

JR

中华人民共和国金融行业标准

JR/T XXXXX—XXXX

证券期货业大数据全生命周期管理指南

Guideline for the full lifecycle management of big data in the securities
and futures industry

(征求意见稿)

XXXX—XX—XX 发布

XXXX—XX—XX 实施

中国证券监督管理委员会 发布

目次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 2

5 基本原则 2

 5.1 全过程域原则 2

 5.2 实用性原则 2

 5.3 安全性原则 2

 5.4 可用性原则 2

 5.5 适配性原则 3

6 总体规范 3

 6.1 数据采集 3

 6.2 数据存储 5

 6.3 数据处理 10

 6.4 数据服务 14

 6.5 数据退役 17

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。

本文件由全国金融标准化技术委员会证券分技术委员会（SAC/TC 180/SC4）提出。

本文件由全国金融标准化技术委员会（SAC/TC 180）归口。

本文件起草单位：南京证券股份有限公司、深圳证券交易所、郑州商品交易所、中国信息通信研究院、南京数字金融产业研究院有限公司、中国国际金融股份有限公司、中泰证券股份有限公司、国金证券股份有限公司、东吴证券股份有限公司、华泰期货有限公司、浙商期货有限公司、南京银行股份有限公司、北京大学、华为技术有限公司、星环信息科技（上海）股份有限公司、恒生电子股份有限公司。

本文件主要起草人：江念南、张之浩、郭枫、谷博、赵延鹏、徐祯、崔一妍、陈莹、李冬、陈绪申、王洪涛、葛菊平、罗庄艮、吴福文、徐小锋、江天玥、沈晴霓、沈玮、张晓明、应雄。

引 言

随着金融科技的发展，证券期货业在交易撮合、行情发布、清算交收、风险监控等业务环节产生和积累了海量数据，包括交易记录、账户信息、持仓数据等结构化数据，以及市场公告、研究报告、音视频资料等多样化的半结构化与非结构化数据。大数据技术的广泛应用使行业机构能够深入挖掘数据价值、提升市场分析能力、优化投资决策、强化风险管控、改进客户服务，数据已成为支撑行业高质量发展的关键生产要素。

证券期货业大数据呈现出体量巨大、类型多样、流转频繁、时效性强等特征，在采集、存储、处理、服务、退役等各个环节面临着质量控制、安全保障、合规管理等多重挑战。随着监管要求的日趋严格和大数据应用的不断深化，建立标准化的大数据管理体系已成为行业发展的迫切需求。实施大数据全生命周期管理，通过对数据从产生到退役的完整过程进行系统化管理，能够明确各阶段的管理要求和技术要求，实现对数据资产的精细化管理。规范的全生命周期管理有助于行业机构优化资源配置，提升管理效率，促进数据价值的有效释放。因此，为提升证券期货业大数据管理水平，特制定本文件。

本文件面向大数据管理，对大数据全生命周期中的数据采集、数据存储、数据处理、数据服务、数据退役等过程域提出管理要求和技术要求，为证券期货业机构建立健全大数据全生命周期管理体系提供标准化指引，供证券期货业机构参考。

证券期货业大数据全生命周期管理指南

1 范围

本文件给出了证券期货业大数据全生命周期管理的指导思路及方法，涵盖了数据采集、数据存储、数据处理、数据服务和数据退役等过程域。

本文件适用于证券期货行业机构（以下简称行业机构）开展大数据全生命周期管理工作，并为行业机构建立和实施有效的大数据管理规范提供参考。

注：行业机构包括承担证券期货市场公共职能的机构、承担证券期货行业信息技术公共基础设施运营的机构等证券期货市场核心机构及其下属机构，以及证券公司、基金管理公司、期货公司、证券期货服务机构等证券期货经营机构。

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 35295-2017 信息技术 大数据 术语

JR/T 0166-2020 云计算技术金融应用规范 技术架构

JR/T 0236-2021 金融大数据 术语

JR/T 0304.1-2024 证券期货业基础数据元规范 第1部分：基础数据元

JR/T 0304.2-2024 证券期货业基础数据元规范 第2部分：基础代码

3 术语和定义

JR/T 0236-2021 《金融大数据 术语》界定的以及下列术语和定义适用于本文件。

3.1

大数据 big data

海量的数据集，其数据在本质上具有体量大、种类多、变化快、变数多的特征，需要一种易扩展的技术来有效存储、计算、管理和分析。

注：1. 大数据通常以多种不同方式使用，例如，作为某种用于处理大数据海量数据集的易扩展技术的名称。

2. GB/T 35295—2017《信息技术 大数据 术语》中2.1.1对大数据的定义为“具有体量巨大、来源多样、生成极快、且多变等特征并且难以用传统数据体系结构有效处理的包含大量数据集的数据”。

3. 在某些非工程性研讨的场合，“大数据”一词的外延可能被扩大到所有的数据。

[来源：JR/T 0236-2021，3.1]

3.2

资源池 resource pool

一组物理资源或虚拟资源的集合。

注：按照一定规则可从池中获取资源，也可释放资源并由资源池回收。资源包括物理机、虚拟机、物理存储资源、虚拟存储资源、物理网络资源和虚拟网络资源等。

[来源：JR/T 0166-2020，3.24]

3.3

分布式文件系统 distributed file system

多个结构化数据集分布在一个或多个服务器集群的各计算节点的文件系统。

注：此类系统中，数据可能分布在文件或数据集层，更为普遍的是在数据块层级分布，同时支持集群中多个节点与大型文件或数据集的不同部分交互。

[来源：JR/T 0236-2021，6.12]

3.4

非结构化数据 unstructured data

特征为除了记录或文件级别外没有任何结构的数据。

注：非结构化数据不是由数据元素组成。

[来源：JR/T 0236-2021, 3.30]

4 缩略语

下列缩略语适用于本文件。

API：应用程序接口（Application Program Interface）

APP：应用程序（Application）

CDC：变更数据捕获（Change Data Capture）

CEP：复杂事件处理（Complex Event Processing）

CPU：中央处理器（Central Processing Unit）

CSV：逗号分隔值（Comma-Separated Values）

FQDN：完全限定域名（Fully Qualified Domain Name）

IP：网际互连协议（Internet Protocol）

JAR：Java归档（Java Archive）

JDBC：Java数据库连接（Java Database Connectivity）

JSON：JavaScript对象简谱（JavaScript Object Notation）

LPA：标签传播算法（Label Propagation Algorithm）

MPI：消息传递接口（Message Passing Interface）

NAS：网络附加存储（Network Attached Storage）

ODBC：开放数据库连接（Open Database Connectivity）

RAID：独立磁盘冗余阵列（Redundant Array of Independent Disks）

SAN：存储区域网络（Storage Area Network）

SDK：软件开发工具包（Software Development Kit）

SQL：结构化查询语言（Structured Query Language）

SSD：固态硬盘（Solid-State Drive）

TLS：传输层安全协议（Transport Layer Security）

XML：可扩展标记语言（Extensible Markup Language）

5 基本原则

5.1 全过程域原则

贯穿数据从产生到退役的完整过程，涵盖数据采集、数据存储、数据处理、数据服务和数据退役五个过程域，将存储资源池、网络资源池、计算资源池管理与高可靠性保障相结合，形成完整的大数据全生命周期管理体系。

5.2 实用性原则

契合证券期货业的业务特征和发展需求，根据数据在交易频率、数据量、时效性、监管要求等方面的特征，结合各业务层次、各环节管理中数据处理的便利和可行性，满足业务管理需求。

5.3 安全性原则

基于数据安全等级的差异，从组织、制度、技术三个方面建立完备的数据安全保护措施，实现数据本体和计算过程的安全，确保数据全生命周期各过程中数据安全，无数据泄漏风险，符合相关法律法规及管理办法中对数据安全的要求。

