## 目标

配置离线任务流程从某数据库落地hive。

下面以雪球数据库为例，做配置方式介绍。

支持三种数据流程：

1. 蜂巢人脸标准表从snowball落库到hive
2. 车辆抓拍表从snowball落库到hive
3. 电子围栏表从snowball落库到hive
4. 人脸抓拍表从snowball落库到hive
5. 行人结构化数据表从snowball落库到hive
6. 主副驾驶人脸表从snowball落库到hive

## 版本要求

1. dts-console、datax、data-integrate-frontend采集前端 要求是2.0.0以上版本；
2. datax必须有hivehdfwriter插件；
3. data-dictionary 要求是分拆后的1.0.0.fbfc55bb以上版本，metadata-frontend元数据前端 要求是1.2.5以上版本；
4. job-console、job-engine要求至少是1.2.1- release以上版本

## 配置数据源以及元数据

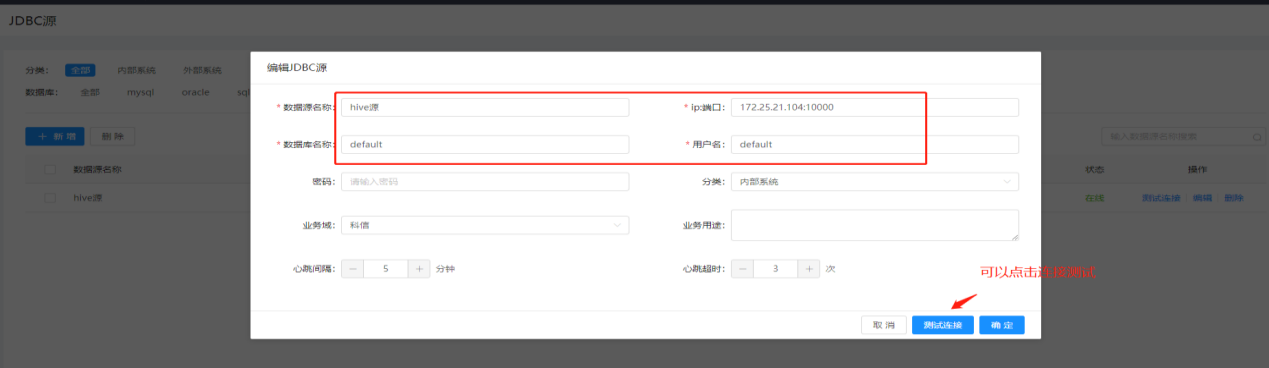
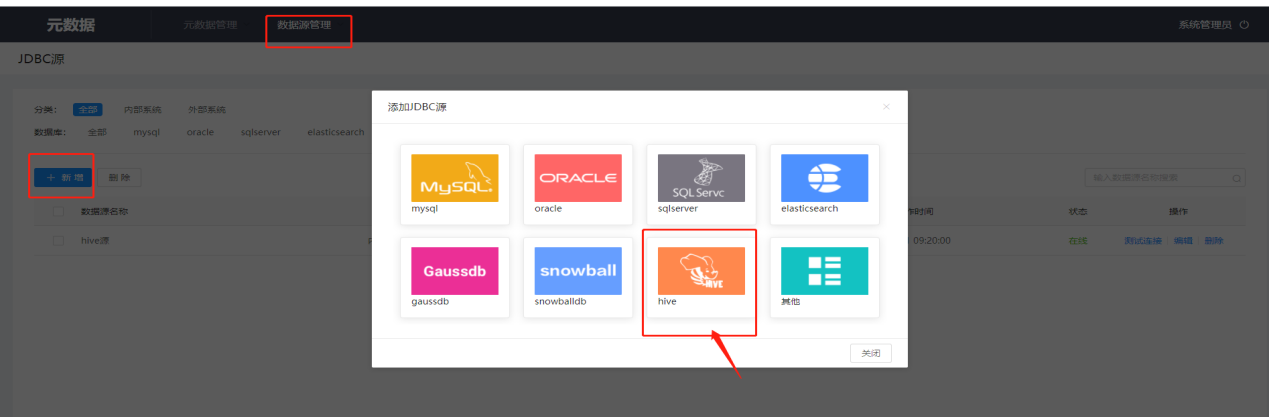
采集流程导入的时候，需要确保采集流程中使用的数据源已经创建了，并且**数据源名称一致**，本系统中使用到的数据源有三个，默认名称分别为：

1. hive:对应元数据文件是“元数据-hive-\*.xlsx”,部署融合分析系统时，自动导入
2. snowball-dts:对应元数据文件是“元数据-snowball-dts-\*.xlsx”,
3. snowball-融合分析专用数据源,部署融合分析系统时，自动导入



