# 一、Kafka\_lob程序说明文档

## 1.功能描述

Kafka\_lob程序主要功能是python多进程消费并解析kafka中order和transaction中的二进制pb数据，通过lob算法生成快照后上传influxdb中，lob中新增两市集合竞价功能，kafka中通过时间戳定位offset保证消费当天数据。

## 2.主要任务

（1）从kafka中获取实时行情数据。

（2）解析pb二进制数据。

（3）集合竞价计算价格。

（4）调用lob算法生成快照。

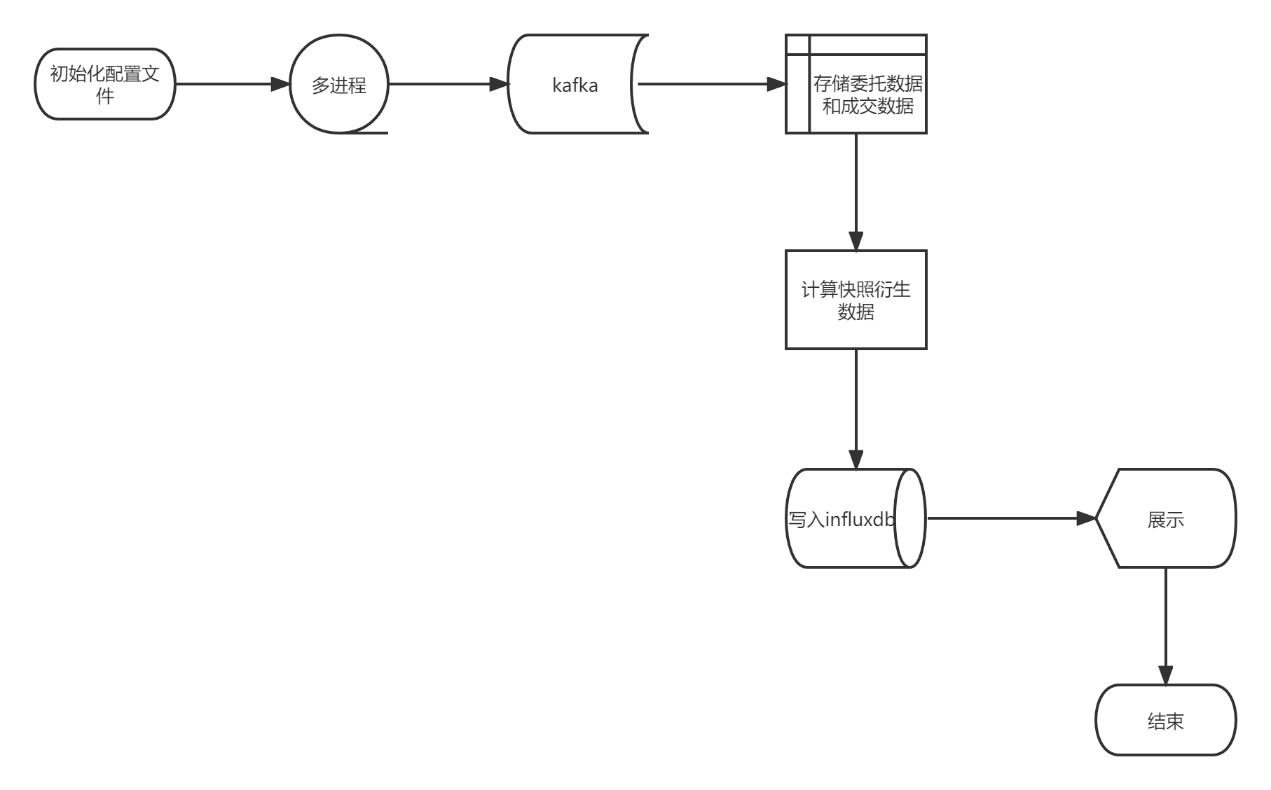
（5）将数据写入influxdb中。

（6）增加处理多只股票。

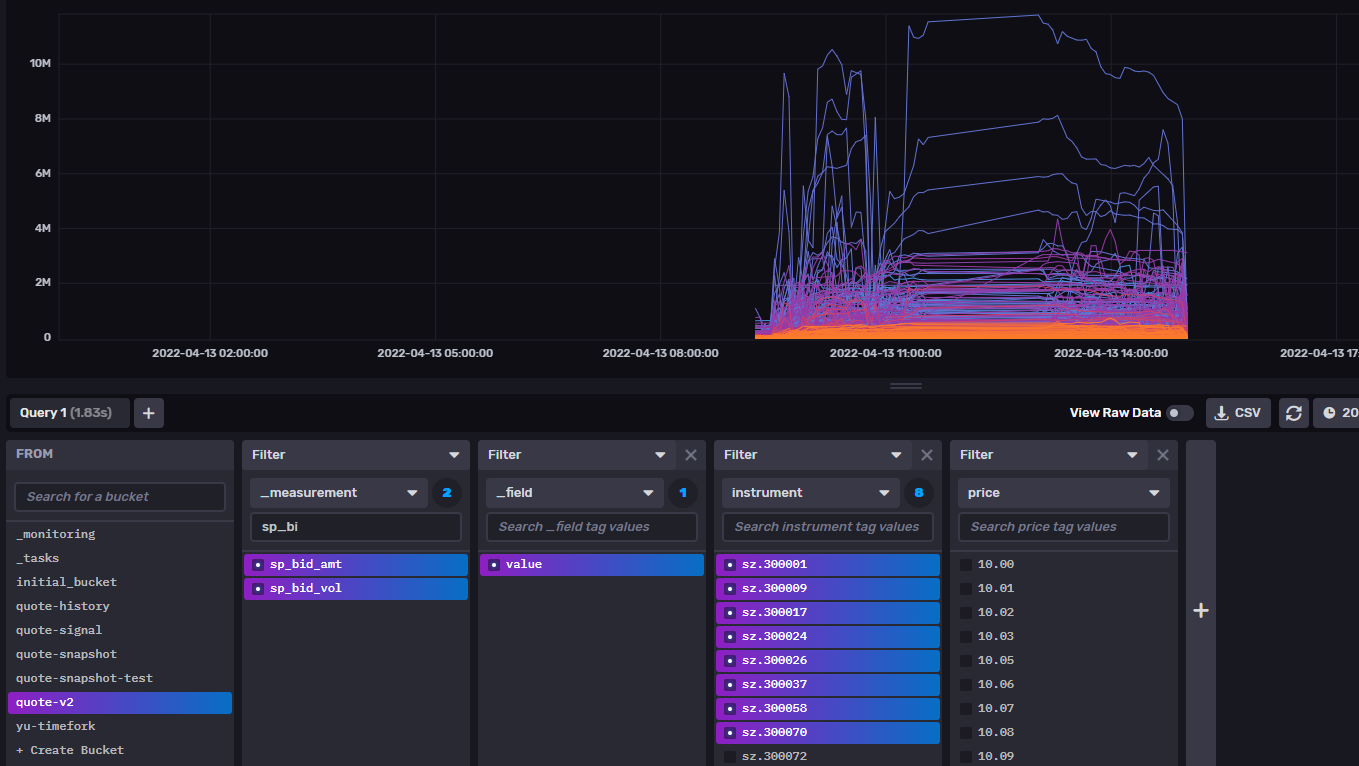
（7）增加自启动功能。

（8）日志功能。

## 3.数据结构图



## 4.效果展示



## 5.中间件

kafka、oracle、influxdb

## 6.源码

<http://192.168.101.210:7990/scm/dat/offline-tickdata.git>

## 7.部署

测试环境：

启动：python lobbuilder.py mode=realtime

## 8.依赖库



# 二、新闻爬虫程序说明文档

## 1.功能描述

新闻程序主要功能是python的scrapy框架爬取新闻信息存储到oracle数据库中，通过搭建web服务器使用http请求展示在Qt界面上，程序使用定时任务，在一周到周五早上9点到下午3点自动触发，爬取重要重要新闻的标题和网址。

## 2.爬取网站

国务院 http://www.gov.cn/guowuyuan/index.htm 动态

银保监会 http://www.cbirc.gov.cn/cn/view/pages/index/index.html 监管动态

证监会 http://www.csrc.gov.cn/ 时政要闻 证监会要闻

央行 http://www.pbc.gov.cn/ 新闻

财政部 http://www.mof.gov.cn/index.htm 财政新闻 政策发布

发改委 https://www.ndrc.gov.cn/ 新闻发布|时政要闻|通知公告

国家统计局 http://www.stats.gov.cn/ 最新发布与解读

上海交易所 http://www.sse.com.cn/ 要闻 本所动态

深圳交易所 http://www.szse.cn/ 时政要闻 深交所要闻

中国证券报

上海证券报

证券时报

证券日报

## 3.主要任务

（1）搭建爬虫框架scrapy 。

（2）模拟浏览器行为 。

（3）获取新闻数据并提取 。

（4）数据去重后写入oracle 。

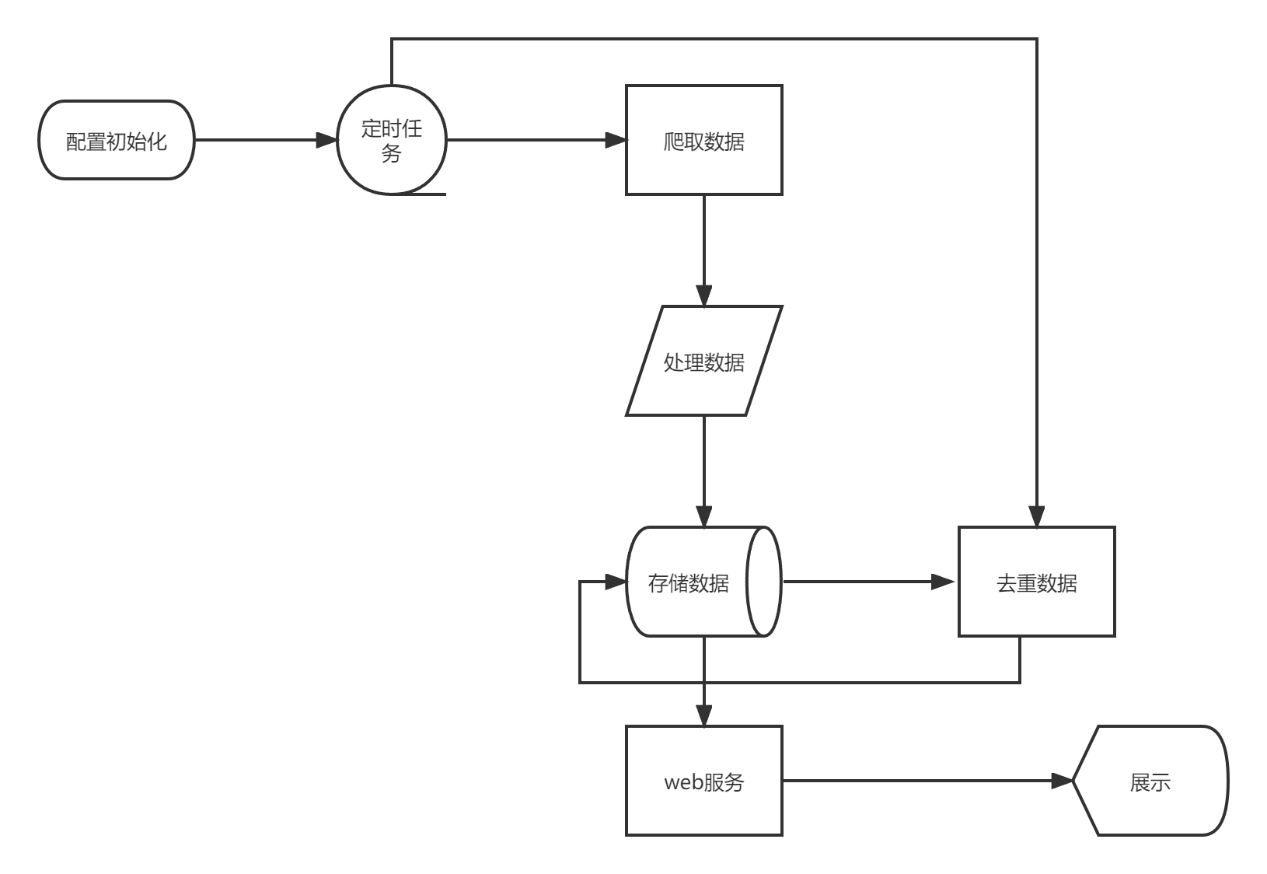
（5）使用tornado框架搭建web服务器 。

（6）通过http请求传递给qt界面 。

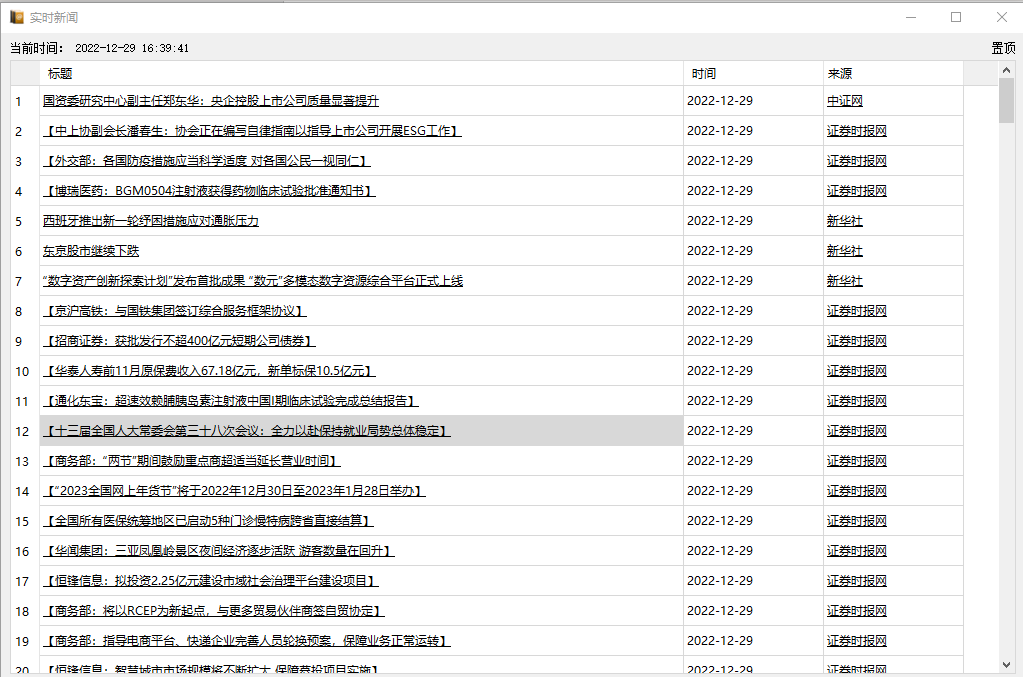
（7）定时任务处理 。

（8）日志功能。

## 4.数据结构图



## 5.效果展示



## 6.中间件

oracle、ChromeSelenium

## 7.源码

<http://192.168.101.210:7990/scm/~wangyanyan/kafka-stock-repo.git>

## 8.部署

测试环境：

启动爬虫程序：./run.sh 停止：stop.sh

启动web服务：./server.sh 停止：server\_stop.sh

日志：./log

## 9.依赖库



# 三、生成经纪商excel文件程序说明文档

## 1.功能描述

生成经纪商excel文件程序主要功能是python定时生成，A股额度.xlsx和全部持仓.xlsx两个文件通过sftp上传到指定服务器下的指定目录下，主要内容包括北上资金说明(亿)，股票至今十大买入卖出(万)，股票近10日买入量(万股)，当日买入为前三日两倍以上(万股)，持仓(万)，预测(万)，观察(万)，股票近10日买入额(万)，股票近10日买入额前五买入(万)，股票近10日买入额连续五日大于0(万)，板块近10日买入额(万)，经纪商近10日持仓持仓市值(万)，股票买卖相对自由流通股近10日变动率(%)，板块买卖相较于北上资金持有板块票总和近10日变动率(%)，板块买卖相较于北上资金持有板块票总和近10日变动率(%)，股票买卖相对昨日持仓变化率(%)，股票近十日股票持仓相对流通股比例(%)。并对excel文件中的单元格数据用涂颜色标记。

## 2.主要任务

（1）北上资金说明(亿)

（2）股票至今十大买入卖出(万)

（3）股票近10日买入量(万股)

（4）当日买入为前三日两倍以上(万股)

（5）持仓(万)

（6）预测(万)

