

ICS 03.060

CCS A 11

Q/GJIT

国 金 证 券 信 息 技 术 标 准

Q/GJIT 024—2023

统一数据采集和交换服务技术标准

Unified data collection and exchange service technical standards

2023 - 08 - 20 发布

2023 - 08 - 25 实施

国 金 证 券 股 份 有 限 公 司 发 布

目 次

前言.....	II
1. 标准编制目的和范围.....	1
2. 规范性引用文件.....	1
3. 术语、定义.....	1
3.1. 数据采集.....	1
3.2. 数据交换.....	1
3.3. 缩略词.....	1
4. 技术标准概述.....	1
4.1. 技术标准目的及意义.....	1
4.2. 技术标准适用对象.....	1
4.3. 技术标准适用范围.....	2
5. 技术过程标准.....	2
5.1. 概述.....	2
5.2. 业务需求阶段.....	2
5.2.1. 阶段活动描述.....	2
5.2.2. 提交文档.....	3
5.3. 分析设计阶段.....	3
5.3.1. 阶段活动描述.....	6
5.3.2. 提交文档.....	7
5.4. 开发实施阶段.....	7
5.4.1. 阶段活动描述.....	7
5.4.2. 提交文档.....	7
5.5. 测试阶段.....	7
5.5.1. 阶段活动描述.....	7
5.5.2. 提交文档.....	8
5.6. 上线维护阶段.....	8
5.6.1. 阶段活动描述.....	8
5.6.2. 提交文档.....	8
6. 管理过程标准.....	8
6.1. 概述.....	8
6.2. 业务需求阶段.....	8
3.2.1. 阶段活动描述.....	9
3.2.2. 提交文档.....	9
6.3. 计划制定阶段.....	9
3.3.1. 阶段活动描述.....	9
3.3.2. 提交文档.....	9
6.4. 开发迭代阶段.....	9
3.4.1. 阶段活动描述.....	10
3.4.2. 提交文档.....	10
参 考 文 献.....	11

前言

本标准是统一数据采集和交换服务技术标准：

本标准由国金证券股份有限公司提出。

本标准由国金证券股份有限公司归口管理。

本标准起草单位：国金证券股份有限公司。

本标准主要起草人：张扬东、徐军辉、李凯、赵靖。

本标准为首次发布

统一数据采集和交换服务技术标准

1. 标准编制目的和范围

本规范用于定义国金证券股份有限公司统一数据采集和交换服务标准,公司统一数据采集和交换服务依据本规范开展,原则上适用于所有系统数据采集和交换服务。

2. 规范性引用文件

下列文件对于本文应用是必不可少的。凡是注日期的引用文件,仅所注日期的版本适用于本件。凡是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 0261—2022 《证券公司客户信息交换规范》

JR/T 0022—2020 《证券交易数据交换协议》

JR/T 0111—2014 《证券期货业数据通信协议应用指南》

3. 术语、定义

3.1. 数据采集

数据采集是指通过结合各种技术从公司业务系统、外部数据、日志等收集、整理存储在统一平台。

3.2. 数据交换

数据交换是指跨系统间数据供给,数据之间交换服务。

3.3. 缩略词

RDBMS: 关系数据库管理系统 (relational database management system)

NoSQL: 非关系型的数据库 (non-relational SQL)

CDC: 变化数据捕捉 (change data capture)

4. 技术标准概述

4.1. 技术标准目的及意义

本技术标准的目的是使统一数据采集和交换服务开发及开发工程阶段清晰,要求明确,任务具体,便于规范化、系统化及工程化。有利于提高数据采集以及数据对外服务的控制及管理,提高数据质量,缩短开发时间,减少其开发和维护成本,使整个数据采集与交换服务更科学、更高效。

4.2. 技术标准适用对象

本设计规范适用于以下相关人员:

需求人员:通过本文档可以了解统一数据采集和交换服务开发技术标准,准确理解需求,并与数据采集和数据服务人员交流,以保障从设计到开发实施遵从本文档,并保证从开发管

理的角度是可控的。

数据采集和交换开发人员：通过本文档了解统一数据采集和交换服务开发技术标准，正确理解和梳理包括数据采集、存储、清洗和交换。保证实际实施不会偏离数据采集和数据交换服务整体设计思路。遵循本文档所规定的架构体系、技术标准及开发管理流程，进行相应的程序设计开发、测试、上线工作。

运行维护人员：通过本文档了解统一数据采集和交换服务开发技术标准，加深理解系统的运行维护工作，保障系统稳定运行。

4.3. 技术标准适用范围

适用于统一数据采集和交换服务数据开发生命周期中全部过程。技术标准分为技术过程标准和管理过程标准。

5. 技术过程标准

5.1. 概述

统一数据采集和交换服务的生命周期可以分为五个顺序实施的阶段，分别为业务需求阶段、分析设计阶段、开发实施阶段、测试阶段、上线维护阶段。在对技术过程规范描述，按阶段内部的活动和产物对五个阶段分别说明。