5.4 可用性原则

确保数据在整个生命周期内保持高可用状态，确保在其全生命周期内获取与处理速度的快捷性、内容的时效性，以及数据内容的完整性，无缺失、损毁，能够满足数据消费需求。

5.5 适配性原则

充分考虑行业机构在规模、业务、技术等方面的差异性，提供灵活可调的管理框架和实施建议，指导各行业机构将大数据全生命周期管理落实到具体的组织架构与岗位职责中，保障符合数据管理相关规范性文件或者流程的要求。

6 总体规范

6.1 数据采集

6.1.1 概述

数据采集是指为满足集中数据治理或数据加工等需求，将指定数据按约定的方式从各类数据源通过采集工具采集到统一的大数据存储设施的过程。

6.1.2 管理要求

6.1.2.1 采集管理定义

数据采集的管理要求，是指在数据采集活动全流程中（事前准备、事中执行、事后管控阶段）应采取的必要管理措施，旨在保障数据采集活动得以安全、规范且有效地开展，数据采集管理活动中涉及的角色如下：

- a) 数据申请方：采集数据的直接或间接使用人员，与所采集数据存在业务关联，负责提出明确的数据采集需求；
- b) 数据采集方：数据采集能力的建设、开发和实施人员，负责根据需求完成数据采集活动；
- c) 技术管理方：数据源系统的开发以及运维管理人员，负责为数据采集方提供必要的技术支持；
- d) 业务管理方：数据源系统承载业务的建设和管理人员，负责定义和解释业务数据标准。

6.1.2.2 采集需求管理

清晰且规范的数据需求管理，是确保所采集数据完整、准确，采集作业稳定、可靠，并满足业务使用需求与运维管理要求的重要依据，具体管理要求如下：

- a) 应与数据申请方确认所需采集数据的业务来源、数据内容、使用目的以及使用期限；
- b) 应与数据源系统的业务管理方、技术管理方明确采集的时机、频率、业务约束条件以及历史数据范围；
- c) 应与数据源系统的技术管理方明确采集的数据源类型、数据格式和采集方式；
- d) 宜在数据采集前与源系统技术管理方评估数据采集环境条件，包括采集活动对源系统性能、功能以及系统间网络带宽资源的影响；
- e) 宜建立数据源认证体系，明确每类具体业务数据的权威源系统，以确保数据从正确数据源采集；
- f) 宜在采集需求变更时，保留变更记录并重新确认以上业务需求信息。

6.1.2.3 采集运维管理

采集运维管理是指在数据采集过程中，为保障数据采集工作的正常开展，并确保异常情况发生时能够得到高效处置，所制定的运维管理方案及对相关必要信息的归集工作：

- a) 应明确所采集数据的管理主体责任，划分数据源系统的业务管理方、技术管理方与数据采集方在数据采集活动中的管理职责；
- b) 应与数据申请方、源系统业务管理方、技术管理方明确数据采集异常时的应急处置方案；
- c) 应从数据源系统的业务管理方获取所采集数据的业务元数据，包括但不限于业务域、业务对象、业务属性、业务规则；
- d) 应从数据源系统的业务管理方获取所采集数据的技术元数据，包括但不限于表名、字段名、字段类型，以及业务对象、业务属性与表、字段的对应关系；

- e) 应及时获取和维护数据源系统的网络地址、采集所需的访问账号与权限；
- f) 应定期对数据源的连通性与采集权限进行检查、异常告警与及时修复。

6.1.2.4 数据质量管理

数据质量管理为所采集数据的可用性提供保障，具体管理要求如下：

- a) 应与源系统业务管理方、技术管理方对采集数据进行质量评估，并确定数据采集的质量检查规则，包括但不限于完整性、一致性、准确性、唯一性、有效性和及时性；
- b) 数据源系统应对生成的落地数据建立符合质量检查规则的检查或管控措施；
- c) 数据采集方应根据确立的质量检查规则，对采集后的落地数据进行质量校验；
- d) 应定期评估数据质量规则的有效性，并及时更新维护质量检查规则。

6.1.2.5 数据安全治理

数据安全治理是为避免数据在采集过程中发生数据泄露等安全风险，具体管理要求如下：

- a) 对采集的数据，应与源系统业务管理方确定数据的分类分级标准，以及脱敏规则；
- b) 对采集的外购数据，应在采购合同中明确数据安全方面的责任与义务；
- c) 对采集的第三方数据，应确保数据获取途径合法、使用范围合规；
- d) 应采用防火墙、网络入侵检测等技术手段保障数据采集的网络安全；
- e) 应制定数据加密规则和静态脱敏规则；
- f) 应保留数据采集的审计日志，包括采集时间、采集对象、采集数据量、采集耗时、传输速度、采集IP和采集账户；
- g) 数据采集活动的全流程中不应明文暴露采集账号密码；
- h) 应在采集活动异常中断、正常结束后清除中间缓存数据。

6.1.3 技术要求

6.1.3.1 技术要求定义

数据采集的技术要求，是指为满足数据采集的各种场景需求，采集平台、工具及其相关配套设施应当具备的技术能力，以保障业务所需的各类数据能够以指定方式和策略高效地采集至大数据存储设施，并保障采集数据的质量与安全。

6.1.3.2 采集能力要求

数据采集平台、工具应兼容多种数据类型的采集能力，具体要求如下：

- a) 应支持对结构化的数据采集，包括但不限于主流的关系型数据库、表格文件；
- b) 应支持对通用数据类型的采集，包括但不限于整型、固定精度数值型、浮点型、定长字符型、变长字符型、布尔型、日期型、时间戳类型；
- c) 应支持半结构化的数据采集，包括但不限于非关系型数据库、JSON、XML、CSV文件等；
- d) 宜支持对半结构化数据中多级嵌套的复杂类型数据的解析和采集，包括对象、字典、列表、数组、结构体等；
- e) 应支持非结构化数据的采集，包括但不限于文档、图像、音频文件、视频文件等常见格式，以及自定义格式的数据；
- f) 应支持基于IP、FQDN的网络通信方式，支持跨防火墙、网间通信区的数据传输能力；
- g) 应支持断点续传机制，支持脏数据容错采集限制；
- h) 应支持对单一采集作业的采集速度、采集通道数配置；
- i) 应支持数据清洗处理，包括但不限于过滤重复数据，过滤无效值、缺失值的记录；
- j) 应支持数据转化处理，包括但不限于配置数据采集的业务约束条件，支持字段类型转化、格式转化、值映射；
- k) 应具备数据质量异常告警能力，以及数据质量处理机制；
- l) 应支持数据加密传输，支持数据静态脱敏传输。

6.1.3.3 采集方式要求

数据采集平台、工具应支持多种数据采集方式，以便于根据源系统特点、数据时效性要求、源系统性能，以及数据量确定适宜的采集策略，具体要求如下：

- a) 应支持对数据库、共享存储、数据接口等数据源介质的数据采集；
- b) 应支持对源系统指定库表、目录、地址的全量数据采集；
- c) 应支持对源系统指定库表、目录、地址的增量数据采集；
- d) 应支持对多个数据源的批量离线采集；
- e) 应支持基于归档日志、时间戳、触发器、快照等方式的CDC实时采集；
- f) 应支持基于增量字段值的周期性准实时数据采集；
- g) 应支持采集方主动拉取与数据源主动推送的数据采集方式。

