

团 体 标 准

T/SAC 004—2024

证券业区块链电子数据存证应用规范

Specification for blockchain based electronic evidence preservation
application in securities industry

2024-07-29 发布

2024-07-29 实施

中国 证 券 业 协 会 发 布

目 录

目 录 I

前 言 II

引 言 III

证券业区块链电子数据存证应用规范 1

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 存证应用区块链技术要求 4

4.1 账本技术 4

4.2 共识协议 5

4.3 智能合约 5

4.4 密码算法 6

5 区块链存证平台 6

5.1 区块链存证平台系统框架 6

5.2 存证业务系统 7

5.3 区块链存证系统 7

5.4 存证证明机构系统 7

6 区块链存证过程 8

6.1 区块链存证平台应用框架 8

6.2 应用原则 8

6.3 参与方 8

6.4 关键过程 9

7 区块链存证模型 10

7.1 存证数据模型定义 10

7.2 业务数据提取 10

7.3 存证数据传输 11

7.4 存证数据隐私保护 11

附录 A （资料性） 单点数据存证模型应用示例 12

附录 B （资料性） 链式数据存证模型应用示例 6

附录 C （资料性） 流式数据存证模型应用示例 19

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国证券业协会归口。

本文件起草单位：国泰君安证券股份有限公司、海通证券股份有限公司、东吴证券股份有限公司、兴业证券股份有限公司。

本文件主要起草人：杨晨旭、詹婷婷、任荣、唐淑艳、石良生、吴鑫涛、应原、杨琦路、张子华、付益明。

引 言

随着金融科技快速发展，区块链技术在金融行业的应用场景逐渐增多，其中区块链存证技术越来越成熟并被行业机构应用在合同签署、版权保护等领域。然而目前证券行业缺乏统一的规范性指导文件，相关标准内容存在一定的局限性且缺乏对行业的指导意义，存证业务面临着数据格式不一致、技术要求不统一、存证流程不完整等问题。鉴于此，进行行业标准化研究，夯实区块链技术标准化发展基础，提升区块链技术标准化研究水平，推进证券行业区块链技术应用标准体系建立至关重要。通过制定证券行业团体标准对存证流程加以规范，对业务场景统一标准，以建设规范的区块链存证应用环境，从而推动区块链存证业务在行业的广泛应用，为加速推进行业数字化高质量转型、有效引领行业金融科技高速发展、有力保障金融科技在行业内的行稳致远贡献力量。

证券业区块链电子数据存证应用规范

1 范围

本文件规定了基于区块链技术的电子数据存证应用的术语和定义、存证应用区块链技术要求、区块链存证平台、区块链存证过程、区块链存证模型。

本文件适用于证券业基于区块链技术的电子数据存证应用。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件。不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 20520—2006 信息技术安全 公钥基础设施 时间戳规范

GB/T 22239—2019 信息安全技术 网络安全等级保护基本要求

GB/T 25069—2022 信息安全技术 术语

GB/T 35273—2020 信息安全技术 个人信息安全规范

GB/T 43572—2023 区块链和分布式记账技术 术语

GB/T 43580—2023 区块链和分布式记账技术存证通用服务指南

GM/T 0024—2014 SSL VPN技术规范

GM/T 0111—2021 区块链密码应用技术要求

JR/T 0184—2020 金融分布式账本技术安全规范

SF/Z JD0400001—2014 电子数据司法鉴定通用实施规范

3 术语和定义

GB/T 20520—2006、GB/T 25069—2022、GB/T 43572—2023、GB/T 43580—2023、JR/T 0184—2020、SF/Z JD0400001—2014 界定的以及下列术语和定义适用于本文件。

3.1

区块链 blockchain

使用密码链接将共识确认的区块按顺序追加形成的分布式账本。

[来源：GB/T 43572—2023，3.6]

3.2

联盟链 consortium blockchain

由一组利益相关的参与者管理和使用，仅有授权节点可接入，接入节点可按规则参与共识和读写数据的区块链部署模型。

3.3

电子数据 digital evidence

以电子手段生成、发送、接收或存储的信息。

[来源：GB/T 43580—2023，3.2]

3.4

存证 preserve evidence

通过技术手段对电子数据进行保存和验证的行为，保证其完整性和真实性并可追溯。

[来源：GB/T 43580—2023, 3.3]

3.5

区块链存证 blockchain preservation

基于区块链技术实现多节点共识的电子数据存证。

[来源：GB/T 43580—2023, 3.4]

3.6

电子数据存证平台 digital evidence preservation platform

可实现电子数据存证相关功能的平台或系统。

3.7

数字签名 digital signature

附加在数据单元上的一些数据，或是对数据单元做密码变换，这种附加数据或密码变换被数据单元的接收者用以确认数据单元的来源和完整性，达到保护数据，防止被人（例如接收者）伪造的目的。

[来源：GB/T 25069—2022, 3.576]

3.8

散列值 hash value

为了标识电子数据的唯一性或完整性而通过特定的散列算法把任意长度的输入数据变换成的固定长度的输出值。散列值又称哈希值或校验码，常用散列算法包括 SM3、MD5 和 SHA256 等。

[来源：SF/Z JD0400001—2014, 3.8]

3.9

共识机制 consensus mechanism

使节点见达成共识的规则和程序。

[来源：GB/T 43572—2023, 3.12]

3.10

默克尔树 merkle tree

区块链的重要数据结构，其本质由节点组成的二叉树，其中每个叶节点代表一段数据，每个非叶（内部）节点代表其子节点的加密哈希，其作用是快速归纳和校验区块数据的存在性和完整性。

3.11

分布式账本技术 distributed ledger technology

实现分布式账本的技术的集合，是密码算法、共识机制、点对点通讯协议、分布式存储等多种核心技术体系高度融合形成的一种分布式基础架构与计算范式。

[来源：JR/T 0184-2020, 3.19]

3.12

智能合约 smart contract

存储在分布式记账技术系统中的计算机程序，该程序的任何执行结果都记录在分布式账本中。

[来源：GB/T 43572—2023, 3.72]

3.13

加密 encipherment encryption

对数据进行密码变换以产生密文的过程。

[来源：GB/T 25069—2022, 3.278]

3.14**可信时间 trusted time**

准确的、值得信赖的当前时间值，这个时间值的来源应是高度权威的。

[来源：GB/T 20520—2006，3.2]

3.15**时间戳 time stamp**

对时间和其他待签名数据进行签名得到的，用于表明数据时间属性的数据。

[来源：GB/T 25069—2022，3.541]

3.16**去标识化 de-identification**

通过对个人信息的技术处理，使其在不借助额外信息的情况下，无法识别或者关联个人信息主体的过程。

[来源：GB/T 35273—2020，3.15]

3.17**匿名化 anonymization**

通过对个人信息的技术处理，使得个人信息主体无法被识别或者关联，且处理后的信息不能被复原的过程。

[来源：GB/T 35273—2020，3.14]

3.18**数据完整性 data integrity**

数据所具有的特性，即无论数据形式作何变化，数据的准确性和一致性均保持不变。

[来源：GB/T 25069—2022，3.574]

3.19**保密性 confidentiality**

信息对未授权的个人、实体或过程不可用或不泄露的性质。

[来源：GB/T 25069—2022，3.41]

3.20**机密性 confidentiality**

采用密码技术保证信息不泄露的性质。

[来源：GB/T 25069—2022，3.259]

3.21**单点数据 single-point data**

静态、相对恒定的数据序列。

3.22**链式数据 chain data**

一组存在逻辑、勾稽关系的数据序列。

3.23**流式数据 streaming data**

一组顺序、大量、快速、连续到达的数据序列。一般情况下，流数据可被视为一个随时间延续而无限增长的动态数据集合。

3.24**业务数据 business data**

存证业务系统运行产生的数据。

3.25

存证数据 preservation data

写入区块链的数据，它可以是一个字符串、一个文件的散列值等形式，用于表示文本、文件等电子数据，确保这些数据在区块链上得到安全存储和验证，以作为有效的法律或商业证据。

