“十三五”规划纲要中指出，实施国家大数据战略。把大数据作为基础性战略资源，全面实施促进大数据发展行动，加快推动数据资源共享开放和开发应用，助力产业转型升级和社会治理创新。

基于此，编者根据多年的经验，通过浅显易懂的方式编写了《一本书搞懂企业大数据》一书，把高深莫测的企业大数据知识普及给读者，供大家参考，以便起到抛砖引玉的作用。

《一本书搞懂企业大数据》主要包括认识大数据、企业迈入大数据时代、企业大数据的技术架构、企业大数据的运营、金融业与大数据、工业与大数据、零售

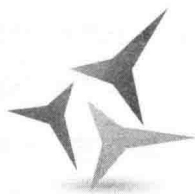
业与大数据、大数据面临的问题及应对等内容。

《一本书搞懂企业大数据》由安德互联数据服务有限公司董事长、高级架构师、大数据分析师董超主编，在编写过程中，深圳市安德互联数据服务有限公司及业内大数据相关公司的专家学者和从业人员参与了本书的资料收集和编写工作，具体参加编写和收集资料的还有卢桂林、傅冬晓、骆相松、王浩鹏、朱祚缤、黄文平、刘诚、罗星宇、史建湘、马静怡、叶坚镇、许泽香、王彦井、杨启江、李维正、郭梅、张杰、智慧、李军、李辉、张海雷、陈超、孙小平、匡仲潇及冯永华，同时特别感谢河海大学陈锴同学做了大量的资料收集和编写工作，阿里巴巴高级技术专家崔亮、小米科技广告部资深研发工程师冀康、58赶集集团商业技术部高级技术经理邓柱中参与了本书的审定工作，最后全书由董超统稿审核完成。在此，对他们一并表示感谢！由于作者水平所限，不足之处敬请读者指正。

编 者

目录

CONTENTS



第一章 认识大数据..... 1

互联网技术的飞速发展和应用，不仅改变了生产和消费的行为及商业模式，同时也带来了信息数据在数量、频度和使用等多方面的巨大变革。从数据角度看，整个世界已跨入“大数据”时代。

第一节 大数据的基本认知..... 3

认知1：大数据的定义..... 3

相关链接 大数据中多种多样的数据单位..... 3

认知2：大数据的类型..... 4

认知3：大数据的特征..... 5

认知4：大数据的作用..... 7

认知5：大数据的商业价值..... 8

第二节 大数据的处理流程..... 11

流程1：大数据的采集..... 11

流程2：大数据的预处理..... 11

流程3：大数据的统计..... 11

流程4：大数据的挖掘..... 12

第三节 大数据的处理技术..... 12

技术1：大数据采集技术..... 13

技术2：大数据预处理技术..... 13

技术3：大数据存储及管理技术..... 14

技术4: 大数据分析及挖掘技术	15
技术5: 大数据展现与应用技术	16
第四节 大数据的发展趋势	17
趋势1: 国家战略推进大数据行动	17
趋势2: 需求导向助力大数据贴近应用	18
趋势3: 将推动智能制造和制造强国建设	19
趋势4: 公众参与推动政务大数据发展应用	20
趋势5: 加速推动信息惠民向纵深发展	20
趋势6: 加快探索交易中心模式走向成熟	21
趋势7: 人工智能市场化、产业化取得新进展	21
相关链接 全球大数据的发展趋势	22

第二章 企业迈入大数据时代

随着网络、特别是移动互联网的快速发展,大数据已渗透到了我们工作与生活的方方面面,也正使得我们生活、工作以及思维方式发生巨大变化。无论是一个国家、一家企业,还是我们每一个人,都将受到大数据时代的影响。

第一节 大数据转化企业思维	29
转化1: 样本思维向总体思维的转变	29
转化2: 精确思维向容错思维的转变	29
转化3: 因果思维向相关思维的转变	30
转化4: 自然思维向智能思维的转变	30
相关链接 中小企业的“数据借力”之道	32
第二节 大数据变革企业管理	33
变革1: 企业管理思想的转变	33
变革2: 企业管理决策的转变	34

变革3: 企业人力资源管理的转变	35
变革4: 企业财务管理的转变	36
相关链接 大数据时代下的管理模式创新	37

第三节 大数据助力企业发展

39

助力1: 帮企业了解用户	40
助力2: 帮企业锁定资源	40
相关链接 啤酒与尿布的故事	40
助力3: 帮企业规划生产	41
助力4: 帮企业做好运营	41
助力5: 帮企业开展服务	42
相关链接 大数据提升企业竞争力	42

第三章 企业大数据的技术架构

45

大数据的爆炸式增长已超出了传统IT基础架构的处理能力,企业应开发新的数据架构,围绕“数据收集、数据管理、数据分析、知识形成、智慧行动”的全过程,开发使用这些数据,让其释放出更多隐藏价值。