6.1.3.4 开发方式要求

完善、便捷、高效的数据采集作业开发功能是大规模、持续性开展数据采集的重要支撑，具体技术要求如下：

- a) 应支持可视化的采集作业配置以及 workflow 编排；
- b) 结构化数据采集应支持多表、整库的批量采集作业配置；
- c) 结构化数据采集应支持数据源表的表结构查询，支持源表与目的表的字段映射，宜支持根据源表结构自动创建目的表；
- d) 非结构化数据采集应支持目录结构复制，宜支持对目的端目录的自定义创建与映射；
- e) 应支持采集作业的测试运行和日志查看，支持调度策略配置、告警规则配置，支持失败处置策略配置，包括但不限于间隔重试、自动跳过、阻断等。

6.1.3.5 采集调度要求

采集作业调度能力是对证券期货业具有行业特点的多业务场景支撑的重要保障，具体技术要求如下：

- a) 应支持自动定时的采集作业调度；
- b) 应支持源系统提供数据就绪状态标志位的触发式采集作业调度；
- c) 应支持多种调度日历管理，包括交易日历、港股交易日历、期货夜盘交易日历、法定工作日、自然日历、自定义调度日历等；
- d) 应支持作业调度并发度设置，包括全局并发度设置以及分源系统的采集并发度设置；
- e) 应支持采集作业所属执行节点的资源组配置；
- f) 应支持采集作业工作流的上下游依赖查询，支持根据依赖关系对某个节点的所有下游节点作业进行批量调度；
- g) 应支持采集作业的批量冻结与重新调度配置。

6.1.3.6 平台架构要求

数据采集平台、工具应具备较完善的高可用性、运维监控能力以及可扩展性，以保障采集作业的稳定运行，具体技术要求如下：

- a) 应具备分布式、高可用架构，支持故障节点的自动作业转移，支持对集群规模进行弹性扩缩容；
- b) 应具备对机器资源、服务状态、网络流量、作业执行的监控能力，支持短信、邮件、电话、APP消息等多种告警消息通知渠道，支持告警接收人配置；
- c) 应支持对新类型数据源的驱动版本的兼容扩展。

6.2 数据存储

6.2.1 概述

数据存储是指证券期货业在提供产品和服务、开展经营管理等活动中，为实现数据的长期、安全、可靠保存，利用各类存储载体与技术手段对数据进行持久化保存的过程。存储形式包括但不限于采用云存储服务、网络附属存储（NAS）、存储区域网络（SAN）以及本地存储设备等多种形态。

6.2.2 管理要求

6.2.2.1 数据分类分级管理

根据数据敏感性和业务重要性对证券期货业数据进行分类分级，明确存储保护策略，实施动态分类管理，防范数据泄露和滥用风险。具体管理要求如下：

- a) 应明确客户信息、交易信息、管理信息和行情数据等分类规则；
- b) 应根据数据的影响范围、影响对象和影响程度进行数据安全级别划分，将数据安全级别从高到低划分为核心、重要、一般三个等级；
- c) 应针对不同分类分级的数据采取不同级别的安全存储措施，包括加密存储、备份策略和访问控制等，确保数据的保密性、完整性和可用性；
- d) 应定期对数据存储过程中可能产生的影响进行风险评估，并采取相应安全防护措施；
- e) 应定期评估数据分级合理性，随业务变化更新存储策略。

6.2.2.2 访问控制与权限管理

通过严格的访问控制与权限管理机制，确保数据存储系统仅被授权人员访问，通过身份认证、授权、访问日志记录等手段，限制数据的读取、修改、删除等操作权限，防止未经授权的访问和数据泄露，保障数据的保密性、完整性和可用性。具体管理要求如下：

- a) 应依据“业务必需、最小权限”的原则，用户权限与其岗位职责匹配，对各类系统用户设计其工作必需的最小访问权限和操作权限；
- b) 应体现职责分离的安全制约原则，实现开发、运维、审计等角色权限分离，禁止同一账户兼任不相容职责；
- c) 应对业务平面和管理平面各自可访问的资源策略进行独立配置，并对业务平面和管理平面的相互访问进行隔离；
- d) 应明确外部机构的保密义务，外部机构访问数据需签署保密协议；
- e) 应以“一事一议”为原则对第三方访问进行授权，并视数据安全等级定期对授权审核审计；
- f) 应运用数字证书或多因素身份认证等技术手段对用户进行严格的身份核实；
- g) 应对访问存储业务的应用程序进行身份验证，赋予应用唯一标识，且系统中不应存在任何能够绕过身份验证机制的访问方式。

6.2.2.3 数据加密与隐私保护管理

通过加密技术对存储的数据进行加密处理，确保数据在存储和传输过程中不被未授权访问或泄露。同时，采取隐私保护措施，如匿名化、去标识化等，防止个人隐私信息被识别或滥用，保障数据的保密性、完整性和可用性，满足法律法规和行业标准对数据隐私保护的要求。具体管理要求如下：

- a) 应根据数据分类分级结果对敏感数据（如客户身份信息、交易记录）和重要数据进行脱敏和加密处理，加密策略应与数据分类分级结果相匹配；
- b) 应采取加密、去标识化等安全技术措施对客户身份证号、银行卡号等个人信息和敏感数据进行保护，防止未经授权的访问以及个人信息泄露、篡改、丢失；
- c) 在线数据和离线数据用于非生产环境时，应对数据进行脱敏处理；
- d) 应将脱敏后的数据与用于还原数据的恢复文件隔离存储；
- e) 应经过严格审批后使用恢复原始数据的技术，并留存相关审批及操作记录。

6.2.2.4 备份与恢复管理

通过制定科学的备份策略与恢复机制，定期备份数据并验证备份数据的完整性和可用性，确保证券期货业数据在灾难或故障场景下可快速恢复，保障业务连续性和数据完整性。具体管理要求如下：

- a) 应根据数据的安全级别和数据对系统运行的影响，制定数据备份策略和恢复策略；
- b) 应明确备份范围、备份方式、备份频度、存放地点、存放时限、有效性验证方式和管理责任人；
- c) 应采取实时备份与异步备份、增量备份与完全备份相结合的备份方式；
- d) 应建立同城与异地数据备份中心的远程数据备份与恢复功能，利用通信网络将关键数据定时批量传送至备用场地；

- e) 应定期对备份数据的有效性和可用性进行检查，定期对主要备份业务数据进行恢复验证，根据介质使用期限及时转储数据，确保数据可用性，如发现问题应采取措施修复备份数据；
- f) 应制定应急预案并根据预案定期组织关键岗位人员开展灾难恢复演练，应对技术方案中关键技术的可行性进行验证测试，并记录和保存验证测试的结果，频率不低于一年一次。

6.2.2.5 日志与监控管理

对数据存储系统中的所有操作行为进行记录、监控和分析，生成详细的日志信息。通过实时监控和事后审计，及时发现异常访问、数据篡改或未授权操作等安全事件，确保数据存储的安全性和合规性。具体管理要求如下：

- a) 应做好数据存储的操作留痕，并形成检查或审计机制；
- b) 应记录内容包括时间、用户、IP地址、操作对象、操作内容、操作行为和操作结果等相关信息，操作日志至少留存6个月；
- c) 应部署实时监控系统，对异常登录、数据异常使用、数据高频导出等用户异常行为进行分析，必要时触发告警，并对异常情况及时处置；
- d) 应定期分析数据存储系统监控日志和操作记录，分析异常情况，形成评估记录，跟踪处理日志分析中发现的异常事件；
- e) 应定期开展风险评估，重点检查权限滥用、数据泄露风险，并向有关主管部门报送风险评估报告。

6.2.2.6 存储介质安全管理

对存储数据的物理设备或载体（如硬盘、磁带、光盘等）进行全生命周期的管理。保证其在物理和逻辑层面的安全性，防止数据因介质损坏、丢失或非法访问导致泄露或损毁，保障数据的完整性和可用性。具体管理要求如下：