4 存证应用区块链技术要求

4.1 账本技术

4.1.1 账本结构

- a) 账本结构由区块头和区块体组成。
- b) 账本结构中区块头应包括上一个区块 ID、版本号、时间戳、难度、默克尔树根等信息。
- c) 账本结构中区块体应包括所有交易信息以及所有交易信息的默克尔树信息。
- d) 账本结构应具备数据防篡改性以及校验数据完整性的功能。

4.1.2 账本数据管理

- a) 账本数据应能正确查询到当前状态信息。
- b) 账本数据应能正确查询到历史更新记录及交易详情。
- c) 账本数据应能对数据来源和变更的操作者身份进行追溯。
- d) 账本数据应能正确查询到状态变更记录所在区块文件的唯一标识和时间戳。
- e) 账本数据应保证节点从异常状态恢复后，仍能够正确完成历史数据溯源。

4.1.3 数据同步

- a) 账本数据同步应保证需同步数据的节点与源节点的数据一致。
- b) 账本数据同步应保证新增节点在同步所需历史数据之后与其他节点数据一致。
- c) 账本数据同步应保证单节点同步增量数据后与其他节点的数据一致，且支持动态切换源数据节点。
- d) 账本数据同步应确保在多节点网络中，节点同步历史数据的一致性和时效性，避免同步错乱。
- e) 账本数据同步应确保账本状态数据可重建，可根据交易序列重建账本状态数据。
- f) 账本数据同步应具备断电、重启、网络波动等异常情况下的断点续传能力。
- g) 账本数据同步应确保节点可通过数据同步纠正本节点的数据异常问题，保证整个网络数据的一致性。

4.1.4 数据归档

- a) 账本应提供数据归档功能，包括但不限于磁盘文件数据、数据库存储数据，存储介质应具备高可靠性。
- b) 归档数据与节点本地存储的数据应完整，不缺失账本数据。
- c) 不同节点归档的数据应存放在不同的存储设备中，防止出现集中数据丢失。
- d) 数据归档应支持用户设置归档数据范围，包括但不限于根据时间、区块序号等范围限定。
- e) 节点在数据归档过程中，应能够继续提供系统服务，可暂时停止历史状态查询功能。
- f) 节点数据归档完成后，应支持对归档数据进行历史状态查询。
- g) 节点从断电、重启、网络波动等异常场景恢复后，应可继续完成数据归档操作。
- h) 数据归档应支持归档数据恢复还原。

4.1.5 数据扩容

- a) 账本应提供存储的扩容技术和方案。
- b) 扩容前，应做好数据归档操作，确保异常情况下的数据可恢复性。

- c) 扩容过程中，应确保用户对账本的操作不会导致节点账本数据文件异常，扩容方案应具备平滑伸缩能力，保障在线系统不必中断运行。
- d) 扩容完成后，节点应能够正常运行、性能稳定，应保证扩容前后数据一致。

4.1.6 数据跨链功能

- a) 账本应建立跨链交易机制，可采用公证人机制、侧链/中继、哈希锁定、分布式私钥控制等实现方式。
- b) 跨链交易过程应保证状态数据修改可正确更新到账本中。
- c) 跨链交易应满足事务完整性、一致性。
- d) 跨链功能应支持基于当前链的账本状态单独验证跨链交易。
- e) 节点从断电、重启、网络波动等异常场景恢复后，应可继续完成被中断的跨链交易。

4.2 共识协议

4.2.1 共识算法

- a) 系统基于声明的所有不同的共识算法，均能实现所有节点达成共识，系统正常运行。
- b) 联盟链可选用的共识机制包括基于投票的共识机制、基于公信力节点的共识机制等。
- c) 共识算法应限定每一次共识出块的最长时间，保证既定的共识出块频率，确保共识机制的可用性。
- d) 共识机制应支持在节点在线、节点离线、网络规模调整等情况下，可切换共识机制能达成全网新的共识。

4.2.2 共识协议一致性

- a) 共识协议应保证每个节点更新账本状态写操作的一致性。
- b) 共识协议应保证每个节点账本的最终状态一致性。
- c) 共识协议应保证节点从断电、重启、网络波动等异常场景恢复后，所有节点交易执行的结果应保持一致。
- d) 共识协议应保证同样的交易在同一或不同节点多次运行，共识结果一致。

4.2.3 共识协议可靠性

- a) 共识协议的安全性要求应符合 GM/T 0111—2021 中第 7.7 款规定。
- b) 共识协议应具备可靠的身份鉴别机制，对可参与共识的节点进行严格准入。
- c) 共识协议应具备一定的抗攻击和识别恶意节点的能力。
- d) 共识协议应保证篡改节点数与故障节点数之和小于最大阈值时可正常达成全网共识。
- e) 共识协议应保证节点从异常场景恢复后数据正常恢复、数据不丢失及正常参与共识流程。

4.2.4 共识协议可拓展性

- a) 共识协议支持动态或静态增加节点。
- b) 共识协议支持动态或静态删除节点。

4.3 智能合约

4.3.1 智能合约正确性

智能合约应保证对其的执行调用能获得与参数输入相对应的正确结果。

4.3.2 智能合约一致性

- a) 智能合约应保证在参数、环境等外部要素一致情况下，在各节点上的执行结果相同。
- b) 智能合约执行时，各节点数据应互不干扰。

4.3.3 智能合约可靠性

- a) 智能合约的安全性要求应符合 GM/T 0111—2021 中第 7.8 款规定。
- b) 智能合约应保证执行的原子性，调用智能合约执行异常操作能够正常回滚，调用智能合约执行有效操作，各节点执行结果准确一致。
- c) 智能合约应保证可向下兼容，在系统环境升级迭代后仍可正常运行。
- d) 智能合约应支持冻结、解冻、销毁、升级和版本管理，以保证智能合约的全生命周期可控。

4.3.4 智能合约隔离性

- a) 智能合约应支持业务数据存储按需隔离，不同智能合约的业务数据传输应互不干扰。
- b) 智能合约应确保没有业务依赖的不同智能合约可并行执行互不干扰。
- c) 智能合约的使用应进行严格授权管理，确保持续管理员或创建者对合约的控制权，参与成员仅可在授权范围内使用和执行智能合约逻辑。

4.3.5 智能合约安全性

智能合约应保证可安全执行，避免因安全漏洞、外部攻击导致的资金损失、数据泄露和合约功能瘫痪等异常情况。

4.4 密码算法

4.4.1 算法基本条件

- a) 算法的采用应符合 GM/T 0111—2021 中第 7.1 款规定。
- b) 算法应符合预期输出并满足正确性和一致性，并能支撑系统正常服务。

4.4.2 算法特性

算法应保障敏感信息的保密性、完整性，应保障实体的真实性。

4.4.3 密钥管理

密钥的使用应符合 GM/T 0111—2021 中第 7.4 款规定。

5 区块链存证平台

5.1 区块链存证平台系统框架

区块链存证平台系统框架包含存证业务系统、区块链存证系统、存证证明机构系统，见图 1：

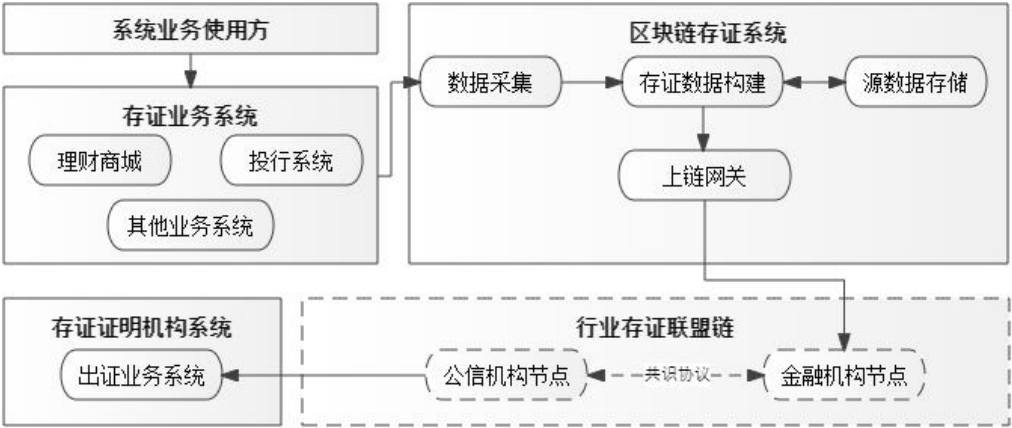


图 1 区块链存证平台系统框架

5.2 存证业务系统

存证业务系统是具有实际存证业务需求的应用系统，包括但不限于财富管理业务系统、投行业务系统、研究所业务系统等，应具有一定的开放能力，可以对接到区块链存证系统进行数据传输上链。

