河北省省级科技计划资助（S&T Program of Hebei）
河北省科普专项项目编号20550301K




# 大数据背景下多模式区块链的认知

闫帅领◎著

中国商务出版社
CCTP
CHINA COMMERCE AND TRADE PRESS

**图书在版编目（CIP）数据**

大数据背景下多模式区块链的认知 / 闫帅领著. -- 北京 : 中国商务出版社, 2022.12

ISBN 978-7-5103-4503-6

Ⅰ. ①大… Ⅱ. ①闫… Ⅲ. ①区块链技术 Ⅳ. ①TP311.135.9

中国国家版本馆CIP数据核字(2023)第014278号

# 大数据背景下多模式区块链的认知

DASHUJU BEIJING XIA DUO MOSHI QU KUAI LIAN DE RENZHI

闫帅领　著

**出　　版**：中国商务出版社

**社　　址**：北京市东城区安定门外大街东后巷28号　　**邮　编**：100710

**责任编辑**：徐文杰　张永生

**总 发 行**：中国商务出版社发行部（010-64208388　64515150）

**网　　址**：http://www.cctpress.com

**印　　刷**：北京九州迅驰传媒文化有限公司

**开　　本**：787毫米 × 1092毫米　1/16

**印　　张**：9.25　　**字　数**：140千字

**版　　次**：2022年12月第1版　　**印　次**：2022年12月第1次印刷

**书　　号**：ISBN 978-7-5103-4503-6

**定　　价**：60.00元

凡所购本版图书如有印装质量问题，请与本社印制部联系（电话：010-64248236）

**版权所有　盗版必究**（盗版侵权举报可发邮件到本社邮箱：cctp@cctpress.com）

# 河北省创新能力提升计划
# 科学普及专项

## “大数据背景下多模式区块链的认知”
## 科普图书创作

### 成员名单

闫帅领　张利民　赵　鑫

宋凌云　房月华　张　蕾

# 前　言

当今快节奏的生活使我们忽略了其中太多重要的数据信息，而要完整地保存繁杂众多的数据信息也并非易事，一方面要解决存储问题，另一方面还要特别关注其数据的真实性。解决该问题的最佳方案无疑是引入区块链技术，因为区块链技术有其独特的技术优势，如无中心、防篡改、可查询等。笔者围绕区块链技术展开，对大数据背景下多模式区块链的认知进行了详细的深入浅出的分析，力争创作一本极具科普意义的图书。本书总共分为五章。

第1章：大数据。本部分主要对大数据的定义进行说明，使读者能够通过对该部分知识的阅读对大数据有更加深入和透彻的理解，发现我们生活周边的数据，并能够从一定程度上学会对数据进行收集、分析和存储，以便以后使用保存的数据，在真正意义上做数据生活的主人。在“初探大数据”阶段，为读者全方位地解析大数据的含义，挖掘大数据背后的秘密，将晦涩难懂的大数据简简单单地呈现在读者的面前，揭开遮在大数据上面的所有面纱。在“生活中的大数据”阶段，分别从日常消费、交通、周边事务变化等方面去捕捉数据，培养读者利用数据生活的习惯，生活产生数据，数据指导生活。在“教你大数据建模”阶段，怎么通过数学思维将收集到的数据进行保存是大数据部分的关键，该阶段和读者一起分享了特征关联的数据保存模型，让大家学会整理我们存储的数据信息。

第2章：区块链。本部分围绕区块链的概念，按照层次递进的方式，分别就区块链的认知、区块链的发展以及区块链技术三个方面进行。在区块链认知中，首先回答了读者比较关心的问题，即什么是区块链。简单的回答、翔实的分析一定会让读者耳目一新。认识区块链的发展史是十分必要的，以图画为引导，展示区块链在近几年的发展状况，梳理出一条清晰的区块链发展脉络，使读者更加深入地了解区块链。区块链使用的技术比较综合，通过抽丝剥茧，将其使用的技术分门别类地进行介绍，让读者从宏观上认识区块链的核心，站在高处俯瞰区块链技术。

第3章：生活中的多模式区块链。本部分以区块链思维为主导，培养读者的区块链思维，在生活和生产中关注类区块链的现象和规律，发现生活中存在的区块链。区块链难以理解，而我们生活中又常常在使用区块链。能够将潜在的事务发觉并应用是一种能力的体现，区块链模式会不会适合我们生活中的大数据呢？本部分的第二阶段对此进行了回答，从大数据角度出发，建立区块链，构造出多模式的区块链。接着对具体如何进行数据在区块链的环境中“上链”操作逐一进行了分析，对区块链中数据的相互关系与影响的展示，让读者了解到区块链内部的数据。尝试利用区块链在生活中的应用，提高我们现有的生活品质是本部分的最后一个分享。

第4章：区块链的高级应用。区块链的乐趣在于其使用，利用区块链技术去解决我们现存的一些问题，而这部分的问题都是之前无法解决或者解决不彻底的。该部分重点讲述了利用区块链技术解决的几类问题，每一类问题都有其显著的代表性。目前来说，区块链技术是解决以下几类问题的最好的方案，如保密问题、信任问题、信息不对称问题、评价问题等。相信在不久的将来，区块链技术的应用势必会普及开来，届时我们现有的很多问题都会迎刃而解，极大地推进我们人类的生产、生活。区块链技术将会掀起新一轮的技术浪潮革命，驱使现有的信息网络向价值网络转变。

第5章：区块链的应用思考。区块链在应用之后所带来的问题需要进一步的去讨论和分析。该部分通过一个成绩预警的实例向读者展示了应用区块链技术后的几项思考方向：第一是数学建模的问题，优质的数学模型是区块链技术应用的核心所在；第二是数据的变化，区块链技术导向的数据自动化的汇聚往往是其技术突破的一个难点；第三是概率预测，不以单个节点为评价依据的区块链技术，其概率更加的趋于稳定。只有在使用了区块链技术之后，才会引发一些问题的思考，希望通过本章在引导读者参与区块链技术革新上能够起到抛砖引玉的功能。

本书由闫帅领所带领的区块链团队成员编写，并得到了河北省科学技术厅的大力支持，被列入了2020年河北省创新能力提升计划项目。

由于编者水平有限，书中难免还存在一些缺点和错误，殷切希望广大读者批评指正。

2021年7月

# 目　　录


第1章　大数据 …… 001

1.1　初探大数据 …… 001

1.2　大数据的特点 …… 004

1.3　生活中的大数据 …… 007

1.4　教你大数据建模 …… 011

第2章　区块链 …… 016

2.1　走进区块链 …… 016

2.2　为什么需要区块链 …… 017

2.3　区块链解决的问题 …… 020

2.4　区块链的发展 …… 026

2.5　区块链的分类 …… 028

2.6　区块链技术讨论 …… 030

第3章　生活中的多模式区块链 …… 042

3.1　生活中的区块链 …… 042

3.2　大数据下的区块链模式 …… 051

3.3　多模式区块的生成 …… 056

3.4　我会“上链” …… 059

3.5　区块链对生活的改变 …… 062

第4章　区块链的高级应用 …… 067

4.1　区块链在Ad Hoc网络中的应用 …… 067

4.2　区块链在教学中的应用 …… 104

4.3　区块链在滑冰运动中的应用 …… 110

4.4　区块链在矿山中的应用 …… 113

第5章　区块链的应用思考 …… 132
5.1　数学建模 …… 132
5.2　数据变化概率 …… 133
5.3　概率预测 …… 133
5.4　小结 …… 134

参考文献…… 135

# 第1章　大数据

## 1.1　初探大数据

数据包含的种类繁多，但是如何衡量各种不同形式的数据呢？为了使用和存储，我们将计算机数据的存储单位作为衡量数据的单位。以比特为最小单位，然后依次是Byte、KB、MB、GB、TB、PB、EB、ZB、YB、BB、NB、DB。我们可以用一张图来表示衡量数据单位的大小关系，如图1.1所示。

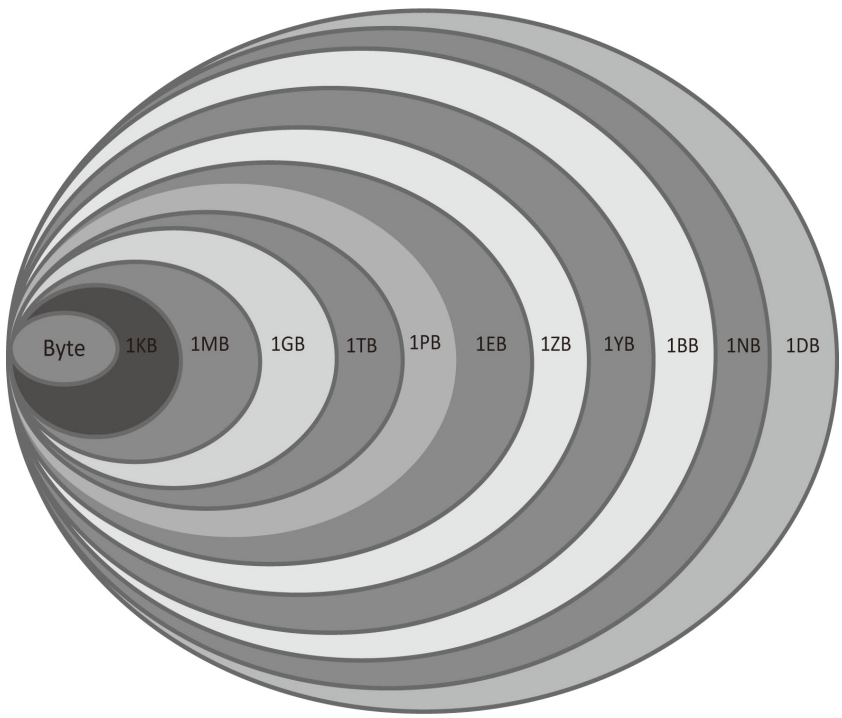



图1.1　数据量的大小关系

这些单位之间的进率为1024，则有：

1 Byte =8 Bits

1 KB = 1,024 Bytes

1 MB = 1,024 KB

1 GB = 1,024 MB

1 TB = 1,024 GB

1 PB = 1,024 TB

1 EB = 1,024 PB

1 ZB = 1,024 EB

1 YB = 1,024 ZB

1 BB = 1,024 YB

1 NB = 1,024 BB

1 DB = 1,024 NB

全称：

1 Bit（比特）=Binary Digit

1 Byte（字节）=8 Bits

1 Kilobyte=1,024 Bytes

1 Megabyte=1,024 Kilobytes

1 Gigabyte=1,024 Megabytes

1Terabyte=1,024 Gigabytes

1 Petabyte=1,024 Terabytes

1 Exabyte=1,024 Petabytes

1 Zettabyte=1,024 Exabytes

1 Yottabyte=1,024 Zettabytes

1Nonabyte=1,024 Yottabytes

1 Doggbyte=1,024 Nonabytes

现在可以使用计算机的数据描述我们生活中遇到的较大数据信息了，这样就大大方便了我们对数据的表达，也更加便捷了我们的后续处理。例如，现在无处不在的“天网”，也就是安装的摄像头。几乎在城市的每一个路口，在商场门口、在公司的办公区域、在公园等场所，我们都能看到

摄像头在工作。平时它们默默地记录着我们生活的点点滴滴，当我们需要还原已经发生过的事情时，它们就发挥了不可替代的作用，这比较像时光倒流，回顾我们过去的生活一样。

那么，摄像头是如何工作的呢？摄像头把所监控范围内的影像信息拍摄下来，然后再把图像和声音信息在计算机内部转化为对应的数据信息，将转化后的数据信息保存在计算机中。当需要回看当初的影像信息时，只需要将计算机中存储的对应数据信息再转化回图像和声音，就可以看到之前发生过的事情，如图1.2所示。

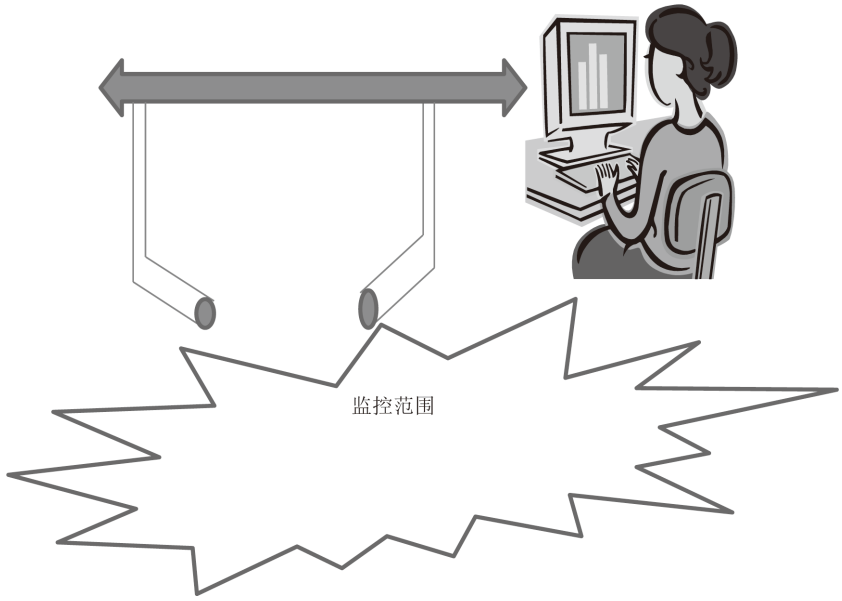



**图1.2　“天网”工作原理**

一个高清摄像头，一个小时大约产生4294967296字节（4GB）左右的数据信息，一天的产生的数据量为96GB，3天产生的数据大约为288GB，一个月大约2TB，一年就是1PB的数据信息。计算机的存储空间是有限的，目前一般的电脑存储空间为1TB，除去操作系统所占有以及其他软件信息占有的空间，剩下的基本就是800GB左右。也就是说，一般的计算机无法存储超过半个月的视频数据信息。所以，大家所看到的摄像头的数据信息的保存时间一般为3天，超过3天会删除之前的数据信息以腾空计算机的存储空间。

