北京邮电大学

硕士研究生学位论文开题报告

学 号: 2019140739

姓 名: 陈铭

学 院: 计算机学院

专业(领域): 软件工程

研究方向: 区块链技术

导师姓名: 杨谈

攻 读 学 位: 工程硕士

2020年11月8日

|  |  |  |  |
| --- | --- | --- | --- |
| 论文题目 | 基于区块链的博物馆数据溯源系统的设计与实现 | | |
| 选题来源 | 中央、国家各部门项目 | 论文类型 | 应用研究 |
| 开题日期 | 2020-11-8 | 开题地点 | 北京邮电大学 |
| **一、立题依据（包括研究目的、意义、国内外研究现状和发展趋势，需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录）（不少于800字）**  **1、研究的目的与意义**  博物馆是在文物及相关行政部门审核批准，得到许可后获得法人资格，并最终面向公众开放的藏品资源丰富、 观赏性和文化价值并存的非营利性社会服务机构。博物馆的存在最初是为了将人类活动和大自然运转的见证物进行收藏、展示、保护、研究。长期以来，人们对博物馆的注重点通常集中于馆内的藏品，很少关注到博物馆的藏品背后的多元化异构数据，比如藏品的出入库，藏品的提借和信息修改，这些都是保证藏品安全，实现数据共享交换的重要过程。  本论文希望系统通过区块链去中心化、不可篡改等特性，一方面以博物馆内业务部门为区块链节点组建馆内私有链，重点解决去中心化的分布式记账问题，实现数据共享交换过程的确权确责和追溯追责问题。针对私有链中共享的数据，将带有数据指纹的标签记录到区块链上，根据实时数据状态通过触发智能合约进行有效管控，实现馆内各业务部门间的基础数据和业务数据的加密安全传输、共享和确权，打破信息孤岛，对馆内事务进行穿透式管理。另一方面，以多家博物馆的对外信息公开部门为节点组建馆际联盟链，对联盟链中共享的公开数据，进行数据请求、审批、交换过程的验证和记录，明确界定共享权责，实现馆际间的公开数据远距离可信可靠共享。  **2、研究现状和发展趋势**  作为更流行的比特币[1]技术的基础技术，自2009年以来，区块链技术受到了广泛关注。从2015年至今，区块链迎来了非常快速的发展。以区块链为关键词的搜索记录一直居榜首，因此2015年被称为区块链元年[2]。经过近几年的发展，区块链技术变得越来越成熟，因此越来越多的区块链框架逐渐诞生。2009年，中本聪提出了比特币模型，该模型将区块链设计的概念传播的更加广泛。 Vitalik Buterin在2013年发布的以太坊模型[3]使得在其他领域使用区块链成为可能；2015年由Linux基金会领导的Hyperledger项目[4]，财团链已经开始进入人们的视野，并且区块链技术已经可以在某些特定领域发挥巨大作用。  与此同时，随着区块链技术的不断发展，各种相关的区块链应用也都日新月异。在电子货币领域，继比特币诞生之后，以太币、点点币[4]等各种各样的电子货币如春笋般的产生，目前已经取得了相当大的成功；在电子支付方面，由于闪电网络技术[5]的发明，使得小金额的快捷支付成为可能，大幅度加速了电子支付的速度，极大的促进了电子货币的发展，后期大量科研团队并在此基础上进行了进一步的发展和应用[6-7]；在区块链的的数据访问控制方面，电子科技大学[8]和中央财经大学[9]的研究团队将传统的数据访问控制技术加以改造，提出了相应的去中心化数据访问控制技术，取得不错的效果；在区块链技术落地方面，其核心技术己经用到了物联网[10]、智能制造[11]、供应链管理[12]、数字资产交易[13]等多个领域，并产生相应的影响。  在区块链存储方面的研究中，太原理工大学的研究团队提出了一种区块链数据的关系查询解决方案FabricSQL，这一方案将 Hyperledger与MySQL两种技术相结合，在SQL数据库中实时存入区块链的有效交易数据 [14]。在农产品应用中，福建师范大学研究团队考虑了数据传入过程中的效率问题，将IPFS与区块链存储相结合，借助联盟链来解决大数据区块链的传播问题，能够完成农产品数据的高效率实时查询，而且安全追溯系统的应用在区块链上，可以面向更多的应用场景[15]。区块链技术难以在大型业务场景应用，由于其存储效率的问题，电子科技大学创新的提出了一种基于门限秘密共享的 区块链分片存储模型[16]，该模型改进了PBFT共识算法，并使用了改进的 Shamir 门限，该模型将要上链的交易数据数字化后，分片存储于各个节点中，保障了数据安全性、隐私性、可靠性的同时，又能够有效减少每个节点所需存储的数据量。在粮油食品供应链的应用中，由于其食品数据链有节点多、供应链长和涉及面广的特点，在区块链上存储传播的效率很慢，因此提出了一种“链上( on-chain) + 链下( off-chain) ”双模存储机制[17]于区块链的粮油食品全供应链信息安全管理原型系统]，通过将供应链上的详细数据信息经过智能合约验证，最终与数据所在的区块链位置信息一同存入关系型数据库中。  在共识算法方面，Hperledger本身采用的是PBFT共识算法，但是由于全节点的确认机制，效率一般。