存证业务系统业务逻辑应完备，业务流程应设计完整。

存证业务系统应给客户揭示相关业务数据散列值将提供区块链存证系统和存证证明机构用于存证用途进行使用。

存证业务系统应具备完善的鉴权机制，同时具备清晰、易用、周全的权限控制能力。

存证业务系统应具备身份鉴别机制，以保障存证操作行为与用户的强关联性。

存证业务系统应完整记录用户的操作日志，日志格式应规范完整、日志存储应进行备份。

存证业务系统应具备保障存证数据完整性、机密性的技术机制。可使用哈希校验、电子签名、加密等技术防止存证数据被篡改，确保存证数据在存储、传输过程中的安全性。

存证业务系统应具备一定的容错性，在存证异常产生时，能及时重试以提高存证成功率。

存证业务系统应具备高可用性，通过同城双活、异地容灾等机制保障业务连续性，建立防御防攻击系统，防御大流量 DDoS 等攻击。应具备完善的应急预案，定期进行应急预案演练。

存证业务系统的密钥管理、证书管理方面的安全要求应符合 GM/T 0111—2021 中第 7.4、7.5 款规定。

5.3 区块链存证系统

区块链存证系统由数据采集模块、数据构建模块、数据存储模块以及对接区块链的上链网关组成，实现过程可根据实际需要构建相应的功能模块，支持扩展。其中，数据采集模块负责从各存证业务系统采集需存证的业务数据；数据构建模块负责基于各业务定义的存证模型构建存证数据；数据存储模块负责存储、归档存证数据；上链网关负责对接外部存证联盟链，对内提供上链存证服务。

区块链存证系统应保证系统安全，应达到 GB/T 22239—2019 的第三级基本要求。

区块链存证系统应保证运行环境安全，应 7*24 稳定运行，应具备完善的监控体系，应具有完整的安全防护措施及应急预案。

区块链存证系统应保证存储安全，应具备冗余备份、存储扩展、异地容灾的能力，确保用户隐私数据安全。

区块链存证系统应保证通信网络安全，应具备防御网络攻击、防病毒的能力，系统应定期检查排除风险。

区块链存证系统应保证数据安全，相关密码、密钥等技术的应用应符合 GM/T 0111—2021 相关规定。

区块链存证系统应可支持国产化、自主可控等国家信创要求。

区块链存证系统应具备跨链对接能力，如适配不同区块链的数据模型转换能力、匹配不同区块链的共识机制兼容能力、开放规范的通用 API 能力等。

区块链存证系统应具备一定的并发性能，以保障正常的业务需求。避免使用过度耗时的算法、性能较差的服务器建设区块链存证系统。

5.4 存证证明机构系统

存证证明机构系统由是中立、具有公信力的机构构建，如法院、公证机构、仲裁机构、审计机构、司法鉴定中心等，可对存证的电子数据出具公证鉴定报告，公证鉴定报告可按不同出证场景、不同出证对象制定相应模板，由各存证证明机构统一维护并遵循。

存证证明机构节点的存证证明源数据与其他节点应保持实时同步,可采用网络专线等方式保障通信。

存证证明机构节点与其他节点的通信通道应做加密处理,节点对外提供的API服务应做加密处理,采用的SSL协议及加密算法应符合GM/T 0024—2014 SSL、GM/T 0111—2021相关规定。

6 区块链存证过程

6.1 区块链存证平台应用框架

区块链存证平台应用框架包含应用原则、相关参与方和关键过程,见图2:

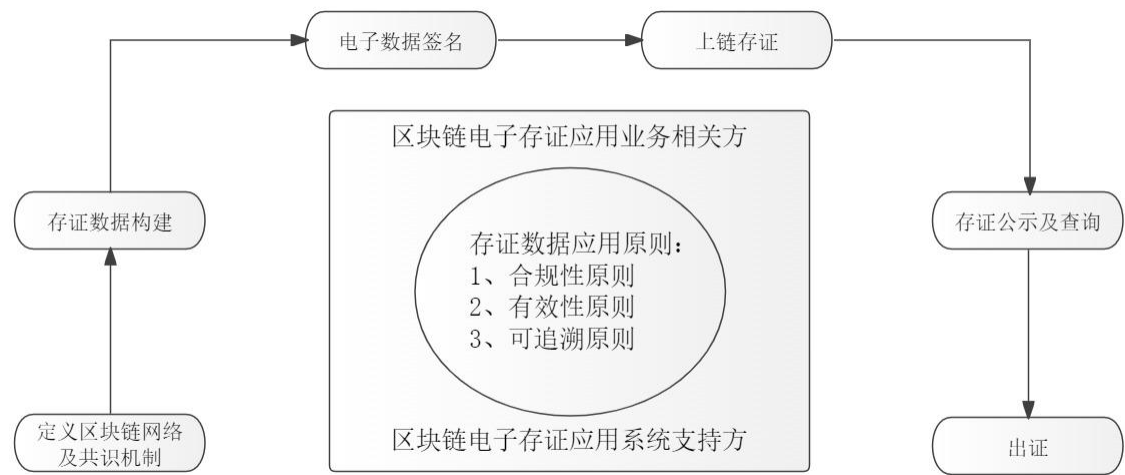


图 2 区块链存证平台应用框架

6.2 应用原则

区块链存证过程应用原则包括合规性原则、有效性原则、可追溯原则。

合规性原则：应保证存证行为合规、存证信息保密合规、存证信息披露合规、取证流程合规。

有效性原则：应保证业务系统数据有效、电子存证数据存取有效、可信时间标识有效、第三方鉴定机构有效、存证核验有效。

可追溯原则：应保证存证数据不可篡改可溯源、存证行为过程可溯源、出证链路可溯源。

6.3 参与方

区块链存证过程参与方分为区块链电子存证应用业务相关方和区块链电子存证应用系统支持方。

区块链电子存证应用业务相关方提供存证场景、存证数据。

区块链电子存证应用系统支持方提供存证服务、存证合法性校验服务、存证查询核验服务、出证服务。

6.4 关键过程

6.4.1 定义区块链网络及共识机制

在建立区块链存证平台时，可按照存证场景及需求目标定义或选择区块链网络的共识机制。适用于联盟链的共识机制包括基于投票的共识机制、基于公信力节点的共识机制等。具体如下：

- a) 基于投票的共识机制：基于区块链节点资源量化等特定算法来确定链上可记账节点的共识机制；
- b) 基于公信力节点的共识机制：由区块链上单个或多个具备公信力的节点进行鉴证，通过该节点执行强制校验的共识机制。

6.4.2 存证数据构建

存证数据构建过程如下：

- a) 存证平台根据业务场景按需预先定义存证模板，根据存证模板规定各业务场景下的存证数据字段，也可定义适用不同业务场景的通用业务模板；
- b) 应对存证模板建立合理的版本管理规范，建立不同版本与存证数据、业务数据的关联关系，模板应具备普适性、时效性和扩展性；
- c) 存证业务系统根据存证模板提取构建存证数据，存证数据应在保证存证有效性的前提下进行结构化及去标识化处理；
- d) 存证数据应满足国家法律法规要求，隐私数据应进行匿名化处理；
- e) 存证数据应具备完整性、真实性、有效性；
- f) 存证数据大小应有最大单元限制，超限时应在保证数据安全、不失真的前提下进行合理的分片；
- g) 针对流式数据，按预设的标准对数据流进行切片，切片的大小决定了流式数据存证的颗粒度。切片的标准包含但不限于固定的时间、固定的大小等。并应针对切片建立索引。详细过程参考可参见附录 C。