（7）观察(万)

（8）股票近10日买入额(万)

（9）股票近10日买入额前五买入(万)

（10）股票近10日买入额连续五日大于0(万)

（11）板块近10日买入额(万)

（12）经纪商近10日持仓持仓市值(万)

（13）股票买卖相对自由流通股近10日变动率(%)

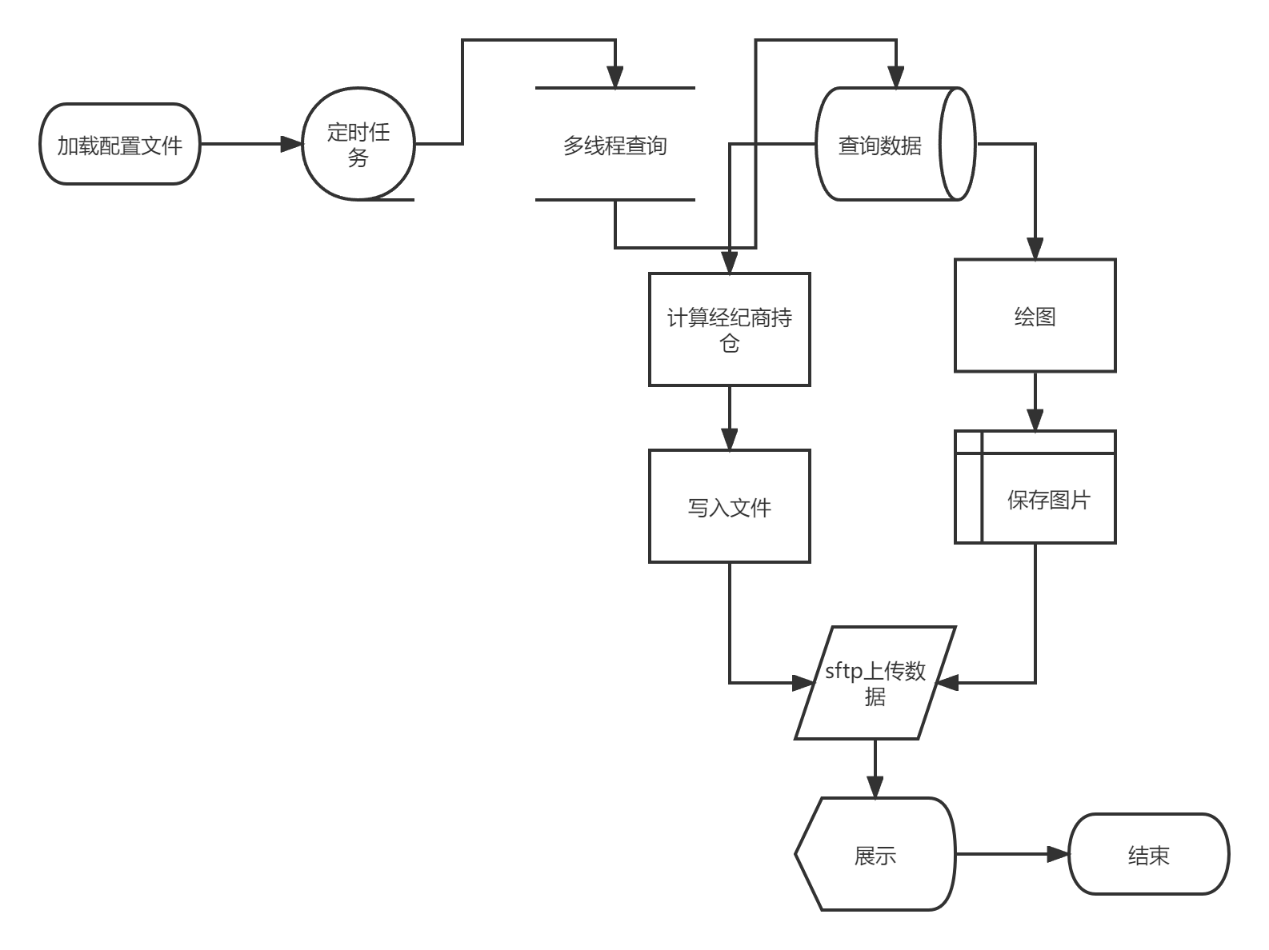
（14）板块买卖相较于北上资金持有板块票总和近10日变动率(%)

（15）板块买卖相较于北上资金持有板块票总和近10日变动率(%)

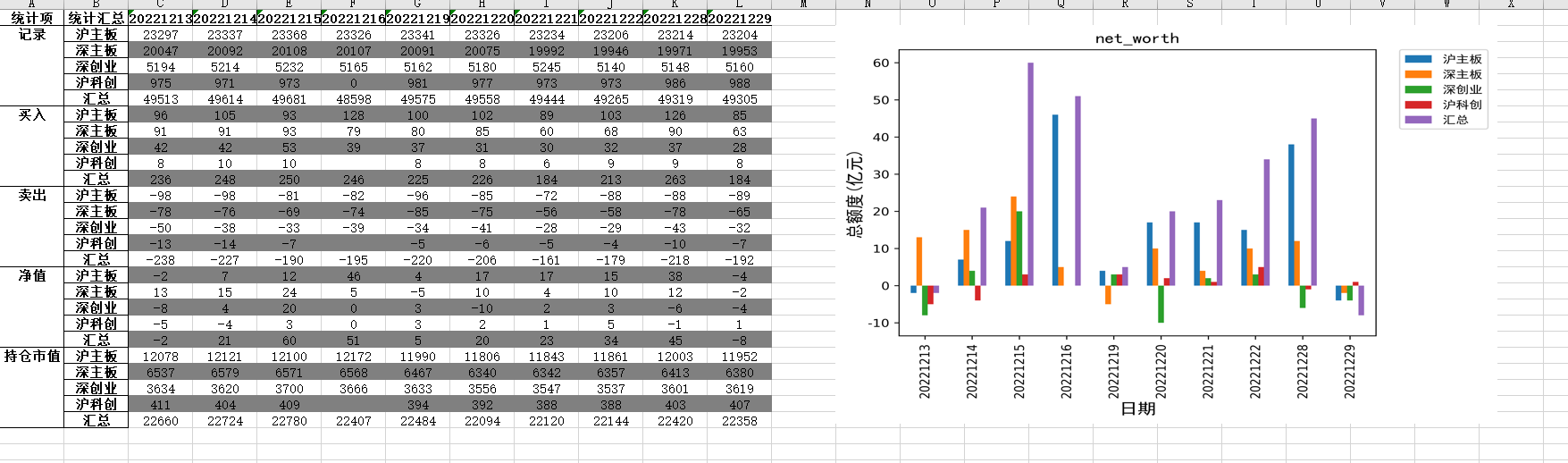
（16）股票买卖相对昨日持仓变化率(%)

（17）股票近十日股票持仓相对流通股比例(%)

