

区块链存储：技术与挑战

蔡振华, 林嘉韵, 刘芳

(中山大学数据科学与计算机学院, 广东 广州 510006)

摘要: 区块链具有不可篡改和去中心化的特性, 将其应用到存储系统中, 有助于提高存储系统的数据安全性和系统可扩展性。区块链技术与存储技术的融合主要分为 3 方面: 基于区块链构建的去中心化存储系统、基于区块链优化已有系统的存储性能和区块链系统自身的存储方法与优化。对这些工作进行了综述, 讨论对比了当前典型的采用了区块链技术的存储系统, 总结了区块链存储面临的主要挑战, 并展望了区块链存储的发展方向。

关键词: 区块链存储; 去中心化; 分布式存储; 数据安全性; 可扩展性

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2020019

Blockchain storage: technologies and challenges

CAI Zhenhua, LIN Jiayun, LIU Fang

School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China

Abstract: Blockchain's characteristics of non-temper and decentralization can be utilized to improve the data security and scalability of storage systems, called blockchain storage. These works that combine storage with blockchain technology, are mainly divided into three categories: decentralized storage systems based on blockchain, improve the performance of existing storage systems based on blockchain, and the underlying storage strategies of blockchain. These works were summarized, the current blockchain storage systems were compared, the main challenges that blockchain and storage being facing were discussed and its future development was forecasted.

Key words: blockchain storage, decentralization, distributed storage, data security, scalability

1 引言

随着大数据、物联网和边缘计算的高速发展, 数据规模呈指数级别增长。存储系统作为数据存

储的基础设施, 面临更大容量、更安全可靠、更低访问延迟等诸多挑战。传统的大规模存储系统存在缺乏安全保障、可扩展性受限等问题, 难以满足大数据应用的需求^[1-2]。

收稿日期: 2019-08-12; 修回日期: 2019-11-20

通信作者: 刘芳, liufang25@mail.sysu.edu.cn

基金项目: 国家自然科学基金(61832020, 61702569); 国家重点研发计划(2016YFB1000302)

Foundation Items: The National Natural Science Foundation of China (61832020, 61702569), The National Key R&D Program of China (2016YFB1000302)

论文引用格式: 蔡振华, 林嘉韵, 刘芳. 区块链存储: 技术与挑战[J]. 网络与信息安全学报, 2020, 6(5): 11-20.

CAI Z H, LIN J Y, LIU F. Blockchain storage: technologies and challenges[J]. Chinese Journal of Network and Information Security, 2020, 6(5): 11-20.

区块链技术具有防篡改和去中心化的特性,已经被广泛应用于金融、银行、供应链等多个领域,相比传统的中心化解决方案,区块链可以有效地保证数据安全和系统架构的去中心化,解决机构间合作的信任和效率问题^[1-2]。

在分布式存储系统中引入区块链技术,可以提高存储系统的安全性和可扩展性。然而,将区块链应用到存储系统中,也带来了新的挑战,如区块链的查询性能瓶颈、存储空间开销增大等。

如图 1 所示,当前区块链技术与存储技术的融合主要有 3 个方向的工作。

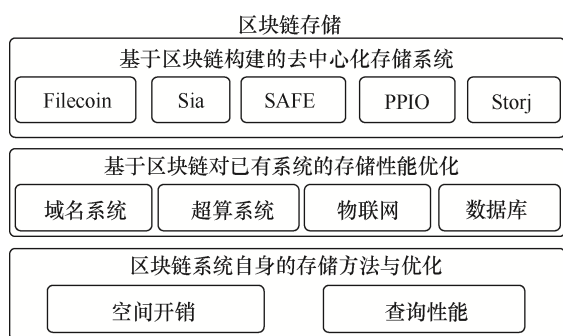


图 1 区块链存储研究分类
Figure 1 Classification of blockchain storage research

1) 基于区块链构建的去中心化存储系统。早期的去中心化存储基于对等网络实现,由个人用户提供存储空间,如 Ocean Store^[3]和星际文件系统(IPFS, inter-planetary file system)等,但是这些系统没有很好地解决存储空间激励的问题,难以让用户长期存储他人的文件。区块链技术的引入,弥补了这方面的不足,利用区块链的虚拟货币(Tokens),可以有效激励用户提供存储空间。具有代表性的基于区块链的去中心化存储系统,有结合 IPFS 与区块链技术的 Filecoin,以及开源项目 Sia、Storj、SAFE Network 等。

2) 基于区块链优化已有系统的存储性能。针对中心化架构系统面临的单点故障、数据安全性低、隐私保护能力不足等问题,将区块链技术应用于域名系统、物联网系统、超级计算系统、数据库系统中,为系统设计去中心化的架构、利用分布式账本提高数据安全性与数据溯源能力。

3) 区块链系统自身的存储方法与优化。针对

区块链的存储空间利用率低、查询性能低等问题进行优化,如采用纠删码降低区块链的存储空间开销^[4-5],使用索引等技术提高区块链系统的查询效率。

2 基于区块链构建去中心化的存储系统

中心化存储系统存在服务器单点故障、数据丢失、隐私泄露等问题,在对等网络等技术相对成熟的背景下,去中心化存储应运而生。在去中心化存储系统中,用户上传到存储系统的文件被切分、加密、分散保存在多个用户提供的存储空间中。目前有大量基于 IPFS 开发的去中心化项目,如匿名交易平台 OpenBazaar、社交媒体网络 AKASHA 等。

