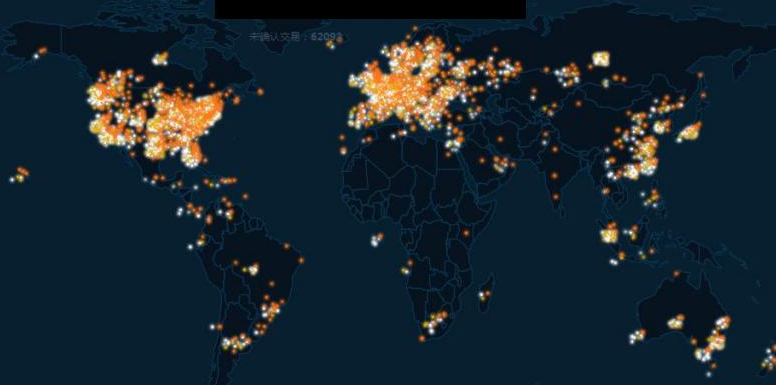
**区块链**

## 一、区块链含义

**1 理解**

**区块链是一种“共识”实现技术，通过区块链可以记录网际间所有的交易，供区块链的用户见证实现“共识”，且链上信息内容“不可篡改”。**而这种“不可篡改”性是通过系统内多个副本的存在增加了内容被恶意篡改的成本。以比特币系统而言，下图中的所有亮点代表一套内容一致的账本。因此，当所有的记录得到公示，就解决了现实生活中的“两表不可测”问题。两表不可测的原因是没有中心，两表数值各异观者无从可知。但中心不是必须，当存在多只表，且多数指针指向一个时间，少数服从多数，观者也就心知肚明。所以区块链要解决的一个问题就是“少数服从多数”，少数的存在有可能是数据生成的错误或者是恶意篡改的内容。也就是说你要篡改成功，就必须要改变系统中51%的副本[ 即51%攻击]，由少数成为多数。可想而知，篡改成本从技术难度、时间消耗、人员使用上都是巨大的。同时，区块链的另一个关键结合，是同智能合约的结合。通过程序执行合约，因为前者已经做到了合约内容的“原汁原味”，后者成为“观察者”用机器执行杜绝了合约双方的违约和执行中断。金融活动从最早的支付到金融衍生品的高度发展，背后都是一个个合约的签订和执行。同时各种金融创新绝大多数也是保证合约的执行和违约偿付。如果合约的达成变得扁平且执行贯彻，那么整个流程将更为简洁和高效，这便是区块链亦将提出就被金融领域重视的一个前提。另外，区块链是一个不断“膨胀”的记录系统，可以记录系统从开始到今后运营中的所以交易的记录。这一机制也便捷了第三方，特别是金融监管者对资产交易信息的追踪、监测和转移。



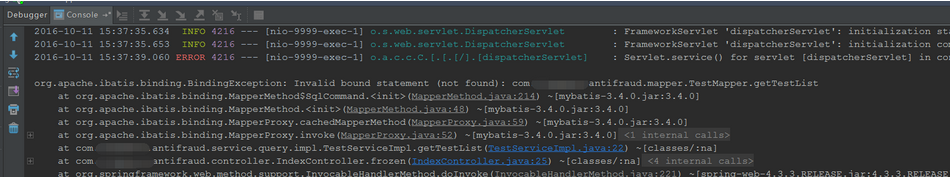
**比特币系统中的区块链副本   
说明：每个亮点都是比特币系统中的终端，每个终端都保持一个账本(副本)**

**2 数据存放的两类模式**

如果说合约是内容，当内容完成数字化，区块链就成为了信息存放的载体。因此，结合梅万妮•斯万意见，数据的存放就形成了两类模式。一类是数据就存放在块中，另一类是将数据压缩形成数字串列，也就是降维哈希过程[ 参见梅兰妮•斯万《区块链新经济蓝图及导读》第21页]。对于前者容易实现，但一致性问题突出。后者只同步数字串列，通过减少同步量解决一致性问题，但哈希过程的降维成为关键。