第一节 企业大数据生态产业链

47

生态1: 大数据提供者	47
生态2: 大数据技术提供者	48
生态3: 大数据服务提供者	49

第二节 企业大数据的建设思路

49

思路1: 数据的获得	49
思路2: 数据的汇集和存储	50
思路3: 数据的管理	50
思路4: 数据的分析	51
思路5: 数据的价值	51

思路6: 数据的使用	52
第三节 企业大数据的基本架构	52
架构1: Hadoop技术平台	52
架构2: 大数据平台系统功能	54
第四节 企业大数据的目标效果	55
效果1: 数据整合	56
效果2: 数据质量管控	56
效果3: 数据共享	56
效果4: 数据应用	57

第④章 企业大数据的运营 59

移动互联网时代给了企业获取海量数据的机会,有了完整数据资源的宝库,企业需要在创新中求变,要通过高效的数据资源分析和利用,在竞争日益激烈的市场环境中作出精准决策,提高数据经营的质量,盘活手中的数据资源。

第一节 构建企业大数据的运营体系	61
构建1: 数据基础平台	61
构建2: 数据报表与可视化	62
构建3: 产品与运营分析	63
构建4: 精细化运营平台	64
构建5: 数据产品	64
构建6: 战略分析与决策	64
第二节 催生企业大数据的商业模式	64
催生1: 数据自营模式	64
催生2: 数据租售模式	66

催生3: 数据平台模式	66
催生4: 数据仓库模式	68
催生5: 数据众包模式	69
催生6: 数据外包模式	70
第三节 实现企业大数据的应用场景	71
应用1: 消费者需求分析	71
应用2: 打通生产竖井	72
应用3: 产品与服务的设计	73
应用4: 开放式的融合创新	74
应用5: 适应性库存管理	74
应用6: 质量管理	75
应用7: 劳动力的数字化	76
应用8: 资产智能管理	76
第四节 实施企业大数据的关键节点	77
实施1: 制定大数据规划找准切入点	78
实施2: 强化高管团队大数据能力	79
相关链接 大数据时代, CDO 崛起	80
实施3: 设计合理的大数据组织架构	80
实施4: 搭建有效的大数据团队	81
相关链接 中央数据部门团队的能力要求	82
实施5: 用制度和文化的保障大数据的实施	83

第 五 章 金融业与大数据

大数据是重塑金融竞争格局的一个重要支撑和抓手,对它的有效利用,将带动整个行业的发展,给整个金融体系带来创新动能。金融行业借助于大数据,将会获得更快的发展速度、更低的成本、更多的先发优势。

第一节 金融大数据的认知	87
认知1: 金融大数据的概念	87
认知2: 金融大数据的特点	87
认知3: 金融大数据的意义	88
认知4: 金融大数据的模式	90
相关链接 京东供应链金融的运营模式	91
第二节 大数据对金融业的影响	92
影响1: 大数据带给金融业的机遇	92
影响2: 大数据带给金融业的挑战	93
相关链接 大数据与金融业的完美结合	94
第三节 大数据在金融业的应用	96
应用1: 银行大数据的应用	97
应用2: 保险行业大数据的应用	100
应用3: 证券行业大数据的应用	104
应用4: 互联网金融大数据的应用	107
第四节 金融行业大数据应用案例	107
案例1: 中信银行玩转大数据, 以客户为“上帝”	107
案例2: IBM助力民生银行应对金融业的大数据挑战	108
案例3: 大数据助力青岛银行提升运营能力	110

第 六 章 工业与大数据

“大数据”的本质是基于互联网基础上的信息化应用, 其真正的“魔力”在于信息化与工业化的融合, 使工业制造的生产效率得到大规模提升。工业大数据已经成为当前制造业转型升级的关键, 也是工业4.0的重要核心。

第一节 工业大数据的认知	115
认知1: 工业大数据的概念	115
认知2: 工业大数据的种类	115
认知3: 工业大数据的特征	116
认知4: 工业大数据的特点	116
认知5: 工业大数据的用途	117
认知6: 工业大数据的发展态势	119
相关链接 中国将推动大数据和制造业相融合	119
第二节 大数据带动制造业转型升级	120
带动1: 数据赋能制造业的全流程	120
带动2: 数据驱动生产和制造柔性化趋势	121
带动3: 利用大数据创新商业模式	122
带动4: 学习国际上工业大数据应用的实践	122
带动5: 从数据化到智慧化要循序渐进	123
相关链接 大数据推动下的工业4.0	123
第三节 大数据在工业企业的应用	125
应用1: 加速产品创新	125
应用2: 高价值大型产品的故障诊断与预测	125
应用3: 工业供应链的分析和优化	126
应用4: 产品销售预测与需求管理	126
应用5: 生产制造环节的改善	127
应用6: 产品质量管理与分析	128
第四节 工业企业大数据应用案例	129
案例1: 德国安贝格西门子智能工厂	129
案例2: 德国博世洪堡工厂	130