### 配置数据源

#### 配置hive源（有此数据源,保证hive可正常连接）



#### 配置snowball-dts数据源,保证连snowball可正常连接

#### snowball-融合分析专用数据源,保证连snowball可正常连接

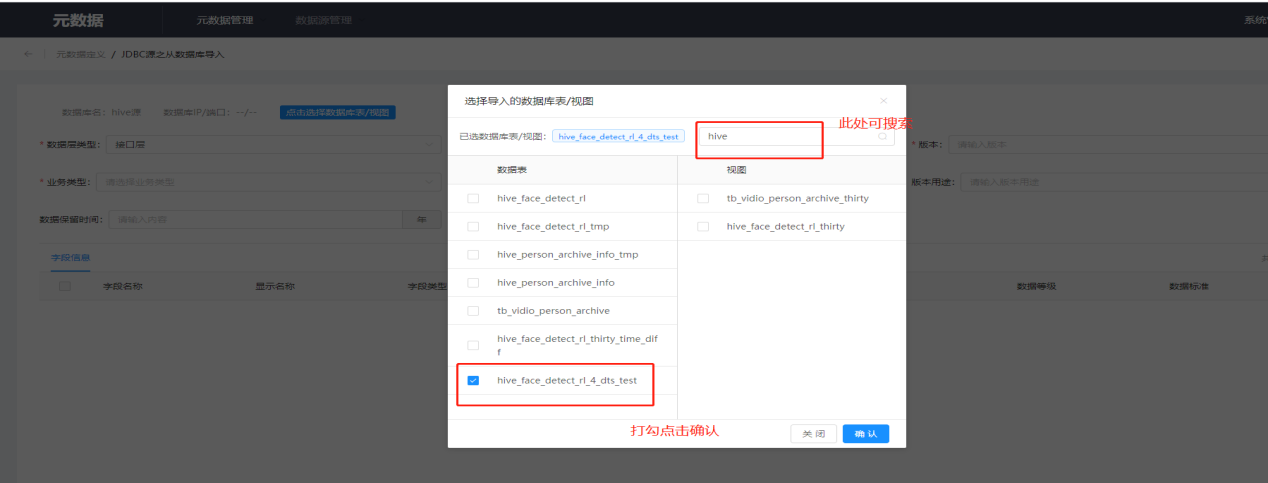


