



区块链让数据真正“放心”流动起来

区块链以其可信任性、安全性和不可篡改性，让更多数据被解放出来。用一个典型案例来说明，即区块链是如何推进基因测序大数据产生的。区块链测序可以利用私钥限制访问权限，从而规避法律对个人获取基因数据的限制问题，并且利用分布式计算资源，低成本完成测序服务。区块链的安全性让测序成为工业化的解决方案，实现了全球规模的测序，从而推进数据的海量增长。

分布式的储存

大数据，指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合，是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。大数据需要应对海量化和快增长的存储，这要求底层硬件架构和文件系统在性价比上要大大高于传统技术，能够弹性扩张存储容量。

区块链，是比特币的底层技术架构，它在本质上是一种去中心化的分布式账本。

区块链技术作为一种持续增长的、按序整理成区块的链式数据结构，通过网络中多个节点共同参与数据的计算和记录，并且互相验证其信息的有效性。从这一点来说，区块链技术也是一种特定的数据库技术。由于去中心化数据库在安全、便捷方面

的特性，很多业内人士看好其发展，认为它是对现有互联网技术的升级与补充。区块链则是纯粹意义上的分布式系统。

分布式计算

大数据的分析挖掘是数据密集型计算，需要巨大的分布式计算能力。节点管理、任务调度、容错和高可靠性是关键技术。Google是这种分布式计算技术的代表，通过添加服务器节点可线性扩展系统的总处理能力，在成本和可扩展性上都有巨大的优势。现在，除了批计算，大数据还包括了流计算、图计算、实时计算、交互查询等计算框架。

区块链的共识机制，就是所有分布式节点之间怎么达成共识，通过算法来生成和更新数据，去认定一个记录的有效性，这既是认定的手段，也是防止篡改的手段。区块链主要包括四种不同的共识机制，适用于不同的应用场景，在效率和安全性之间取得平衡。以比特币为例，采用的是“工作量证明”，只有在控制了全网超过51%的记账节点的情况下，才有可能伪造出一条不存在的记录。

大数据与区块链的相同点

大数据，需要应对海量化和快速增长的存储，这要求底层硬件架构和文件系统在性价比上要大大高于传统技术，能够弹性扩张存储容量。谷歌的GFS和Hadoop的HDFS奠定了大数据存储技术的基础。另外，大数据对存储技术提出的另一个挑战是多种数据格式的适应能力，因此现在大数据底层的存储层不只是HDFS，还有HBase和Kudu等存储架构。

分布式存储：HDFS vs 区块

区块链，是比特币的底层技术架构，它在本质上是一种去中心化的分布式账本。

区块链技术作为一种持续增长的、按序整理成区块的链式数据结构，通过网络中多个节点共同参与数据的计算和记录，并且互相验证其信息的有效性。

从这一点来说，区块链技术也是一种特定的数据库技术。由于去中心化数据库在安全、便捷方面的特性，很多业内人士看好其发展，认为它是对现有互联网技术的升级与补充。

分布式计算：MapReduce vs. 共识机制

大数据的分析挖掘是数据密集型计算，需要巨大的分布式计算能力。节点管理、任

务调度、容错和高可靠性是关键技术。

Google和Hadoop的MapReduce是这种分布式计算技术的代表，通过添加服务器节点可线性扩展系统的总处理能力（Scale Out），在成本和可扩展性上都有巨大的优势。现在，除了批计算，大数据还包括了流计算、图计算、实时计算、交互查询等计算框架。

区块链的共识机制，就是所有分布式节点之间怎么达成共识，通过算法来生成和更新数据，去认定一个记录的有效性，这既是认定的手段，也是防止篡改的手段。区块链主要包括四种不同的共识机制，适用于不同的应用场景，在效率和安全性之间取得平衡。以比特币为例，采用的是“工作量证明”（Proof Of Work，简称POW），只有在控制了全网超过51%的记账节点的情况下，才有可能伪造出一条不存在的记录。

大数据与区块链的不同点

2011年，“大数据”第一次上榜，位于技术萌芽期的爬坡阶段，当时还统称为“Big Data and Extreme Information Processing and Management”（“大数据”和极端信息处理和管理）。2012年更进一步，并在2013年几乎达到了过热期顶峰。经历了2014年的下滑，从2015年开始，“大数据”突然从曲线中消失，可解读为Gartner对大数据的定位已从“新兴”转为“主流”。当前，大数据对于企业的意义已从能力要素上升为战略核心。相对而言，“区块链”直到2016年才第一次出现在《技术成熟度曲线》中，并直接进入“过热期”。总的来看，“大数据”和“区块链”所处的生命周期阶段大不相同，两者约有5年左右的差距。

大数据通常用来描述数据集足够大，足够复杂，以致很难用传统的方式来处理。区块链能承载的信息数据是有限的，离“大数据”标准还差得很远。

大数据与区块链的差异：

结构化 vs 非结构化：区块链是结构定义严谨的块，通过指针组成的链，典型的结构化数据，而大数据需要处理的更多的是非结构化数据。

独立 vs 整合：区块链系统为保证安全性，信息是相对独立的，而大数据着重的是信息的整合分析。

直接 vs 间接：区块链系统本身就是一个数据库，而大数据指的是对数据的深度分析和挖掘，是一种间接的数据。

数学 vs 数据：区块链试图用数学说话，区块链主张“代码即法律”，而大数据试图用数据说话。

匿名 vs 个性：区块链是匿名的（公开账本，匿名拥有者，相对于传统金融机构的公开账号，账本保密），而大数据有意的是个性化。

在区块链中使用大数据技术

区块链是一种不可篡改的、全历史的分布式数据库存储技术，巨大的区块链数据集包含着每一笔交易的全部历史，随着区块链技术的应用迅速发展，数据规模会越来越大，不同业务场景区块链的数据融合会进一步扩大数据规模和丰富性。

区块链以其可信任性、安全性和不可篡改性，让更多数据被解放出来，推进数据的海量增长。

区块链的可追溯性使得数据从采集、交易、流通，以及计算分析的每一步记录都可以留存在区块链上，使得数据的质量获得前所未有的强信任背书，也保证了数据分析结果的正确性和数据挖掘的效果。

区块链能够进一步规范数据的使用，精细化授权范围。脱敏后的数据交易流通，则有利于突破信息孤岛，建立数据横向流通机制，形成“社会化大数据”。基于区块链的价值转移网络，逐步推动形成基于全球化的数据交易场景。

区块链提供的是账本的完整性，数据统计分析的能力较弱。大数据则具备海量数据存储技术和灵活高效的分析技术，极大提升区块链数据的价值和使用空间。