

DB 5101

四川省成都市地方标准

DB 5101/T XXXX—XXXX

公共资源交易数据规范 第3部分：集成与治理

(征求意见稿)

XXXX - XX - XX 发布

XXXX - XX - XX 实施

前 言

DB5101/T XXXX-XXXX《公共资源交易数据规范》分为4个部分：

- 第1部分：资源体系；
- 第2部分：元数据管理；
- 第3部分：集成与治理；
- 第4部分：共享服务。

本部分为DB5101/T XXXX的第3部分。

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。

本文件由成都市城市运行和政务服务管理办公室提出。

本文件起草单位：成都市公共资源交易服务中心。

本文件主要起草人：章建国、李晓莺、刘蕊梅、马艳红、周涛、李运飞、蔡俊。

公共资源交易数据规范 第3部分：集成与治理

1 范围

本文件规定了公共资源交易的数据集成和数据治理的整体流程、数据安全与质量和成果的主要内容。

本文件适用于公共资源交易的数据集成和数据治理。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据集成

通过数据集成平台，抽取中心不同来源、格式、特点性质的数据进行集中存储。

3.2

数据治理

按照数据规范对抽取至数据集成平台的数据进行清洗，将不完整数据、重复数据以及错误数据进行治理，得到符合要求的数据，并加载到数据仓库中进行存储。

3.3

ETL

数据抽取（Extract）、转换（Transform）、装载（Load）。

4 整体流程

4.1 概述

中心数据首先通过数据源配置、ETL任务开发和完成数据调度实现数据集成，再通过新建数据仓库库表、创建数据治理任务和设置任务执行周期实现数据治理，整体流程如图1所示。

