

ICS 03.060

CCS A 11

Q/GJIT

国 金 证 券 信 息 技 术 标 准

Q/GJIT 023—2023

市场资讯数据中心数据标准及技术规范

Market Information Data Center Standards and Specifications

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. 范围.....	1
2. 规范性引用文件.....	1
3. 术语和定义.....	1
3.1. 资讯中心 Information Center.....	1
3.2. 资讯数据 Information data.....	1
3.3. 数据主题 Data Subject.....	1
3.4. 数据服务 Data Services.....	1
3.5. 缩略语.....	1
4. 市场资讯数据中心.....	2
4.1. 定义及架构.....	2
4.2. 数据源.....	2
4.3. 数据链路及数据服务.....	3
4.4. 衍生指标计算&多源比对融合.....	3
4.5. 应用系统交互规范.....	4
4.6. 统一监控与管理.....	4
5. 数据标准.....	5
参 考 文 献.....	7

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。

本文件所有内容应符合强制性国家标准、行业标准及地方标准，若与其相抵触时，以国家标准、行业标准、地方标准为准。

本企业对本文件的合法性、真实性、准确性、技术合理性和实施后果负责。

本标准由国金证券提出。

本标准由国金证券归口管理。

本标准起草单位：国金证券股份有限公司。

本文件主要起草人：刘畅、才东旭、林玉鸳、赵靖、熊友根、王洪涛。

市场资讯数据中心数据标准及技术规范

1. 范围

本规范用于规范公司市场资讯数据中心相关数据和技术工作，市场资讯数据应用可依据本规范开展，适用于公司所有应用系统。

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 0275—2023 证券期货业机构内部接口资讯数据

JR/T 0176.1—2019 证券期货业数据模型 第1部分：抽象模型设计方法

JR/T 176.3—2021 证券期货业数据模型 第3部分：证券公司逻辑模型

3. 术语和定义

3.1. 资讯中心 Information Center

资讯中心，指国金证券市场资讯数据中心。

3.2. 资讯数据 Information data

通过媒体，网站，公告等方式进行公开的数据，包括但不限于基础信息数据（公司、证券、市场、行业）、股票数据、基金数据、债券数据、理财产品数据、金融衍生品数据、宏观数据、行业数据等数据内容。

3.3. 数据主题 Data Subject

市场资讯数据，主要包括：行情数据（Market Data）、新闻资讯数据（News Data）、参考数据（Reference Data）、经济数据（Economic Data）等主题。

3.4. 数据服务 Data Services

资讯数据服务，主要包括：存储服务、交换服务、计算服务、多源比对等。

资讯应用库，为C端资讯数据应用服务系统提供数据存储服务和查询服务。

3.5. 缩略语

MPP:大规模并行处理架构(Massively Parallel Processing)。

GP:PB级的数据库（Greenplum）。

ODS:数据交换中心的交换层（Operational Data Store）。

CDC:实时捕获数据库变更传输技术（Change Data Capture）。

CDH:Hadoop发行版（Cloudera's Distribution Including Apache Hadoop）。

4. 市场资讯数据中心

4.1. 定义及架构

资讯中心定义：

市场资讯数据中心，覆盖资讯数据采集、存储、服务、应用全链路，支持数源监控、统一存储计算、标准化加工处理、个性化指标计算。

对外部数据源进行统一的中台管控，推进降本、增效、提质、增值，形成统一管控、长效管理、持续优化的市场数据运营服务机制。

形成对各个部门应用系统的支撑，提供不同场景的资讯数据服务，如：实时交互、离线处理和多源比对等，详见4.3章节。

资讯中心整体架构分为资讯源、核心功能和应用系统三层，其中资讯源层主要用于接收不同来源的三方数据，进行数据存储。核心功能层包含存储服务、计算服务、交换服务、数据质量控制、数据组件扩展、数据链路管理、数据资产管理和数据运维监控八大核心功能点，将数据全生命周期管理，更好的为上游系统提供数据支持与服务。应用系统层主要分为C端应用、B端应用和专项应用三大类，通过下两层来支撑全司资讯相关应用系统。