然而,去中心化存储系统面临存储激励问题,即如何激励用户提供存储空间,长期保存其他用户的文件。将区块链技术作为去中心化存储系统的激励层,可以有效缓解该问题。引入区块链之后,一方面,用户通过提供存储空间获得虚拟货币奖励,为租用存储空间、访问文件支付费用,建立了良好的经济模式^[6];另一方面,利用分布式账本保存元数据信息、控制访问权限,提高了系统的数据安全性^[7-9]。

2.1 基于区块链的去中心化存储工作方式

以文件的上传和下载过程分析去中心化存储系统的工作方式。基于区块链的去中心化存储的文件上传过程如图 2 所示。

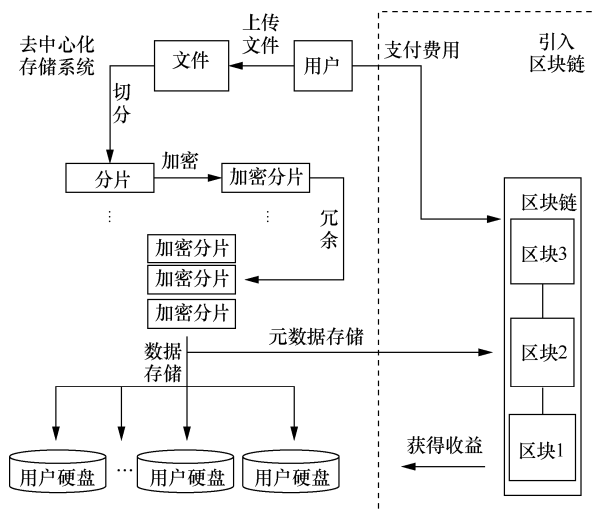


图 2 基于区块链的去中心化存储的文件上传过程
Figure 2 File upload process of blockchain based decentralized storage

虚线框部分代表引入区块链机制之后的变化有两方面。一方面，区块链账本存储文件分片的元数据，可保证数据防篡改、可溯源。另一方面，将区块链发行的虚拟货币作为奖励，可以激励用户提供存储空间。

文件上传的算法如下。

算法 1 文件上传

chunks = Cut(file)//把文件切分成相同大小的分片

For c in chunks:

c = Encrypt(c)//对文件分片进行加密

cs = Replicate(c)//将加密分片通过纠删码编码成冗余分片

m = Store(cs)//将冗余分片存储到分布式节点中

hash = Hash(c)//生成加密分片的哈希值

StoreLedger(hash, m)//将分片哈希值与元数据等信息存储到区块链

基于区块链的去中心化存储的文件下载过程如图 3 所示。

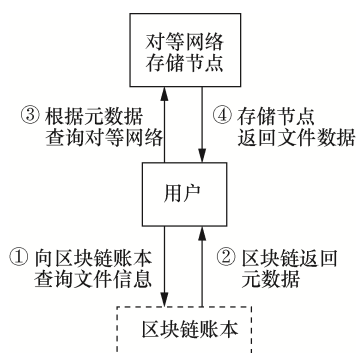


图 3 基于区块链的去中心化存储的文件下载过程

Figure 3 File download process of blockchain based decentralized storage

虚线框部分代表存储元数据的区块链账本。用户通过文件名的哈希值向区块链账本发起查询，后者返回文件的元数据信息；用户根据元数据到对等网络的存储节点中获取数据，后者返回

文件数据。

2.2 基于区块链的去中心化存储的特点

去中心化存储系统引入区块链后，不仅鼓励了用户更积极地提供存储空间，还提高了数据安全性。中心化存储、去中心化存储和基于区块链的去中心化存储系统对比如表 1 所示。

由表 1 可知，基于区块链的去中心化存储有如下特点。

1) 隐私保护能力强。经过分片、加密后存储的文件，在其他用户硬盘中显示为部分分片的密文数据，只有通过数据拥有者的密钥才可以查看完整数据，有效地防止了数据泄露。同时，通过区块链控制数据访问权限，增强了隐私保护能力。

2) 数据安全性强。去中心化的存储系统架构缓解了中心化存储面临的单点故障等问题。此外，区块链技术的防篡改与可溯源特性，提高了数据安全性。

3) 响应速度快。由于去中心化的分布式架构，存储设备分散在不同地区，可以同时快速响应多地的设备请求，有利于提高系统的数据收集和处理能力，加快物联网应用和边缘计算的响应速度。

4) 下载速度快。文件分片被存储在不同节点，用户下载文件时能够以并行方式进行。

5) 闲置存储资源利用率高。激励用户提供存储资源，可以提高网络上大量闲置存储空间的利用率。

6) 存储激励能力强。在去中心存储系统中，需要设定良好的经济模式，以持续激励用户提供存储空间。区块链发行的虚拟货币，为存储激励提供了一个可行的解决方案。中心化存储系统不涉及激励。