## 3.数据结构图



## 4.效果展示



## 5.中间件

Oracle，pandas

## 6.源码

<http://192.168.101.210:7990/scm/~wangyanyan/xlsm_img_sftp_upload.git>

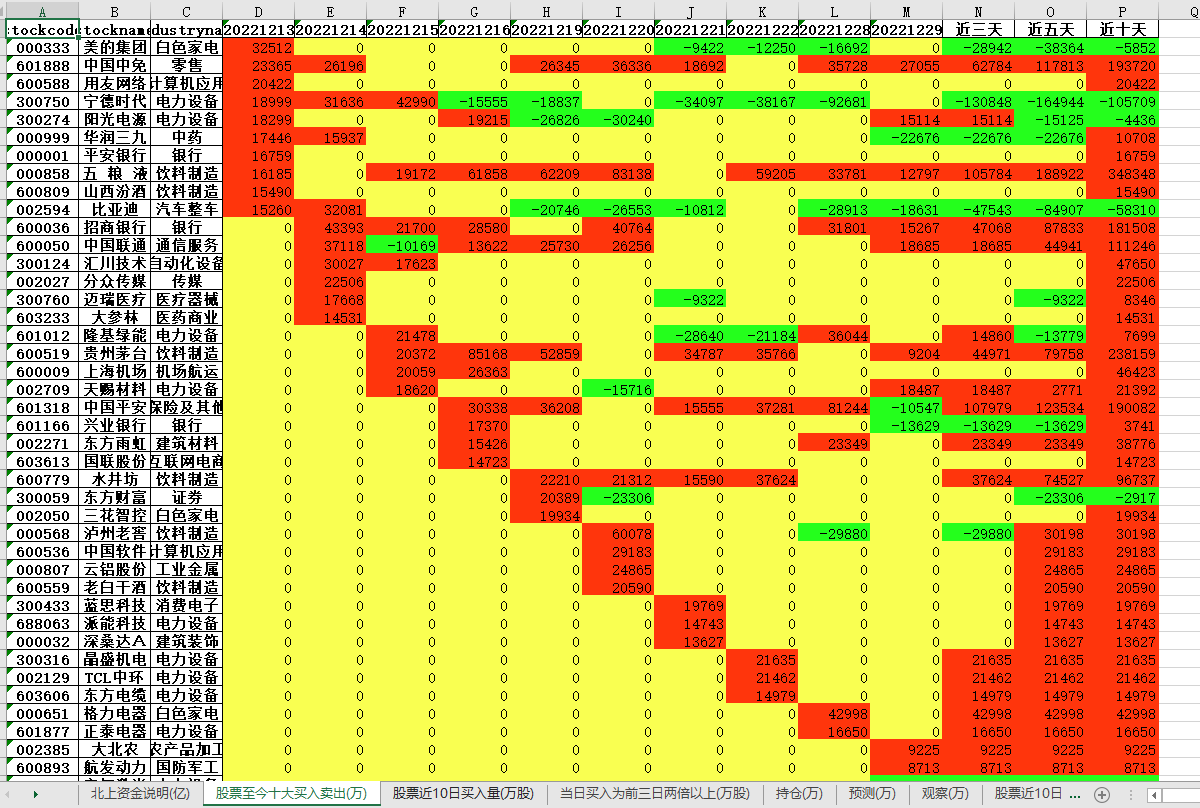
## 7.部署

测试环境：

启动：./start.sh 停止：./stop.sh

## 8.依赖库



****

# 四、influxdb\_img生成灰度图程序说明文档

## 1.功能描述

influxdb\_img生成灰度图程序主要功能是查询influxdb中的- sp\_ask\_amt

- sp\_ask\_vol

- sp\_bid\_amt

- sp\_bid\_vol，整合时间戳和价格后生成一个二维矩阵再通过matplotlib生成灰度图。

## 2.主要任务

（1）导入股票的涨停和跌停价格 。

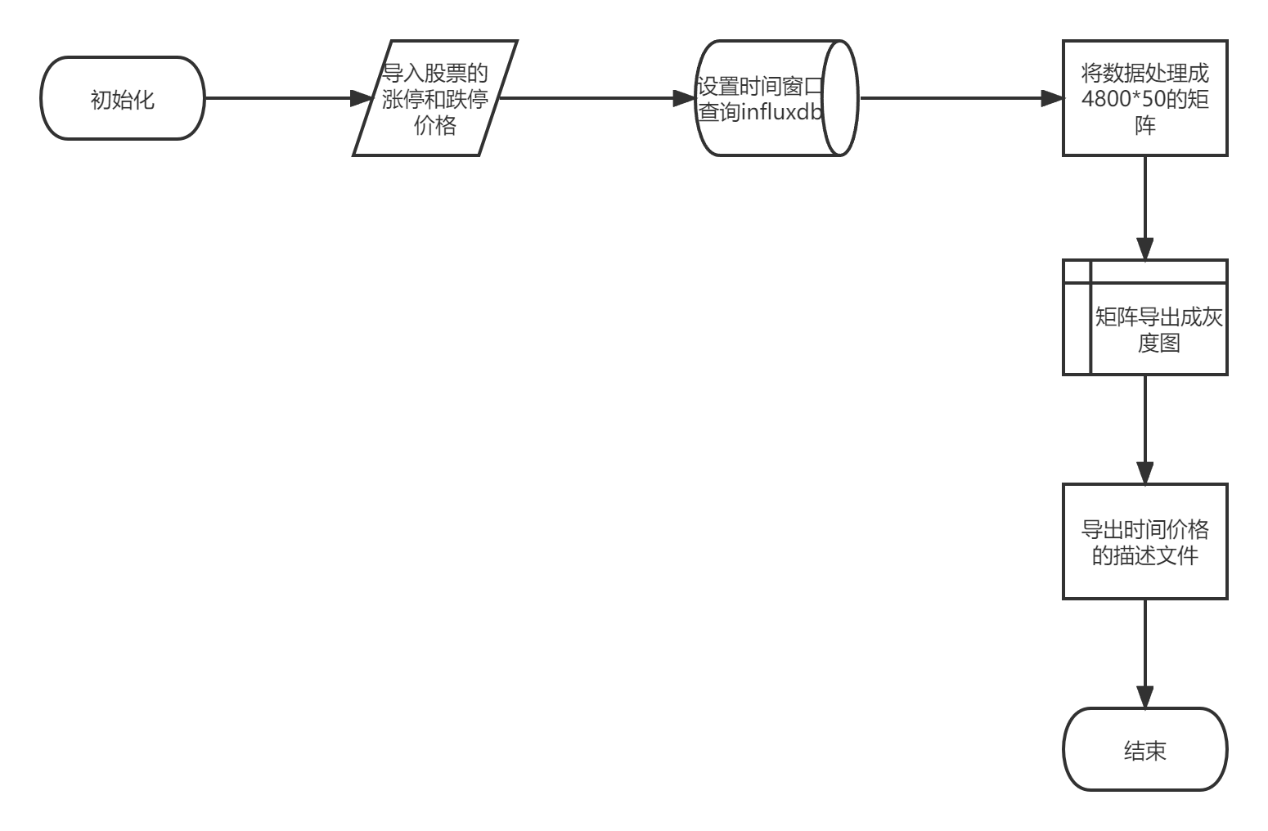
（2）设置时间窗口查询influxdb 。

（3）将数据处理成4800\*50的矩阵。

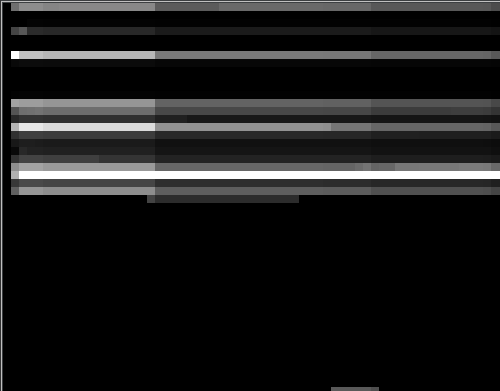
（4）矩阵导出成灰度图 。

（5）导出时间价格的描述文件。

## 3.数据结构图



## 4.效果展示



## 5.中间件

Influxdb，matplotlib

## 6.源码

http://192.168.101.210:7990/scm/~wangyanyan/influx-img.git

## 7.部署

测试环境：

启动程序：python influx\_img\_reval\_v2.py

## 8.依赖库



# 五、influxdb\_parser\_save\_bin命令行导出二进制文件程序说明文档

## 1.功能描述

## 2.主要任务

（1）定义相关命令 。

（2）处理单个文件 。

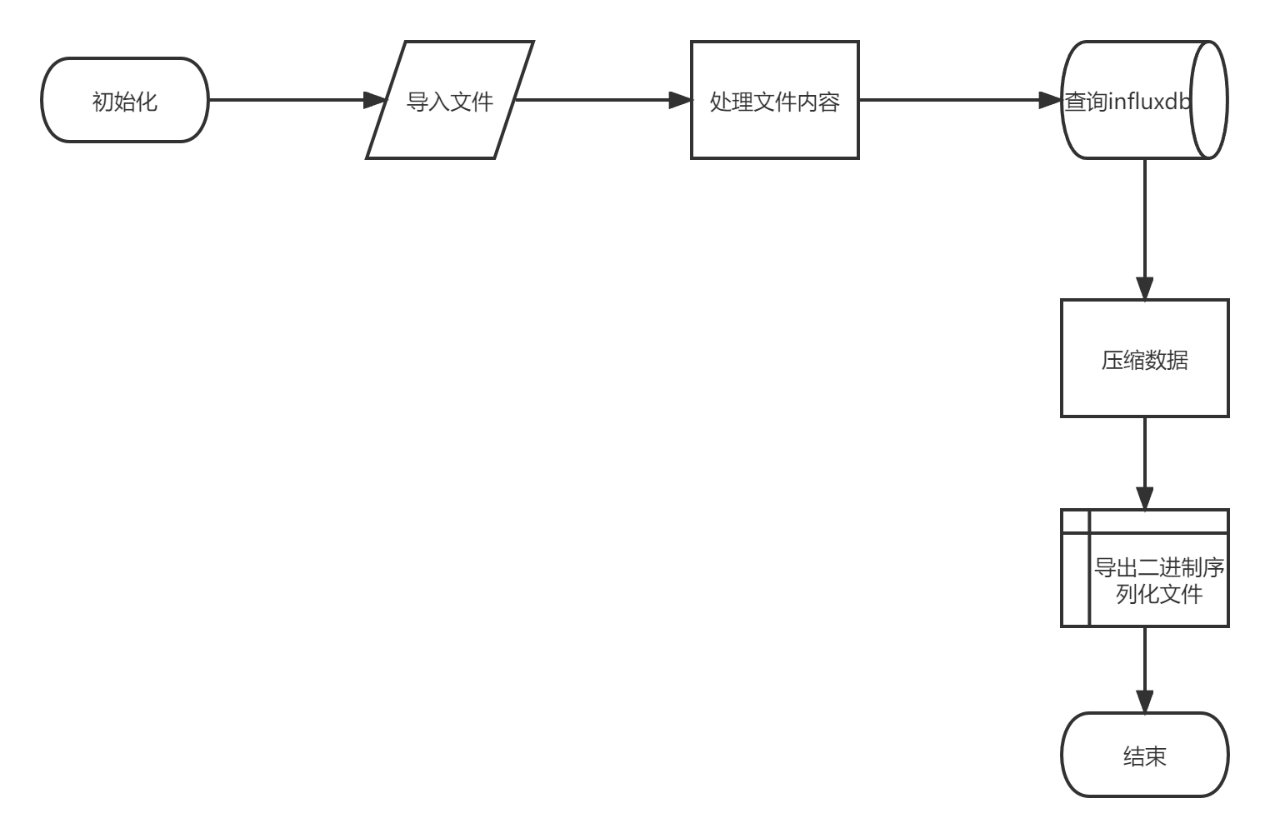
（3）批量获取文件 。

（4）查询influxdb 。

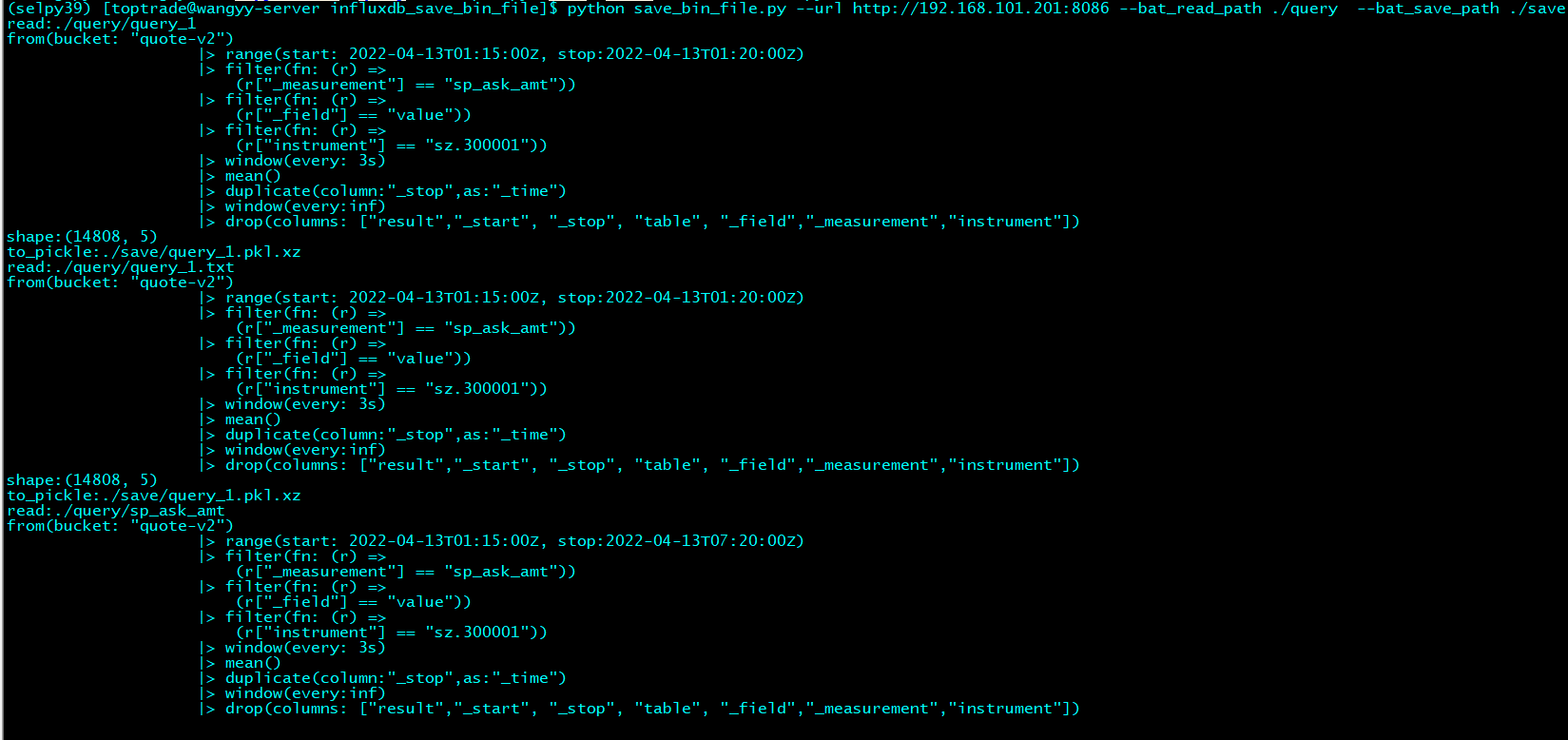
（5）压缩数据 。

（6）导出二进制序列化文件 。

## 3.数据结构图



## 4.效果展示



## 5.中间件

Influxdb，matplotlib

## 6.源码

http://192.168.101.210:7990/scm/~wangyanyan/influxdb\_save\_bin\_file.git

## 7.部署

测试环境：

启动程序：python [save\_bin\_file.py](http://192.168.101.210:7990/users/wangyanyan/repos/influxdb_save_bin_file/browse/save_bin_file.py)

## 8.依赖库



# 六、TWAP-PoV服务程序说明文档

## 1.功能描述

本程序为服务端程序，通过接受kafka消息中间件中的TWAP-PoV的JSON命令消息后，根据相关算法调用柜台交易接口，实现TWAP-PoV算法下单，订单查询，成交查询，持仓查询，撤单交易算法的处理。

## 2.主要任务

（1）设计接口格式JSON结构。

（2）异步接入kafka，向指定topic写入请求数据。

（3）获取命令数据对接算法处理。

（4）实现TWAP算法下单。

（5）实现PoV算法下单。

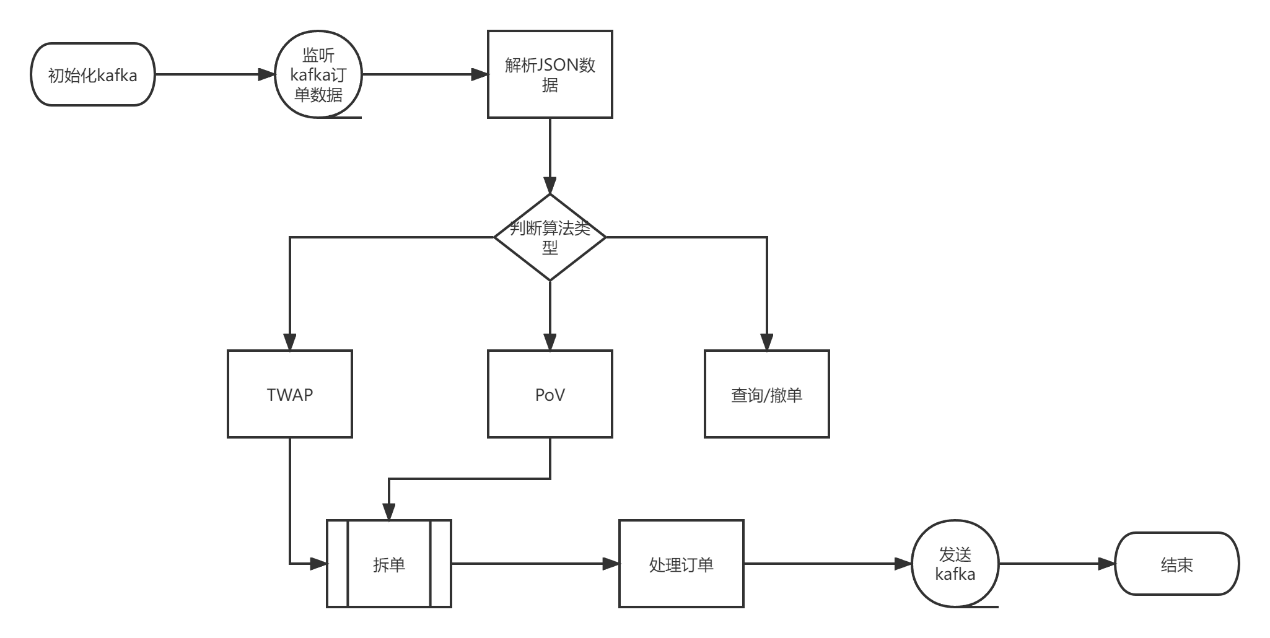
（6）实现订单查询。

（7）实现批量撤单。

（8）实时获取全市场成交量数据。

（9）实时获取股票价格。

## 3.数据结构图



## 4.接口说明

## 4.1 时间优先（TWAP）

请求JSON结构：

{  
    **"id"**:**1**, //job id   
    **"vol"**:**100**, //下单数量

**"period"**:**10000，** //下单周期 毫秒

**"step"**:5000, //步长 毫秒  
    **"cancel "**:**5000**, //撤单周期 毫秒  
    **"bsdir"**:**1**, //报单方向 1 买入 2 卖出  
    **"price"**:**8.0**, //报单价格 price为空时，获取市场价  
    **"type"**:**1**, //算法类型 1 TWAP 2 PoV  
    **"stockcode "**:**"300081"**, //证券代码  
    **"exchangeid"**:**1**, //交易市场 1 上交所 2 深交所   
    **"limit"**:**1** //0 非限制完成 1 限制完成   
}

## 4.2成交量优先（PoV）

请求JSON结构：

{  
    **"id"**:**1**, //job id   
    **"vol"**:**100**, //下单数量

**"period"**:**10，** //成交熟虑 10%

**"step"**:5000, //步长 毫秒  
    **"cancel "**:**5000**, //撤单周期 毫秒  
    **"bsdir"**:**1**, //报单方向 1 买入 2 卖出  
    **"price"**:**8.0**, //报单价格 price为空时，获取市场价  
    **"type"**:**2**, //算法类型 1 TWAP 2 PoV  
    **"stockcode "**:**"300081"**, //证券代码  
    **"exchangeid"**:**1**, //交易市场 1 上交所 2 深交所   
    **"limit"**:**1** //0 不计入本账户交易量 1 计入本账户交易量   
}

## 5.中间件

pandas，redis，fastapi，kafka，nginx，django

## 6.源码

<http://192.168.101.210:7990/scm/~wangyanyan/async_kafka_rep.git>

http://192.168.101.210:7990/scm/~wangyanyan/django\_admin.git

## 7.部署

cd到[src](http://192.168.101.210:7990/users/wangyanyan/repos/async_kafka_rep/browse/src)/[fast\_api](http://192.168.101.210:7990/users/wangyanyan/repos/async_kafka_rep/browse/src/fast_api)/[fastapi-react-dev](http://192.168.101.210:7990/users/wangyanyan/repos/async_kafka_rep/browse/src/fast_api/fastapi-react-dev)/[app](http://192.168.101.210:7990/users/wangyanyan/repos/async_kafka_rep/browse/src/fast_api/fastapi-react-dev/app)/**app**/

启动程序：./start.sh

停止程序：./stop.sh

django\_admin：

启动程序：./start.sh

停止程序：./stop.sh

## 8.依赖库



# 七、板块开闭市涨幅档次排名说明文档

## 1.功能描述

本程序为服务端程序，读取年初至今板块开闭市涨幅档次排名，包括板块代码，板块名称，涨幅，现价，总金额，总市值，流通市值等数据，生成每日板块排名文档后上传，开闭市涨幅，开闭市档次，开闭市排名，开市档次着色处理，闭市档次着色处理，开市涨幅排名着色处理，闭市涨幅着色处理，开市排名着色处理，闭市排名着色处理。