市场资讯数据中心可提供批量及实时数据同步、数据清洗加工、数据指标计算、多源比对融合等多种数据服务，服务对象包括C端及B端各业务条线。

基于国金数据港数据中台建设，使用MPP数据库GP进行数据加工处理，使用大数据平台CDH进行大规模指标计算，通过DATAx及ODS进行批量数据交换，通过CDC进行实时数据同步，使用多源比对融合平台进行数据质量监控。

4.2. 数据源

覆盖通用外部数据源，覆盖股债基等全部品种，覆盖行情、新闻、财务、研究等各类数

据。数据厂商包括，万得、聚源、东财、同花顺、朝阳永续、财联社等，详细表单及数据字典可见数据门户。

4.3. 数据链路及数据服务

数据服务如下：

数据采集及数据管理	系统功能	技术支持 (to应用)	数据服务 (to应用)
<ul style="list-style-type: none"> ■ 通用数据源 ■ 专用数据源 ■ 数据库同步 ■ SFTP文件同步 ■ 数据API □ RPA采集 □ 爬虫采集 ■ 数据源统一管理 ■ 数据标准管理 ■ 数据资产管理 	<ul style="list-style-type: none"> ■ 单源校验监控 ■ 多源数据比对 ■ 数据一致性校验 □ 全链路技术监控 ■ 统一存储 ■ 质控与标准化 ■ 弹性计算 (按需) ■ 多源融合 (按需) ■ 资源管控 ■ 权限管控 ■ 技术规范管控 ■ 元数据管理 	<ul style="list-style-type: none"> ■ 数据库同步 ■ ESB (REST API) ■ SFTP ■ JDBC/ODBC ■ GP查询 ■ SmartBI/FineBI 	<ul style="list-style-type: none"> ■ 数据资产目录 ■ 原始数据服务 ■ 衍生指标数据服务 ■ 融合数据服务 ■ 离线数据服务 ■ 实时数据服务 □ 数据分析工具 ■ 创新数据服务 (沙盒)

图 1 数据服务功能清单

离线原始数据需求场景：资讯数据源通过ODS交换，离线同步至应用系统数据库。

实时原始数据需求场景：资讯数据源通过CDC交换，实时同步至应用系统数据库。

离线加工数据需求场景：资讯数据源通过数据同步工具（DATAx）进入资讯批计算中心根据业务需求定义进行指标加工批计算，再通过ODS交换，离线同步至应用系统数据库。

实时加工数据需求场景：资讯数据源通过CDC交换，进入实时数据中心（KAFKA+FLINK）根据业务需求定义进行指标的实时计算，再通过CDC交换，实时同步至消息队列（KAFKA）应用方进行实时数据订阅消费。

多源比对需求场景：根据业务需求，进行多源比对融合及监控预警开发。

4.4. 衍生指标计算&多源比对融合

衍生指标计算规范：

- 1) 资讯指标类：需要进行数据的清洗、转换和计算等（同比，环比）；
- 2) 模型计算、数据挖掘类：需要使用机器学习算法类指标；
- 3) 历史数据处理类：指标计算所用数据含有大量历史数据；
- 4) 查询性能成本大：指标关联逻辑复杂程度，无法满足查询性能要求；

- 5) 弹性计算类：计算量大且需要实时计算场景，如需要与客户进行交互的指标，且无法预先计算。

多源比对融合规范：

- 1) 功能介绍：多源比对融合服务，包括单源校验、多源比对、多源融合、监控预警等功能；
- 2) 需求规范：多源比对需求需提供业务系统使用单源或多源表名及需要比对的字段；多源融合需求需提供需要使用的字段。

4.5. 应用系统交互规范

离线数据交互规范：

由数据部门统一将数据从源库按原始结构提取到中台ODS或者数仓，对于同步到中台ODS任务，中台ODS定位只存储临时快照数据，需先清空中台ODS对应表的数据，再进行数据灌入，同时数据写入成功后，将写入成功的标志位写入标志位表中。

