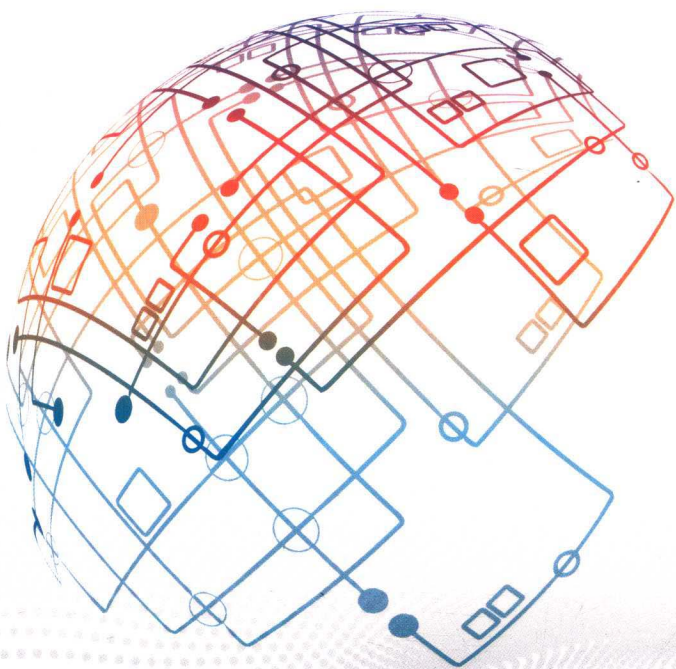


BLOCK CHAIN +BIG DATA

杨永强 蔡宗辉 刘雅卓 著



突破瓶颈，
开启智能新时代

区块链+大数据

全面解析区块链与大数据融合的理论实践
咖联合推荐

区块链 + 大数据： 突破瓶颈，开启智能新时代

杨永强 蔡宗辉 刘雅卓 著



机械工业出版社

继大数据、工业 4.0 之后，区块链作为一种新的互联网技术，以新的时代趋势出现，而且区块链与大数据、工业 4.0 呈现相融相生的格局，特别是区块链和大数据两者更是不可分割，区块链必须与大数据深度融合才能发挥其真正的价值。另外，大数据本身也面临一些很难突破的困境，而区块链技术的迅速崛起，势必会帮助大数据突破困境，让两种技术在相互融合中共生共长。

为了让读者更好地了解区块链的实际应用，本书讲述了区块链在大数据、金融领域、医疗行业、物联网、知识产权管理、能源等方面的应用，同时还对区块链及大数据在各方面的融合发展趋势进行分析和预测。本书属于普及类的大众读物，逻辑清晰、案例新颖，即使是区块链零基础的读者也可阅读。

图书在版编目 (CIP) 数据

区块链 + 大数据：突破瓶颈，开启智能新时代 / 杨永强，蔡宗辉，刘雅卓著. —北京：机械工业出版社，2019.5

ISBN 978-7-111-62633-6

I. ①区… II. ①杨… ②蔡… ③刘… III: ①电子商务—支付方式—研究 IV. ①F713.361.3

中国版本图书馆 CIP 数据核字 (2019) 第 082372 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑：孙业 责任编辑：孙业 秦菲

责任校对：徐红语 责任印制：张博

三河市国英印务有限公司印刷

2019 年 6 月第 1 版第 1 次印刷

169mm × 239mm · 11.5 印张 · 173 千字

0001—3000 册

标准书号：ISBN 978-7-111-62633-6

定价：39.80 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

服务咨询热线：010-88361066

机工官网：www.cmpbook.com

读者购书热线：010-68326294

机工官博：weibo.com/cmp1952

金书网：www.golden-book.com

封面防伪标均为盗版

教育服务网：www.cmpedu.com



杨永强

互联网技术专家，教育区块链产业联盟副理事长，斗云科技董事长，清华大学、长江商学院EMBA。乐视集团创始成员，原乐视网CTO，带领技术团队，从零走向乐视网上市，并主导孵化乐视智能硬件、乐视云、乐视互娱等多项创新业务。在行业内最早提出VaaS架构模式，搭建了国内领先的云视频平台和跨屏多终端互动娱乐体系，拥有多项技术发明专利。曾被CSDN评为中国最具价值和最具人气CTO。



