



中华人民共和国金融行业标准

JR/T XXXXX—XXXX

证券期货业大数据灾备指南

Guidelines for big data disaster recovery for 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 III

引言 IV

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 总体原则 2

5.1 合规性与战略性原则 2

5.2 业务导向原则 2

5.3 分级分类原则 2

5.4 技术先进性原则 2

5.5 经济适用性原则 2

5.6 持续运维原则 2

6 技术参考模型 3

6.1 灾备体系架构 3

6.2 大数据灾备对象 3

7 灾备能力等级要求 3

7.1 等级一：数据冷备份级 3

7.2 等级二：数据热备份级 3

7.3 等级三：应用温备级 3

7.4 等级四：应用热备级 4

7.5 等级五：业务持续级 4

8 实施流程 4

8.1 调研与评估阶段 4

8.2 方案设计阶段 4

8.3 建设与部署阶段 5

8.4 测试验证阶段 5

8.5 运维优化阶段 5

9 关键技术要求 5

9.1 数据复制与同步 6

9.2 网络连接 6

9.3 计算与存储 6

9.4 切换与恢复 6

9.5 安全与合规 6

10 组织与运维管理 6

10.1 组织架构 6

10.2	制度流程	7
10.3	运维演练	7
10.4	培训与意识	7
附录 A	（资料性附录）大数据平台灾备 RTO/RPO 指标建议表	8
附录 B	（资料性附录）常见大数据组件灾备技术方案	9
附录 C	（资料性附录）数据一致性校验方法	10
C.1	校验和比对法	10
C.2	记录数比对法	10
C.3	关键字段抽样比对法	10
C.4	业务规则验证法	10
参考文献	11

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范》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：兴业证券股份有限公司、深圳证券交易所、北京证券交易所、郑州商品交易所、中金所数据有限公司、中信建投证券股份有限公司、南京证券股份有限公司、阿里云计算有限公司。

本文件主要起草人：蒋剑飞、林宇、林泉、张承芳、郑煊、季常青、李建、邵辰、高森、车江涛、胡志华、郭枫、刘远。

引 言

随着证券期货业数字化转型深入，大数据平台已成为支撑经营决策、风险管理、客户服务、监管合规的核心基础设施。在大数据时代，数据已成为证券期货行业的核心资产和命脉。交易、行情、客户信息、风控模型等海量数据不仅是日常运营的基础，更是构成企业核心竞争力的关键。因此，针对大数据平台进行高规格的灾备建设，已从“技术备选项”变为“行业必选项”，其必要性极为突出。为应对自然灾害、设备故障、网络攻击等可能导致的数据丢失和服务中断风险，建设健壮可靠的大数据灾难备份体系至关重要。大数据灾备建设是证券期货行业应对内生风险和外生威胁、满足监管合规、保障投资者利益并支撑未来发展的战略性、基础性工程，必须得到最高程度的重视和投入。本文件行业实际出发，规范大数据灾备建设的技术要求和管理流程，为行业机构提供全面、可操作的指南。

证券期货业大数据灾备指南

1 范围

本文件定了证券期货业机构开展大数据系统灾难备份（以下简称“灾备”）建设的总体原则、技术参考模型、灾备等级要求、组织管理、实施流程、运维演练及关键技术建议。

本文件适用于证券、期货、基金公司等持牌经营机构，以及证券期货信息服务提供商等机构的大数据平台、数据中台及重要数据应用的灾备体系建设与改造。

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。

JR/T 0059—2024 证券期货业信息系统备份能力规范

JR/T 0236—2021 金融大数据 术语

JR/T 0237—2021 金融大数据平台总体技术要求

3 术语和定义

JR/T 0059—2024、JR/T 0236—2021、JR/T 0237—2021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、变量多的特性，需要一种易扩展的技术来有效存储、处理、管理和分析。

[来源:JR/T 0236—2021, 3.1]

3.2

数据湖 data lake

数据湖是一个集中式存储，允许存储海量的结构化和非结构化数据。按原样存储数据，并允许不同类型的分析，包括大数据处理、实时分析和机器学习等。

3.3

恢复点目标 recovery point objective; RPO

信息系统和数据必须恢复到的时间点要求。

[来源:JR/T 0059—2024, 3.3]

3.4

恢复时间目标 recovery time objective; RTO

信息系统从停顿到必须恢复的时间要求。

[来源:JR/T 0059—2024, 3.4]

3.5

数据一致性 data consistency

数据一致性是指在分布式系统中，多个节点中存储的数据副本之间保持相同的状态和值，以确保数据的准确性和可靠性。

3.6

灾备能力等级 disaster recovery capability level

根据恢复能力和资源保障要求划分的灾难恢复级别。

4 缩略语

BIA - 业务影响分析 (Business Impact Analysis)