## 2.主要任务

（1）开闭市涨幅

（2）开闭市档次

（3）开闭市排名

（4）开市档次着色处理

（5）闭市档次着色处理

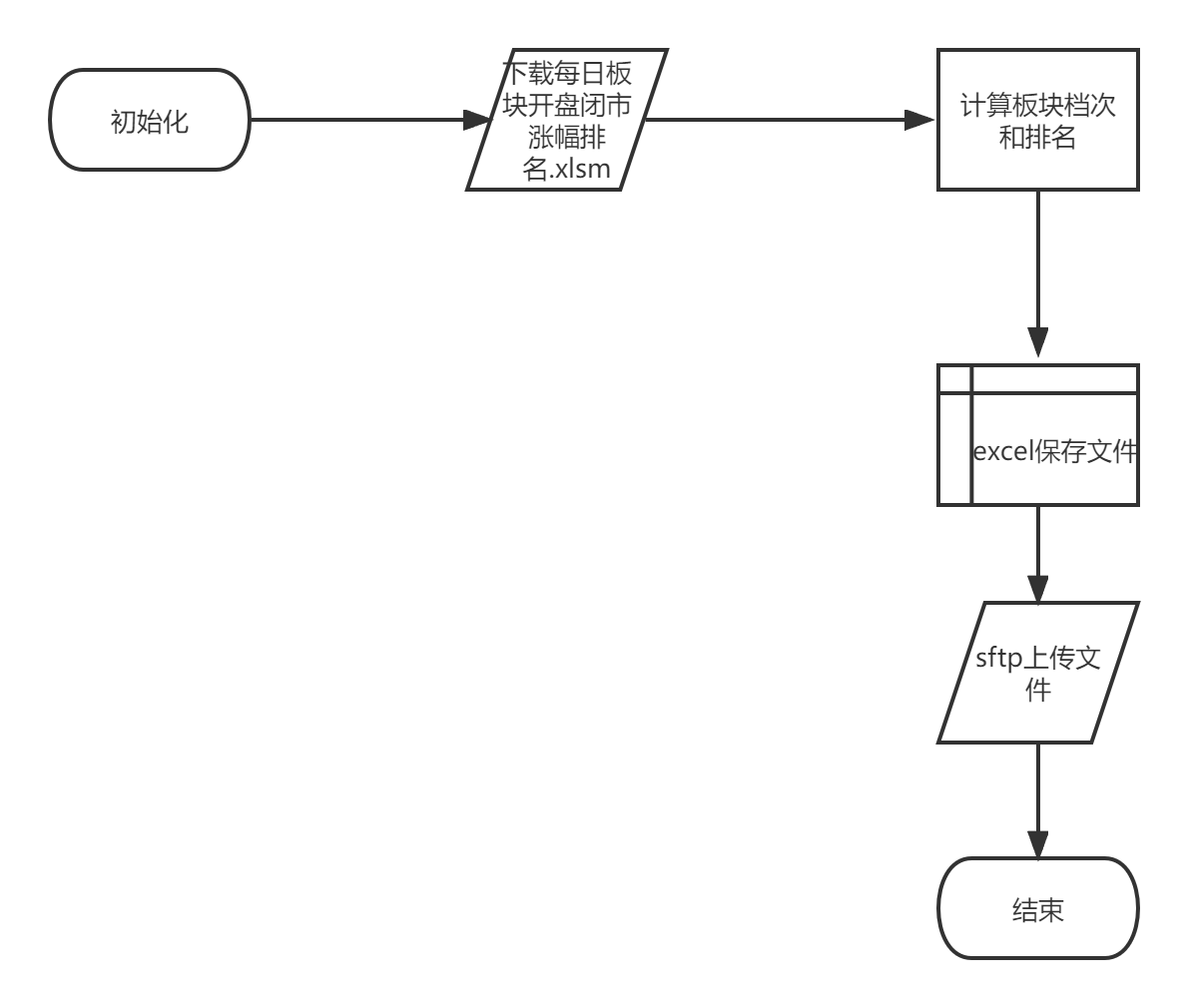
（6）开市涨幅排名着色处理

（7）闭市涨幅着色处理

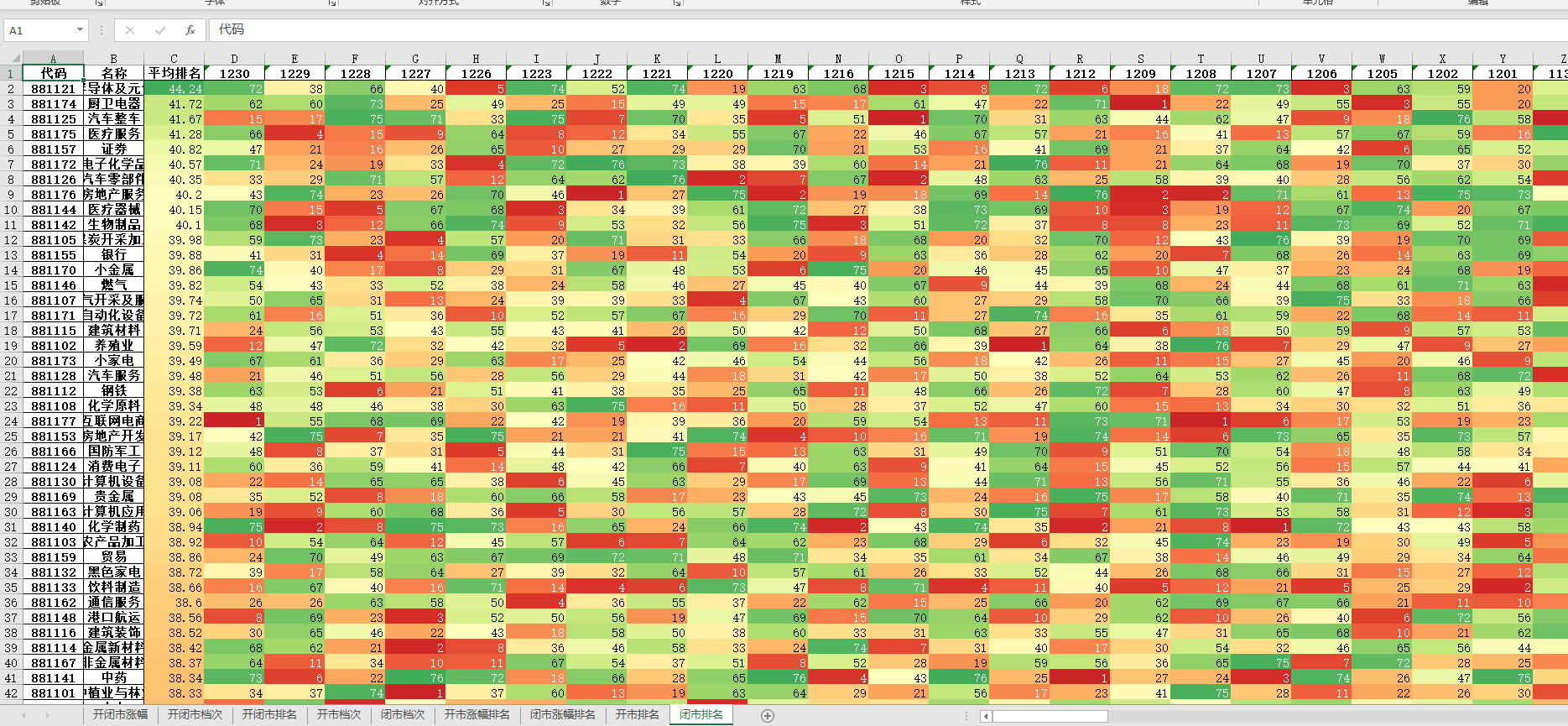
（8）开市排名着色处理

（9）闭市排名着色处理

## 3.数据结构图



## 3.效果展示



Influxdb，matplotlib

## 4.中间件

Influxdb，matplotlib

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 **HK\_data**

启动程序：./start.sh

停止程序：./stop.sh

## 7.依赖库



# 八、板块分钟线涨幅档次排名说明文档

## 1.功能描述

交易日盘中实时抓取同花顺问财二级行业,涨幅数据，计算排名和档次数值，后写入oracle和influxdb中配置Grafana进行数据展示。

## 2.主要任务

（1）盘中爬取同花顺问财财二级行业,涨幅数据

（2）异步爬虫，绕开反爬机制。

（3）设置时间节点重复获取错误数据。

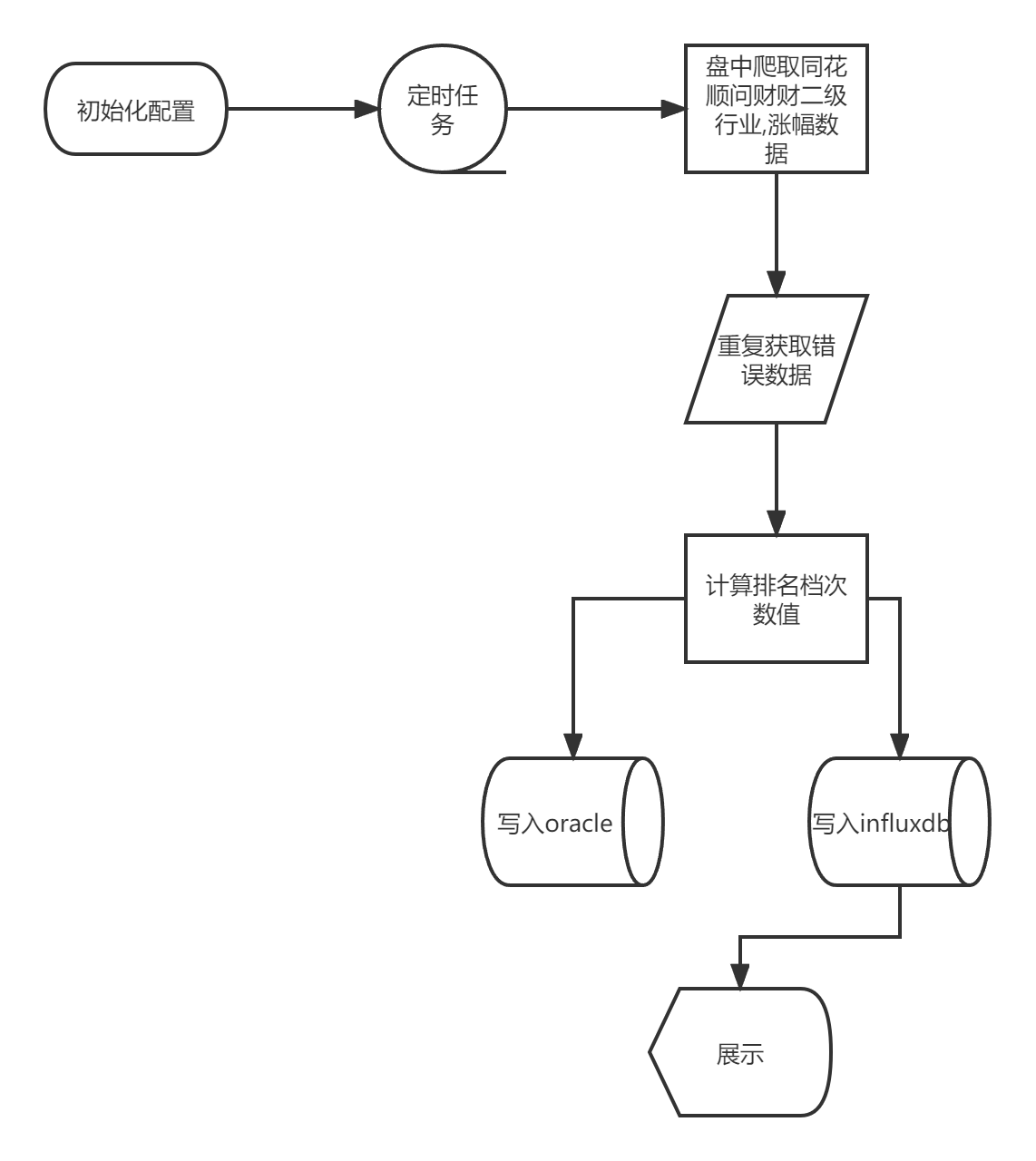
（4）板块代码，板块名称，涨幅，现价，总金额，总市值，流通市值等数据写入orale。

（5）计算排名。

（6）计算档次数值。

（7）板块名称，排名，档次数值数据写入influxdb。

## 3.数据结构图



## 3.效果展示



## 4.中间件

Influxdb，matplotlib

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 **main**

启动程序：./start.sh

停止程序：./stop.sh

## 7.依赖库



# 八、TWAP-PoV前端界面说明文档

## 1.功能描述

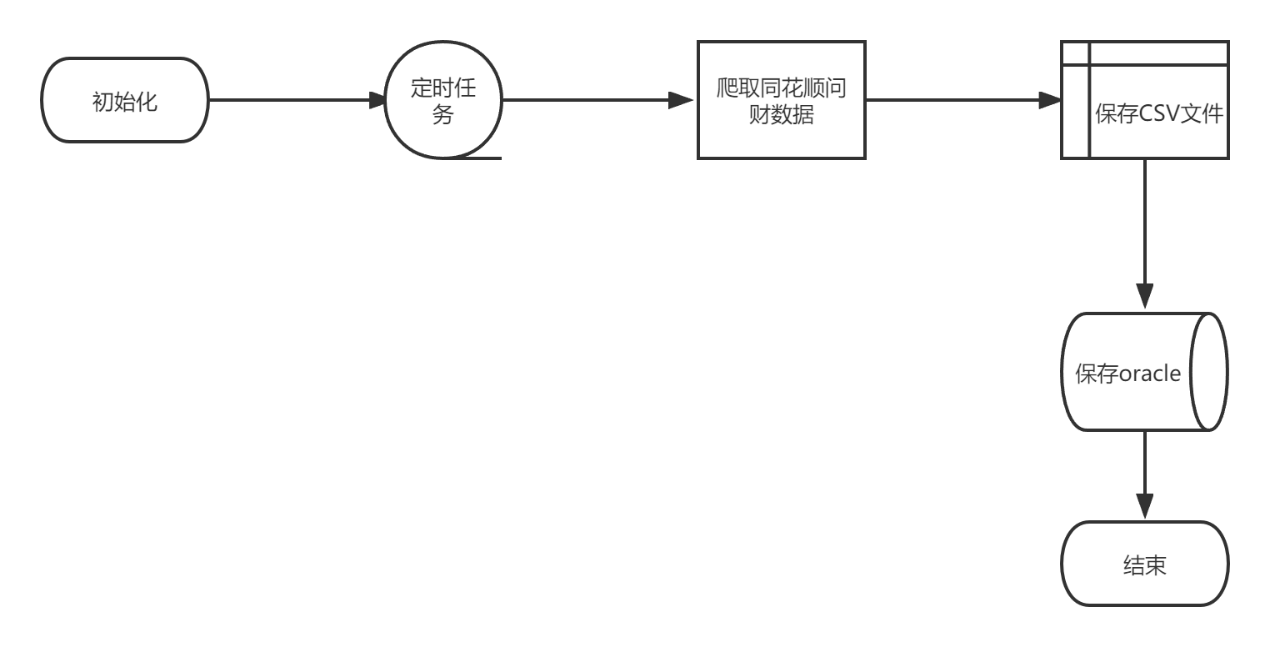
设计前端页面，绘制原型图。

# 九、同花顺板块指数-成交量-总额股票行业分类-主营-概念文件导入说明文档

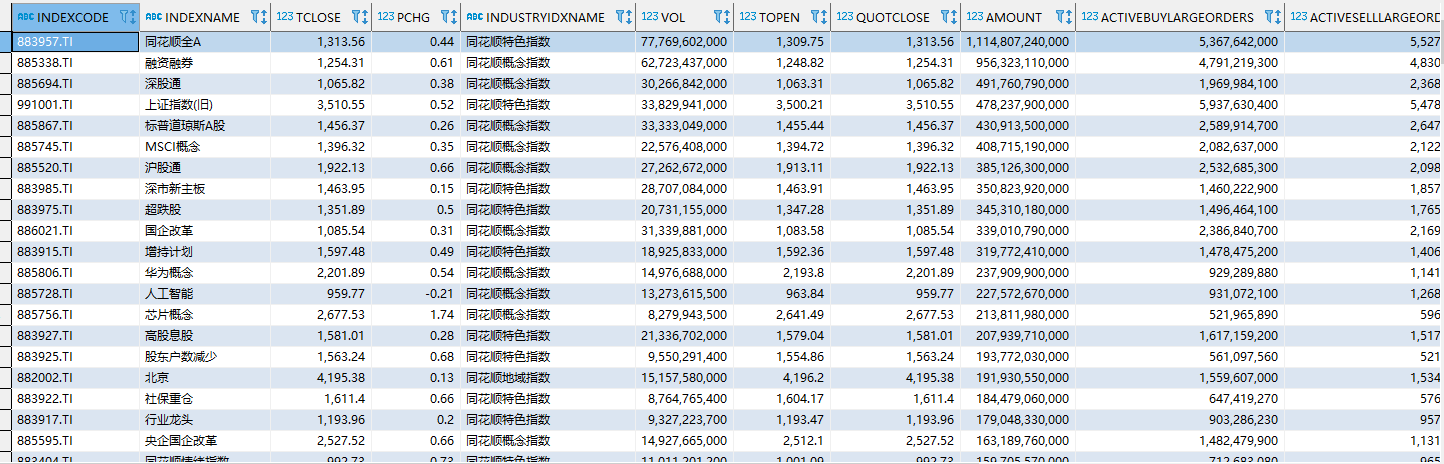
## 1.功能描述

同花顺板块指数-成交量-总额股票行业分类-主营-概念文件导入数据库，使用python aiohttp每日手动跑爬取同花顺问财数据，不存到本地csv文件中，检查数据正确后再导入oracle数据库中。

## 2.数据结构图



## 3.效果展示





## 4.中间件

pandas，oracle

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 **excel\_to\_oracle**

启动程序：python [excel\_oracle\_file\_data.py](http://192.168.101.210:7990/users/wangyanyan/repos/ths-hk-data/browse/excel_to_oracle/excel_oracle_file_data.py)

