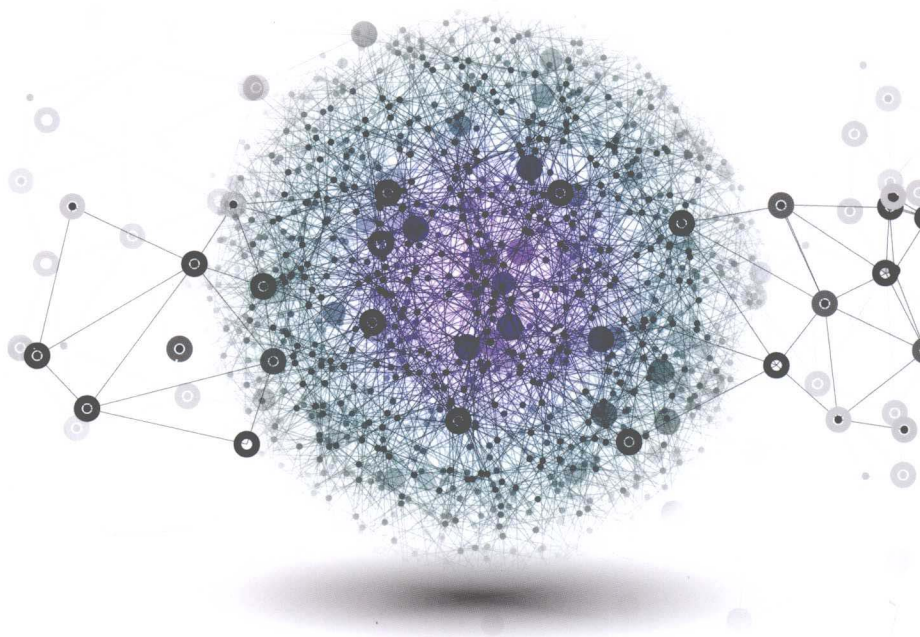




康路晨◎著

Big Data Era

一本书读懂

大数据时代

大数据=大决策，带你从巨量资料中找到人潮、钱潮、资讯潮！

图解、案例、策略与实战，一本书彻底读懂大数据时代！

中恒国际管理咨询机构推荐阅读

民主与建设出版社

大数据=大决策，带你从巨量资料中找到人潮、钱潮、资讯潮！

图解、案例、策略与实战，一本书彻底读懂大数据时代！

中恒国际管理咨询机构推荐阅读

一本书读懂 大数据时代

Big Data Era 康路晨◎著

民主与建设出版社

图书在版编目（CIP）数据

一本书读懂大数据时代 / 康路晨著. — 北京 : 民主

与建设出版社, 2015.1

ISBN 978-7-5139-0562-6

I. ①一… II. ①康… III. ①互联网络—基本知识

IV. ①TP393.4

中国版本图书馆CIP数据核字（2015）第016418号

出 版 人：许久文

责任编辑：李保华

整体设计：久品轩

出版发行：民主与建设出版社有限责任公司

电 话：(010)59419778 59417745

社 址：北京市朝阳区曙光西里甲六号院时间国际8号楼北楼306室

邮 编：100028

印 刷：固安县保利达印务有限公司

版 次：2015年4月第1版 2015年4月第1次印刷

开 本：16

印 张：15.75

书 号：ISBN 978-7-5139-0562-6

定 价：39.80元

注：如有印、装质量问题，请与出版社联系。

从王永庆的“原始大数据”到《纸牌屋》的成功

在很多人还不知道“大数据”这个词的含义时，大数据时代已经悄然到来。事实上，大数据并不是什么新鲜事物，人们对于大数据的无意识的零散应用，更是一直持续进行着的。

台湾“经营之神”王永庆，在卖大米的时候就曾应用过大数据的方法。

当时，王永庆的米店有一项送米上门业务。在为新顾客送米时，他总是会注意观察这家人米缸的容量，他还会尽量多跟顾客聊一下：问问家里有多少人、每顿饭要做多少米？这样一来，他就会根据所获得的信息判断大致何时该再次送米上门。果不其然，等他送米上门的时候，顾客往往也正准备去买米。王永庆的米店服务如此贴心，顾客自然乐意光顾。不要小瞧了这种小生意。假设王永庆有500个顾客，1户人家按4口人算，每户每月用米平均100斤，500户一月用米就是50000斤。一年就是300吨。这是一个非常惊人的数字。付款方式一般分为三种：现收现付、先付后收、先收后付。为了挤出更多的头寸，王永庆一方面需要找到对自己更有利的供货商，可以让自己适当延期付款；另一方面，为了拓展