由此可见，当数据信息量较大的时候，计算机都会因无法存储而不得不删除过期的数据信息，然而在较大数据信息当中可能蕴藏着很多有价值的数据信息，也可能是一些无用的垃圾数据。那么，怎么区分这些呢？在此，我们一起来看看“大数据”这个词汇。其实所谓的大数据就是针对上面的含有有价值数据信息的海量数据的一种表达。大数据并不是单纯指数据信息量大，它有很多制约因素，不过我们可以认为，数据量大不一定是大数据，但是数据量小肯定不是大数据，如图1.3所示。

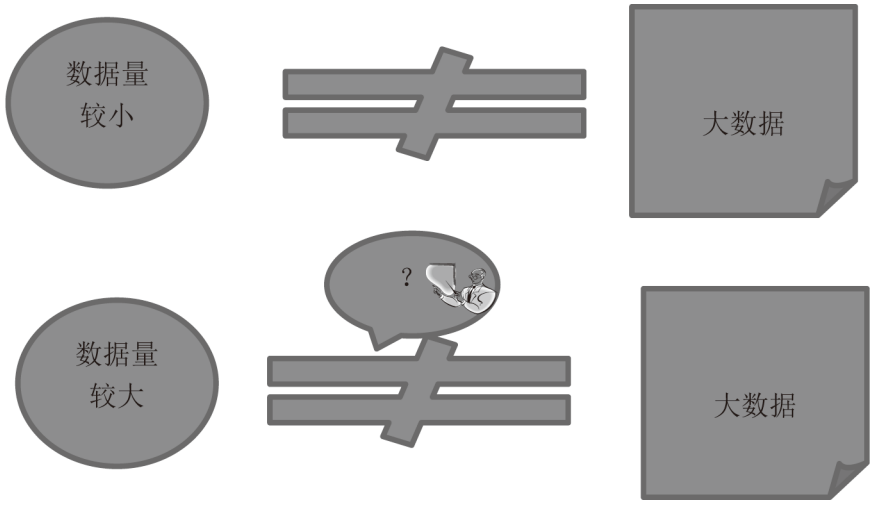



图1.3　大数据初步理解

## 1.2　大数据的特点

为了更加全面地理解大数据的定义，我们不妨从挖掘有价值数据信息的角度去思考，一般有价值的数据信息是汇总各种情况的结果。由此可见，数据的获得如果太过单一，必然会失去一些重要的信息，所以数据的来源一定要多样。有价值的信息一定是获得后才能够感知到，这样就要求信息一定是能够被有效整理的，长期提取不到的有价值信息即便有也等于没有。核心价值数据信息相较于蕴含其中的众多数据信息来说，量是极其微小的，这就涉及低蕴含的考量。大数据的特征如图1.4所示。

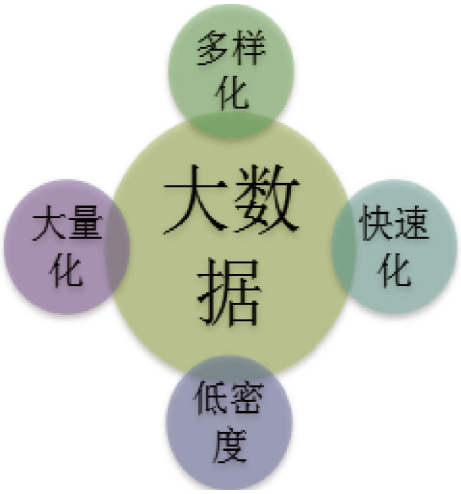



图1.4　大数据特征

1．大量化（volume）

信息量大是大数据的首要特征。目前，全球总共拥有35ZB左右的数据量。如何处理、分析海量的数据信息，是摆在大数据面前的最主要的问题。在现实生活中，往往需要通过消息去反映一个事物的发展变化或者状态信息，并且只有尽可能多的信息才能描述更贴近真实的情况。

如需要描述一个人的面部表情变化，其中的信息量需要包含面部各个器官的微小变化、眼睛的聚集程度、眼光的离散规律、面部肌肉的跳动、呼吸的次数与周边肌肉的关系等，这些都需要进行考虑，并且需要及时地捕捉到每一个变化量，而这些变化量就是面部变化的信息值。很明显，这个信息量是巨大的。假如再加上其他方面对面部的影响，如心脏的跳动与心情的变化对其的影响、天气的影响、温度的影响、湿度的影响、噪声的影响等，那么势必会使分析面部表情的信息量急剧增加。实际上，上述因素确实是研究面部表情变化需要考虑的。因此，面部表情变化所产生的数据就是大量化的。

2．多样化（variety）

信息的出现不是以单一的某种形式，如常见的文本形式，它往往是由多种样式组合而成的，如文本、表格、订单、日志、音频、视频等多种样式。正是因为信息形式的多样化，造成了研究该类信息时无法使用现有

的软件进行处理分析，而必须考虑综合因素，找到一种可以概括其全面的新的形式进行处理，而这种新的处理形式就是大数据所需要的。如目前的Hadoop分布式处理模式，信息的处理打破常规的在一处进行汇总分析，把数据分散到很多的计算机上去完成，让特定的计算机去处理特定的信息，发挥各自的长处，最终达到大数据信息处理操作的目的。

数据格式必须具有多种展现形式，单一的数据格式即使信息量再大，也无法称为大数据。也正是因为大数据的多样化，大数据中所包含的信息来自多个不同的通道，信息更为综合，所能反映出的事物更加全面。

3．快速化（velocity）

大数据的信息量大，形式多，那么在对于信息的处理上也并不是无限制容许长时间地进行，而有些大数据信息是需要在短时间内要求处理结果的。这样就存在信息量大且复杂与快速处理所带来的矛盾问题。而大数据的快速化主要指数据信息必须要有很强的变化性，获取的数据是在不断发展变化的，在一定时间内数据会失去它的效能。对于静止不变的数据信息，那样可以长时间地分析处理，即便数据再复杂、处理再困难，也构不成大数据的快速化。

大数据反映的是不断变化事物的信息描述，正因为这一特点，只有不断变化的数据才具有实际意义。变化速度越快，表明数据的可靠性越高，数据的有效性越好。

4．价值密度低（value）

价值密度低主要是由海量数据造成的，而有效信息数据就隐藏在海量数据中。大数据中的多数信息是垃圾数据，只有个别的数据才是真正能反映出事物的特征值。如何能高效地提取到有效的数据，在大数据的研究中也一直是很重要的一部分。但是，不可能采集的数据都是有效的，大数据就是在海量的垃圾数据中寻找符合一定要求的数据信息，进而从中找到我们所希望得到的信息。

数据挖掘正是基于这一大数据特征而存在的数据提取技术。在数据仓库中，寻找具有一定特征规律的数据，发现杂乱无章的数据中潜在的有规律的行为，为决策者提供可靠的数据支撑。对于数据仓库是不是值得挖掘

以及是否能挖掘出有价值的信息，在没有实施挖掘之前，都无法下定论。这也正是数据挖掘的奥秘所在。

## 1.3　生活中的大数据

数据是从何时开始产生的呢？这需要追溯到人类的诞生时期。自有人类以来就存在着各种各样的数据，只不过那时的人类并没有意识到这些数据的存在，简单的生活方式注定对产生的数据视而不见。然而，随着人类的进步，数据慢慢出现在了人们的意识中，其中以结绳记事（图1.5）为代表，人类才会使用数据，进而也为生产、生活带来了巨大的便利。正是借助于有效的数据，我们人类才能迅速发展并达到了当今的文明。

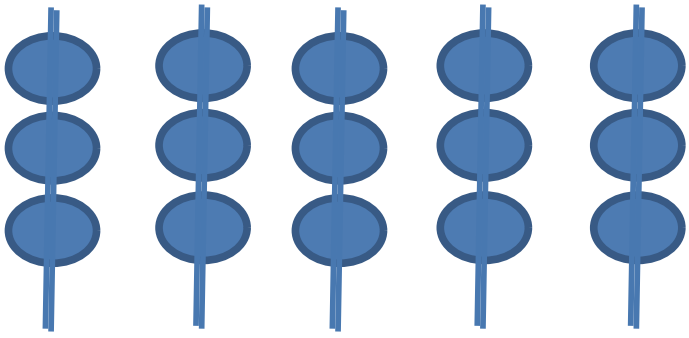


图1.5　结绳记事

在日新月异的今天，数据更是每天以令人吃惊的程度产生。如图1.6所示，大数据涉及很多方面，即大数据中可以包含各行各业的数据信息，同时表明大数据信息可以通过多种手段进行获取，如Wi-Fi、电脑网络等，从而在综合数据信息中发掘一些人们没有发现的潜在规律或者结果。麦肯锡全球研究所认为，大数据是一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合。它需要新处理模式才能具有洞察发现力、流程优化能力和更强的决策力来适应海量、多样化和高增长率的信息资产。

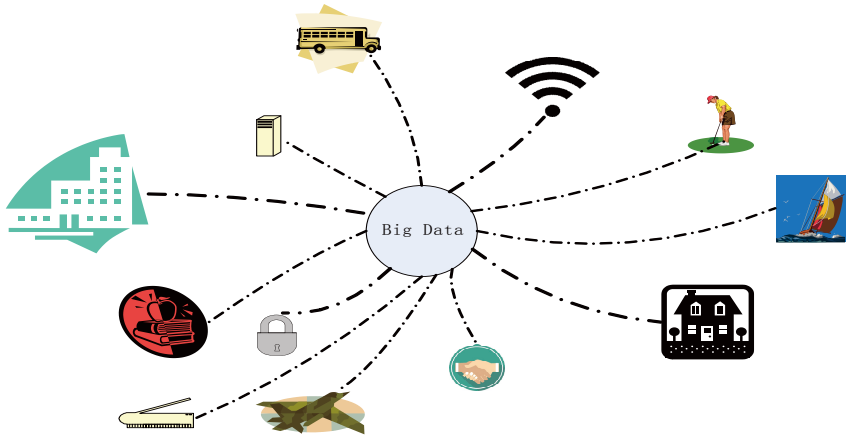



图1.6　大数据示意图

以一个简单的汽车出行为例，对数据的产生进行观察。如图1.7所示，有一辆需要通过高速公路出行的汽车。首先，汽车在驶入高速公路时会产生入口信息，该信息会记录下汽车的车牌和驶入时间；当行驶在高速公路上时会有阶段性的计费信息和速度检测信息，该信息会根据驶入地点扣除高速费的消费状况信息和通过瞬时速度而进行与最高限速相比较而产生的检测信息；而根据车载导航或手机导航，在经过每一个高速公路上的路口时会产生位置提示和建议车道行驶提示以及距离目的地的路程信息；驶下高速时会产生出口信息，该信息记录下汽车的车牌、驶入时间以及计算出该车行驶的路程信息。在汽车行驶高速公路的整个过程中，期间所产生的数据包含数字、图像、语音、记录、文本等。

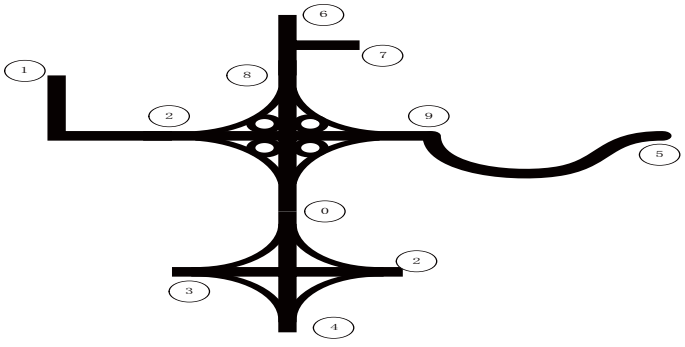



图1.7　高速公路示意图

其大数据特点对应如下：

（1）大量化：高速公路收集到的汽车行驶速度信息、路程计算数据信息、车牌信息等，由此产生海量的高速公路数据信息。

（2）多样化：高速公路中的数据信息包含有路程信息、车速信、路程计算数据、图像信息等，且各种数据信息之间有相互的关联。

（3）快速化：海量的高速公路数据信息主要支撑汽车高速公路行驶的计费、是否违规驾驶、路况监测等，能够在汽车驶离高速时迅速及时地提取到行驶的路程信息，可以在监控室远程实时地监测高速路面的情况等，这些都是大数据快速化的表现。

