

从“条数据”到“块数据”

全面阐述大数据时代的颠覆性变革
一个新的时代到来了

块数据

大数据时代
真正到来的标志

大数据战略重点实验室◎著



中信出版集团 · CHINA CITIC PRESS

块数据

大数据时代
真正到来的标志

大数据战略重点实验室◎著

图书在版编目 (CIP) 数据

块数据 / 大数据战略重点实验室著. —北京: 中信出版社, 2015.5

ISBN 978-7-5086-5043-2

I. 块… II. 大… III. 经济管理—数据管理—通俗读物 IV. F2-39

中国版本图书馆CIP数据核字 (2015) 第026446号

块数据

著 者: 大数据战略重点实验室

策划推广: 中信出版社 (China CITIC Press)

出版发行: 中信出版集团股份有限公司

(北京市朝阳区惠新东街甲4号富盛大厦2座 邮编 100029)

(CITIC Publishing Group)

承 印 者: 北京盛源印刷有限公司

开 本: 880mm×1230mm 1/32 印 张: 8.75 字 数: 170千字

版 次: 2015年5月第1版 印 次: 2015年5月第1次印刷

广告经营许可证: 京朝工商广字第8087号

书 号: ISBN 978-7-5086-5043-2 / F · 3325

定 价: 45.00元

版权所有·侵权必究

凡购本社图书, 如有缺页、倒页、脱页, 由发行公司负责退换。

服务热线: 010-84849555 服务传真: 010-84849000

投稿邮箱: author@citicpub.com

大数据对经济发展的推动已经超出了所有最伟大的预言家的预测，而本书认为大数据对社会发展的推动更将是一个史诗般的剧目。这出剧的序幕正准备拉开或者说刚刚拉开，而拉开序幕的推手就是块数据。

——大数据战略重点实验室

编委会名单

编委会主任：陈刚

编委会常务副主任：刘文新

编委会副主任：聂雪松 高卫东 徐昊 连玉明

康克岩 李瑶

撰稿人：王黔 唐振江 刘春辉 陈栋

王鹏飞 徐先文 李祥 王麒

赵刚 杨达 白延成 张清

谌贻建 张洁

受摩尔定律驱动的信息技术的不断廉价化、互联网的普及以及其延伸所带来的信息技术无处不在的应用，催生了大数据时代的到来。进而使信息化进入以数据的广泛关联、跨域融合和深度应用为特征的智慧化阶段。在当前的大数据热潮中，相关的书籍、文章可谓车载斗量，共识与各种观点的争鸣共存。本书却从一个新颖的视角——块数据，来看待大数据及其未来的发展，颇有创意，也发人思考。读完本书，感受有三。

其一，“条数据”和“块数据”的划分，师法自然，抓住了数据的本质。从其定义和静态的角度看，“条”是一个领域或行业内纵深数据的集合，可以反映本领域或行业的规律，无疑具有很大价值。“块”是一个物理区域或行政区划内众多“条数据”的集合，更能反映现实世界和社会的极度复杂性，其综合应用

无疑会带来数据价值的显著提升。从动态的视角看，重视“块数据”是为了避免仅仅关注“条数据”而可能带来的新的数据孤岛现象，更体现了一种对信息化建设的发展性思维。

其二，讨论大数据，很多人都会谈到当前价值密度高的数据大多掌握在政府手中，期待政府依法依规开放数据并推动数据共享。本书恰恰是从政府管理和社会治理的视角，使用了在管理和治理体系中常用的“条”“块”概念，着眼于当前“最有价值”的一部分数据的有效、高效应用，其意义不言而喻。

其三，本书源于贵阳市领导在推进大数据应用及其产业发展过程中的思考和实践，既有理论的阐述，也有实践的支撑，呈现了对诸多方面极有价值的参考。期待贵阳市在本书指引下，在其大数据驱动的智慧城市建设中，积极探索，稳步前行，为信息化建设的新阶段——智慧化，树立示范。