- a) 应对包括备份介质在内的存储介质出入库采取措施进行出入库控制，并由指定岗位人员完成，未经授权，任何存储介质不应带离库房；
- b) 应执行严格的授权审批程序针对第三方机构人员对于存储系统服务器与带库区域的访问；
- c) 应做好介质检查工作，制定备份介质抽检计划，检查备份介质可用性和备份数据的有效性；
- d) 应将介质应存放于防火、防磁、防潮的专用保险柜或机房，对各类介质进行控制和保护，实行存储环境专人管理，并根据存档介质的目录清单定期盘点；
- e) 应部署电子门禁、视频监控等措施控制、鉴别和记录进入存储系统服务器与带库等设备机房的人员；
- f) 应制定数据销毁指引，明确销毁方式和销毁要求，设置销毁相关监督角色，监督操作过程，并对审批和销毁过程进行记录控制，从而实现对数据的有效销毁；
- g) 应做好数据删除的检查和验证，对数据内容进行清除，防止因对存储介质上数据内容的恶意恢复而导致的数据泄露风险。

6.2.2.7 合规审计与改进管理

定期对数据存储管理流程和措施进行合规性审查，确保其符合法律法规和行业标准要求。通过评估审计结果，识别潜在风险与不足，制定并实施改进计划，持续优化数据存储管理策略，提升数据安全性和管理效能，保障数据存储的长期合规与高效运行。具体管理要求如下：

- a) 应定期开展数据存储管理工作专项审计，频率不低于每年一次，确保三年内完成数据存储管理全部事项的审计工作；
- b) 应当委托外部专业机构开展数据存储管理工作的全面审计，频率不低于每三年一次；
- c) 应跟踪审计发现问题的整改情况，并将审计报告提交相关部门审议；
- d) 应妥善保存审计报告，保存期限不得少于二十年，包括但不限于纸质材料、电子版材料、离线备份材料；
- e) 应建立数据存储管理评估制度，结合行业技术趋势更新存储策略。

6.2.3 技术要求

6.2.3.1 基础要求

6.2.3.1.1 架构

- a) 数据存储系统应构建分层分布式架构，底层设置大容量、低成本的归档存储层，用于长期保存低频访问的历史数据；中层为通用的在线存储层，支持日常业务的数据读写访问；上层为高性能缓存层，以加速高频访问数据的读取。各层之间应支持高效的数据迁移机制，以适应数据访问热度的变化；
- b) 应采用分布式存储技术，将数据分散存储于多个存储节点，规避单点故障风险，同时实现存储资源的弹性扩展。通过分布式文件系统或对象存储系统，应实现对海量数据的统一管理和访问，确保数据的一致性、完整性与可用性。

6.2.3.1.2 性能

- a) 查询吞吐率：在正常业务负载下，对于实时查询请求，存储系统应保证每秒处理一定数量的查询操作，具体指标需根据业务规模和复杂程度确定。对于复杂的多表关联查询或涉及大数据量的查询，响应时间应控制在可接受范围内；
- b) 交易吞吐率：在交易高峰期，存储系统应能够支持每秒处理大量的交易数据写入和读取操作。鉴于证券期货交易数据量大且交易频繁，存储系统应具备高并发处理能力，具体指标需根据业务规模和复杂程度确定，以确保交易的实时性和数据的准确性；
- c) 数据写入速度：针对业务系统持续产生的大量新数据，存储系统应具备快速的数据写入能力，确保数据能够及时持久化存储。平均每秒的数据写入速率具体指标需根据业务规模和复杂程度确定，对于突发的数据写入高峰，应能够在短时间内维持更高的写入速度，避免数据积压。

6.2.3.1.3 可靠性

- a) 可用性效率：存储系统的整体可用性应达到极高水平，年度可用性应不低于99.99%，以确保业务的连续性。应通过冗余设计、故障自动切换等技术手段，保障存储系统在部分组件出现故障时仍能正常运行；
- b) 宕机恢复时间：当存储系统发生故障导致宕机时，应具备快速恢复机制。从故障发生到系统恢复正常运行时间应尽可能短，关键业务数据存储系统的恢复时间和普通业务数据存储系统的恢复时间应符合相关要求，以减少对业务的影响；
- c) 异地容灾：应建立异地容灾中心，通过数据实时复制技术，将本地数据同步到异地容灾中心。异地容灾中心应具备与本地数据中心相同的存储能力和处理能力，能够在本地数据中心发生重大灾难（如自然灾害、火灾等）时，迅速接管业务，确保数据不丢失，业务不中断。同时，宜定期进行异地容灾演练，以检验和提升容灾系统的有效性和可靠性；
- d) 冗余策略：应基于多冗余策略进行数据备份，可采用磁带、光盘、磁盘镜像、磁盘冷备、热备等技术实现，遵循至少保存三个数据副本、使用两种不同存储介质、将其中一个副本异地存放的基本原则，确保所有副本数据可恢复且无损坏。

6.2.3.1.4 扩展性

- a) 节点扩展：存储系统应能方便地进行节点扩展。在不影响现有业务正常运行的前提下，应可在线添加存储节点，扩展存储容量和处理能力。存储系统应具备良好的线性扩展能力，即随着节点数量的增加，系统的整体性能和存储容量应能够按比例提升；
- b) 加密算法扩展：存储系统应具备加密算法扩展能力。应能够方便地升级或更换加密算法，以适应新的安全标准和要求，保障数据在存储过程中的安全性。同时，在更换加密算法时，应确保对现有数据的加密和解密操作不受影响，数据的完整性和可用性应得到保障。

6.2.3.1.5 兼容性

- a) 操作系统：存储系统应兼容主流的操作系统，包括但不限于国内外主流操作系统。应确保在不同操作系统环境下，业务系统能够无缝地与存储系统进行交互，实现数据的存储和访问，避免因操作系统兼容性问题导致的数据读写错误或性能下降；

- b) 编程语言：应支持多种常用编程语言对存储系统的访问接口。通过提供丰富的编程接口和开发工具包（SDK），应方便开发人员在不同的应用开发场景中，根据业务需求灵活地使用存储系统，实现数据的高效存储和处理；
- c) 数据库兼容：应能够与常见的国内外关系型数据库和非关系型数据库良好兼容。存储系统应能满足不同数据库对存储性能、数据一致性等方面的要求，支持数据库的数据存储、备份、恢复等操作，确保数据库应用在存储层面的稳定运行。

6.2.3.1.6 成本效率

- a) 存储分层：应依据数据的访问频率、重要性和保存期限等因素，实施存储分层策略。将高频访问、重要性高的数据存储在高性能存储介质上，以保障业务性能；将低频访问、历史久远的数据迁移至低成本的存储介质中，降低存储成本。通过合理的存储分层，应在满足业务需求的同时，优化存储资源的利用效率，降低总体存储成本；
- b) 压缩去重：应采用数据压缩和去重技术，减少数据的存储空间占用。利用去重技术识别并删除重复的数据块，提升存储效率，降低存储成本，在不影响数据使用的前提下，应实现存储资源的高效利用。

6.2.3.2 分布式文件系统

分布式文件系统的功能要求如下：

- a) 应支持文件的上传、下载、读写、复制、移动、删除、访问控制等；
- b) 应支持文件的搜索、批量操作、回收站、快照等；
- c) 应支持文件容错机制和系统高可用机制，包括备份、容灾、系统快速恢复等；
- d) 应支持文件数据的校验和同步，保证数据的完整性与一致性；
- e) 应支持金融大数据平台对国内主流操作系统的兼容性；
- f) 应支持接口对分布式文件系统进行基本操作；
- g) 宜支持对不同型号CPU的兼容性；
- h) 宜支持弹性扩展功能、动态添加操作以及删除节点操作；
- i) 宜支持分级存储，同一节点支持配备不同类型磁盘。

