区块链系统中的分布式数据管理技术——挑战与展望

于 戈 聂铁铮 李晓华 张岩峰 申德荣 鲍玉斌

(东北大学计算机科学与工程学院 沈阳 110169)

摘 要 区块链是在数字加密货币的应用基础之上发展起来的一种分布式数据库技术.区块链系统具有去中心化、不可篡改、分布共识、可溯源和最终一致性等特点,这使其可以用于解决不可信环境下数据管理问题.区块链独特的数据管理功能已经成为各领域应用中发挥区块链价值的关键.本文基于对比特币、以太坊、超级账本等代表性区块链系统的研究分析,阐述区块链系统中分布式数据管理技术.首先,深入讨论区块链系统与传统分布式数据库系统之间的异同点,从分布式部署模式、节点角色、链拓扑结构等多个方面给出区块链的分类.然后,详细分析各类区块链系统所使用的数据存储结构、分布式查询处理与优化技术及其优缺点.最后,总结区块链系统的分布式数据管理技术在各专门领域应用中所面临的挑战和发展趋势.

关键词 区块链;分布式数据管理;数据存储;查询处理

中图法分类号 TP311 **DOI 号** 10.11897/SP.J.1016.2021.00028

The Challenge and Prospect of Distributed Data Management Techniques in Blockchain Systems

YU Ge NIE Tie-Zheng LI Xiao-Hua ZHANG Yan-Feng SHEN De-Rong BAO Yu-Bin

(*School of Computer Science and Engineering, Northeastern University, Shenyang* 110169)

Abstract Blockchain is a technique of distributed database which is developed with the applications of digital encrypted currency. A blockchain system has the characteristics of decentralization, non-tampering, distributed consensus, provenance and eventual consistency, which makes it be applied to solve data management problems of the untrusted environments. The data management function of a blockchain system has already become the important feature for playing its value in the applications of different domains. Blockchain systems make every node contain a complete copy of ledger data, and use distributed consensus algorithms to ensure the consistency of data. Therefore, a blockchain system is a new kind of distributed data management systems compared with traditional distributed database systems. With analyzing the representative blockchain systems including Bitcoin, Ethereum and Hyperledger Fabric, this paper focuses on the distributed data management techniques in existing blockchain systems, which covers query processing, smart contract, network communication, and data storage layers in the architecture of blockchain systems. This paper first discusses the main differences and similarities between a blockchain system and a traditional distributed database system. Just like a distributed database system, a blockchain system has features of distribution, transparency, autonomy and scalability on managing data, but it is

收稿日期:2019-03-22;在线发布日期:2019-10-31.本课题得到国家重点研发计划项目(2018YFB1003404)、国家自然科学基金(U1811261,61672142)、辽宁省科学技术基金(20180550321)资助.**于 戈**,博士,教授,中国计算机学会(CCF)会员,主要研究领域为分布式数据库、分布与并行计算、区块链.E-mail:yuge@mail.neu.edu.cn.**聂铁铮**(通信作者),博士,副教授,中国计算机学会(CCF)会员,主要研究方向为数据库、数据集成、区块链.E-mail:nietiezheng@mail.neu.edu.cn.**李晓华**,博士,讲师,中国计算机学会(CCF)会员,主要研究方向为信息安全、区块链.**张岩峰**,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为分布式数据处理、云计算.**申德荣**,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为分布式数据库、数据集成.**鲍玉斌**,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为数据仓库、OLAP.

also different from most of distributed database systems on topologic, data distribution, query processing, consistency and security mechanism. Then, this paper presents the classification of blockchain systems on different aspects of distributed deployment styles, node roles and topological structures. With the development of blockchain technology, blockchain systems are designed to adapt blockchain application environments. The models of public blockchain, consortium blockchain and private blockchain are proposed, and functions of blockchain are regrouped and deployed to make nodes play different roles in a system. Moreover, the multiple topologic structures of blockchain are proposed. Besides the chain structure of traditional blockchain, the DAG structures, such as Tangle and Lattice, are applied to improve the efficiency of blockchain systems. Thirdly, this paper analyzes the techniques of distributed data storage management, distributed query processing and optimization used in blockchain systems and discusses their advantages and disadvantages. Specifically, the data storage techniques of existing blockchain systems are deeply analyzed, including the data structures of storage, as well as the organization of data files and optimization techniques. Key-Value databases such as LevelDB are usually used in blockchain systems to improve the efficiency of accessing ledger data and state data. Currently, more research works focus on using different methods, including database, index and distributed storage, to optimize the storage of blockchain. This paper also analyzes various queries in the existing blockchain systems and classifies them into three types: account query, transaction query and contract query. The distributed query processing techniques used in blockchain systems are discussed. Fourthly, this paper points out the challenges and development trends of distributed data management techniques for blockchain systems, including distributed storage for blockchain data, efficient and secure consensus mechanism for blockchain transactions, high available query processing, distributed management of smart contracts, privacy protection for blockchain data, data audit and monitoring in blockchain system. Finally, this paper shows the distributed data management of blockchain systems in various domain-specific applications, such as finance, manufacture, network storage, credit and other fields.

Keywords　blockchain; distributed data management; data storage; query processing

1　引　言

在“互联网＋”应用日益普及的大环境下，大量应用需要将发生的事件、行为、状态持久地记录在分布式环境中以用于日后的查询，即进行分布式记账. 分布式记账已经逐渐成为互联网应用中的一项重要功能. 以电子商务交易系统为例，客户需要执行提交订单、通过电子银行向商家支付货款、从物流公司收货等操作，商家需要执行接受订单、通过物流公司发货、通过电子银行收款等操作，电子银行方需要执行从客户收款、向商家付款等操作，物流公司需要执行从商家收货、收取物流款、向客户发货等操作. 客户、商家、物流公司、电子银行共四方处于一个分布式环境中，在每一个环节都需要记录相关的操作和信息. 由于各方之间并不存在完全信任关系，最终以哪一方记录的账目为确认信息是一个重要的问题. 传统的方法采用由电子商务交易服务平台作为公正的第三方进行统一记账，所有的交易信息的查询操作全部在这个平台上进行处理，物流公司和银行的部分数据也以接入的方式添加至交易服务平台. 在这种传统集中式记账方式里，主要的交易信息存储在单一的记账方，这是一种“逻辑”上的集中式存储模式，即交易数据存储在唯一的某业务参与方并由其负责管理. 集中式记账方式存在的问题包括：(1) 记账方为了保证可靠性需要存储数据的多个副本，从而造成了数据存储的性能瓶颈；(2) 交易数据可能被记账方篡改且无法验证，因此各参与方需要完全信任记账方；(3) 记账方受到攻击后数据难以恢复. 因此，传统集中式记账方式存在着存储效率低、可信性差、易受攻击等弊端.

为了解决以上难题，采用分布式记账方式的比

特币系统(Bitcoin)[1]在2008年被首次提出,并受到广泛关注.随后,区块链技术作为比特币系统所采用的底层技术逐渐引起工业界与学术界的重视,比特币系统所具有的分布共享性、共识性、不可篡改性、可溯源性和最终一致性等特点均来源于区块链技术.在基于区块链技术的分布式记账方式中,所有参与方都可以保存一份相同的完全账本,新加入的参与方可以下载完全账本并验证账本的正确性.这种方式降低了传统集中式记账方式中记账方的多副本数据维护成本,同时参与方也可以通过访问本地数据提高访问效率.此外,在区块链系统中,交易的账目采用数字签名和加密算法处理,从而提高了系统中数据的安全性,而区块之间通过哈希值串联的数据关联方式和基于共识算法确认区块的数据写入机制也使得区块链上的数据极难被篡改.

起初,区块链技术所支撑的比特币系统仅是一个专用的交易系统,并不支持虚拟货币交易以外的其他功能,这严重限制了区块链技术在分布式数据管理上的应用.随着区块链技术的发展,产生了大量新型区块链系统.2014年由Buterin基于区块链技术推出了以太坊(Etherum)平台[2].以太坊提供了基于智能合约的编程功能,支持区块链应用的二次开发,这标志着区块链2.0时代的诞生.超级账本(Hyperledger Fabric)[3]则是基于IBM早期贡献出的Open Blockchain为主体搭建而成的Linux基金会的区块链项目,其主要目的是发展跨行业的商用区块链平台技术.在超级账本框架中,包括了Hyperledger Fabric①、Hyperledger Burrow②、Hyperledger Sawtooth③和Hyperledger Iroha等多个项目,构成了完整的生态环境.区块链3.0时代[4]则是将区块链技术的应用范围扩展到各类应用之中,服务领域除金融、经济之外,还包括政府、健康、科学、文化等领域.区块链技术将支持各类资产交易与登记的去中心化可信处理,并与物联网等技术融合.

未来,区块链技术将会与其他新兴技术相结合用于各类应用之中,诸如区块链+科学、区块链+医疗、区块链+教育、区块链+能源等应用将会迅速发展.目前,区块链技术已应用于多个领域之中.在数字货币服务领域,支持支付、兑换、汇款、交易功能;在金融服务领域,支持清算、结算、安全监管、反洗钱等功能;在B2C服务领域,支持无人管理的商亭等新业务;在P2P租赁管理领域,支持无需中介的货物交换、租赁等共享经济新业务;在供应链管理领域,支持物理资产签名、物流跟踪和交付等功能;在知识产权保护领域,用于建立不可篡改的权利和拥有权;在征信管理领域,支持身份认证、日志审计和监管等;在溯源管理领域,支持数据鉴别与存证、防伪溯源等功能.

区块链技术是一种建立在多种技术之上的分布式共享账本技术,而区块链本质上是一种多方参与共同维护的分布式数据库.相对于集中式数据库管理系统,区块链系统采用去中心化或者弱中心化的数据管理模式,没有中心节点,所有参与节点均可以存储数据,而事务的持久性则依靠参与节点共同维护的不断增长的数据链和非集中式的共识机制予以实现,保证了数据在基于验证基础上的可信性.此外,相比于传统的分布式数据库和分布式数据存储系统,区块链系统的参与节点可以获得完整的数据副本,而非部分数据的副本.区块链系统的特殊数据存储机制和一致性共识机制是其不同于传统分布式数据库系统的主要原因.

区块链的数据存储结构和数据组织方式不同于其他数据存储系统.区块链将数据记录组织成区块(Block),并在每个区块的区块头中通过记录前一区块的哈希值将区块组织成链式结构.这种结构使区块链的数据存储具有不易篡改性、可溯源性和可验证性.然而,区块链的存储结构和基于密码学算法的共识机制也为数据管理带来了交易确认效率低和查询不便等诸多弊端.例如在记录交易的吞吐量方面,使用区块链技术的比特币系统仅支持每秒处理7笔交易数,并且还需要经过1小时以上时间才可以确认写到区块(相关研究表明43%的比特币交易未能在一小时内得到处理④).此外,区块链的数据记录按时间顺序存储在区块中,这为交易数据的查询处理带来了挑战,当前很多数字货币系统的查询处理都要依赖于某种键值数据库系统.

其次,区块链的共识机制也不同于分布式数据库系统.区块链系统为了在P2P网络环境下保证交易操作符合事务特性,需要维护数据一致性,并避免"双重支付"(Double Spends)的发生,这是区块链共

① Hyperledger Fabric. https://www.hyperledger.org/projects/fabric

② Hyperledger Burrow. https://www.hyperledger.org/projects/hyper ledger-burrow

③ Hyperledger Sawtooth. https://www.hyperledger.org/projects/sawtooth

④ Study: 43% of Bitcoin Transactions Aren't Processed after First Hour. 2017. https://www.ccn.com/43-bitcoin-transactions-not-processed-one-hour-study-says

识机制的主要考虑的问题. 同时，由于区块链网络本身是一个去中心化的网络，参与节点完全自治，并没有统一的节点负责管理和维护，为此区块链节点之间需要使用 P2P 技术实现数据广播以更新节点的状态信息和账本信息.

区块链系统公认的基础架构模型[5]主要分为 6 层，本文在其基础上增加了查询层，以便对区块链系统的查询处理机制进行分析. 这样，区块链系统架构扩展为 7 层，如图 1 所示，主要包括：(1) 应用层. 基于区块链的各类应用，如数字货币、区块链金融、区块链征信等；(2) 查询层. 实现对交易账本数据的访问和验证，以及对账号状态的查询；(3) 合约层. 由脚本、算法机制和智能合约所构成的可编程基础框架；(4) 激励层. 负责为奖励记帐工作而进行货币发行、交易费用分配任务；(5) 共识层. 封装网络节点的 PoW、PoS、DPoS 和 PBFT 等各类共识算法，实现分布式共识机制；(6) 网络层. 封装 P2P 组网机制，数据传播机制和数据验证机制；(7) 数据层. 封装底层数据区块的数据结构和加密机制. 当前的区块链系统大多基于该系统架构进行实现，其中数据层、网络层、共识层和查询层是区块链系统的必要元素.

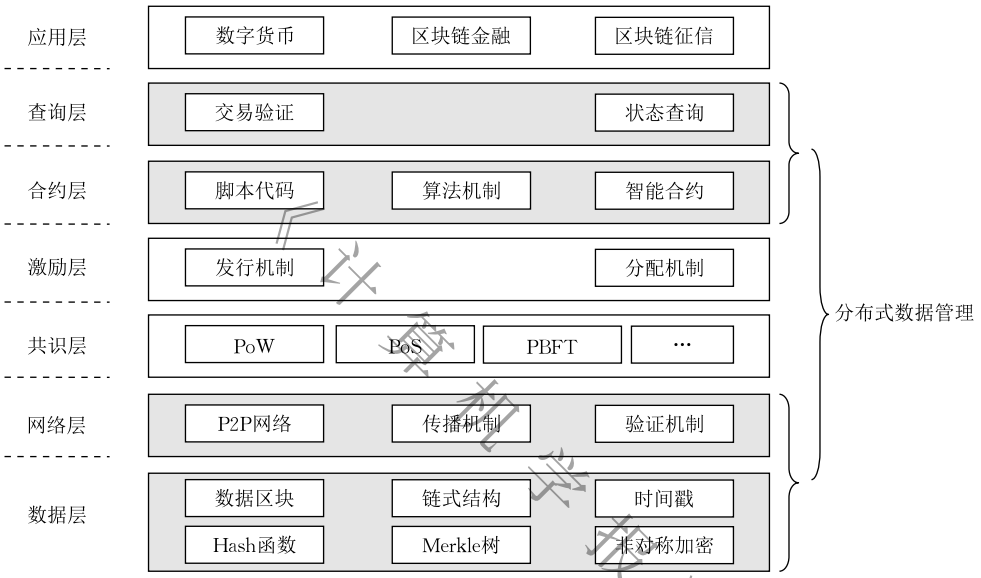


图 1　区块链系统的架构模型

现有相关工作针对区块链系统不同层次的技术和区块链在各领域上的应用进行了大量研究与综述. 对于区块链系统所包含的关键技术和研究现状，以及未来的发展趋势，袁勇等人[5]在区块链的基础架构模型方面对比特币的原理和技术进行了系统的阐述，何蒲等人[6]结合比特币系统介绍了区块链的概念和技术，并对前景进行了展望，邵奇峰等人[7]对比特币、以太坊和超级账本等多个区块链平台进行分析，总结了区块链的优势、劣势和发展趋势. 在应用层方面，文献[8]对区块链在数字货币上的应用进行了全面的综述，刘敖迪等人[9]介绍了区块链技术在信息安全领域的研究现状和进展. 由于区块链具有健壮的数据存储能力，因此相关研究工作在数据存储系统上进行区块链技术的应用[10-11]. 对于合约层，贺海武等人[12]结合多个领域应用场景对智能合约技术的概念、关键技术和面临的问题进行了阐述. 此外，对于共识层、网络层和数据层，已有研究分别对区块链系统的共识机制[13-14]、安全机制[15-16]、网络协议[17]、可信数据管理[18]和查询处理[19]进行了整理和综述.