（4）价值密度低：能够使用的高速公路数据信息与收集到的海量数据信息相比，微乎其微，整体使用率不足十万分之一。只有足够的数据量才能支持现有的便捷应用。

图1.8是排球比赛。当今的体育赛事如果想取得比较好的结果，离开数据分析是万万不行的，那么比赛的数据来自哪里呢？第一，比赛前对方队员的个人能力数据，包括出场次数、擅长扣球的方位和角度等；第二，比赛中运行数据，在比赛过程中统计双发发球失误的次数，在哪一个轮次的时候胜球的概率较大，每一场次各个队员的扣球次数等；第三，比赛后的结果数据，依据赛前的个人的身体状况，逐一对比每一个队员的数据，从而形成一个发挥因果关系表，调整个人的状态，这样可以指导以后的每一场比赛，力争最大概率的胜利。

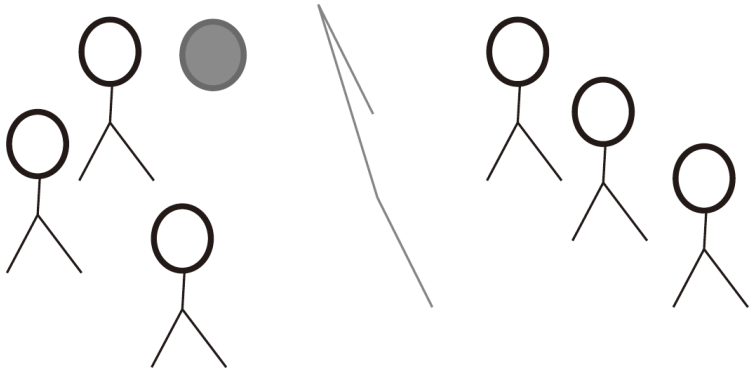


图1.8　排球比赛

其大数据特点对应如下：

（1）大量化：排球比赛中，运动员的每一次发球、扣球、拦网、传球、失误等，以及相对应的次数、位置、力量大小、情绪值等均产生了数据信息，该信息量较大。

（2）多样化：数据信息从多个渠道获得，发球为一个获得渠道；扣球为一个获得渠道；拦网为一个获得渠道等。不同的数据信息交织在一起，相互制约、相互关联。

（3）快速化：能够从大量的数据信息中获得某一个运动员的现场发挥状态，能够帮助教练及时调整队员，或者更改排球的打法和战术，以期能够获得最佳的效果。这种折射结果需要迅速获得。

（4）价值密度低：判断一种战术的好坏或者一个运动员的状态信息，并不是由某一次失误或者效果较差而决定。往往需要一个场次的观察以及大量对应数据信息的印证，才能得出好或者坏的结果，好比需要一个人做1000道百分的题，而结果只有及格和不及格，给定结果之前必须要对这1000道题批改完毕。由此可见，价值密度之低。

图1.9为某旅游景区。在旅游景区中可以按照天、周、月或者年为单位统计游客数据，其中以旅游团的形式到达，以自驾的方式到达或者其他等方式进行数据的收集，分析数据中各个年龄段游客的数量。景区内各种商品的销售数据也至关重要，如那些数据在什么时间段内产生，各个商品数据的峰值和低谷值的出现节点等。采集与景区相关的所有数据，势必会得到大量的数据信息，而这些数据信息之间或多或少地存在着一定的关联性，如果能找出一条或者几条重要的关联关系，肯定会大大提高景区的服务质量以及景区整体的收入。

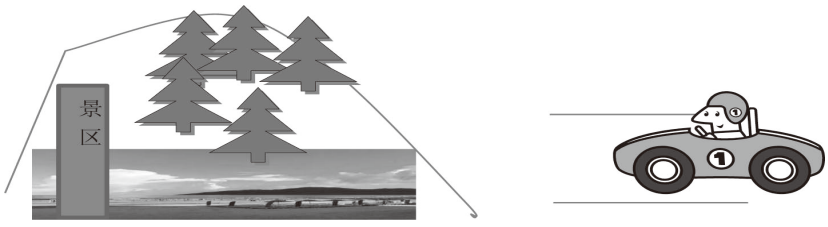



图1.9　某旅游景区

其大数据特点对应如下：

（1）大量化：管理景区的数据主要来源于游客信息的统计，包括游客的年龄、地域、出行方式、时间段以及消费情况等。其中蕴含着大量数据信息，这方面的数据信息是海量的。

（2）多样化：管理景区的数据信息要从各个管理者那里获得。如需要从停车管理员处获得车辆信息；需要从售票处获得人次数和大致年龄段信息；需要从景区商品经营者那里获取销售额信息等，将全部数据信息收集到一起，提供给景区的信息管理者。这些形成了该景区的大数据的多样化。

（3）快速化：在收集了景区的所有可能数据信息后能够快速得到相应的结果。该结果直接影响景区的管理水平以及游客的体验感，如各个景区经营者是否需要预备更多的商品、商品的类型等；景区的停车场是否需要扩建；景区的售票窗口在什么时间段开放多少个窗口合适。不断优化的景区显然是大数据快速化的效果。

（4）价值密度低：必须事无巨细地收集游客的各种信息，才能够更加全面地对景区的管理给予决策的支撑。景区的一个决策往往是大量数据的信息的支撑，是有价值数据的体现。相对于收集到的数据信息，其有价值数据信息的含量极低。

由上可见，数据并不是指单纯的数字，数据的形式多种多样。社会从来不缺数据，我们拥有强大的数据源，但是如何有效地获取、分析、利用如此浩瀚的数据源一直是人类研究的焦点。只有掌握了数据，利用好数据，发挥数据的应有效用才能进一步地推动人们的生产、生活。

## 1.4　教你大数据建模

“建模”这个词比较陌生，下面首先介绍什么是建模。建模就是搭建起一个模型，然后将现有事务与模型进行比对，如果现有事务符合模型的要求，那么就认定该事务属于模型类型，否则就不属于模型类型。当然，我们都可以通过建模来更好地普及或者应用各种各样的事务，使复杂的事务简单化、步骤化。

最初的建模是房屋，可以先设计出房屋的模型，然后按照房屋的模型进行施工，图1.10为一个二层楼房模型，已经给出了房屋的具体结构，包括窗户、门等，可以根据这种模型很容易地建造出所要求的房子。将该模型给了不同的施工队，如果严格按照模型进行操作，那么最后盖的房子效果是一致的。由此可见，通过模型做到了简单化、步骤化和一致化。

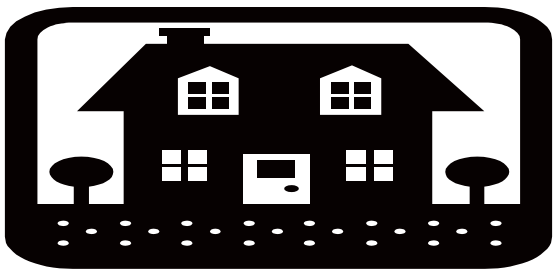


图1.10　房屋模型

大数据是一个抽象概念，它可以容纳整个人类生产、生活，也可以仅仅指人们生活的一个方面。最主要的是利用大数据思维去指导人们的生产、生活，让我们周围的大数据“活”起来，以便能够更好地服务人们，让生活变得轻松舒适。如何让大数据“活”起来呢？那就必然要先收集数据，再整理数据、分析数据，最后使用有价值的数据。我们就以此来建立大数据模型，具体如图1.11所示。

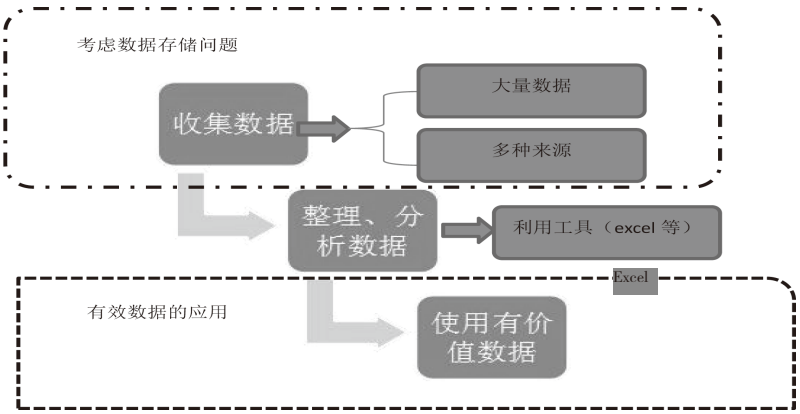



图1.11　大数据建模型

1．收集数据

大数据最重要的一步就是收集数据，然后将其进行存储。根据大数据的特点，在收集数据的时候应注意两点。第一就是数据的信息量一定要大，因为足够大的数据信息量才有可能发现更多的有价值数据信息，数据量太小的话，出现的有价值数据信息可能是假的有价值，从而误导得出一些错误的结果；第二需要从不同数据源得到数据，切忌数据信息来源于一个地方，否则往往可能因为数据考虑得不全面而影响对整个有价值数据的提取，只有从多个方面得到的数据才能够相互佐证，得出有效的有价值的数据信息。

关于数据的存储，可以考虑采用集中式和分布式两种方式，根据数据量的大小以及计算机存储容量的大小进行衡量。

集中式的一种数据存储方式为数据库的存储，目前比较流行的数据库有SQL Server、MySQL、DB2、Oracle等，也就是关系数据库。这种数据存储有着较大的限制，数据之间必须能够写入二维表中，即数据必须是有结构的，否则将无法进行存储。一旦能够进入数据库存储，那么对于后续的数据的整理、分析将大大减少工作量，可以利用结构化查询语言进行各种各样的操作。而对于无结构的数据信息往往采用记事本、Word、音频文件、视频文件等进行存储。能否有效地存储非结构化的数据信息是制约大数据的数据收集的一个瓶颈。

目前比较流行分布式存储方式，所谓的分布式就是让数据信息存储在不同的协调工作的多台计算机中，这些计算机相互协作完成数据的存储工作，这样可以大大扩大单台计算机的存储容量，使大数据能够不断延伸。当数据信息量大时，只需更多的计算机参与进来存储即可，不必对单台计算机进行无限的扩容。

2．整理、分析数据

数据被收集完成后，下一步就是对收集的数据进行整理、分析。海量数据的整理过程：首先，删除明显错误数据，将数据采集过程中有缺失的数据和重复数据删除，以便第一次过滤数据；其次，对数据进行分类，可以按照关联强弱进行划分，建议划分为三类——强关联、中关联和弱关

联，并对其进行数据标记。在此可借助相关软件或者通过编程等方式处理。由于数据量比较大，这里一般不要手动去操作，否则极容易引起数据处理疲劳，也容易出错。建议采用自己熟悉的工具进行，并且根据数据的类型，可以采用不同的软件进行处理，以便达到更好的效果。

在数据的分析过程中，要明确有效数据的类型以及性质，进而有针对性地对海量数据进行筛选。这类似于我们去菜市场买菜，首先要确定自己想要购买的蔬菜，其次去除自己不需要的，最后在剩余的自己希望的蔬菜中进行挑选，以期达到购买的目的，如图1.12所示。

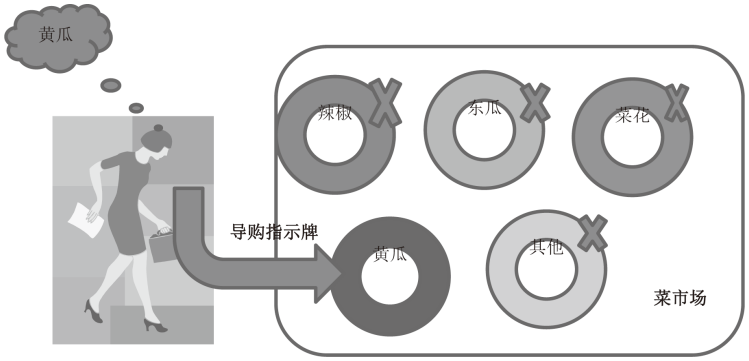



**图1.12　菜市场挑选蔬菜**

如图1.12所示，在偌大的菜市场中，首先，排除不需要的辣椒、冬瓜、菜花等，留存自己所需要的黄瓜。其次，集中在黄瓜的摊位前进行询价和挑选自己喜欢的黄瓜品种。这样，菜市场就是我们的大数据，而顾客的黄瓜即是有价值的数据信息，顾客进入菜市场寻找黄瓜就是在进行数据的整理、分析，而最终的黄瓜摊位则是有价值的数据信息，需要从海量的数据信息中找到需要获取的东西。而进入菜市场后，利用导购指示牌能够更快地找到所需要的蔬菜就是我们在数据处理过程中需要采用的工具，那样会更加便捷地进行数据的整理、分析。