梅 宏

全国信标委大数据标准工作组组长、
中国科学院院士、上海交通大学副校长

2015年2月

云计算、大数据现在已经成了科技发展的代名词，我和不少政府官员讨论云计算、大数据对社会经济发展和资源配置的影响，深感贵州政府是对新技术最敏锐、最大胆的实践者之一。

数据是驱动商业向前发展的核心，更是人类社会的未来。我们看到这样一个趋势：以控制为出发点的IT时代正在走向以激活生产力为目的的DT（数据技术）时代。从组织内部角度来看，DT会改变一个组织的沟通、生产、消费方式，驱动它的架构、文化的变革；从跨组织角度来看，由于DT时代的“利他”思维取代了IT时代的“利我”思维，组织与组织的合作将远大于竞争，跨组织的协同会频繁发生，而且将变得越来越敏捷，越来越高效。这不仅仅是技术的升级，更是思想意识的巨大变革，我可以肯定地说，这无疑是一个更波澜壮阔的大时代。

阿里巴巴集团于2014年4月17日与贵州省政府签订了“全面战略合作协议”，项目之一“云上贵州”已经取得一些成绩，成为政府运营云计算和大数据的最佳实践。政府作为一个组织，生来就是一个极为重要的数据生产和交换平台。数据本身并不能创造价值，只有让更多的人对其进行分析和运用，才能成倍地创造价值。贵州省政府、贵阳市政府、贵安新区就是本着这样一个开放的心态，将自身打造为推动地区发展的数据引擎——而非数据仓库，将政府数据公开，鼓励那些有创造力的年轻人在这个平台上创业，挖掘新的商业机会，既提高了执政效率又促进了民生服务。这不仅是对云计算领域的预判和布局，也给本地区注入了新的活力，让当地年轻人有机会站在时代的最前沿。

很多地方都在谈数字经济、云计算产业，但贵州下如此大的决心，善于利用阿里巴巴集团业已搭建的云计算和数据平台，跟我们一起干、坚持干、务实干，一起探索未来，这样的勇气和魄力值得敬佩。

作为一位专研型官员，贵阳的陈刚书记既是上述实践的推动者、实施者，也是大数据理论的研究者。我不懂技术，但深知一边做，一边想，每天努力，未来才会到来。

祝愿我们理想成真。

马云

阿里巴巴集团董事局主席

2015年新春于杭州

块数据是一个新名词。到目前为止，人类生活、生产产出的大数据，无论是搜索引擎、电商、社交平台形成的企业大数据，还是天气预报形成的科学大数据，还是各政府部门形成的行业大数据，都可以定义成“条数据”。而块数据则是以一个物理空间或行政区域形成的涉及人、事、物的各类数据的总和。

在此书完稿后，不少人士甚至是专业人士都问我，你能用一个形象的比喻把块数据说清楚吗？我想举个例子说明：一个百货商场卖出许许多多的商品，每个商品有从原材料到加工成商品的生产过程的数据，也有品牌设计、广告营销和销售数量的数据，以及产品售后服务、商户反馈等服务数据，这些都是以产品为中心的“条数据”，而百货商场销售的商品种类、数量，男女老少在商场的购物、娱乐情况，天气、公交和停车场对商场经营情况的影

响等却易被忽视。事实上，当商场的影院播映一部聚人气的大片时，商场的销售量也会随之上升，这些数据可以称为块数据，这个“块”是指这个商场，商场这个物理空间产生的数据总和就是商场的块数据。当块扩大到社区和城市层面时，在这个块上形成的数据总和就是本书所指的块数据。

现任国家大剧院院长的陈平先生，在21世纪初时任北京市东城区区委书记期间，探索并构建了一整套的“网格化治理”城市的模式，时任微软总裁比尔·盖茨先生由衷评价：“东城区网格化是人类发明计算机以来在社会领域应用方面最伟大的发明。”国家建设部在全国推广复制这一模式。本人在北京朝阳区工作期间，学习借鉴东城经验，并将这个模式从对“物”的管理，转向对社会治理中动态的“人”、“事”和“物”的管理，探索了“全模式”社会治理的路径。由于受技术的局限和历史因素限制，这些探索尽管已经足够了不起了，仍属于初步探索。例如，“万米网格”给所有公共物件进行人工数字编码，然后靠网格员手持掌上电脑（PDA），依靠人的巡查去控制各个公共物件的运转情况，把这些公共物件的信息汇聚在一起，形成物联网（IoT）的初级阶段，也可以称为手工物联网。