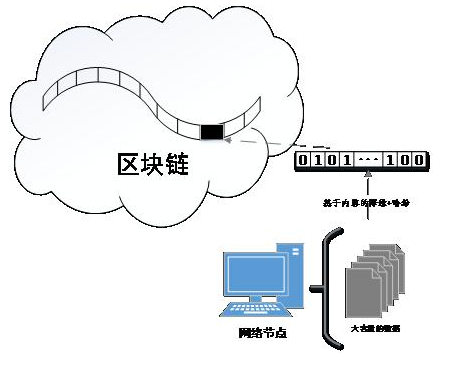
**3一致性问题**

一致性问题是分布式存储系统一个比较难解决的问题，通常这一个环节起容灾备份功能，同步操作大多放在“闲时”，是一种后台操作。区块链采用“前置”的同步操作，当一个块生成，块链的更新需要所有副本完成内容一致的复制。如果内容存储在块内，那么更新的数据量为更新的内容乘以副本的数量，再加上副本间的传输距离和带宽，整个操作是耗时的，整体的效率体现在系统中最长的同步时间，如图2。

因此采用这种方式，将影响对高频应用的响应，例如诸多的交易场景。所以解决一致性问题的办法一是控制更新内容的大小，二是限定副本的数量。比特币系统的实现是采用前者，私有链是应用了后者。

**4降维哈希问题**

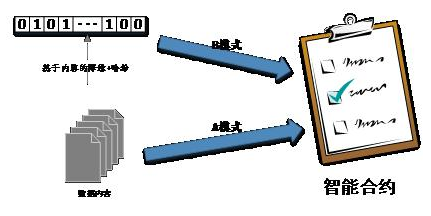
一致性问题的确是区块链技术未来真正应用的关键，如何降“魔”？梅万妮•斯万给出了他的思路，如果同步的数据量仅仅是64位或者是128位的数字串列，那么一致性起码是可控的。在此模式下，区块链服务的任何数字资产(例如文件、图片、视频录像、监控记录、软件等)数据最终不需要存放在区块链上，更新的内容只是对应的数列。因此如何降维是这一模式下的问题所在，共识的内容是自然语义。当自然语义过渡到程序识别的数字表达，多维的自然语义和元数据需要完成一对一的关系映射，这一过程需要对信息进行降维，其中会用到哈希、编码等技术。例如一个9G的基因组文件，就需要降维为1维的二进制数列来平衡区块链同步前置的负外部性，最终使智能合约的执行算法能够以此甄别对象的“真伪”。需要提醒的一点是梅万妮•斯万这种方式哈希过程不是地址(地址是数字资产数据的存储地址，例如指基因组文件的存放地址。要区别比特币中的哈希地址，其实质是明确比特币的所有权，类似于交易的甲乙方，是交易的内容)映射，如果区块链当中保存的是地址信息。



一是违背了内容共识的机制，二是大大降低了系统的安全性。一旦非法入侵者获得任何一个副本，就可以找到合约内容的存储空间进行篡改。降维哈希这一过程类似于“压缩”，需要基于数据内容、文件特征、时间戳等特定信息进行压缩。这样智能合约的执行算法才能够以此判定指定内容的“真伪”。所以这个过程是复杂而艰巨的，工程量浩大。

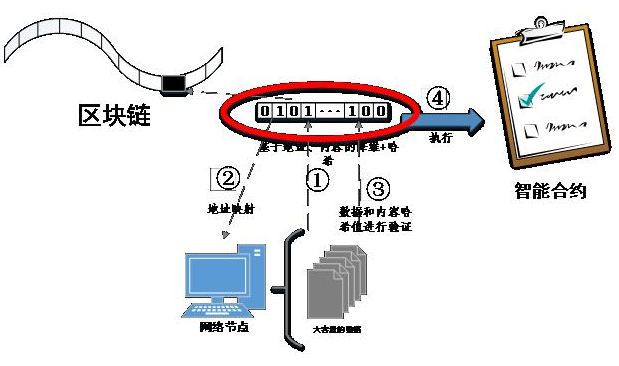
**5智能合约的衔接**

智能合约的核心是利用程序算法替代人执行合同，从这个概念上理解合约的执行可以类比“支付宝”。当买家和卖家在淘宝网达成了某笔交易，钱货两讫将按照支付宝程序执行，买卖双方更多是配合程序执行的贯彻。因此，智能合约的前提就必须是签订的交易在执行前不能被篡改。区块链因此成为智能合约在去中心环境下的天然基石，该机制保证了智能合约执行程序前能得到“原汁原味”的合约。