5.2. 业务需求阶段

业务需求阶段主要指的是业务需求调研，包括配合业务人员制定数据采集与交换服务业务方案和落地对应技术实施方案，最终形成业务口径《数据采集和交换需求设计说明书》需详细列出业务人员对数据使用的口径说明。

业务需求阶段是采集交换开发的开始，该阶段工作成绩的优劣、工作质量的高低，将直接影响采集交换实施的成败。

5.2.1. 阶段活动描述

- 1) 明确用户需求，与业务人员就口径进行沟通，形成业务口径的《数据采集与交换服务业务需求说明书》。
- 2) 完成数据采集与数据交换服务定义，根据数据标准模板要求对数据采集或交换的标准服务，包括业务、技术与管理属性。
- 3) 完善业务口径《数据采集与交换服务业务需求说明书》，并与业务人员进行确认，并形成技术口径文档。

5.2.2. 提交文档

梳理整体需求的业务和技术口径，确定《数据采集与交换服务业务需求说明书》发起需求流程，提交对应文档。

5.3. 分析设计阶段

数据采集和交换服务分析设计主要分为需求指标分析、数据源接入、数据存储与管理。同时遵守其命名标准与代码标准。

1) 需求指标分析

进行数据采集与交换前，需要对需求进行详细的分析和探讨，包括数据采集和交换的频率、来源、格式、数据量、业务条件等方面的要求。

数据采集和交换的频率：取决于具体的数据采集和交换需求场景，一般分为实时和离线两个方面。实时场景在于数据源发生变化时，基于 CDC 捕获其变化应用到目标系统中，保证数据的实时同步性；离线场景一般应用于数据更新速度慢，定义为小时级别以上的数据采集与同步，通过定时或事件触发方式进行数据采集与交换服务。

数据来源：数据采集的来源主要为公司业务系统数据库，业务日志数据，外部资讯数据等；数据交换服务数据主要为业务原始数据或数仓加工的结果性数据，便于下游业务系统进行跨系统间原始数据拉取以及数据指标结果入库。

数据格式：数据格式主要为 RDBMS 关系型数据库存储、大数据存储、NoSQL 数据存储、非结构化数据存储等。

数据量：大数据量的数据采集交换可能会影响系统性能、网络带宽、平台资源限制。数据采集与交换需要将业务场景与系统性能和资源限制等因素结合进行综合考虑。

数据业务条件：数据采集与交换时，商定好数据采集交换的数据业务条件，增加业务约束条件。

2) 数据源接入

根据需求分析，选择合适的数据源，并实现对不同数据源的接入和采集，包括数据抓取、数据清洗、数据同步等过程，确保数据来源的完整性和可靠性。

实时主要数据源类型明细详见表 2.3-1

类型	数据源	读(Reader)	写(writer)
RDBMS 关系型数据库	Mysql	√	√
	Oracle	√	√

	PostgreSQL	√	√
NoSQL 数据存储	MongoDB	√	√
	HBASE	√	√
	Kudu	√	√
消息中间件	KAFKA	√	√

表 2.3-1

离线支持的主要数据源类型明细详见表 2.3-2

类型	数据源	读(Reader)	写(writer)
RDBMS 关系型数据库	Mysql	√	√
	Oracle	√	√
	OceanBase	√	√
	SqlServer	√	√
	PostgreSQL	√	√
	DRDS	√	√
	Kingbase	√	√
	达梦	√	√
	通用 RDBMS (支持所有关系型数据库)	√	√
NoSQL 数据	MongoDB	√	√
大数据	Hive	√	√
非结构化数据	TxtFile	√	√
	FTP/SFTP	√	√
	HDFS	√	√
	Elasticsearch	√	√

表 2.3-2

3) 数据存储与管理

基于需求分析，数据采集入数据仓库系统或数据交换中心，统一存储。数据对外服务场景，通过数据交换中心进行对外分发。

数据采集：需要对于数据进行入仓进行管理，按业务类型进行分类和标签管理，对于业务系统和业务库进行逻辑编码方式进行数据入仓管理，实现对数据的高效存储和管理。同时对于数据采集条件中全量、增量、快照、拉链等逻辑分类方式进一步按固定维度进行分类存储。