### 配置元数据（有需要的元数据可跳过本步骤）

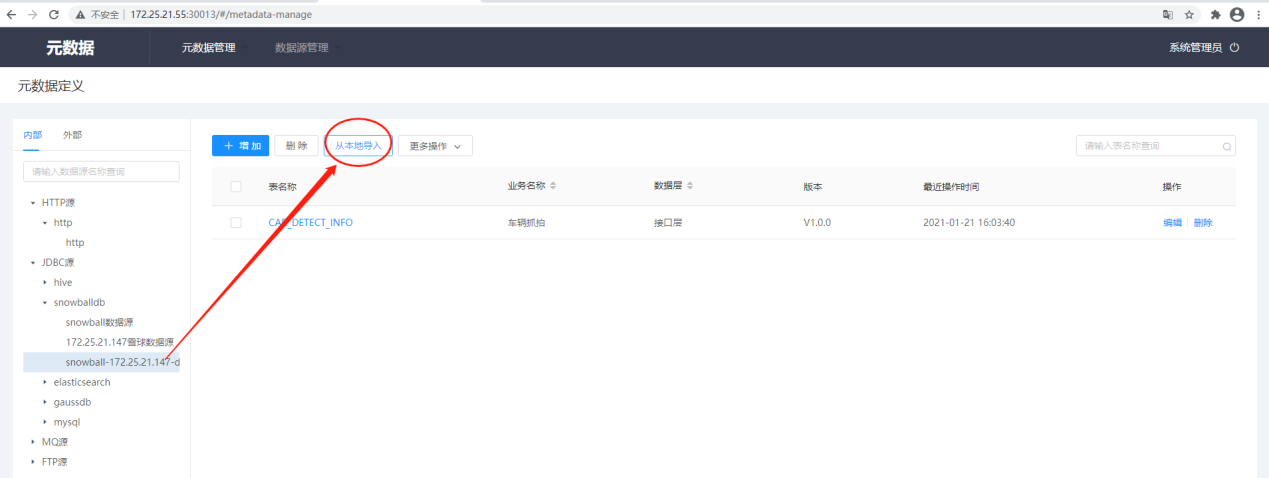
snowball-dts 数据库



需要被抽取的数据库元数据若不存在也请新建。



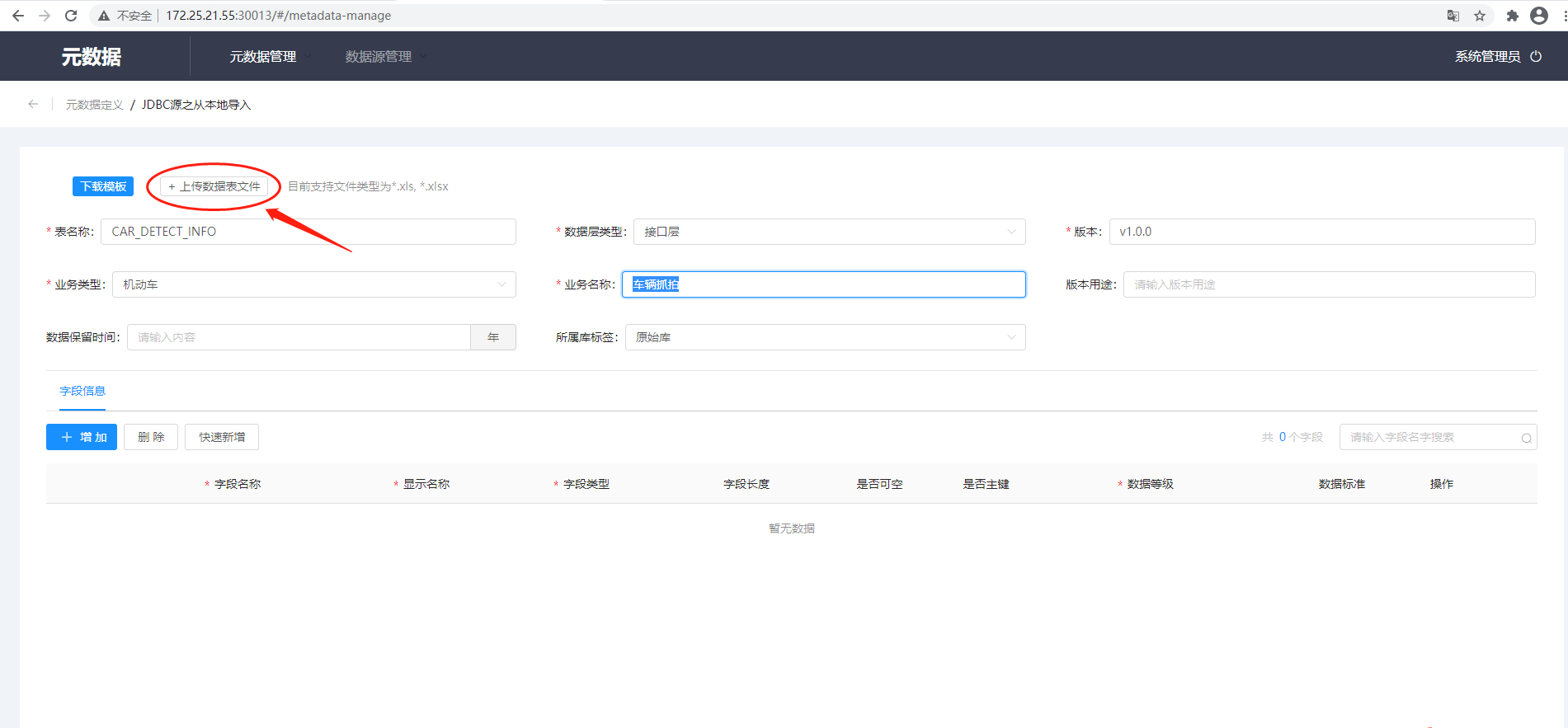
也可以导入已经准备好的元数据文件，如下所示：