因此有学者研究通过一致性哈希算法（Consistent Hashing）[18]对区块链运行中出现的问题进行了研究和优化，较好地解决了区块链存储空间问题、重复验证问题和数据分层管理问题，有效节约了网络存储资源。针对传统区块链共识算法存在效率低和大量资源浪费的问题，吉林大学研究团队提出一种使用基于ID分类的有向无环图数据结构取代传统区块链链式数据结构的算法[19]，解决了传统区块链项目无法在短时间内确认大量交易的问题。  **参考文献：**  [1] 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.  [2] 王元地,李粒,胡谍.区块链研究综述[J].中国矿业大学学报(社会科学版),20(3):74-86  [3] WOODG ,et al.Ethereum:A secure decentralized generalized transaction ledger[J].Ethereum project yellow paper,2014,151:1-32.  [4] CACHIN C. Architecture of the hyperledger blockchain fabric[C]/Workshop on distributed cryptocurrencies and consensus ledgers:volume 310.2016.  [5] PENG S. Bitcoin: Cryptography, economics, and the future[J].Senior Capstone Thesis School of Engineering and Applied Science, University of Pennsylvania, 2013.  [6] POON J, DRYJA T. The bitcoin lightning network: Scalable off-chain instant payments[Z] 2016.  [7] LUU L, NARAYANAN V, ZHENG C, et al. A secure sharding protocol for open blockchains[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.2016:17-30.  [8] 陈永丰.基于区块链的数据访问控制方法及应用研究[D].成都:电子科技大学,2018  [9] 王秀利,江晓舟,李洋.应用区块链的数据访问控制与共享模型[J].软件学报,2019(6):1661-1669.  [10] SUN H, HUA S, ZHOUE, et al. Usingethereum blockchain in internet of things: a solution for electric vehicle battery refueling[C]//International Conference on Blockchain. Springer,2018:3-17.  [11] VIRIYASITAVAT W, DA XU L, BI Z, et al. Blockchain-based business process management(bpm) framework for service composition in industry 4.0[J]. Journal of Intelligent Manufacturing,2018:1-12.  [12] ELMESSIRY M, ELMESSIRY A.Blockchain framework for textile supply chain management[C]//International Conference on Blockchain. Springer,2018:213-227.  [13] SWAN M.Blockchain:Blueprint for a new economy[M].2015.  [14] 余涛,牛保宁,樊星. FabricSQL：区块链数据的关系查询.[A] 太原理工大学 ，2020  [15] 王可可 ，陈志德 ，徐 健，基于联盟区块链的农产品质量安全高效追溯体系，计算机应用，2019，39( 8) : 2438 － 2443  [16] 张国潮，王瑞锦，基于门限秘密共享的区块链分片存储模型，计算机应用，2019，39( 9) : 2617 － 2622  [17] 许继平，孙鹏程，张新，王小艺，孔建磊，赵峙尧.基于区块链的粮油食品全供应链信息安全管理原型系统.农业机械学报，2020年2月，第51卷第2期.  [18] 苏畅，基于一致性哈希算法的区块链优化模型.Computer Knowledge and Technology电脑知识与技术.第 15 卷第 14 期 (2019 年 5 月).  [19] 王壹铭，初剑峰，王永军，陈彦东. 基于有向无环图的高效区块链共识算法.吉林大学学报.2020年9月第58卷第5期. | | | |