6.2.3.3 分布式列数据库

分布式列数据库的功能要求如下：

- a) 应支持数据以键值形式进行存储；
- b) 应支持基于表级和列级的用户权限管理；
- c) 应支持数据压缩；
- d) 应支持多级索引，支持除行键（Rowkey）以外列值进行索引；
- e) 宜支持双集群双读，保证查询的稳定，降低随机读毛刺现象；
- f) 宜支持区域服务器（RegionServer）单节点多实例部署，提高资源利用率。

6.2.3.4 分布式图数据库

分布式图数据库的功能要求如下：

- a) 应支持由节点及边组成（即节点间关系）的数据模型；
- b) 应支持图查询、图遍历及图分析；
- c) 应支持可视化查询；
- d) 应支持主流开发接口和主流图查询语言；
- e) 应支持单节点多标签的数据模型；
- f) 宜支持图与图之间逻辑隔离；
- g) 宜支持单节点、多节点带有过滤条件的多跳查询；
- h) 宜支持事务机制，提供事务回滚、事务提交操作。

6.2.3.5 分布式关系型数据库

分布式关系型数据库的功能要求如下：

- a) 应支持结构化数据的存储;
- b) 应支持SQL实现数据的查询, 包括并发事务控制、存储过程等;
- c) 应支持多表关联;
- d) 应支持多副本, 支持主副本与从副本之间的数据同步;
- e) 应支持与ODBC、JDBC接口的兼容性;
- f) 应支持对数据库进行全备、部分备、增备;
- g) 应支持事务的原子性、一致性、隔离性、持久性;
- h) 应具备数据库用户管理和权限控制能力。

6.2.3.6 向量数据库

向量数据库的功能要求如下:

- a) 应支持大规模向量检索能力;
- b) 应支持多样化检索策略;
- c) 应支持混合检索, 支持文本、时间、空间等结构化字段与向量字段的混合检索;
- d) 应支持包括欧式距离、汉明距离、余弦距离等多种距离度量方式;
- e) 应提供标准化访问接口。

6.2.3.7 时序数据库

时序数据库的功能要求如下:

- a) 应支持高吞吐持续写入, 具备高效数据压缩机制;
- b) 应支持多精度存储策略;
- c) 应支持流式处理能力, 支持高效时间窗口计算;
- d) 应支持标准JDBC接口, 支持类SQL查询语言快速分析数据;
- e) 宜支持在线扩容。

6.3 数据处理

6.3.1 概述

数据处理是基于市场分析、业务优化、风险管控等需求对数据进行清洗、转换、分析、挖掘等操作, 涵盖数据研发管理、数据交付管理、数据运维管理以及数据安全合规管理的完整过程。

6.3.2 管理要求

6.3.2.1 数据研发管理

6.3.2.1.1 概述

数据研发管理应以研发治理一体化为目标, 构建标准化的数据开发流程, 确保研发过程安全合规、质量可控。

6.3.2.1.2 需求管理

- a) 应建立统一的需求管理平台, 规范数据处理需求提出、分析、评审、确认的全过程管理, 明确需求管理责任部门与岗位, 建立跨部门协作功能, 建立需求优先级动态调整机制;
- b) 需求分析应明确数据处理目的、涉及数据范围、技术实现要求、量化需求价值指标、安全保障措施及风险评估方案;
- c) 应建立需求反馈闭环机制, 定期回顾需求实现效果, 及时发现需求变更的必要性, 保障数据处理需求管理的动态性和有效性;
- d) 应明确需求变更的管理流程, 规范需求变更的审批和记录, 保障需求变更过程透明可控;
- e) 应实施需求追踪管理, 确保需求从提出到落实全过程可追溯, 避免需求遗漏或误解;
- f) 针对人工智能应用场景的数据需求, 应明确模型训练所需的数据规模、数据多样性、数据标注精度要求。

6.3.2.1.3 设计管理

- a) 应制定数据处理设计规范，明确数据模型、数据流程和接口定义要求，形成统一的设计标准；
- b) 设计阶段应进行安全风险评估和合规审查，确保数据处理方案符合安全和监管要求；
- c) 应详细记录设计方案及变更内容，保障设计方案的可审计性和持续维护性；
- d) 应定期组织设计评审，确保数据处理设计的有效性、合规性和安全性；
- e) 应明确设计文档和成果的管理制度，建立设计资料的版本控制和变更记录。

6.3.2.1.4 开发管理

- a) 应明确数据处理开发的编码标准、代码审查制度，包括安全漏洞检测和依赖包安全审查；
- b) 开发阶段应实施严格的单元测试、安全测试和性能测试，保障开发质量；
- c) 代码上线前应进行安全审计和风险评估，防止安全漏洞进入生产环境；
- d) 应明确开发过程中的日志记录与异常管理机制，及时发现和修复开发过程中的问题；
- e) 应定期回顾开发质量与效率，持续改进开发管理流程和工具；
- f) 人工智能数据开发过程中，应建立数据标注、增强、缩减、合成的标准化流程。

6.3.2.1.5 测试管理

- a) 应建立多维度的测试计划，包括功能、性能、安全和兼容性测试；
- b) 应实施独立测试环境，避免风险传导至生产环境；
- c) 应明确测试过程的自动化管理，确保测试覆盖度与效率；
- d) 测试结果和缺陷修复过程应完整记录，提供后续审计依据；
- e) 应定期进行测试结果分析，优化测试方案和提高测试质量；
- f) 应建立人工智能数据集的基准测试体系，通过标准测试集评估模型训练效果。

6.3.2.2 数据交付管理

6.3.2.2.1 概述

数据交付管理应以提升交付效率和质量为目标，建设持续测试和交付的能力，提升数据产品交付的自动化水平，加快交付速度，提高交付质量。

6.3.2.2.2 配置管理

- a) 应建立统一配置管理平台，详细记录所有配置项的变更和历史；
- b) 配置变更前应进行风险评估和审批，确保配置变更安全；
- c) 配置项管理应提供快速恢复和回退机制，应对异常情况；
- d) 应定期开展配置项审计，确保配置项的准确性和完整性；
- e) 应明确配置权限的管理流程，避免配置的非授权变更。

6.3.2.2.3 部署与发布管理

- a) 部署发布流程应明确变更申请、审批、实施与回退计划；
- b) 应详细记录部署实施和发布后的系统验证过程；
- c) 部署发布前后应进行全面的安全风险和性能评估；
- d) 发布后应实施有效的监控机制，及时发现并处理系统异常；
- e) 应定期开展发布后回顾分析，持续优化部署发布流程。

6.3.2.3 数据运维管理

6.3.2.3.1 概述

数据运维管理应以全面立体的持续监控、发现、处理数据问题为目标，构建全链路可观测能力，确保数据处理系统稳定运行。

6.3.2.3.2 监控管理

- a) 应实施全面监控，涵盖系统资源、任务运行和性能指标；
- b) 应建立统一的告警管理平台，明确告警响应和处理流程；

- c) 应定期审计监控日志，及时发现监控盲点和监控误报情况；
- d) 应持续优化监控规则，提高异常情况的准确检测和响应速度；
- e) 应明确监控权限的管理制度，防范非授权的监控数据访问；
- f) 应对人工智能训练数据集的使用情况进行监控，包括数据访问频率、模型训练任务关联、数据使用分布。

6.3.2.3.3 资源管理

- a) 应明确资源动态监测和容量规划的管理制度；
- b) 应提供资源自动扩缩容策略，确保资源使用的高效性；
- c) 应建立资源使用情况的定期分析与报告机制，优化资源分配；
- d) 应实施资源使用权限管理，防范资源的非授权使用；
- e) 应定期开展资源配置审计，确保资源利用的合规性和经济性。