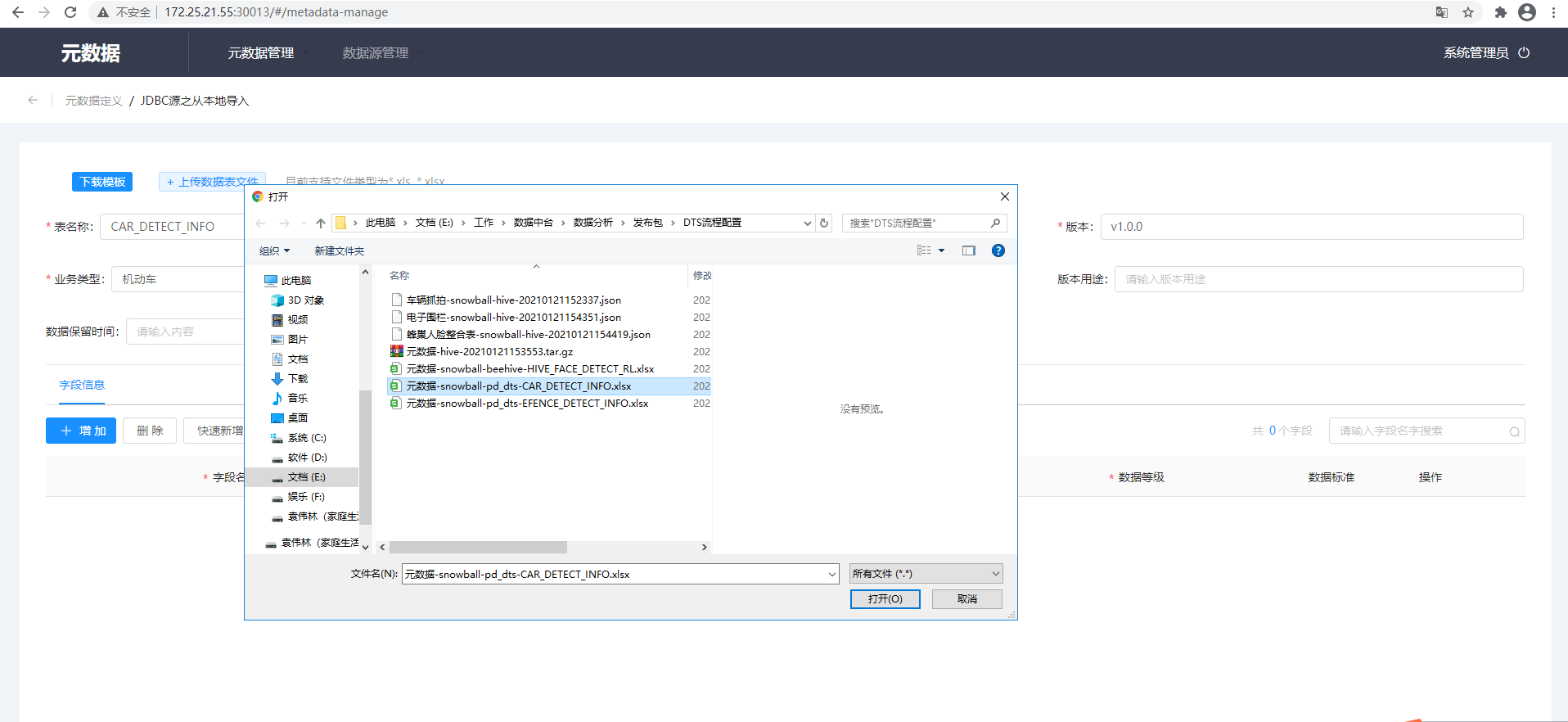


选择已新建的目标数据源，点击“从本地导入”，导入元数据

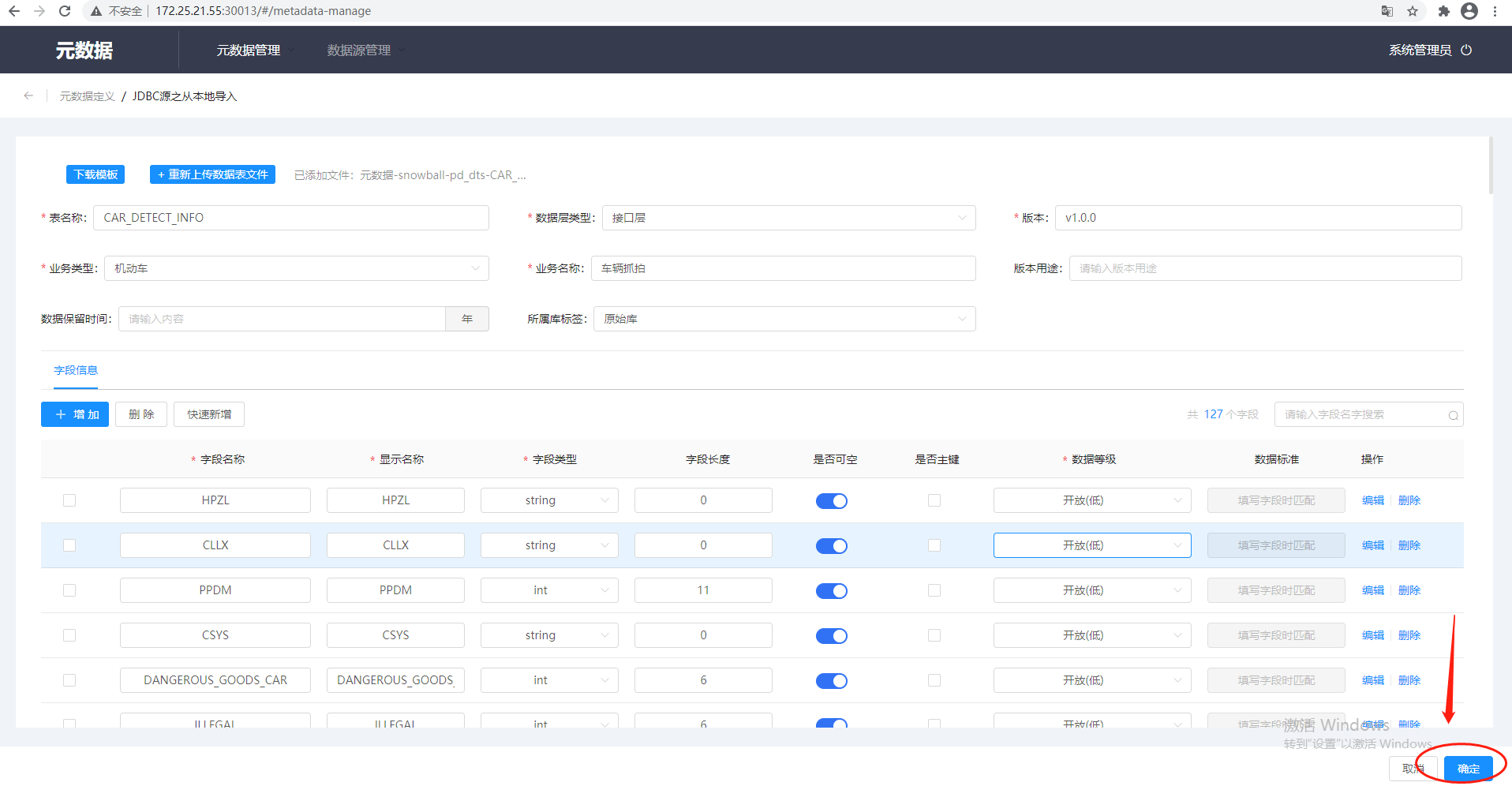


填写好表信息，点击“上传数据文件”上传元数据信息





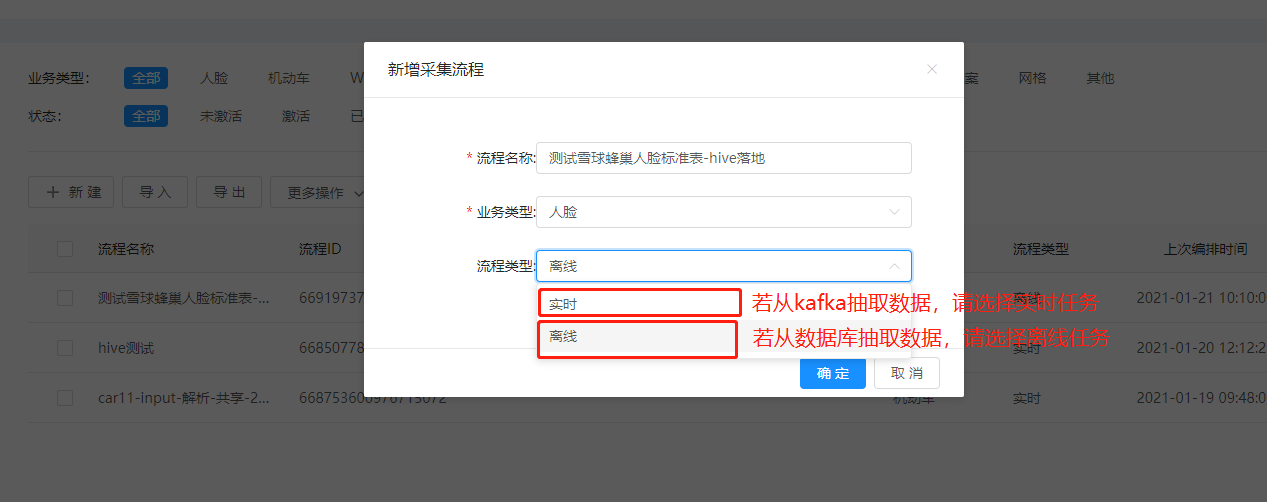
点击“确定”