RTO - 恢复时间目标 (Recovery Time Objective)

RPO - 恢复点目标 (Recovery Point Objective)

DNS - 域名系统 (Domain Name System)

5 总体原则

5.1 合规性与战略性原则

大数据灾备建设应遵循国家法律法规、网络安全要求及行业监管规定，纳入机构整体发展战略和业务连续性管理体系，获得高级管理层的支持和承诺。

5.2 业务导向原则

以保障关键业务连续运营为核心目标，基于业务影响分析确定灾备保护优先级和恢复指标，确保资源投入与业务价值相匹配。

5.3 分级分类原则

根据数据敏感度、业务重要性和监管要求，对大数据资产进行分类分级，实施差异化的灾备策略，优化资源配置。

5.4 技术先进性原则

采用行业主流、成熟稳定的技术方案和产品工具，鼓励使用自动化、智能化、云原生技术提升灾备效率和可靠性。

5.5 经济适用性原则

综合考虑风险承受能力、成本投入和效益回报，选择与机构规模、业务特点相适应的灾备方案，避免过度投资和资源浪费。

5.6 持续运维原则

建立完善的灾备运维管理体系和长效机制，包括定期评估、演练测试、优化改进，确保灾备体系持续有效。

6 技术参考模型

6.1 灾备体系架构

6.1.1 生产中心

承担正常业务运营的大数据平台和环境。

6.1.2 同城灾备中心

与生产中心距离适中（通常 ≤ 50 公里），实现数据同步复制，用于应对局部故障和快速恢复。

6.1.3 异地灾备中心

与生产中心距离较远（通常 ≥ 100 公里），用于应对区域性重大灾难。

6.1.4 多云/混合云架构

支持公有云、私有云和混合云环境下的灾备部署模式，确保架构灵活性和扩展性。

6.2 大数据灾备对象

6.2.1 数据层灾备

包括源数据采集层、数据存储层（数据湖、数据仓库）、数据处理层及元数据管理系统的备份保护。

6.2.2 平台层灾备

涵盖计算引擎、存储系统、资源调度、数据集成、数据开发等基础平台的容灾能力建设。

6.2.3 应用层灾备

面向数据分析、风险监控、客户画像、监管报送等关键应用的业务连续性保障。

6.2.4 管理层灾备

包括权限体系、配置管理、作业调度、运维脚本等管理元素的备份与恢复。

7 灾备能力等级要求

7.1 等级一：数据冷备份级

要求：实现周期性的数据离线备份和异地存放， $RT0 > 24$ 小时， $RPO > 24$ 小时。

适用场景：非核心数据、历史归档数据、测试开发环境。

7.2 等级二：数据热备份级

要求：实现关键数据的在线备份和异地复制，具备基本恢复能力， $RT0 \leq 24$ 小时， $RPO \leq 24$ 小时。

适用场景：重要程度较低的业务数据。

7.3 等级三：应用温备级

要求：灾备中心具备基本硬件和网络环境，数据定期同步，可在较短时间内恢复应用运行， $RT0 \leq 12$ 小时， $RP0 \leq 6$ 小时。

适用场景：一般业务应用和非实时数据处理系统。

7.4 等级四：应用热备级

要求：灾备平台实时处于就绪状态，业务数据实时或近实时同步，可快速完成业务切换， $RT0 \leq 1$ 小时， $RP0 \leq 5$ 分钟。

适用场景：核心业务系统、重要实时数据处理业务。

7.5 等级五：业务持续级

要求：采用双活或多活架构，实现业务自动负载均衡和无缝切换，接近零中断， $RT0 \approx 0$ ， $RP0 \approx 0$ 。

适用场景：极高可用性要求的实时交易和风控业务。

8 实施流程

8.1 调研与评估阶段

8.1.1 业务影响分析(BIA)