区块链在设计之初就是以进行防篡改的数据存储和管理为目的，分布式数据管理是区块链系统的主要功能之一. 区块链技术中涉及分布式数据管理的部分主要集中在区块链架构的查询层、合约层、网络层和数据层，其中查询层和合约层在区块链系统中负责实现对数据的处理操作，如图 1 所示. 本文主要以分布式数据管理为视角，基于对当前主流的区块链系统分析，对比不同区块链系统在数据管理上的差异，对其中分布式数据管理所涉及的数据存储技术、查询处理机制和算法进行阐述和分析，并对区块链研究中涉及分布式数据管理的挑战进行探讨，对各领域的应用进行展望.

本文第 2 节对区块链系统的分布式数据管理机制进行分析，对比区块链系统和传统分布式数据管

理系统的异同;第 3 节介绍区块链系统的分类;第 4 节介绍区块链系统中的数据存储技术,包括物理存储结构,对比不同区块链系统在物理存储机制上的差异,以及区块链系统所采用的数据存储优化技术;第 5 节介绍区块链系统的数据查询处理技术;第 6 节探讨区块链系统在分布式数据管理方面所面临的研究挑战和发展方向;第 7 节展望区块链所支持领域应用的场景和待解决的问题;第 8 节总结全文.

2 区块链系统的分布式数据管理

区块链系统作为一种分布式数据库管理系统,主要以解决数字货币的货币转移、兑换和支付功能而被提出.区块链的特征主要体现在数据的公开透明、不可篡改和网络结构的去中心化等几个方面.由于区块链主要面向的是不可信数据存储环境下的记账应用,因此在数据存储上采用了去中心化、全副本的分布式方式,即所有参与方均通过 P2P 网络结构连接,并可以存储完整的共享账本.由此可见,区块链系统在管理交易记账上虽然使用了分布式数据管理方式,但与传统的集中式数据管理和分布式数据库系统管理数据的方式均有所差别.本节主要将区块链系统与传统数据管理方式进行对比和分析,并阐述彼此间的共同点和差异性.

2.1 区块链与传统分布式数据库的共同点

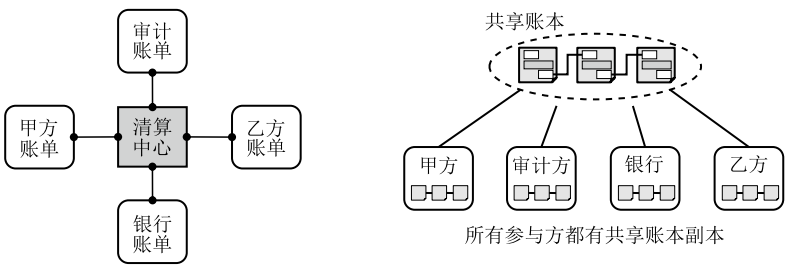
区块链技术主要是针对现有金融机构的集中式记账系统的信任问题而被提出的,其本身是由分布式存储、P2P 网络、加密算法、共识机制等多种技术所构成的.中本聪基于区块链技术设计并发行了数字货币"比特币",用以解决美国次贷危机中所展现的金融机构信任问题.相比于金融机构的集中式记账系统,基于区块链技术的交易记账系统具有公开透明、去中心化、可溯源查询和不可篡改等诸多的优势,从而避免了集中式记账方式中账本的真实性高度依赖于对记账方信任的弊端.这里以电子商务的交易记账应用为例,对传统基于清算中心的集中式记账方式和基于区块链的分布式记账方式的记账业务流程进行对比.传统集中式记账方式如图 2(a)所示,交易相关的账目数据集中存储在清算中心的数据库中,交易的参与各方如果需要调用完整的交易信息需要访问清算中心,其弊端主要体现在完全依赖于对清算中心记账方的信任,一旦记账方失信或遭受攻击,其保存的数据也随之失去可信性.区块链的分布式记账方式如图 2(b)所示,其中账本数据是整体共享的,以区块为单位通过密码学算法链接在一起,且网络中任何一个参与方均可以存储完整的共享账本副本,而数据的安全性则也是基于密码学算法予以保证.由于所有参与方均保存有共识后的共享账本,因此任何一个参与方进行双重支付或篡改账本数据的难度变得极大,从而保证账本数据在不可信环境中的可信性.区块链系统的分布式记账方式使其在数据存储管理的方式上与分布式数据库相同,即存储结构化的数据集合,这些数据逻辑上属于同一系统,物理上分布在计算机网络的各个不同场地上[17].区块链系统同样具有分布式数据库所具有的诸多特性:

(1)分布性

区块链系统与分布式数据库系统在数据的存储方面都是物理上分散、逻辑上统一的系统.区块链系统中具有全局统一的数据模式,数据以副本形式存储在参与节点中,每个参与节点存储的是数据模式相同且数据一致的共享账本.

(2)透明性

区块链系统在数据访问上具有透明性,用户看到的共享账本是全局数据模型的描述,就如同使用集中式数据库一样,在记录交易数据时也不需要考虑共享账本的存储场地和操作的执行场地.在数据复制方面,区块链系统的共享账本存储在各个参与节点上,并通过共识机制自动维护数据的一致性.



(a) 传统集中式记账方式 (b) 区块链的分布式记账方式

图 2 记账业务流程对比

(3) 自治性

区块链系统的参与节点具有高度的自治性. 在通信方面,参与节点可以独立地决定如何与其他参与者进行通信;在查询方面,参与节点本地就保存了完整的共享账本,可以在本地执行对账本数据的访问.

(4) 可伸缩性

区块链系统支持参与节点规模的任意扩展. 区块链系统允许参与节点在任意时刻加入和退出系统. 而且,由于区块链的参与节点保存的是完整共享账本,因此对于参与节点重新加入区块链系统后,仅需要从其他节点更新缺失的区块数据即可完成数据的重新分布,不会影响整体的系统性能.

2.2　区块链与传统分布式数据库的差异

区块链系统原始的设计目的之一是解决非信任环境下数据的可信性问题. 所谓的非信任环境是指负责数据存储的节点可能随意篡改数据而其他参与节点又无法识别,这将造成参与节点之间的互不信任问题. 对于传统分布式数据库管理系统而言,系统建立在信任环境,其中参与节点采用统一管理的方式,节点之间具备完全相互信任的关系. 因此区块链与传统的分布式数据库在数据管理方式上又具有显著的差异,如图 3 所示,具体体现在以下几个方面:

(1) 去中心化拓扑结构

在参与节点的网络拓扑结构方面,区块链系统的去中心化结构采用了基于 P2P 的分布式模式,这种结构与基于 P2P 网络结构[20]的数据库系统(P2PDBS)[21-22]相似. 如图 3(b)所示,区块链节点通过通信控制器(CM)仅基于邻居地址进行通信,其加入和退出都是随意和动态的. 传统分布式数据库虽然数据分布在不同的场地,但是通常采用中心化的主从结构,由全局的网络管理层存储各个局部数据库节点的地址和局部数据的模式信息,以用于查询处理时进行全局优化和调度,如图 3(a)所示.

(2) 数据分布方式

分布式数据管理的数据存储方式,通常分为两类[23]:① 分割式. 数据被划分成若干个不相交的分片,分别保存在不同的节点上,数据的划分方法分为水平分片和垂直分片;② 复制式. 同一个数据分片保存在一个以上的节点上,复制方式分为部分复制和全复制. 分割式能够节省数据的存储空间,查询时需要在节点间传输数据,虽然使用半连接等算法可进行优化,但效率依然较低. 复制式通过多节点的数据冗余存储可提高查询效率,但耗费存储空间且需要维护数据一致性. 区块链系统的数据分布采用的是全复制式,即每个参与节点都在本地复制了具有全局模式的全部数据. 因此,数据在区块链系统中是全局共享的,如图 3(b)所示. 相比于区块链系统,传统分布式数据库的分布方式主要基于在全局模式创建局部模式,再对数据进行垂直分片和水平分片,如图 3(a)所示,每个节点存储的是全局数据分片的副本,再通过数据分片的元信息管理实现全局数据的访问和查询处理. 当前很多基于分布式数据库技术的大数据存储系统,如 HBase① 等,均采用集中式的元信息管理节点管理数据副本的分布信息.

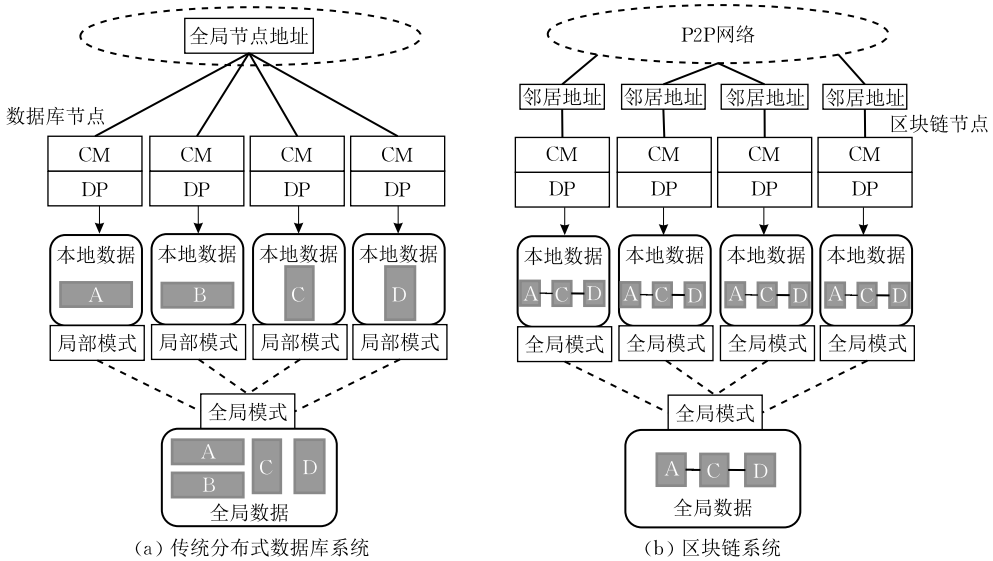


图 3　区块链系统与传统分布式数据库系统对比

(3) 数据查询处理

区块链系统中对账本信息的查询处理通常在存

① Apache HBase Reference Guide. http://hbase.apache.org/book.html

储了完整共享数据的参与节点本地执行.由于区块数据采用基于文件的存储方式且本身缺少索引结构,因此在区块链上直接执行对账本查询只能使用顺序扫描的方式访问所有区块数据.目前区块链系统常用的查询优化方式是将账本记录存储在Key-Value数据库中,以提高数据的访问效率.当前,比特币和以太坊等系统都使用了LevelDB① 存储和检索数据.需要说明的是,在以太坊这类支持智能合约的第二代区块链系统中,智能合约代码的执行处理是嵌入在区块链记账功能中的.因此,对智能合约代码的调用是在所有参与进行共识验证的节点上执行.传统分布式数据库的查询处理主要基于数据副本的大小和分布场地进行优化[24],而在面向大数据的分布式数据库上则采用基于并行计算思想的查询优化方法[25].

(4)数据一致性维护

数据一致性是保证数据正确性和可信性的关键,区块链系统采用共识机制来保证各节点上数据的一致性.在数字货币的应用中通常采用工作量证明机制(PoW)通过算力竞争保证分布式的一致性[26],如解决基于SHA256、Ethash② 等算法的数学难题,而从节约能耗的角度,则会采用权益证明机制(Proof-of-Stake,PoS)和授权权益证明机制(Delegated Proof-of-Stake,DPoS)等③方法.其中,使用工作量证明机制进行一致性维护的最大问题在于共识的效率过低,一个区块的一致性需要在其后生成一定长度的后续区块之后才能够被确认.分布式数据库系统通常采用包括实用拜占庭容错PBFT[27]、Paxos[28]、Gossip[29]、RAFT[30]等高效的算法维护数据的一致性,而这些算法也被一些面向联盟链应用的区块链系统所采用.

(5)数据安全性机制

区块链系统在安全性方面主要为用户提供了数据篡改验证、数据溯源和加密安全机制.数据的篡改可以通过校验前后区块的哈希值进行验证,因此要篡改数据并被所有参与者认可就需要在算力上付出高昂代价以重新生成区块,其难度相比传统的集中式和分布式数据库都要大很多.但是在数据的可访问性上,由于区块链的共享性,所有用户均可访问完整数据,而传统数据库管理系统则基于用户身份验证方式控制数据的访问.为了解决共享数据上的隐私安全性问题,区块链采用了基于非对称加密的交易方式实现匿名交易,其优点是很好地保护了用户隐私,缺点是一旦密钥丢失,用户的账号信息将无法恢复.

综上所述,区块链系统相比传统分布式数据库系统,在记账方式上提供了更好的分布性、透明性和可信性,在功能上提供了防篡改验证机制和智能合约机制,因此更加适合在非可信环境下的匿名使用.另一方面,相比传统的分布式数据库系统,区块链系统在网络结构、数据存储和访问方式上也具有显著的差异.

3 区块链系统的分类

3.1 区块链系统部署方式的分类

区块链系统根据其分布式部署方式和开放对象被划分为三种:"公有链"(Public Blockchain)、"联盟链"(Consortium Blockchain)和"私有链"(Private Blockchain).三类区块链系统的对比如表1所示.

表1 各区块链系统类型对比

	公有链	联盟链	私有链
网络结构	完全去中心化	部分去中心化	(多)可信中心
节点规模	无控制	可控	有限
加入机制	随时可以参加	特定群体或有限第三方	机构内部节点
记账方	任意参与节点	预选节点	机构内部节点
数据读取	任意读取	受限读取	受限读取
共识机制	容错性高、交易效率低(PoW或PoS等)	容错性和交易效率适中(PBFT,RAFT)	容错性低、交易效率高(Paxos,RAFT)
激励机制	有代币激励	无代币激励	无代币激励
代码开放	完全开源	部分开源或定向开源	不开源

(1)公有链

公有链是对所有人开放的,任何互联网用户都能够随时加入并任意读取数据,能够发送交易和参与区块的共识过程.比特币和以太坊等虚拟货币系统就是典型的公有链系统.公有链是完全去中心化的结构,其共识机制主要采用PoW、PoS或DPoS等方式,将经济奖励和加密算法验证相结合,以保证经济奖励和共识过程贡献成正比.此外,公有链中程序开发者对系统的代码是完全开源的,而且开发者无权干涉用户.

在分布式数据管理方面,公有链系统的优势和缺陷主要包括以下几个方面:

① LevelDB. http://leveldb.org/
② Ray J. Ethash. https://github.com/ethereum/wiki/wiki/Ethash
③ Bitshares. Delegated Proof of Stake. http://docs.bitshares.org/bitshares/dpos.html

优势：

① 数据透明性高. 公有链是任何人都可以未经许可加入的，同时公有链的数据也是开放给所有参与者的. 这使得在公有链上存储的数据具有很高的透明性，加入系统的参与者都可以下载和读取区块链数据，并对其中的账本记录或智能合约内容进行验证，从而无需依靠可信的第三方.