7) 存储空间开销大。从用户设备的角度来看，中心化存储并不需要额外的存储空间开销，

表 1 中心化、去中心化、基于区块链的去中心化存储对比
Table 1 Comparison of centralized storage, decentralized storage and decentralized storage based on blockchain

存储类型	隐私保护	数据安全性	响应速度	下载速度	闲置存储资源利用	存储激励能力	存储空间开销
中心化存储	弱	弱	慢	慢	低		小
去中心化存储	中	中	快	快	高	弱	中
基于区块链的去中心化存储	强	强	快	快	高	强	较大

注：空白表示不适用。

去中心化存储需要保存其他用户的数据, 引入区块链之后需要保存分布式账本, 存储空间开销较大。

2.3 基于区块链的去中心化存储开源项目

基于区块链构建的去中心化存储系统引起了广泛的关注, 本小节介绍几个具有代表性的去中心化存储开源项目。

Filecoin 是普林斯顿大学的研究项目, 它是构建于 IPFS 上层的去中心化的存储网络, 作为 IPFS 的激励层存在。PPIO 是 PPLabs 团队开发的去中心化的数据存储和分发平台, 致力于提高数据的分发能力。Storj 是 StorjLabs 公司发起的去中心化存储网络项目, 它根据用户提供的存储空间容量, 按月发放虚拟货币奖励, Storj 兼容 AmazonS3 存储系统, 使第三方平台更容易集成、迁移到 Storj。SAFENetwork (safe access for everyone network)^[6] 是 MaidSafe 公司开发的项目, 主张建立一个任何人都能安全访问的去中心化网络, 保护用户的隐私和数据自治权。Sia^[4] 是 Nebulous Labs 公司创建的去中心化存储项目, 通过保存在区块链上的智能合约, 自动维护存储提供方与用户的合约关系, 无须依靠可信第三方。

从技术手段、空间利用率、数据安全性、项目状况、应用场景等方面对上述 5 个项目进行比较, 如表 2 所示。

1) 技术手段。在数据冗余方面, Filecoin、PPIO、Storj、Sia 采用了纠删码技术, 提高了数

据的可靠性。在存储空间开销方面, Filecoin、PPIO、Storj、SAFENetwork 采用了重复数据删除技术, 提高了存储空间利用率。

2) 空间利用率。在存储系统中, 重复数据删除技术可以有效降低数据冗余, 纠删码在一定程度上增加了数据冗余。SAFE Network 采用了重复数据删除技术, 空间利用率最高; Filecoin、PPIO、Storj 两者兼用, 次之; Sia 采用了纠删码技术, 空间利用率最低。

3) 数据安全性。除了 SAFE Network 未提及, 其他系统均采用了纠删码的数据冗余方式。除了 Filecoin 需要用户自行加密, 其他系统均采用了加密算法, 保证了数据的可靠性和私密性, 安全性高。

4) 项目状况。各项目的关键时间节点如图 4 所示。

PPIO 仅开放部分代码, 其他项目均已在 GitHub 或 GitLab 开源。Sia 在 2016 年发布了正式版本, 其他项目均处于测试阶段。

5) 应用场景。虽然上述去中心化存储系统都可用于存储数据、管理数据和出租存储空间, 但侧重点不同。如图 5 所示, 用户可以根据自身需求和存储场景选择合适的系统。

Filecoin 和 Sia 适合作为存储基础设施。其中, Filecoin 希望建立一个去中心化的存储市场, 将存储与数据的定价交给市场而不是企业, 它是基于 IPFS 构建的, 适合基于 IPFS 进行开发的用户;

表 2 5 个代表性的去中心化存储项目比较
Table 2 Comparison of 5 representative decentralized storage projects

项目	适用场景	操作系统	重复数据删除	纠删码	加密方式	用户访问接口	目标用户	编程框架	兼容亚马逊 S3	空间利用率	数据安全性	项目状况
Filecoin	数据存储	Linux、MacOS	支持	支持	用户自行加密	API	一般用户	Go、Rust	不兼容	中	高	GitHub 开源、内部测试
PPIO	数据分发	Windows、MacOS、Linux	支持	支持	系统加密	命令行	开发者	Node.js、Android	不兼容	中	高	GitHub 部分开源、内部测试
Storj	数据存储	Linux、MacOS、Windows	支持	支持	系统加密	命令行、图形界面	一般用户、开发者	C	兼容	中	高	GitHub 开源、内部测试
SAFE Network	数据自治	Linux、MacOS	支持	不支持	系统加密	命令行、浏览器	一般用户、开发者	Node.js、Web、C#、Java、Rust	不兼容	高	中	GitHub 开源、内部测试
Sia	数据存储	Windows、MacOS、Linux	不支持	支持	系统加密	命令行、图形界面	一般用户	Go	不兼容	低	高	GitLab 开源、正式发布