通常对智能合约的“触发”有两种自然的解决方式，模式A和模式B，如图3。前者通过数据本身的内容形成触发事件来驱动智能合约的执行，因此在执行之前需要验证数据内容和哈希值是否匹配。如果不匹配，就意味原始数据被篡改后续需要匹配容灾恢复机制。而采用模式B，则需要智能合约实现对哈希值的解读。如果说降维哈希是“压缩”，那么与智能合约之间的衔接就需要执行“解压”，需要从数字串列的特征值上进行解读，实现执行算法的“事件触发”。因此，区块链和智能合约的应用需要对数字资产内容进行标准化，通过标准的组件配置实现资产从注册、确权、转移等功能的量身定制。

**6区块链的检索**

如果说一致性问题的解决是区块链“写”操作的实践，那么区块链内容的检索就是一个“读”过程。因此，对于应用的需求而言，切不可为了区块链而区块链。在内容实现广泛共识的基础上，我们可以相机选择合适的方式来优化区块链上的遍历。中心化的检索或者区块链本身复合检索功能都是可以参考的方式。国内开发的唯链Vechain是采用后者的一种方式，将地址信息和内容信息同时进行哈希。但这一方案反而降低了系统的安全性，因为区块链系统存在太多的副本。假如非法入侵者获取到任何一个副本就得到了系统一个完整的地址表，入侵者就很容易找到原始数据的存放空间进行“篡改”。所以需要强调的是：区块链的“不可篡改性”，仅仅指区块链上存储的内容(如图4红色圈出的内容，图中的序号说明了整个流程)难以篡改，而非其他数据。同时，复合地址信息从另一个角度来看也为未来系统的负载均衡设置了障碍。由于的区块链不能“瘦身”，存储信息日益膨胀使得“读”效率会逐步降低，未来客户端是否需要装载完整的区块链？毕竟对于单个用户而言涉及到的信息比例很低，在区域内建立区块链服务器也不为一个预期的方向。这样“去中心”的区块链将过渡到“非中心”的区块链，由区块链服务器保留完整的链信息。

传统金融的基础设施是中心化的，交易过程中信息的汇总、合同的内容、信用的背书等等操作都必须以金融中介机构为中心。由它为双方达成的合约背书，也就默认了凡是中介机构承载的交易信息都是正确的。因此，传统中心化的服务方式，系统本身是不容错的。一旦中介机构本身存在道德风险或者中介机构被黑客恶意修改合同，那么交易合同履行存在漏洞，参与者会出现财产损失。这进一步说明了，最近几年金融机构的服务器屡受攻击的动机所在。区块链的存在允许了系统的容错，通过“去中心的共识”实现系统的统一数字表达，将供需拉得更近。持续的被关注，本身就是对传统金融现状的“反抗”。金融本质是价值的流动，需求、供给和中介成为了实现流动的三个环节。传统的金融中介现阶段难以完成供需的直接满足，区块链强化供需、削弱中介来实现供需效率匹配的优化，这也是当前供给侧改革一种技术实现。

## 二、区块链的举例

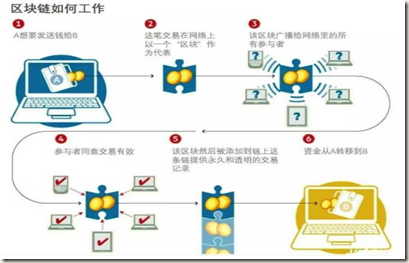
**区块链的本质是一个分布式的公共账本，任何人都可对这个账本进行核查，但不存在单一的用户可以对它控制。在区块链系统中的参与者共同维持账本的更新：它只能按照严格的规则和共识进行修改。**