## 配置流程

### 新建方式：

### （1）新建任务

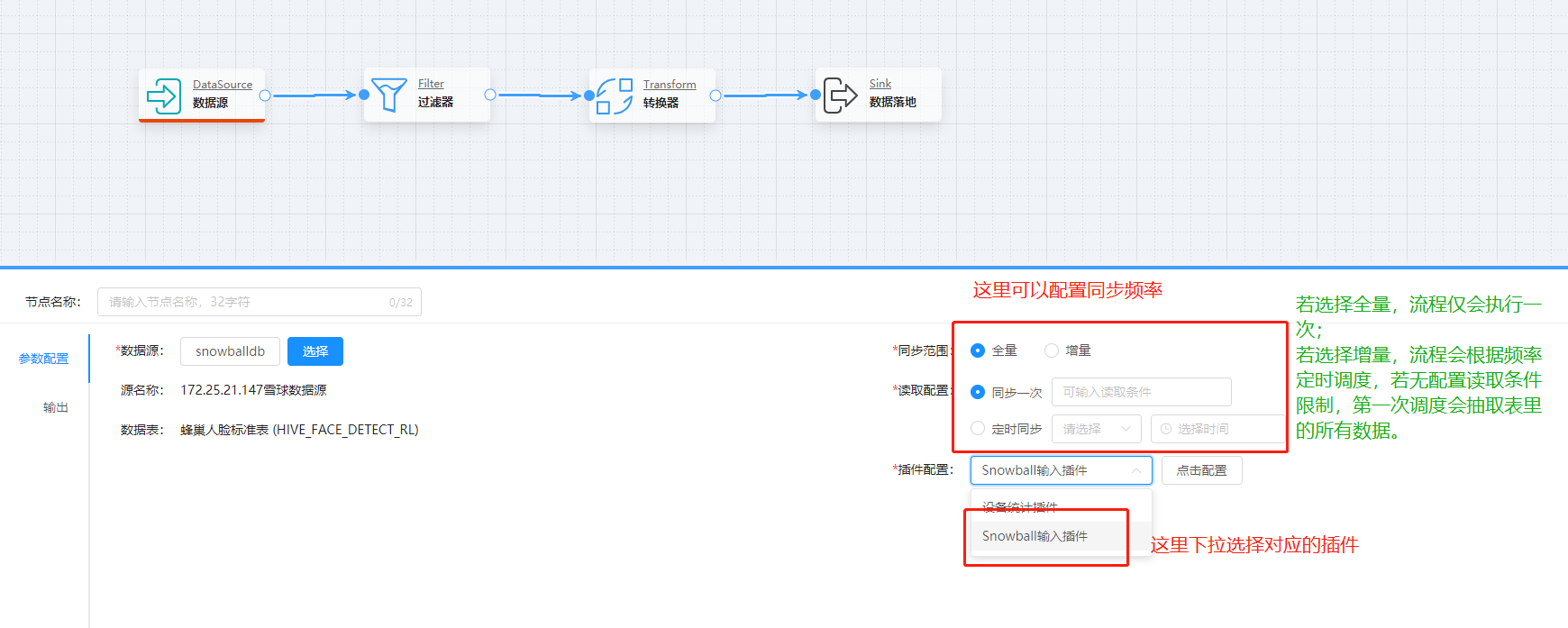


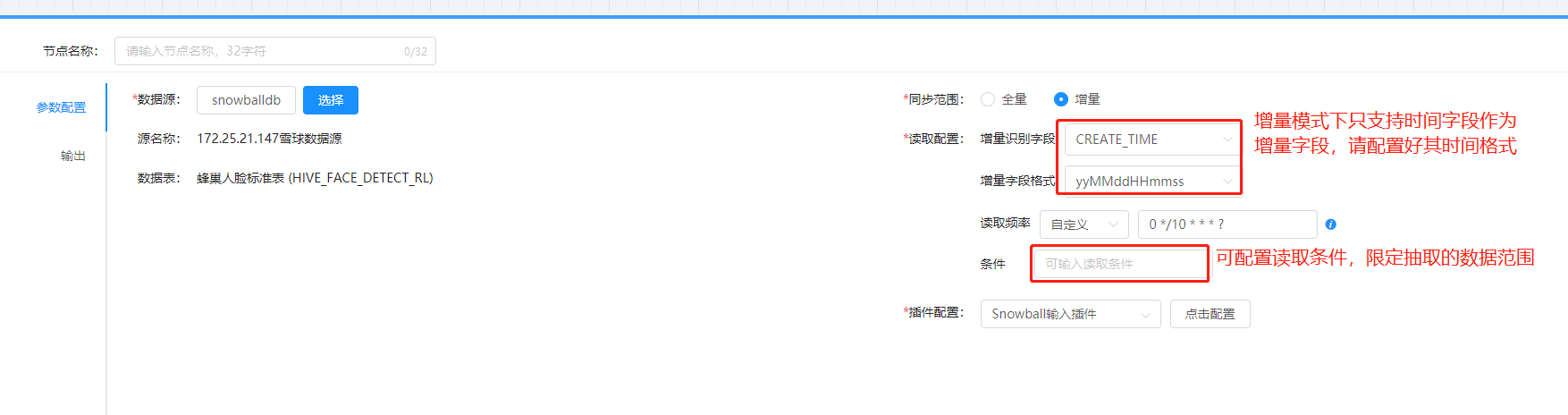
示例这里选择离线任务。



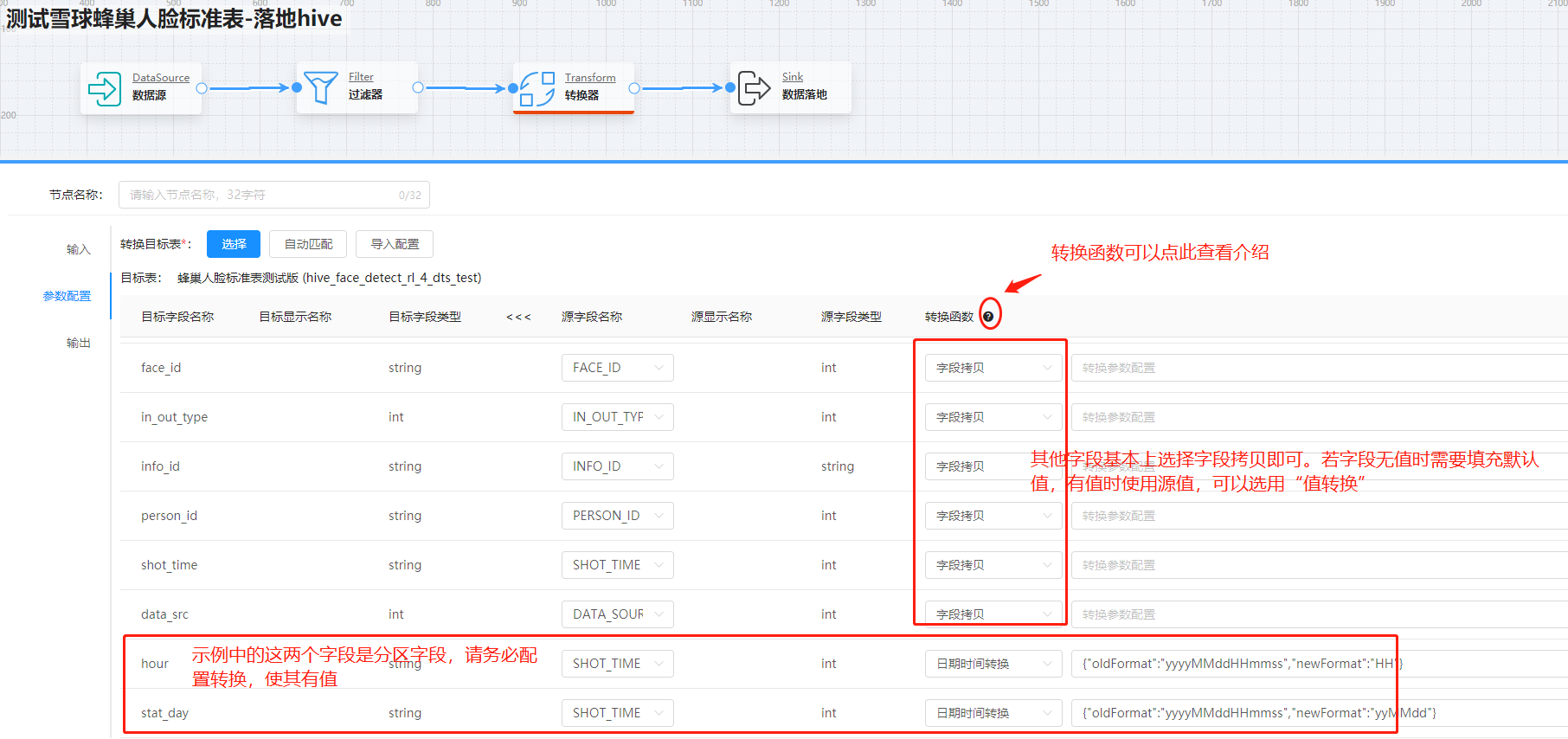