6.4.3 电子数据签名

区块链存证平台对存证内容进行电子数据签名时：

- a) 应使用合法授权的身份数字证书进行签名，证书私钥本地保存，公钥及证书信息公开；
- b) 证书私钥应具有有效期，过期无效；
- c) 应使用存证主体区块链节点私钥对数据进行数字签名后发送共识；
- d) 已签名数据可在区块链网络内各节点进行合法性验证。

6.4.4 上链存证

上链存证过程如下：

- a) 存证业务系统将存证数据提交上链节点进行上链存证，可根据实际需要灵活使用实时上链、异步上链、定时批量上链等上链机制，可使用分布式事务、补偿事务等方式避免数据单边问题；
- b) 应确保存证过程中的数据提取、构建、存储、传输所依赖的软硬件环境安全、可靠；
- c) 数据上链可根据不同场景要求采用不同的存证信息生成算法，可包括原文存证、Hash 存证、默克尔树存证（可参考附录 C）等；
- d) 区块链共识可采用白名单机制，只允许白名单内的节点进行存证数据传输共识，传输信道可靠；
- e) 存证电子数据存储于区块链节点数据库中，每个共识节点均保留有区块链账本的完整备份，数据应定期共识同步避免异常或状态单边；
- f) 针对流式数据，将数据的切片内容、位置以及索引信息同时上链存证。详细过程可以参考附录 C。

6.4.5 存证公示及查询

区块链存证数据应支持所有区块链网络参与者查询本机构节点的存证公示信息：

- a) 业务层存证结果查询：存证平台区块链浏览器可采用网站形式构建，参与者可登录区块链浏览

器查询链上的存证数据散列值以及相对应的区块信息等；

- b) 系统层存证结果查询：存证平台应提供存证查询 API 接口，区块链网络内各节点可采用接口调用的方式查询获取电子存证数据结果；
- c) 节点层存证结果查询：每个区块链节点均提供服务查询本节点的存证数据散列值，区块链浏览器以及存证业务系统对接均可指定某一特定区块链节点进行存证数据散列值以及存证上链结果的查询；
- d) 针对流式数据，查询时应利用分片索引优化查询效率，并保证流式数据的完整性以及连续性。详细过程可以参考附录 C。

6.4.6 出证

区块链存证平台应提供出证功能，相应的公证鉴定报告由链上存证证明机构出具。

区块链网络参与方可通过存证平台区块链浏览器网站、API 调用等方式进行出证申请获取存证数据，存证数据直接来源于区块链共识账本。

出证过程应确保数据提取、构建、存储、传输所依赖的软硬件环境安全、可靠。

出证数据提取过程应可重现，提取过程的记录是连续的：同一电子存证数据支持无限次出证提取，区块链的不可篡改特性保证每次提取的内容相同。每次出证提取行为，各区块链节点的行为数据区块链存证平台均会记录留痕。

出证数据应包括：链上存证信息、存证信息加密算法、存证各参与方身份信息、存证可信时间戳、存证节点网络信息、出证结论、其他必要信息。

7 区块链存证模型

7.1 存证数据模型定义

存证数据模型定义应具备根据不同业务类型定义不同存证数据模板的能力并支持灵活扩展。

存证数据模型定义可建立通用的标准化模型，提供一组标准的业务数据模型的定义，模型应具有通用公共信息字段，可根据实际业务需求，参考使用本文附录中的存证数据模型示例。

存证数据模型定义应提供符合跨链技术标准的数据模型，包含对业务元素的存储模型规范，包括单点数据模型、链式数据模型、流式数据模型。

单点数据模型：单点数据是指静态、相对恒定的数据集，基于单点数据构建存证标准化数据模型的场景包括但不限于静态文本、静态图片、投行底稿文件等。本文定义的单点数据存证标准化模型参考附录 A。

链式数据模型：链式数据是指一组存在逻辑、勾稽关系的数据集，基于链式数据构建存证标准化数据模型的场景包括但不限于理财产品购买、投顾服务签约等。本文定义的链式数据存证标准化模型参考附录 B。

流式数据模型：流式数据是指一组顺序、大量、快速、连续到达的数据集，基于流式数据构建存证标准化数据模型的场景包括但不限于交易日志、委托记录等。本文定义的流式数据存证标准化模型参考附录 C。

7.2 业务数据提取

业务数据提取过程对于标准化模型的通用公共信息字段宜应填尽填。

业务数据提取手段应安全可靠，可采用加密、备份冗余、数字签名、时间戳、消息摘要等技术手段对业务数据第一时间进行固化，保证业务数据的保密性、完整性和真实性，确保业务数据提取后不被篡

改和破坏。

业务数据提取应保证数据的可管理性、可查询性、可追溯性和可还原性，应制定详细的存储规范明确数据的存储位置、存储格式、存储期限等关键要素，可采用分布式存储、云存储等多种存储方式确保数据的可靠和安全。

7.3 存证数据传输

存证数据传输应保证存证数据的安全性，可通过加密的方式保证存证数据的机密性，可通过数字证书私钥签名保证存证数据的完整性。

存证数据传输应保证传输信道的安全性，可通过使用符合 GM/T 0111—2021 相关规定的加密算法加密通信信道，来保证存证数据的安全传输。

7.4 存证数据隐私保护

存证数据隐私保护应符合 GB/T 35273—2020 相关规定。

存证数据构建前，应对存证原始数据中的个人敏感信息进行去标识化处理。

存证数据和业务系统中的客户信息等数据应分开存储、隔离保存。

存证数据的访问及提取应具备完善的鉴权机制，同时具备清晰、易用、周全的权限控制能力。

存证数据上链以及传输过程均应采用加密的方式进行交互，可通过使用符合 GM/T 0111—2021 相关规定的算法进行数据摘要信息提取、数据加解密、数字签名以及加密通信信道，保证数据安全传输。

附录 A
(资料性)
单点数据存证模型应用示例

对于区块链存证在证券行业内的应用场景，各机构可根据实际业务、技术需求灵活选择数据模型。单点数据存证模型示例以投行底稿文件业务的存证模型定义进行说明。

投行底稿文件存证模型包含存证主体信息、底稿数据信息两个部分。属性分类信息如表 A.1 所示：

表 A.1 投行底稿文件存证模型表

一级分类	二级分类
存证主体信息	主体基本信息
	主体区块链节点信息
底稿数据信息	项目信息
	底稿信息

a) 存证主体信息

存证主体用于记录行业区块链存证平台各参与方，主要包括行业联盟链上各类有存证需求的机构。存证主体信息划分为一到二级属性分类，属性分类应符合表 A.2 的内容。二级子类包括存证主体的基本信息、区块链节点信息，基于该主体的实际业务填写必要或非必要的所有源数据，如：企业名称、机构类型、行政区域、机构性质等。

表 A.2 存证主体信息表

一级分类	二级分类
存证主体信息	主体基本信息
	主体区块链节点信息

表 A.2 中具体信息的说明如下：

主体基本信息见表 A.3：

表 A.3 主体基本信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填 (Y)	枚举值说明
主体名称	subject_company_name		CHARACTER	256	Y	
主体英文名称	subject_company_english_name		CHARACTER	256		
主体行政区域	subject_main_administrative_region		NUMBER		Y	枚举-境内外类别
主体信息维护时间	subject_create_time	业务操作时间	TIME		Y	
机构性质	subject_organization_nature		NUMBER			枚举-机构性质
所有制性质	subject_ownership_nature		NUMBER			枚举-所有制性质
证件类型及证件号码	subject_document_information		对象数组			枚举-机构代码类型