② 存储容错性高. 在公有链中有大量的参与节点存储着完整的区块链数据用以执行区块的共识和交易验证，相当于在分布式存储系统的每个节点上都保存了数据的完整副本. 这确保了系统在数据存储上的高容错性，即数据存储在公有链系统中几乎不会丢失，更无法被轻易篡改.

③ 系统创建成本低. 与传统分布式数据库等存储系统不同，公有链系统无需设计基础架构系统，而是运行在分散式应用程序之上. 因此，公有链创建者无需维护大量的存储服务器以及聘用系统管理员，从而降低了创建系统的成本，系统运行的成本由参与者负担.

缺陷：

① 数据处理速度低. 区块链交易记录写入区块等同于数据库中的事务提交. 数据写入区块链并在节点间达成一致性确认的效率是由区块链的共识机制所决定的. 公有链由于参与节点数量庞大，且采用各种证明机制(PoW 或 PoS 等)实现数据的一致性共识，这导致了共识达成的时间效率低下，也造成了在数据存储上的低效问题.

② 数据存储吞吐量低. 公有链的协议中限制了区块的大小，这使得每个区块所能存储的交易记录的数量受到限制. 由于公有链中区块的生成和确认时间较长，从而导致区块链作为存储系统具有较低的吞吐量.

③ 易产生硬分叉. 当新版本软件定义了新协议规则且与旧版本不兼容时，运行新旧不同协议软件的节点将在原有区块链基础上产生两条基于不同规则的区块链硬分叉. 公有链系统中的参与者具有高度的自治性，因此一旦有新协议的软件出现，就极易产生基于新旧协议的硬分叉出现，从而分裂成两个永远不会合并的区块链系统. 比特币和以太坊系统都发生过硬分叉.

④ 运行依赖代币. 公有链系统为了激励节点能够参与共识过程，需要采用代币奖励的激励机制. 而参与者在公有链上发送交易或智能合约等操作都需要依赖于其持有的代币才能够执行，这将导致参与者交易成本的增加.

(2) 联盟链

联盟链是仅对特定的组织团体开放的区块链系统，主要特点是其共识过程受到预选节点控制. 联盟链的数据可能允许所有用户可读，或者只允许受限于参与者读取. 在结构上联盟链采用“部分去中心化”的方式，将节点运行在组成联盟共同体的有限数量的机构中. 联盟链的加入机制相对更加严格，节点间具有一定的信任性，不需要激励机制，因此共识机制不需要工作量证明等资源消耗较大的方法，更多的是采用容错性和性能效率适中的算法，如实用拜占庭容错算法(PBFT)和 DPoS 等. 联盟链常用于银行、保险、证券、商业协会、集团企业及上下游企业. 典型的联盟链系统有超级账本[3]和 R3 的 Corda 系统①. 联盟链系统通常采用部分开源和定向开发的方式.

在分布式数据管理方面，联盟链系统的优势和缺陷主要包括以下几个方面：

优势：

① 数据写入吞吐量高. 联盟链的共识机制会采用更加节能和高效的拜占庭容错类算法，这使得系统数据处理的效率显著提升：数据写入系统的吞吐量大幅提升，写入区块的响应时间也明显缩短. 超级账本系统的每秒钟处理事务个数(TPS)可以达到 300 至 500，远高于比特币这类公有链系统.

② 提供具有隐私保护的数据共享. 联盟链通常创建于明确的组织机构之间，参与成员之间既要共享数据又要保护隐私. 因此联盟链系统在数据管理上会采用基于通道方式的数据隔离机制，并结合高强度的加密处理和零知识证明等隐私保护方法对数据进行保护.

③ 支持共识协议扩展. 联盟链系统需要在系统规模和稳定性等方面进行动态平衡，为此许多联盟链系统在设计上支持共识模块可插拔，允许共识机制的切换操作，从而支持根据节点数量、网络平衡情况、吞吐量等指标进行共识机制的调整. 超级账本、FISCO BCOS② 和 Quorum③ 等都采用了此种设计.

缺陷：

① 系统规模扩展性差. 由于联盟链系统会采用高效的共识协议以提升系统数据处理效率，同时也导致了系统规模扩展性的限制. 如采用 BFT 类共识

① Corda. http://www.corda.net/
② FISCO BCOS. http://www.fisco-bcos.org/
③ Quorum. https://www.goquorum.com/

协议的系统在节点数量超过一定水平时，由于需要在节点间传输大量的消息，会造成系统吞吐量的显著下降.

② 部署和运维代价高. 面向联盟链应用设计的区块链系统虽然具有较高的技术成熟度，但由于这类系统不像公有链开放性高，因此相关的第三方支持工具较少，联盟成员如果要实现特殊的数据管理功能需要自行开发工具. 这将增加系统部署和运维的成本.

(3) 私有链

私有链是指区块记账权限由某个组织或机构控制的区块链，其读取权限不对外开放或进行某种程度的限制. 私有链通常采用具有可信中心的部分去中心化结构. 私有链由于不需要复杂的共识机制，通常采用容错性低、性能效率高、不需要代币的 Paxos 和 RAFT 等算法，因此其记账效率要远高于公有链和联盟链系统. 私有链相比于传统的数据库系统，能够提供更好的隐私保护、更低的交易成本，且不易被恶意攻击.

在分布式数据管理方面，私有链系统的优势和缺陷主要包括以下几个方面：

优势：

① 数据处理速度和读写吞吐量高. 私有链中节点数量少且具有更高的互相信任，共识协议则采用 Paxos 类高效的算法，不需要每个节点参与认证，因此数据的写入在区块链中能够迅速被确认，同时数据处理的吞吐量也随之显著提高.

② 组织的数据隐私保护性更好. 私有链中的数据内容并不对组织以外的用户开放，相比于公有链其数据隐私能够得到更好的保护.

③ 交易成本大幅降低. 私有链不需要使用代币作为奖金激励，也不需要运行成本高昂的 PoW 等共识算法，因此也不需要为记录和存储数据而收取费用.

缺陷：

① 依赖可信中心. 私有链中会生成一个可信的中心节点负责处理共识机制，这与区块链的去中心化思想有所出入，一旦这个中心节点固定或过于中心化，那么就与传统中心化的分布式数据库系统没有区别了. 因此私有链没有完全解决信任问题，通常用于改善可审计性.

② 数据容错性低. 私有链使用中心化的数据库系统，同时参与节点较少，因此相比于公有链和联盟链，私有链的数据容错性较低.

不同类型的区块链系统在分布式结构上也具有一定的差异. 由于各自具有不同的优势和缺陷，因此相应的应用场景也不同. 公有链是完全的去中心化结构，参与节点尽量具有平等的权利，通常用于搭建开放式的共享记账系统. 联盟链是部分去中心化的分布式结构，由参与联盟的多个机构形成多中心的分布式系统，通常用于行业机构间构成权利相对对等的组织团体以共享数据. 私有链在公司或机构内部形成小范围的可信中心化结构，省略激励层以提高性能效率，用于企业和机构内的数据共享管理.

3.2 区块链系统节点角色的分类

区块链系统中，根据在系统中具有的不同功能，节点对应着不同的角色，不同类型的区块链系统所设定的角色也有所不同. 有些节点能够下载完整的区块链数据副本，有些节点只能查看区块的部分数据，有些节点负责生成区块记账. 随着当前区块链系统的复杂性不断增加，节点分工更加明确，节点角色也逐渐多样化且对应多种不同的权限和功能.

3.2.1 区块链系统节点功能分类

在区块链系统中，节点所具有的系统功能包括以下几种：

(1) 记账功能

区块链系统的记账功能是节点通过计算新区块的哈希值获得记账权利，并将新区块加入区块链中. 在多数的公有链区块链系统中，区块的记账权是通过工作量证明机制进行确认的，即由最先发现生成区块哈希值随机数的节点获得记账权和代币激励. 在记账过程中，节点会将未打包的交易进行排序生成区块，而在以代币激励为基础的公有链中，节点会将记账奖励高的交易优先进行打包.

(2) 数据存储功能

区块链系统的数据存储功能是对系统中的区块数据、区块链元数据和状态余额等进行存储并验证有效性. 不同的区块链系统的区块结构、元数据和状态余额的存储方式是不同的. 如比特币的状态余额采用的是未交易消费输出(Unspent Transaction Outputs，UTXO)方法，而以太坊的状态余额采用的是账户余额模式.

(3) 交易提交功能

交易提交功能是由用户向区块链系统提交交易记录，包含交易输入和交易输出. 交易提交后由记账功能节点将其加入到区块中. 在以太坊中，用户可以将智能合约作为一个账户并与其进行交易.

(4)路由功能

路由功能用于节点在 P2P 网络中发现邻居节点并进行通信和同步区块.比特币系统采用基于 TCP 与 DNS 种子通信并逐渐扩大邻居节点列表的路由方式,而以太坊系统则采用基于 UDP 的 Kademlia[31] 通信协议进行路由.

(5)钱包功能

钱包功能用于生成并保存公有链账号对应的私钥、公钥和规范地址.目前在公有链系统中对钱包进行管理的方法主要包括基于密码的 Keystore 方式和助记词方式.

(6)成员管理

在联盟链中,成员管理用于提供成员注册和对成员身份证书进行管理.如超级账本中的 MSP(Membership Service Provider)会建立一套根信任证书体系对成员身份进行认证和验证用户签名.

3.2.2　区块链系统节点角色分类

区块链系统中,并不是每个节点都需要包含所有功能,根据所具有的不同功能组合节点对应着不同的角色,而角色本身也是一个逻辑概念,同一个物理节点可以运行不同的区块链系统角色.本文将区块链系统的节点角色分为系统内角色和系统外角色两类.

系统内角色主要存在于区块链网络之内,负责维护区块链系统的运行,包括以下几种类型:

(1)全节点

全节点主要具有数据存储功能和路由功能,其任务是验证记账节点生成区块的有效性,并对交易进行确认,同时也帮助其他节点进行路由.全节点的特点是存储区块链的完整数据副本,因此全节点上可以提供对交易记录的查询服务.

(2)记账节点

记账节点通常包括记账功能、数据存储功能和路由功能,主要任务是对交易进行排序打包生成区块,并将区块加入到区块链中.在公有链系统中记账节点需要通过解决 PoW 等难题挖掘区块,以获得相应的代币奖励,此类记账节点具有钱包功能.而在基于其他共识机制的区块链系统中则不需要基于算力生成区块,仅对区块进行打包和维护.

(3)简单支付验证节点

简单支付验证(Simplified Payment Verification,SPV)节点的任务主要对交易进行验证,但不存储区块数据.SPV 节点主要存在于比特币系统中,虽然支持数据存储功能但并不存储区块链的完整数据,而是存储全部区块的区块头数据.验证交易时,SPV 节点通过 Bloom Filter[32] 等方法过滤掉不可能存在交易的区块,再从全节点下载可能存在交易的区块进行验证.这种方式降低了节点验证交易的相应时间和负荷.

(4)排序服务节点

排序节点的任务是对接收到的交易进行排序并打包生成区块.在基于代币激励的公有链系统中,排序服务功能将基于每笔交易所提供的交易费用进行排序,交易费用高的交易将被优先打包到区块之中从而先于其它交易记录到区块中,这样能够有效地保证记账节点的收益.对于超级账本这类联盟链系统,排序服务的目的是保证区块链上的节点接收到相同的信息并且交易具有相同的逻辑顺序记入区块链.

区块链系统的外部角色是存在于区块链网络之外的,与区块链网络的内部节点进行通信实现相关功能,具体包括:

(1)交易客户端

交易客户端具有交易提交功能和钱包功能,主要用于用户在区块链系统上提交交易操作,如进行账户间转账、提交智能合约等.交易客户端可以是用户使用的应用程序,运行于区块链系统之外,也可以与 SPV 节点机制结合实现轻量化的数据货币钱包.

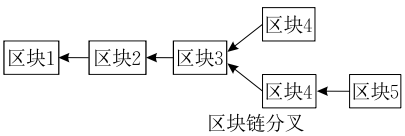
(2)纯矿工节点

纯矿工节点仅负责解决 PoW 难题的计算任务,不具有区块链的其他任何功能,主要存在于使用工作量证明机制的区块链系统之中,与被称为“矿池”(Mining Pool)的记账节点相连.由于在比特币等区块链系统中,全网络的算力已达到极高的数值(比特币全网算力峰值接近 80EH/s),因此独立的记账节点凭借自身的算力只能获得极低的记账代币奖励概率.为了稳定记账节点收益,降低收益方差,便产生了“矿池”.“矿池”使得每个纯矿工节点仅贡献自己的算力而无需存储数据,“矿池”将分散的算力聚合,在获得收益后再按提供算力比例将收益分发到矿工节点对应的钱包地址.

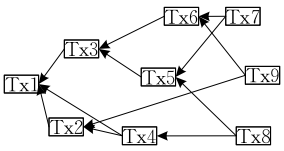
在以上区块链系统角色中,对数据进行分布式存储和管理的任务主要是由记账节点和全节点负责,其中全节点还负责执行对数据的查询服务.相对的,简单支付验证节点因为只存储区块头数据,因此在执行查询服务时只能实现过滤功能,并无法提供交易信息.因此,也可以认为区块链系统根据节点在系统内对应的角色对数据进行了划分.

3.3 区块链结构拓扑的分类

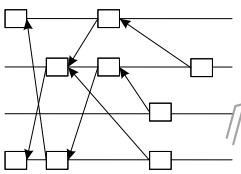
区块链技术的发展使得区块组织的拓扑结构上衍生出了多种类型.区块的拓扑结构不仅局限于全序的链式(Chain)结构,还有基于有向无环图结构的DAG区块链.不同区块链拓扑结构如图4所示.



(a) 链式结构



(b) Tangle结构



(c) Lattice结构

图4 区块链的拓扑结构

(1)链式结构区块链

链式结构是当前区块链系统所主要使用的数据结构.如图4(a)所示,在链式结构的区块链中,交易集被打包为区块,系统的每次提交都是以区块为单位且只能添加一个区块,每个区块包含了用于验证其有效链接于上一个区块的哈希值作为数字凭证.链式结构中区块之间具有全序关系,能够很容易进行数据验证和数据溯源.链式结构在实际运行中可能会出现区块链分叉的情况,具体分为硬分叉和软分叉两种情况.其中,在硬分叉时区块链系统将分裂为两个独立的区块链系统.比特币、以太坊和超级账本的区块链所采用的都是链式结构.

(2)有向无环图结构区块链

采用有向无环图(DAG)结构的DAG区块链系统是将原有的链式结构替换为有向无环图结构,同时组成区块链的单元也不再是由交易集合所构成的区块,而是使用更细粒度的交易作为基本单元.DAG区块链中一笔交易接着另外一笔交易,这意味着每笔交易都能够为下一笔后续交易提供证明,这样所有交易就构建成了一个有向无环图.DAG区块链系统为了提高交易的确认效率,交易之间存在着大量的偏序关系,这样可以实现交易的写入和确认的异步执行,并可以并行执行验证过程.DAG区块链的拓扑结构可以分为Tangle结构和Lattice结构两种,如图4(b)和图4(c)所示.其中,Tangle结构的区块链系统具有大量的偏序关系,这使其交易确认效率受到一定的影响,而Lattice结构消减了大量偏序关系,交易验证效率有所提高但是存在着被篡改的风险,安全性问题也比较多.