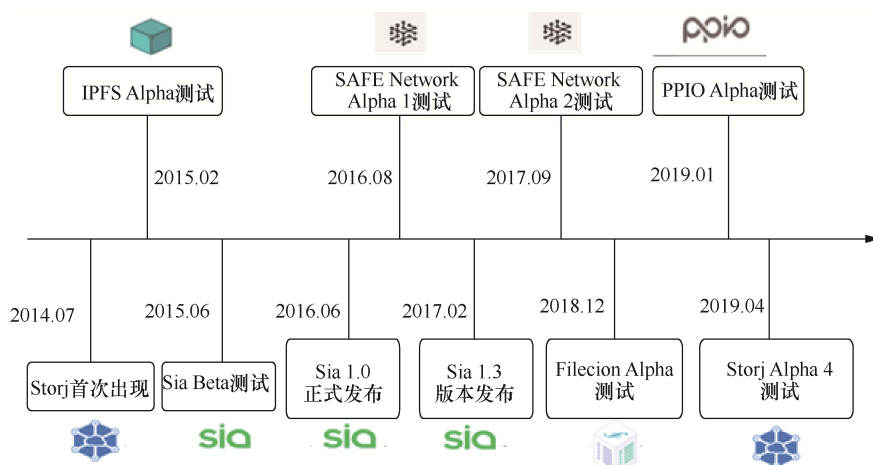


图 4 基于区块链的去中心化存储项目进展
Figure 4 Progress of decentralized storage projects based on blockchain

Sia 专注于为去中心化存储提供一个功能丰富的平台，发布了正式版本，能够提供更稳定、成熟的服务，在静态内容，如视频、音频等数据的存储方面表现良好。

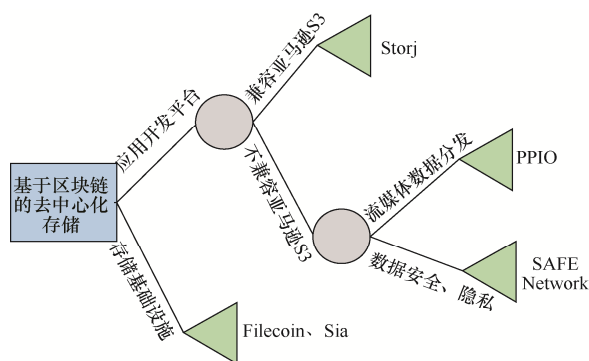


图 5 系统应用场景决策树
Figure 5 System application scene decision tree

PPIO、Storj 与 SAFE Network 可以作为应用开发平台。PPIO 关注流媒体数据的服务质量，适合音视频应用的开发者；Storj 兼容亚马逊 S3 存储系统，适合希望从 S3 迁移的用户；SAFE Network 看重用户的隐私与数据自主权，希望建立一个言论自由、隐私受保护、安全的通信网络，适合对隐私与安全要求更高的场景。

3 基于区块链优化已有系统的存储性能

由于区块链技术的去中心化和防篡改特性，人们研究如何应用区块链技术优化已有系统的存储性能，这些研究工作涉及域名系统、超算系统、

数据库、物联网等领域。

针对域名系统的单点故障和信任问题，普林斯顿大学研发了基于区块链的域名系统 Blockstack^[10]，该系统具有去中心化和无须信任的特性，可以提高域名系统的安全性和可扩展性。在该系统中，区块链记录域名及其对应地址的路由信息、地址信息被存储在第三方云存储服务器上。

Blockstack 架构如图 6 所示。

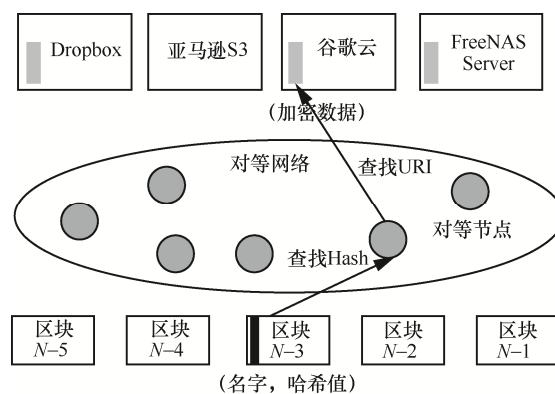


图 6 Blockstack 架构
Figure 6 Architecture of Blockstack

数据的查询路径为：在区块链上查找域名对应的哈希值；通过哈希值查找对等网络节点的路由文件；根据路由文件中的地址信息访问云存储服务器获取地址信息。Blockstack 不仅可以应用于域名系统，也可以用于存储其他类型的数据。

超级计算系统中，通常需要进行数据溯源。

针对现有基于数据库与文件系统的方法准确度不高、I/O 开销大等问题, In_memory Blockchain^[11] 将账本放置在超级计算系统的计算节点内存中, 在计算节点和存储节点(对应易失内存与非易失存储)上验证数据的合法性, 防止内存数据丢失影响数据的准确性, 该方案具有更高的数据溯源性能。

数据库通常是中心化的管理方式, 由管理员控制整个系统, 存在数据冗余度高、安全性不足等问题。区块链技术有助于解决上述问题, 数据库和区块链技术的结合, 可以构建既有数据管理能力, 又具备去中心化、防篡改等特性的区块链数据库。