## 7.依赖库

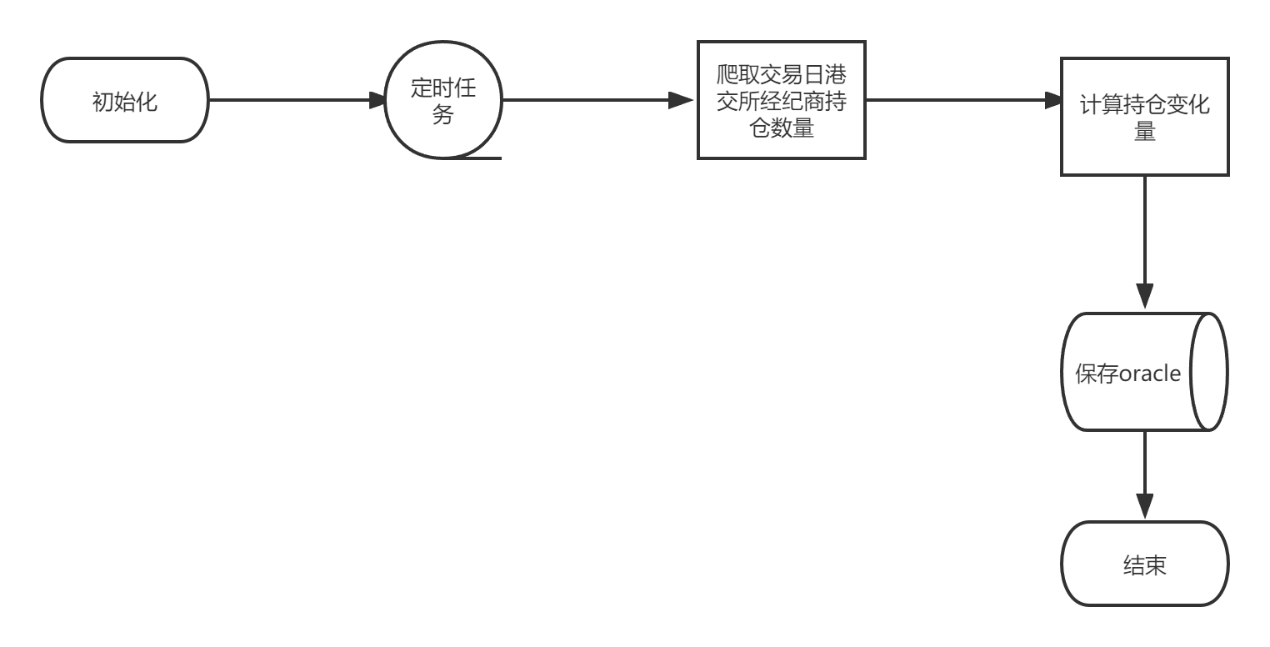


# 十、港交所经纪商持仓数量爬取说明文档

## 1.功能描述

交易日港交所经纪商持仓数量爬取后数据入库，首先保证数据的正确性以及没有重复数据后调用oracle数据库函数进行数据去重，使用merge方法计算持仓变化数量。

## 2.数据结构图



## 3.效果展示



## 4.中间件

pandas，aiohttp，oracle

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 **HK\_data**

启动程序：./start.sh

停止程序：./stop.sh

## 7.依赖库

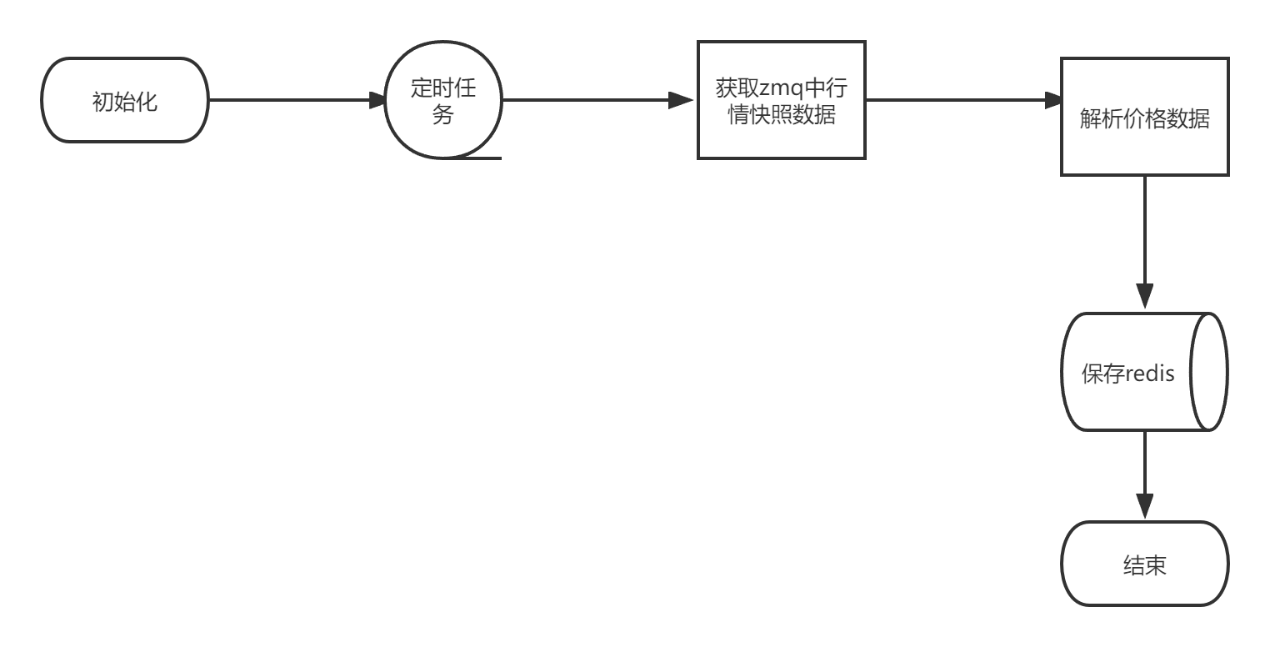


# 十一、zmq\_redis同步实时行情股票价格，交易量说明文档

## 1.功能描述

通过zmq订阅全市场股票的实时快照数据，将股票价格，交易量以异步的形式写入到redis中，后续为市价单提供数据。

## 2.数据结构图



# 十二、交易系统后端接口说明文档

## 1.功能描述

使用python中fastapi框架搭建交易系统后台服务主要包括登录，注册，token登录，资金查询，持仓查询，逐笔查询，撤单，新增策略，更新策略，更新策略状态，删除策略，查询策略列表，添加组合名称，查询组合列表，获取空行id，更新组合，更新组合状态，删除组合，下单，退出，用户列表，用户账号列表，用户拥有账号查询，用户权限增加，用户权限更新，用户权限状态更新，用户权限状态删除，股票列表模糊匹配，实时用户订单查询，文件导入，文件导出，单用户登录。使用sqlalchemy orm建立数据库对象关系映射，登录策略使用jwt生成token保存到redis中。算法部分整合TWAP\_PoV算法，连接kafka进行交易下单，资金查询，订单查询。