举例：如果A借了B 100块钱，这个时候，A在人群中大喊“我是A，我借给了B 100块钱！”，B也在人群中大喊“我是B，A借给了我100块！”此时路人甲乙丙丁都听到了这些消息，因此所有人都在心中默默记下了“A借给了B100块钱”。这个系统中不需要银行，也不需要借贷协议和收据，严格来说，甚至不需要人与人长久的信任关系（比如B突然又改口说“我不欠A钱！”，这个时候人民群众就会站出来说“不对，我的小本本上记录了你某天借了A100块钱！”）。

**1这是一个去中心化的系统**，整个系统中没有了权威的中心化代理，信息的可信度和准确性便会面临问题。

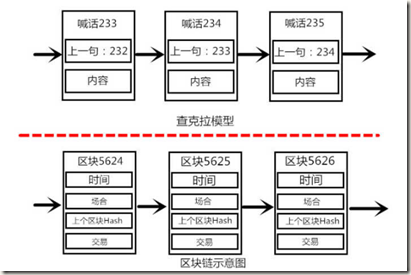


**2可能你已经发现了，在上述的模型中，所谓的“100块钱”已经不重要了。**换句话说，任何东西都可以在这个模型中交换，甚至你可以凭空杜撰一个东西，只要大家承认，你就可以让你杜撰的东西流通。比如：我在人群中高喊一声“我创造了10个查克拉！”，我甚至不需要知道查克拉是什么，也不需要关心世界上是不是真的有查克拉，只要大家都听到，然后在自己的小本本上记下“LaiW3n有10个查克拉”，于是我就真的有100个查克拉了。从此以后，我便可以声称我给了某人1个查克拉，只要路人甲乙丙丁都收到并且承认了这一信息，那我就算完成了这次交易，哪怕世界上没有查克拉。又比如：“比特币”，区块链是比特币的底层技术，真正的区块链和比特币比上述的模型复杂太多，细节也丰富太多。



**3但是存在以下问题：**

假设过了很长一段时间，凭空创造的查克拉已经在这个系统中流通了起来，大家都开始认可了查克拉。但是这个系统中一共就只有10个查克拉，于是有人动了坏心思，他在人群中高呼“我有10个查克拉！”怎么办？大家是直接在本本上记下他有10个查克拉么，这样不是人人都可以伪造查克拉了么？**为了防止这种现象发生，**决定在我创造查克拉的时候给我的查克拉打上标记（更准确地说，我是给我喊的那句“我创造了10个查克拉”打上标记，比如标记为001），这样以后在每一笔交易的时候，我在高喊“我给了某某1个查克拉！”的时候，会附加上额外的一句话：“这1个查克拉的来源是记为001的那条记录，我的这句话标记为002！”。我们再抽象一点，某人喊话的内容的格式就变成了：“这句话编号xxx，上一句话的编号是yyy，我给了某某1个查克拉！”，这样就解决了伪造的问题。其实上述模型就变成一个简化的中本聪第一版比特币区块链协议。看到这里基本已经能够生动形象又不涉及任何细节地解释区块链了。但是任然存在以下疑问：



**4 “凭啥？”**

“凭啥你喊一句话我就帮你记？我的小本本不要钱么？”。为了激励大家帮我传话和记账，我决定给第一个听到我喊话并且记录在小本本上的人一些奖励：第一个听到我喊话并记录下来的人，你就凭空得到了1个查克拉，这个查克拉是整个系统对你幸苦记账的报酬，而你记录了这句话之后，要马上告诉其它人你已经记录好了，让别人放弃继续记录这句话，并给你自己的记录编号让别人有据可查，然后你再把我的话加上你的记录编号一起喊出来，供下一个人记账。

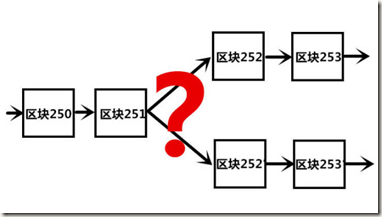
当这个规则定下以后，这个系统中一定会出现一批人，他们开始竖着耳朵监听周围发出的声音，以抢占第一个记账的权利。比如概念 “比特币挖矿”。毕竟1比特币=1K-nK美元。

比特币挖矿机，就是用于赚取比特币的电脑，这类电脑一般有专业的挖矿芯片，多采用烧显卡的方式工作，耗电量较大。用户用个人计算机下载软件然后运行特定算法，与远方服务器通讯后可得到相应比特币，是获取比特币的方式之一。