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填 (Y)	枚举值说明
注册日期	subject_registry_date		DATE			
主营业务	subject_company_business		CHARACTER	2048		
公司简介	subject_company_profile		TEXT			
注册资本	subject_registered_capital		DECIMAL			
注册资本币种	subject_registered_capital_currency		CHARACTER	16		枚举-币种类型
实收资本	subject_paid_in_capital		DECIMAL			
实收资本币种	subject_paid_in_capital_currency		CHARACTER	16		枚举-币种类型
注册地址	subject_registered_address		CHARACTER	512		
办公地址	subject_office_address		CHARACTER	512		
联系地址	subject_contact_address		CHARACTER	128		
联系电话	subject_contact_number		CHARACTER	512		
企业传真	subject_fax		CHARACTER	64		
邮政编码	subject_postal_code		CHARACTER	32		
互联网地址	subject_internet_address	企业 官网 地址	CHARACTER	256		
电子邮箱	subject_mail_box		CHARACTER	64		
主管单位	subject_regulator		CHARACTER	128		

主体区块链节点信息见表 A. 4:

表 A. 4 区块链节点信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填 (Y)	枚举值说明
节点所属主体 ID	node_subject_id		CHARACTER			
节点名称	node_name		CHARACTER			
节点英文名称	node_english_name		CHARACTER			

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填 (Y)	枚举值说明
节点类型	node_type		NUMBER			枚举-节点类型
节点接入时间	node_access_time		TIME			
节点接入方式	node_access_method		CHARACTER			
节点接入网络	node_access_network		CHARACTER			
节点机器数量	node_machine_numbers		CHARACTER			
节点机器类型	node_machine_type		NUMBER			枚举-机器类型
节点 IP	node_ip		CHARACTER			
节点物理位置	node_position		CHARACTER			
节点系统版本	node_system_version		CHARACTER			
节点备注信息	node_remark		CHARACTER			

b) 底稿数据信息

底稿数据信息用于记录各参与主体在行业联盟链上进行区块链存证的底稿数据信息。底稿数据信息划分为一到二级属性分类,属性分类应符合表 A. 5 的内容。其中项目信息用于记录存证的投行项目信息,如项目编号、项目名称、项目类型、项目状态、项目发行人信息、项目当前阶段、项目负责人、项目行业、项目主办部门等。底稿信息用于记录存证的底稿文件信息,如底稿文件名称、文件大小、文件上传时间、底稿文件散列值等。

表 A. 5 底稿数据信息表

一级分类	二级分类
底稿数据信息	项目信息
	底稿信息

表 A. 5 中具体信息的说明如下:

二级分类项目信息见表 A. 6:

表 A. 6 项目信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填 (Y)	枚举值说明
项目编号	project_code		CHARACTER		Y	
项目名称	project_name		CHARACTER		Y	
项目类型	project_type		CHARACTER			
项目状态	project_status		CHARACTER			
项目发行人信息	project_issuer_infor mation		CHARACTER			

项目当前阶段	project_current_stat e		CHARACTER			
项目负责人	project_pr		CHARACTER			
项目行业	project_industry		CHARACTER			
项目主办部门	project_host_departm ent		CHARACTER			
项目描述	project_description		CHARACTER			

二级分类底稿信息见表 A. 7:

表 A. 7 底稿信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填 (Y)	枚举值说明
底稿所属项目目标号	working_papers_project_c ode		CHARACTER		Y	
底稿所属项目名称	working_papers_project_n ame		CHARACTER		Y	
底稿文件名	working_papers_name		CHARACTER		Y	
底稿文件描述	working_papers_descripti on		CHARACTER			
底稿文件大小	working_papers_size		CHARACTER			
底稿文件上传时间	working_papers_upload_ti me		TIME			
底稿文件散列值	working_papers_hash		CHARACTER		Y	

c) 上链数据示例

```
{
  "transaction_info_array": [{
    "transaction_info": {
      "transaction_detail_info": {
        "biz_data": {
          "bizName": "投行底稿文件存证",
          "businessData": [{
            "StageHash":
"60c8847b8a2e3f8e71420d1c7756d943132e174f80674683fb70bc86c4362baa",
            "StageName": "存证主体-主体基本信息"
          }, {
            "StageHash":
"52c8847a842e369efe656b00d128505501f9418ea9641cc5c0308c8c4361d1fb",
            "StageName": "存证主体-主体区块链节点信息"
          }, {
```

```
        "StageHash":
"46e1eb2d8785badc69fe850386f02e174f80674683fb70bc9f8f1cd1d1fb6658",
        "StageName": "项目信息"
    }, {
        "StageHash":
"af8205ffbbcc5c0308c8af8c1fc3a90b0674e9efda69efe656b00d128505501f",
        "StageName": "底稿信息"
    }],
    "necessaryData": {
        "CreateTime": "2022-08-12 12:01:22",
        "PackageHash":
"588750387566580de8ba6e6a19bccfdf5cb4bfad7dadfef6bc1aef4ac1346866",
        "UploadTime": "2022-08-12 13:02:42"
    }
}
}
```

附录 B
(资料性)
链式数据存证模型应用示例

对于区块链存证在证券行业内的应用场景，各机构可根据实际业务、技术需求灵活选择数据模型。链式数据存证模型示例以公募基金委托下单业务的存证模型定义进行说明。

公募基金委托下单业务存证模型包含存证主体、存证数据两个部分，其中存证数据包含实际业务过程中的账号、产品以及事务环节信息。属性分类信息如表 B.1 所示：

表 B.1 公募基金委托下单业务存证模型表

一级分类	二级分类	三级分类
存证主体信息	主体基本信息	--
	主体区块链节点信息	--
存证数据信息	客户账号信息	--
	产品信息	--
	事务环节信息	登录信息
		浏览记录信息
		适当性匹配信息
		协议签署信息
		委托下单信息

a) 存证主体信息

存证主体用于记录行业区块链存证平台各参与方，主要包括行业联盟链上各类有存证需求的机构。存证主体信息划分为一到二级属性分类，属性分类应符合表 B.2 的内容。二级子类包括存证主体的基本信息、区块链节点信息，基于该主体的实际业务填写必要或非必要的所有源数据，如：企业名称、机构类型、行政区域、机构性质等。

表 B.2 存证主体信息表

一级分类	二级分类
存证主体信息	主体基本信息
	主体区块链节点信息

表 B.2 中具体信息的说明如下：

主体基本信息见表 B.3：

表 B.3 主体基本信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
主体名称	subject_company_name		CHARACTER	256	Y	
主体英文名称	subject_company_english_name		CHARACTER	256		
主体行政区域	subject_main_administrative_region		NUMBER		Y	枚举-境内外类别
主体信息维护时间	subject_create_time	业务操作时间	TIME		Y	

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
机构性质	subject_organization_nature		NUMBER			枚举-机构性质
所有制性质	subject_ownership_nature		NUMBER			枚举-所有制性质
证件类型及证件号码	subject_document_information	只可选1或者同时提供2/3/4,枚举值只用于标记type类型。数据格式为: [{"type": 1, "code": "123" }, {"type": 1, "code": "123" }]	对象数组			枚举-机构代码类型
注册日期	subject_registry_date		DATE			
主营业务	subject_company_business		CHARACTER	2048		
公司简介	subject_company_profile		TEXT			
注册资本	subject_registered_capital		DECIMAL			
注册资本币种	subject_registered_capital_currency		CHARACTER	16		枚举-币种类型

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
实收资本	subject_paid_in_capital		DECIMAL			
实收资本币种	subject_paid_in_capital_currency		CHARACTER	16		枚举-币种类型
注册地址	subject_registered_address		CHARACTER	512		
办公地址	subject_office_address		CHARACTER	512		
联系地址	subject_contact_address		CHARACTER	128		
联系电话	subject_contact_number		CHARACTER	512		
企业传真	subject_fax		CHARACTER	64		
邮政编码	subject_postal_code		CHARACTER	32		
互联网地址	subject_internet_address	企业官网地址	CHARACTER	256		
电子邮箱	subject_mail_box		CHARACTER	64		
主管单位	subject_regulator		CHARACTER	128		

主体区块链节点信息见表 B. 4:

表 B. 4 区块链节点信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
节点所属主体ID	node_subject_id		CHARACTER			
节点名称	node_name		CHARACTER			
节点英文名称	node_english_name		CHARACTER			
节点类型	node_type		NUMBER			枚举-节点类型
节点接入时间	node_access_time		TIME			
节点接入方式	node_access_method		CHARACTER			
节点接入网络	node_access_network		CHARACTER			
节点机器数量	node_machine_numbers		CHARACTER			
节点机器类型	node_machine_type		NUMBER			枚举-机器类型
节点 IP	node_ip		CHARACTER			
节点物理位置	node_position		CHARACTER			

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
节点系统版本	node_system_version		CHARACTER			
节点备注信息	node_remark		CHARACTER			

b) 存证数据信息

存证数据用于记录各参与主体在行业联盟链上进行区块链存证的数据信息。存证数据信息划分为一到三级属性分类，属性分类应符合表 B. 5 的内容。例如二级子类包括客户账号信息、产品信息、事务环节信息。存证数据信息对公募基金委托下单进行全生命周期管理，通过事务环节唯一编码将客户下单公募基金的全流程数据串联形成链式证据，进行区块链存证。

表 B. 5 存证数据信息表

一级分类	二级分类	三级分类
存证数据信息	客户账号信息	--
	产品信息	--
	事务环节信息	登录信息
		浏览记录信息
		适当性匹配信息
		协议签署信息
		委托下单信息

表 B. 5 中具体信息的说明如下：

二级分类客户账号信息见表 B. 6：

表 B. 6 客户账号信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
账号所属主体	account_subject		CHARACTER		Y	
客户账号	account_code		CHARACTER		Y	
账号类型	account_type		NUMBER		Y	枚举-账号类型
账号用途	account_purpose		CHARACTER		Y	
账号状态	account_status		NUMBER		Y	枚举-账号状态
账号信息维护时间	account_create_time		TIME		Y	
账号创建日期	account_establish_date		DATE			
账号开户日期	account_opening_date		DATE			
账号开户文件	account_opening_doc		FILE_OBJECT			
账号销户日期	account_closing_date		DATE			
账号销户文件	account_closing_doc		FILE_OBJECT			
账号冻结日期	account_frozen_date		DATE			
账号冻结文件	account_frozen_doc		FILE_OBJECT			

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
账号解冻日期	account_thaw_date		DATE			
账号解冻文件	account_thaw_doc		FILE_OBJECT			
账号关联关系	account_association		NUMBER			枚举-关联关系
账号关联对象引用	account_associated_account_ref		CHARACTER			

二级分类产品信息见表 B. 7：

表 B. 7 产品信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
产品代码	product_code		CHARACTER		Y	
产品全称	product_name		CHARACTER		Y	
产品简称	product_name_abbreviation		CHARACTER			
产品类型	product_type		NUMBER		Y	枚举-公募基金产品类型
产品风险等级-适当性	product_risk_level_suitability		NUMBER		Y	枚举-风险等级
产品投资品种-适当性	product_invest_kind_suitability		NUMBER		Y	枚举-投资品种
产品投资期限-适当性	product_invest_term_suitability		NUMBER		Y	枚举-投资期限
产品可承受亏损-适当性	product_affordable_of_loss_suitability		NUMBER		Y	枚举-可承受亏损
产品最新净值	product_new_nav		CHARACTER			
产品累计净值	product_net_value		CHARACTER			
产品净值日期	product_nav_date		CHARACTER			
产品成立以来收益率	product_rate		CHARACTER			
产品规模	product_scale		CHARACTER			
信息披露方式	product_info_disclosure_way	公募基金均为公开披露	CHARACTER			
产品规模币种	product_scale_currency		CHARACTER			
产品规模总额	product_scale		CHARACTER			
联系人	product_issuer_contact_person		CHARACTER			

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
联系信息	product_issuer_contact_info		CHARACTER			
备注	product_remark		CHARACTER			
产品信息维护时间	product_create_time		TIME			
投资产品类型	product_investment_products_type		CHARACTER			
投资比例范围	product_investment_proportion_range		CHARACTER			
投资产品详情	product_Investment_product_details		CHARACTER			
投资产品详情说明文件清单	product_detailed_description_document		CHARACTER			
发行成功日期	product_release_date		DATE			
基金成立日期	product_foundation_date		DATE			
基金状态	product_fund_status		NUMBER		Y	枚举-基金状态
募集开始日期	product_raise_start_date		DATE			
募集结束日期	product_raise_end_date		DATE			

三级分类登陆信息见表 B. 8：

表 B. 8 事务环节登录信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
事务环节唯一编码	transaction_link_unique_code	一次存证环节的唯一编码	CHARACTER		Y	
登录账号	login_account		CHARACTER		Y	
登录时间	login_time		TIME		Y	
登录系统	login_application	客户登录的应用系统	CHARACTER		Y	
登录 IP	login_ip		CHARACTER			
登录终端信息	login_terminal	PC： MAC/移动端： 设备标识码	CHARACTER			

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
登出时间	logout_time		TIME			

三级分类浏览记录信息见表 B. 9:

表 B. 9 事务环节浏览记录信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y/N）	枚举值说明
事务环节唯一编码	transaction_link_unique_code	一次存证环节的唯一编码	CHARACTER		Y	
一级页面名称	primary_page_name	例如：产品详情页	CHARACTER		Y	
一级页面内容	primary_page_content		CHARACTER		Y	
一级页面点击操作	primary_page_click		CHARACTER			
一级页面浏览开始时间	primary_page_begin_time		TIME			
一级页面浏览结束时间	primary_page_end_time		TIME			
二级页面名称	secondary_page_name		CHARACTER			
二级页面包含内容	secondary_page_content		CHARACTER			
二级页面点击操作	secondary_page_click		CHARACTER			
二级页面浏览开始时间	secondary_page_begin_time		TIME			
二级页面浏览结束时间	secondary_page_end_time		TIME			

三级分类适当性匹配信息见表 B. 10:

表 B. 10 事务环节适当性匹配信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
事务环节唯一编码	transaction_link_unique_code	一次存证环节的唯一编码	CHARACTER		Y	
客户账号	account_code		CHARACTER		Y	
适当性匹配结果	suitability_match_result		NUMBER		Y	枚举-是否
风险等级匹配结果	risk_match_result		NUMBER		Y	枚举-是否

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
投资品种匹配结果	invest_kind_match_result		NUMBER		Y	枚举-是否
投资期限匹配结果	term_match_result		NUMBER		Y	枚举-是否
可承受亏损匹配结果	affordable_of_loss_match_result		NUMBER		Y	枚举-是否
产品风险等级-适当性	product_risk_level_suitability		NUMBER		Y	枚举-风险等级
产品投资品种-适当性	product_invest_kind_suitability		NUMBER		Y	枚举-投资品种
产品投资期限-适当性	product_invest_term_suitability		NUMBER		Y	枚举-投资期限
产品可承受亏损-适当性	product_affordable_of_loss_suitability		NUMBER		Y	枚举-可承受亏损
客户风险等级-适当性	cust_risk_level_suitability		NUMBER		Y	枚举-风险等级
客户投资品种-适当性	cust_invest_kind_suitability		NUMBER		Y	枚举-投资品种
客户投资期限-适当性	cust_invest_term_suitability		NUMBER		Y	枚举-投资期限
客户可承受亏损-适当性	cust_affordable_of_loss_suitability		NUMBER		Y	枚举-可承受亏损
适当协议名称	suitability_agreement_name		CHARACTER			
适当性协议编号	suitability_agreement_number		CHARACTER			
适当性协议散列值	suitability_agreement_hash	协议摘要信息	CHARACTER			
适当性协议签名串	suitability_agreement_signature	客户签署协议形成的密文串	CHARACTER			

三级分类协议签署信息见表 B. 11：

表 B. 11 事务环节协议签署信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填（Y）	枚举值说明
事务环节唯一编码	transaction_link_unique_code	一次存证环节的唯一编码	CHARACTER		Y	
客户账号	account_code		CHARACTER		Y	

协议名称	agreement_name		CHARACTER		Y	
协议编号	agreement_number		CHARACTER		Y	
协议散列值	agreement_hash		CHARACTER			
客户证书 sn	cust_certificate_sn		CHARACTER			
客户证书 dn	cust_certificate_dn		CHARACTER			
协议签名串	agreement_signature		CHARACTER			
签署时间	sign_time		TIME		Y	