识别关键业务过程，分析中断影响，确定 $RT0/ RP0$ 指标。

8.1.2 大数据资产梳理

全面盘点数据资产，依据JR/T 0158进行数据分类分级。

8.1.3 风险分析

评估技术、管理、环境等方面存在的风险点。

8.1.4 现状评估

评估现有基础设施、技术能力、管理流程对灾备需求的支撑程度。

8.2 方案设计阶段

8.2.1 技术选型

选择适合的数据复制、备份恢复、系统监控等技术方案。

8.2.2 架构设计

设计灾备中心架构，包括网络拓扑、存储规划、系统配置等。

8.2.3 资源规划

规划所需硬件、软件、网络和人力资源。

8.2.4 预算估算

编制详细的建设预算和运维成本预测。

8.3 建设与部署阶段

8.3.1 环境准备

准备灾备中心基础设施，包括机房、网络、硬件设备等。

8.3.2 数据同步部署

部署数据复制和同步工具，配置同步策略和链路。

8.3.3 平台部署

安装配置大数据平台组件，确保与生产环境一致性。

8.3.4 应用部署

部署业务应用系统，配置连接参数和依赖服务。

8.3.5 切换工具开发

开发或配置自动化切换工具和脚本。

8.4 测试验证阶段

8.4.1 单元测试

测试数据复制、平台功能、应用功能等各个组件的有效性。

8.4.2 切换演练

模拟灾难场景，执行完整切换流程，验证RTO/RPO达标情况。

8.4.3 数据校验

采用校验工具或方法验证灾备数据的一致性和完整性。

8.4.4 性能测试

验证灾备系统在切换后的性能表现是否满足业务要求。

8.5 运维优化阶段

8.5.1 监控机制

建立常态化监控机制，对灾备系统运行状态进行实时监控。

8.5.2 定期评估

定期评估灾备体系有效性，根据业务变化和技术发展进行优化调整。

8.5.3 文档体系

完善文档体系，记录灾备配置、操作流程和应急预案。

9 关键技术要求

9.1 数据复制与同步

大数据灾备建设在数据复制与同步阶段，应遵循以下要求：

- a) 应根据RPO要求选择适当的数据复制技术，如基于存储阵列、主机卷、数据库日志、文件系统或应用层的复制方式。
- b) 重要业务数据应实现实时或近实时同步，确保数据丢失风险在可接受范围内。
- c) 应建立数据一致性保障机制，确保复制数据的完整性和可用性。
- d) 应定期验证数据复制链路的健康状态和同步性能。

9.2 网络连接

大数据灾备建设中涉及数据中心内部以及跨数据中心的网络传输，应遵循以下要求：

- a) 生产中心与灾备中心之间应建立安全、可靠、高速的网络连接，带宽满足数据同步和业务恢复需求。
- b) 应采用加密技术保护数据传输过程中的安全性，防止数据泄露。
- c) 应实现网络链路冗余，避免单点故障。

9.3 计算与存储

对大数据主集群及灾备集群的计算与存储资源，应遵循以下要求：

- a) 灾备中心计算和存储资源应能满足业务恢复需求，可根据需要采用弹性伸缩架构。
- b) 应定期进行容量规划评估，确保资源充足性。

9.4 切换与恢复

在发生灾备切换场景下，对大数据灾备切换与恢复的要求如下：

- a) 应制定详细的切换流程和操作手册，明确切换步骤和责任人。
- b) 应尽可能实现切换过程自动化，减少人工操作风险和恢复时间。
- c) 应实现DNS/IP地址的快速切换能力，确保业务访问可快速重定向到灾备中心。
- d) 应建立完善的回切流程，确保生产中心恢复后能平稳切换回主系统。

9.5 安全与合规

大数据灾备环境建设中应遵守相应的安全与合规要求，具体如下：

- a) 灾备中心安全防护等级应不低于生产中心，建立全面安全防护体系。
- b) 敏感数据在传输和存储过程中应进行安全处理，包括但不限于加密、脱敏处置等，密钥管理符合监管要求。
- c) 应建立灾备中心访问控制机制，严格限制人员权限。
- d) 对灾备环境的数据使用，应与主环境要求保持一致，根据数据的敏感性采取适当脱敏措施。
- e) 应建立完整的审计日志体系，记录所有灾备相关操作。

10 组织与运维管理

10.1 组织架构

在大数据主环境及灾备环境的建设过程中，对组织架构应做到明确要求，具体如下：

- a) 应建立明确的灾备管理组织架构，包括决策层、管理层和执行层。

- b) 应设立灾备负责人和相应岗位，明确各岗位职责和分工。

10.2 制度流程

制度和流程的相关要求如下：

- a) 应制定完善的灾备管理制度，涵盖建设、运维、演练、评估等各个环节。
- b) 应建立严格的变更管理流程，确保生产与灾备环境的配置一致性。
- c) 应制定详细的应急响应预案，明确灾难宣告条件和处置流程。

10.3 运维演练

定期进行灾备演练，有利于场景覆盖、查缺补漏，具体要求如下：

- a) 应建立灾备系统日常监控机制，定期检查系统健康状态和数据同步情况。
- b) 应制定年度演练计划，包括桌面推演、模拟切换和实战演练等多种形式。
- c) 演练应覆盖不同灾难场景，包括基础设施故障、网络中断、数据损坏、网络安全事件等。
- d) 每次演练后应进行全面评估，总结经验教训，持续优化改进。

10.4 培训与意识

培训和业务连续性意识教育的要求如下：

- a) 应定期开展灾备知识培训和技能考核，提高相关人员专业能力。
- b) 应组织全员业务连续性意识教育，提高对灾备工作重要性的认识。

附 录 A

(资料性附录)