3．使用有价值的数据

在经过整理、分析后，就是在海量数据中筛选有价值的数据信息了。该类数据信息往往能够支撑起对某一个事务规律的发现或者印证。在此，我们以一个餐馆的销售大数据为例，通过收集顾客的消费信息、天气状

况、物价波动等数据信息，再经过整理、分析，梳理出其中的哪类顾客在什么时间段容易到店里消费，或者哪一类顾客喜欢以外卖的形式消费等。这样，这些支撑起顾客趋向性消费的数据信息就是本餐馆大数据中有价值的数据信息，应当从海量数据信息中提取出来，找出规律，然后根据相关的规律性进行餐馆的进一步优化经营，如图1.13所示。

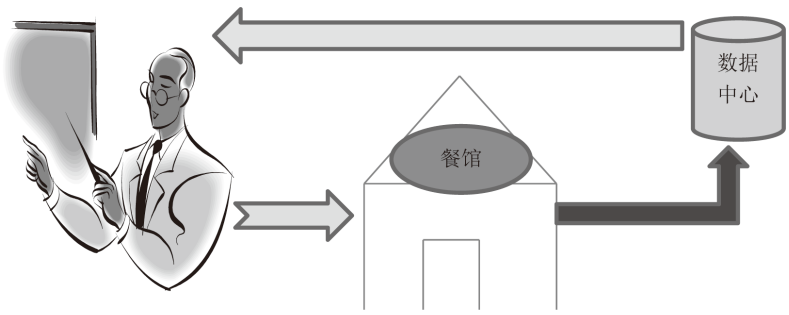



**图1.13　餐馆大数据有价值数据信息提取框架**

有顾客消费大数据支撑的餐馆将会在餐馆性能提高、餐馆菜品的发展、未来规划等方面都有据可依。在现如今餐馆竞争激烈的时代，与时俱进的餐馆经营策略尤为重要，而要做到及时快速地紧跟时代步伐，唯有采用大数据战略。未来的餐馆行业竞争，其实也是大数据信息的竞争，谁能最先掌握顾客的消费现状、消费心理，谁就能立于不败之地。

**本章思考：**

我们生活中还有哪些大数据，你能找到吗？不妨找找看。

# 第2章　区块链

## 2.1　走进区块链

在走进区块链之前，我们首先需要认识区块，区块链是由一个个的区块构成的，也就是说，区块是构成区块链的基本要素。那么，区块是什么样子的呢？如图2.1所示。

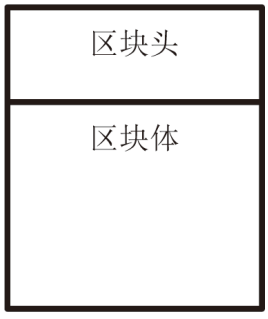



图2.1　区块示意图

区块头中放置的是一些状态信息，也就是属性值等；而区块体是存放数据的。区块头和区块体之间有一个连接点，该连接点存放了数据信息的汇总加密结果值。区块体中可以存放很多条记录信息，在区块体内有着一定的存储结构。

那么，一个区块是如何构成区块链的呢？首先我们来看一个游戏——老鹰抓小鸡，如图2.2所示。在这个游戏当中，“鸡妈妈”掩护“小鸡”的队伍就是一种链式结构，它的形成就是后面的人总是抓住前面的人，始终保持着一定的顺序不变。如果我们把一个个“小鸡”当作区块，那么整个这种链式结构就是一个区块链。我们不难发现，在这种结构中总是一个一个链接的，也就是说先有第一个，然后是第二个，第三个，第四个……以

此类推，形成这种链接。同理，区块链也必须一个区块一个区块地上链，进而形成一个牢固的区块链，区块链结构如图2.3所示。

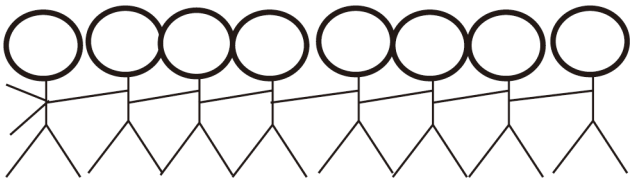


图2.2　“老鹰捉小鸡”游戏

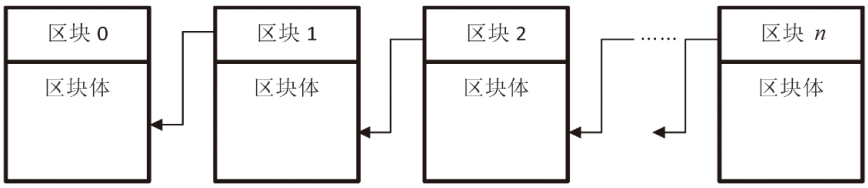



图2.3　区块链结构

## 2.2　为什么需要区块链

“区块链”是2008年才出现的一个技术名词，它来自一个名叫中本聪的人，他于当年的11月1日发表了一篇关于数字货币的文章《比特币：一种点对点的数字货币》。之所以创造出这个名词，主要是因为目前网络的发展十分迅速，很多人都在利用网络进行交易，可是在网络交易的时候因为看不到实际的人，那么交易的双方会出现相互不信任的情况。最开始的解决办法为仿照现实社会的思维，互相不信任的双方要想进行交易，那么必须找一个可靠的双方都信任的第三方进行担保或者见证，但是这样就会有一种潜在的风险，一旦这个第三方出现问题或者有恶意的行为，那么交易就会面临很大的不确定性。

比如，张三需要向李四借500元钱，而他们之间又不认识，那么张三就会找一个认识李四的人叫马六，如图2.4所示，让他去向李四进行担保。而李四由于和马六很熟悉，自然就相信马六，将钱借给张三。其中马六就担当了一种第三方的角色，而一旦马六消失，那么张三向李四借钱的事情

就再无别人知道了，张三是否还李四的钱完全取决于张三自身，如图2.5所示。很明显，这种行为框架自然存在着问题。

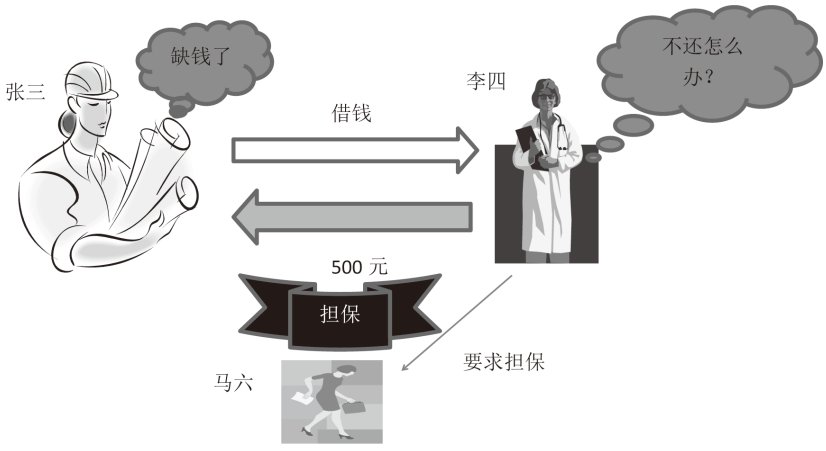



图2.4　常规担保式的借钱

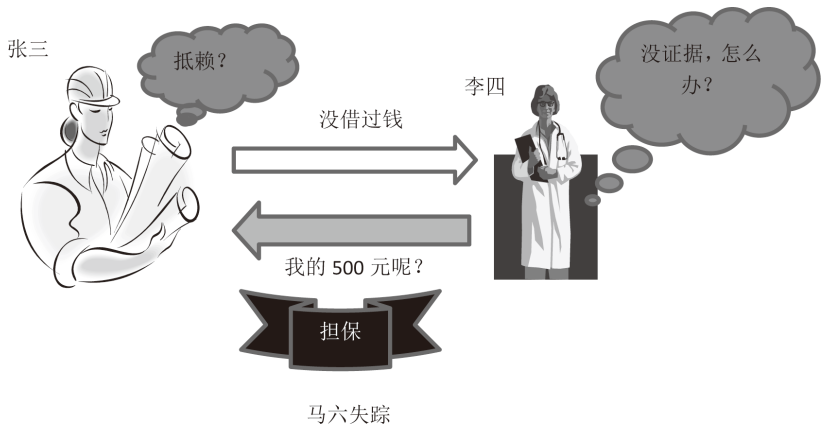



图2.5　常规担保式借钱可能出现的情况

当图2.5的情况出现的时候，李四真的找不到其他证据能够说明张三确实借过钱，这时便无法证明借钱的事情发生过。而李四只能无奈地接受这样的事实。因为除了担保人，再没有别人知道借钱的事情了。那么，如果是借钱的人忘记了呢？他本身不想抵赖，可是他真的忘记了借过别人钱，而此时恰恰担保人又失踪了，这种情况也是经常发生的。再有一种情况是如果借钱人确实借过钱，他也记得且诚信，他及时地把欠款归还给当事人

了，可是如果出借人忘记他已经还了，又重新去要当初借出的钱，而担保人又失踪了，无法证明还钱的事情，这样就会出现重复还钱的现象。这种重复还钱的现象在经济学领域被称为“双花”问题。由此可见，担保人式的借钱，一旦担保人出现问题，将引起一系列的问题，导致问题不可能得到很好的解决，承担问题的瓶颈明显。

正是为了解决这种薄弱第三方的问题，区块链就出现了。那么，区块链到底是什么呢?

对于区块链定义的理解，很多人将区块链定义为一种分布式的账本，也就是说，区块链类似于账本集合，通过重复分开记录以保证数据的证据性和不可抵赖性，此时我们首先可以把它理解为一种共同见证。为了保证这种证据的有效性，这种见证需要很多人参与实施，如图2.6所示（以借钱的四账本为例）。

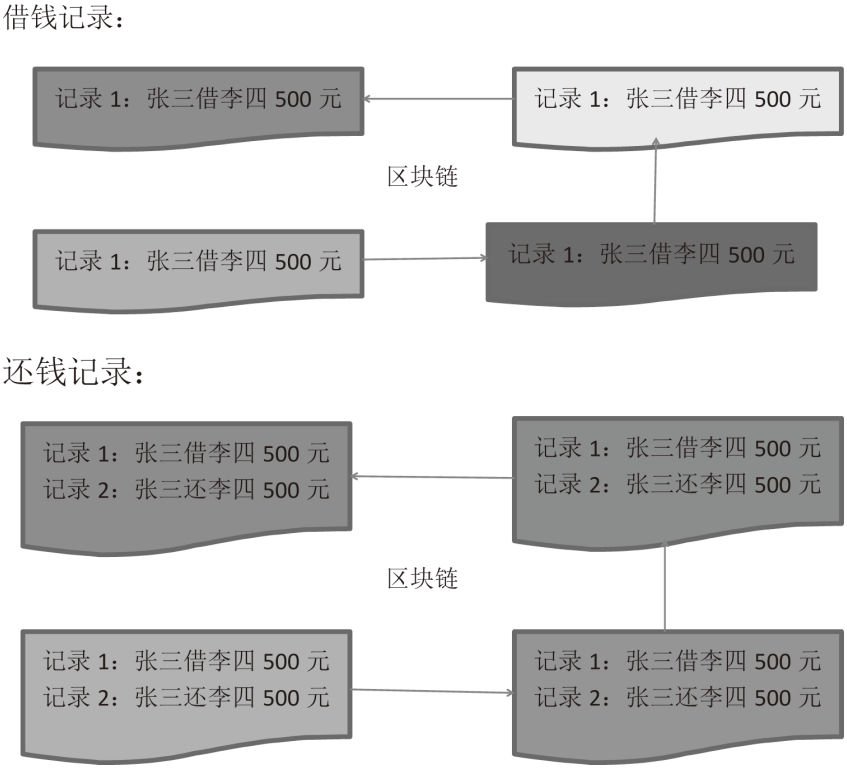



图2.6　区块链解决借钱问题

在图2.6中，展示了四个区块的区块链，该区块链记录的为张三借李四500元，将借钱的事情作为一条记录，记录在了一个区块当中。每一个区块的形成需要很多人去证明并由双方确认，一旦区块形成便为一种固定的记录方式，任何人不得对账本中的内容进行更改，在账本形成后，所有人都可以查看账本以确认是否发生过该事件。当然，张三还李四500元的事件也同样作为一条记录被保存在了区块当中，在还钱的过程中可由一定数量的人见证并确认该事件确实发生过，之后再记录到区块链当中，一旦进入区块链就称为已发生过的事实，不得篡改和否认。

由此可见，区块链涉及几个主要问题：一是如何形成区块？因为区块的形成必须要有一个人去实施，究竟让谁负责一个区块的形成呢？信息记录发生在每时每刻，显然由专门的一个人负责区块的形成是不现实的。二是在区块形成的过程中，由谁去验证区块内记录数据是正确的呢？这个证明人是否可靠呢？假如证明人不可靠怎么办呢？解决证明人问题也是十分关键的，因为它决定记录的准确性，一旦区块内数据记录的信息是错误的，势必会造成十分严重的后果。三是如何保证数据的不可更改呢？如果记录在区块中的数据被别人恶意篡改，区块链将失去其存在的意义，保证数据的可靠性是一个基本要求。另外，区块链中还涉及一些个人隐私保密、区块形成时间、区块链的安全性等问题。解决以上的这些问题，需要一些综合的技术进行实现，我们稍后再和大家一起分享。