整体上DAG区块链比链式区块链具有更高的交易处理效率,并且降低了图灵完备智能合约在世界状态维护上的代价,但安全性问题也十分突出.当前,基于Tangle结构账本的DAG区块链有IOTA和Byteball[33],采用Lattice结构账本的Vite项目[34]和InterValue项目[35].

4 区块链系统的数据存储技术

区块链系统的数据存储结构包括数据在区块链系统中逻辑组织的数据结构和数据在外存储器中物理存储结构.区块链系统根据所支持的交易数据形式对区块的数据结构进行设计,并将区块链中所包含的不同数据部分以不同的文件组织方式存储在外存储设备上.本节主要对不同区块链系统的数据存储结构和数据组织进行对比,同时介绍现有区块链系统上存储优化技术的相关研究.

4.1 区块链系统的主要数据结构

区块链本身是一串通过密码学算法生成的前后连接的数据区块,每个区块包含区块头和区块体两个部分[1].区块头部分用于构成区块之间的链接关系,主要的组成部分包括上个区块哈希值(Prev Hash)、时间戳、随机数(Nonce)和交易的根哈希(Root Hash),其中上个区块哈希、根哈希和随机数用于生成当前区块的哈希值,而交易的根哈希用于验证交易使得交易不可伪造.区块体部分则主要存储交易计数和交易详情,而交易数据的结构设计则决定了区块链系统所能够支持的功能.

当前,大量以发行数字货币为目的的区块链系统均基于比特币的区块链数据结构进行设计,即采用了一棵Merkle树[36]的结构生成交易的根哈希值.而支持智能合约的区块链系统则在比特币基础之上对数据结构进行了一些调整,其中最具有代表性的系统是以太坊.以太坊的区块链系统针对其交易数据中包含的三种对象设计了三棵Merkle Patircia树(MPT)[37],分别是状态树、交易树和收据树.这些数据结构能够使得以太坊的客户端支持一些简单的查询.以太坊的Merkle Patircia树是结合了Merkle树和Trie树[38](也称为Radix树)两种数

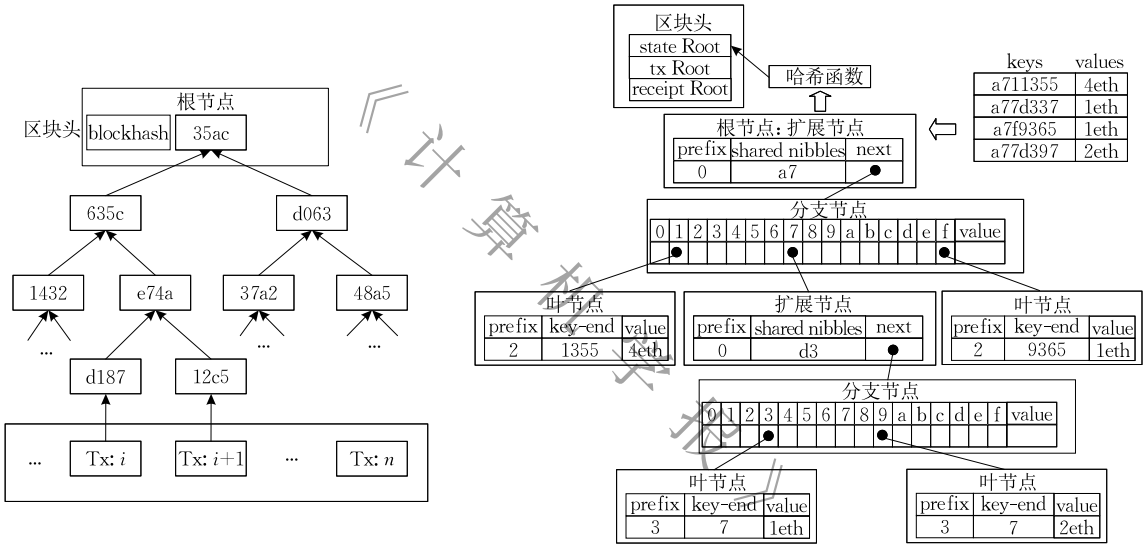
据结构的特点而设计的.

(1) Merkle 树

该树型数据结构可以是二叉树也可以是多叉树，在比特币中使用的是二叉树结构，如图 5(a)所示. Merkle 树采用自底向上的方式构建，在区块链中叶节点为基础交易数据，每个中间节点是它的子节点的哈希，根节点是最终的哈希值. 基于 Merkle 树结构可以使得每个数据集对应一个唯一的哈希根，如果要验证一个交易仅需存储从交易到根节点的分支即可.

(2) Merkle Patircia 树

该数据结构是以太坊对 Merkle 树和 Trie 树进行结合和改进后的数据结构，如图 5(b)所示. Trie 树的优点包括具有相同前缀的值在树中位置更加靠近、不会有哈希冲突等，但是自身存在着存储不平衡的问题. Merkle Patircia 树在结构上以 Trie 树的结构为基础，键值基于 Merkle 树的方式生成. Merkle Patircia 树的每个节点通过它的哈希值被引用，并在 LevelDB 中通过查询访问，其中 Key 为节点 RLP 编码(Recursive Length Prefix)的 SHA3 哈希值，Value 为节点的 RLP 编码. Merkle Patircia 树的结构中引入了多种节点类型，包括空节点、叶节点、扩展节点和分支节点. 其中，叶节点的结构是键值对的列表，Key 为特殊十六进制编码，Value 是 RLP 编码的数据内容；扩展节点也是键值对列表，Value 是其他节点的哈希值，用于在 LevelDB 中链接到其他节点；分支节点则是一个长度为 17 的列表，前 16 个元素对应 Key 编码中的十六进制字符.



(a) Merkle树　　(b) Merkle Patircia树

图 5　Merkle 树与 Merkle Patircia 树的结构

从图 5(a)中可以看到 Merkle 树的叶节点存储交易，通过哈希函数逐层生成上层节点的哈希值，如果底层交易记录发生篡改，则 Merkle 树根值也会变化，因此 Merkle 树能够有效检测底层交易记录的变化. 图 5(b)中是基于 Merkle Patircia 树的状态树(stateTree)，用于存储键值映射，其中键是地址，从根节点到叶节点的路径存储的就是 Key 值，值是账户的声明、余额和 nonce 等信息，存储于叶节点的 Value 中. 这里由于状态数据不同于历史交易记录，会有频繁的账户插入和余额的改变，而 Merkle Patircia 树更加适合数据更新，能够有效地发挥其所改变节点快速计算到根节点的特点，因此可以避免重新计算整棵树的哈希值. 在数据存储结构上，不同的区块链系统中包含的根哈希结构、根哈希数量和存储编码等均有所不同，如表 2 所示. 这也导致了各区块链系统不同的功能设定和数据访问性能.

表 2　典型区块链系统存储结构对比

	比特币	以太坊	超级账本
根哈希结构	Merkle 树	Merkle Patircia树	Merkle 树
根哈希数量	1	3	1
数据存储编码	Base58Check 编码	RLP 编码	Json 编码转 Protobuf 格式
数据存储系统	LevelDB	LevelDB	LevelDB/CouchDB
数据库数量	2	3	4
数据库存储内容	UTXO 数据区块的元数据	账户状态区块头和交易收据信息	状态数据库 索引数据库 历史数据库 账本数据库
数据索引	Bloom Filter	Bloom Filter	Key-Value
区块链数量	单链	单链	多链

4.2 区块数据的存储组织方式

现有区块链系统使用数据文件方式存储区块头,而区块数据及元数据主要使用基于键值模型的数据存储系统存储.区块链系统主要使用 LevelDB 这类键值数据库,通过 LMS-tree 结构[39]用以提高对交易的存储写入效率和查询访问效率.其中比特币和以太坊的数据存储于 LevelDB 数据库中,而超级账本 Fabric 的状态数据可以在转换成 Json 格式后选择存储在 CouchDB① 之中.其他基于区块链的存储系统,如 Storj、Filecoin、BigchainDB 等系统也均采用了 LevelDB 或 MongoDB② 等基于键值模型的数据库系统存储元数据信息[19].文献[19]较好地总结了现有区块链系统的数据库存储方案.在区块链系统中,不仅要存储由交易数据生成的区块头和区块体数据,还需要根据功能设计管理状态数据、索引数据和元信息等数据,因此在数据组织方式上也具有较大的差异,主要的区块链系统存储组织对比如表 2 所示.下面将介绍主要区块链系统的交易数据、索引数据和其他元信息的组织方式.

(1) 比特币

比特币的区块链系统将数据分为四个部分分别存储在 LevelDB 和文件系统中.其中,区块头和区块体数据以 blk*.dat 文件的形式存储,而包含交易花费 out 信息的区块"undo"数据以 rev*.dat 文件存储,用于区块链发生回滚时进行恢复,比特币的状态数据和区块元数据则采用 LevelDB 存储.状态数据中存储了所有当前未花费交易输出及相关元数据,这样可以不通过扫描全部区块数据就能够验证新加入的区块和交易.区块的元数据记录着区块在磁盘上存储的位置,也同样使用 LevelDB 进行物理存储以提高访问效率.在仅保存区块头的简单支付验证节点(SPV 节点)中,为了能够在本地进行验证而且节省存储空间和网络传输,还使用了 Bloom Filter 数据结构过滤不属于钱包的状态数据.

(2) 以太坊

在以太坊的区块链系统中,数据最终存储形式是基于 Key-Value 的键值对,并使用 LevelDB 作为底层数据库存储数据.在数据结构上,以太坊区块头的成员变量中包含了三个 Merkle Patircia 树的根哈希,分别对应状态树、交易树和收据树,此外还包含了 Bloom Filter 变量用于快速判断一个日志对象是否存在于区块日志集合中.而以太坊的区块体中则包含的是交易记录和 Uncles 成员.区块头和区块体的成员变量最终会转换成 RLP 编码的 Key-Value 键值对形式存储在 LevelDB 之中.

以太坊中共建立了三个 levelDB 数据库,分别是 BlockDB、StateDB 和 ExtrasDB.其中,BlockDB 存储的是区块头和交易记录,StateDB 存储的是账户的状态数据,ExtrasDB 则存储收据信息和其他辅助信息.以太坊中用于支持查询处理的交易树、状态树和收据树就分别存储在以上三个数据库中,每个 Merkle Patircia 树的存储内容和功能如表 3 所示,每个区块包含了整个状态树的根哈希,其中状态树需要经常进行更新.

表 3 以太坊的 Merkle Patircia 树

	状态树	收据树	交易树
Key	账号地址	交易编号	索引编号
Value	账户内容	交易内容	收据内容
存储数据库	StateDB	BlockDB	ExtrasDB
唯一性	区块链整体一棵	每个区块独立一棵	每个区块独立一棵
所支持的查询	指定账户的余额 账户是否存在 合约交易的输出	交易是否存在于区块中	地址的事件实例

(3) 超级账本

超级账本系统与比特币和以太坊的最大区别就是支持多链,其中每个链对应一套账本.其中每个账本包含的数据包括区块数据、区块索引、状态数据和历史数据.为此,超级账本系统中的每个节点会通过维护 4 个 LevelDB 数据库管理这些数据,其中区块数据以文件形式存储,如图 6(a)所示,具体包括:

① idStore.存储账本编号,用于用户快速查询节点中存储了哪些账本;

② stateDB.状态数据库,存储世界状态数据,默认使用 LevelDB,可以替换为 CouchDB;

③ historyDB.历史数据库,存储状态数据库中 Key 的版本变化;

④ blockIndex.索引数据库,存储区块的文件索引.

超级账本相比于比特币和以太坊具有更加明显的分布式数据库特征.在超级账本中通过排序服务实现了多通道机制,这与 Kafka 消息系统的 Topics 相同,能够实现通道间数据隔离,如图 6(b)所示.每个通道对应一条区块链账本,排序服务节点会根据交易中账本信息决定添加到哪个通道队列中,生成区块后在广播给加入这个通道的记账节点.每个记

① CouchDB. http://couchdb.apache.org/
② MonogoDB. https://www.mongodb.com/

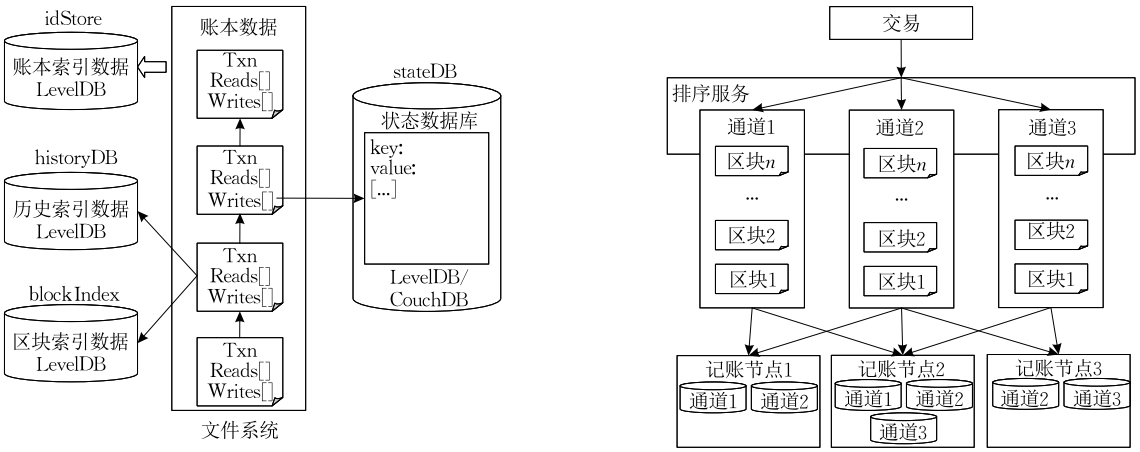


图 6　超级账本的分布式存储结构

账节点可以加入多个通道，一旦加入通道就可以接收通道的区块信息，而没加入的通道则记账节点不会收到这个通道的数据. 例如图 6(b)中记账节点 1 仅加入了通道 1 和通道 2，因此只能接收这两个通道所包含的数据. 但是，这里的数据隔离是对记账节点的，排序服务节点则可以接收所有通道数据. 根据数据的分布情况，超级账本区块链中分为系统账本和子账本，在所有记账节点上都存在的账本被称为系统账本，在部分记账节点存在的则成为子账本.

4.3　区块链系统的存储优化技术

区块链本身的数据结构和存储方式并不适合实时的交易记录查询或用户账号地址余额查询. 如果访问区块链数据需要对所有区块进行遍历，那么区块链系统的存储效率和执行效率显然十分低下. 为此，区块链系统都在存储结构和数据组织方式上采用了优化技术提高数据的访问效率和存储效率. 当前关于区块链系统的研究中所采用的存储优化方法主要包括三种：使用数据库系统管理区块数据、使用高效的索引结构提高数据访问效率和利用分布式存储策略减少节点存储负载.

(1) 数据库存储系统

利用高效的数据库系统提高数据存取性能是区块链系统普遍所采用的方法，这样不仅能够提高数据访问性能，也能够提高系统数据的存储能力. 从表 2 中可以看出 LevelDB 数据库被比特币和以太坊等区块链系统用以存储元数据和区块数据，甚至索引数据也存储在其中. BigchainDB 系统[40]为了提供更大规模的数据存储能力，直接以分布式数据库系统为基础增加区块链特性，在 BigchainDB2.0 版本中底层存储使用了 MongoDB 数据库从而支持高吞吐量和大容量. 而在数据库领域，则直接设计并实现了面向区块链系统的数据存储系统，如 ForkBase 存储系统[41]设计了高效的索引结构和数据模型，这使其能够与超级账本系统结合，UStore 系统[42]也支持为区块链系统提供底层存储.