蔡宗辉

数字科技创新研究院执行院长，毕业于北京邮电大学，专注5G通信、大数据营销、产业互联网等方向。曾在中国移动、中信集团、中南大学、童伴教育等企事业担任高管或创始成员，拥有超过14年行业应用管理和互联网营销经验。



刘雅卓

数字经济教育委员会副理事长，斗云科技合伙人，404学院创始人，中国人民大学EMBA。专注新经济、创新科技领域教育及产业研究，专注互联网及互联网教育领域15年，中关村智汇移动互联网教育产业联盟秘书长，互联网教育产业基金主席，科技服务业创新发展试点工程项目负责人。曾先后担任国育科技CEO，奥鹏教育、孔子学院高管等职，并担任全国信息技术标准化技术委员会教育技术委员、特聘专家。

前言

PREFACE



大数据从诞生到今天，已经取得了非常多的成绩，几乎所有的行业都需要大数据来支撑。随着互联网、云计算、物联网等技术的逐步发展，数据的采集、分析也变得日益简单，在这些技术的帮助下，大数据也将发挥出更大的价值。

大数据在处理过程中，与传统数据处理方式相比有三个特点：大数据是对全部数据的处理分析而不是抽样分析；在数据分析时，整体的效率提升，结果更加精确；数据中的相关关系比因果关系更重要。

大数据虽然取得了明显的进展，但是也面临一定的困境。目前，数据已成为企业重要的资源，也越来越受到关注。在数据资源的开发利用中，数据的开放、共享、流通、隐私保护成为痛点。

大数据的核心价值就在于做出精准的预测，高质量的数据是前提，但是，在行业中优质可用的数据非常少，这给数据分析增加了难度。大数据的一大明显特征就是关联性比较强，但是，企业之间、政企之间都把数据紧紧地握在自己手中，不对外共享，形成了严重的信息壁垒。

此外，在进行数据处理时，由于对问题梳理得不够全面，低估大数据的复

杂程度，并且缺乏大数据的分析能力，导致大数据项目应用频频失败。在大数据的实践应用中，缺乏专业的数据分析人才、现有模型数据相对陈旧、数据建模对场景细分不够等问题都给大数据产业的发展带来了困难。

大数据建立在云计算的基础上，通过云管理的手段进一步实现大数据的管理应用是必然的，但现实中屡屡发生的云管理失误案例让大数据行业在企业管理方面的应用受到了质疑，这也是大数据的发展一直停留在困境中的主要原因。

因此，如何解决以上问题是未来大数据发展的重点。而区块链技术的出现为解决大数据面临的问题提供了新的思路。大数据技术是建立在集中式核心思想上的，而区块链的本质是分布式系统。集中式的大数据与分布式的区块链相结合必然产生更大的价值。

区块链具有数据不可篡改、可追溯的特点，能够解决数据共享开放中的很多问题。利用区块链技术，能够建立一个可以记录时间先后、不可篡改、可信的数据，该数据库是去中心化的，能够保证数据的安全。利用区块链技术，不再需要第三方中介组织，没有建立信任关系的双方能够直接实现交易。

在数据的流通领域，隐私保护问题一直存在，一旦数据交易触及了法律，就很难再举证。利用区块链技术，建立一个数据交易追溯平台，然后把每一笔交易数据都存储到该平台中，交易双方可以获得一个交易凭证，交易凭证中记录着这笔交易的数字证书以及交易信息存储的地址，用户可以直接在该平台中完成交易数据的确权。