### （2）配置数据源







### （3）配置转换器

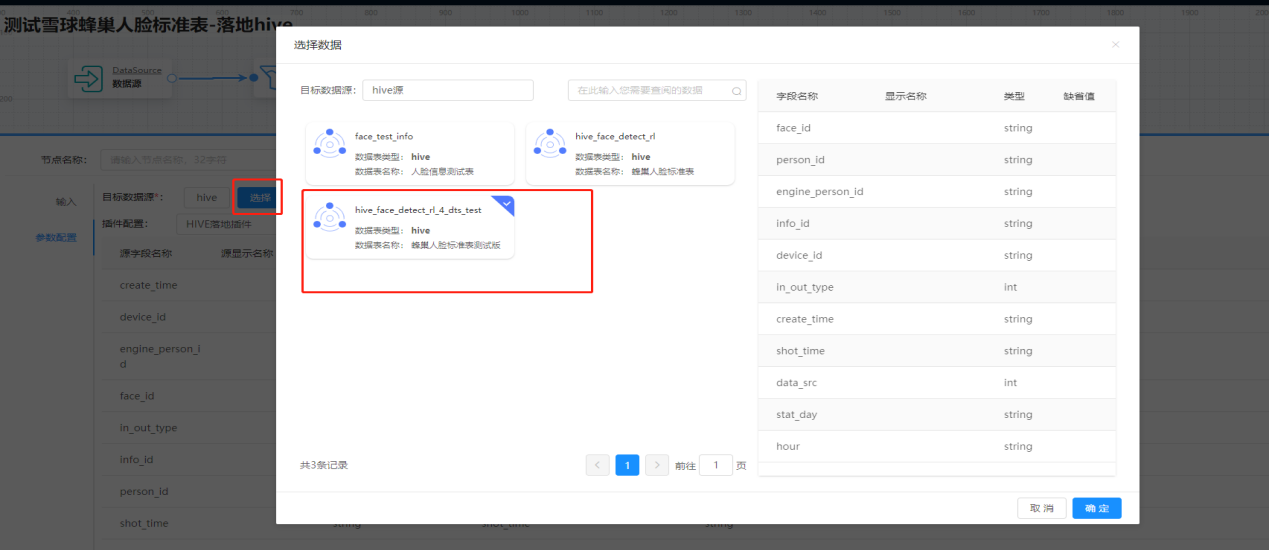


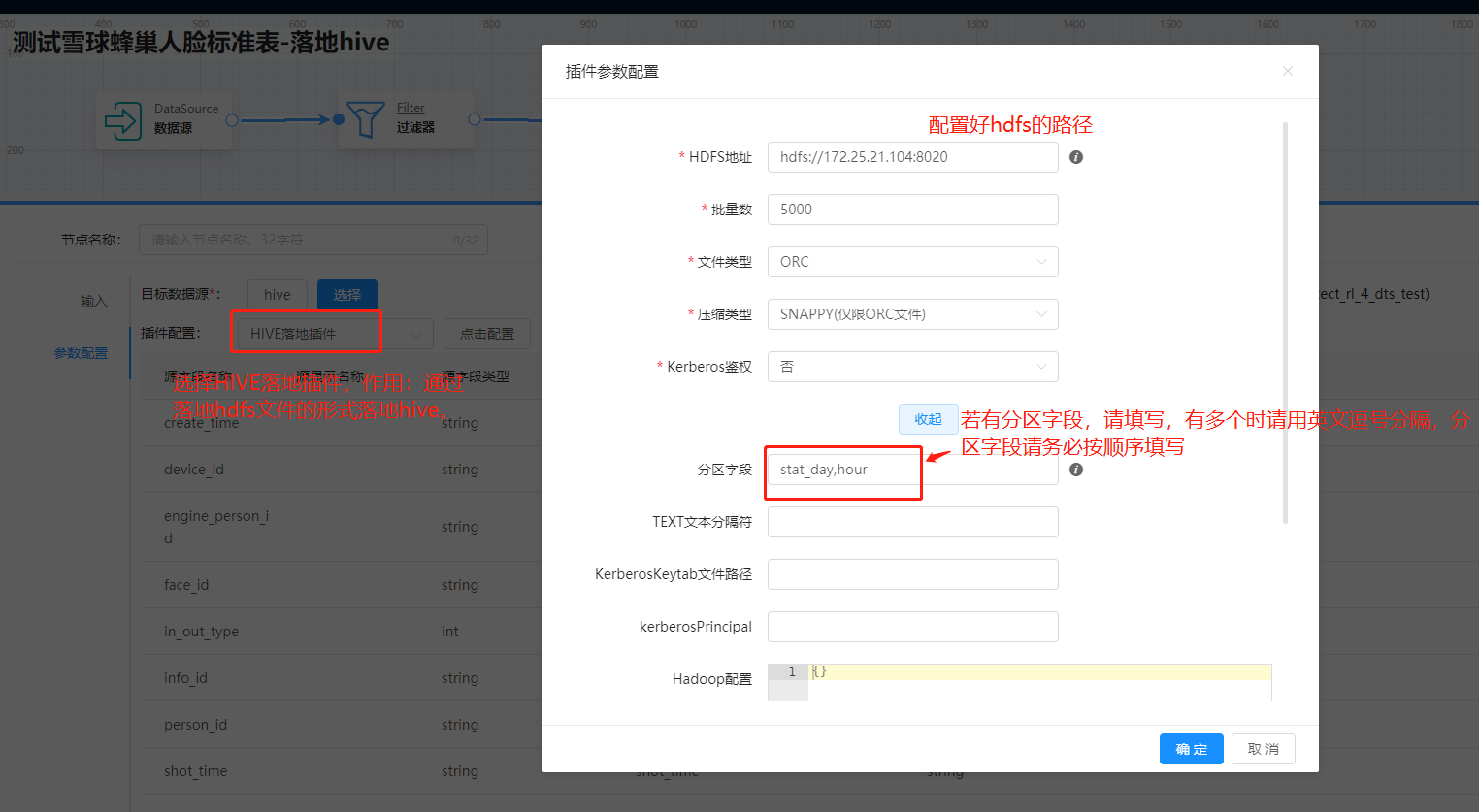
### （4）配置过滤器



### （5）配置数据落地

选择落地的hive数据源的数据集：



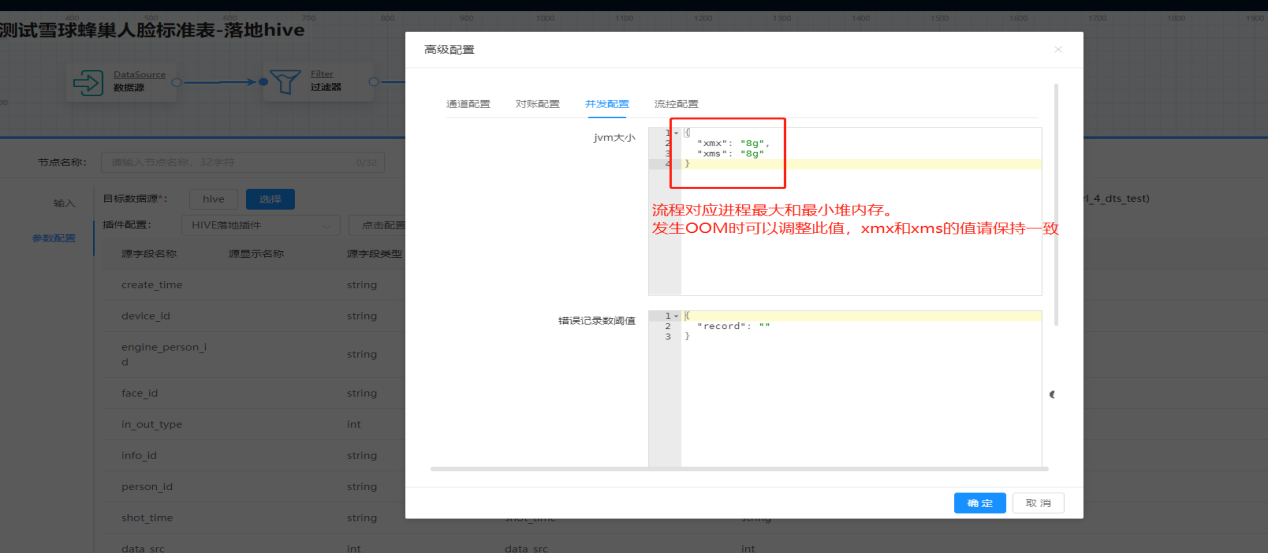