三级分类委托下单信息见表 B. 12:

表 B. 12 事务环节委托下单信息表

中文名称	英文名称	详细说明	数据类型	数据长度	是否必填 (Y)	枚举值说明
事务环节唯一编码	transaction_link_unique_code	一次存证环节的唯一编码	CHARACTER		Y	
客户账号	account_code		CHARACTER		Y	
产品代码	product_code		CHARACTER		Y	
产品全称	product_name		CHARACTER			
产品类型	product_type		NUMBER			枚举-公募基金产品类型
委托编号	entrust_number		CHARACTER		Y	
委托时间	entrust_time		TIME		Y	
委托金额	entrust_amount		CHARACTER			
委托终端	entrust_terminal	客户委托下单的平台终端	CHARACTER			

枚举项信息见表 B. 13:

表 B. 13 枚举项信息表

枚举项	枚举值	说明
枚举-境内外类别	0	空值
	1	境内
	2	境外
	255	其他
枚举-机构性质	0	空值
	1	企业
	2	国家行政机关
	3	事业单位
	4	合作组织
	5	社会团体
	6	农民专业合作社
	7	个体工商

枚举项	枚举值	说明
	255	其他
枚举-所有制性质	0	空值
	1	中央国有控股
	2	地方国有控股
	3	外资
	4	民营
	255	其他
枚举-机构代码类型	0	空值
	1	社会统一信用代码
	2	组织机构代码证
	3	营业执照
	4	税务登记证
	255	其他
枚举-币种类型	156	人民币
	840	美元
	036	澳元
	978	欧元
	276	德国马克
	344	港币
枚举-节点类型	1	全节点
	2	轻节点
枚举-机器类型	1	物理机
	2	虚拟机
枚举-账号类型	1	证券账号
	2	资金账号
	3	游客账号
枚举-账号状态	1	正常
	2	注销
	3	冻结
	4	未实名认证
枚举-关联关系	0	空值
	1	资金账号与证券账号关联
枚举-公募基金产品类型	1	股票型
	2	债券型
	3	混合型
	4	货币型
	255	其他
枚举-是否	1	是
	2	否
枚举-风险等级	1	低风险
	2	中低风险

枚举项	枚举值	说明
	3	中风险
	4	中高风险
	5	高风险
枚举-投资品种	1	债券、货币市场基金、债券基金等固定收益类投资品种
	2	股票、混合型基金、偏股型基金、股票型基金等权益类投资品种
	3	期货、期权等金融衍生品
	4	其他产品或服务
	255	其他
枚举-投资期限	1	没有经验
	2	少于 2 年
	3	2-5 年
	4	5-10 年
	5	10 年以上
	255	其他
枚举-可承受亏损	1	10%以内
	2	10%-30%
	3	30%-50%
	4	超过 50%
	255	其他
枚举-基金状态	1	可申购赎回
	2	发行
	3	停止申购赎回
	4	停止申购
	5	停止赎回
	6	基金终止
	7	基金封闭
	255	其他

c) 上链数据示例

```
{
  "transaction_info_array": [{
    "transaction_info": {
      "transaction_detail_info": {
        "biz_data": {
          "bizName": "公募基金委托下单业务存证",
          "businessData": [{
            "StageHash":
"24fc216d15df3a4fe7c913f1f1feb4653d0438115b6fc59e91ca99f56637a70c",
            "StageName": "存证主体-主体基本信息"
          }], {
```


附录 C
(资料性)
流式数据存证模型应用示例

C.1 存证过程消息流程图

存证过程消息流程图给出了存证过程中的主要消息流向和交互过程的示例，如图 C.1 所示（红色箭头表示）：

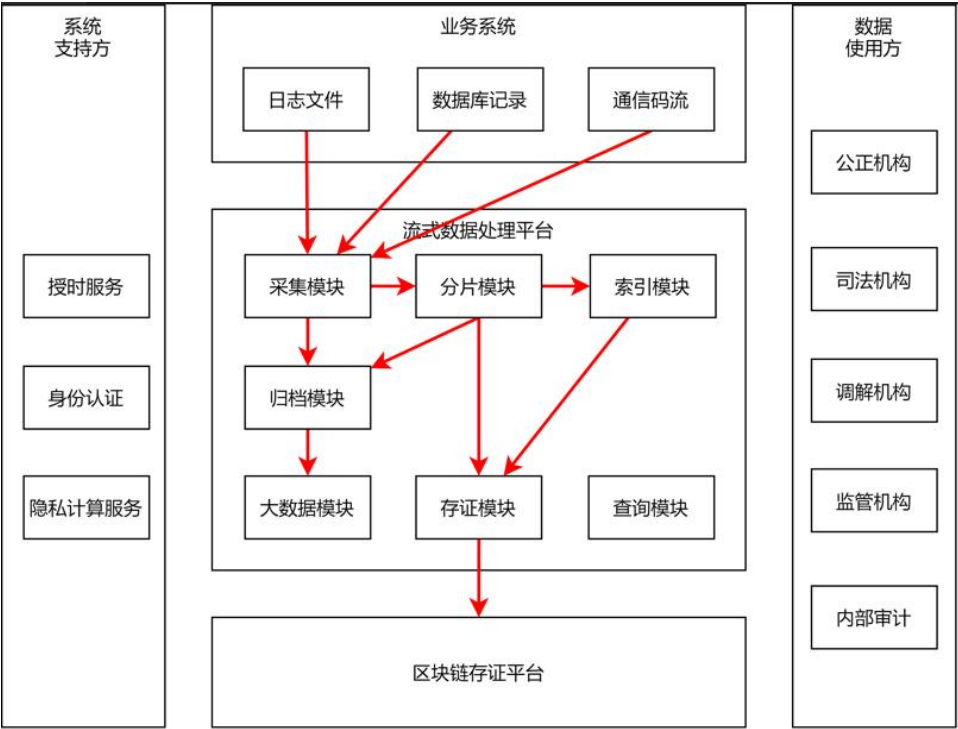


图 C.1 存证过程的消息流程图

存证过程主要包括以下几个方面：

- a) 流式数据处理平台实时从业务系统上采集流式数据；
- b) 流式数据处理平台实时将数据存档至大数据模块；
- c) 利用分片模块进行实时分片，并将分片归档，同时将分片存证至区块链存证平台；
- d) 计算分片索引，并将索引归档，同时将索引存证至区块链存证平台。

C.2 取证过程消息流程图

取证过程消息流程图给出了取证过程中的主要消息流向和交互过程的示例，如图 C.2 所示（红色箭头表示，箭头上的数字表示顺序）：

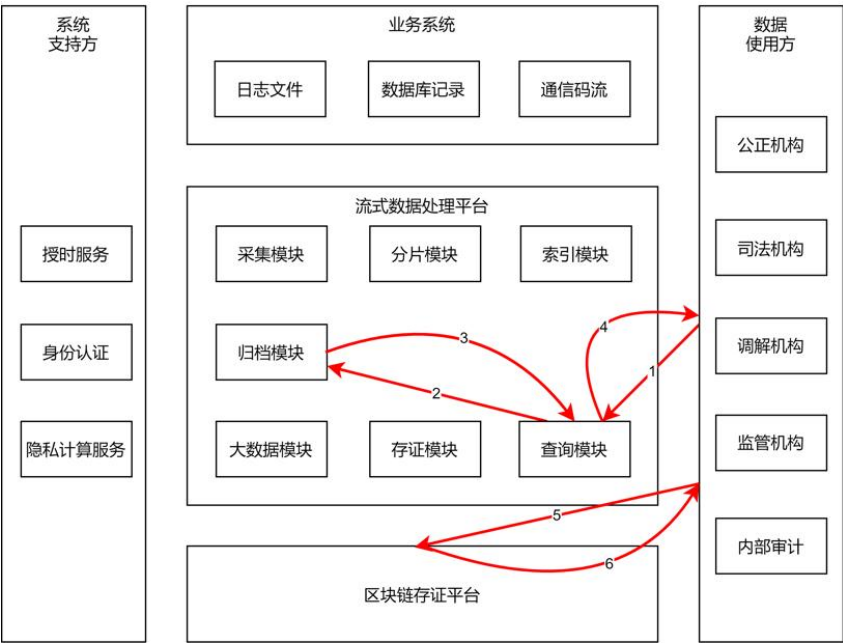


图 C.2 取证过程的消息流图

主要包括以下几个方面：

- a) 数据使用方请求流式数据处理平台查询相关流式数据；
- b) 查询模块请求归档模块返回相关信息；
- c) 归档模块返回相关索引、分片信息；
- d) 查询模块组织查询结果，并返回；
- e) 数据使用方请求区块链存证平台返回存证信息。