第七章 零售业与大数据..... 133

通过对零售行业数据进行挖掘分析，能够给零售企业带来巨大的商业价值以及服务创新，诸如能够更好地了解 and 洞察消费者，从而实现精准化营销，或者变革供应链模式，实现货品精细化管理等。

第一节 零售业大数据的认知..... 134

认知1：零售业大数据的范畴..... 135

认知2：零售业大数据的发展构想..... 135

认知3：零售业大数据的价值..... 136

认知4：零售业大数据的应用需求..... 138

第二节 大数据对消费行为的改变..... 139

改变1：消费者购买决策更为理性..... 140

改变2：消费者品牌依赖度弱化..... 140

改变3：消费者需求更加个性化..... 141

改变4：消费者在参与互动中获得消费体验..... 142

第三节 大数据在零售企业的应用..... 143

应用1：用户画像让商家读懂用户..... 143

应用2：提供个性化的服务与商品推荐..... 144

应用3：大数据与用户忠诚度管理..... 145

应用4：重要客户的识别与维护..... 145

应用5：提升用户洞察力，优化产品与服务..... 146

相关链接 大数据在零售行业的创新性应用..... 146

第四节 零售企业大数据的应用案例..... 149

案例1：ZARA把消费者声音转化成大数据..... 149

案例2：7-Eleven的App，既是数据搜集器也是行动店员..... 150

案例3：沃尔玛利用大数据颠覆零售业..... 152

第八章 大数据面临的问题及应对..... 155

随着互联网+行动计划进一步推进实施,大数据将加速从互联网向更广泛的领域渗透。我们在享受大数据带来各种方便的同时,大数据也给我们带来了新的问题,如安全问题、隐私问题等。

第一节 大数据与安全挑战..... 156

挑战1: 信息管理成本显著增加..... 157

挑战2: 信息有效性验证工作大大增加..... 157

挑战3: 安全防御边界有所扩展..... 158

挑战4: 独立决策的比例显著降低..... 159

挑战5: 攻防双方地位的不对等性大大降低..... 159

挑战6: 安全加固策略的复杂性有所降低..... 160

相关链接 大数据安全问题频现..... 160

第二节 大数据与隐私保护..... 163

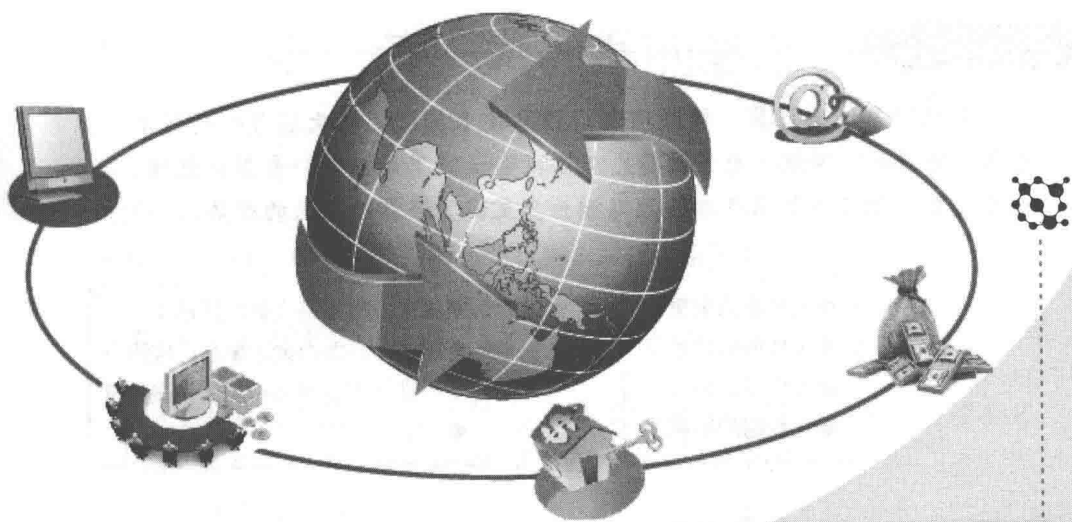
要点1: 保护隐私成新趋势..... 163

要点2: 隐私的概念及量化..... 163

要点3: 大数据生命周期的隐私保护模型..... 165

相关链接 大数据时代如何进行有效的隐私保护..... 166

参考文献..... 169



第一章 认识大数据



互联网技术的飞速发展和应用，不仅改变了生产和消费的行为及商业模式，同时也带来了信息数据在数量、频度和使用等多方面的巨大变革。从数据角度看，整个世界已跨入“大数据”时代。