**5 “听谁的？”**

在这个系统中，如果我和另一个人C几乎同时地喊出一句：“为了艾泽拉斯！”。由于听众所处的位置不同，一定会有人先听到我说的那句话，而另外一些人则先听到C的那句话，如果我们规定只能有一个人说出这句话，那到底这句话是谁说的？如果不加任何条件，那么上述的情况一定会这样发展：一部分人认为这句话是我说的，在听到这句话之后开始记账，之后他们所做的所有事情都是基于这个事实，并且随着这个信息一次次的传下去，这条信息链会越来越深；而另外一群认为是C先说这句话的人，也会按照这样的趋势发展。这样，原本是一条唯一的信息链，在我们喊出“为了艾泽拉斯”这句话之后，分叉了？

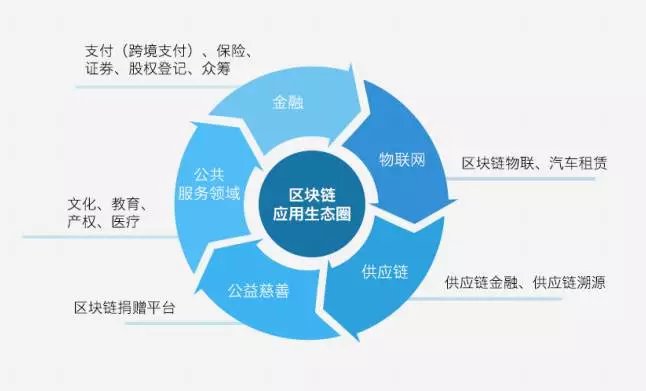


这会导致怎样的情况呢？按照我们的设想，应该每个人的小本本上记录的东西都是一样的，都是一条可以把所有信息串联起来的链条。但是在这一刻，他们小本本上记录的东西不一样了？以后还怎么确定交易和信息的真实性？为了解决这个问题，又追加了新的规则，增加记录编码的难度，即比特币挖矿难度，保证记录的唯一性（保证节点之间的同步）。”不知道用唯一性是否准确？

## 三、区块链的应用

**1 区块链几乎可以应用到任何领域，在金融、物流、公共服务等领域都有大量案例。**中国央行早在2016年就表态支持区块链。2016年11月，中国政府正式把区块链列入十三五规划纲要中。

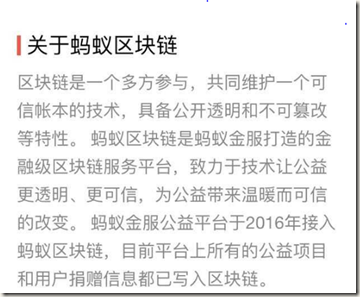
比如贵州超过3000亿的脱贫脱贫攻坚基金，要通过300多个单位的运作，惠及493万贫困人口，如何确保扶贫资金募款，投资，使用管理等环节正确运作呢？区块链具有多方共识，交易溯源，不可篡改等技术特点，使它在确保信息可信、安全、可追溯等方面具有传统技术不可比拟的优势。以城市为单位，第一个发布区块链白皮书的就是贵阳市。



**2 如果仔细观察，我们会发现马云最近两年在公共场所提及最多的概念便是信用体系。**互联网在未来不会是创业主体，而是像空气一样成为我们生活的一部分。

蚂蚁金融便是马云最为看重的业务体系，在马云的构想下，未来人们的生活将无现金，并可以用信用兑换商品。这一实现，十分需要一种革命性的信用机制。这一机制在2016年的支付宝App里已经悄然 上线---蚂蚁区块链