区块链技术使大数据的交易变成了对数据使用权的交易，这样，数据在产生时就可以被加密存储到区块链中。区块链能够明确交易历史以及各方的贡献，量化数据价值。数据在计算以及输出的每一个环节都被记录在区块链中，如果交易参与者对数据源头存在质疑，就可以利用区块链的追溯特性来核实。

区块链与大数据结合起来，能够打破数据孤岛，建立一个开放的数据共享生态体系。区块链作为一种基础性技术，将会通过与大数据结合的方式被应用到各个行业，为行业带来模式创新，重塑现有的商业模式。

在大数据中使用区块链技术，能够保证数据不被篡改，通过区块链与大数据的结合，区块链中的数据会更加具有价值，实现大数据的预测分析落实到实

践中，促进数字经济时代的发展。

区块链实现了数据的开放共享，扩大了数据规模，规范了数据管理，而大数据能够极大提高区块链的数据价值。区块链数据库记录了每一笔交易数据，并以其可信任、不可篡改的特性让更多的数据被挖掘出来，不同行业中的数据相融合扩大了数据规模，逐步形成全球化的数据交易模式。

目前，区块链技术的发展已经引起了社会各界的广泛关注，人们普遍认为，该技术能够推动互联网的发展，有效解决互联网中的信任问题，从而推动大数据的快速发展。

目录

CONTENTS



前 言

第1章 CHAPTER ONE

大数据发展面临窘境，区块链技术异军突起 / 1

- 1.1 大数据产业发展的三大基石：云计算、物联网、移动互联网技术 / 1
 - 1.1.1 云计算 / 2
 - 1.1.2 物联网 / 4
 - 1.1.3 移动互联网技术 / 7
- 1.2 大数据发展面临五大困境 / 10
 - 1.2.1 优质可用数据少 / 10
 - 1.2.2 信息壁垒严重 / 13
 - 1.2.3 数据处理有困境 / 16
 - 1.2.4 实践应用障碍多 / 18

- 1.2.5 云管理失误多 / 19
- 1.3 区块链发展的三个阶段：1.0—2.0—3.0 / 20
 - 1.3.1 区块链 1.0：货币 / 21
 - 1.3.2 区块链 2.0：合约 / 22
 - 1.3.3 区块链 3.0：新应用 / 24
- 1.4 区块链技术的七大行业应用 / 26
 - 1.4.1 区块链在金融行业的五大应用场景 / 26
 - 1.4.2 区块链在医疗行业的六大发展机遇 / 28
 - 1.4.3 区块链融合物联网，打造智慧地球 / 31
 - 1.4.4 区块链与知识产权管理，让维权更容易 / 32
 - 1.4.5 区块链让智慧能源迎来新契机 / 34
 - 1.4.6 区块链在供应链中消除平台壁垒 / 35

第 2 章
CHAPTER TWO

区块链与大数据的联系与区别 / 39

- 2.1 区块链是大数据的安全载体 / 39
 - 2.1.1 区块链确保数据安全性 / 39
 - 2.1.2 区块链保障数据私密性 / 41
 - 2.1.3 区块链保护数据相关权益 / 45
- 2.2 区块链和大数据的共同关键词：分布式 / 46
 - 2.2.1 分布式：让大数据和区块链从技术权威向去中心化转变 / 47
 - 2.2.2 分布式存储：HDFS VS 区块 / 48
 - 2.2.3 分布式计算：MapReduce VS 共识机制 / 50
- 2.3 区块链与大数据生命周期大不相同 / 52
 - 2.3.1 大数据：萌芽期—过热期—幻灭期—复苏期—成熟期 / 52
 - 2.3.2 区块链：技术萌芽期—过热期 / 55
- 2.4 区块链与大数据在数据方面的六个差异 / 57

- 2.4.1 数据集：足够大 VS 有限 / 58
- 2.4.2 数据结构：结构化 VS 非结构化 / 59
- 2.4.3 信息：独立 VS 整合 / 61
- 2.4.4 表达方式：数学 VS 数据 / 63
- 2.4.5 数据本质：直接 VS 间接 / 65
- 2.4.6 数据特点：匿名 VS 个性化 / 67