6.3.2.3.4 变更管理

- a) 应明确变更管理的申请、审批和实施流程，严格控制变更风险；
- b) 变更实施应详细记录操作日志，提供追溯和审计功能；
- c) 应提供变更的快速回退能力，应对变更失败或异常情况；
- d) 应定期回顾变更实施情况，分析变更成功率和失败原因；
- e) 应明确变更管理权限，防止非授权变更行为的发生。

6.3.2.3.5 异常管理

- a) 应建立异常事件快速响应和处置流程，明确异常处理责任；
- b) 应记录异常发现、诊断、处置和恢复的详细日志信息；
- c) 应定期进行异常处置效果的评估分析，持续提升异常管理水平；
- d) 应实施异常事件的实时告警机制，确保快速发现和响应异常情况；
- e) 应提供异常事件处理的知识库，积累异常处理经验和方法。

6.3.2.3.6 持续优化

- a) 应定期开展数据处理运维复盘，分析问题根因，提出优化措施；
- b) 应持续分析和优化系统性能，降低数据处理的延迟和资源消耗；
- c) 应明确持续优化的责任部门和人员，建立持续优化的反馈和评估机制，跟踪优化措施的落实情况；
- d) 应持续更新和优化运维文档与操作手册，确保运维操作规范统一；
- e) 应建立人工智能数据集的反馈优化机制，持续改进数据标注、增强策略。

6.3.2.4 数据安全合规管理

6.3.2.4.1 概述

数据安全合规管理需明确安全管理和合规管理职责，确保数据处理活动合法合规。

6.3.2.4.2 安全管理

- a) 应建立数据处理安全分类分级制度，根据数据敏感程度实施差异化安全控制措施；
- b) 应定期进行数据处理安全风险评估，识别和防范数据处理过程中的安全风险；
- c) 应在数据处理过程中实施严格的数据访问控制、加密和脱敏技术；
- d) 应建立数据处理安全审计机制，明确审计内容、流程和责任主体，确保安全审计全面覆盖；
- e) 应定期开展数据处理安全意识教育培训，提升全员数据安全意识。

6.3.2.4.3 合规管理

- a) 应定期开展数据处理活动的合规性自查和审计；
- b) 应建立数据处理合规管理制度，明确违规行为的认定标准和处理机制；
- c) 应建立数据处理安全事件报告机制，明确报告内容、流程和时限要求；

- d) 应建立法律法规跟踪机制，及时更新数据处理合规管理要求；
- e) 应实施数据处理操作合规性监督，确保处理过程持续符合法律法规和监管要求；
- f) 应建立人工智能数据处理的伦理审查机制，防范算法歧视风险。

6.3.3 技术要求

6.3.3.1 批处理

- a) 应支持多种数据类型的离线分析，包括结构化数据、半结构化数据、非结构化数据；
- b) 应支持离线计算任务进度与状态的实时上报；
- c) 宜支持执行多节点离线任务联动；
- d) 宜支持多种语言分析任务的开发接口；
- e) 宜支持分散-聚集的处理方式；
- f) 支持多租户管理，提供弹性的分布式资源共享，保障数据、资源、应用间的安全隔离；
- g) 应支持基于标准化数据库语言的开发；
- h) 应支持批处理/交互式融合分析、跨源数据分析、跨域数据协同分析能力；
- i) 应支持数据并行计算，且集群性能保持稳定；
- j) 宜支持多种计算优化策略，包括但不限于基于代价优化、基于规则优化、基于物化视图优化、SQL过程间优化器，提高计算性能；
- k) 应支持批处理任务的运维监控查看，包括任务运行消耗资源，任务执行计划等；
- l) 宜支持容器化资源管理与调度技术，可以进行资源动态调整，应用租户资源物理隔离。

6.3.3.2 流处理

- a) 应支持数据的实时获取、处理、输出和持久化；
- b) 应支持用户级别的访问控制；
- c) 应支持对消息处理任务进行全生命周期管理，包括创建、浏览、中止、激活、去激活等；
- d) 应支持多种窗口方式，包括但不限于滚动窗口、滑动窗口；
- e) 应提供SQL或类SQL的数据操作接口；
- f) 应支持在出现故障情况下，使用容错机制处理事件；
- g) 应具备确保消息不丢失、不重复的数据容错处理能力，应具备高容错能力，如节点、进程等出现异常时，能够重新部署该处理单元；
- h) 宜支持基于多种分布式流处理引擎的查询语言；
- i) 宜提供基于流计算引擎的画布、SQL和JAR等多种流计算开发模式；
- j) 应支持实时统计分析和复杂事件处理；
- k) 应支持流量控制功能，防止数据发生阻塞，结合数据流量弹性扩缩容，以满足不同场景的业务需求；
- l) 宜支持流批一体化统一开发框架；
- m) 应支持流处理引擎对实时数据与其他数据关联分析能力；
- n) 宜支持基于快照的表级别增量数据读取功能；
- o) 宜支持实时增量聚合功能，能够基于增量数据进行聚合处理，快速提供计算结果。

6.3.3.3 图计算

- a) 应支持同步计算模型或异步计算模型编写迭代算法；
- b) 应支持基于属性图模型的图数据表达，包含结点或边上的标签和属性类型定义；
- c) 应支持内置常用图指标计算功能，以描述图的拓扑结构特征；
- d) 应支持实现水平扩展的分布式图计算和图查询；
- e) 宜支持索引，提供在线图分析和图查询功能；
- f) 宜支持图复制能力；
- g) 宜支持社区发现、最短路、LPA等多种图算法能力。

6.3.3.4 内存计算

- a) 应支持负载均衡和水平扩展；
- b) 应支持对多种数据类型的离线分析，包括结构化数据、半结构化数据、非结构化数据；
- c) 应支持高度抽象算子，快速构建分布式的数据处理应用；
- d) 宜支持标准SQL语法；
- e) 宜支持读取非关系型数据库数据；
- f) 应支持数据存储在内存中，同时支持内存加SSD的混合存储架构，使得数据不仅能存储在内存中也能存储在SSD中，提升计算性能；
- g) 应支持数据列式存储，通过内存加速分析，秒级别响应，对数据进行交互式探索挖掘；
- h) 宜支持行列混存的数据格式，支持向量化计算引擎。

6.3.3.5 批流融合计算

- a) 宜支持统一查询SQL语言；
- b) 宜支持多种场景下的流式SQL，如位置信息分析等；
- c) 宜支持常用时间窗口，包括滚动窗口、滑动窗口等；
- d) 宜支持基于SQL的批流数据的模式识别；
- e) 宜支持事件驱动的流处理，降低处理延迟；
- f) 宜支持处理乱序事件流、窗口计算、CEP等；
- g) 宜支持对复杂任务的调度，如支持深度学习的训练、MPI任务等；
- h) 宜支持批流使用统一的SQL语法，再通过底层分别开启流与批作业，简单易上手；
- i) 宜支持同时运行批任务和流任务，满足计算资源不足场景下的实时需求；
- j) 宜支持批流共享存储，流作业实时数据入库即可查询，兼容数据的一致性和实时性，支持多种存储结构元数据统一管理，并通过SQL接口进行统一读写；
- k) 宜支持统一数据采集工具，统一配置接入实时和离线数据，包括日志方式采集多种数据库的CDC数据，历史数据与增量数据都走实时链路进行同步。