过去的10年，人类在互联网、云计算、大数据、物联网等领域突飞猛进，这些技术的进步成就了硅谷的苹果、Google（谷歌）和中国BAT（百度、阿里巴巴、腾讯）等这样一批互联网企业，人类目前已经认识到，“大数据”将在经济领域彻底颠覆人类自工业革命以来积累形成的经济模式和商业模式。

块数据的贡献，将会使人类在大数据领域初步探索并逐步形成的条数据基础上，在一个物理空间形成汇聚，形成条数据和块数据的融合式发展。这正是本书以“块数据——大数据时代真正到来的标志”为题目的原因。只有条数据和块数据融合式发展，人类所描述的大数据的时代特征才真正开始实现。

本书对块数据的基本概念、基本特征进行了阐述，对块数据的聚集路径进行了设计，进而特别提出了块数据有别于条数据的最大特征——共享性和可交易性。与此同时，本书指出块数据对经济发展的推进，特别是数据共享和可交易性为时下推崇的“大众创业、万众创新”的创新驱动发展模式，插上了腾飞的翅膀。更为重要的事情是，块数据是提高社会治理能力、更好地服务民生的“共性需求”和“长尾需求”的一条现实的可实现路径。此外，大数据让“世界变成透明的”，块数据在对监督政府依法行政行为和倒逼政府职能改革方面可实现其他任何外力手段不能达到的推动力。

需要说明的是，本书所讲的块数据不是过去这些年各地方政府的“智慧政府”、“智慧城市”的“集成版”。与后者有别，块数据构建的模式是开放的、共享的和可自更新的，是一种可持续发展的模式。它解决的不仅仅是“数据孤岛”和统一平台标准问题，更多的是一种“互联网思维”模式，由政府、社会和个体共同以“众包”的模式构建一个面向未来的数据平台。本书对这方面的内容从机制上和构架上进行了阐述。

本书适合互联网的创业者阅读，读者会从本书内容中判断互

联网的未来发展方向和商业趋势。本书适合企业管理者阅读，当下对大数据概念的炒作可能让你对互联网发展方向有点儿迷茫，本书可以告诉你互联网大数据发展的方向，让你从一个全新的视角重新认识大数据。本书也适合金融家、投资者阅读，你会从中找到投资的方向和发现成千上万的新投资机会，或许你就是下一批“BAT”的早期股东。本书同样适合政府公职人员阅读，社会治理能力提升的根本出路在于信息化，而块数据是实现社会信息化的必由之路。你会从本书中找到提升政府治理能力的方向。

最后还要说明的是，本书不是互联网、大数据行业的入门读物。当然，如果你有志于关注互联网、了解大数据的话，这本书或多或少会对你有帮助。大数据是新生事物，现在不存在真正意义上的大数据专家。若干年可能甚至若干月以后，你会发现这本书的有些观点已经过时、落伍，甚至是错误的，那么本书出版的目的也许就实现了，因为这意味着大数据特别是块数据又取得了新的进展和突破，而这也正是本人所希望看到的。

感谢在本书创作、完稿过程中各方的支持者。

陈刚

2015年2月于贵阳

当前，世界正处于以数据驱动发展的转折点上，数据的大爆炸炸出了一个创新涌动、风云变幻的大时代，在由无数的树木、建筑、道路、车辆等实体物质构成的现实世界背后，一个由众多计算机、应用软件和数据库通过网络构建的“虚拟帝国”正在成为世界的主宰。在这个“帝国”里，以互联网的广泛深入应用为基础，基于数据汇集、处理、分析基础上的大量商业模式、社会治理模式创新不断涌现，人类的生产方式、生活方式、行为方式甚至价值观念、道德观念正在发生深刻变化。可以说，数据，已经成为充满科幻感的达摩克利斯之剑，具有决胜未来的神奇力量。