第3章

CHAPTER THREE

区块链与大数据技术创新融合 / 69

- 3.1 区块链+大数据 VS 大数据+区块链 / 69
 - 3.1.1 在区块链中融入大数据技术 / 70
 - 3.1.2 在大数据中融入区块链技术 / 72
- 3.2 区块链助力大数据采集、存储与分析 / 73
 - 3.2.1 数据采集：区块链解放更多数据 / 74
 - 3.2.2 数据存储：区块链是强背书的数据库存储技术 / 77
 - 3.2.3 数据分析：区块链授权研究机构才能访问数据 / 79
 - 3.2.4 区块链推进基因测序大数据产生 / 82

第4章

CHAPTER FOUR

区块链在大数据领域的运用 / 85

- 4.1 区块链加速数据流通产业确权时代的到来 / 85
 - 4.1.1 建立去中心化的数据流通平台 / 86
 - 4.1.2 突破信息孤岛，建立数据横向流通机制 / 88
 - 4.1.3 跟踪数据交易的全过程 / 91
 - 4.1.4 保证数据交易不可篡改 / 93
 - 4.1.5 确保链上数据来源可靠性 / 96
 - 4.1.6 对确权数据做登记和验证 / 98
- 4.2 区块链如何助力大数据实践应用 / 101
 - 4.2.1 区块链为大数据分析提供机会 / 101
 - 4.2.2 区块链让实时数据分析成为可能 / 104

4.2.3 区块链揭示交易数据 / 106

4.2.4 区块链揭示用户倾向 / 109

第5章 | **大数据在区块链网络中的交易 / 111**
CHAPTER FIVE

5.1 大数据交易中的瓶颈 / 111

5.1.1 数据提供方有哪些疑虑 / 112

5.1.2 数据需求方有哪些疑虑 / 113

5.1.3 监管层有哪些疑虑 / 115

5.2 区块链破解大数据交易难题的方法 / 116

5.2.1 利用不对称加密技术,对售卖数据签名 / 116

5.2.2 静态数据隔离验证,保护用户隐私数据 / 118

5.2.3 以动态数据最小交易单元为交易标准 / 120

第6章 | **区块链智能合约和大数据促进社会
共治 / 122**
CHAPTER SIX

6.1 智能合约概况 / 122

6.1.1 什么是智能合约 / 122

6.1.2 智能合约与区块链的关系 / 125

6.1.3 区块链智能合约的构建、存储和执行 / 127

6.2 区块链与大数据促进社会共治方法 / 131

6.2.1 基于大数据的信用体系建设是社会共治的基础 / 131

6.2.2 智能合约通过代码和数据实现程序操作自动化 / 134

6.2.3 区块链带来共享、共治性质的大数据 / 136

6.2.4 智能合约破解大数据风控难题 / 138

6.2.5 海尔经营智能合约平台,合同形成从22天减至
2天 / 140

第7章 | **“区块链+大数据”的未来趋势 / 142**
CHAPTER SEVEN

7.1 “区块链+大数据”技术开发趋势 / 142

- 7.1.1 底层系统开发：去中心化+区块链构建+数据安全 / 143
- 7.1.2 上层应用开发：社交区块链+金融区块链+信用区块链 / 145
- 7.2 “区块链+大数据”助力人工智能 / 148
 - 7.2.1 区块链优化后的大数据：人工智能的优质油田 / 148
 - 7.2.2 区块链重塑生产关系反作用于人工智能 / 151
 - 7.2.3 “区块链+大数据”与人工智能深度融合 / 153
- 7.3 共享经济日益繁荣 / 156
 - 7.3.1 区块链缔造崭新的共享经济 / 157
 - 7.3.2 “区块链+共享经济”促进经济创新 / 159
 - 7.3.3 Smartshare：借区块链为共享经济插上翅膀 / 162
- 7.4 打造新职业 / 164
 - 7.4.1 区块链研究员、高级顾问 / 164
 - 7.4.2 区块链视觉设计师 / 165
 - 7.4.3 区块链数据科学家 / 167
- 7.5 “区块链+大数据”：与欺诈说再见 / 167
 - 7.5.1 “区块链+大数据”：值得信赖的第三方调解人 / 168
 - 7.5.2 如何避免身份欺诈：建立“分享+流动+信用”机制 / 169
 - 7.5.3 反思脸书“数据门”：IT巨头重视区块链数据加密 / 172