## 2.3 区块链解决的问题

区块链是一个综合性很强的技术融合体，拥有区块链技术将会解决现有的一些令人很苦恼的问题。

1．信息不对称问题

现在李明想买一辆汽车，但是由于资金有限，所以他决定先购买一辆二手汽车。那么这其中就会出现一些问题，首先是二手汽车的价格问题，正常情况下质量好一些的二手汽车自然价格高一些，而质量差一些或者比较旧一些的二手汽车自然要便宜一些。如果单从汽车的时间考虑就比较简

单了，但是影响价格的还有一个重要因素就是该二手汽车是否出过事故，事故车相比一般二手车的价格要低很多。但是，汽车的事故目前只有汽车拥有者本身知道，而购买者并不清楚，如图2.7所示。

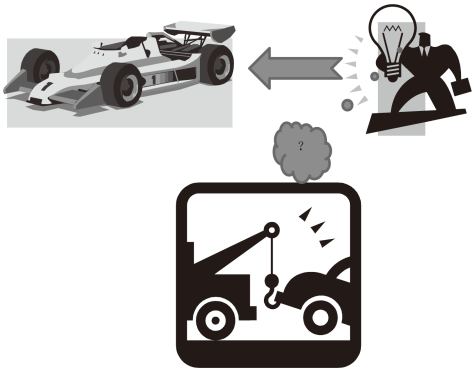


图2.7　信息不对称举例

图2.8中显示了出售的二手汽车在以往行驶过程中出现事故的上链过程，每一次事故都会形成一个事实事故区块链，以如实记录该车辆的行为轨迹。该记录可由交管部门、事故双方或单方车主、见证者等多方组成，当见证者足够时，便可将保存记录的区块链接到该辆汽车的事故区块链中，保存在事故区块链中的记录信息真实可靠。

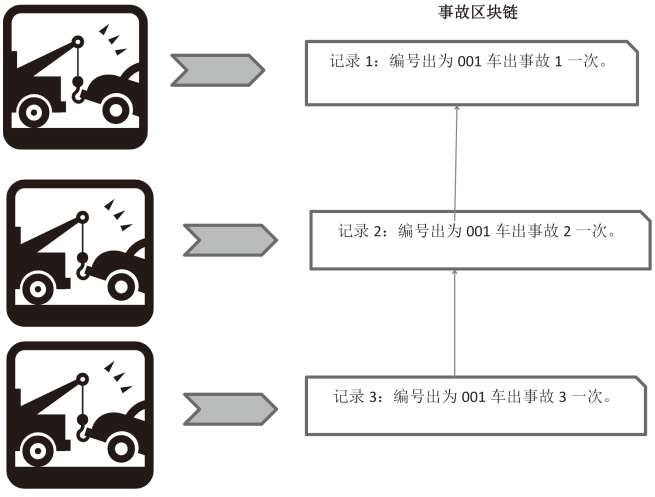



图2.8　事故区块链形成

图2.9展示了拥有事故区块链的二手汽车的售卖过程。购买者可通过查询该辆汽车的事故区块链以确定该汽车出过多少次事故，以及每次事故造成的损伤等信息，再依据其他的一些信息确定该辆二手汽车目前的价格。因为区块链中记录是经过多次确认真实可靠的，所以在购买的过程中没有必要再过多考虑事故的真实性问题，真正做到了信息的对称，为构造公平公正的二手车交易市场奠定了坚实的基础。有了事故区块链，我们就可以大胆地去购买二手车了。

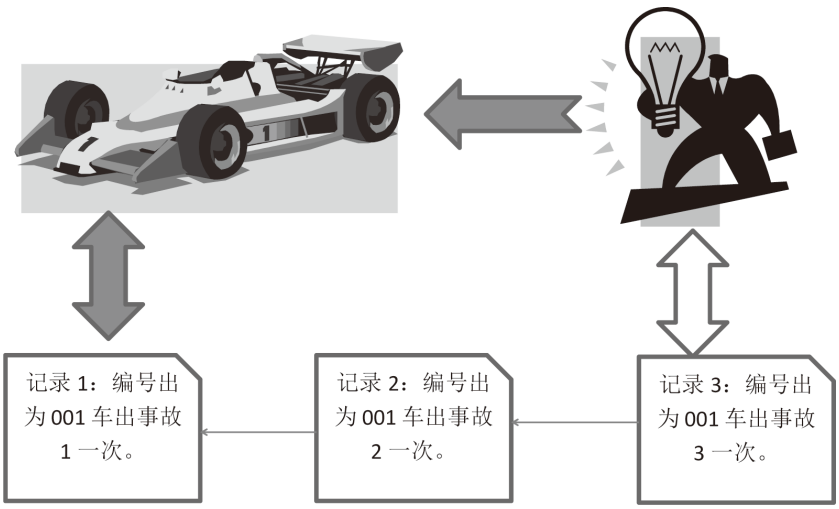



**图2.9　二手车事故区块链查询**

2．信任问题

在和陌生人进行交易的时候最重要的就是信任问题，彼此建立信任是一件特别困难的事情。最为典型的就是很多销售人员向某一个客户推销自己的产品效果，如产品功能多、使用方便等，但是客户对销售人员的话一般是不会完全相信的，最主要的就是没有证据证明他们说的是真的，如图2.10所示。

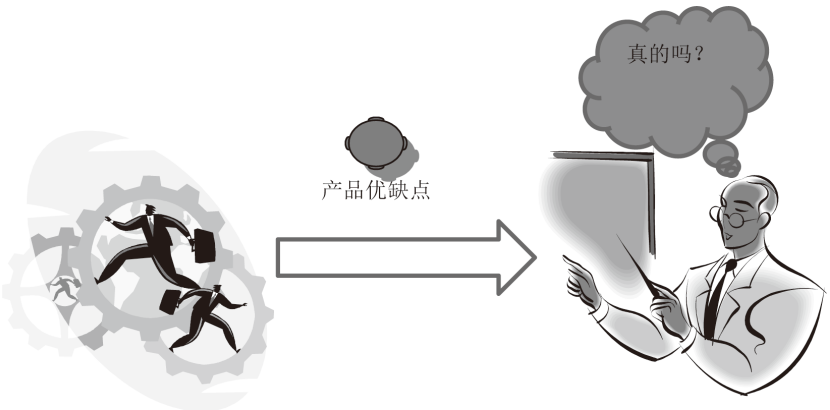



图2.10　常规式推销产品

在图2.11中，在产品使用过的顾客中所形成的真实性评价以记录的形式出现，并经过产品的使用者、围观者以及产品检验者等以顾客给出的评价进行真实性的检验，通过检验的评价进行产品的记录区块的上链，上链后的评价真实可靠且不可更改，可供所有的人进行查询。

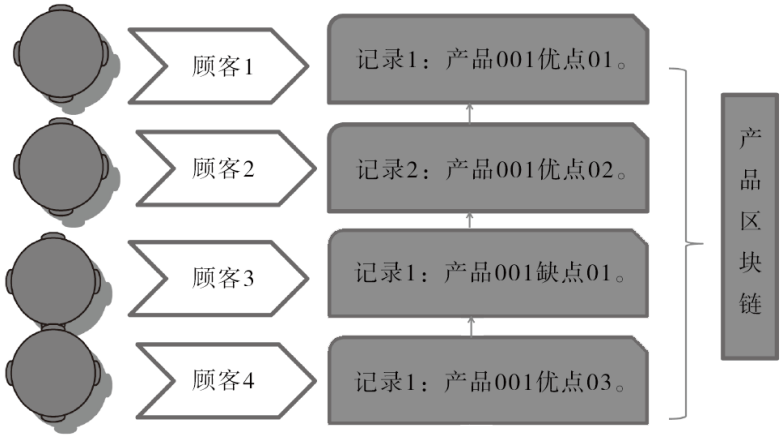



图2.11　产品历史记录区块链

在图2.12中，顾客可通过查阅产品评价区块链验证销售人员所说的是否属实，销售人员也可以通过该区块链尽快地与顾客解决相互的信任问题。大多数情况下，一旦信任问题解决了，很多事情便会迎刃而解。如果我们所有的人都相互信任，没有欺骗，那么我们生活的环境将会得到大大的改

善，骗子可能再也没有办法工作了，如图2.13所示。

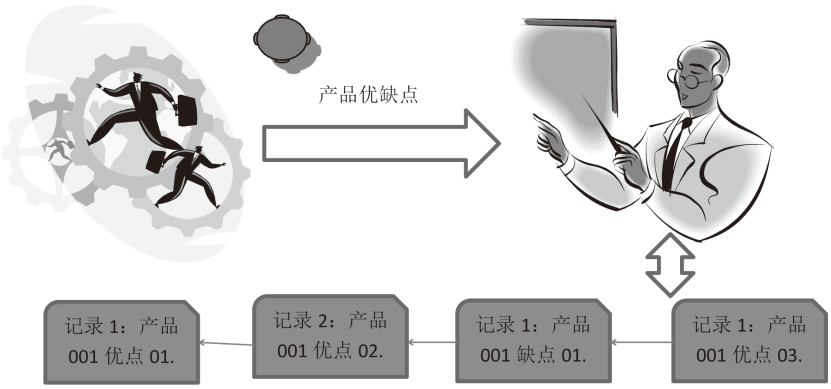



图2.12　区块链后的产品销售

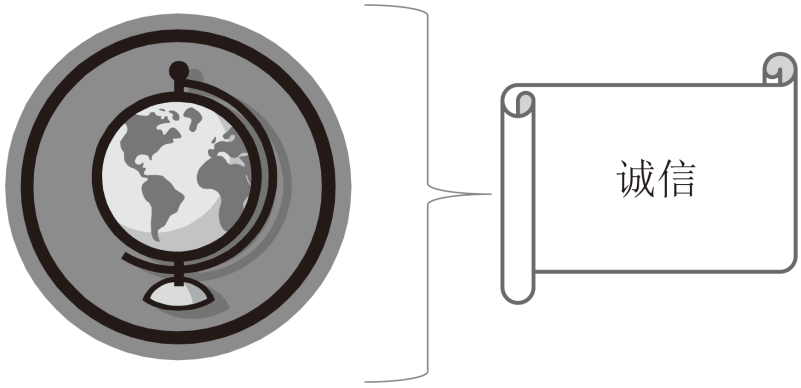



图2.13　区块链下诚信的环境

3．评价问题

评价问题是困扰我们人类的一个十分棘手问题，因为评价问题会受评价者主观想法的左右而使评价并不是特别客观，甚至有些评价是在几个人或者一个人手中进行，那样会让评价不可避免地出现一些偏颇性。例如，公司中对一个员工的评价，往往由部门经理依靠几个指标完成度去进行，而这常常忽略员工的一些其他方面。某公司对自己部门员工的年度考核如图2.14所示。

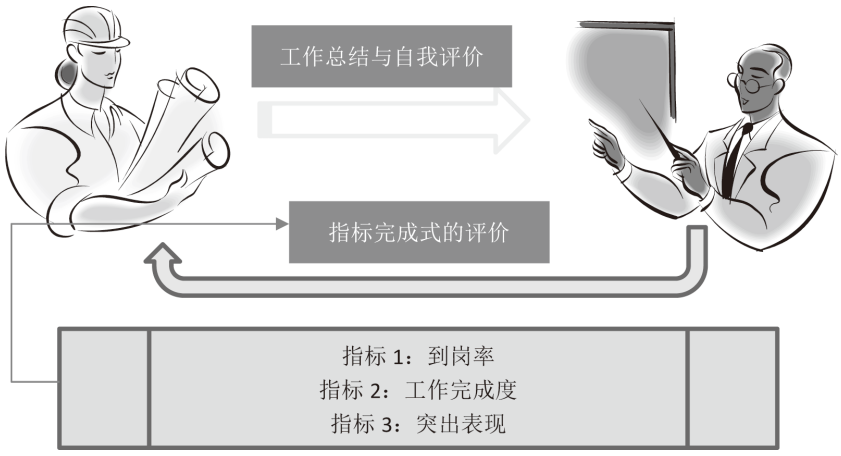



图2.14　常规员工评价

在图2.14中，仅仅依靠三项指标去评价一个员工显然是不全面的，而假如设置的指标过于繁杂，会直接影响到对员工评价的效率问题。不能因为考核员工而浪费掉大量的时间，这显然是所有公司所不愿意的。而真正理想的考察便是对员工的行为可以实时记录，在确保真实的前提下进行永久性的保存，并且这种评价不是以某个人为标准，而是依靠常规的行为准则、行业规范、工作能力或者社会监督等多种形式展开，这可通过区块链进行解决，如图2.15所示。