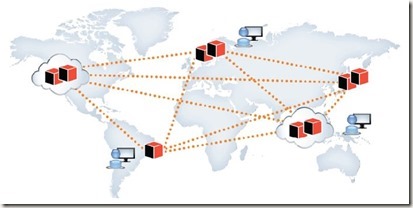
## 四、从技术角度简单理解区块链

**1 区块链的本质**

区块链是一种特殊的分布式数据库。

**首先**，区块链的主要作用是储存信息。任何需要保存的信息，都可以写入区块链，也可以从里面读取，所以它是数据库。

**其次**，任何人都可以架设服务器，加入区块链网络，成为一个节点。区块链的世界里面，没有中心节点（去中心化），每个节点都是平等的，都保存着整个数据库。你可以向任何一个节点，写入/读取数据，因为所有节点最后都会同步，保证区块链一致。



**2区块链的最大特点**

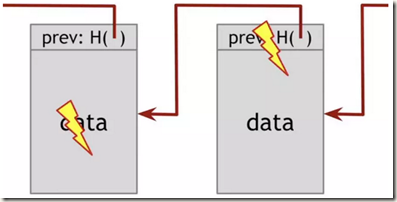
区块链没有管理员，它是彻底无中心的。其他的数据库都有管理员，但是区块链没有。如果有人想对区块链添加审核，也实现不了，因为它的设计目标就是防止出现居于中心地位的管理当局。没有了管理员，人人都可以往里面写入数据，怎么才能保证数据是可信的呢，这就是区块链奇妙的地方。

**3 区块**

区块链由一个个相连的区块（block）组成。区块很像数据库的记录，每次写入数据，就是创建一个区块。每个区块包含两个部分：

区块头（Head）：记录当前区块的元信息

区块体（Body）：实际数据



**区块头包含了当前区块的多项元信息**

生成时间

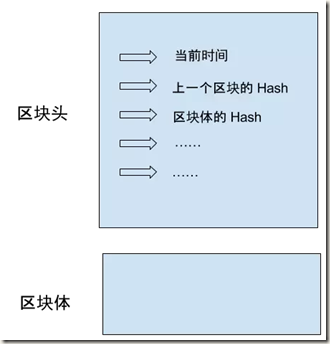
实际数据（即区块体）的 Hash

上一个区块的 Hash

......

Hash 就是计算机可以对任意内容，计算出一个长度相同的特征值。区块链的 Hash 长度是256位，不管原始内容是什么，最后都会计算出一个256位的二进制数字。而且可以保证，只要原始内容不同，对应的 Hash 一定是不同的。

举例来说，字符串123的 Hash 是a8fdc205a9f19cc1c7507a60c4f01b13d11d7fd0（十六进制），转成二进制就是256位，而且只有123能得到这个 Hash。



**4 Hash 的不可修改性**

区块与 Hash 是一一对应的，每个区块的 Hash 都是针对”区块头”（Head）计算的。

**Hash = SHA256（区块头）**

区块头包含很多内容（包括上一个区块的Hash、当前区块体的Hash等，见上图）。这意味着，如果当前区块的内容变了，或者上一个区块的 Hash 变了，一定会引起当前区块的 Hash 改变。如果有人修改了一个区块，该区块的 Hash 就变了。为了让后面的区块还能连到它，必须同时修改后面所有的区块，否则被改掉的区块就脱离区块链了。Hash 的计算很耗时，同时修改多个区块几乎不可能发生，除非有人掌握了全网51%以上的计算能力。