|  |
| --- |
| **二、研究内容和目标（说明课题的具体研究内容，研究目标和效果，以及拟解决的关键科学问题。此部分为重点阐述内容）（不少于2500字）**   1. **研究内容**   本论文计划针对博物馆藏品数据，分析了博物馆藏品从征集到入账以及修改信息、藏品提借等的业务流程，通过区块链解决去中心化的分布式记账问题，实现藏品数据共享，保证藏品本书数据以及藏品移动过程数据可溯源并且不可篡改。本论文针对博物馆系统工作流程构建数据结构，确定需要上传到区块链的藏品相关数据。实现藏品的征集到入账需经过征集、登编入库、入账等多个步骤中的关键数据能够共享和溯源，在原系统的基础上，基于区块链数据存储优化，采用“IPFS+区块链”的存储模式，将区块链节点存储优化，减少链上存储空间，提升区块效率。同时根据馆际联盟链节点优化区块链共识算法，减低全网负载，加快博物馆间数据交换的效率。  以下说明具体的三个工作：   1. 设计博物馆区块链数据结构，优化链上链下数据模型。   国家博物馆本身采用微服务架构设计将系统模块划分为藏品管理、藏品查询、数据共享等多个服务，分别独立部署，对外提供服务。藏品从征集到入账需经过征集、登编入库、入账等多个步骤，入账后根据需要也会对藏品信息进行修改。为了保留每个阶段的藏品数据信息。系统采用多套藏品数据并存的机制。本文根据业务流程划分和对应接口，提取出需要上链的藏品相关数据。   1. 藏品从征集到入账需经过征集、登编入库、入账等多个步骤，这一流程涉及：   征集信息审核阶段：从获取征集单及征集藏品信息开始，至征集藏品信息审核完成。在本阶段设置征集单数据表和藏品信息数据表。保留征集阶段的数据，征集信息不能进行修改。  藏品登编入库阶段：从藏品登编至藏品入库。在本阶段将复制征集信息审核阶段的数据，在此基础上进行更新。  藏品入账：藏品入库完成后自动入账。在本阶段将复制登编入库阶段的数据。作为藏品的总账信息进行管理。    图1 藏品登编入藏流程图  基于这一流程，定义数据结构，包括征集单编号、征集时间、征集藏品编号、入库时间，由于这一流程只用于溯源，无法修改，因此将具体数据存于区块链节点数据库，将数据哈希化后的Hash值存入区块链链上，提高溯源效率。  2）入账后的藏品修改，包括藏品信息更新、藏品注销、藏品提借不能在原有藏品总账信息上更新，将修改后的信息保存为新的记录，通过藏品编号与原记录建立关联关系，每次修改都会保留原有记录，创建新的记录，便于查看藏品信息修改历史。    图2 藏品信息修改流程图  基于藏品信息修改，定义数据结构，包括藏品编号、藏品修改时间、藏品历史信息和修改日志。 链上存储藏品编号、藏品修改时间，藏品历史信息和修改日志的hash，链下存储具体的藏品历史信息和修改日志。  3）藏品提借包括提借申请权限、藏品展览提借、藏品修复提借、藏品照相提借。    图3 藏品展览提借流程图  基于藏品信息提借定义数据结构，包括提借单编号，申请部门信息，申请人信息，策展人信息，出库时间，提取事由，藏品清单。提借单用于记录，无法修改，因此在链下存储具体提借单信息，链上存储hash值。   1. 采用IPFS优化区块链数据存储   考虑到目前区块链所面临的数据膨胀问题，采用分布式存储与区块链相结合的方法，IPFS星际文件系统是一个旨在创建持久且分布式存储和共享文件的网络传输协议。它是一种内容可寻址的对等超媒体分发协议。因此将具体的藏品信息和业务流程信息存入IPFS数据库中，再将IPFS网络生成的内容相关hash存储到区块链上，保证了区块链上的hash不可篡改，通过该hash访问的IPFS文件也不会改变。    图3 “区块链+IPFS”存储架构  Hyperledger平台本身的状态数据库采用的LevelDB和CouchDB本身是键值对存储的，虽然CouchDB增加了复查询的功能，但是查询方法还是受到限制，IPFS本身是以文件hash形式存储，也并没有增加查询多样性，结合IPFS分布式存储之后在存储上有优化，并且区块链数据查询效率方面由于需要寻址查找，效率没有提升，而且在此基础上增加缓存数据库，结合关系型和非关系的综合特性，提高区块链查询的优化效率。    图4 系统架构   1. 优化区块链共识算法   Hyperledger本身采用PBFT共识算法，PBFT是一种基于状态机复制的一致性算法，是解决分布式中数据一致性问题的传统方法。在PBFT共识算法中，首先由客户端调用主节点，利用主节点广播信息给所有的从节点，之后所有的节点将自己根据验证策略处理结果发送给客户端，客户端根据返回的结果判断是否达成数据的一致性。该算法的优点在于这是一个确定性的共识算法，共识速度快，吞吐量小；但是缺点在于容错率较低，在完全去信任的环境下不够足够安全。  由于博物馆藏品数据除了馆内流通，还有馆际间的数据共享，在馆际联盟链中，每个博物馆对应一个节点，由于默认的一致性算法，如果全部节点都保持相同状态数据库，其数据量很大，而且一旦众多博物馆加入馆际联盟链，链上节点数目增多，区块链网络负载变大，可能导致延迟，降低效率。  计划采用区块链共识节点分块，构建节点集群，改善的共识算法分块实现，各集群内部先达成共识，选出中心子节点，各个集群间的中心子节点再进行共识算法，实现全馆际的数据共享。  **2、研究目标与效果**  针对上述三项具体研究内容，对应的研究目标和效果如下：  （1）设计博物馆区块链数据结构，优化链上链下数据模型。  数据结构分为基于业务流程的数据和基于藏品信息的数据，根据对应的数据结构编写相应的智能合约，这样可以更好的利用区块链可溯源的特性，将重要的数据按不同方式存入区块链中，保证数据不可篡改。   1. 采用IPFS优化区块链数据存储   结合第一点的数据结构，将具体的藏品数据和业务流程数据存入IPFS文件中，将IPFS对应的相关hash存入区块链上，减少区块链存储量，提升区块查询效率。   1. 优化区块链共识算法   优化后 PBFT 分层分块共识，应用于各博物馆间藏品的数据交换共享，同一时间多个子节点集群的共识中心节点同时执行共识，这避免了节点统一通道集中广播，降低了区块链网络占用率，共识耗时时间缩短。   1. **拟解决的关键问题**   （1）设计博物馆区块链数据结构，优化链上链下数据模型。  初步构建的数据结构可能不够准确，在编写智能合约和实例化过程中需要深入研究各流程的关系，结合存储方式改善需要存储到区块链上和链下分布式数据库的数据。   1. 采用IPFS优化区块链数据存储   采用IPFS存储本身没有改变区块链的查询方式，区块链状态数据库仍然是采用键值对方式存储，查询方法单一，思考是否建立缓存数据库，利用关系型数据库方式进行查询，加快效率。   1. 优化区块链共识算法   优化共识算法采用的是构建多个集群，要考虑集群的容量和各集群的协调，研究如何分开才能更为有效的提高每个集群的平均数据交换效率。 |