新的顾客群，他需要尽量增加先付款后拿米的客户。

王永庆结合自己的业务发展，来关心顾客的需要和供货商的需要，并且将顾客的各种数据集合起来以提高自己的业务量，这也算是对大数据的一点浅显应用吧。

近年来，随着社交网络的逐渐成熟，移动带宽迅速提升，云计算、物联网应用更加丰富。更多的传感设备、移动终端接入网络，数据出现了爆炸式增长，数据发掘的软硬件环境也越来越成熟。而关于大数据的最新故事，是美国电视剧《纸牌屋》的成功。有报道说，《纸牌屋》能够迎合多数观众的口味和兴趣，甚至连美国总统奥巴马都在Twitter上留言“明天纸牌屋回归，请大家不要剧透”。《纸牌屋》之所以能取得这样的成功是因为其剧情设计是基于Netflix的网站观看和搜索行为数据分析。

那么，大数据究竟“大”到了什么程度呢？

一项针对甲骨文公司的独立用户的调查发现，90%的企业的数据量在迅速上涨，其中16%的企业的数据量每年增长率达到50%或更高。不少企业已经感受到失控的数据增长对绩效造成的冲击，其中87%的受访者将企业的应用程序性能下降归咎于不断增长的数据量。另一项调查则显示，全球数据量在2011年已达到1.8ZB，在过去5年里增加了5倍。1.8ZB是什么样的概念呢？如果把所有这些数据都刻录存入普通DVD光盘里，光盘的高度将等同于从地球到月球的一个半来回也就是大约72万英里。这相当于每位美国人每分钟写3条推特微博，而且还要不停地写2.6976万年，是不是很恐怖？这还不是最恐怖的，一个权威调查机构还预测全球数据量大约每两年翻一番，2015年全球数据量将达到近8ZB，到2020

年，全球将达到35ZB。这个数据不可谓不大！

海量数据成了现实世界的量化反映，通过人们口袋里和包里的手机、平台电脑，人与人被密切连接在一起；你每天上网玩游戏、用手机、住宾馆、逛超市、买车票都会产生数据痕迹；通过谷歌眼镜和穿戴型数据终端以及形形色色的数据传感装置，我们能将物理界、生物界和社会界的万事万物数据化并加以存储。

获取和利用大数据，寻找过去现实世界中没有的全新生活方式、社会治理机制和经济发展途径，开始成为社会方方面面关注投入的中心。

在未来，数据将会像土地、食物和水一样，成为我们生活的必需资源：

一个零售商可以根据顾客购买的婴儿内裤颜色、尺寸，预测未来提供儿童玩具的时间和种类；

百度地图推出了“景区热力图”，以热力图的形式直观地展示众多热点商业区的人流量，这对商业选址、旅游出行都有指导意义；

华尔街资本市场通过分析股民微博账户留言来判断其情绪，这些人看多时买股票，看空时抛售股票；

招商银行利用大数据分析客户信用卡经常使用的场所，兴业银行用数据开发新客户，平安集团利用数据分析消费者行为……政府、电信、教育、医疗、娱乐等行业都在利用大数据获取信息。