### （6）高级配置

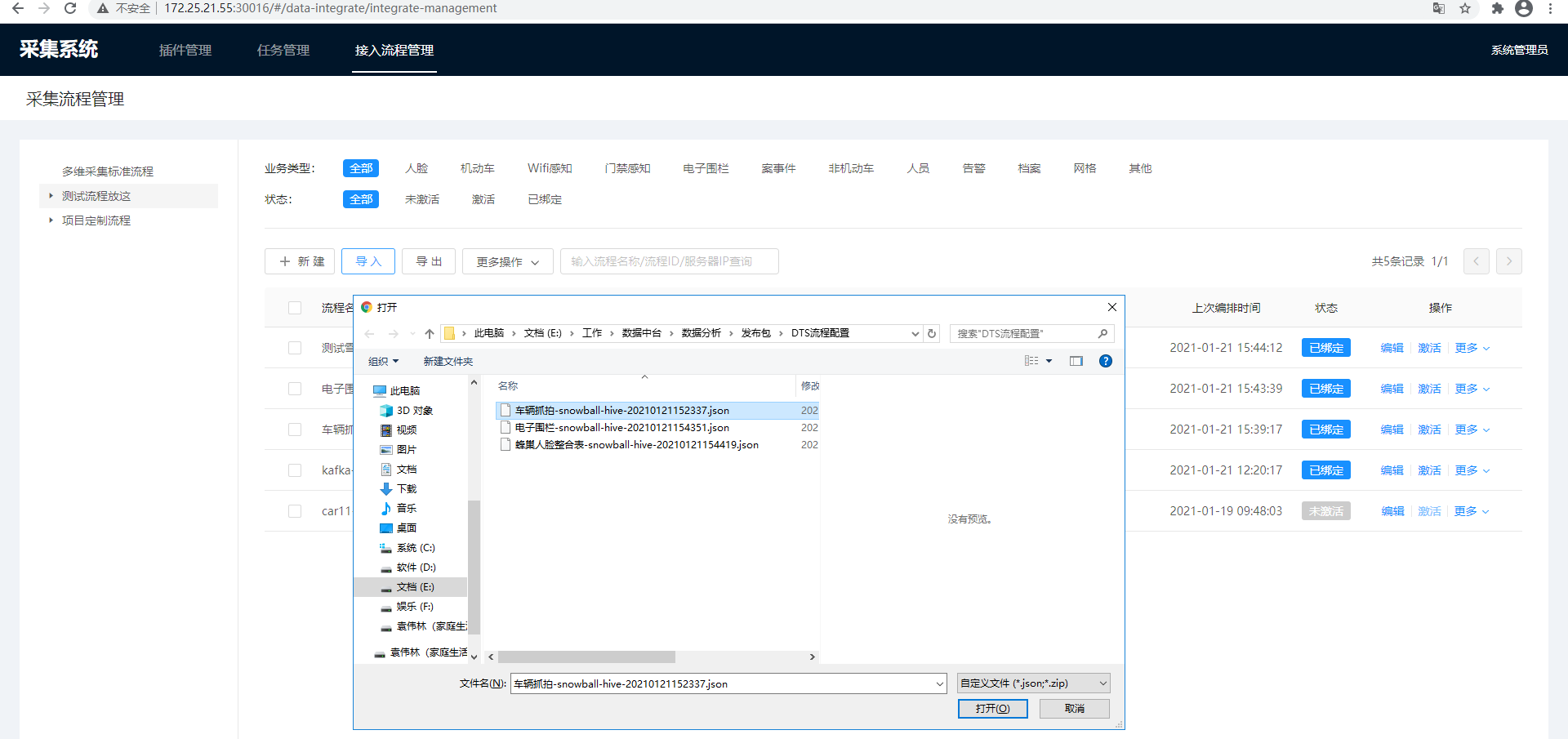
一般使用默认值即可。



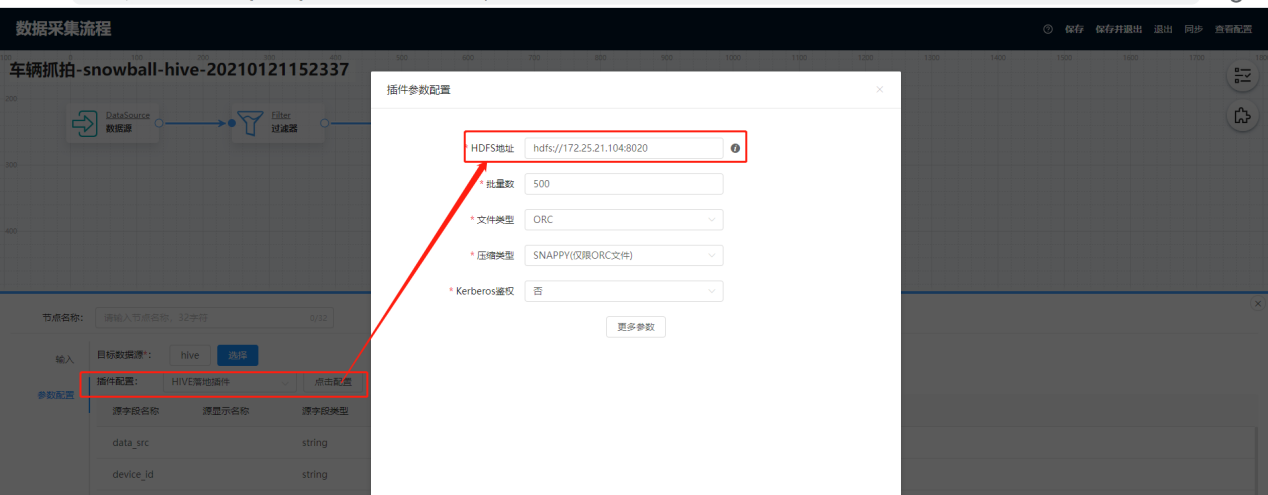




### 导入方式



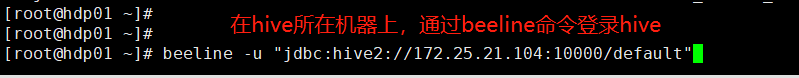
流程导入后，需要根据现场情况修改集群地址：



修改后，保存，之后激活流程即可。