使用Django-admin开发交易系统后台管理平台，交易员账户管理，用户权限管理，角色管理，资金账号管理，超级管理员用户管理，配置管理。

## 2.主要任务

（1）交易员登录，

（2）token登录，

（3）资金查询，

（4）持仓查询，

（5）逐笔查询，

（6）撤单，

（7）新增策略，

（8）更新策略，

（9）更新策略状态，

（10）删除策略，

（11）查询策略列表，

（12）添加组合名称，

（13）查询组合列表，

（14）获取空行id，

（15）更新组合，

（16）更新组合状态，

（17）删除组合，

（18）下单，

（19）退出，

（20）用户列表，

（21）用户账号列表，

(22)用户拥有账号查询，

(23)用户权限增加，

(24)用户权限更新，

（25）用户权限状态更新，

（26）用户权限状态删除，

（27）股票列表模糊匹配，

（28）实时用户订单查询，

（29）文件导入，

（30）文件导出，

（31）单用户登录

（32）交易员账户管理，

（33）用户权限管理，

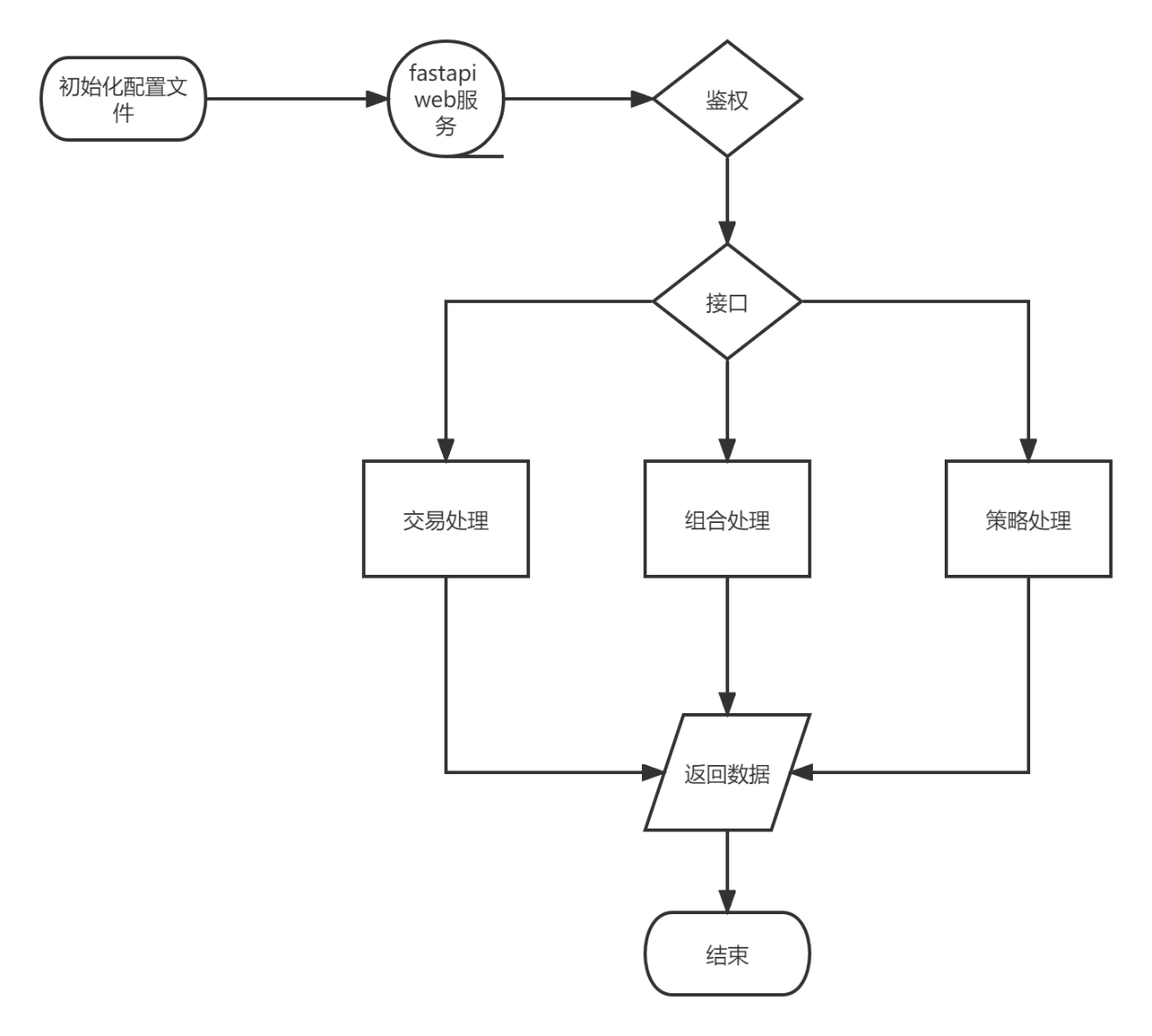
（34）角色管理，

（35）资金账号管理，

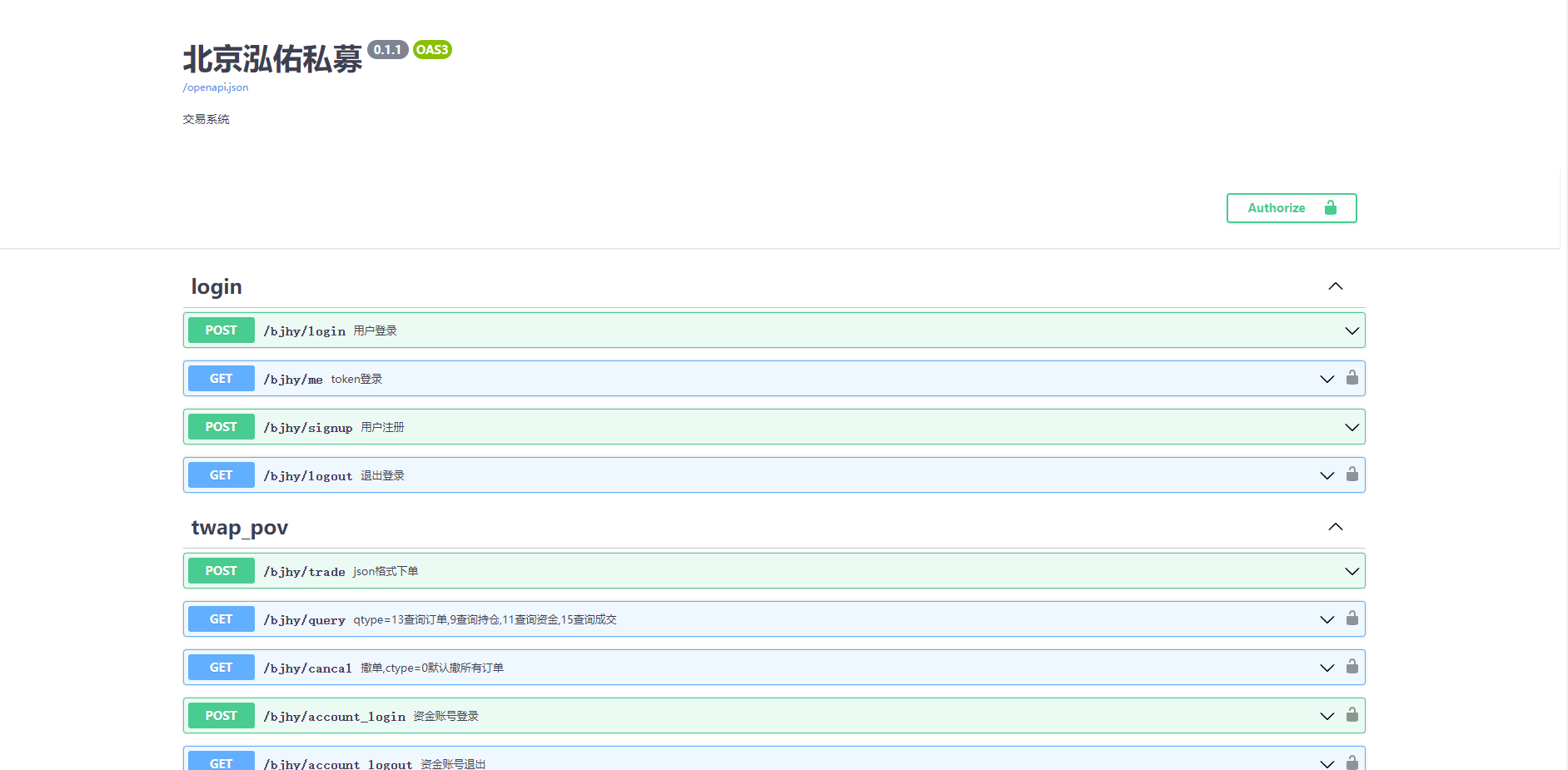
（36）超级管理员用户管理，

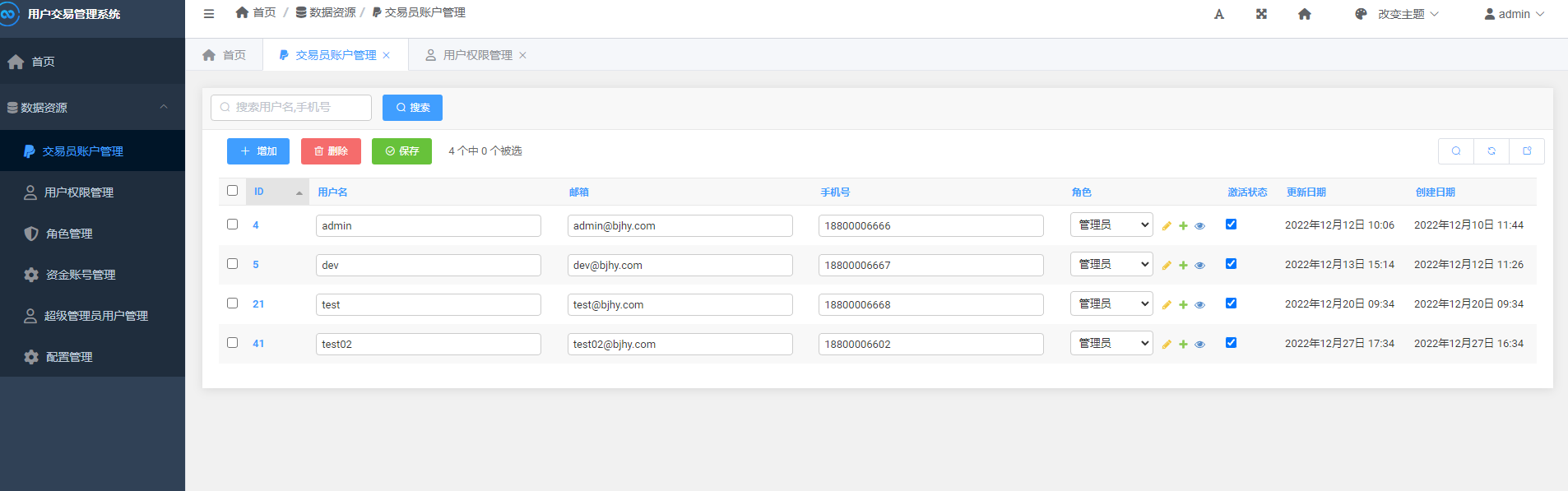
（37）配置管理

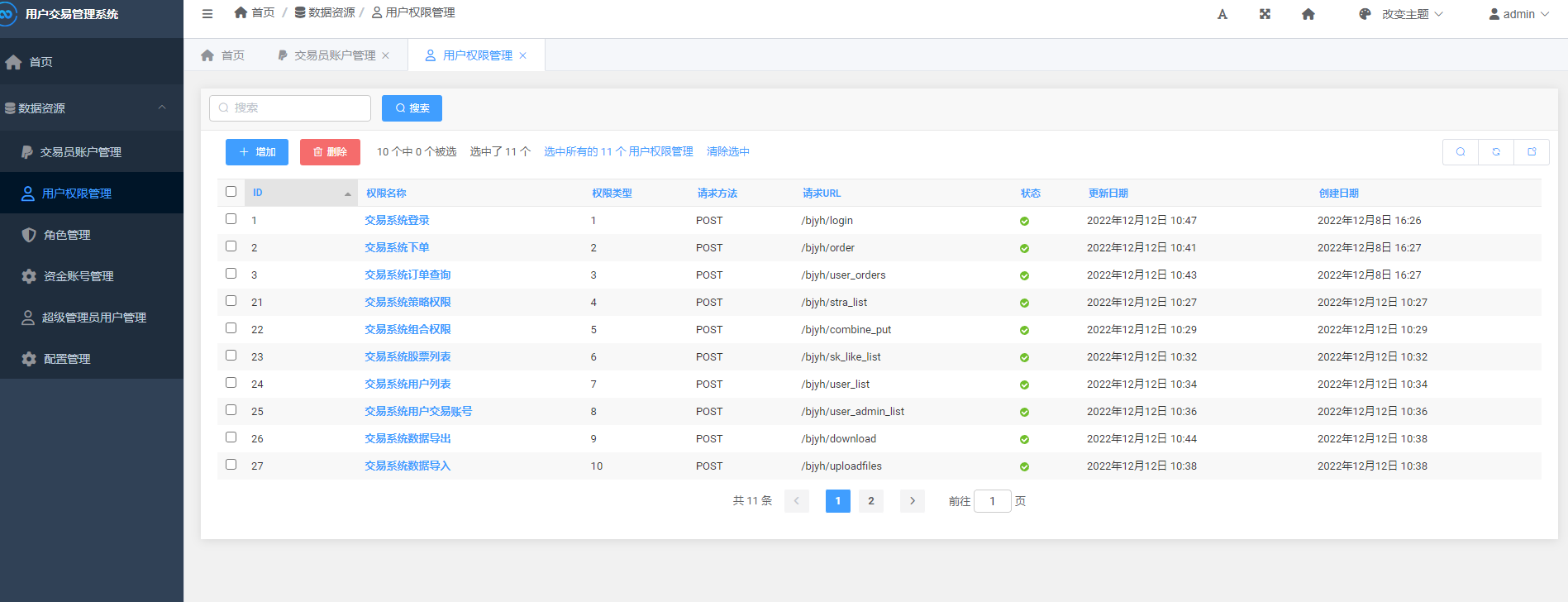
## 3.数据结构图

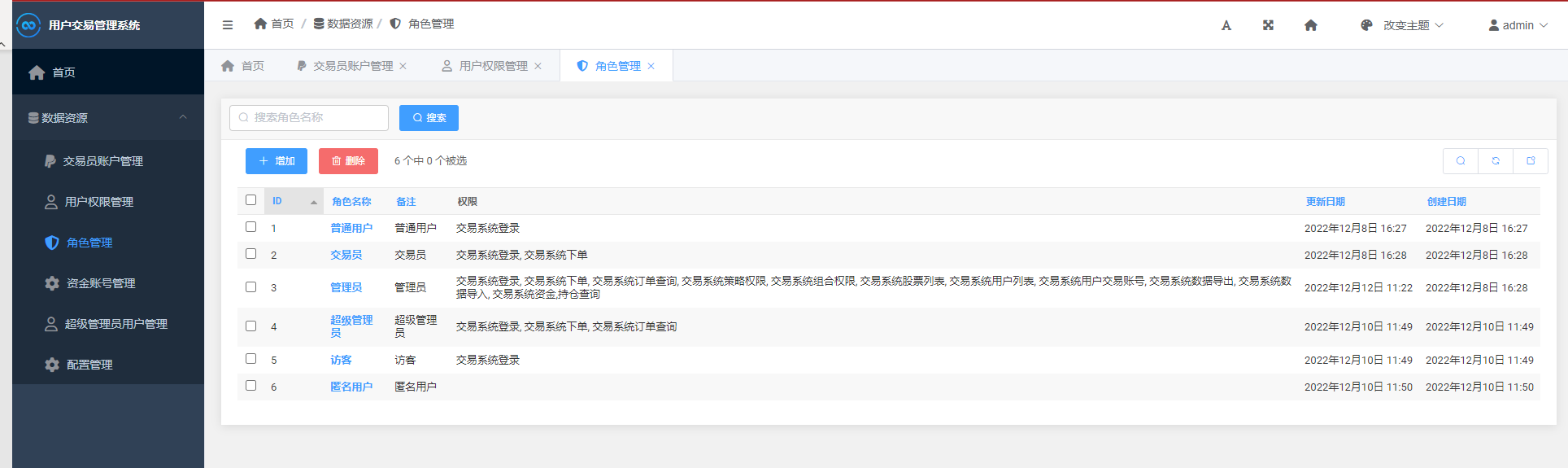


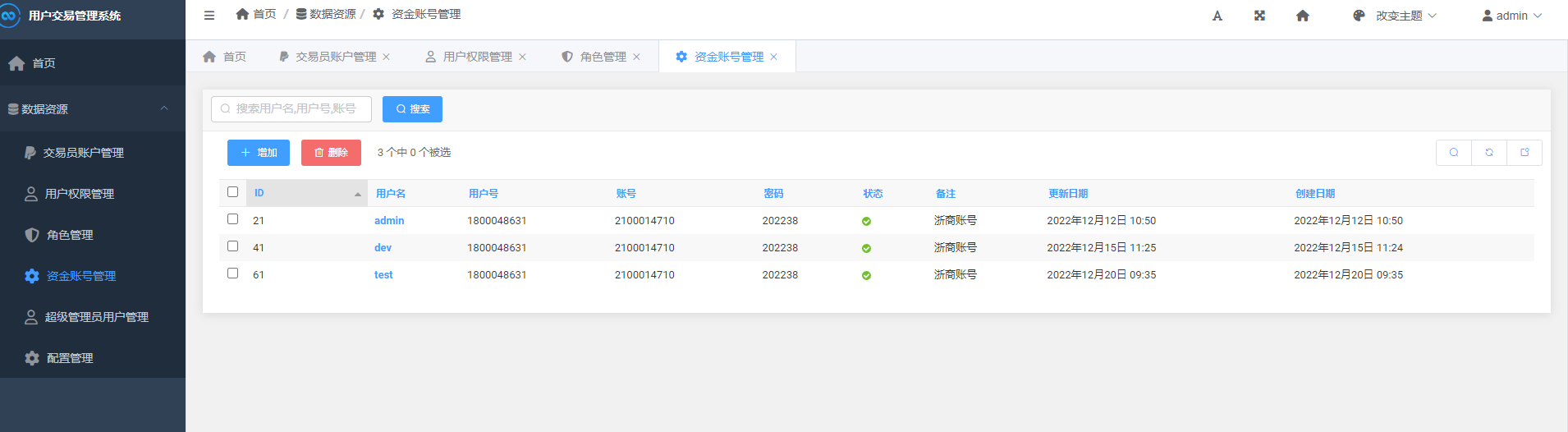
## 4.效果展示

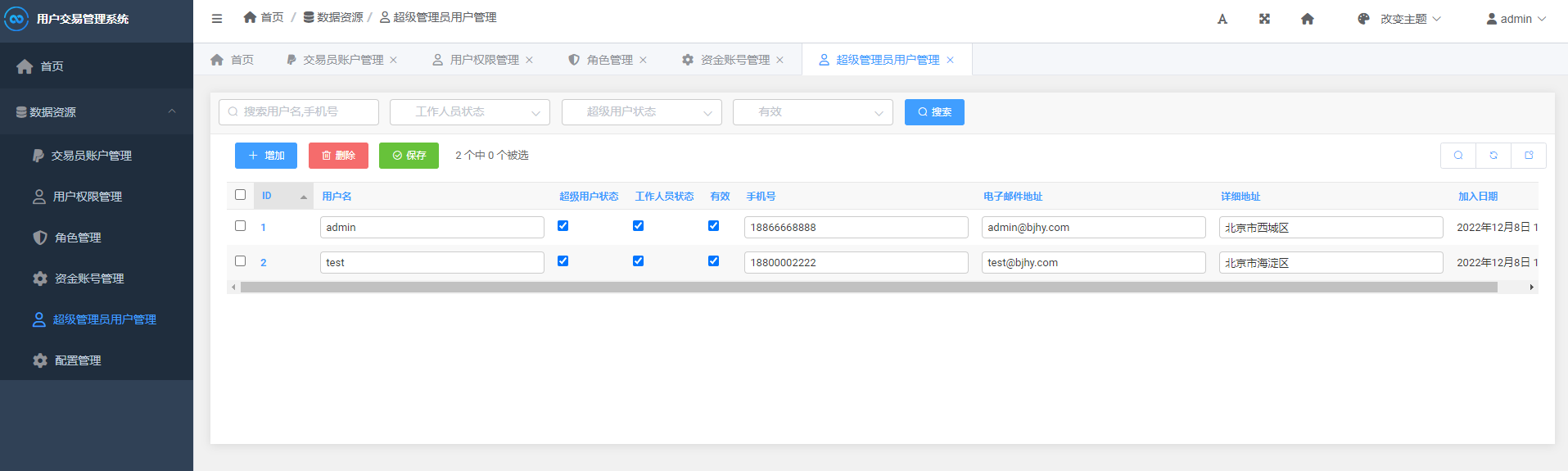


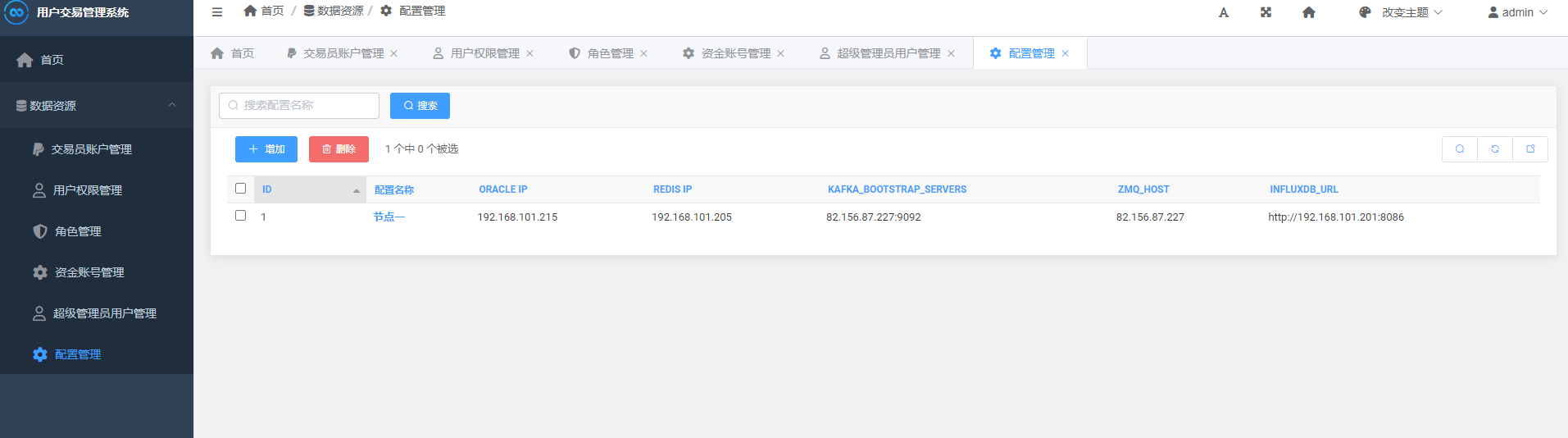












## 5.中间件

pandas，redis，fastapi，kafka，nginx，django

## 6.源码

<http://192.168.101.210:7990/scm/~wangyanyan/async_kafka_rep.git>

http://192.168.101.210:7990/scm/~wangyanyan/django\_admin.git

## 7.部署

cd到[src](http://192.168.101.210:7990/users/wangyanyan/repos/async_kafka_rep/browse/src)/[fast\_api](http://192.168.101.210:7990/users/wangyanyan/repos/async_kafka_rep/browse/src/fast_api)/[fastapi-react-dev](http://192.168.101.210:7990/users/wangyanyan/repos/async_kafka_rep/browse/src/fast_api/fastapi-react-dev)/[app](http://192.168.101.210:7990/users/wangyanyan/repos/async_kafka_rep/browse/src/fast_api/fastapi-react-dev/app)/**app**/