对于个人来说，手机也不再仅是一个通话终端，根据你的通话记录、浏览网站、流量记录，使用范围记录，手机信号强弱记录，运营商就可以分析出你的收入水平、兴趣爱好等。一句话，大数据将无处不在，无物不读。

欢迎来到大数据时代，一个崭新的、神奇的大数据时代！

第1章 读懂大数据时代的N个关键词

大数据是什么 / 3

4V特征 / 3

云计算与物联网 / 4

移动互联网 / 5

大数据思维 / 6

大数据技术 / 7

感知技术 / 7

存储技术 / 8

云计算 / 9

分布式技术 / 10

第2章 大数据前夜的数据探索

从小数据走向大数据 / 15

啤酒与尿布的销售关联 / 15

数据与分牛难题 / 18

从流行病说现场调查 / 23

大数据的前世今生 / 26

庞涓与小数据的失败 / 26

诞生于大瘟疫的统计学 / 29

问卷调查与小数据的局限 / 34

第3章 0和1的数据“大爆炸”

互联网新世代 / 39

复杂计算的烦恼 / 39

从织布机到计算机 / 42

电子时代到来 / 45

“电脑”的由来 / 48

全新的技术革命 / 51

互联网的兴起 / 54

从云计算到大数据 / 58

数据激增 / 58

数据如何“大”起来 / 60

大数据的商业价值在哪里 / 64

第4章 大数据时代的思维革命

让数据走向更多、更快、更好 / 71

免费的才是最贵的 / 71

一切皆可数据化 / 74

一切都可以量化 / 77

大数据≠大价值 / 80

数据是最宝贵的资产 / 82

数据就是机遇 / 82

第5章 数据社会与数据人

工作中的大数据 / 87

搜狗热词的秘密 / 87

猿题库的创业 / 89

生活中的大数据 / 93

量化自我的健康生活 / 93

大数据的交通红利 / 97

大数据致富之路 / 100

垃圾达人的数据挖掘 / 100

卖掉自己的大数据 / 102

大数据帮助政府决策 / 105

世纪选举的背后 / 105

旧数据的新用途 / 107

大数据与国家安全 / 110

拉登是怎么死的 / 110

大数据的火眼金睛 / 113

卡特里娜飓风 / 116

未来的福尔摩斯 / 120

第6章 大数据的趋势与障碍

谁会成为大数据时代的“新贵” / 125

大数据时代——存储的世界 / 125

数据增值的关键——数据分析 / 127

大数据的局限 / 130

大数据时代的6个非技术问题 / 130

个人隐私的战争 / 134

算法不能代替判断 / 137

计算机的危机 / 139

大数据的未来走向 / 143

第7章 大数据时代的机遇与挑战

盘活数据资产 / 147

淘宝数据王国——海量数据称王 / 147

亚马逊数据库——精确营销制胜 / 151

大数据时代的线上线下融合 / 153

大数据时代面临的挑战 / 154

第8章 大数据时代的跨界与颠覆

教育行业的变革 / 159

新闻媒体的转型 / 161

影视行业的探索 / 165

制造业的大数据之路 / 167

餐饮旅游业的未来 / 169

传统农业的展望 / 171

第9章 大数据时代的商业营销

大数据让营销更精准 / 177

更智能的广告 / 177

塔吉特的“读心术” / 181

大数据与品牌代言 / 183

大数据的用户体验 / 185

用户体验的威力 / 185

LinkedIn的成功 / 189

大数据的粉丝经济 / 193

大悦城的大数据营销 / 193

小米的崛起之路 / 196

可口可乐的昵称瓶 / 199

第10章 大数据时代的企业创新

大数据另辟蹊径 / 205

大数据与流感预测 / 205

错误数据的用处 / 208

数据也甜蜜 / 210

IBM的美味机器 / 214

“预言帝”的诞生 / 218

大数据的破坏式创新 / 222

余额宝的大数据思维 / 222

帮人怀孕的手机软件 / 225

传统企业的大数据 / 228

小钱包做大事情 / 228

防疲劳驾驶的尝试 / 231

大数据改变篮球比赛 / 233

后记

第1章

读懂大数据时代的N个关键词

麦肯锡最早预言了大数据时代的到来：“数据，已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来。” IBM 则最早定义了大数据的特征：规模性Volume、多样性Variety、价值性Value和高速性Velocity。