|  |
| --- |
| **三、研究方案设计及可行性分析（包括：研究方法，技术路线，理论分析、计算、实验方法和步骤及其可行性等）（不少于800字）**  **1、研究方法**  本论文采用分阶段的方法，从调查研究，到梳理业务数据，到应用设计再到应用优化，最后进行性能测试。从理论基础到技术选型、从底层到应用层、从模拟到验证。重要模块的分析应该结合相关文献、已有系统和工程实际进行深度分析，必要时要经过相关实验和操作的验证。  具体的方法如下：  （1）调查研究，梳理业务逻辑  先了解博物馆的业务逻辑，了解区块链技术发展历程，了解Hyperledger平台的原理，分析博物馆哪些业务数据能够应用于区块链。  （2）应用设计  搭建Hyperledger平台，配置节点，选择合适的编程语言，编写智能合约。  选择合适的框架搭建应用，根据已有接口获取数据，同时选择合适的SDK对接平台，编写代码。  （3）应用优化  研究区块链的存储方式和共识算法，研究优化的合理性，测试优化后的效率。  **2、技术路线**  （1）首先要进行需求调研，调研博物馆的业务逻辑，然后形成需求分析报告。  （2）接着对区块链的原理和使用方式进行研究，需要注意总结和对比不同区块链技术之间的差别，要清楚不同区块链技术具体的原理和运作流程，并且先自己进行实际配置和运行，对其功能和性能进行评估。  （3）然后对目前采用的Hyperledger平台的架构进行研究，重点分析其可扩展性和提供的区块链存储结构，思考基于此平台搭建的架构设计。  （4）从底层开始进行层次模块设计，先搭建Hyperledger平台，根据已调研的业务逻辑，结合区块链选出合适的数据结构，编写对应智能合约。在平台基础上使用框架搭建应用，使用平台提供的接口设计功能模块，并且对重点模块进行模拟，验证可行性。对不可行的设计进行调整和重新设计。  （5）应用测试成功后进行Hyperledger平台数据存储和共识算法的优化。进行整合设计，进行集成验证。经过调整和优化达到要求后输出详细设计报告，接着进入实现阶段。  （6）在实现阶段首先要确保有可用的环境和实验资源，然后配置好所需的实验环境，包括基础的云环境和开发环境。  （7）在开发过程中需要进行单元测试和集成测试，及时记录发生的问题和遇到的瓶颈，开发完成之后将系统部署到云端，最后生成测试报告。    图5 技术路线图  **3、可行性分析**  本论文的工作核心是将区块链技术应用于博物馆数据中，利用区块链的分布式存储，不可篡改和去中心化等特性，以博物馆内业务部门为区块链节点组建馆内私有链，解决博物馆内各部门数据传输、共享的困难，对馆内事务进行穿透式管理。同时以多家博物馆的对外信息公开部门为节点组建馆际联盟链，对联盟链中共享的公开数据，进行数据请求、审批、交换过程的验证和记录，明确界定共享权责，实现馆际间的公开数据远距离可信可靠共享。区块链的平台的可扩展性已经足够成熟，博物馆数据也有提供相应的接口，这些功能都是基于事实现状且有很大的实用性，因此可行性较高。 |