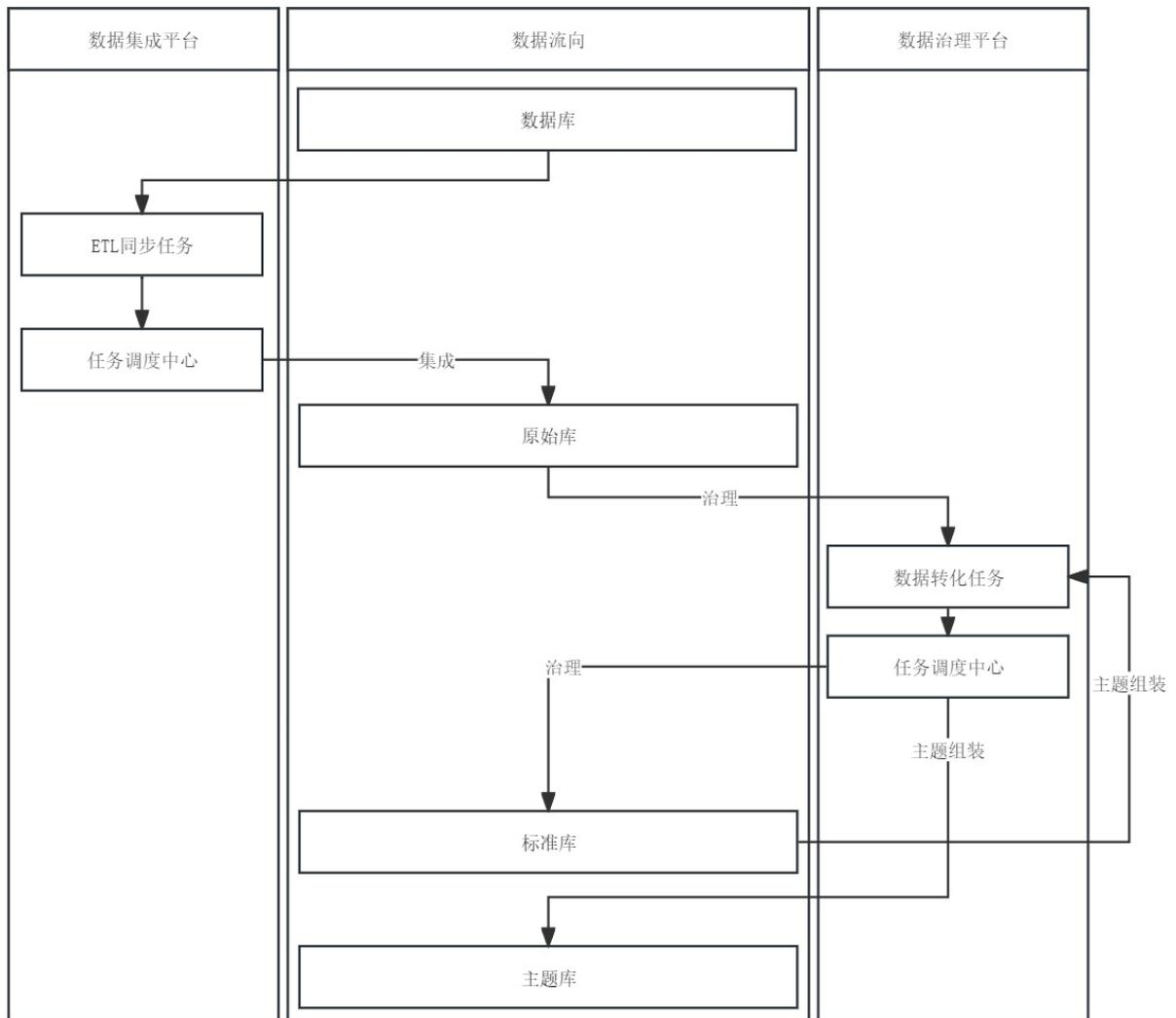


图1 整体流程图

4.2 数据集成

4.2.1 数据源配置

4.2.1.1 选择要抽取的数据源类型，设置读写权限、IP 地址、端口号等信息，完成数据源连通性测试。

4.2.1.2 数据源类型包含主流的 oracle、mysql、sqlserver 等，且支持国产化数据库，如 dameng、oceanbase、kingbase 等。

4.2.2 ETL 任务开发

通过图形拖拽方式，创建ETL任务，由以下三段图形组件构成：

- a) 数据抽取组件，包括增量表读取、sql 读取、表全量读取；

- b) 数据转换组件，可视化配置转换，将特殊编码字符替换过滤，优先使用表直接读取，遇到目标数据与集成数据库数据结构有冲突使用 SQL 拉取；
- c) 数据加载组件，包括增量表加载、表全量加载。

4.2.3 数据调度

4.2.3.1 根据 ETL 任务，从配置的数据源自动调度数据。

4.2.3.2 通过数据调度中心可查看 ETL 任务执行情况，用于监控数据同步。

4.3 数据治理

4.3.1 新建数据仓库库表

根据数据集成的数据源建立原始数据库表，根据数据规范建立数据标准库表，数据仓库数据库表可在线管理。

4.3.2 创建数据治理任务

通过图形化组件拖拽方式，按照数据规范，创建数据治理任务，常用以下两种方式：

- a) 数据处理 sql 组件，通过编辑 sql 组件，编写标准化处理的 sql 语句，实现将原始数据处理成标准数据，适用于空值处理，多系统不同码值转换、字段转换等；
- b) 数据清洗转换组件，使用图形化的操作，配置原始库表到标准库表的清洗转换，可为字段配置不同的处理方法，适用于原始数据跟标准数据区别不大的情形。

4.3.3 设置任务执行周期

按需设置数据治理任务执行周期，周期可设置为一次性、每天、每月、每几小时等，系统按照设置的任务执行周期进行数据治理。

5 数据安全与质量

5.1 数据安全

对数据仓库中的数据进行加密、解密、脱敏等安全的操作，有效避免数据泄密、数据滥用、数据丢失等风险，具体操作步骤如下：

- a) 上传密钥私钥，用于加解密；
- b) 新增脱敏、加密、解密任务，配置脱敏、加解密规则；
- c) 可设置任务执行周期（一次性、每天、每月、每几小时等），系统按照设置的任务执行周期进行数据安全治理。

5.2 数据质量

对数据仓库中的数据进行数据质量校验，保障数据的完整性、准确性，具体操作步骤如下：

- a) 新增数据质量模型，选择需要校验的主表，为其字段配置质量校验规则；
- b) 为数据质量模型设置任务执行周期；
- c) 质量模型按设置的周期执行，形成数据质量分析报告，报告包含数据质量得分、趋势统计、问题占比、质量时间趋势等；

- d) 根据质量分析报告跟踪数据质量问题并处理。

6 成果

通过数据集成和治理，即可抽取不同来源、格式、特点性质的公共资源交易相关数据，进行规范化治理得到公共资源交易数据标准库，并可根据各类数据应用需求建立不同主题库。