第 1 章

大数据发展面临窘境，区块链技术异军突起

随着大数据行业的飞速发展，行业的技术窘境也日益明显。由于技术的深入，人们发现现有的大数据产业面临着优质可用数据少、信息壁垒严重、数据处理有困境、实践应用障碍多、云管理失误多五大困境。而这些困境是由现有技术手段的不足造成的，因此，必须依靠新技术才能够突破大数据行业的瓶颈。

在大数据的发展面临窘境的同时，区块链技术异军突起，经历了三个阶段后成功为各大行业带来了新的曙光。本章将从源头开始分别介绍大数据和区块链技术，并对两者的应用现状进行详细阐述。

1.1 大数据产业发展的三大基石：云计算、物联网、移动互联网技术

大数据从第一次出现，就和移动互联网紧密地联系在一起，可以这么说，正是因为移动互联网的飞速发展带动了互联网数据的海量增长，从而推动了大数据行业的产生和发展。随着大数据的发展，云计算、物联网也随之出现并成为了大数据产业的坚强后盾。在当今“数据为王”的社会中，大数据产业的三大基石都具有举足轻重的作用。

1.1.1 云计算

云计算是指通过从云端来获取所需要的服务内容，所谓“云端”就是指网络资源。一般来说，“云”中的资源是可以无限扩展的，使用者可以随时按需获取和使用这些资源，也可以随时扩展资源内容，然后按照对资源的使用情况付费。由于云计算的这种特性类似于日常生活中的水电资源服务，因此它也被称作 IT 基础设施。

如果将云计算的概念扩大化，把它扩张到服务领域，那么所有通过网络来满足用户需求并且易扩展的服务都可以称作云计算，这种服务可以是互联网相关的硬件、软件，也可以是存储、下载等其他服务。

云计算的典型应用就是苹果 iCloud。在 iCloud 上，苹果用户不仅可以上传各种资料以节省硬盘空间，还能够同步不同苹果设备中的文件、日程等，如果用户在某一台设备上对云端的资料进行了修改，iCloud 还能够帮助用户同步到其他苹果设备并对旧文件进行备份以备用户需要。iCloud 为用户提供了免费的 5G 云端硬盘空间，如果用户有需求，可以付费扩容并享受更多服务。

从 iCloud 的使用模式可以看出，云计算服务提供了一种分布式架构——多个智能终端共同同步云端资料，而大数据的应用处理必然无法依靠单独的计算机，必须使用分布式架构，因此云计算刚好能够为大数据处理提供便利条件。

从技术上看，大数据和云计算的关系就像一枚硬币的正反面一样相辅相成、密不可分。依托云计算的分布式架构和云端存储、虚拟化技术，大数据能够充分发挥它对海量数据的挖掘能力。从整体上看，云计算为大数据处理提供了计算资源的底层架构，是上层数据分析处理软件的基础。