(2) 高效的索引结构

数据索引是解决数据查询访问性能的关键技术，区块链系统也同样使用索引来提高交易记录的访问效率.

由于区块链中的交易数据具有不可改变的性质且以区块为单位进行存储，这使得 Bloom Filter 索引结构在各类区块链系统中广泛应用. 如前所述，比特币系统在简单支付验证节点上通过保存区块的 Bloom Filter 数据来过滤区块访问，以太坊也在区块头中增加了 Bloom Filter 结构以实现对日志的高效搜索. Bloom Filter 结构能够快速判断一个元素是否存在于一个集合之中，同时其所需的存储空间远小于元素集合的存储空间，因此在访问数据时可以实现对区块的快速过滤. 在超级账本系统中也创建索引数据来支持对区块的各种查询，其索引结构采用键值模型并使用 LevelDB 存储.

超级账本系统中创建了多个区块索引，包括区块编号、区块哈希、交易编号、同时区块编号和交易编号、区块交易编号、交易验证码等索引，所有索引均存储在 blockIndex 数据库中，不同类型索引的键值通过键值的命名规则进行区分，例如区块编号索引的键值是“n”＋blockNum，索引项的 Value 部分是一个文件位置指针，指针的内容包括文件编号、文件偏移量和区块占用空间，存储时序列化成字节保存在 LevelDB 数据库中.

(3) 分布式存储策略

区块链系统通过使用分布式数据库系统的存储

策略来提高存储空间的使用效率.在比特币和以太坊系统的网络中均存在“轻节点”,如前面提到的比特币简单支付验证节点,这些节点并不存储全部的区块数据,而是仅存储区块头和 Bloom Filter 索引数据,在进行交易验证时根据索引的决定需要访问哪些区块的区块体数据,再通过网络访问从全节点下载对应区块在本地进行验证.这样“轻节点”可以在存储少量数据的情况下对交易进行验证.

超级账本系统则是使用多通道的模式实现多链的数据存储,如图 5 所示,每个账本对应的区块链存储在一组记账节点之中,而每个节点只接收和存储其加入的通道数据.为了解决区块链存储效率问题,文献[43]提出了基于共识单元的分布式区块链存储策略,其目的是实现将多个分布式区块链存储节点合并为一个单元,并保证这个单元中至少包含一个完整的区块链副本.

4.4 区块链数据存储的保密机制

区块链系统的完全共享账本方式增加了交易透明性,但同时也带来了隐私泄露的风险.参与节点可以下载全部数据,并利用交易规律分析用户的身份和位置等特征信息.然而,区块链系统不仅要避免个人信息因交易发布而泄露,在一些应用场景还需要对交易内容提供隐私保护[16].对于数据存储层的隐私保护主要是在满足基本共识机制前提下尽可能隐藏数据语义和背后的知识.目前,已经出现了很多针对区块链数据的隐私保护方法,为区块链系统提供不同程度的保密机制.

(1)数据加密保护机制

加密保护机制是利用密码学算法对数据进行加密以实现只有相关方才能够查看数据.这是隐私保护的最常用方法,通常采用对称加密算法和非对称加密算法结合的方法.在区块链系统中使用加密机制对区块存储和交易传输过程进行加密以保证安全性,但必须保证节点能够在加密数据上完成交易验证任务,否则无法实现共识.在这方面典型的应用是门罗币(XMR)和 Zcash 的加密机制.其中,门罗币通过对输出地址使用加密算法生成随机的新地址从而隐藏了资金的真正去向,Zcash 则将交易全部信息使用非对称加密,实现只有密钥持有者才能查看交易内容,同时使用零知识证明技术 zk-SNARKs 来验证交易.加密保护机制的局限性主要在于加密后验证效率低,同时加密算法的安全级别随着技术进步将逐渐降低.

(2)数据隔离机制

数据隔离是将具有不同隐私需求的交易账本存储在不同的分布式账本上.超级账本使用的就是此种保护机制.如图 6(b)所示,超级账本中建立多条通道分别对应多个分布式账本,记账节点只接收并维护与其相关的通道数据从而保证消息的隔离性和私密性.对于共识节点,为防止其因接收所有通道交易数据造成隐私泄露,可以采用只传输共识所需部分信息的方式再结合同态加密技术实现共识.

(3)链外存储机制

链外(Off-Chain)存储的基本思想是将需要保密的隐私数据存储在区块链之外,仅将可公开的内容发布在区块链上.在这种方式中,主区块链数据仅作为交易事件发生证明,完整的交易信息则采用链外消息传递或状态通道机制存储在相关节点或侧链上.摩根大通的 Quorum 系统中私人消息采用链外中继方式,即区块链中仅记录消息的加密指纹.比特币体系下的闪电网络则属于典型的状态通道机制,其中交易明细不作为记录存储在分布式账本上,仅作为有争议发生时的单据,从而实现隐私保护的效果.

(4)其他隐私保密机制

部分明文机制是将区块中的交易数据分为敏感数据和非敏感数据.典型的是 Corda 系统的类盲签名技术,其中通过去掉敏感数据后生成 Merkel 树的方式实现交易的隐私保护.

身份混淆机制是将交易者的身份进行匿名化的隐私保护方法.超级账本通过使用一次一密的方法对用户匿名化,从而实现交易之间具有无关联性.其他的身份混淆技术还有群签名和环签名,其基本思想都是任意一个成员以匿名方式代表一个群体或一组人进行签名而不泄露实际签名者的信息.

4.5 区块链数据存储扩展技术

虽然区块链系统具有非交易数据的分布式存储功能,但现有区块链系统为了保证链上存储空间的使用效率,都设计了相应的存储扩展机制实现对附加数据的存储,以避免非交易信息过多的占用区块空间.

(1)比特币系统的 OP_RETURN 类型

比特币系统在交易的 output 部分提供了两种类型:P2PH 类型和 OP_RETURN 类型,前者用于记录交易的接收方地址,后者用以记录交易地址以外的其他数据.OP_RETURN 类型本身的数据结构

与 P2PH 类型是相同的. 由于存储过多的非交易信息会极大地影响区块链性能，因此这部分的大小在比特币的区块链系统中被限制为 40 字节. 目前，比特币区块的这一存储空间主要与元数据功能相结合，被用于资产的发行和转移活动，以及版权声明等信息的永久存储.

(2) 多链存储机制

由于区块链上存储大量非交易数据不仅造成了严重的网络拥堵，同时还会提高区块链上数据存储的成本. 为此相关机构提出了多链架构技术，典型的是 EKT 框架[①]，该技术采用并行多主链结构，分为 Token 链和 DApp 链，其中不同的主链可以采用不同的共识机制. 这样，用户可以将账户信息存储在主链用户系统上，而基于智能合约的 DApp 存储在其他主链上. EKT 框架为解决跨链资源交换问题设定的基本原则是不同主链上的代币交易所消耗手续费由交易发生主链决定. 多链技术为区块链的存储容量扩展提供了有效的解决方案.

(3) IPFS 的文件存储

IPFS(Inter-Planetary File System)是一个可快速索引的版本化的 P2P 文件系统. 存储在 IPFS 上的文件将被自动分片并加密分散存储，同时自动消除重复文件，这保证了文件存储的安全性和高效性. IPFS 的核心是基于 DAG 结构的 MerkleDAG. MerkleDAG 由节点和链接组成，节点存储数据及数据的下级链接关系，链接存储的是数据的 Hash 值. IPFS 是为了用于替代 HTTP 协议，实现在网络中存储文件而被提出的.

现在，由于区块链系统的区块空间资源非常宝贵，很多相关工作将 IPFS 作为底层存储以扩展现有区块链系统的数据存储能力. 以太坊＋IPFS 的存储方案中就是将数据的 IPFS Hash 值存储在以太坊区块链的状态数据库中，而数据本身存储在 IPFS 系统中.

5　区块链系统的分布式查询处理机制

区块链系统为了支持对交易记录的验证等操作需要提供相应的数据查询功能. 由于现在区块链系统的数据结构和存储策略均不相同，因此系统所支持的查询功能也不同，相应的查询处理策略也具有较大的差异. 本节对区块链系统的查询类型和处理策略进行说明和分析，并进一步对区块链系统中工作量证明共识机制在应用中所涉及的分布式处理技术进行说明.

5.1　区块链系统的查询类型

各个区块链系统所支持的查询操作主要由区块链系统的数据内容和存储结构所决定. 根据查询的数据对象可以把区块链系统的内部查询功能分为账户查询、交易查询和合约查询. 这里所讨论的查询是基于各区块链系统原生数据结构和存储所提供的查询功能.

(1) 账户查询

账户查询是指对区块链系统中的账户地址相关状态进行的查询，通常用于交易验证. 交易验证主要是验证账户的数据货币余额，或者验证是否存在双重支付，通常区块链中的记账节点需要执行此功能. 比特币和以太坊系统均使用了代币以作为记账奖励，因此账户查询主要是对账户地址所拥有的数字货币的查询. 比特币系统并没有账户的概念，而是使用由交易所生成的状态数据 UTXO 来计算出账户地址所对应的比特币余额，其查询需要访问存储在 LevelDB 中的 UTXO 数据. 以太坊则是使用账户余额方式存储账户状态，相关的状态数据存储在状态数据库中，并通过状态树与区块关联. 基于 Merkle Patircia 树结构的状态树，以太坊能够支持查询账号是否存在和查询指定账号余额. 而对于超级账本系统，由于其并没有采用代币作为记账的激励机制，因此其账号状态并不包含余额信息. 超级账本的账号是一组证书和密钥文件，在排序服务节点、记账节点、客户端、CLI 接口等操作都需要使用账号执行相关操作，由此可见账号在超级账本系统中十分重要. 为此超级账本系统使用了专门的 CA 认证节点来实现账号信息的管理.

(2) 交易查询

交易查询是指对区块链交易进行访问，可以实现支付验证. 支付验证用于在提交一笔交易后判断该笔支付交易是否已经得到了区块链节点共识验证. 比特币系统的交易查询是验证一个交易在提交后是否有效，即检查交易是否被记录在区块中，主要采用基于交易的哈希值访问交易所在区块数据的方式实现. 在以太坊系统中，则是基于交易树访问交易数据实现对交易的验证. 超级账本系统在内部基于其区块索引结构实现快速查找区块和交易的功能，区块索引能够支持多种交易数据的访问，包括基于交易编号、交易验证码和区块编号获取区块和交易.

① EKT. https://ekt8.io/

在超级账本中区块数据在提交到账本前,执行基于状态数据的区块验证,在验证过程中需要访问状态数据.

(3)合约查询

合约查询是指对智能合约的相关信息进行查询.以太坊中的收据树使其能够支持查询一个智能合约在过去一段时间发生某类型事件的所有实例.而使用状态树并结合虚拟客户端的方法,以太坊能够查询智能合约中交易的输出.超级账本的智能合约也称为链码(Chaincode),不同于以太坊的智能合约,链码并没有将智能合约代码序列化后存储到区块中,而是运行于 Docker 容器之中.链码执行智能合约主要是在链码实例化后通过与背书节点(Endorser node)建立连接来通讯,再通过背书节点访问账本中合约相关数据.

5.2 区块数据的分布式处理策略

在早期的区块链系统体系结构中,节点之间是对等的,每个节点都会存储全部区块数据,因此对区块和交易的查询处理都是本地执行.然而,随着各类区块链系统的发展,节点功能逐步细化且节点角色逐渐多样化,同时交易数据开始使用分布式数据管理方式存储和访问.为此,当前区块链系统中的交易记账处理和交易查询处理均使用了分布式处理策略执行.此外,为了实现对区块链中数据的高效访问,也有相关研究在区块链系统之上构建查询处理框架.

(1)区块链数据的本地查询处理

对于区块链的全节点来说,其中集成了区块链系统的全部功能也存储了区块链的全部数据,因此全节点可以在本地执行全部的数据查询处理.区块链系统在全节点上对区块数据和交易的查询处理采用与集中式数据库相同的处理策略,即对区块扫描和使用索引结构提高访问效率.很多区块链系统使用 LevelDB 数据库存储区块数据或元信息,其目的也是为了使用键值数据库的 LMS-tree 结构提高写入和查询数据的效率.相比遍历区块的方法,使用键值数据库的查询功能显然更加高效.超级账本这类区块链系统,更是创建了多个区块索引,以提高对区块和交易的访问效率.

(2)区块链数据的分布式查询处理

由于区块链系统将区块相关的数据分布式地部署在节点中以减轻节点负载,诸多分布式查询处理技术被应用于区块链系统之中.

比特币系统使用的简单支付验证(SPV)机制就属于分布式查询技术.简单支付验证节点在进行交易的支付验证时,首先从全节点下载存储最长区块链的区块头至本地,再从区块链获取待验证支付对应的 Merkle 树哈希认证路径,最后根据哈希认证路径,计算 Merkle 树的根哈希值,将计算结果与本地区块头中的 Merkle 树的根哈希值进行比较,定位到包含待验证支付的区块.如果定位的区块包含在区块链中且区块高度符合确认要求则认为该支付验证通过.简单支付验证的执行过程如图 7 所示,其查询原理与分布式数据库的半连接算法类似.其中,由于验证交易需要发送用户地址从而造成隐私安全问题,比特币系统在 2012 年的 BIP37 中引入了布隆过滤器(Bloom Filter)对传输区块进行过滤.将交易地址转换为布隆过滤器可以有效避免直接发送地址而造成的隐私泄露,而全节点也仅需根据布隆过滤器结果返回部分区块信息即可.简单支付验证节点为了保证验证结果的可信性,会与多个全节点通信对交易进行验证.

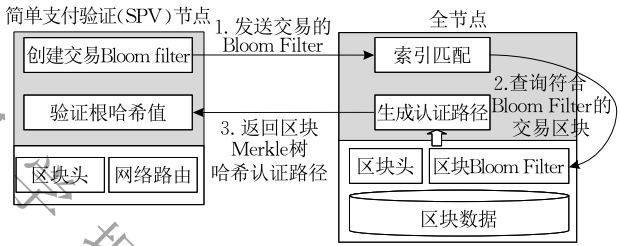


图 7 简单支付验证的分布式查询处理

(3)基于分布式计算的矿池技术

在区块链系统生态环境中,矿池技术是一个典型的分布式数据处理技术的应用.对于比特币和以太坊这类使用代币作为激励机制的区块链系统,需要使用工作量证明(PoW)等方式来实现记账的共识机制.随着参与共识的节点数量的增加,全网算力水准也不断增加,这导致单个记账节点具有过低的概率获得区块的代币奖励.为此,矿池机制通过将工作量证明任务划分成大量的 share 分发给具有算力的矿工节点,矿工节点运算后将 share 结果提交给矿池,最终挖出的区块由矿池节点作为整体的记账节点提交区块到网络中.矿池中的每个矿工节点仅提供算力而无需存储数据,这使得提供算力的节点能够轻量化.此外,矿池能够有效地通过分布式处理的方式聚集网络中零散的算力资源来提高挖矿效率,同时通过相应的分配模式使得提供算力的节点都能够获得奖励.