C.3 切片算法与数据结构示例

流式数据由于自身的特性实际处理时应对其进行分片处理，如图 C.3 所示：

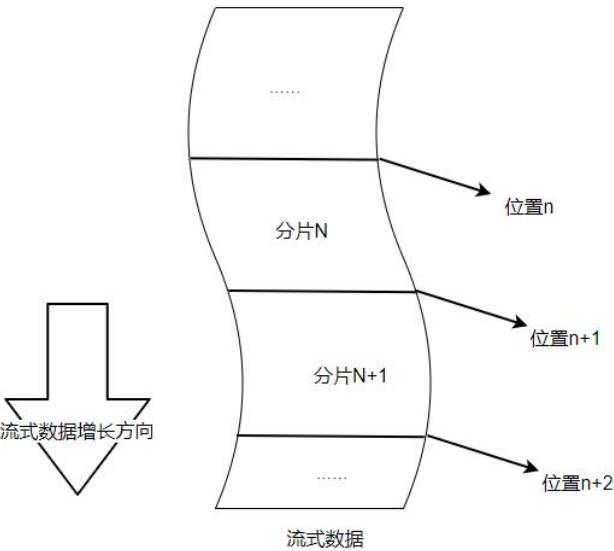


图 C.3 流式数据分片示意图

随着流式数据的不断采集，我们利用分片策略持续生成分片，可选的分片策略有：

- a) 以固定的频率进行分片；
- b) 以固定的大小进行分片；
- c) 以数据流业务上的属性进行分片（包括指定的日志行数、网络交互的数据包等）；
- d) 分片应该足够小，这样能够保证上链的实时性，分片应包含以下字段：
 - 流式数据标识：唯一标识流式数据；
 - 分片的标识：唯一标识分片；
 - 分片生成的时间；
 - 分片的起始位置；
 - 分片的截止位置。

C.4 索引算法与数据结构示例

为了能够加速查询、确保流式数据完整性，流式数据分片后应建立索引。利用索引可以在不了解原文的情况下确认分片是否有相关信息，并且利用索引信息可以保障数据完整无缺失。建立索引的关键过程如下：

- a) 根据业务系统特点得到关键字集合 $\{k | k \text{ 属于业务上可供查询的关键字}\}$ ，记作 K ；
- b) 针对集合 K 中的所有关键字 k ，计算出分片中关键字出现的次数，从而得到集合 $\{(k, n) | k \in K, n \text{ 为分片中 } k \text{ 出现的次数}\}$ ，标识成集合 KN ，集合 KN 即为此分片的索引；
- c) 索引信息应包含以下字段：
 - 流式数据标识：唯一标识数据流；
 - 分片的标识：唯一标识分片；
 - 关键字及出现次数列表。

C.5 存证信息生成算法

分片以及索引文件中可能存在用户隐私数据，根据实际业务场景以及隐私保护要求，可以选择不同的存证信息算法，从而达到隐私保护的目标。这里我们列举两种常见的算法：

- a) Hash 存证：存证时计算原文的 Hash 值并上链，取证的过程中利用 Hash 校验完整原文，可以得到原文是否被篡改结论。Hash 存证信息算法简单，但是实际取证的过程中需要有完整的原文信息，这里有隐私泄露的风险；
- b) 默克尔树存证：存证时计算默克尔树，并将根节点上链，取证时只需要原文中少部分信息，连同默克尔树部分节点即可得到信息是否被篡改结论，默克尔树算法相对复杂，但是更有利于隐私保护。

下面详细介绍默克尔树在分片、索引信息存证中的应用示例，默克尔树生成的具体的过程如下：

- a) 假定分片中的内容可以从业务上分为不同的条目（日志文件的单行，数据库的单条记录等），索引信息中的 (k, n) 也可以看成一个条目，这些条目记作 $item$ ，分片或者索引信息可以看成条目的列表，记作 $[item]$ ；
- b) 将 $[item]$ 中的元素进行 Hash 运算得到序列 $[H]$ 其中每个元素 $H_i = \text{Hash}(item_i)$ ，利用 $[H]$ 中的元素生成默克尔树的根节点；
- c) 根节点两两组合并计算组合后的 Hash 函数得到第二层节点；
- d) 重复 b) 步，直到生成一个根节点 H_{root} ；
- e) H_{root} 的值即为索引文件的存证值，上传到区块链中进行存证。

索引文件默克尔树结构如图 C.4 所示：

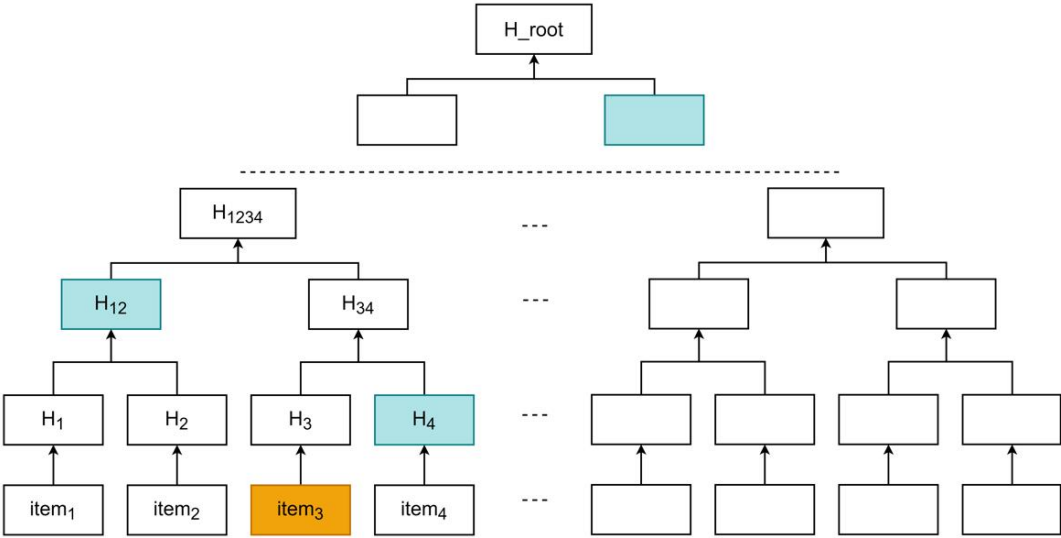


图 C.4 索引文件默克尔数示意图

- 取证时，利用默克尔树验证数据的真实完整性过程如下：
- a) 需要原始数据中待验证的条目 $item$ （如图 C.4 中 $item_3$ ）记作 $item_s$ ；
 - b) 以及相关的默克尔树节点（如图 C.4 中 H_4 、 H_{12} 等）这些节点可以同 $item_s$ 计算出默克尔树根节点；
 - c) 计算出默克尔树根节点 H_{root} ，该根节点值即为索引信息的存证值；
 - d) 通过区块链存证平台可以验证 H_{root} 的有效性从而证实 $item_s$ 的真实可靠。
- 存证信息生成算法的选择可以参考表 C.1：

表 C.1 存证信息生成算法选择参考

场景	分片存证信息算法	索引存证信息算法
业务不涉及隐私	Hash 存证	Hash 存证
分片涉及隐私, 索引不涉及隐私	默克尔树存证	Hash 存证
分片、索引均涉及隐私	默克尔树存证	默克尔树存证