那么，云计算为何能够帮助大数据将庞大的数据信息转化成经济效益呢？这里主要包括以下四个方面的原因，如图 1-1 所示。

1. 作为提取大数据的前提

在数据量不断增长的信息社会，获得足够多的数据才是企业从大数据中获

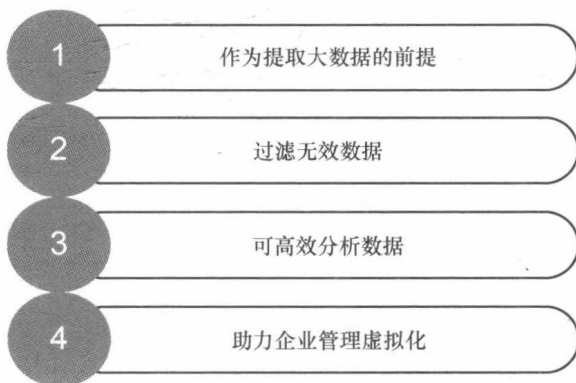


图 1-1 云计算帮助大数据的具体表现

得利益的前提。而想要提取出大数据，来自于各种云端强大的云计算能力必不可少。云端不仅为提取大数据提供了足够的硬盘空间，还能够以较低的成本提取尽量多的数据资源，这一点在大数据产品普遍偏贵的条件下显得尤为重要。

2. 过滤无效数据

在大数据的初次收集中，有接近 90% 的数据属于无效数据，这是由互联网数据本身的特点决定的。既然无法保证数据全部有效，那么就必须要找到一种技术过滤掉无效数据。一般来说需要重点过滤掉的无效数据有两大类，一是大量的临时缓存信息，二是公司防火墙外的网络数据。

由于云计算可以按照需求进行扩展计算和存储资源，所以经过一定的设计后云计算就可用来过滤这些无效数据，常见的公有云就是用于过滤来自公司防火墙外部的无效网络数据的最佳工具。

3. 可高效分析数据

云计算能够为大数据计算提供分布式软件处理方式，用以高效快速地进行数据分析。如果将公有云和私有云结合，就可以在数据分析完成后利用私有云将数据分析结果导入公司内部，方便公司进行下一步的运营决策。

4. 助力企业管理虚拟化

随着市场更新的速度也越来越快，企业管理模式也追求突破有形界限的虚拟

化管理，希望在有限的资源条件下实现资源效率的最大化。由于云计算就是硬件资源的虚拟化，因此当企业运用大数据分析结果指导决策时，如果加上云平台的使用，就能够通过云端应用决策指导所需软件，将决策顺利转化到企业现有的管理系统中，助力企业管理虚拟化。

从美国国家标准与技术研究院对云计算的定义来看，云计算是一种按使用量付费的服务模式。由于它能够快速为用户提供资源，减少交互所需步骤和时间，用于计算时能够实现每秒 10^{14} 次的运算速度，可用来模拟核爆炸、预测市场等。

从用途上看，云计算的应用潜力和大数据不谋而合，云计算与大数据如同手心手背的关系，二者相辅相成。云计算的存在，为大数据技术挖掘数据背后的价值提供了平台。

简单来说，大数据拥有三层架构体系，包括数据存储体系、数据处理体系 and 数据分析体系。数据存储体系是大数据收集并存储数据资源的支撑；数据处理体系包括无效数据的过滤和基础建模等工作；数据分析体系则会根据具体情况做出结论预测，产生相应价值。

云计算能够从存储到处理再到分析给大数据技术提供全面的技术支持，云计算的并行计算和分布式计算能力都在大数据体系中具有不可或缺的重要作用。除了技术层面，云计算还能够利用其分布式架构的特点，极大地降低企业在挖掘数据背后价值时的成本投入。

云计算通过自身和大数据的重合特点以及强大的运算能力为大数据技术提供了发展的平台。虽然云计算概念比大数据概念提出得早，但正是大数据的出现让云计算的优势有了发挥的空间。与此同时，云计算和大数据结合也为大数据提供了牢固的基石，让大数据能够在技术上充分发挥数据的价值。

1.1.2 物联网

物联网概念由麻省理工学院的 Kevin Ashton 于 1999 年第一次提出，他认为，物联网是通过射频识别（RFID）技术和传感器技术结合运用于日常生活中形成