大数据平台灾备 RTO/RPO 指标建议表

表 A.1 大数据平台灾备恢复目标建议

系统/业务类型	RTO (恢复时间目标) 建议	RPO (恢复点目标) 建议	关键业务影响说明
核心交易与结算类** (如集中交易、网上交易、融资融券、结算系统)	≤30分钟	≈0 (零数据丢失)	直接影响市场交易秩序和投资者资产安全,需最高级别保障。
实时风控与监控类** (如实时风险控制、异常交易监测、市场监管)	≤1小时	≤1分钟	风险事件发现与处置的及时性至关重要,数据延迟将导致风险敞口。
行情服务类** (如Level-1/2行情发布、指数计算)	≤15分钟	≤10秒	行情中断和数据延迟将直接影响投资决策和市场公平性。
数据分析与报表类** (如监管报送、每日业务报表、投资决策支持)	≤4小时	≤24小时	影响非实时性的监管合规和内部管理决策,允许一定时间窗口。
历史数据查询与备份类** (如历史数据归档、客户行为日志查询)	≤8小时	≤48小时	主要用于审计和事后分析,对业务连续性要求相对较低。

附 录 B
(资料性附录)
常见大数据组件灾备技术方案

大数据组件	主流灾备技术方案	优点	缺点/注意事项
分布式文件系统 (如HDFS等)	1. 跨集群数据分发 (如 DistCp定期同步) 2. NameNode 元数据高可用 (HA) + JournalNode 共享存储 3. HDFS Federation + Balancer 4. 商业解决方案 (如 Cloudera Manager 备份/恢复)	方案成熟, 社区活跃, 工具开源。	DistCp为周期性同步, RPO较大; 架构复杂度随集群规模增加。
分布式 NoSQL 数据库HBase (如Hbase, 建立在HDFS之上的、面向列的NoSQL数据库)	1. Replication (基于WAL的主从异步复制) 2. Snapshot (快照) + Export (导出) 3. CopyTable	复制功能内置, 配置灵活, 可实现跨集群实时同步。	异步复制可能带来毫秒级延迟, 存在数据一致性延迟风险。
分布式的发布-订阅消息系统(如Apache Kafka等)	1. MirrorMaker 2.0 (集群间消息同步工具) 2. 跨地域集群镜像 3. 将消息同步至备端对象存储 (如S3)	专为跨集群数据同步设计, 性能高, 延迟低。	需要精细配置以避免消息重复或丢失, 需监控同步延迟。
分布式搜索分析引擎 (如 Elasticsearch等)	1. Cross-Cluster Replication (CCR) 2. Snapshot and Restore (快照与恢复至对象存储)	CCR支持主动-被动模式的持续复制; 快照适合冷备份。	CCR对网络带宽和延迟要求较高; 快照恢复耗时较长。
流处理引擎 (如 Spark/Flink 等高性能、高吞吐、低延迟的流处理框架)	1. Checkpoint状态快照定期持久化到异地 (如HDFS、S3) 2. Savepoint手动保存作业状态 3. 代码与配置的版本化管理 (Git)	保证计算应用状态的可恢复性, 与底层存储解耦。	应用层恢复需要与数据层恢复协调一致, 流程复杂。

附 录 C
(资料性附录)
数据一致性校验方法

C.1 校验和比对法

描述：对源端和灾备端的特定数据集（如表、文件、目录）计算MD5、SHA-256等校验。
适用场景：静态数据、归档文件、备份文件的完整性验证。适用于HDFS文件、导出文件等。
优点：快速、简单，能有效发现比特位损坏。
缺点：无法识别数据顺序变化，海量数据计算耗时。

C.2 记录数比对法

描述：统计源端和灾备端数据表或分区的总记录数。
适用场景：快速初步校验，排除大规模数据丢失。
优点：非常简单高效。
缺点：若记录数相同但内容不同，则无法发现。

C.3 关键字段抽样比对法

描述：根据业务规则抽取关键字段（如主键、更新时间戳、金额字段），在两端进行精细化比对。
适用场景：数据库表（HBase）、核心业务表的一致性验证。
优点：能在性能和数据质量间取得平衡，精准发现不一致记录。
缺点：抽样规则设计复杂，可能存在漏抽风险。

C.4 业务规则验证法

描述：在灾备端运行简单的业务查询或汇总报表，将结果与生产端运行结果进行比对。
适用场景：数据仓库、报表系统，验证数据逻辑正确性。
优点：从业务视角验证数据有效性。
缺点：对计算资源有要求，方法复杂。

参 考 文 献

- [1] GB/T 20988-2007 信息安全技术 信息系统灾难恢复规范
 - [2] GB/T 35274-2023 数据安全技术 大数据服务安全能力要求
 - [3] JR/T 0158-2018 证券期货业数据分类分级指引
-