数据交换服务：与应用系统进行数据交换时，数据存储和数据交换中心，数据交换中心存储的数据应与来源系统保持一致，存在交换逻辑互斥场景，增加业务逻辑后缀区分。

数据交换中心存储近期的源系统交换数据，不保留全量历史数据。

4) 命名标准

定义统一数据采集命名标准，公司数据采集相关表命名遵循以下标准展开，详见表 2.3-3。

说明	<所属层次>	STG（贴源层）
	<来源系统简称>	来源系统编码
	<数据库表名称>	系统编码_数据库编码_系统表名
	<计算周期>	年\半年\季\月\周\日\时\分\实时 Y\HY\Q\M\W\D\H\MNT\R。其中缺省值为 D
	<存储策略>	拉链 H，增量 I，全量 A，快照 S。其中 I、A 均为分区表，全量 A 是以时间为分区，每天存一份全量数据。H、S 表均为非分区表，快照 S 表只存一份全量数据，注意和全量 A 做区分。
	<后缀>	临时表的序号 (01,02···)，外部表的 ext，历史表的 his，该项属于可选。
其他约束	<p>所有数据库对象名称均使用英文字母、下划线或数字来命名，不得以下划线或数字开头。</p> <p>表名称由字母、数字、下划线组成。</p> <p>表名与源系统保持一致，并采用大写。</p>	

表 2.3-3

定义统一数据交换服务命名标准，公司数据交换服务相关表命名遵循以下标准展开，详见表 2.3-4。

说明	<所属层次>	ODS（数据交换层）
----	--------	------------

	<应用分类>	数据交换中心
	<模式名>	源系统库
	<自定义表名>	与源系统采集的保持一致，存在交换逻辑互斥场景，增加业务逻辑后缀区分。
	<计算周期>	无计算周期
其他约束	<p>表名与采集端的系统简称保持一致。</p> <p>对于异构数据库存在数据表长度超长问题，源表名截位，并保留采集策略，如截位出现命名规则一样的，增加系统标识区分，符合见表知源表原则。</p> <p>建表时不可使用双引号限制表名、字段名的大小写。</p>	

表 2.3-4

5) 代码标准

标准代码梳理是数据采集与交换服务的一个非常重要的环节，代码开发时遵循的设计原则见下表：

序号	原则
1	数据采集与交换服务涉及应用库连接信息，统一从数据源模块，通过获取变量编码方式获取。
2	代码中禁止出现硬编码明文业务系统参数。
3	相关脚本命名应与调度平台任务名保持一致，统一大写编码。
4	如果代码有外部标准（国际标准，国家标准，行业标准），则以外部标准为基础按实际应用需要进行删减。
5	如果代码在行业中有明确约定，则沿用该约定。

表 2.3-5

5.3.1. 阶段活动描述

- 1) 根据技术口径《数据采集与交换服务业务需求说明书》，确定数据来源，从测试系统中抽取数据，进行数据口径统一及命名标准化定义，与业务人员确认，验证技术口径《数据采集与交换服务业务需求说明书》准确性。

5.3.2. 提交文档

- 1) 《数据采集与交换服务业务需求说明书》

5.4. 开发实施阶段

将业务人员的需求变成真正可用的数据采集与交换服务脚本,是通过编码和系统实现阶段来完成的。编码的途径和实现的具体方法对程序的可靠性、可读性、可测试性和可维护性产生深远的影响。

开发过程还应该进行单元测试工作,确保数据采集和交换服务的正确性和性能满足业务需求。

5.4.1. 阶段活动描述

- 1) 根据技术口径《数据采集与交换服务业务需求说明书》,进行数据开发设计(包括开发资源确认,工作任务分解,开发进度分解,质量目标等)。
- 2) 进行单元测试,形成《单元测试报告》。

5.4.2. 提交文档

- 1) 《单元测试报告》

5.5. 测试阶段

测试阶段在整个数据采集和交换服务中是占据总工作量最大的一个环节。测试工作还伴随着诊断、改正错误、调试等复杂过程。所以应该包括几乎所有的开发人员。

这个阶段,必须严格把关,确保数据采集和交换服务的数据结果的高可靠性。经过测试、稳定的版本包括相关的文档可以进行打包,同时需包含备份脚本和回退脚本,作为软件开发的出口。

5.5.1. 阶段活动描述

- 1) 测试人员根据前面提供的资料,进行测试用例编写,形成《测试用例文档》。
- 2) 测试人员根据《测试用例文档》对每一个指标进行一系列的测试,开发人员针对测试发现的问题进行修正;
- 3) 集成环境进行集成测试,开发人员针对测试发现的问题进行修正;
- 4) 将版本发布到 UAT 环境进行 UAT 测试,开发人员针对 UAT 测试发现的问题进行修正;
- 5) 编制《测试报告》,提供《版本发布包》、《版本发布说明书》、《数据备份脚本》、《数据回退脚本》。
- 6) 发布软件版本。
- 7) 其中 2、3、4 是循环执行的。

5.5.2. 提交文档

- 1) 测试用例文档
- 2) 测试报告
- 3) 版本发布脚本
- 4) 版本发布说明书
- 5) 数据备份脚本
- 6) 数据回退脚本

5.6. 上线维护阶段

数据采集和交换服务脚本运行过程中，会出现新增业务需求、需求变更和需求下线的情况，在该阶段需保证已运行的数据采集与交换服务数据准确性，并分析新需求和需求变更的可行性。