大数据是什么

4V特征

大数据(big data), 也被称为巨量资料, 其被定义为无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的数据集合。

麦肯锡最早预言了大数据时代的到来: “数据, 已经渗透到当今每一个行业和业务职能领域, 成为重要的生产因素。人们对于海量数据的挖掘和运用, 预示着新一波生产率增长和消费者盈余浪潮的到来。” IBM则最早定义了大数据的特征: 规模性Volume、多样性Variety、价值性Value和高速性Velocity。

4个“V”反映了大数据在四个层面的特点: 第一, 数据体量巨

大。从TB级别，跃升到PB级别，大数据的起始计量单位至少是P（1000个T）、E（100万个T）或Z（10亿个T）；第二，数据类型繁多，比如，网络日志、视频、图片、地理位置信息等；第三，价值密度低、商业价值高，只要合理利用数据并对其进行正确、准确地分析，将会带来很高的价值回报；第四，处理速度快，可从各种类型的数据中快速获得高价值的信息，这一点和传统的数据挖掘技术有着本质的不同。

云计算与物联网

云计算是一种基于互联网的计算方式，通过这种方式，共享的软硬件资源和信息可以按需求提供给计算机和其他设备。而物联网则是物物相连的互联网。具体来说，物联网的核心和基础仍然是互联网，是在互联网基础上的延伸和扩展的网络；其用户端延伸和扩展到了任何物品与物品之间，进行信息交换和通信。物联网技术带给事物的是一种技术上的革新，而它的载体还是原有的事物。

那么，云计算和物联网之间是什么关系呢？

第一，云计算是实现物联网的核心。物联网时代，所有设备将实现互联互通，随之而来的则是巨大的数据量，而运用云计算模式，将使物联网中数以兆计的各类物品的实时动态管理、智能分析变得可能。物联网通过将射频识别技术、传感器技术、纳米技术等新技术充分运用在各行各业之中，将各种物体充分连接，并通过无线等网络将采集到的各种实时动态信息送达计算处理中心，进行汇总、分析和处理。

第二，云计算将成为物联网的重要环节。云计算是以应用为目的，

通过互联网将大量必须的软硬件按照一定的形式连接起来，并且随着需求的变化而灵活调整的一种低消耗、高效率的虚拟资源服务的集合形式。物联网强调物物相连，设备终端与设备终端相连，云计算能为连接到云上设备终端提供强大的运算处理能力，以降低终端本身的复杂性。二者都是为满足人们日益增长的需求而诞生的。

移动互联网

近几年来，智能手机、平板电脑等移动终端的普及，让移动互联网成为我们最贴身的媒体。毫不夸张地说，移动互联网已经极大地改变了我们的生活方式。

移动互联网的定义比较简单，它是互联网的技术、平台、商业模式和应用与移动通信技术结合并实践的活动的总称，是一种通过智能移动终端，采用移动无线通信方式获取业务和服务的新兴业务，包含终端、软件和应用三个层面。

移动互联网的适用场合非常广泛，可以满足形式简单的需求；它拓展了互联网的用途，扩大了互联网的市场，却创造不了PC互联网那样的工业价值。

移动互联网一共有三个入口：浏览器、App、二维码。移动互联网诞生之初，浏览器成为用户在手机端延续桌上互联网的行为习惯，但是随着 iPhone 横空出世，苹果通过“IOS+App Store”重新定义底层结构，手机用户开始学习使用本地 App 连接丰富的网络服务，Android（Google Play）也随之跟进，共同确立了“操作系统搭台、应用程序唱戏”的游