启动程序：./start.sh

停止程序：./stop.sh

django\_admin：

启动程序：./start.sh

停止程序：./stop.sh

## 8.依赖库

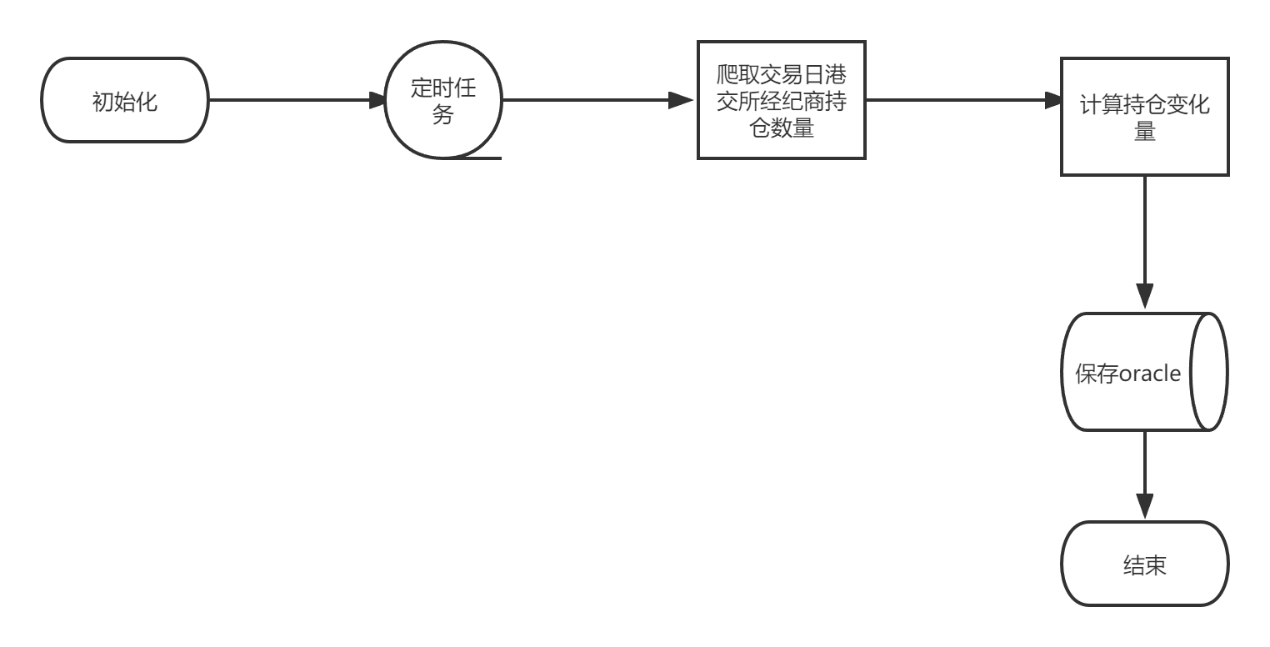


# 十三、东方财富解禁个股明细爬取程序

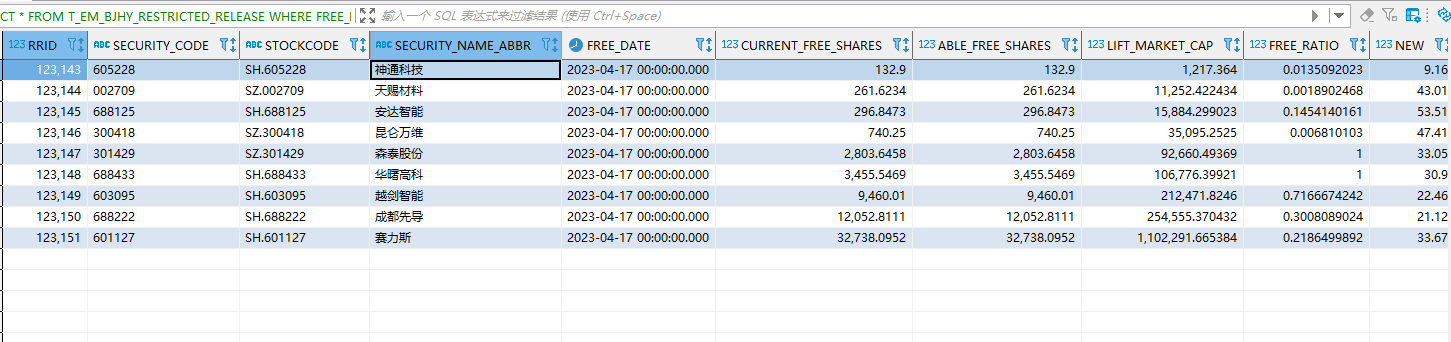
## 1.功能描述

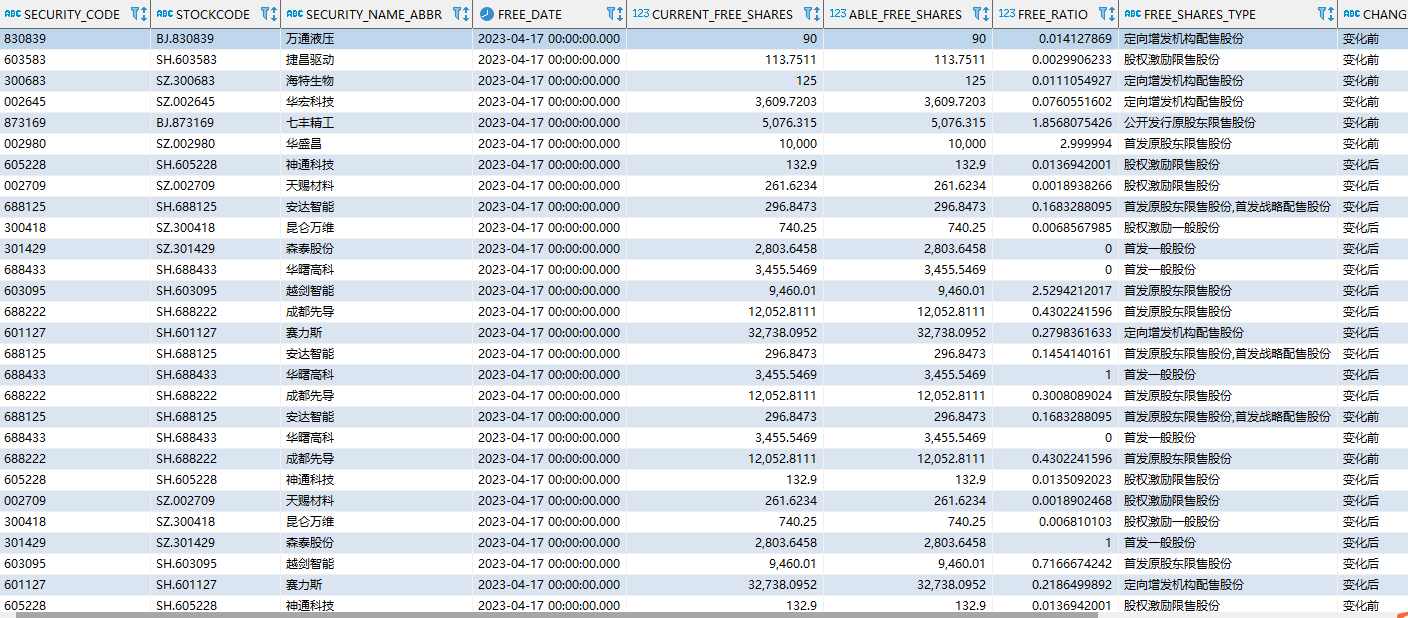
爬取东方财富解禁个股明细，包括股票代码，股票简称，解禁时间，限售股类型，解禁数量，实际解禁数量，实际解禁市值，占比，收盘价，涨跌等，定时任务爬取交易日前后30天的数，以及解禁信息的变化数据入库，爬取历史数据，前后十年的数据入库，每日8,12,16点抓取数据。

## 2.数据结构图



## 3.效果展示





## 4.中间件

pandas，aiohttp，oracle

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 eastmoney

启动程序：./start.sh

停止程序：./stop.sh

## 7.依赖库

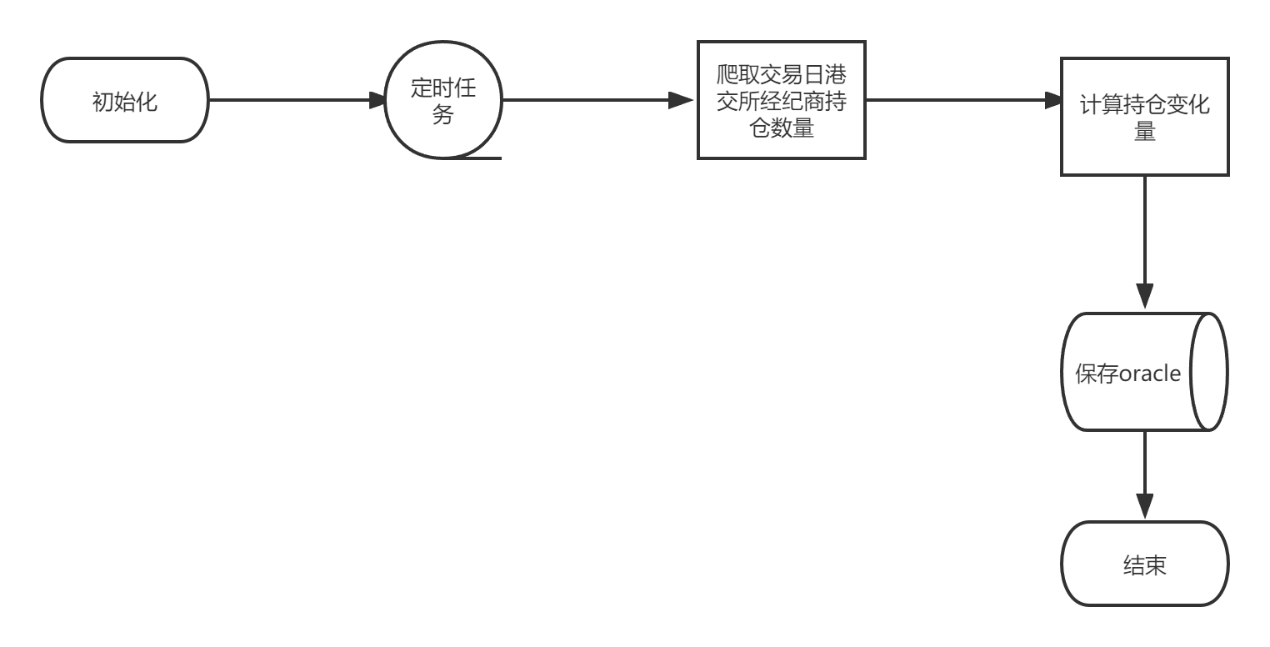


# 十四、券源信息入库和解禁日可用券源信息生成

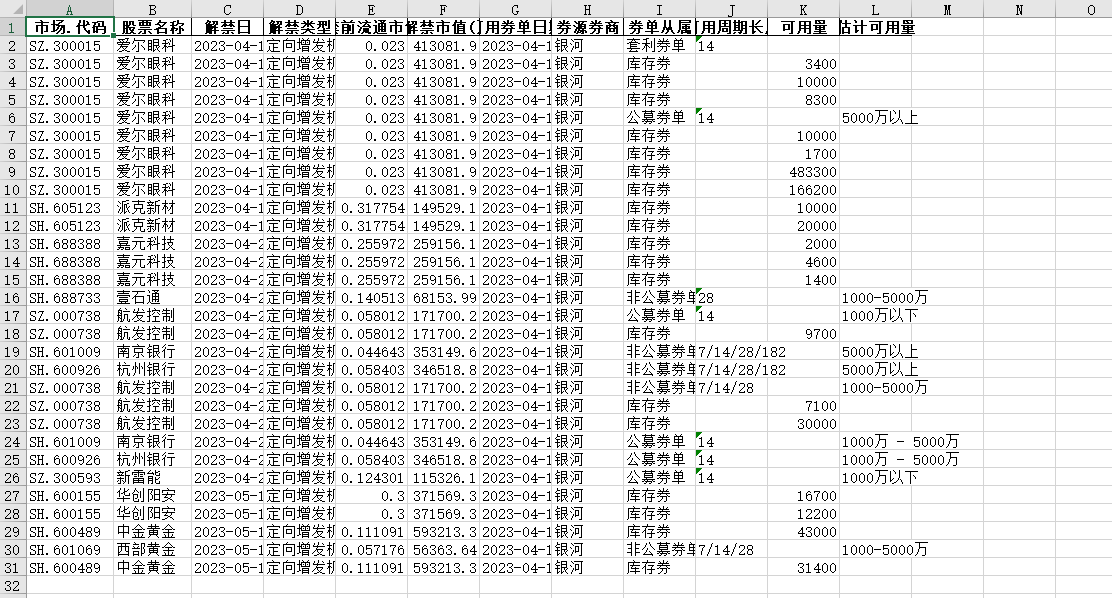
## 1.功能描述

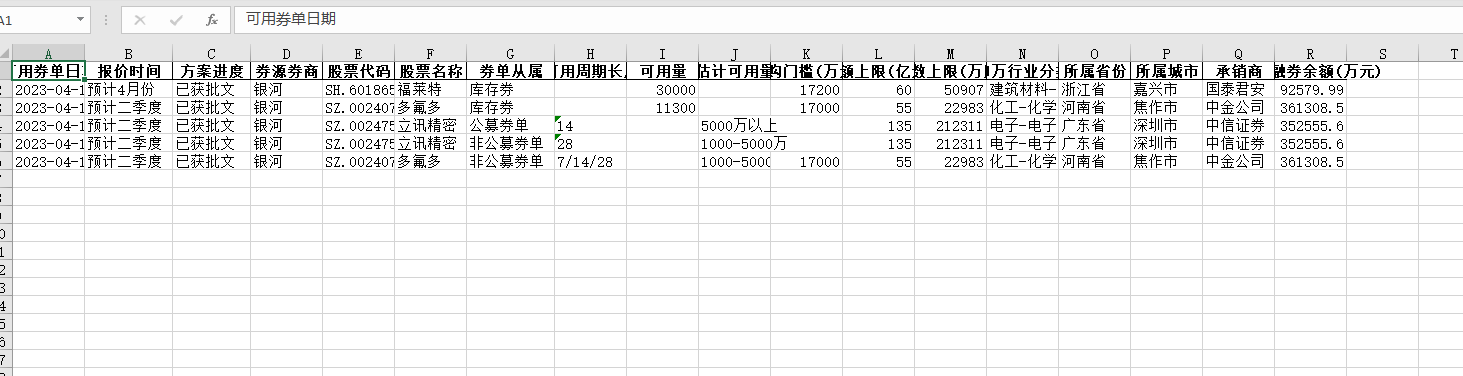
手动从excel文件中将，中信，中信建投，中金，广发，银河等券商券源入库，包括T+0券源模板，T+1券源模板，库存券源(审批)模板，库存券源模板，实时锁券券单，券源供给信息，专项券源名单，套利券单，T0券单，日间券单，定增项目信息表等，股票代码，股票简称，类型，券源日期，券息，可提供期限，数量，到期日期，可用天数，参考费率，成本价，行业，所属指数，期望概率，预计市值，交易市场，可用券时间。导出数据包含，市场，股票名称，解禁日，解禁类型，占比，实际解禁市值，可用券单日期，券源券商，券单从属，可用周期，可用量，估计可用量，每周定增项券源信息导出。

## 2.数据结构图



## 3.效果展示





## 4.中间件

pandas，oracle，sqlalchemy

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 stock\_margin\_detail

启动程序：python parse\_excel\_margin\_db.py

## 7.依赖库

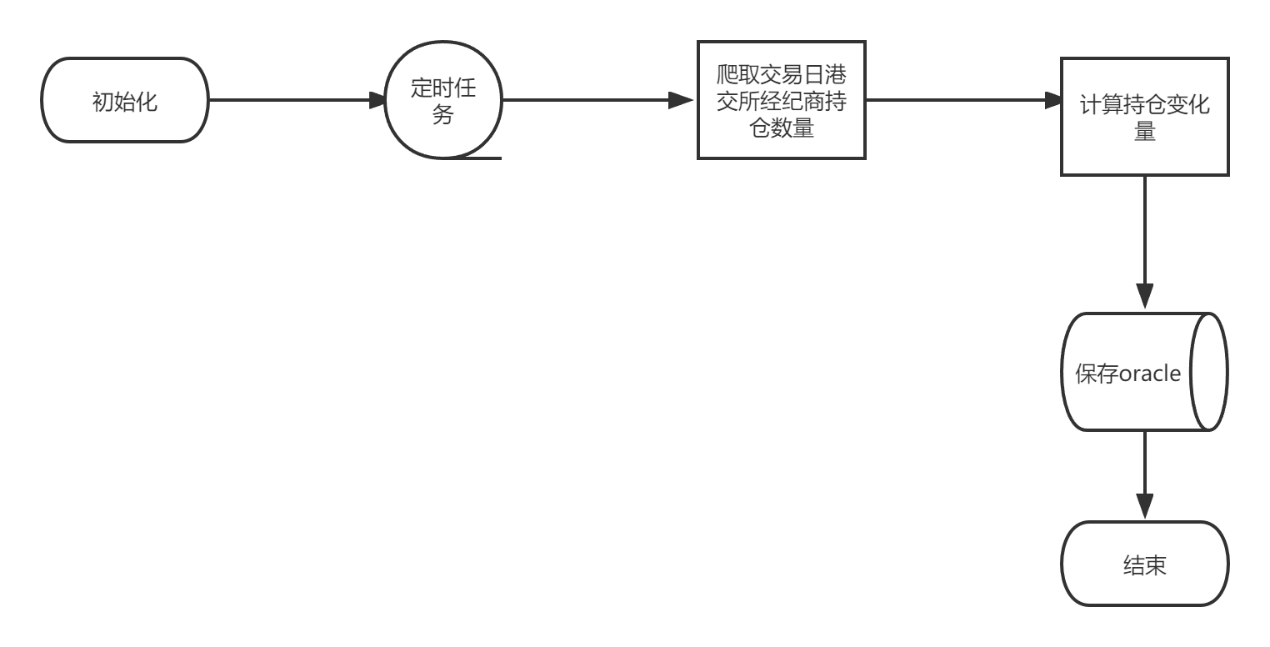


# 十五、通达信每日分钟线入库

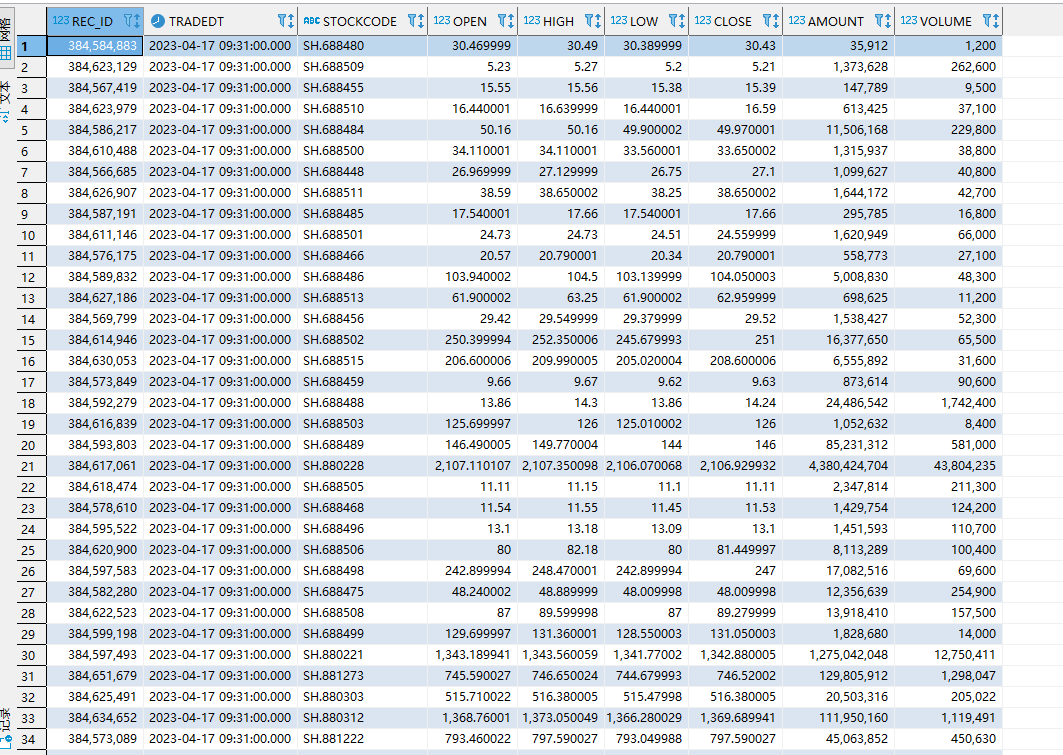
## 1.功能描述

交易日每天下午15.45手动下载通达信盘后1分钟线数据，然后批量解析下载的lc1文件，获取，代码，"trade\_time", "open", "high", "low", "close", "vol", "amount"等信息，使用pandas批量将数据入库。