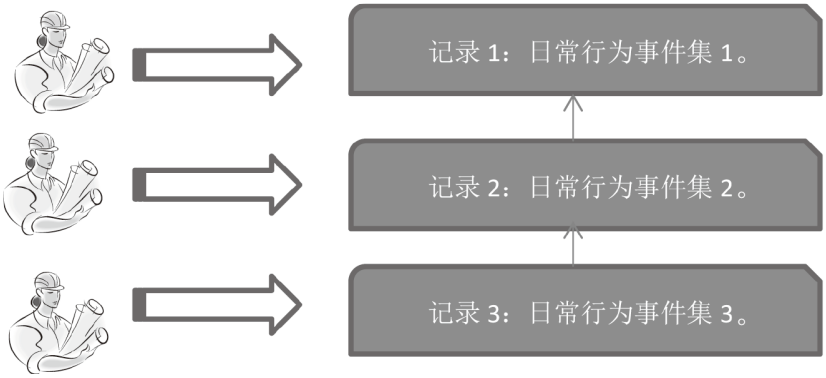



图2.15　员工日常行为区块链

图2.15中每一个记录都是收集着员工最近日子的一些特定行为，并有条

理地形成记录，存储于区块当中。随着时间的推移，区块中记录员工的行为事件在不断累加。当需要对员工进行评价时，可通过查询区块链中该员工的收集记录进行全面分析，如图2.16所示。这样比传统中的几个指标项要全面很多，且在区块链中可利用计算机技术进行统计分析，更容易得出员工的评价结果值。

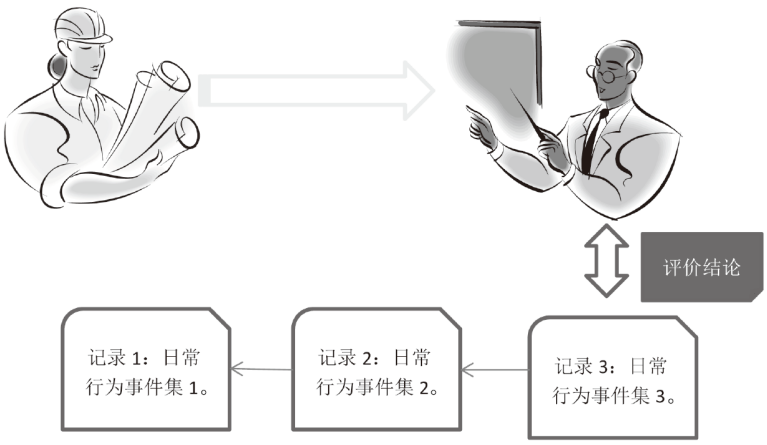



图2.16　区块链下对员工的评价

## 2.4　区块链的发展

区块链的发展离不开一个事物的产生，那就是比特币。什么是比特币呢？

比特币是一种虚拟货币，它区别于大家所熟知的普通货币，如图2.17所示。依靠于计算机的算力，也就是计算机运行的速度和性能，在网络中通过猜测以类似于暴力猜数字的办法计算出网络所期望的数值，全网中的计算机进行比赛，先猜测出结果的计算机获胜，获胜者可获得一定数量的比特币，全网有且仅有一台计算机能够获得单次的奖励，如图2.18所示。这种行为通常被称为“挖矿”，从事这种操作的计算机称为“矿机”。这种挖矿的行为是很难得出正确的结果的，但是如果有正确的结果，那就比较容易验证。

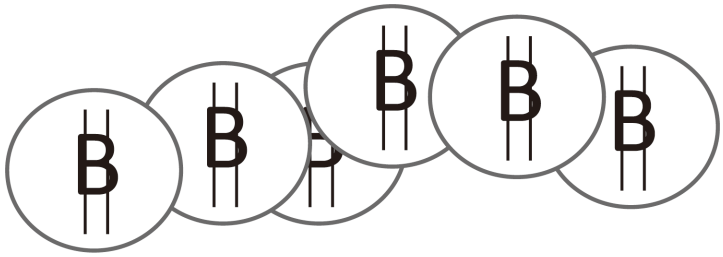


图2.17　比特币

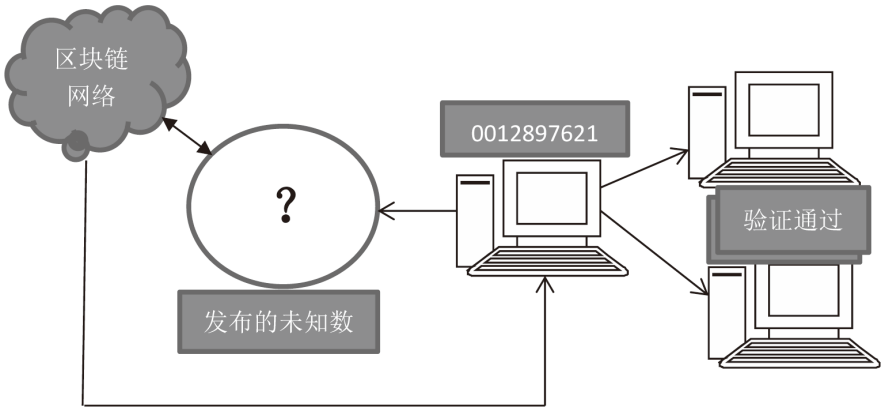



图2.18　比特币“挖矿”示意图

区块链技术是比特币能够正常运行的主要核心因素，比特币也使区块链得以被认可，自此以后人们便开始关注区块链技术的发展，也因此将支持比特币的区块链称为区块链1.0时代。此时的区块链比较纯粹，它的应用领域仅仅涉及金融领域，也是区块链应用最为成功的案例。它的本意是打破世界各国的货币流通不便，使人们的生产、生活更加便捷， 实现了货币的编程。但是，由于其太过理想化，并没有考虑到实际当中遇到的问题，如安全问题、区块链分叉、效率低等。

比特币虽然存在这样或那样的问题，但是支撑其发展的区块链技术得到了迅猛发展，在区块链1.0阶段之后，便因以太坊和超级账本的应用，使得区块链进入了区块链2.0时代。该时代最明显的标识就是智能合约的使用。区块链2.0使区块链更加智能和客观，能够根据相关条件的触发自动执行相关指令，不以某人或者部门为依赖。但是，此时的区块链依旧将应用

的中心放在了经济领域。

区块链技术所带来的优势不仅仅局限于金融领域，在其他一些领域也可以有着广阔的应用空间。所以，区块链2.0又逐渐向社会化应用的区块链3.0时代迈进，区块链3.0时代是真正的区块链应用时代。社会生活的方方面面都逐渐有了区块链的影子，区块链也正在悄悄地改变着人们的生产和生活。如今千家万户的互联网也正慢慢地被价值互联网所取代，这必然是一场全新的革命。在司法、物流、医疗、艺术等多个领域，区块链技术都正在悄然发芽。

区块链由1.0时代快速发展到3.0时代，如图2.19所示，是由于区块链技术切合现如今各行各业的技术发展需求，由最初的金融领域渗透到现在社会生活的方方面面。区块链、大数据和云计算是现如今的三大黑科技，之所以称其为黑科技，是因为该三项技术将颠覆现有技术的范围，给人以全新认识事务的角度。

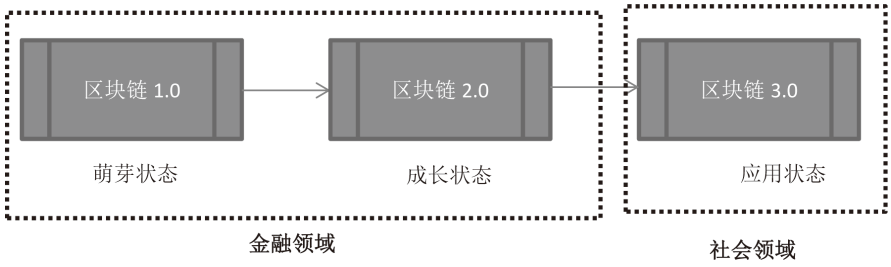



图2.19　区块链的发展历程

## 2.5　区块链的分类

随着区块链的发展，区块链也逐渐被人们进行分类。目前，区块链广为人们所接受的分类为三种：公有链、私有链和联盟链。

1．公有链

以比特币应用为典型代表的公有链，允许所有用户自由出入区块链网络，真正实现随时入网随时退网的功能。现有的区块链网络上用户是不稳定的，如同在市场上走动的人们，如图2.20所示。

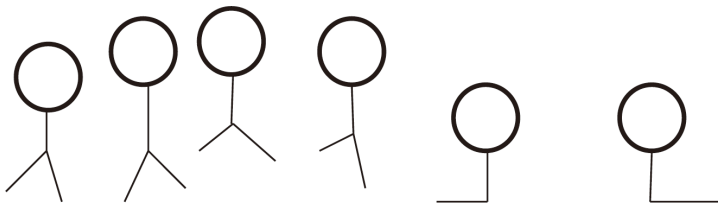


图2.20　市场上忙碌的人们

在市场中的人会各自忙碌着自己的事情，随时都有到市场上来的人，但是随时也有离开市场回家的人，也有的人在市场中逛着，悠闲地看着自己喜欢的商品等，人们对于市场是自由的，是否留在市场，离开或者进入都是由每一个人自己决定的。市场依旧是市场，市场不以某个人或者某些人的离开或到来而失去市场的功能。市场就如同区块链的网络，而市场里的人就是区块链网络中的成员（计算机），网络中成员可以根据自己的情况决定是否离开，或到来，或留存，成员的离开或到来不影响整个区块链网络的正常运行。公有链适合公共事务的系统或者平台的搭建。

2．私有链

私有链是相对于公有链而言的。私有链中的成员不能自由进出网络，它必须得到允许才可能进入区块链网络内，没有得到允许的成员是无法进入网络内部的。私有链类似于登录一个带有用户名和密码的系统，当我们进入系统时，必须输入由管理员分配给我们的用户名和密码，否则是无法进入的，如图2.21所示。

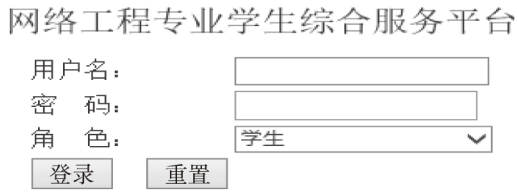



图2.21　平台登录界面

在图2.21中，如果想登录网络工程专业学生综合服务平台，必须要有正确的用户名和密码。该服务平台类似于私有链网络，而用户即为网络中的

成员。通常情况下，私有链的应用范围比较小，速度快，保密性好，整个区块链网络的安全性高。由于成员都是经过授权进入网络的，所以整个私有链网络中的成员均为可靠成员，私有链网络管理相对容易。

3．联盟链

联盟链是介于公有链和私有链之间的一种区块链，它是由几个部门或单位组成的，而成员需要为联盟单位的成员，联盟单位外的成员是不允许进入联盟链中的，如图2.22所示。它的使用范围通常要大于私有链的应用范围，而核心成员需要包含联盟单位的成员，在保证安全性和速度的前提下具有更加广泛的成员组成。

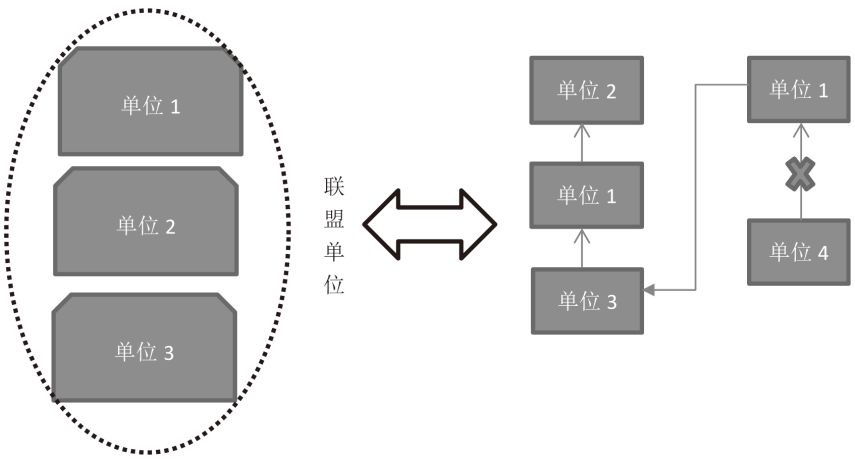



图2.22　联盟链示意图

在图2.22中，由于单位4并非联盟单位，所以单位4的成员不允许加入联盟区块链网络中，这样可以最大限度地保证联盟成员的权利与义务，确保联盟链网络的安全。

## 2.6　区块链技术讨论

区块链技术主要围绕区块链的几个特点展开，区块链的特点如图2.23所示。

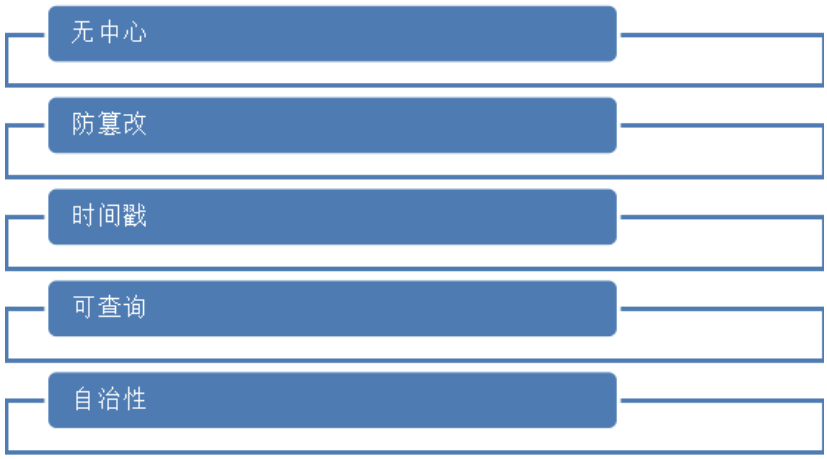



图2.23　区块链主要特点

1．无中心

无中心是区块链技术最主要的一个技术特征。没有中心也就意味着汇总后的数据存储不再依靠某一台或者一处的设备，转而采用了分布式的存储技术，将汇总后的数据分别存储在不同的地方，这样可以避免设备损坏而引起的数据丢失问题产生。比如，现在大学里存储学生成绩的往往是一台服务器，如图2.24所示。

