# 区块链是什么

1. 区块链是一个放在非安全环境中的分布式数据库（系统）。
2. 区块链采用密码学的方法来保证已有数据不可能被篡改。
3. 区块链采用共识算法来对于新增数据达成共识。

具有以上三个性质的系统，就是区块链。



图一.区块链宏观理解

这张图的意义为的是帮助大家在宏观上先快速理解区块链中所涉及到的相关名词以及他们的层级关系。

## 非安全环境中的分布式数据库

这里的要点有两个：

* 分布式
* 非安全环境。

首先，这是一个分布式的，去中心化的系统。所以，只有一个中心服务器或者节点的，不是区块链。节点都是安全的，无恶意的，那这不是区块链。同理，从应用的角度讲，如果你的应用必须要使用中心节点（例如要用超级计算机做深度学习）或者没必要考虑节点不安全的情况（例如某个安全的工厂里的传感器），那么并不需要考虑区块链技术。

至于后面的词“数据库”，目前大部分成熟的区块链都是数据库，例如比特币就是一个分布式账本，而账本其实就是数据。然后，根据数据的格式，又可以分三种：

1. 数据是完全不相关的，只是达成的共识，没有有效无效之分；（例如：保全链）
2. 数据有某些逻辑结构，例如账本中，一笔交易实际上除了金额，还有输入和输出，连接到之前的交易，这些数据需要通过逻辑验证（例如交易中，节点需要验证输入的交易是否有效）；（例如：比特币系统）
3. 数据拥有图灵完备的逻辑，而验证的时候需要通过节点使用算力运算，每笔交易可以有不同的输出和状态，每个节点要做的不仅仅是验证交易的真实性和输入的正确性，还要根据交易里的逻辑读入数值，进行验算然后再验证结果。（例如：SimpleChain、以太坊）

## 采用密码学的方法来保证已有数据不可能被篡改

这个是误解最多的部分，因为很多人一提到区块链就只觉得是这个。诚然，这部分很重要，而且确实区块链也因此得名，但这只是区块链的定义的一部分。

这个部分的两个核心要点是：

* 哈希函数
* 非对称加密。

两个都是密码学的基础概念，网上都有非常清晰的定义，这里简单描述：

哈希函数：一个函数Y=H(X)，有如下性质：

1. 由X可以很容易算出Y；
2. 由Y不可能算出X；
3. 有Y不可能找到另一个X'使得H(X')=Y；3.5，如果X和X'相差很小，H(X)和H(X')则完全不相关。

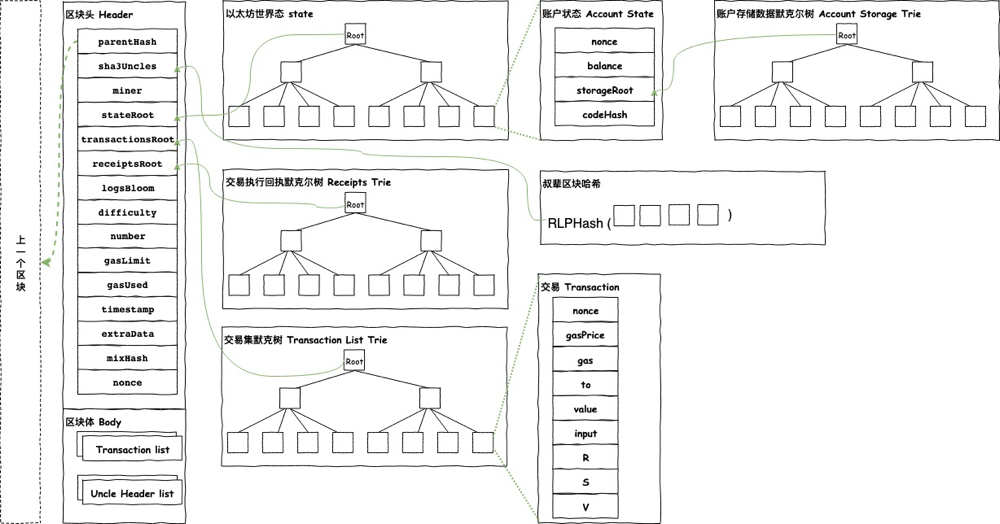
哈希函数主要用于验证信息完整性。在一个信息后面放上这个信息的哈希值，这个值很小（256bit），而且计算方便。收到信息之后收信人再算一遍哈希值，对比两者就知道这条信息是否被篡改过了。如果被篡改过，整个哈希值就会截然不同。而根据哈希函数的性质，没有人能够伪造出另一个消息具有同样的哈希值，也就是说篡改过的数据完全不可能通过哈希校验。

非对称加密：非对称加密就是有两把钥匙，一把叫公钥，一把叫私钥，用其中一把加密的话，只能用另一把解密，反之亦然。另一个重要的性质是，给你密文，明文和其中一把钥匙，你还是解不出来另一把钥匙是啥。原理是基于密码学数学难题，例如因数分解和离散对数，区块链常用的有Diffie-Hellman和ECC。

非对称加密除了和对称加密一样用于信息加密之外，还有另一个用途，就是身份验证。因为通常情况我们假设一对公私钥，公钥是公开的，而私钥只有本人有，于是一个人如果有对应的私钥，我们就可以认定他是本人。其中一个重要的应用就是数字签名——某个消息后面，发信人对这个消息做哈希运算，然后用私钥加密。接着收信人首先对消息进行哈希运算，接着用相应的公钥验证数字签名，再对比两个哈希值，如果相同，就代表这个消息是本人发出并且没有被篡改过。

以上是基础知识，至于区块链怎么实现的，很简单：

交易（数据）写在区块里。第一个区块叫创世区块，写啥都行。从第二个区块开始，每个区块的第一部分有前一区块的哈希值。此外，区块里的每一笔交易（数据），都有发起人的数字签名来保证真实性和合法性。于是，先前区块里的任何数据都不可被篡改。



图二.以太坊区块链结构

到这为止有人可能会问：为什么要弄个链啊？直接所有数据加个哈希值不就行了？

因为——这个数据库并不是静止的啊。

数据库的数据是会增加的，而每次增加的数据，就是一个区块，于是这些生成时间不同的区块，就以这种形式链在一起了。

至于如何增加区块，就涉及到第三个部分——共识算法。

## 采用共识算法来对于新增数据达成共识

共识算法的目的，就是让所有节点对于新增区块达成共识，也就是说，所有人都要认可新增的区块。对于有中心的系统，这事很简单，中心说什么大家同意就好了，但是放到去中心化系统里，尤其是当有些节点有恶意的时候，这东西非常复杂，计算机科学里有个相应的问题，叫做“拜占庭将军问题”或者“拜占庭容错”（BFT，这里暂不展开）。

很多年后，比特币的诞生，中本聪从某种意义上简化了这个问题，在比特币中，同样是共识问题，中本聪引入了一个重要的假设——奖励。他之所以能这样做的原因是，他考虑的是一个数字货币，也就是说共识这个东西是有价值的。

所有挖矿，矿工，最长链，分叉等，都可以归结为一句话：说话是要有代价的，说真话是有好处的，说假话是要扣钱的——工作证明机制。

在该共识机制设置之下，所有参与方都倾向于做诚实的参与者，共同维护新增数据的真实性，增加区块。