矿池所使用的收益分配模式包括 Pay Per Last

N Shares 模式(PPLNS)、Pay-Per-Share 模式(PPS)和 PROP 模式等. PPLNS 模式是根据过去的 N 个股份来支付收益，即矿池一旦发现了一个区块，所有提供算力节点将根据各自贡献的股份数量占比来分配区块中的货币，其特点是收益延迟，需要等到矿池产生区块获得奖励后才能分配收益. PPS 模式立即为每一个 share 支付报酬，其中的支出来源于矿池现有的数字货币，因此提供算力的矿工节点不用等待区块生成完毕或者确认即可获得收益. PPS 模式需要矿池对节点的算力进行估算，以计算其在矿池中的算力比例并按此比例分配收益. PPS 模式对于算力提供节点具有收益稳定的特点，因此也是当前各矿池所主要采用的收益分配模式. PROP 模式是基于 PPLNS 模式的，但是需要等区块确认后才会分配收益，这导致收益的滞后性更长.

虽然矿池技术能够为提供算力的节点带来稳定的收益，但是，矿池技术的应用也会为区块链系统带来弊端. 矿池将大量的算力集中到一个记账节点用以计算工作量证明问题，一方面，过高的算力导致区块生成的难度值不断增加，矿池之间的竞争需要更大的算力，从而造成了能源的浪费，另一方面，过于集中的算力很可能发生垄断记账权，造成 51%攻击的双重支付问题. 现在，已经有多个基于区块链的数字货币系统发生了单一矿池算力高于全网算力 51%的情况，为此系统维护团队不得不通过增加验证确认区块数量的方式以临时应对.

(4) 区块链的查询处理辅助系统

面对区块链系统本身的查询功能有限的问题，大量研究工作和应用系统分别提出了链上和链下的解决方案. 链上解决方案指基于区块链存储结构进行改进并提高性能，链下解决方案是在区块链的数据层之上构建了查询层以扩展区块链的查询功能并为用户提供更加灵活高效的区块数据访问. 当前大量的区块链查询系统均是采用的链下解决方案. Etherscan① 则提供了一个以太坊区块浏览器，任何在以太坊区块链上发生的交易都可以在 Etherscan 上访问查看. Etherchain② 同样提供了在线的以太坊的区块、交易和账户的数据访问功能. Blockchain.info③ 和 Toshi④ 提供了对比特币数据的 RESTful 服务访问方式，其底层采用了将区块数据存储与关系数据库系统中的方式以提供结构化的查询功能. 以上这些系统提供的是基本的区块数据查询功能，近期的研究主要致力于面向区块链提供更复杂的查询. EtherQL[44] 提出了一个基于以太坊系统的查询处理层系统，实现对区块中数据的范围查询和 top-k 查询等更高级的功能，以及支持上层应用的区块链数据分析. Gupta 等人[45] 提出了面向超级账本的时序查询处理模型，这些模型将时序相近的事件集中存储并共享元信息. Bartoletti 等人[46] 提出了一个区块链数据分析通用框架并设计实现，该框架能够支持比特币和以太坊的数据分析. 此外，也有研究工作和应用系统利用区块链的存储特性扩展数据库系统. ChainSQL⑤ 将区块链与传统数据库结合，实现一种全新的基于区块链技术的数据库应用平台，通过配置数据库的区块链节点，提高数据库的恢复效率.

针对区块链的处理性能，BLOCKBENCH[47] 提供了一个面向私有链的基准测试框架，提供了用于整体性能评估的宏观评价基准和用于每层性能评估的微观评价基准. BLOCKBENCH 对以太坊、Parity 和超级账本进行了分析，并与分布式数据库进行了对比.

6　区块链系统的分布式数据管理挑战

区块链系统作为数字货币的底层技术得到了快速发展，现在已经从资产数字化与智能合约的区块链 2.0 进入到通过广义资产交互与各行业深度融合的区块链 3.0 阶段. 区块链系统在分布式数据管理上的优势已得到认可，但同时制约区块链技术应用的挑战也逐渐显现. 本节将对区块链的分布式数据管理技术挑战进行概述并展望发展趋势.

6.1　区块链数据的分布式存储机制

区块链系统在记账的全节点上保存了全部区块数据，随着系统持续的运行，区块链中的数据将不断累积增加，这造成了节点高昂的存储代价. 即使是对于交易速度有所限制的比特币和以太坊，其公有链数据也已经达到了较大规模. 在 2019 年初，每秒 7 笔交易的比特币区块链系统的数据规模已经达到 200 GB，以太坊的公有链也达到了 1 TB 以上的数据规模，一旦技术上解除了交易速率的限制，这些区块链系统的数据规模还将以更高的速率增长. 对于大规模的区块链系统而言，很难保证每个存储节点都

① EtherScan. EtherScan.io. https://etherscan.io/
② Etherchain. https://etherchain.org
③ Blockchain.info. https://blockchain.info/
④ Coinbase: Toshi project. https://github.com/coinbase/toshi
⑤ Chainsql. http://www.chainsql.net/PDF/ChainSQL 技术白皮书 v1.0.pdf

能够应对如此高昂的存储负载.区块链的全副本分布式存储机制在增加本地查询处理代价的同时,也提高了区块链系统网络通信代价.在未来的应用中区块数据的存储模式将是制约区块链发展的主要问题.

在区块链中使用分布式存储策略是解决区块链系统的存储负载问题的方案,能够有效地降低各个记账节点的负载.区块链的分布式存储策略的目标是保证记账节点不必存储全部区块数据,同时系统中的节点依然能够保存大量完整区块链副本.当前的区块链系统的分布式存储策略依然还存在弊端.比特币虽然使用了轻量级的简单验证节点,但全节点依然要存储全部区块链数据.超级账本的存储策略中每个记账节点可以通过制定通道从而仅存储特定通道上的区块链数据.这种方法虽然避免了参与节点存储系统的全部数据,但由于每个通道对应一个独立的账本区块链,因此超级账本区块链在分布式存储结构上依然是全副本存储.

在区块链系统使用分布式存储策略的主要挑战包括以下几个方面:

(1)区块链的存储划分策略

存储划分策略是指记账节点仅保存区块链的部分子链片段,从而降低存储代价,其挑战是保证在指定节点集合内区块链具有完整副本.这种方案能够降低节点的存储负载,但需要保证子链片段在系统中的副本数量,否则极有可能发生低成本攻击问题,从而降低系统的安全性.

(2)划分存储的区块数据查询处理模型

区块链分片存储后需要有高效的查询模型保证交易验证的执行.现有的查询都是基于完全账本所设计的,而在存储分片后,会产生子链结构,这将导致大量交易时跨组交易,也就是需要通过主链进行中转.这种处理方式将严重降低系统在 TPS 上吞吐效率,一种可行的解决方案是将主链承担的中转任务再次进行分解,这也为区块链数据的查询处理提供了新的思路.

(3)构建新型区块链数据结构

区块链系统的数据组织逐渐多样化,分布式存储的数据内容不仅是区块数据,还包含状态数据等.新型的区块链数据结构将有助于解决现有区块链在数据存储方面局限性.一类方案是将账本的结构改变,例如变成有向无环图(DAG).基于 DAG 结构的区块链系统,处理的粒度由区块变成了交易,而且有多个头部可以同时记录交易,这样可以并行确认.另一类方案是将交易分散到多条链上,例如基于以太坊的分片(Sharding)项目.这类方案可以通过将用户分组,每组用户单独有一条子链的方式提高交易速度.这些新型区块链结构为了提高交易速度,都采用了一些中心化的组件,因此如何完全实现去中心化将是这些新型区块链结构要解决的问题.

6.2 区块链交易数据的共识机制效率与安全性

共识机制是区块链系统中事务达成分布式共识的算法,是保证各记账节点管理交易数据一致性的关键.

当前公有链系统普遍采用以 PoW 和 PoS 为基础的共识算法.基于 PoW 的算法通过计算复杂的数学问题防止女巫攻击,但造成了共识效率的下降和计算资源的浪费,导致区块链系统整体的吞吐量和扩展性均较低.而随着矿池技术不断将算力集中,少量几个矿池的算力就可以超过全网算力的 51%,可以看到算力聚合的模式下 PoW 也无法保证能够抵御 51%攻击.PoS 共识机制能够解决算力 51%攻击问题,但潜在着股权 51%攻击问题,且在区块分叉时依然存在安全性问题.DPoS 共识机制通过股权投票的方式选取 101 个记账权益对等的节点,这解决了由于矿池机制产生 PoW 的算力 51%攻击和 PoS 的股权 51%攻击问题,但依然存在着持股人参与投票度不高而导致的恶意节点 51%攻击问题.

可见,共识算法是限制当前区块链应用和发展的因素之一.对于公有链系统,矿池机制将算力汇聚到少量节点之上,导致 51%攻击发生的概率显著提高,其本质原因是计算资源更容易通过分布式处理环境被集中用于解决可分解的集中式问题,可见分布式处理技术反而为区块链技术带来了安全性问题.在联盟链和私有链的区块链系统中,可以通过使用 PBFT 和 Pasox 这类分布式一致性算法提高交易验证的吞吐量,同时降低资源的消耗.在安全性方面,联盟链和私有链系统所面临的问题主要是拜占庭容错问题,这要求网络中的恶意节点要低于节点总数 1/3.此外,当前出现了大量的新型共识算法被提出以解决区块链系统中交易验证的效率和安全性问题,包括容量证明算法 PoC(Proof of Space,也称为 PoSp)、授权拜占庭容错算法 dBFTP、Hash-Graph 和分层共识算法 HashNet[48] 等.容量证明算法由 BitTorrent 创始人提出,核心思想是通过证明节点存储数据量来决定获得记账权的概率,从而解决算力上的资源浪费并提高验证效率,但潜在着存储空间的 51%攻击问题.HashGraph 和 HashNet

算法主要面向 DAG 区块链系统. HashGraph 采用 Gossip 通讯协议和虚拟投票的拜占庭协议实现以交易为粒度的共识，其最大的特点是共识达成是异步的. HashGraph 共识机制面临的主要问题在于多轮的投票验证可能会降低共识的效率.

这些新型共识机制主要针对共识的效率、资源消耗、区块链存储结构进行优化，以保障区块链系统的稳定运行. 表 4 对当前区块链系统所使用的主要共识算法在分布式数据管理相关的记账效率、安全性、能源消耗和适用的区块链类型进行了对比.

表 4　主要共识算法在分布式数据管理功能上的对比

共识算法	记账效率	安全性	资源消耗	适用类型
PoW	10 分钟	算力 51%攻击	高能耗	公有链
PoS	秒级	股权 51%攻击	中能耗	公有链
DPoS	秒级	恶意节点 1/2 攻击	低能耗	联盟链
PoC	秒级	存储 51%攻击	低能耗、高存储	公有链
Pasox	秒级	节点 51%失效	低能耗	私有链
PBFT	秒级	恶意节点 1/3 攻击	低能耗	联盟链
HashGraph	秒级	恶意节点 1/3 攻击	低能耗	私有链

为此，共识算法需要具备以下性质才能够保证区块链中分布式数据管理的安全性，并在此基础上提高交易验证的效率：

(1) 一致性. 节点收到相同顺序交易数据时，每个节点发生的状态数据改变是一致的.

(2) 存活性. 在没有通讯故障的情况下，每个非故障节点最终都能够接收到提交的所有交易.

(3) 可用性. 区块链网络中存在网络延时或网络故障时，能够确保有效的交易被验证并且节点之间数据一致.

(4) 容错性. 区块链网络中存在部分恶意节点试图篡改交易或攻击网络时，依然能够确保交易数据在节点上的数据一致性.

新的共识算法将以更低的延时、更低的能源消耗、更加公平、更加安全为目标进行设计. 共识算法的改进需要与区块链系统的应用场景相结合. 具体包括以下路线：

(1) 改变达成共识所需消耗资源形式

针对 PoW 机制能源消耗过高的问题，改进的思路主要集中在改变达成共识所需要消耗的资源形式，即由背后的电力能源转变为其他资源. 早期的 PoS、DPoS 及其衍生的共识算法是将共识所需的能源转变为持有代币的潜在权益，之后的空间证明 PoC 则将消耗的资源变为磁盘空间，而有益工作证明 PoUW，则将消耗的能源用于进行有价值的运算. 以此思路改进的共识算法是以有效降低能耗为主要目的，同时与激励机制紧密耦合的，在提高交易处理效率方面则需要一定程度上牺牲去中心化的特性才能够实现.

(2) 改进传统分布式一致性算法

传统的分布式一致性算法以非拜占庭容错的 Paxos 族共识算法为主，这类算法主要用于维护分布式数据库的副本一致性[49]，但由于难以应用于公有链系统中，所以主要服务于联盟链系统. 对于非拜占庭容错算法的改进，主要是使其支持拜占庭容错，从而提升安全性. Tangaroa① 算法和 AlgoRand② 算法都是属于此类共识算法. 基于此思路改进的共识算法在性能上能够满足高吞吐量和快速共识的需求，在能源消耗上也大幅降低，并且不需要代币激励从而降低了交易成本，但保障安全性则成为待解决的核心问题.

6.3　区块链系统的高可用分布式查询处理

区块链系统中对区块数据的查询处理局限性在于仅支持对交易的简单查询访问功能且查询效率较低，要将区块链系统融合到各领域应用之中，区块链的查询处理机制还有待进一步的改进. 区块链系统的查询处理机制需要以提高查询性能、扩展查询功能和保证用户信息安全性为主要目标. 对于区块链查询处理机制的研究挑战主要包括以下几个方面：

(1) 构建高效的数据索引

通过对区块链中区块数据在记账节点上的逻辑存储和物理存储上的优化，实现对扩展查询功能的支持. 面向区块数据创建区块索引是一种有效的优化方法. 现有的区块链系统大多也是采用此类方法，但索引的对象还只局限于交易相关数据. 设计新型的区块数据存储结构和索引结构能够有效扩展查询功能和查询性能. 这对于 DAG 区块链这类新型系统将能带来更显著的性能提升. 区块链的数据索引不仅可以服务于实时的交易验证，也可以提高对交易历史信息的访问效率. 因此，在性能上区块链的数据索引访问机制将向着提供更高的实时访问效率方向进行设计，在功能上区块链的索引结构将面向支持基于历史数据溯源查询的方向发展.

(2) 查询处理层的嵌入

在区块链系统的体系结构中嵌入查询处理层同

① Tangaroa: A byzantine fault tolerant raft. http://www.scs.stanford.edu/14aucs244b/labs/projects/copeland_zhong.pdf

② Scaling byzantine agreements for crypto currencies. http://eprint.iacr.org/2017/454

样是扩展区块链系统查询功能的有效方法.查询处理层可以构建于数据层之上,在节点上直接支持对区块链完整数据的快速访问.嵌入的查询处理层需要为区块链提供更多的交易数据查询功能和数据分析功能,而不是仅仅局限于对交易记录的访问,这就需要查询处理层提供更多的 API 接口,并设计面向复杂查询的优化策略.此外,查询处理层还需要支持智能合约的相关查询,以便应用于更加复杂的业务需求.

(3)区块的分布式查询算法

区块链系统如果采用分布式存储策略,系统中部分节点将仅存储区块链的部分数据,此时需要使用分布式查询处理算法实现在不同节点之间对区块链数据的查询.区块链中的分布式查询将主要以查询性能优化为主要目标.对于基于分片(Sharding)的分布式存储,子链间的交易需要通过主链周转,而 DAG 结构和多链系统则需要跨链访问数据.在更加复杂的应用场景中,分布式查询还需要处理不同区块链间的跨链访问需求.因此新的分布式查询算法要解决多链之间的数据访问,实现在不同链上、不同区块结构上的分布式统一查询处理,并通过并行性提高性能.