|  |
| --- |
| **四、本研究课题可能的创新之处（不少于500字）**  1、区块链查询和查询方式的优化  目前采用的IPFS+区块链存储的方式，利用了分布式网络的特性，很好地与区块链本身相结合，IPFS本身是基于内容的寻址，只关心文件的内容，通过其生成唯一的哈希标识，节约了空间，与区块链相结合，在存储空间上有一定的优化，解决了区块链存储空间问题、重复验证问题和数据分层管理问题，有效节约了网络存储资源。另一方面，Hyperledger平台本身的状态数据库采用的LevelDB和CouchDB本身是键值对存储的，虽然CouchDB增加了复查询的功能，但是查询方法还是受到限制，IPFS本身是以文件hash形式存储，也并没有增加查询多样性，结合IPFS分布式存储之后在存储上有优化，并且区块链数据查询效率方面由于需要寻址查找，效率没有提升，而且在此基础上增加缓存数据库，结合关系型和非关系的综合特性，提高区块链查询的优化效率。   1. 区块链共识算法的优化   目前超级账本普遍采用的PBFT共识算法已经足够完善，该算法的优点在于这是一个确定性的共识算法，共识速度快，吞吐量小；但是缺点在于容错率较低，在完全去信任的环境下不够足够安全。而且对于大量节点的联盟链，查询速度会有所下降，本论文对共识算法的优化，将众多节点分集群调度，区块链全网参与共识节点的分层分块构建了多个参与区块链共识的多中心节点集群，各共识集群互不干扰、相辅相成，共同作用。优化后的 PBFT 实现了多中心子节点集群分块共识，能够解决传统 PBFT 算法中网络堵塞问题，减少了区块链网络广播资源浪费、降低了区块链共识通信成本，保证区块链共识安全的同时大大地提升了通信和共识效率。 |
| **五、研究基础与工作条件（1.与本项目相关的研究工作积累基础 2.包括已具备的实验条件，尚缺少的实验条件和拟解决途径）（不少于500字）**  1、与本项目相关的研究工作积累  （1）在项目开展前期阅读了区块链的资料，了解了基于区块链3.0的Hyperledger平台的工作原理。通过阅读文献和源代码了解了基于区块链的应用实例。  （2）通过阅读文献了解博物馆的数据和业务逻辑，了解了哪些数据方面可以结合区块链进行应用。阅读了项目文档，梳理了可用于区块链存储的博物馆数据结构。  （3）搭建了Hyperledger平台并进行网络测试，学习了智能合约的编写以及应用时的实例化和使用。  （4）学习了Spring系列框架的使用，学习了前端知识。  2、已具备的实验条件  本课题依托于智慧博物馆大数据模型与私有云架构关键技术研究项目，  已具备的条件：  （1）实验所需数据类型：依据“智慧国博”接口相关规范，提供了藏品管理系统与“智慧国博”项目中其他业务系统之间的接口。  （2）数据存储软件：超级账本区块链存储数据库CouchDB，IPFS数据库。  （3）算法运行平台：多台Windows/Linux操作系统PC机，Linux操作系统Docker容器。  3、尚缺少的实验条件和拟解决途径  尚缺少的实验条件：  （1）馆际联盟链的共享和交换的具体数据缺少实例 。  （2）目前项目提供了藏品管理系统与“智慧国博”项目中其他业务系统之间的接口，还未提供具体的数据。  拟解决途径：  （1）从馆内数据类型提取出馆际联盟链的数据共享类型，  （2）先按照接口规范编写相应数据结构，拟定具体数据进行测试。 |