## 如何查找hive分区字段？如何查询hive表记录？

### （1）beeline登录



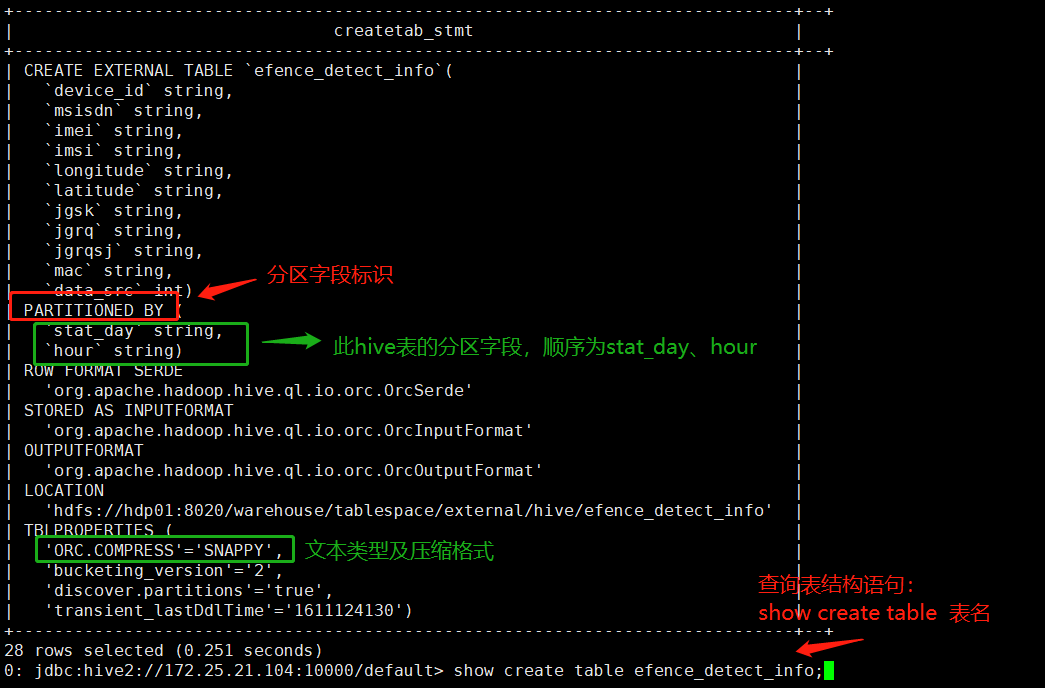
需要有jdbc路径.

### （2）如何查找hive分区字段？

输入sql：

show create table 表名

示例输出如下：



### （3）如何查询hive表记录？

直接执行sql即可，如：