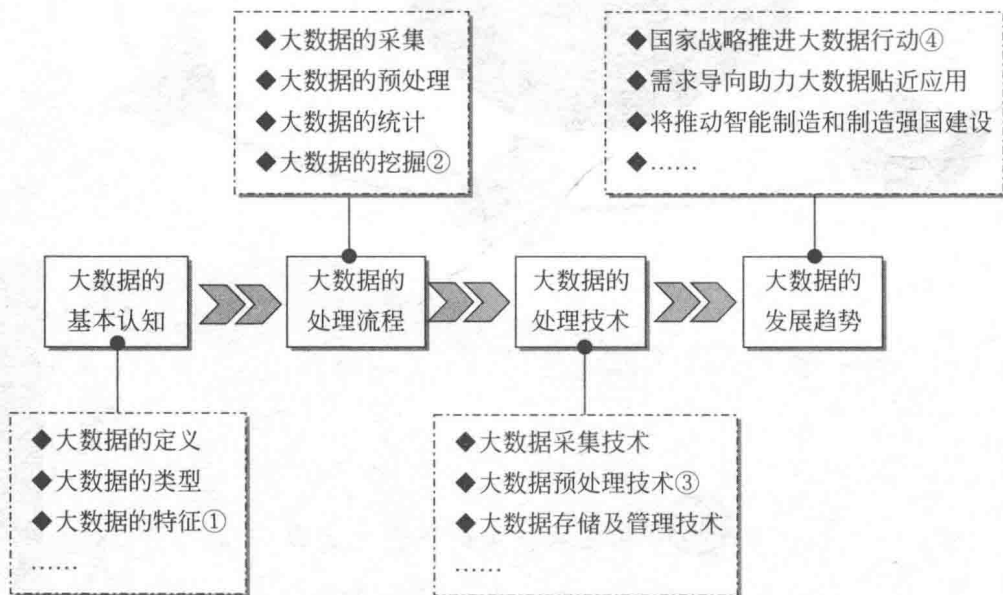


一本书搞懂

企业大数据

阅读指引

大数据是继云计算、物联网之后信息技术领域又一次颠覆性的技术变革。随着社交网络、电子商务、互联网和云计算的兴起,音频、视频、图像、日志等数据量正在以指数级增长,呈现了爆炸性增长的趋势。



图示说明:

① 大数据意味着包括交易和交互数据集在内的所有数据集,其规模或复杂程度超出了常用技术按照合理的成本和时限捕捉、管理及处理这些数据集的能力。大数据通常有如下特征:数据体量巨大、数据种类繁多、流动速度快、价值密度低。

② 大数据的挖掘是指对之前已经做好统计的大数据基于不同的需求,利用数据挖掘算法进行挖掘。

③ 为了提高数据挖掘的质量而产生的数据预处理技术有如下四种方法:数据清理、数据集成、数据变换、数据归约。

④ 在大数据战略的引领下,依托纲要的贯彻实施,我国将在大数据产业规划编制、地方大数据试点、大数据关键产品研发和产业化、大数据基础设施建设、大数据标准体系建设等方面取得显著进展。



第一节 大数据的基本认知

随着新一代信息技术的迅猛发展和深入应用，数据的数量、规模不断扩大，数据已日益成为土地、资本之后的又一种重要的生产要素，与各个国家和地区争夺的重要资源，谁掌握数据的主动权和主导权，谁就能赢得未来。

认知1：大数据的定义

大数据还没有统一的标准定义。尽管大数据的定义各家歧异，但基本上，大数据领域里的每个人都同意一点：大数据不仅仅是指更多资料。

1. 百度搜索的定义

百度搜索认为：大数据（Big Data），指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合，是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产。

2. 互联网周刊的定义

互联网周刊的定义为：“大数据”的概念远不止大量的数据（TB）和处理大量数据的技术，而是涵盖了人们在大规模数据的基础上可以做的事情，而这些事情在小规模数据的基础上是无法实现的。

换句话说，大数据让我们以一种前所未有的方式，通过对海量数据进行分析，获得有巨大价值的产品和服务，或深刻的洞见，最终形成变革之力。

3. 研究机构的定义

研究机构认为：“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

从数据的类别上看，“大数据”指的是无法使用传统流程或工具处理或分析的信息。它定义了那些超出正常处理范围和大小、迫使用户采用非传统处理方法的数据集。



相关链接 ▶▶

大数据中多种多样的数据单位

适用于大数据的技术，包括大规模并行处理（MPP）数据库、数据挖掘电网、分布式文件系统、分布式数据库、云计算平台、互联网和可扩展的存