**宿迁市政务信息采集平台**

**技术对接规范**

# 一、编制说明

（一）编制目的

为指导宿迁市政务信息采集平台的技术对接工作，实现全市政务数据资源的归集、治理和共享，特编制本规范。

（二）采集平台简介

基于分布式架构的宿迁市政务信息采集平台提供数据抽取、集中、加载、转换、传输及共享等功能，可为不同系统间执行信息数据交换、信息数据共享以及信息数据业务协同等任务提供基础。支持多源异构数据采集与管理，多种加密、解密算法配置等功能，从而可以保障数据在传输中的安全以及数据在存储过程中的安全。为不同地域、不同部门、不同平台、不同应用系统和不同数据库之间的互连互通提供了包含提取、转换、传输和加载等操作的数据整合服务，实现了扩展性良好的“松耦合”结构的应用和数据集成。进一步发挥信息数据资源的效能，提升信息数据资产对业务产品和后续数据仓库的重要支撑作用。

（三）内容概要

本规范从网络接入、配置、采集方式、库表交换规范、文件交换规范、数据源格式规范、实施步骤系等技术方面，介绍了数据采集平台对外开放的技术规范与要求。

# 二、数据对接注意事项

1. 数据源单位需接入电子政务外网；
2. 数据源单位需已完成宿迁市政务数据共享交换平台账号申请；
3. 通过库表交换对接的，数据源单位需保障本单位侧网络及前置库的安全及相关维护工作；
4. 数据源单位前置库建议开启白名单访问限制（大数据中心采集平台IP地址段信息，请见附件1 数据采集平台白名单IP表）；
5. 数据源单位负责提供对接资源表结构和数据字典等信息，以及基于资源目录的数据库查询脚本。

# 三、对接流程

1. 数据源单位完成宿迁市政务信息共享交换平台完成账号申请；
2. 数据源单位进行政务信息资源目录编制工作；
3. 数据源单位与大数据中心协商确定数据对接及数据更新方式；
4. 数据源单位完成数据准备；
5. 数据中心配置采集任务；
6. 双方检查数据同步准确性，完成对接。

# 四、数据采集方式

数据采集支持库表交换和文件交换两种方式。具体如下：

## （一）库表交换

库表交换是以前置交换节点上的数据库表作为数源平台与数据采集平台之间的数据交换接口，数据源单位向指定交换库表中写入要推送的数据，数据采集平台从交换表中抽取数据。

数据源单位应明确告知大数据中心数据同步方式（增量、全量、实时），同时明确数据同步频率，确保大数据中心可以按时采集到更新的数据。推荐采用增量或实时方式进行数据同步。

交换数据库目前支持oracle、mysql、sql server、DB2、Sybase等多种关系型数据库的适配接入。

## （二）文件交换

文件交换是指数源单位通过将excel文件上传至共享交换平台，数据采集平台，数据采集平台将读取、解析上传文件中的数据。

文件交换支持全量更新和基于时间的增量同步方式。建议采用增量同步方式上报数据。

文件交换目前仅支持excel格式的推送文件。

# 五、库表交换规范

5.1 全量同步

确定数据同步时间，并确保大数据中心进行数据在同步时间前完成数据更新。

5.2 增量同步

增量更新主要以时间戳更新方式为主，增量同步表须设置业务主键、时间戳（时间戳格式需满足附件2 时间戳规范要求）和删除标记字段change\_type，若表没标记字段change\_type时，则需添加字段change\_type，默认值为0，类型为int， (0为新增、1为删除)。各委办局有对数据进行删除操作时，无需删除整条数据，只需把该条数据字段change\_type值改为1，表示此条数据已废弃。

5.3 实时同步

实时同步目前支持MYSQL 和ORACLE数据库,MYSQL基于Binlog 日志、ORACLE基于归档日志实现数据同步。数据源单位在创建数据库时需开启以上功能，并对提供大数据中心的数据库账号进行相关授权。（相关配置请看附件3实时采集数据源端配置手册）

# 六、文件交换规范

（1）文件交换目前支持Excel文件上报；

（2）初次数据以全量方式上报，后续建议基于时间戳的增量方式推送数据；

（3）文件命名格式

数源单位推送数据文件时，对文件命名需使用“资源名称\_YYYYMMDDHHMM+文件后缀”的格式。例：等级裁判员\_201811201010.xls、等级裁判员\_201811201010.xlsx。

（4）Excel规范及要求

受Excel文件每个Sheet页行数限制，建议每个文件只有一个Sheet页，且行数不超过6万行。Excel表中尽量不要使用单元格格式，以文本形式录入数据才能保证数据入库的正确性。

# 七、账号权限相关要求

建议以满足业务需要的情况下，最小化权限为大数据中心创建数据库账号。除必要的数据库权限外，还需授予以下权限。

7.1 oracle授权需求

grant select on dba\_free\_space to test;

grant select on dba\_data\_files to test;

grant select on user\_tablesto test;

grant select on user\_tab\_columns to test;

7.2 sqlserver授权需求

sys.partitions

sys.allocation\_units

sys.internal\_tables

sys.syscolumns

sys.sysobjects

sys.sysindexes

附件一、数据采集平台白名单IP表

|  |  |
| --- | --- |
| 数据采集平台白名单IP表 | |
| 数据源端配置白名单需放开的地址 | 172.22.137.90、172.22.137.91、172.22.137.92、172.22.137.93、172.22.137.94 |

附件二、时间戳规范



附件三、实时采集配置手册