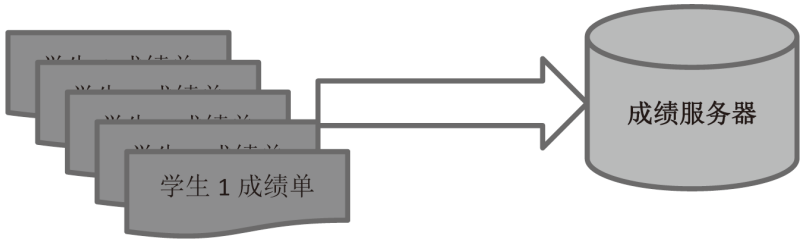



图2.24　某大学中心化存储成绩示意图

图2.24为一个存在中心化的典型应用。全校学生的成绩都存储在一台服务器上，这样这台服务器就尤为重要，一旦这台服务器出了问题，那么全校学生的成绩将会丢失，从而造成严重的后果。这就是我们经常所说的“瓶颈”，在学校的成绩管理中，这台存储成绩的服务器就是一个最薄弱的环节。为保证成绩的有效性，一种很简单有效的办法就是做好数据的备

份，也就是把数据复制出来保存，一旦服务器出问题，可以保证数据的不丢失。基于此点的问题考虑，很多大学在做成绩存储的时候都进行了中心化存储的改进，如图2.25所示。

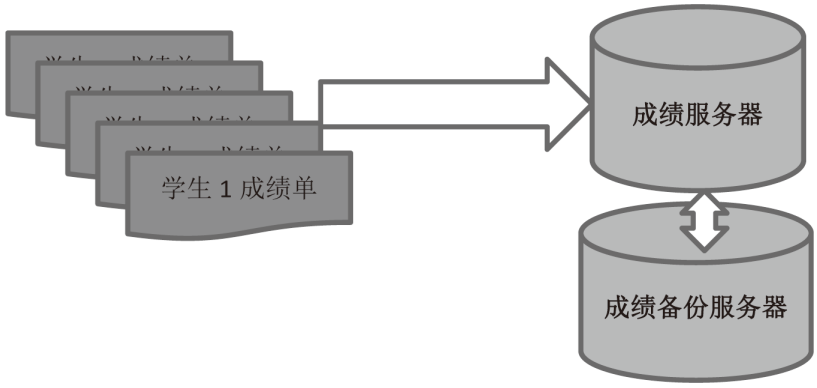



图2.25　某大学改进中心化存储成绩示意图

图2.25虽然对图2.24中的中心化进行了改进，从一定程度上保障了成绩的有效存储，但是仍然存在两台服务器都失效的可能性，问题依然存在。其实这也是一种中心化的数据存储方式。举一个简单的实例，如图2.26所示，我们常说的把鸡蛋放在了一个篮子里，一旦篮子漏了，将会失去全部的鸡蛋。而把鸡蛋分装两个篮子，相对于放入一个篮子更好些，但是依然有鸡蛋全部打碎的可能，而且可能性很大。

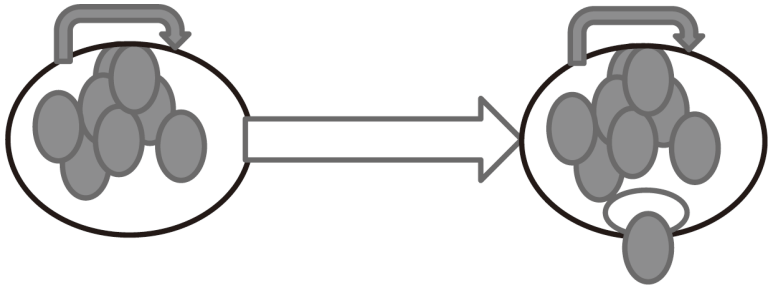


图2.26　鸡蛋放进一个篮子的危险性

在图2.26中可以看到，放在一个篮子里的鸡蛋会因为篮子的漏洞而丢失掉所有的鸡蛋。中心化问题在此可见一斑。

区块链技术中去中心化实现的核心观点就是分布式存储，所谓分布式

存储，就是将数据信息放在很多不同的地方存储（见图2.27），而且数据都是一致的，一个地方改变，其他任何地方都需要改变（见图2.28）。好比把一份数据复制了很多份，而且每一份数据信息都是一样的，那么当其中的一份丢失或者存储器失效都不会带来数据的丢失，除非全部的存储器同时损坏。而全部存储器同时损坏的概率极低，因此分布式存储能够最大限度地保障数据的有效性，防止单点故障。

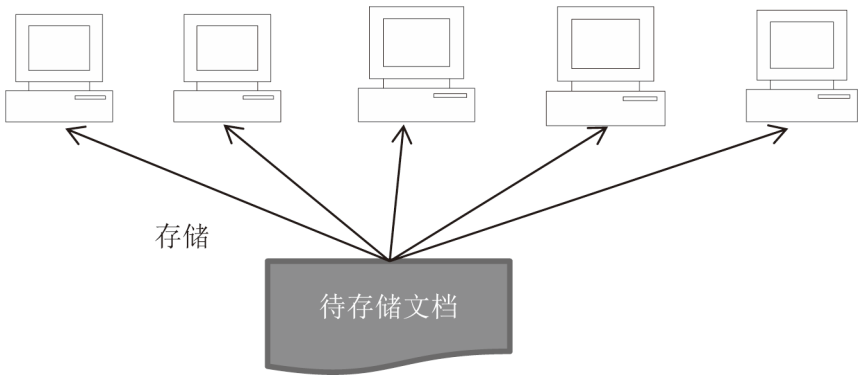



图2.27　分布式存储

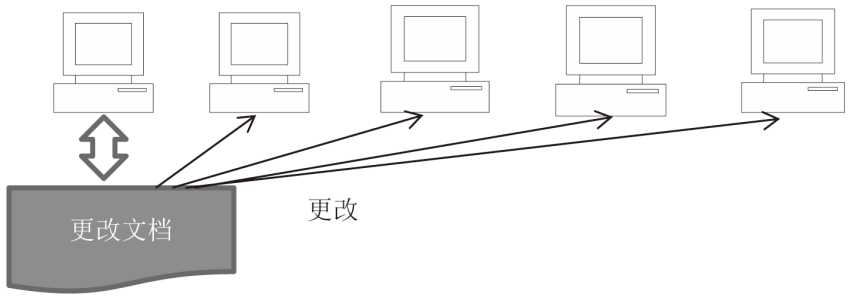



图2.28　分布式存储的数据更改

当数据信息被存储到世界各地的很多台计算机中时，我们就认为该数据是绝对安全的，因为几乎没有人可以让全世界的计算机同时都失去存储的数据信息，那么这样的分布式存储方式就可以解决中心化数据存储的单点失效可能带来的灾难问题。

2．防篡改

区块链技术实现防篡改主要是利用加密技术来实现的。最为典型的就是

Hash函数加密。那么，什么是Hash函数加密呢？它是通过一系列的算法，将可识别的数据信息变换成一种不可识别的数据信息形式，进而形成对数据保密的效果，我们称加密后的数据为哈希值。需要注意的是，这种数据加密是不可逆的。不可逆就是指不可能通过不可识别的数据信息形式变换为可识别的形式。这有点类似于一个人戴上面具便不知道他是谁一样，Hash函数就是那面具，而数据就是那个人，戴上面具的人就是Hash函数加密后的数据。一个你熟知的人，一旦他戴上面具，你便不会认出他。如果把面具摘下来便显示了本来的面目，这就是所谓的解密。但是，Hash函数具有不可逆性，就是虽然你知道他如何戴上面具，但是你是无法通过他戴面具的方法将面具摘下。

Hash函数有一个显著的特点就是无论需要加密多长的数据信息，加密后的数据都具有固定的长度。对1个字符的加密或者对1000万个字符信息的加密，其结果都是具有固定长度的效果，如图2.29所示，对1加密的结果为0xds78wed123，共12位。对“Hello word, my name is LiLei.”加密后的结果为ert146dgkb7u，共12位。从这个例子中我们不难发现，该加密的固定长度为12位。

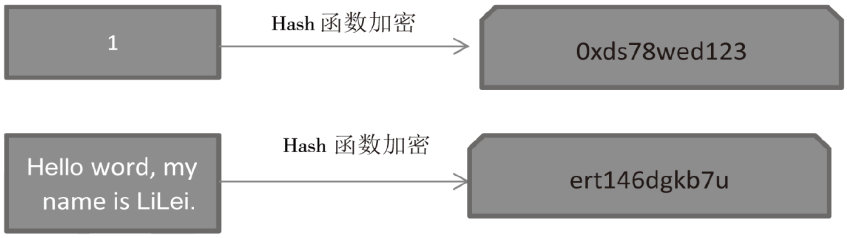



图2.29　Hash函数加密示例

另外，值得一提的是，不同的加密后的结果肯定对应不同的数据信息，但是反过来不成立，也就是说，不同的数据信息有可能对应出相同的加密结果。这种情况出现的概率极低，所以大家可以忽略，通常认为相同的数据信息会出现相同的加密结果，而不同的数据信息会出现不同的结果。

了解了Hash函数加密，那么接下来的区块链的防篡改就变得很简单了，首先我们从第一个区块开始分析。第一个区块在区块链中通常被称为创世区块，其前面不连接任何区块，创世区块的区块体可以由创建者自行

确定。创世区块仅含有一个区块号，区块号为区块中所有数据Hash函数加密后的哈希值，创世区块的区块号为自身区块的区块号。而非创世区块都含有两个区块号，其中一个为与之连接的前一个区块的区块号，另一个为自身的区块号。需要注意的是，前一个区块的区块号是作为该区块本身数据存在的。我们以图2.30为例进行说明。

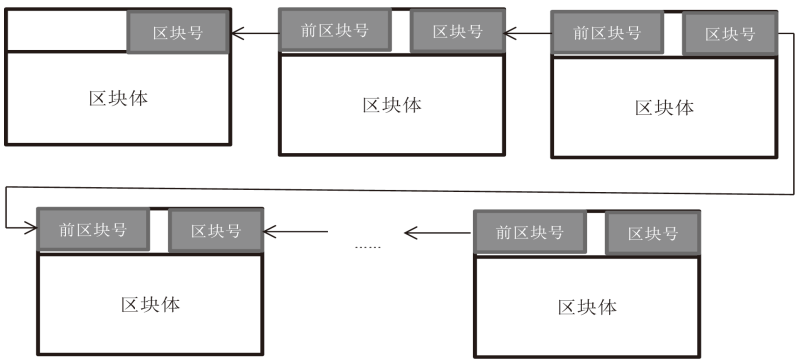



**图2.30　区块链的区块连接示意图**

在图2.30中，第二个区块包含了第一个区块的区块号，如果第一个区块中的信息被篡改，势必会引起区块号的改变，而这个区块号已经存在于第二个区块中了，所以如果想修改第一个区块中的信息值，必须先修改第二个区块中的第一个区块号。那么，新的问题又来了，如果第二个区块信息修改又会引起第二个区块的区块号变动，而第二个区块号又同时包含在了第三个区块内，这样又必须去处理第三个区块。这样一直修改下去，直到最后一个区块。可见，如果想更改一个区块内的信息值，必须修改它之后的所有区块的信息。

依据以上表述，如果前面的区块信息不容易篡改，那么篡改最后一个区块是不是就会很简单了呢？事实并非如，如图2.31、图2.32所示，一旦一个区块链形成后，它的区块连接又会很快，如比特币在10分钟左右的时间便会有一个新的区块形成，如果想改变一个区块内的信息值，必须攻破区块自身设置的重重保护，对保护的攻破又需要花费大量时间。当你攻破一个区块时，却会发现它的后面又连接了很多新的区块，进而使你无法攻击成功。这样就保护了区块链中的数据信息，达到了无法篡改的目的。