## 2.数据结构图



## 3.效果展示



## 4.中间件

pandas，oracle，sqlalchemy, Reader

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 tdx\_1mins\_line/dist

启动程序：File[1minkline\_to\_oracle\_v2.exe](http://192.168.101.210:7990/users/wangyanyan/repos/ths-hk-data/browse/tdx_1mins_line/dist/1minkline_to_oracle_v2.exe)

或者 python [1minkline\_to\_oracle\_v2.py](http://192.168.101.210:7990/users/wangyanyan/repos/ths-hk-data/browse/tdx_1mins_line/1minkline_to_oracle_v2.py)

## 7.依赖库

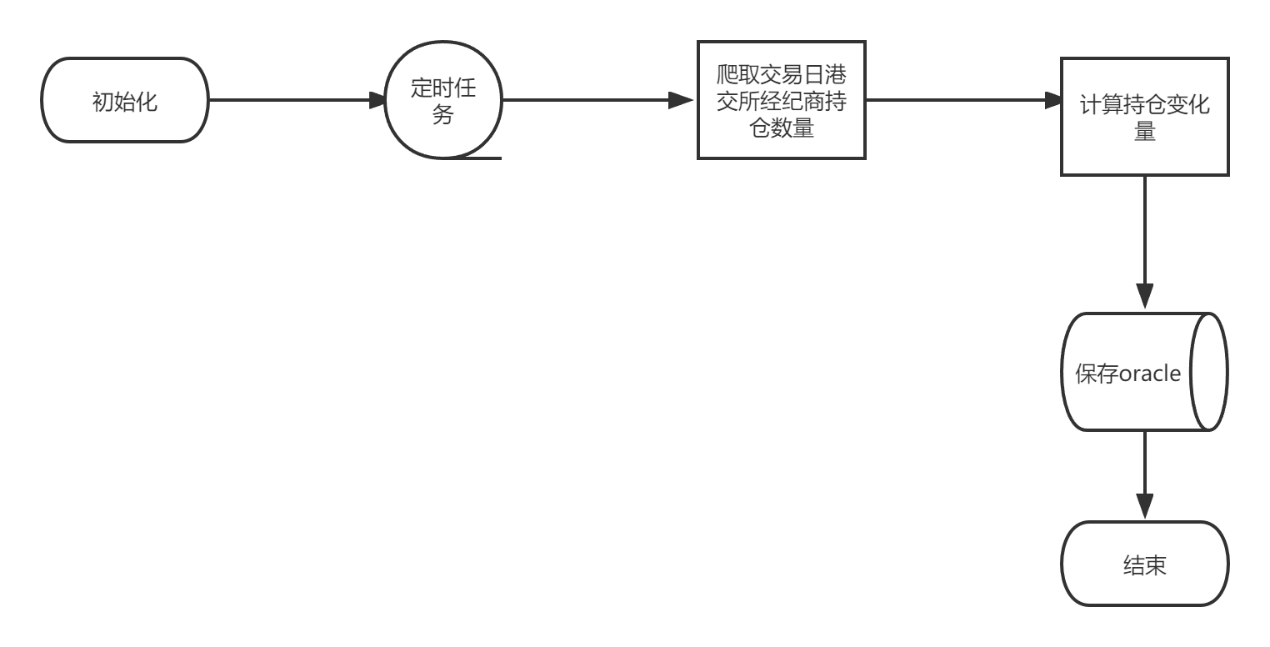


# 十六、通达信每日分钟线生成csv文件上传

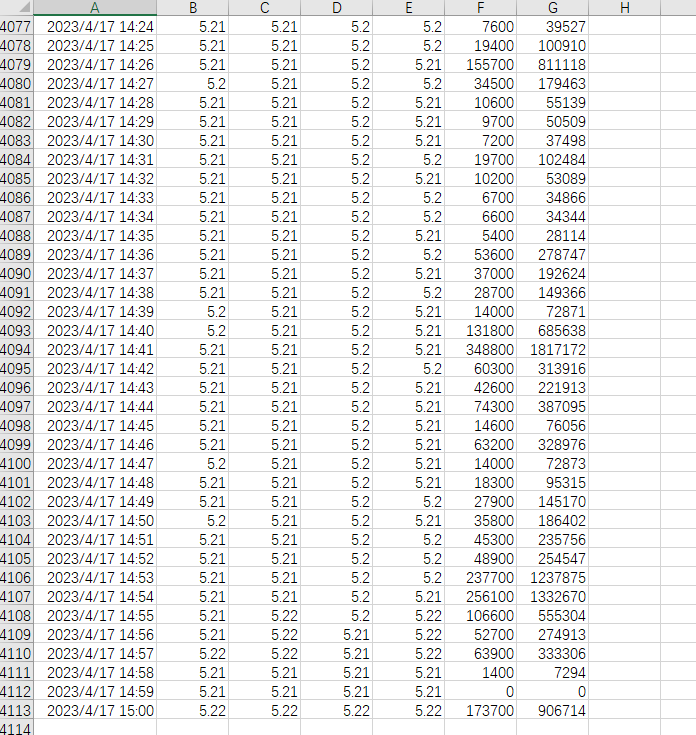
## 1.功能描述

解析每日通达信1分钟线数据，市场，股票代码，"trade\_time", "open", "high", "low", "close", "vol", "amount"。使用pandas导出成csv文件后sftp上传到共享文件中。文件校验，只上传增量文件。

## 2.数据结构图



## 3.效果展示



## 4.中间件

pandas，oracle，sqlalchemy, Reader

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 tdx\_1mins\_line/dist

启动程序：File[1minkline\_to\_sftp\_csv.exe](http://192.168.101.210:7990/users/wangyanyan/repos/ths-hk-data/browse/tdx_1mins_line/dist/1minkline_to_sftp_csv.exe)

或者 python [1minkline\_to\_sftp\_csv.py](http://192.168.101.210:7990/users/wangyanyan/repos/ths-hk-data/browse/tdx_1mins_line/1minkline_to_sftp_csv.py)

## 7.依赖库

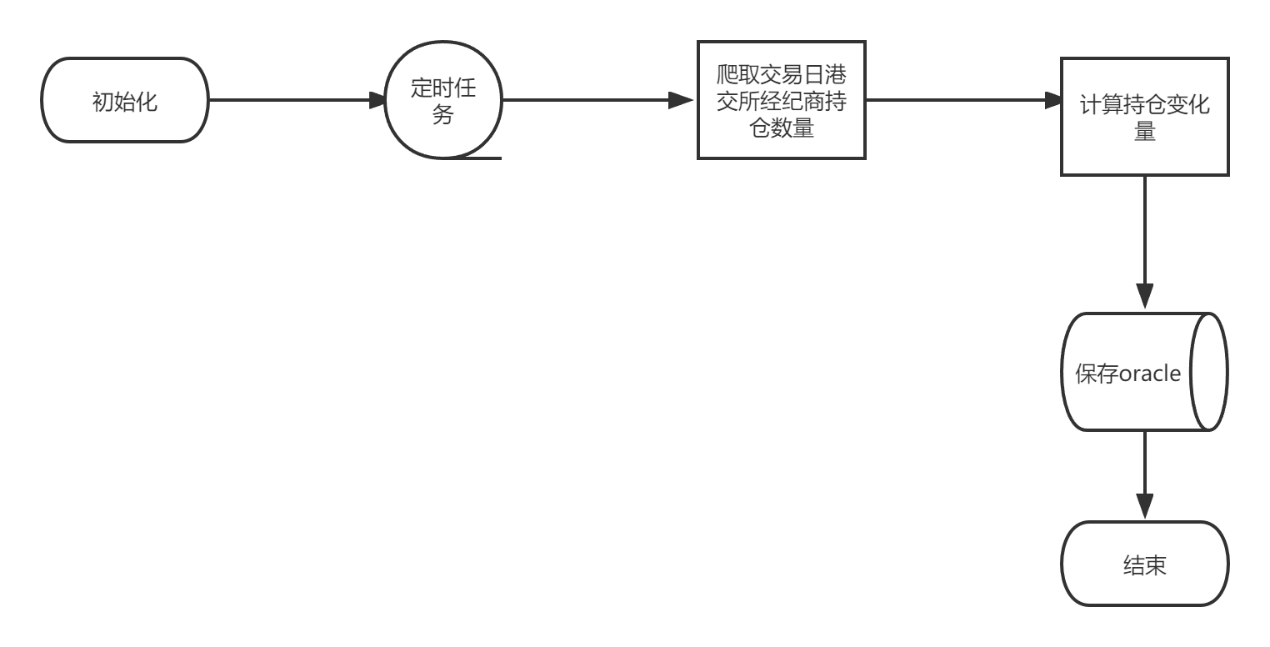


# 十七、股票历史分钟线行转列处理生成parquet文件上传

## 1.功能描述

获取2017-2023年历史1分钟线，将当日收盘价并进行除权，除权后，计算"open", "high", "close", "low"除以除权后的收盘价，vol量除以当天量的总和，amount除以流通市值，然后将当天数据翻转，行转列成一行数据，历史数据以此类推，float类型数据转float32后导出成parquet文件scp上传。使用redis缓存历史交易日以及交易日的复权因子和流通市值，使用多进程快速计算生成文件。

## 2.数据结构图



## 3.效果展示

0 2017-01-03

1 2017-01-04

2 2017-01-05

3 2017-01-06

4 2017-01-09

...

1523 2023-04-11

1524 2023-04-12

1525 2023-04-13

1526 2023-04-14

1527 2023-04-17

Name: tradedate, Length: 1528, dtype: datetime64[ns]

0 0.623245

1 0.999386

2 0.998163

3 1.000000

4 1.003708

...

1523 1.002774

1524 1.002786

1525 1.001387

1526 1.004138

1527 0.998625

Name: open/lclose-000, Length: 1528, dtype: float32

## 4.中间件

pandas，oracle，sqlalchemy, redis

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 sftp\_1mins\_k\_line/oracle\_to\_redis\_data

启动程序：python [ltdxdy\_dailykline\_parquet.py](http://192.168.101.210:7990/users/wangyanyan/repos/ths-hk-data/browse/sftp_1mins_k_line/oracle_to_redis_data/ltdxdy_dailykline_parquet.py)

## 7.依赖库

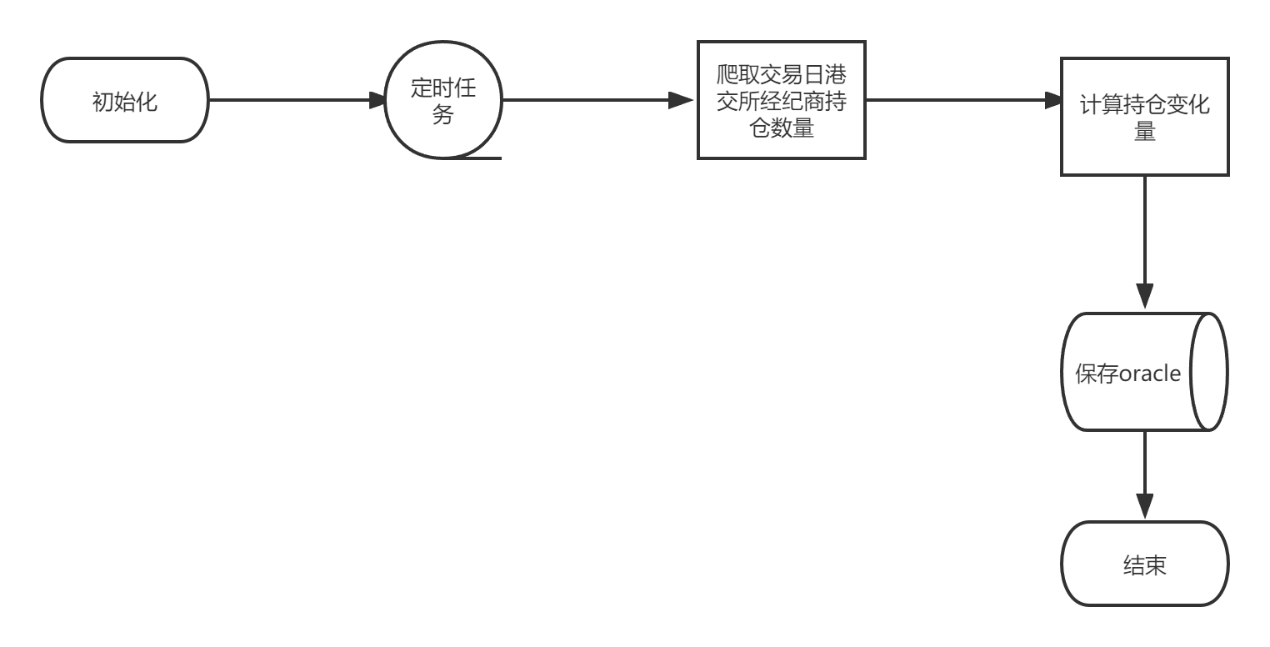


# 十八、金融界新闻内容获取

## 1.功能描述

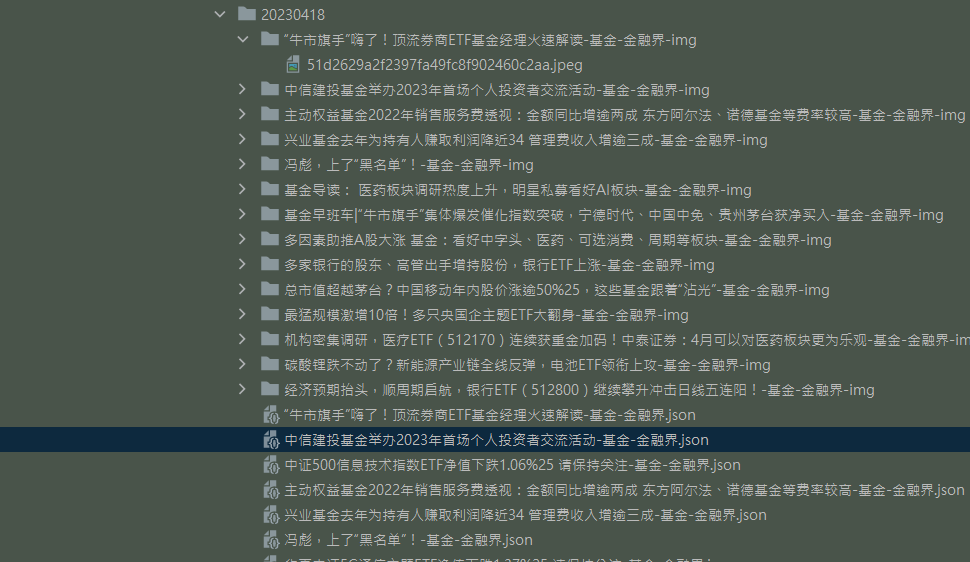
爬取金融界历史(2010-2023)以及每日，基金，外汇，港股，美股，财经，银行新闻，包含url，news\_source，title，article\_source，news\_date\_time，keywords，content，recommend，imgs，article\_auth。如果新闻中有图片，则下载图片并保存在同一目录，并在新闻内容中保留图片位置上填充本地图片路径。启动定时任务每日8,12,15,21点爬取数据。

## 2.数据结构图



## 3.效果展示





## 4.中间件

aiohttp，BeautifulSoup，lxml, apscheduler，json

## 5.源码

<http://192.168.101.210:7990/scm/~wangyanyan/ths-hk-data.git>

## 6.部署

cd到 jrj\_news

启动程序：./start.sh

停止程序：./stop.sh

## 7.依赖库