6.3.3.6 向量计算

- a) 应支持存储稠密向量、稀疏向量，并且能够基于稠密向量、稀疏向量进行距离计算和近似检索；
- b) 应支持2至4096维度的稠密向量创建，包括2至4096维度的索引创建和查询；
- c) 应支持至少字符串、浮点数、整数、布尔值、时间戳等基础标量数值类型；
- d) 应支持向量数据存储压缩，支持压缩算法；
- e) 应支持近似检索和精确检索，如近邻检索算法，支持召回率为100%的向量检索；
- f) 应支持标量与向量的融合查询，包含向量检索和标量过滤同时检索，标量检索支持常用的运算符，包括比较运算符、逻辑运算符等；
- g) 应支持标量数据和向量数据的增删改查，并且保证操作的原子性要求；
- h) 应支持主键查询，根据主键查询对应的标量和向量数据；
- i) 应支持数据导入导出，数据按单条、批量等至少一种导入/导出方式；
- j) 应支持内积、欧氏距离、余弦等三种相似距离计算；
- k) 应支持至少一种索引的构建与删除，如标量索引、向量索引等；
- l) 应支持如强一致性、最终一致性、会话一致性、有界一致性等至少一种数据一致性；
- m) 应支持全文索引、向量索引、标量索引的创建，并支持全文、向量融合检索的能力；
- n) 应支持节点动态扩缩容，包括计算或存储能力的动态扩容，动态扩容方式可以为水平扩展或垂直扩展；
- o) 应支持多模态数据向量化能力；
- p) 应支持对关键指标进行监控与存储，如资源利用率、查询延迟、吞吐量等；
- q) 应支持告警通知，支持用户自定义告警阈值。

6.4 数据服务

6.4.1 概述

数据服务是指通过标准化接口、规范化流程及模块化技术，将数据转化为可复用、可组合的服务形态，向内部业务系统、外部合作机构或终端用户提供数据支持的功能层。

6.4.2 管理要求

6.4.2.1 数据服务规范管理

- a) 应制定切实可行与组织发展阶段相匹配的数据服务规划，明确数据服务需求目标；
- b) 应区分业务等级，明确可降级的非关键数据项，保障关键数据项的及时可用；
- c) 应制定公司级的数据服务定义规范，明确名称、接口、参数、数据格式及服务等级等核心要素的标准；
- d) 数据服务工具应符合分布式架构、合规性要求，具备监控与优化功能；
- e) 宜建立数据服务评价机制，制定相应的指标体系对服务成效进行评价；
- f) 宜建立覆盖数据服务全生命周期管理体系的相关规范；
- g) 应界定数据服务应用场景边界：数据服务覆盖证券期货业机构内部业务系统间跨节点数据传输场景；跨机构数据流转需以满足监管合规要求、签订正式数据共享与使用协议为前提，仅适用于经内部审批的合规合作业务（如跨机构风险联防、监管数据报送等），非通用数据服务场景，需额外履行专项合规审查程序。

6.4.2.2 数据服务资产管理

数据服务资产管理具体要求如下：

- a) 应建立服务资产目录，将接口、交换、分析等服务纳入资产清单，明确唯一标识、所属业务域、责任部门及管理岗；
- b) 应实施资产登记，记录基本信息、关联数据资产、依赖关系及生命周期状态，由责任岗动态维护更新；
- c) 应建立价值评估体系，定期评估业务价值、技术价值、经济价值，为资源投入和优化提供依据；
- d) 应实施分类分级管理，根据服务重要性、敏感程度、使用范围划分类别与级别，针对不同类别级别采取差异化管理策略；
- e) 应规范资产全生命周期状态转换流程，新增、变更、停用、注销需经审批并记录，避免无效资产占用资源；
- f) 应定期开展资产盘点，核查实际状态与登记信息一致性，清理无效或冗余资产，优化结构。

6.4.2.3 数据服务权限管理

数据服务权限管理旨在通过严格的权限控制机制，防止未授权访问和数据泄露，具体要求如下：

- a) 应遵循“最小权限”“按需分配”原则，按岗位职责与业务需求授予必要权限，明确可访问服务、操作类型及数据范围；
- b) 应建立权限申请、审批、授予流程，明确各环节责任人及标准；申请需说明使用目的及必要性，审批过程记录存档，确保可追溯；
- c) 应实施角色化管理，按业务场景与职责定义标准化角色，通过角色分配权限，简化流程、降低成本；
- d) 应建立权限定期复核机制，及时回收岗位变动、业务终止等场景下的冗余权限，避免权限滥用；
- e) 应记录权限变更历史，包括时间、原因、操作人、审批人及变更内容，确保权限变更可审计；
- f) 对于涉及高敏感数据的服务，应采取多因素认证、访问IP限制、操作日志实时审计等增强措施，提升数据安全性。

6.4.2.4 安全合规管理

数据服务的安全合规性要求如下：

- a) 遵守相关法律法规和监管要求，确保数据服务的合规性，防止出现违规行为和风险；
- b) 通过数据分类分级、跨境传输管控、定期合规审计，使其符合行业及地区法规要求；

- c) 记录全链路日志，并通过操作溯源、版本控制等完整记录数据流动和操作历史，保障数据服务的可追溯性；
- d) 签订数据保护协议，明确数据用户的保密义务等；
- e) 响应用户对个人信息的相关权力，涉及个人信息时支持用户行使删除权和查询权等；
- f) 跨节点数据传输需纳入监控，建立链路异常检测机制，对传输中断、数据篡改、非授权访问监测告警，确保可追溯；
- g) 整体遵循“数据分类分级，最小必要，全程可控”原则，保障数据服务过程中的数据安全。

6.4.3 技术要求

6.4.3.1 基础要求

6.4.3.1.1 及时性

数据服务及时性的建议如下：

- a) 应支持数据预热，高峰前预加载热点数据；
- b) 应按业务场景分级设响应时效，根据不同时效性要求进行数据更新；
- c) 应支持业务高峰动态扩容资源，保障时效稳定；
- d) 应定期校验时效达标情况，未达标及时优化。

6.4.3.1.2 完整性

数据服务的完整性能力应具备：

- a) 数据同步完整性校验规则，数据服务请求方应具备根据校验规则校验数据完整性的条件（如哈希校验、数字签名等）；
- b) 数据传输防丢失机制，数据服务请求方应具备数据传输防丢失校验机制，如采用可靠传输协议、确认机制、重试机制和多路冗余等；
- c) 数据服务完整性保障措施，数据服务应具备对输入输出的完整性保障。如输入参数的完整性保障（边界值、值域校验等）、输出数据的完整性校验算法及机制。

6.4.3.1.3 可靠性

数据服务的可靠性要求如下：

- a) 应支持展示所有已发布服务及实例健康状态、运行状态等信息；
- b) 应支持服务进程挂起后自恢复的能力；
- c) 应支持对服务进行升级，及升级失败后的自动回滚；
- d) 应支持检测服务的可用性及发现问题时发送告警；
- e) 应支持停止或重启某个服务时，上层服务联动一起停止或重启或给出提示；
- f) 应支持通过各种策略（如分批、主备依次等）重启单个服务，同时不中断业务；
- g) 应支持服务和数据自动恢复到新增或者是更换之后的服务器；
- h) 宜支持服务冗余部署；
- i) 宜具备集群互备能力。

6.4.3.1.4 扩展性

数据服务的扩展性要求如下：

- a) 应支持对分布式服务进行在线水平扩容或扩容的能力；
- b) 应支持对分布式服务进行离线水平扩容或扩容的能力；
- c) 应尽可能选择标准化的部件，利于灵活替换和容量扩展；
- d) 应支持在容量扩展时，服务的负载性能同步得到提升。