具有代表性的区块链数据库有 BigchainDB、ChainSQL、RecordsKeeper 等项目, 具有吞吐量大、防篡改、可溯源和去中心化的特点。BigchainDB 能够存储 PB 级别数据, 写性能达到每秒 100 万次, 延迟在亚秒级别。ChainSQL 将对数据表的所有操作以日志形式记录在区块链账本上, 提供数据防篡改服务, 支持多种数据库如 MySQL、Oracle 等。RecordsKeeper 是一个开源的 Key-Value 数据存储平台, 通过分布式账本与数据加密技术提供可验证和安全的数据存储服务。

BeeKeeper^[12]将区块链技术引入物联网系统中, 使设备能够自由加入和退出物联网, 提高了系统的可拓展性。Shafagh 等^[13]针对物联网中存在的数据隔离问题, 将物联网设备产生的数据存储到云端, 提高数据共享能力, 并利用区块链控制权限, 保证数据安全。对于自动化销售系统, Yoo 等^[14]引入区块链和智能合约, 利用智能合约代替人力执行操作, 降低了管理成本, 提高了系统的安全性。针对物联网数据分析任务的网络传输带宽要求高的问题, Xu 等^[15]提出在物联网设备上执行特定数据分析操作的方法, 减少设备间的数据传输, 加快数据分析任务的执行速度。Wu 等^[16]在电网系统中应用区块链技术, 基于联盟链构建了智能电网系统, 缓解了单点故障、数据被篡改等问题。

4 区块链系统的存储方法与优化

区块链系统自身的存储同样面临着挑战。在

存储空间开销方面, 全副本和区块链膨胀问题导致了庞大的存储开销。在 I/O 性能方面, 大部分区块链使用了 LevelDB 作为底层数据库, 该数据库读速度慢^[17], 影响了区块链的查询性能。因此, 需要对区块链系统的存储方法进行优化。本节主要介绍降低存储空间开销、提高查询性能两个方面的工作。

4.1 降低区块链的存储空间开销

当前主流的区块链系统内主要有全节点和轻节点两种节点, 以匹配不同节点的存储容量^[18]。由于区块链上的数据不可删除, 设备需要存储的数据量将不断增长。将纠删码引入区块链系统, 有助于减小全节点的负担, 只存储编码区块的节点称为纠删码节点。一个完整区块被分成 K 块, 采用纠删码编码成 $K+R$ 个编码区块, 任意 K 个编码区块都可以恢复出完整区块。纠删码节点只需存储其中一个编码区块, 即完整区块的 $\frac{1}{K}$, 因此降低了存储负担。

4.1.1 基于纠删码的区块链

区块链系统通过将纠删码的编/解码过程嵌入区块链的打包与同步过程来集成纠删码。

1) 嵌入编码过程。纠删码节点在同步新的区块文件时, 需要先对文件进行编码操作, 再存储编码后的区块文件, 如图 7 所示。区块链系统中参与挖矿的节点负责记账、产生区块、并将新的区块广播到全网。为了方便后续的解码操作, 还需生成相应的解码信息, 一并存入纠删码节点中。

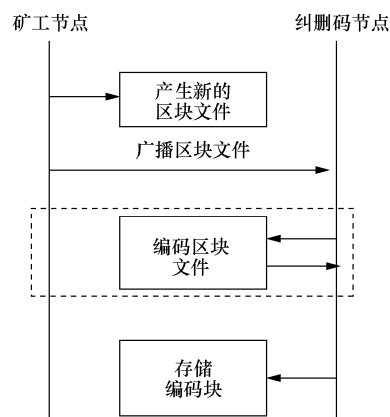


图 7 在区块链中嵌入纠删码编码过程(虚线框内为引入纠删码技术后的过程)

Figure 7 Encoding process of erasure code in blockchain (the additional process after the integration of erasure code is in the dotted frame)

2) 嵌入解码过程。当新的节点加入系统时,它需要从邻近节点处同步区块链。由于邻近节点可能只保存了部分编码块,无法提供完整的区块文件,因此在进行区块同步时,需要首先判断邻近节点的存储数据类型,即存储的是编码块还是完整区块,如果不是完整区块,则需要先进行区块文件解码,恢复出完整区块再进行同步,如图8所示。

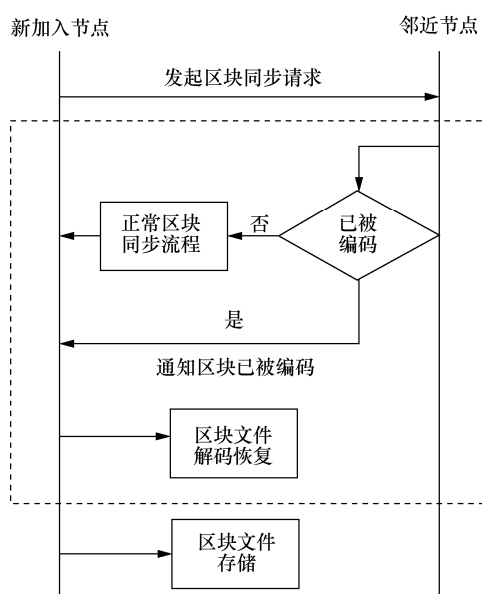


图8 在区块链中嵌入纠删码解码的过程(虚线框内为集成纠删码技术后添加的过程)