(4)支持隐私保护的高效查询算法

公有链的区块链系统中,数字货币的钱包用户通常不会维护完整的区块链数据,因此用户对账号的查询需要通过存储有区块链数据的记账节点实现.现有的查询技术虽然通过布隆过滤器技术实现了查询时对用户账号隐私的保护,但随着区块链数据的增长,现有方式的查询负载将逐渐提高.因此,未来的区块数据查询算法将面临查询效率与隐私保护安全性之间平衡性的挑战,因为更好的隐私保护效果意味着更高的查询代价和存储代价.

6.4 分布式的智能合约管理机制

智能合约是一种计算机可读取和可执行的程序代码,能够自我执行和自我强制,不需要可信的第三方干预.在区块链系统中,智能合约作为一种特殊的交易事务被执行,合约的定义和输入输出都将被记录在区块链之中.然而,各区块链系统的智能合约依然不够成熟,在很多方面都需要逐渐的完善:

(1)智能合约的修改机制

智能合约是作为一个独立的程序在区块链系统中运行的,对于程序而言通常潜在着修改和升级操作以适应应用环境的变化.但是对于以太坊这类写入区块链的智能合约而言是不可篡改的,这样就限制了合约的更新.

(2)智能合约的权限管理

智能合约的现有权限管理机制还是相对较粗粒度的级别,并没有基于角色的权限访问机制(RBAC)那样能够精确到细粒度的资源,这一点对于像超级账本这类区块链系统尤其明显.智能合约的代码是向区块链网络内所有共识节点公开的,这对于很多金融贸易、企业交易来说是个巨大的弊端.为此,对于智能合约需要合理的权限管理来部署、审核和授权合约的访问与执行.

(3)智能合约的图灵完备性

区块链的智能合约编程语言分为图灵完备和非图灵完备两种.图灵完备的智能合约可形式化为状态机模型,并保证世界状态在区块链网络中的所有节点保持数据一致性.在设计图灵完备的智能合约时保持数据一致性是主要的挑战,往往需要在数据一致性和交易的确认效率之间做出权衡.

6.5 区块链数据隐私保护机制

区块链系统为了保障分布式存储中交易数据不会发生隐私泄露,其采用的保密机制可以分为两类:基于加密方法的保密机制和基于限制发布的保密机制.而要解决已发布数据的保密性问题,最终还是需要借助加密方法进行实现.在区块链数据加密方面主要的研究方向和挑战包括以下几个方面.

(1)基于同态加密的隐私保护机制

同态加密技术是一种密码学技术,其效果是将数据进行处理获得输出,将这一输出进行解密,其结果与用同一方法处理为加密数据得到的输出结果是一样的.使用同态加密技术能够更好地降低加密数据被破解的可能性,但其本身技术还不成熟,目前还停留在理论阶段,缺少真正可用的全同态加密算法.

(2)基于安全多方计算的隐私保护机制

安全多方计算允许参与方无需将数据汇集在一起进行分析,原理是允许一组用户基于他们的输入进行联合计算,而不需要每个用户显示其输入值,从而保证用户数据的隐私性.例如实现在不公布两个账户余额的情况下比较余额的多少.安全多方计算为交易数据的验证提供了隐私保护方案.然而,其面临的挑战在于安全多方计算在实际应用中执行效率极低,将严重影响交易的吞吐量.

(3)智能合约的隐私保护问题

交易记录数据可以通过加密方式实现隐私保护,而智能合约则无法通过加密措施实现隐私保护.其原因在于当前智能合约是公开的程序代码,其内

容无法进行混淆处理，因此其中的发送方和接收方地址，以及交易金额均无法匿名化. 这已经成为限制智能合约公有链应用的最大障碍. 已有研究工作中，对于解决智能合约隐私保护问题的思路包括放弃合约的可编程性和将智能合约与共识机制分隔等方法. 因此，隐私智能合约的出现都将有助于推动区块链系统的实际应用.

6.6　区块链系统的数据监管技术

区块链系统具有去中心化、不可篡改和伪造、数据高度冗余等诸多特性，这些特性保障了信息的不变性、安全性、透明性，使其在应用中发挥重要作用，但区块链技术是具有中立性的，这种不可回滚的数据管理模式也产生出了严重的数据监管问题[50]，而这一问题正日益引起金融和政府机构的重视. 在金融领域，在缺少数据监管措施的情况下，区块链系统的任何细小漏洞都会造成难以估量的损失. Facebook 面向金融领域开发的 Libra 区块链系统①就是由于缺少健全的监管机制而被美国政府要求停止开发. 而在公有链系统中，由于缺少数据修正机制，一旦隐私信息或违法数据被写入区块链发布，则会因为难以撤销而失去对数据监管的控制. 虽然政府机构普遍认识到需要对区块链系统构建谨慎的监管法规，但这需要建立在区块链的数据管理等技术成熟的前提下. 为此，区块链系统可以通过以下数据监管技术完善系统中数据监管功能.

(1) 区块链数据的回滚机制

通过使用基于事务的数据回滚机制，数据库管理系统可以将数据恢复到某一时刻的一致性状态. 目前的区块链系统虽然具有信息不可篡改特性，但依然可以通过硬分叉的方式实现数据回滚的功能，当前的以太坊系统(ETH)和以太坊经典(ETC)就是由于为了回滚数据而进行硬分叉的结果. 可见，现有区块链系统的数据回滚操作不仅代价高、效率低，而且无法从根本上解决回滚问题. 为此，未来区块链系统的分布式数据存储架构需要设计更加轻量级的数据回滚机制，而智能合约技术的发展可能为数据回滚机制带来新的解决方案.

(2)“以链治链”机制

区块链系统本身所具有的特性经常被用于在各领域中对现有系统的数据和运行流程进行监管，中国人民大学杨东教授提出了“以链治链”思想. 同样，我们也可以使用一个区块链系统监管另一个区块链系统的数据操作，即实现“以链治链”机制. “以链治链”机制需要合理设计出监管区块链和数据区块链之间的数据组织结构、数据溯源流程和监管操作规则. 文献[51]提出的可监管数字货币模型就采用了此种机制. 对于数字货币以外的区块链系统的监管还需要进一步设计相应的基于区块链的监管机制.

(3) 面向区块链数据内容的访问控制机制

在现有区块链系统缺少有效数据回滚机制的前提下，依然需要解决系统中出现的恶意数据的扩散问题. 一种有效的解决方法是面向区块链中的数据内容构建基于数据监管系统的访问控制机制，即对数据内容的访问需要通过监管系统审核方可执行. 这种机制能够在区块链中发现恶意数据后及时阻止其继续被访问，其主要挑战在于数据的发布、监管和访问控制间的数据内容加密、存储和管理机制与流程的设计.

7　区块链分布式数据管理对各领域应用展望

基于区块链系统的生态环境，使用区块链系统的分布式数据存储功能能够为大量领域应用系统带来新的应用模式. 目前，各应用领域已经在区块链技术上形成了初步的积累，逐步将区块链系统功能与原有业务系统相结合，利用区块链特性解决业务系统弊端，同时改进区块链系统自身的不足与局限. 现在，各领域的应用已经对区块链系统提出更多新的挑战.

7.1　泛金融应用领域

在泛金融应用领域，区块链系统能够为跨境支付、供应链金融、数字票据等金融支付与管理场景提供更安全、更便捷的自动交易和记账服务. 在跨境支付应用上已有诸多企业开始进行相关尝试，早期工作主要集中在利用虚拟货币为中小企业提供低成本的跨境汇款，而当前银行等金融机构则致力于利用联盟链的高吞吐量和高效确认特性进一步提高跨境支付的效率. 相比于早期 VISA 系统的 10% 费率和一周的到账时间，基于区块链结构的蚂蚁金服相关系统已能够实现 1 分钟到账，平均效率可提升 1 万倍，同时也具有更加低廉的支付成本. 对于票据与供应链金融业务，金融机构可以利用区块链技术实现票据资产的分布式管存，即将资产数据存放在区块链中，使资产的相关人员实现点对点的价值传输，无需实物票据和中心化系统进行控制和验证. 基于区

① https://libra.org

块链的供应链机制减少了中间人工参与成本，并实现端到端的透明化，从而提高资产信息处理效率的同时降低了管理成本.

金融领域中应用区块链系统的分布式存储、信息共享和不可篡改等特性管理数据，将交易记录、支付记录、数字资产等数据分布式地以全副本的方式存储在参与节点中.金融领域主要使用联盟链系统，通过参与方准入机制保证安全性，更简单的共识机制保证系统效率.然而，区块链系统依然需要解决交易处理效率、跨链交互和复杂行为自动处理等方面的局限性，才能够更好地支持金融服务.

首先，区块链系统在交易验证的实时性和吞吐量上还需要进一步提升.现有的 VISA 系统可以支持每秒数千笔交易，支付宝系统更是达到了 8.59 万笔/秒的交易峰值，而使用 DPoS 共识机制的 EOS 系统目前的 TPS 是每秒 3000 笔，基于 DAG 结构的 IOTA 系统的 TPS 极限是每秒 3000 笔.可见区块链系统处理能力还不足以替代现有的集中式交易系统.因此，区块链系统还需要不断改进共识机制以适应高吞吐的金融交易应用.

其次，未来区块链系统需要支持跨链数据交互功能.现有金融领域存在多个金融组织，这也意味着未来金融领域中将存在着多个区块链系统，各区块链系统在存储上是异构的数据结构.不同金融组织间的支付交易需要通过跨链交互实现，为此要解决不同区块链异构数据访问的问题.

7.2 产业供应链领域

产业供应链系统中往往存在有多达数百的加工环节，如此庞大的节点数量为供应链的追踪管理带来了极大的挑战.在供应链中，商品相关的生产、运输和销售等环节都需要记录大量的过程信息.使用区块链系统存储并管理供应链系统数据，能够有效地提供供应链过程信息的深度溯源、查询和验证等核心功能，从而提高行业的透明度和安全性.以饮料供应链为例，在原有机制下生产商在代理商拿走货物后就无法获得后续的销售信息.而通过区块链系统可以很轻松地将生产数据和销售数据融合，一方面方便最终消费者对商品溯源，另一方面为厂商提供透明的数据管理.在实际应用中，区块链还需要进一步提升信息安全性、构建溯源机制和建立错位激励措施来支撑产业供应链应用.

在安全性方面，由于区块链系统中信息共享的透明性，导致其中的隐私信息存在着泄露的风险.现有机制主要是利用数字签名和公私钥加密解密机制保障信息安全，这就需要区块链系统实现更加高效的智能合约机制，一方面简化供应链流程提升业务处理效率，另一方面提供自动化的签名验证防止个人信息等隐私数据泄露.

在数据溯源方面，区块链系统还需要进一步构建与产业供应链过程更加切合的溯源模型，能够及时、高效地发现供应链过程各个环节的作弊行为.溯源机制与智能合约的结合，将能够实现基于技术和算法的自动管理体系.

在激励机制方面，区块链需要实现错位激励措施.供应链系统中要为所有参与者设计合理的奖励措施，才能够有效地保证参与者加入区块链系统，并保证录入到区块链数据的质量.

7.3 网络文件存储领域

在网络文件存储领域，区块链系统可以构建公有链存储平台，提供个人数据存储和分享服务，实现数据资源的共享机制、溯源机制和确权机制.当前已有一些面向文件存储的公有链系统，如 IPFS 和 Sia.这些系统主要目标是构建一个去中心化的分布式数据存储，降低网络文件的带宽和存储成本.通过公有区块链系统与分布式数据存储平台相结合，可以构建包括文件数据存储、分享、治理和增值的文件存储生态系统.现有区块链分布式存储机制实现了去中心化，同时解决了安全性和吞吐率问题，但是在可编程性、可扩展性和可确权性方面仍然存在不足[52]，这也是基于区块链的网络文件存储面临的主要挑战.

可编程性主要是指把文件数据交易中的执行过程写入智能合约的可编程语言，并通过智能合约的代码强制运行机制保证交易执行的自动性和完整性.当前的智能合约主要支持转账交易的自动执行，而对于支持数据治理的智能合约及可编程语言则较少被提出.

可扩展性是指随着系统中节点数量规模扩大、文件数量增长，系统提供的吞吐率性能不会明显下降.由于网络文件存储主要以公有链形式运行，相应的区块链系统在节点达到一定规模后将无法使用高效的共识算法，此问题如果不解决将造成吞吐率的下降和可扩展性的降低.

可确权性主要指网络中的文件在经过复制等操作后能够基于区块链的溯源机制确认文件的原始拥有者和相关的授权信息，从而避免文件存储系统中的版权纠纷等问题.文件的确权管理需要区块链将溯源机制与分布式存储的元信息管理机制相结合，

这对网络文件存储提出了又一个新的挑战.

7.4　征信管理领域

在征信管理领域，原有中心化系统需要将各参与机构的共享信用数据汇聚、整理并发布，而其中还存在着第三方存证可信性的问题. 基于区块链管理征信信息，参与机构可以构建联盟链架构，并共享所拥有的信用数据. 此时，区块链系统自身具有的信息不可篡改、数据加密授权保护、智能合约等特性能够有效地解决原有征信系统中信用信息孤岛问题、提高系统安全性和降低征信运营成本. 但是，由于传统征信领域现存系统的封闭性，使用区块链系统替代依然具有较大的风险，区块链在征信管理中也面临诸多技术上的挑战，具体包括以下几个方面：

首先是数据的隐私保护问题. 基于区块链的征信平台通过全副本机制使得所有节点参与数据维护，这使得 51%恶意攻击难以实施从而保证了系统具有防数据篡改的安全性，但是由于链上数据全部公开，即便使用公钥私钥的加密方法依然具有隐私数据泄露的风险，因此系统需要支持对链上数据的控制机制，避免数据共享交易参与方之外的无关第三方获得数据.

其次是区块链上的数据质量控制. 由多参与方提供的征信数据潜在着数据不一致、冗余数据等数据质量问题. 如何在不可篡改的区块链征信系统上实现征信数据的消除冗余数据、保证数据一致性，是区块链分布式数据管理研究面临的又一挑战性问题.

最后是多样性数据的链上管理. 征信数据具有多模态的特性，覆盖结构化记录、文档文件、公正证书、甚至语音和邮件等数据类型. 为此，区块链系统需要提供多种数据类型上链存储的功能，并能够基于这些文件进行自动的验证处理.

7.5　其他应用领域

除了上述几个主要应用领域之外，区块链还能够应用于各类具有分布式数据存储及相关需求的领域：

(1) 分布式社交网络

分布式社交网络基于 P2P 技术构建，采用去中心化的结构，同时所有数据分散在网络节点之中. 基于区块链系统构建分布式社交网络能够通过去中心化社交避免单点失效，实现多样化的激励机制，具有较好的隐私保护机制，并通过激励机制降低运营成本，如 Steem 系统. 然而，分布式社交网络需要使用公有链模式运行，大规模的社交网络区块链系统将面临严重的共识效率和去中心化的平衡问题. 以 Steem 系统为例，当前使用 DPoS 共识机制通过 21 位证人来达成共识以保证系统的吞吐率，但其中潜在着中心化的风险.

(2) 公益慈善领域

区块链系统基于分布式存储和共识算法所构建的新型信任机制，能够实现公益信息的透明公开、有效监管，帮助公益慈善领域解决信任问题. 区块链系统可以记录公益流程中的捐赠项目、募集明细、资金使用和受益人反馈等全部信息，同时由多家公益组织、支付机构、审计机构构成多方参与的联盟链以提高系统信息的安全性. 其应用挑战主要来自于复杂的公益场景与简单的交易记账方式之间的不匹配问题，这需要通过构建符合复杂公益流程的智能合约来解决.