6.4.3.2 数据接口服务

数据接口服务为跨系统数据交互提供标准化通道，技术要求如下：

- a) 宜采用RESTful等行业通用协议，数据传输格式宜使用JSON、XML等标准化格式；接口命名与参数定义应符合JR/T 0304 《证券期货业基础数据元规范》，确保与行业通用数据单元属性规范一致；
注：REST指的是一组架构约束条件和原则，满足这些约束条件和原则的应用程序或设计称为RESTful。
- b) 宜建立身份认证机制，支持数字证书、API密钥等方式；敏感数据接口宜部署请求频率限制、IP白名单等防护措施；
- c) 传输过程宜采用TLS 1.3及以上加密协议，支持哈希校验、数字签名等数据完整性验证方式；
- d) 应提供完整接口文档，支持版本管理，旧版本接口宜预留合理过渡期。

6.4.3.3 数据交换服务

数据交换服务宜根据业务需求提供多样化同步能力，确保数据一致性与时效性：

- a) 宜支持实时、准实时、批量等多种同步模式，实时同步宜采用CDC技术结合消息队列，支持断点续传与数据重放；批量同步宜按固定周期，数据交换前宜进行数据一致性校验；
- b) 异构系统同步宜提供数据格式转换、编码适配能力，宜通过中间转换层实现不同数据模型的映射；
- c) 宜具备同步监控与异常处理机制，同步日志需记录时间、数据量、异常信息，留存大于等于6个月；
- d) 宜支持同步策略灵活配置，可调整频率与并发数；
- e) 宜具备水平扩展能力，提升数据同步吞吐量。

6.4.3.4 数据应用服务

数据应用服务应基于标准化的数据服务能力，具体要求如下：

- a) 应支持与核心业务系统的无缝集成，提供标准化的数据接入接口，确保数据应用的实时性和准确性；
- b) 宜具备灵活的应用配置能力，支持根据业务需求快速构建个性化的数据应用；
- c) 应保障数据应用过程中的数据安全性，对敏感数据应用实施严格的权限控制和脱敏处理，防止数据泄露；
- d) 宜支持数据应用效果的跟踪与评估，提供应用指标分析功能，如应用频率、数据使用量、业务提升效果等；
- e) 宜支持低代码或无代码开发模式，降低使用门槛，提升数据应用的开发效率；
- f) 应符合证券期货业相关业务规则，如客户适当性管理、风险控制指标等要求，确保数据应用的合规性；
- g) 跨机构数据共享的应用场景，宜采用隐私计算技术，实现数据可用不可见，降低泄露风险；
- h) 跨机构数据交换应部署双向身份认证、细粒度权限控制，全链路日志留存大于等于12个月。

6.4.3.5 数据分析服务

数据应用服务应支撑业务场景，技术要求如下：

- a) 宜具备多种数据分析模型和算法，满足不同业务场景的分析需求；
- b) 宜提供支持SQL、编程语言的标化工具接口，支持分布式计算框架处理千亿级数据；
- c) 应保障数据分析过程中的数据安全性，对分析数据进行分类分级管理，实施必要的加密和脱敏处理，防止敏感数据泄露；
- d) 应支持数据分析结果的可视化展示，提供图表、报表等多种展示方式，便于理解和使用分析结果；
- e) 宜支持实时分析与离线分析相结合的模式，满足不同时效性要求的业务场景；
- f) 对涉及客户隐私、交易敏感信息的分析场景，应采用隐私计算技术，确保数据分析过程中原始数据不落地、不泄露，同时满足结果可用性；

6.5 数据退役

6.5.1 概述

数据退役是对历史数据的管理，根据法律法规、业务、技术等方面需求对历史数据的保留和销毁，执行历史数据的归档、迁移和销毁工作，确保组织对历史数据的管理符合外部监管机构和内部业务用户的需求，而非仅满足信息技术需求。

6.5.2 管理要求

6.5.2.1 数据退役需求分析

- a) 应向公司管理层、各领域业务用户调研和收集内部、外部对于数据退役的需求；
- b) 应明确数据分类分级要求；
- c) 应明确外部监管和内部数据应用的数据保留和销毁要求；
- d) 宜兼顾信息技术对存储容量、访问速度、存储成本等需求。

6.5.2.2 数据退役设计

- a) 应综合考虑合规、业务和信息技术需求，设计数据退役标准和执行流程；
- b) 应明确不同类型数据的退役策略，包括保留期限、保留方式、销毁方式等；
- c) 应建立数据退役的工作流程和操作规程，确保数据退役符合标准和流程规范。

6.5.2.3 数据退役执行

- a) 应根据数据退役设计方案执行数据退役操作，完成数据的归档、迁移和销毁等工作；
- b) 应满足法规、业务和技术需要，同时根据需要更新数据退役设计。

6.5.2.4 数据退役检查

- a) 应对归档数据制定数据恢复检查机制，定期检查退役数据状态；
- b) 应进行完整性和可用性验证，确保数据在需要时可恢复；
- c) 应根据业务管理或监管需要，对归档数据的访问请求进行管理，并按需恢复相关数据以供使用；
- d) 应对销毁数据进行抽查，确保销毁方式符合要求，销毁数据不可恢复。

6.5.2.5 合规管理与审计监督

- a) 应建立符合《证券基金经营机构信息技术管理办法》（证监会令第152号）的安全保障措施，实施数据退役全流程记录与审计，严格执行访问控制和权限管理；
- b) 应定期评估数据退役策略的有效性和合规性。

6.5.3 技术要求

6.5.3.1 数据归档

- a) 应采用开放、稳定、长期可读的数据格式；
- b) 应包含完整、结构化的元数据，描述数据内容、背景、结构、来源、保留期限等关键信息，确保归档数据可被理解和正确解读；
- c) 应根据数据价值、访问频率、合规要求和成本效益，选择合适的存储介质（如磁带、磁盘、云存储等）；
- d) 采用磁带归档时，应在适当的温度湿度环境中存储，并定期进行数据读取验证；
- e) 采用磁盘归档时，宜配置RAID阵列或其他冗余技术提高数据可靠性，同时配备有效的冷却系统；
- f) 采用云存储归档时，应支持传输中和静态数据加密，确保数据在云环境中的安全性。

6.5.3.2 数据删除

6.5.3.2.1 逻辑删除

- a) 应通过修改数据的状态或标记将数据标记为不可用，或者调整文件入口标识使操作系统无法正常识别到数据文件，原数据仍保留在存储介质中；

- b) 采用修改数据状态或标记删除数据时，应明确定义数据的状态标记方式（如设置“已删除”标志位、更新删除时间戳等），以确保数据在系统中被正确标识为不可用；
- c) 逻辑删除操作后，应更新访问控制机制，确保任何用户或程序无法通过正常途径访问已被逻辑删除的数据。

6.5.3.2.2 物理删除

- a) 应通过对存储介质中的数据进行多次覆写、格式化或ATA Secure Erase指令等方式，使存储区域中的数据被真正清除。必要时，可采用数据删除工具；
注：ATA Secure Erase是一种设备级数据安全清除指令，由硬盘固件在设备内部执行。
- b) 采用覆写的方式删除数据时，应明确数据填充方式与覆写次数，并确保完全覆盖存储数据区域；
- c) 采用格式化方式删除数据，宜使用低级格式化的方式将磁盘清空，恢复出厂状态；
- d) 删除存储在固态硬盘中的数据时，宜采用ATA Secure Erase指令的方式进行；
- e) 执行物理删除操作后，宜通过数据恢复工具进行数据的尝试恢复及检查，验证数据删除效果；
- f) 针对物理删除失败的存储介质，应采用介质销毁的方式进一步处理。

6.5.3.3 介质销毁

- a) 应根据存储介质的类型，选择适当的介质销毁技术，如消磁、粉碎、焚毁等；
- b) 应在固定场所执行介质销毁，对介质销毁过程进行监控与记录；
- c) 应对销毁的介质进行抽查，确保介质销毁的效果。