正是通过这种联动机制，区块链保证了自身的可靠性，数据一旦写入，就无法被篡改。这就像历史一样，发生了就是发生了，从此再无法改变。

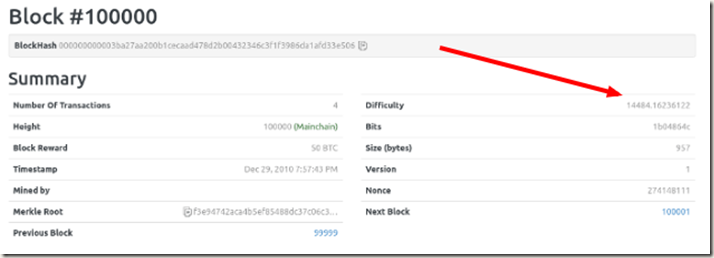


**5 采矿**

由于必须保证节点之间的同步，所以新区块的添加速度不能太快。试想一下，你刚刚同步了一个区块，准备基于它生成下一个区块，但这时别的节点又有新区块生成，你不得不放弃做了一半的计算，再次去同步。因为每个区块的后面，只能跟着一个区块，你永远只能在最新区块的后面，生成下一个区块。所以，你别无选择，一听到信号，就必须立刻同步。所以，区块链的发明者故意让添加新区块，变得很困难。他的设计是，平均每10分钟，全网才能生成一个新区块，一小时也就六个。这种产出速度不是通过命令达成的，而是故意设置了海量的计算。也就是说，只有通过极其大量的计算，才能得到当前区块的有效 Hash，从而把新区块添加到区块链。由于计算量太大，所以快不起来。这个过程就叫做采矿（mining），因为计算有效 Hash 的难度，好比在全世界的沙子里面，找到一粒符合条件的沙子。计算 Hash 的机器就叫做矿机，操作矿机的人就叫做矿工。

**6 难度系数**

你可能会有一个疑问，人们都说采矿很难，可是采矿不就是用计算机算出一个 Hash 吗，这正是计算机的强项啊，怎么会变得很难，迟迟算不出来呢？（比特币挖矿机，就是用于赚取比特币的电脑，一般有专业的挖矿芯片，多采用烧显卡的方式工作）原来不是任意一个 Hash 都可以，只有满足条件的 Hash 才会被区块链接受。这个条件特别苛刻，使得绝大部分 Hash 都不满足要求，必须重算。区块头包含一个**难度系数**（difficulty）这个值决定了计算 Hash 的难度。举例来说，第100000个区块的难度系数是 14484.16236122。区块链协议规定，使用一个常量除以难度系数，可以得到目标值（target）。显然，难度系数越大，目标值就越小。Hash 的有效性跟目标值密切相关，只有小于目标值的 Hash 才是有效的，否则 Hash 无效，必须重算。由于目标值非常小，Hash 小于该值的机会极其渺茫，可能计算10亿次，才算中一次。这就是采矿如此之慢的根本原因。区块头里面还有一个 Nonce 值，记录了 Hash 重算的次数。第100000 个区块的 Nonce 值是 274148111，即计算了 2.74 亿次，才得到了一个有效的 Hash，该区块才能加入区块链。



**7 难度系数的动态调节**

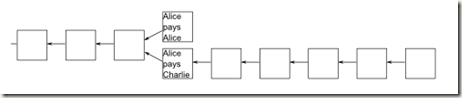
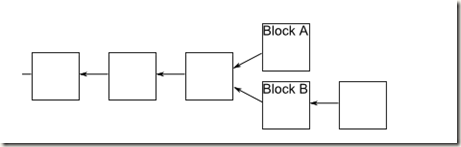
就算采矿很难，但也没法保证，正好十分钟产出一个区块，有时一分钟就算出来了，有时几个小时可能也没结果。总体来看，随着硬件设备的提升，以及矿机的数量增长，计算速度一定会越来越快。为了将产出速率恒定在十分钟，区块链发明者还设计了难度系数的动态调节机制。他规定，难度系数每两周（2016个区块）调整一次。如果这两周里面，区块的平均生成速度是9分钟，就意味着比法定速度快了10%，因此难度系数就要调高10%；如果平均生成速度是11分钟，就意味着比法定速度慢了10%，因此难度系数就要调低10%。难度系数越调越高（目标值越来越小），导致了采矿越来越难。

**8 区块链的分叉**

即使区块链是可靠的，现在还有一个问题没有解决：如果两个人同时向区块链写入数据，也就是说，同时有两个区块加入，因为它们都连着前一个区块，就形成了分叉。这时应该采纳哪一个区块呢？

现在的规则是，新节点总是采用最长的那条区块链。如果区块链有分叉，将看哪个分支在分叉点后面，先达到6个新区块（称为”六次确认”）。按照10分钟一个区块计算，一小时就可以确认。

现在的规则是，新节点总是采用最长的那条区块链。如果区块链有分叉，将看哪个分支在分叉点后面，先达到6个新区块（称为”六次确认”）。按照10分钟一个区块计算，一小时就可以确认。



为了保证数据的可靠性，区块链也有自己的代价。一是效率，数据写入区块链，最少要等待十分钟，所有节点都同步数据，则需要更多的时间；二是能耗，区块的生成需要矿工进行无数无意义的计算，这是非常耗费能源的。