**学位论文工作计划**

|  |  |  |
| --- | --- | --- |
| 时间 | 研究内容 | 预期效果 |
| 2020.09-2020.11 | 熟悉项目需求，了解相关技术，整理需求，准备开题 | 撰写开题报告，完成开题答辩，整理需求分析 |
| 2020.12-2021.05 | 前端系统及后台管理平台的原型设计，实现基本溯源功能，做为进一步研究的依据 | 完成前端系统及后台管理平台的原型。基本实现博物馆的数据存储功能 |
| 2021.5-2021.5 | 系统主要模块单元测试，以及模块功能的完善和代码优化。 | 系统主要模块单元测试，以及模块功能的完善和代码优化。 |
| 2021.9-2021.9 | 撰写阶段报告 | 完成阶段报告 |
| 2021.10-2021.12 | 撰写毕业论文 | 提交初稿、修改稿、定稿 |
| 2022.1-2022.5 | 论文送审、修改及答辩 | 论文送审、修改及答辩 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 评  定  小  组  成  员 | 姓 名 | 职 称 | 单位名称 | 职务 |
| 雷友询 | 副教授 | 北京邮电大学 | 组长 |
| 崔毅东 | 副教授 | 北京邮电大学 | 成员 |
| 漆涛 | 教授 | 北京邮电大学 | 成员 |
| 杨谈 | 副教授 | 北京邮电大学 | 成员 |
|  |  |  |  |
|  |  |  |  |
| 导师意见： | | | | |
|  | | | | |
| 导师（签名）：  日期： 年 月 日 | | | | |
| 开题报告小组意见： | | | | |
| 组长（签名）：  日期： 年 月 日 | | | | |
| 学院意见（签章）： | | | | |
| 负责人：  日期： 年 月 日 | | | | |