Figure 8 Decoding process of erasure code in blockchain (the additional process after the integration of erasure code is in the dotted frame)

4.1.2 纠删码节点的优势

全节点、轻节点和纠删码节点的比较如表3所示。

相比全节点和轻节点,纠删码节点有以下4个优势。

1) 存储需求低。纠删码不仅降低了对节点存储资源的需求,而且有效改善了区块链系统的可扩展性,即节点可以根据自身容量大小,选择存储的编码块数量,即使是小型设备也能

够参与区块链系统,有助于将区块链技术扩展到物联网。

2) 网络带宽开销小。通过纠删码降低对设备存储容量的要求,有助于增加物联网设备的数量,从而在初始化和同步节点时减少网络带宽开销。

3) 数据可用性高。降低节点的存储需求,允许更多低存储容量的节点参与存储,增加系统中参与区块重构的节点比例,进而提高区块链系统的可用性。

4) 数据安全性高。引入纠删码后,若要篡改纠删码节点中存储的数据,需要保证篡改前后区块的哈希值一致,并且使该编码块与其他任意编码块的组合的解码结果相同,这是难以实现的。相当于多了一种区块链验证手段,这样有利于维护区块链的安全性。

4.2 区块链的查询性能优化

区块链系统常用 LevelDB 数据库存储数据,该数据库以日志合并树(LSM-Tree, the log-structured merge-tree)作为数据结构,该数据库的读速度慢,且一次查询可能需要执行额外多次内部查询,影响了区块链的查询性能。下面介绍两种提高查询性能的方法,分别是在区块链上构建查询层和在区块链内部加入辅助索引。

4.2.1 外联数据库

外联数据库的思想来自 EtherQL^[19],通过在区块链外部设置数据库,监听并同步区块链数据到数据库;通过数据库接口进行数据查询,该数据库架构如图9所示。

4.2.2 内置索引

内置索引的方法在区块链系统内部建立查询层。王千阁等^[20]提出为主键建立辅助索引,提高区块链系统的查询性能。在该方法中,主键指向的是数据的物理地址,而辅助索引指向的是数据对应的主键,架构如图10所示。

表3 3种区块链节点的比较
Table 3 Comparison of 3 types of nodes in blockchain

节点类型	是否参与验证	对信任的依赖	存储内容	存储空间需求	存储容量需求	是否参与区块重构
全节点	参与	低	整个区块链	高	高	参与
轻节点	不参与	高	区块头部	低	低	不参与
纠删码节点	参与	低	部分编码块和解码信息	中	低	参与

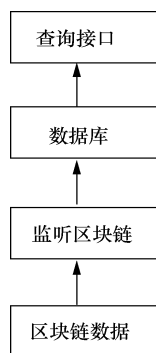


图 9 外联数据库架构
Figure 9 Architecture of external database

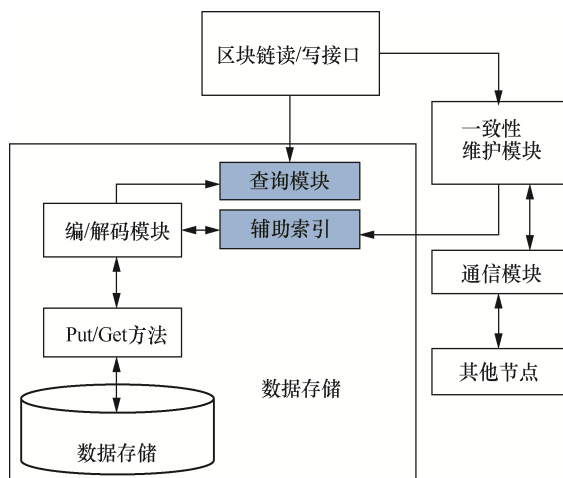


图 10 内置索引架构
Figure 10 Architecture of built-in index

焦通等^[21]扩展了区块链中数据交易的数据结构,增加了交易信息的索引记录,相比使用哈希值进行查找交易的方法,该方法丰富了区块链系统的查询语义。

4.2.3 查询性能优化方法对比

本节从查询速度、额外空间开销、性能影响、可扩展性等方面对比了外联数据库与内置索引方法,如表 4 所示。

1) 查询速度。使用外联数据库和内置索引的方式有助于提高区块链的数据查询速度。以 select 为例,理论分析单次查询的时间复杂度, $Ts(k)$ 为系统读取耗时, k 为选择度, $T(N)$ 为 LevelDB 执

