# 系统简介

本人是一名曾就职于证券公司的IT人员，日常工作主要为数据处理、项目管理与程序开发。因工作需要，平时经常与SQL与python语言打交道，一个偶然的机会得知python不仅可以进行数据的加工与分析，还可以进行web应用的开发，遂对此产生了浓厚的兴趣，于是抱着学习以及对平时工作内容进行回顾的目的利用业余时间开发了finance-calculator系统。finance-calculator系统是一个简易的web应用，用于对股票持仓进行存储、加工、统计与风险指标的计量，通过采集-计量-制表等三个基本的处理流程实现数据的可视化。技术选型方面，系统前端页面通过react框架进行搭建，后台程序通过python的flask框架进行编写，数据库采用的是开源的PostgreSQL数据库。系统仍在不断开发迭代中，目前实现的功能有：数据的采集与增删改、数据的统计与VaR指标的计量，下面我将就这几块内容进行详细介绍。  
**数据的采集与增删改**

**采集**

系统目前支持**自动采集**与**手工导入**两种数据获取模式，分别对应“ETL任务”和“数据导入”功能。其中，**自动采集**通过python的BackgroundScheduler进行实现，可以在每天的固定时间从固定文件夹中读取固定名称的excel文件，并根据日期条件进行判断，自动采集程序运行日期前一天的数据，并且在采集之前会先对数据库中现有的对应日期的数据进行删除处理，确保对应日期数据的唯一性。

**手工导入**可以看做是自动采集的补充，当有大量的历史数据需要导入或者不方便调用自动采集任务的时候，可以用该功能进行数据导入。数据导入之前会自动判断是否和现有的数据的日期有冲突，只有在日期没有冲突的情况下数据才会导入成功。

在数据采集和导入的时候，系统会自动调用“证券宝”API自动获取所在日期持仓证券的收盘价与所属行业，算出证券的市值并给证券打上所属行业的标签。

**增删改**

“数据编辑”功能用于处理导入后数据的增删改。当自动采集的数据或者手工导入的数据有误的时候，可以通过该功能进行在该功能下，可以通过该功能进行补救。在该功能下，可以根据数据的各种标签进行筛选，并且对选中的数据进行修改和删除处理，以纠正数据采集或导入时候的错误。

**数据的统计**

系统目前支持明细数据统计与汇总数据统计两个模式，分别对应“持仓展示”与“持仓分析”功能。其中，持仓展示是对明细数据的统计，可以根据股票持仓数据对应的日期、部门、投资组合、所属的行业与证券代码进行筛选。在“持仓分析”功能中，可以选择根据“公司”、“部门”、“投资组合”等三个不同的层级进行统计，这里主要通过python的pandas对数据进行处理，在分别加工出查询日期当天的汇总后前十大股票的信息、查询日期同一年的每一天市值汇总值的列表数据、查询日期当天汇总的市值的集中度、查询日期当天以行业汇总的市值的集中度后，通过json形式传给前端，前端的react通过react-chartjs-2组件根据数据绘制出对应的折线图与饼图。

**VaR指标的计量**

该部分内容对应系统中“VaR计量”功能，系统目前支持通过历史模拟法、参数法与蒙特卡洛模拟法三种方法进行VaR指标的计量，并可以自由选取置信度、预测天数、历史选择日期的长度，在蒙特卡洛模拟方法下还可以选择模拟路径的条数。在计量完成后，将以可下钻的二维表格与折线图、频率分布直方图的形式对结果进行展示。该功能实现原理如下：  
**计量**

1. 后台根据前端传来的日期参数从数据库中提取对应日期的持仓数据；
2. 计算出单只股票的市值在投资组合、部门、公司中所占的权重；
3. 调用证券宝接口，获取每一只股票在所选历史日期区间的每一天的收盘价，计算出每一只股票每一天的对数收益率；
4. 与之前的权重数据结合，计算投资组合、部门、公司每一天的对数收益率；
5. 根据所选择的方法进行计量  
    （1）历史法：直接用对应层级的收益率根据分位数进行划分，得出对应置信度下的VaR  
    （2）参数法：求出对应层级收益率的均值与标准差，并根据固定公式算出对应置信度下的VaR  
    （3）蒙特卡洛模拟法：求出对应层级收益率的均值与标准差，根据输入的路径数进行随机模拟，对模拟后的收益结果根据分位数进行划分，得出对应置信度下的VaR

**下钻表格**

将不同层级的var通过关联条件组合成一张包涵公司、部门、投资组合、个券等各个层级的大宽表，通过json形式传给前端，前端react通过递归的方法实现数据的下钻展示

**图表**

将必要的参数通过json形式传给前端，前端react用react-chartjs-2与chart.js进行频率分布、正态分布并根据此绘制出对应的图表

本系统为业余时间开发，受本人技术、业务与时间上的局限的影响，仍然有很大的优化空间，比如VaR的算法可以考虑相关性问题，证券品种方面可以扩充至债券及衍生品，自动采集的数据可以支持更多的数据库，前端页面可以做的更漂亮，后台的架构也应更多的考虑效率问题。将此系统开源出来主要目的是为了抛砖引玉，希望和大家进行交流沟通，后续有时间我将持续维护该系统。