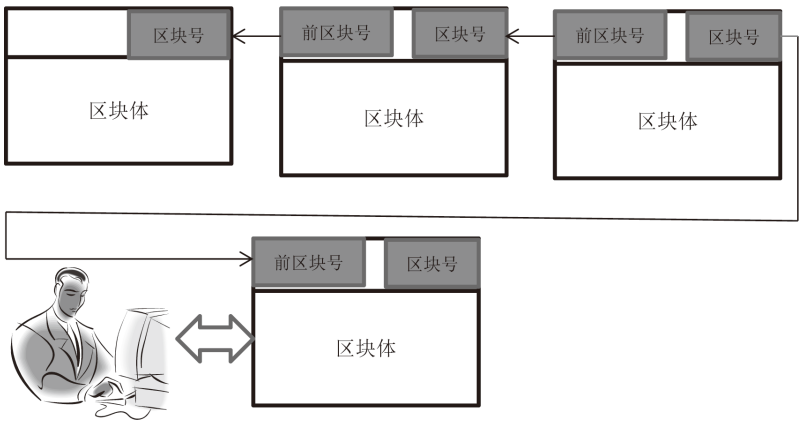



图2.31　篡改区块链最后一个区块开始

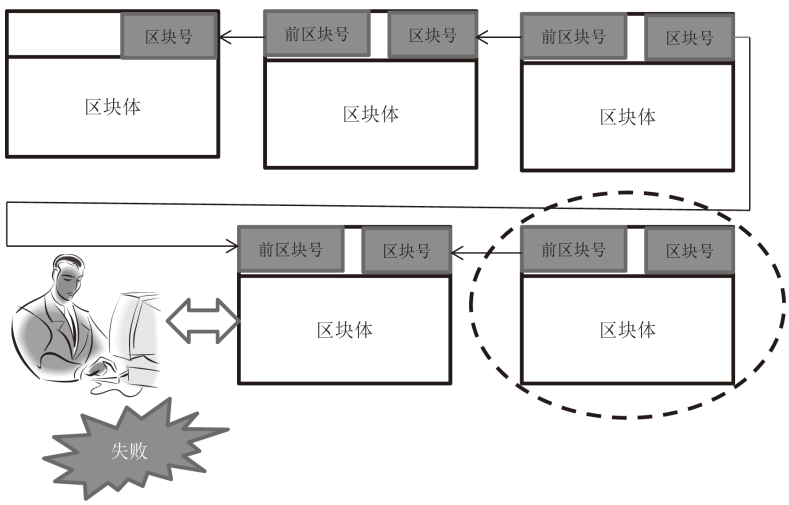



图2.32　篡改区块链时新区块形成

3．时间戳

在区块链形成的过程中，在每一个区块的头部都含有一个时间戳的字段。这个字段主要是记录该区块形成的时间，是竞争区块上链的时间证明。如比特币中的挖矿就是以计算出大数的时间来判定哪个计算机获胜的。当计算机A和计算机B都算出了正确的结果并广播让区块链网络中的其他计算机确认时，不但要验证其结果是否正确，而且要看计算出正确结果的时间，时间早的计算机赢得本次区块上链的权利，而验证的时间就是一

个区块中其头部的时间戳。该时间戳是在生成结果后由系统自动获取的网络系统时间，不能人为地去更改，从而保证了时间戳的公平公正，防止他人作弊。

下面以上课时学生的抢答过程来理解时间戳，如图2.33所示。

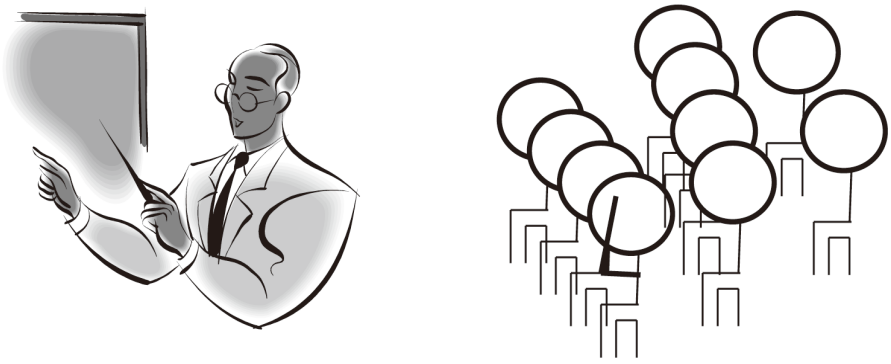


图2.33　课堂抢答问题

在课堂上，数学教师给出一个抢答题，这一个抢答题是很容易验证的，看看这一个题谁先算出来。那么，教师发出开始指令后，所有学生开始了计算，一旦算出结果就大声说出答案。而其他同学在听到有同学说出答案后，首先记录下时间，然后进行验证。当验证结论不正确时，就忽略说出答案的学生，继续计算那道抢答题。如果验证结论正确，就鼓掌祝贺计算出答案的学生。当有多名学生说出答案时，依次记录相应的时间，验证前面正确就忽略之后的学生答案。那么，在这个过程中，我们将学生记录的时间认定为时间戳，每一个学生的答案由一个数对组成，而不单纯是结果，该数对为（时间，结果）。最后获胜者是数对的获胜，而不单是单个结果获胜，也就是说，算得又对又快的学生才是最终的胜利者。

区块链的时间戳不仅仅在区块生成的过程中起到了不可替代的作用，如果有人试图篡改区块中的数据，时间戳也会自动更改为修改数据的时间，这样势必会打乱按照时间顺序连接的区块，那样整个区块链也会阻止相应的修改的结果。同时，在区块链中提供的查询功能，可以方便用户得到区块内记录的事件发生的时间，起到更好的记录功能以及防止事件参与者的抵赖。

4．可查询

区块链形成后便会在区块链网络内的每一台电脑中存储这条区块链，如图2.34所示，然后可从最新生成的区块开始按照连接的顺序一个一个地进行查询。在每一个区块内部，数据的保存形式可以自行设计，其设计的原则一般是便于用户快速查询。

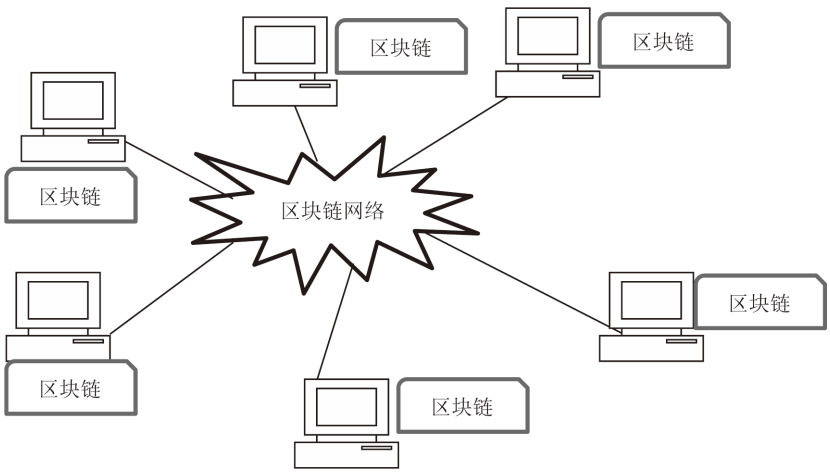



图2.34　区块链存储形式

由图2.34可知，在区块链网络内，如果一台电脑需要查询，仅需在自己电脑内操作即可，因为网络内每一台电脑中保留的副本都是相同的。知道了区块链宏观的查询，那么我们进一步讨论，在一个区块内是如何查询的呢？在此我们不妨先构建一个区块结构，然后进行分析，如图2.35所示。

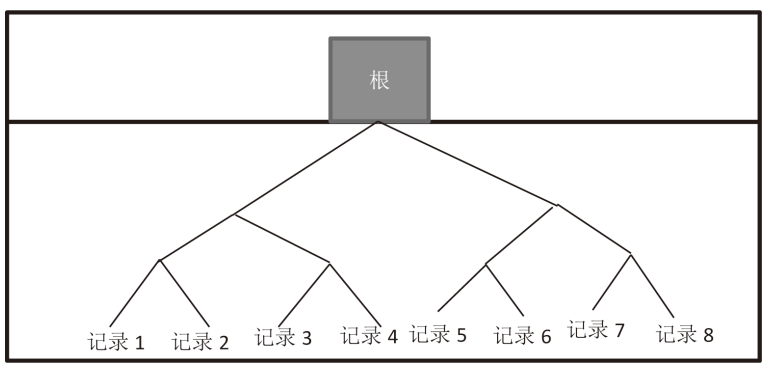



图2.35　区块内数据记录存储

一个区块可包含多个数据记录，为了能在一个区块内更快地找到目的数据记录，在图2.35中采用了一种树形层状数据存储方式，在这个树形存储结构中，一个数据记录就是该树的一个叶子节点，叶子节点是树的最底层。每两个叶子节点可形成一个高一级的汇总节点，在汇总节点中描述了其下节点的共同性质特征，该汇总节点又可与其他汇总节点再生成一个更高一级的汇总节点，直至汇聚为一个节点，该节点被称为树根。所以，树根拥有所有节点的共同特征，即树根具有该区块中所有记录的一个共同特征。

当需要在一个区块内查找指定的数据记录时，只要根据其特征不断地细化，逐层向下进行查找即可，最后会到达一个叶子节点的具体数据记录，该数据记录便是所需要查找的数据记录。这样的存储方式可起到快速查找数据记录的目的。下面我们以一个存储车辆的数据记录为例进行更进一步的剖析。

某汽车的4S店，将店内的汽车以区块链的形式进行存储，那么每一个汽车可当作一个记录存储到区块内，然后区块连接即可形成。为存储方便，我们可为六辆汽车分别进行编号，即1、2、3、4、5、6，然后按照编号存储于区块内，如图2.36所示。

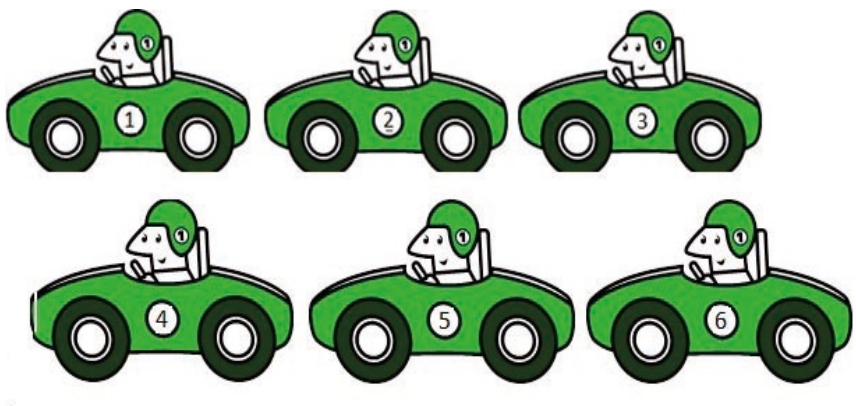



图2.36　存储车辆的数据记录

在汽车以编号的形式进行存储后，我们便可以快速地找到所指定的车辆。如果我们要在这个区块中寻找3号汽车的相关信息，那么首先根据该区

块的树根节点提示，因为3＜5，所以我们向树的左边进行搜索，到达下一层的节点。到达这一层后，我们再根据“<3左”的提示，由于3==3，所以不能左搜索，而是向相反的右侧进行查找，而此时会进入第三层。这一层不是叶子节点，同样它又出现的新指示信息，即<4左，那么因为3<4，所以我们可向左进行而到达叶子节点，该叶子节点正是我们所要寻找的3号汽车相关数据信息。

综上所述，在区块内对于数据的存储形式可以根据数据的类型和特点进行灵活设置，安全快速查询到指定的数据信息一直是人们不断探讨的问题。大家可根据不同需求设计出相对完美的区块内数据存储结构。

5．自治性

自治性指的是存在于区块链网络中的计算机，而这些计算机都是独立的，都有各自独立的工作空间和工作自由度。区块链网络不会强制性地要求计算机本身去做什么。那么，为了更好地维护区块链网络的正常运行，到达共同维护区块链中信息的内容，区块链网络引入了奖惩机制以引导网络内的计算机正常有序工作。如比特币应用中，对能够最快速地计算出大数据的计算机给予奖励（一定数量的比特币），而对恶意破坏网络规则的计算机则给予处罚，处罚的形式多种多样，在比特币中表现为警告节点计算机需要处罚比特币或者移除出区块链网络。

区块链网络中的自治性是引导下的自治，这样可确保区块链中数据的有效运维，维护区块链长期稳定地运行。如果失去奖惩机制，将全部的依托都放在每一个区块链网络的维护者或者网络内计算机用户的自觉性，那么该区块链网络是脆弱的，是经不起长期考验的。这如同学生在学校学习是一样的，如图2.37所示，学生学习是为了自身，但是由于学生自身缺乏自制力，在学校学习的过程中经常会被其他东西诱惑，甚至出现厌学的情况。对此，学校便会出台一系列的校规校纪，去帮助、引导学生在学习的道路上正常有效前进。如果离开了校规校纪，学生能够自我约束吗？所以，在学校中，学生自我约束、自我加压学习的效果自然离不开学校的校规校纪。




图2.37　学生自治性的形成