这是一个前所未有的时代。有人为之欢欣鼓舞，认为人类正迎来自工业革命以来甚至是人类有史以来最深刻、最伟大、最彻底的大变革，随之而来的是既有利益格局被打破、重构和新生，而正是在这种深度的调整

中，给在上一轮发展大潮中落在后面的人和地区以全新的机会实现后发赶超。也有人不为所动，认为就像历史上那些昙花一现的所谓发明创造一样，泡沫之后，便是深渊。总之，是众说纷纭、莫衷一是。

但是无论如何，凭借着20世纪60年代以来数据存储技术的发展，90年代以来社交网站的兴起，以及近年来数据挖掘技术的突破，我们事实上已经打开了大数据时代之门。

这个时代是大数据与小数据并驾齐驱的时代。一方面，万物的数据化带来的是数据之海的高速扩张，数据量级呈现出跳跃式增长的趋势。在对海量数据进行分析的基础上，我们正在具备发现隐藏在大量细节背后的规律，并用规律预测未来的能力。另一方面，与日常生活紧密相连的一些数据，如消费、健康、教育等数据本身的量级并不大，往往是以M（兆）为单位，但是也可以用于改善每个人的生活，推动产业模式的升级与变革，也就是说，小数据也具有无可取代的价值。

这个时代是虚拟与现实深度交互的时代。网络为人们在现实世界之外创造了一个虚拟的世界，在这个世界的帮助下，现实世界里很多事情变得前所未有的快速、高效、精确和便捷。而在这种变化的背后，是两个世界交流的日趋频繁，是虚拟和现实界限的日益模糊。越来越多原本必须在现实世界里完成的活动正在借助数据模拟技术放到网络上进行。3D（三维立体）电影、游戏、打印技术大行其道，正在把人类送入一个虚拟的“真实”世界。

这个时代，人们对事物的关注点从逻辑性转向逻辑与关联并

重。大数据最重要的特征之一就是在混沌、失序、杂乱的数据中寻找看似毫无关系数据之间的关联性，在挖掘、分析中重构事物的线性关系，进而建立以此为基础的数据模型并用于指导实践，这是对既有强调因果逻辑思维方式的冲击。当然，从发展的趋势来说，数据的分析在某种程度上也必然需要考量因果关系，且不说编码本身的数据基础就是逻辑关系，即使是在比如深度学习、人工智能等技术的发展上也都需要在不断“喂食”数据的同时建立完善的逻辑反馈机制。可以说，关联性和逻辑性都是新时代的基本特征。

这个时代是数据的融合成为数据价值提升关键的时代。单一的数据价值始终是有限的，但如果能够把各个单一的数据串联起来，通过多维度的挖掘比对，像GPS（全球定位系统）一样，就能够实现对事物规律的精准定位。我们知道，大数据在数据的精确度上比不上传统的统计学抽样，但它的优点就在于不需要抽样，所有的数据都是样本。而在将不同领域、不同行业的数据交叉、对比、融合的过程中，对事物的预测、对规律把握的精准性就会不断提升，甚至能够发现以往未能发现的新规律。

无论你信或不信，我们已经站在了大数据时代的风口之上。正是在这样的时代大背景下，块数据应运而生。

推荐序一 // XI

推荐序二 // XIII

序 // XV

引 言 // XIX

上篇

➤ 块数据的定义与特征

第一章 块数据出现——时也，势也

（一）条数据——越长大越孤单 // 005

（二）块数据——坐拥天时和地利的宠儿 // 013

（三）“块”思辨——块数据哲学体系初步构建 // 017

第二章 块数据素描——“金矿”是怎么炼成的

（一）高度关联性：“你”“我”“他”与“们” // 023

（二）立体性：块数据的维度结构 // 025