C端应用：

应用系统通过资讯应用库统一提供资讯数据，上层通过资讯数据应用服务提供统一的资讯数据API，为C端各个业务系统提供资讯数据支持。

B端应用：

应用系统原则上不可直连资讯源库，通过中台ODS等拉取数据到本地应用库，或对接应用前置库，需按照采集模板提交需求。

4.6. 统一监控与管理

统一监控管理系统，包括数据源管理及任务统一监控管理。

数据源管理，包括表清单和数据字典等。

数据源监控，包括单源质量校验、多源比对融合和预警通知等。

任务监控，包括数据同步任务和数据比对融合任务等。

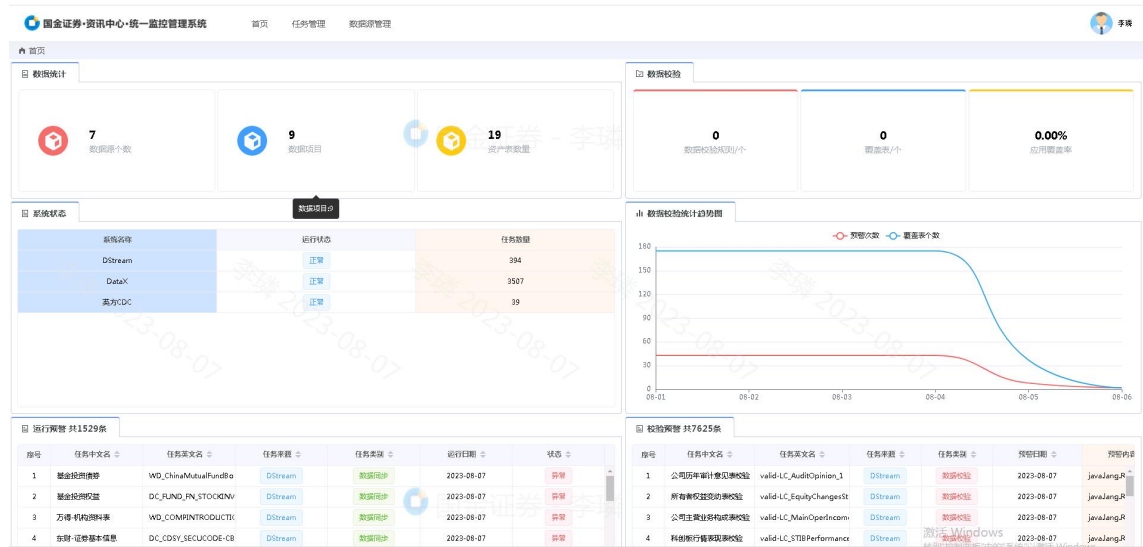


图 2 资讯中心统一监控管理平台

5. 数据标准

资讯源库数据标准：依据供应商数据标准执行。

ODS:

〈数据库表名称〉	系统英文标识_数据库英文标识（示例:NEWFX_FXGLBDB）
〈存储策略〉	与源系统采集的保持一致，增加后缀区分采集策略（I：增量 A：全量 M：Merge操作）。

CDH:

〈数据库表名称〉	源系统数据库编码_系统表名
〈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。

Greenplum:

〈数据库表名称〉	贴源层_源系统Schema_源表名
〈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。

融合中心库：

〈数据库表名称〉	MDC_业务含义字母简称
〈存储策略〉	与源系统采集的保持一致，表名由业务含义字母简称及下划线组成。

资讯应用库：

〈数据库表名称〉	源系统Schema_源表名
〈存储策略〉	与源系统采集的保持一致，增加后缀区分采集策略（I：增量 A：全量 M：Merge操作）。

参 考 文 献

- [1] JR/T 0275—2023 证券期货业机构内部接口资讯数据
 - [2] JR/T 0176.1—2019 证券期货业数据模型 第1部分：抽象模型设计方法
 - [3] JR/T 176.3—2021 证券期货业数据模型 第3部分：证券公司逻辑模型
 - [4] 金融资讯数据服务平台建设实践《交易技术前沿》2021年第三期（总第45期）
-