行一次随机查找的耗时,外联数据库较内置索引更快^[20]。

2) 存储空间开销。外联数据库需要存储区块链的全部数据和索引,内置索引只需要增加索引信息,因此前者的存储空间开销较大。

3) 性能影响。外联数据库对数据的查找和更新不影响区块链,而内置索引方法需要同步索引信息,降低了系统的吞吐量。

4) 可扩展性。外联数据库的部署、扩展容易。内置索引需要修改系统内部代码,可扩展性较低。

5 区块链存储的挑战与趋势

区块链存储提高了系统的存储性能,但面临着诸多挑战。

1) 分布式账本的存储空间开销大。分布式账本是区块链技术的核心,然而,账本的多副本特性需要大量的额外存储空间,增加了存储成本,导致空间利用率有限,难以支撑大规模应用。如大规模物联网中包括大量异构的小容量设备,节点存储能力有限,难以持续满足分布式账本不断膨胀的存储需求。

2) 区块链系统查询性能的提高带来了新的问题。使用索引、外联数据库等技术,能够提高区块链存储的查询性能。然而,这些技术通过空间换取时间,降低了查询时间,却增加了存储空间开销,影响了系统的可扩展性,难以较好地满足多方面的需求。

3) 去中心化存储系统不够稳定、可靠。去中心化存储通过引入区块链技术得到改进,然而,去中心化存储仍面临着稳定性不足、冷启动引起的系统性能抖动等问题。当在线的存储节点较少时,可用文件的数量随之减少,使数据可靠性、服务质量降低,影响用户体验。同时,目前大部分的去中心化存储系统并未给出存储的耐久性数据,难以证明存储系统的可靠性。

表 4 外联数据库与内置索引的比较
Table 4 Comparison of external database method and built-in index

方法	查询速度	外空间开销	性能影响	可扩展性
外联数据库	$O(\log(N + Ts(kN)))$	大	小	高
内置索引	$O(\log(N) + kT(N))$	小	大	低

4) 数据隐私与监管方式尚不成熟。区块链上记录的数据交易是公开透明的, 对于数据的共享和验证有利, 却不利于用户的隐私保护。用户可能通过匿名平台进行非法交易、发布与传播非法内容, 对于这些内容需要监管。

5) 区块链存储结构和系统发展的可扩展性不足。区块链的链状存储结构可能演变为树状、网状、格点状等结构, 区块链也发展出了分片、可编辑技术。存储系统需要适应新技术的不断演化, 因此要求区块链存储系统具有良好的可扩展性, 能够以较小的修改开销满足对新特性的支持。

6) 区块链技术对非易失存储系统的寿命与性能的挑战。非易失存储介质的写/擦除次数有限, 存在使用寿命有限的问题, 且写延迟和写功耗相对较高。面对读写不平衡、随机小写频繁的区块链应用, 非易失存储系统的寿命和平均 I/O 性能问题将更为突出, 需要研究如何减少写次数和写入数据量, 提高系统整体访问性能和延长非易失存储介质的使用寿命。

由这些挑战可预见, 区块链存储未来研究的趋势主要有以下 4 方面。

1) 低存储空间开销的区块链存储系统。在大规模网络中, 设备的存储能力常常是异构的, 需要考虑小容量设备的存储能力。低空间开销的区块链存储系统, 有利于区块链部署到大规模异构网络中, 适应更多类型的系统与场景。纠删码技术能够降低存储空间开销, 然而, 纠删码与区块链技术结合的工作尚不够完善。关于纠删码节点的比例、节点容量溢出、编码系数设置和区块同步等问题亟待解决。

2) 高效的区块链存储查询技术。区块链存储系统中需要对数据进行查询, 然而, 其查询性能受限于区块链的数据结构、查询方法、架构设计等多个方面。在保证低存储空间开销的同时, 提高其查询性能将有利于区块链存储在更多场景的应用。

3) 可靠稳定的去中心化存储服务。基于区块链的去中心化存储系统是未来的发展趋势, 对于该存储系统的数据可靠性、服务稳定性的研究, 将成为重要部分, 使去中心化存储能够取代中心化的存储系统。

4) 隐私数据的保护与非法内容监管。在区块链系统中, 数据记录是公开透明的, 用户身份是匿名的。信息世界的内容传播速度快, 对于用户的隐私需要进行保护, 同时, 对于非法的监管至关重要。需要平衡用户隐私与监管力度, 防止非法内容的传播。

6 结束语

区块链存储结合区块链技术和存储技术, 提高了存储系统的安全性和可扩展性, 在许多应用领域有着巨大潜力, 对未来的全球存储融合、去中心化存储趋势有着巨大的推动作用。本文总结了目前区块链与存储技术相结合的 3 方面的工作: 基于区块链构建的去中心化的存储系统; 基于区块链优化已有系统的存储性能; 区块链系统自身的存储方法与优化。对以上 3 方面的工作进行分析与比较, 展示了这些工作的进展和特点。最后, 展望了当前区块链存储所面临的挑战。

参考文献:

- [1] 沈鑫, 裴庆祺, 刘雪峰. 区块链技术综述[J]. 网络与信息安全学报, 2016, 2(11): 11-20.
SHEN X, PEI Q Q, LIU X F. Survey of block chain[J]. Chinese Journal of Network and Information Security, 2016, 2(11): 11-20.
- [2] HU W H, AO M, SHI L, et al. Review of blockchain-based DNS alternatives[J]. 网络与信息安全学报, 2017, 3(3): 71-77.
HU W H, AO M, SHI L, et al. Review of blockchain-based DNS alternatives[J]. Chinese Journal of Network and Information Security, 2017, 3(3): 71-77.
- [3] KUBIATOWICZ J, BINDEL D, CHEN Y, et al. OceanStore: an architecture for global-scale persistent storage[J]. ACM SIGPLAN Notices, 2000, 35(11): 190-201.
- [4] DAI M J, ZHANG S L, WANG H, et al. A low storage room requirement framework for distributed ledger in blockchain[J]. IEEE ACCESS, 2018, 6: 22970-22975.
- [5] PERARD D, LACAN J, BACHY Y, et al. Erasure code-based low storage blockchain node [C]//IEEE Symposium on Recent Advances on Blockchain and Its Applications. 2018: 1622-1627.
- [6] 陈晓, 吕欣冉, 刘志. 基于隐私保护的法定“数字货币”激励机制[J]. 网络与信息安全学报, 2019, 5(2): 30-39.
CHEN X, LYU X R, LIU Z. Truthful incentive mechanism for digital currency based on privacy protection[J]. Chinese Journal of Network and Information Security, 2019, 5(2): 30-39.
- [7] FU Y. Meta-key: a secure data-sharing protocol under blockchain-based decentralised storage architecture[J]. arXiv Preprint arXiv:1710.07898, 2017.

- [8] LI J X, LIU Z S, CHEN L, et al. Blockchain-based security architecture for distributed cloud storage [C]//IEEE International Symposium on Parallel and Distributed Processing with Applications(ISPA). 2017: 408-411.
- [9] DO H G, NG W K. Blockchain-based system for secure data storage with private keyword search[C]//IEEE World Congress on Services. 2017.
- [10] ALI M, NELSON J, SHEA R, et al. Blockstack: a global naming and storage system secured by blockchains[C]//USENIX Annual Technical Conference. 2016: 181-194.
- [11] AL-MAMUN A, LI T, SADOOGHI M, et al. In-memory blockchain: toward efficient and trustworthy data provenance for HPC systems[C]//IEEE International Conference on Big Data (Big Data). 2018: 3808-3813.
- [12] ZHOU L J, WANG L C, SUN Y R, et al. BeeKeeper: a blockchain-based IOT system with secure storage and homomorphic computation[J]. IEEE Access, 2018, 6: 43472-43488.
- [13] SHAFAGH H, BURKHALTER L, HITHNAWI A, et al. Towards blockchain-based auditable storage and sharing of IoT data[C]//The 9th Cloud Computing Security Workshop. 2017: 45-50.
- [14] YOO M, WON Y. Study on smart automated sales system with blockchain-based data storage and management[C]//Advances in Computer Science and Ubiquitous Computing - CSA/CUTE. 2017: 734-740.
- [15] XU Q Q, AUNG KMM, ZHU Y Q, et al. A blockchain-based storage system for data analytics in the internet of things[M]//New Advances in the Internet of Things, Studies in Computational Intelligence, Springer International Publishing. 2018: 119-138.
- [16] WU Z Q, LIANG Y H, KANG J W, et al. Secure data storage and sharing system based on consortium blockchain in smart grid[J]. Journal of Computer Applications, 2017, 37(10): 2742-2747.
- [17] KAIYRAKHMET O, LEE SONGYI, NAM B, et al. SLM-DB: single-level key-value store with persistent memory[C]//The 17th USENIX Conference on File and Storage Technologies(FST). 2019: 191-205.
- [18] 袁勇, 王飞跃. 区块链技术发展现状与展望[J]. 自动化学报, 2016, 42(4): 481-494.
YUAN Y, WANG F Y. Blockchain: the state of the art and future trends [J]. Acta Automatica Sinica, 2016, 42(4): 481-494.
- [19] LI Y, ZHENG K, YAN Y, et al. EtherQL: a query layer for blockchain System[C]//International Conference on Database Systems for Advanced Applications. 2017: 556-567.
- [20] 王千阁, 何蒲, 聂铁铮, 等. 区块链系统的数据存储与查询技术综述[J]. 计算机科学, 2018, 45(12): 12-18.
WANG Q G, HE P, NIE T Z, et al. Survey of data storage and query techniques in blockchain systems[J]. Computer Science, 2018, 45(12): 12-18.
- [21] 焦通, 申德荣, 聂铁铮, 等. 区块链数据库:一种可查询且防篡改的数据库[J]. 软件学报, 2019, 30(9): 2671-2685.
JIAO T, SHEN DR, NIE T Z, et al. Blockchain DB: a querable and immutable database[J]. Journal of Software, 2019, 30(9): 2671-2685.

[作者简介]



蔡振华 (1995-), 男, 福建泉州人, 中山大学硕士生, 主要研究方向为存储系统、计算机体系结构。



林嘉韵 (1998-), 女, 广东珠海人, 主要研究方向为存储系统、存储系统可靠性。



刘芳 (1976-), 女, 四川绵阳人, 博士, 中山大学副教授、博士生导师, 主要研究方向为计算机系统结构、新型存储、边缘计算。