(3) 教育就业领域

教育就业领域中需要对学生的学籍信息、学历信息、就职信息进行管理. 原有分散的教育就业信息管理体系存在着大量的信息孤岛，各教育机构仅管理自身所涉及学生的信息，从而导致就业和升学过程中面对学历造假、学籍造假、简历造假等问题，单位和院校缺少有效的验证机制. 区块链系统的数据透明化和不可篡改特性能够有效地构建可信的教育存证机制，通过将学生教育证明信息加密存储在区块链上，从而支持学生教育就业信息的受控访问和自动验证. 由于教育数据也可以看作是一种征信数据，所以，区块链在教育就业领域面临与征信管理领域类似的挑战.

8　结　论

数据管理技术是支撑区块链系统的核心，本文梳理了区块链系统中的分布式数据管理涉及的主要技术，并对其面临的挑战和发展趋势进行了展望. 区块链系统与分布式数据库系统之间既有共同点也有着显著的差异，这些差异体现了区块链系统在存储层面的新特性. 随着区块链技术的发展，区块链系统被设计为不同的类型以适应不同的应用场景，同时区块链系统的数据存储也在区块结构拓扑、物理存储结构、数据组织方式和优化技术上不断发展，进行着各种创新性的尝试. 在数据的查询处理方面，虽然传统区块链系统对于区块数据的查询功能较少且性

能有限,但新型区块链系统的研究工作正在从多个环节上实现着对区块链数据查询处理的性能提升和功能扩展.区块链系统的分布式数据管理技术为当前各领域应用的数据管理提供了安全、可共享、可溯源的新特性.为了适应各种应用的发展,区块链的分布式数据管理技术也不断面临着新的机遇和挑战,其发展也势必会为各领域带来革命性的影响.

参 考 文 献

[1] Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. White Paper, 2008

[2] Buterin V. Ethereum: A Next Generation Smart Contract and Decentralized Application Platform. White Paper, 2013

[3] Christian C. Architecture of the hyperledger blockchain fabric //Proceedings of the Workshop on Distributed Cryptocurrencies and Consensus Ledgers. Chicago, USA, 2016: 14-17

[4] Zhang Zeng-Jun, Dong Ning, Zhu Xuan-Tong, Chen Jian-Xiong. Deep Exploration Block Chain: Hyperledger Technology and Application. Beijing: China Machine Press, 2018(in Chinese)
(张增骏, 董宁, 朱轩彤, 陈剑雄. 深度探索区块链: Hyperledger 技术与应用. 北京: 机械工业出版社, 2018)

[5] Yuan Yong, Wang Fei-Yue. Blockchain: The state of the art and future trends. Acta Automatica Sinica, 2016, 42(4): 481-494(in Chinese)
(袁勇, 王飞跃. 区块链技术发展现状与展望. 自动化学报, 2016, 42(4): 481-494)

[6] He Pu, Yu Ge, Zhang Yan-Feng, Bao Yu-Bin. Survey on blockchain technology and its application prospect. Computer Science, 2017, 44(4): 1-7(in Chinese)
(何蒲, 于戈, 张岩峰, 鲍玉斌. 区块链技术与应用前瞻综述. 计算机科学, 2017, 44(4): 1-7)

[7] Shao Qi-Feng, Jin Che-Qing, Zhang Zhao, et al. Blockchain: Architecture and research progress. Chinese Journal of Computers, 2018, 41(5): 969-988(in Chinese)
(邵奇峰, 金澈清, 张召等. 区块链技术: 架构及进展. 计算机学报, 2018, 41(5): 969-988)

[8] Tschorsch F, Scheuermann B. Bitcoin and beyond: A technical survey on decentralized digital currencies. IEEE Communications Surveys & Tutorials, 2016, 18(3): 2084-2123

[9] Liu Ao-Di, Du Xue-Hui, Wang Na, Li Shao-Zhuo. Research progress of blockchain technology and its application in information security. Journal of Software, 2018, 29(7): 2092-2115(in Chinese)
(刘敖迪, 杜学绘, 王娜, 李少卓. 区块链技术及其在信息安全领域的研究进展. 软件学报, 2018, 29(7): 2092-2115)

[10] Storj Labs, Inc. Storj: A Decentralized Cloud Storage Network Framework. White Paper, 2018

[11] Wilkinson S, Lowry J. MetaDisk: Blockchain-Based Decentralized File Storage Application. White Paper, 2014

[12] He Hai-Wu, Yan An, Chen Ze-Hua. Survey of smart contract technology and application based on blockchain. Journal of Computer Research and Development, 2018, 55(11): 2452-2466(in Chinese)
(贺海武, 延安, 陈泽华. 基于区块链的智能合约技术与应用综述. 计算机研究与发展, 2018, 55(11): 2452-2466)

[13] Yuan Yong, Ni Xiao-Chun, Zeng Shuai, Wang Fei-Yue. Blockchain consensus algorithms: The state of the art and future trends. Acta Automatica Sinica, 2018, 44(11): 2011-2022(in Chinese)
(袁勇, 倪晓春, 曾帅, 王飞跃. 区块链共识算法的发展现状与展望. 自动化学报, 2018, 44(11): 2011-2022)

[14] Yang Guang-Yu, Zhang Shu-Xin. Review and research for consensus mechanism of block chain. Journal of Information Security Research, 2018, 4(4): 369-379(in Chinese)
(杨宇光, 张树新. 区块链共识机制综述. 信息安全研究, 2018, 4(4): 369-379)

[15] Lin Iuon-Chang, Liao Tzu-Chun. A survey of blockchain security issues and challenges. International Journal of Network Security, 2017, 19(5): 653-659

[16] Zhu Lie-Huang, Gao Feng, Shen Meng, et al. Survey on Privacy Preserving Techniques for Blockchain Technology. Journal of Computer Research and Development, 2017, 54(10): 2170-2186(in Chinese)
(祝烈煌, 高峰, 沈蒙等. 区块链隐私保护研究综述. 计算机研究与发展, 2017, 54(10): 2170-2186)

[17] Pass R, Seeman L, Shelat A. Analysis of the blockchain protocol in asynchronous networks//Proceedings of the Theory and Applications of Cryptographic Techniques. Paris, France, 2017: 643-673

[18] Qian Wei-Ning, Shao Qi-Feng, Zhu Yan-Chao, et al. Research problems and methods in blockchain and trusted data management. Journal of Software, 2018, 29(1): 150-159(in Chinese)
(钱卫宁, 邵奇峰, 朱燕超等. 区块链与可信数据管理: 问题与方法. 软件学报, 2018, 29(1): 150-159)

[19] Wang Qian-Ge, He Pu, Nie Tie-Zheng, et al. Survey of data storage and query techniques in blockchain systems. Computer Science, 2018, 45(12): 12-18(in Chinese)
(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. 计算机科学, 2018, 45(12): 12-18)

[20] Tanenbaum A S, Steen M V. Distributed Systems: Principles and Paradigms. 2nd Edition. Upper Saddle River, USA: Pearson Prentice Hall, 2007

[21] Bonifat A, Chrysanthis P K, Ouksel A M. Distributed databases and peer to peer databases: Past and present. SIGMOD Record, 2008, 37(1): 5-11

[22] Ng W S, Ooi B, Tan K, Zhou A. PeerDB: A P2P-based system for distributed data sharing//Proceedings of the International Conference on Data Engineering. Bangalore, India, 2003: 633-644

[23] Özsu M T, Valduriez P. Principles of Distributed Database Systems. 3rd Edition. New York, USA: Springer, 2011

[24] Sacco M S, Yao S B. Query optimization in distributed database systems. Advances in Computers, 1982, 21: 225-273

[25] Afrati F, Ullman J D. Optimizing joins in a map-reduce environment//Proceedings of the International Conference on the Extending Database Technology. Lausanne, Switzerland, 2010: 99-110

[26] Rosenfeld M, Rosenfeld M, Rosenfeld M, et al. Proof of activity: Extending Bitcoin's proof of work via proof of stake [Extended Abstract]. ACM SIGMETRICS Performance Evaluation Review, 2014, 42(3): 34-37

[27] Castro M, Liskov B. Practical byzantine fault tolerance and proactive recovery. ACM Transactions on Computer Systems, 2002, 20(4): 398-461

[28] Lamport L. The part-time parliament. ACM Transactions on Computer Systems, 1998, 16(2): 133-169

[29] Alan D, Dan G, Carl H, Wes I, et al. Epidemic algorithms for replicated database maintenance//Proceedings of the 6th Annual ACM Symposium on Principles of Distributed Computing. New York, USA, 1987: 1-12

[30] Ongaro D, Ousterhout J. In search of an understandable consensus algorithm//Proceedings of the 2014 Annual USENIX Technical Conference. USENIX Association, Philadelphia, USA, 2014: 305-320

[31] Maymounkov P, Mazières D. Kademlia: A peer-to-peer information system based on the XOR metric//Proceedings of the International Workshop on Peer-to-Peer Systems. Berkeley, USA, 2002: 53-65

[32] Burton H B. Space/time trade-offs in hash coding with allowable errors. Communications of the ACM, 1970, 13(7): 422-426

[33] Churyumov A. Byteball: A Decentralized System for Storage and Transfer of Value. White Paper, 2017

[34] Liu C, Wang D, Wu M. Vite: A High Performance Asynchronous Decentralized Application Platform. White Paper, 2018

[35] InterValue Team. InterValue Techniqual Whitepaper. White Paper, 2018

[36] Merkle R C. Protocols for public key cryptosystems//Proceedings of the 1980 Symposium on Security and Privacy. Oakland, USA, 1980: 122-133

[37] Wood G. Ethereum: A Secure Decentralized Generalized Transaction Ledger Byzantium Version. Yellow Paper, 2019

[38] Brass P. Advanced Data Structures. New York, USA: Cambridge University Press, 2008

[39] O'Neil P, Cheng E, Gawlick D, O'Neil E. The Log-Structured Merge-Tree (LSM Tree). Acta Informatica, 1996, 33(4): 351-385

[40] BigchainDB GmbH. BigchainDB 2.0 The Blockchain Database. White Paper, 2018

[41] Wang S, Dinh T A, Lin Q, et al. ForkBase: An efficient storage engine for blockchain and forkable applications//Proceedings of the 44th International Conference on Very Large Data Bases. Rio de Janeiro, Brazil, 2018: 1137-1150

[42] Dinh A, Wang J, Wang S, et al. UStore: A distributed storage with rich semantics. ArXiv. Feb 2017, abs/1702.02799: 1-21

[43] Xu Zi-Hua, Han Si-Yuan, Chen Lei. CUB, a consensus unit-based storage scheme for blockchain system//Proceedings of the IEEE 34th International Conference on Data Engineering. Paris, France, 2018: 173-184

[44] Li Y, Zheng K, Yan Y, et al. EtherQL: A query layer for blockchain system//Proceedings of the International Conference on Database Systems for Advanced Applications. Suzhou, China, 2017: 556-567

[45] Gupta H, Hans S, Aggarwal K, et al. Efficiently processing temporal queries on hyperledger fabric//Proceedings of the IEEE 34th International Conference on Data Engineering. Paris, France, 2018: 1489-1494

[46] Bartoletti M, Bracciali A, Lande S, et al. A general framework for blockchain analytics//Proceedings of the 1st Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers. Las Vegas, Nevada, USA, 2017: 11-15

[47] Dinh T A, Wang J, Chen G, et al. BLOCKBENCH: A framework for analyzing private blockchains//Proceedings of the 2017 ACM International Conference on Management of Data. Chicago, USA, 2017: 1085-1100

[48] Cao Zhang-Jie, Long Ming-Sheng, Wang Jian-Min, Yu P S. HashNet: Deep learning to hash by continuation//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2018: 5609-5618

[49] Guo Jin-Wei, Chu Jia-Jia, Cai Peng, et al. Low-overhead Paxos replication. Data Science and Engineering, 2017, 2(2): 169-177

[50] Wang Jun-Sheng, Li Li-Li, Yan Yong, et al. Security incidents and solutions of blockchain technology application. Computer Science, 2018, 45(z1): 352-355, 382(in Chinese)
(王俊生，李丽丽，颜拥等. 区块链技术应用的安全与监管问题. 计算机科学，2018，45(z1)：352-355，382)

[51] Zhang Jian-Yi, Wang Zhi-Qiang, Xu Zhi-Li, et al. A regulatable digital currency model based on blockchain. Journal of Computer Research and Development, 2018, 55(10): 127-140(in Chinese)
(张健毅，王志强，徐治理等. 基于区块链的可监管数字货币模型. 计算机研究与发展，2018，55(10)：127-140)

[52] Cao Yuan, Zhang Chong, Ding Zhao-Yun, Jiang Xin-Wen. Blockchain Technology on DAG Principle and Practice. Beijing: China Machine Press, 2018(in Chinese)
(曹源，张翀，丁兆云，姜新文. DAG 区块链技术原理与实践. 北京：机械工业出版社，2018)

YU Ge, Ph. D., professor. His research interests include distributed database, distributed and parallel computing, and blockchain.

NIE Tie-Zheng, Ph.D., associate professor. His research interests include database, data integration and blockchain.

LI Xiao-Hua, Ph. D., lecturer. Her research interests include information security and blockchain.

ZHANG Yan-Feng, Ph. D., professor. His research interests include distributed data processing and cloud computing.

SHEN De-Rong, Ph. D., professor. Her research interests include distributed database and data integration.

BAO Yu-Bin, Ph. D., professor. His research interests include data warehouse and OLAP.

Background

This paper surveys the state of art techniques of distributed data management in blockchain systems. Blockchain is proposed to construct transaction systems that are decentralized, non-tampering, distributed consensus and final consistency for managing transaction records. So blockchain is also a kind of distributed database management system, and it will be widely used in the field of financial, banking, education, and so on.

Blockchain system has significant differences with traditional distributed database systems, since the participants of blockchain are organized in a peer-to-peer network. The differences mainly reflect on the pattern of data storage and query processing. Blockchain organizes records in blocks with a chain structure to link them, and allows every participant download the complete record collection. Based on blockchain technology, some blockchain systems have been constructed to support digital currency transactions or smart contracts, such as Bitcoin, Ethereum, Hyperledger fabric. These blockchain applications are also different each other, and adopt various techniques to manage data in their distributed platform. This paper focuses on discussed the techniques of distributed data management used in blockchain platforms according to the classification of blockchain, the topologic structure of blocks, the storage structure of data, and the mechanism of distributed query processing. This paper also points out the challenges and trends of existing blockchain system, which will be the future research issues for blockchain technology, and also for distributed data management.

This work is partially supported by the National Key Research and Development Program of China (No. 2018YFB1003404), which aims to develop distributed data storage and management techniques on heterogeneous computing architecture, the National Natural Science Foundation of China (No. U1811261 and No. 61672142), which aim to research on accurate entity fusion techniques for big data, and construction methods for big data based interactive and personalized instructional environments, respectively, and the Liaoning Science and Technology Foundation (No. 20180550321), which aims to research on smart contract concurrency controlled mechanisms for blockchain systems.

Our research group has been working on database systems, distributed data management, distributed and parallel computing, data integration, information security for many years. Related works were published in good-reputation journals and conferences, such as *Chinese Journal of Computers*, *VLDBJ*, *TKDE*, *TPDS*, *ICDE*, *VLDB* and *SIGMOD*.