5.6.1. 阶段活动描述

- 1) 上线安装及管理。
- 2) 持续了解新的业务需求。
- 3) 分析新的需求及需求变更可行性。
- 4) 对新的需求以及需求变更进行设计开发。

5.6.2. 提交文档

- 1) 业务需求新增/变更单
- 2) 用户验收报告

6. 管理过程标准

6.1. 概述

数据采集和交换服务的计划和进度安排，采用由粗到细、由简到繁的方式，首先制定开发过程总体阶段和迭代的开发计划，而后根据所划分的迭代过程，在每个迭代开始时，对该迭代过程进行详细的任务分配和进度规划。

本规范中所提到的开发计划，包含了开发计划、质量管理计划、技术支持计划等多项内容，但主要以开发计划为主，其他计划视开发具体情况确定是否制定。

6.2. 业务需求阶段

各业务部门在提出数据采集和交换服务需求时需填写《需求申请表》，经数据采集交换负责人确认后纳入开发计划。

数据采集交换负责人根据资源情况和公司需求组织开发人员进行进一步的需求讨论会，

最终形成《业务需求说明书》。

3.2.1. 阶段活动描述

- 1) 根据《需求申请表》进行需求确认和需求讨论
- 2) 分析数据采集和交换服务需求,形成《业务需求说明书》

3.2.2. 提交文档

- 1) 业务需求说明书

6.3. 计划制定阶段

可交付结果进行分类与划分。将不同需求分解出多个开发任务,每一个任务对应一个可交付结果。由采集交换负责人根据需求情况和制定《数据采集交换开发计划》,并提交研发负责人反馈确认,如无异议按照《数据采集交换开发计划》执行。

采集交换负责人对开发情况进行持续跟踪、监控,并定期组织工作会议、统计工作度量数据,跟踪工作进度执行情况。

3.3.1. 阶段活动描述

- 1) 明确开发阶段和迭代次数,同时完成其他开发相关的计划内容,形成《数据采集与交换开发计划》。

3.3.2. 提交文档

- 1) 数据采集与交换开发计划
- 2) 工作任务进度计划书,工作任务进度跟踪表

6.4. 开发迭代阶段

采集交换负责人根据《数据采集与交换开发计划》的进度定期进行一次分析汇报,确定开发的状态、分析风险和对策。《数据采集与交换开发计划》需按照实施过程分解为需求分析、系统设计、开发编码和测试运行阶段。

1) 需求阶段

业务团队完成《需求规格说明书》交由采集交换负责人审阅并修正。

需求分析过程完成后,如果需求变更,需求提供部门须书面提出《需求变更通知书》,数据采集交换开发团队在完成分析反馈,确定需求变更,采集交换负责人变更对应开发计划内容,并安排对应的开发变动。

2) 设计阶段

采集交换负责人依据技术实现方案进行详细设计工作,并按计划完成技术口径《数据采集和交换需求数据设计说明书》,确保数据采集和交换服务能满足业务需求。

采集和交换服务发生变更时，采集交换负责人遵循开发中变更流程组织实施变更。

3) 开发编码阶段

在完成需求和设计阶段后，采集交换负责人按照《数据采集与交换开发计划》中的周期计划编写编码计划任务书，确定开发周期，开发顺序，以及任务责任人。

根据开发的情况，要求开发人员对相关代码进行质量和性能测试、单元测试和集成测试等，并输出单元测试报告。

需求发生变更时，采集交换负责人遵循开发中心变更流程组织实施变更，确保需求、设计、测试等相关文档和代码的变更保持一致性，变更流程的完整性和可被追溯。

4) 测试阶段

软件测试是保证软件质量的重要工作环节。测试前要详细制定《测试计划》，记录测试中发现问题及解决情况。测试完成后，形成《测试报告说明书》。

3.4.1. 阶段活动描述

- 1) 根据《数据采集与交换开发计划》，结合具体的开发状况和资源获取情况，确定在一个迭代期间的开发任务，进度安排，质量目标，资源分配形成《任务单》，描述任务要求，明确开发人员职责，并更新《数据采集与交换开发计划》。
- 2) 根据本次迭代开发的完成情况和提交的成果，并根据实际情况，提出《变更请求》。

3.4.2. 提交文档

- 1) 数据采集与交换开发计划
- 2) 变更请求
- 3) 测试报告说明书

参考文献

- [1] JR/T 0261—2022 证券公司客户信息交换规范
 - [2] JR/T 0022—2020 证券交易数据交换协议
 - [3] JR/T 0111—2014 证券期货业数据通信协议应用指南
 - [4] JR/T 0158—2018 证券期货业数据分类分级指引
-