## 本阶段成果

上个阶段，我完成了主要的算法算法研究，包括PCA分析提取债券主成分、Monte Carlo模拟法计算在险价值和Nelson-Siegel模型拟合收益率曲线。关于软件开发的基础算法部分，在上一个阶段已经全部完成，本阶段的主要任务就是基于上一阶段的研究成果，完成软件的开发。

根据任务书要求，所要设计的软件主要包括以下功能：

1. 收益率曲线数据
2. 中国国债历年信息
3. 美国国债历年信息
4. SHIBOR历年信息
5. LIBOR历年信息
6. 最新市场债券信息
7. PCA分析
8. VaR在险价值

（1）中债VaR数据展示

（2）VaR计算器

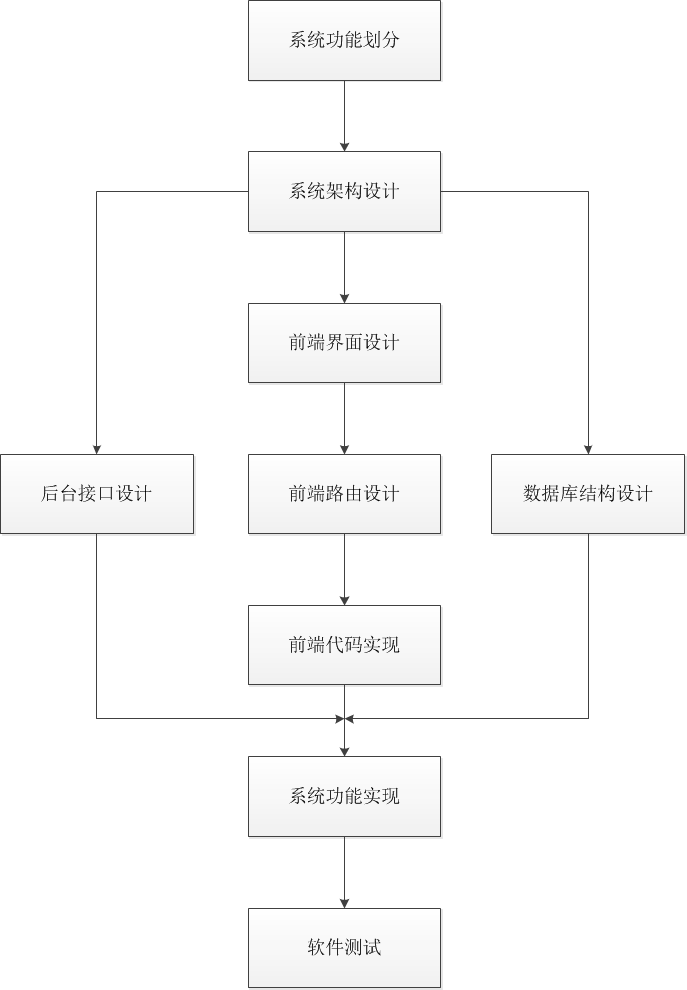
1. 中债估值数据展示
2. 数据查询
3. 自定义收益率曲线拟合

围绕这些功能，首先进行架构设计，主要架构设计如下：



其中前端界面负责与用户进行交互、Node框架Express搭建的HTTP服务器负责控制路由、派发请求，Python脚本负责对数据进行处理、MongoDB负责数据存储。

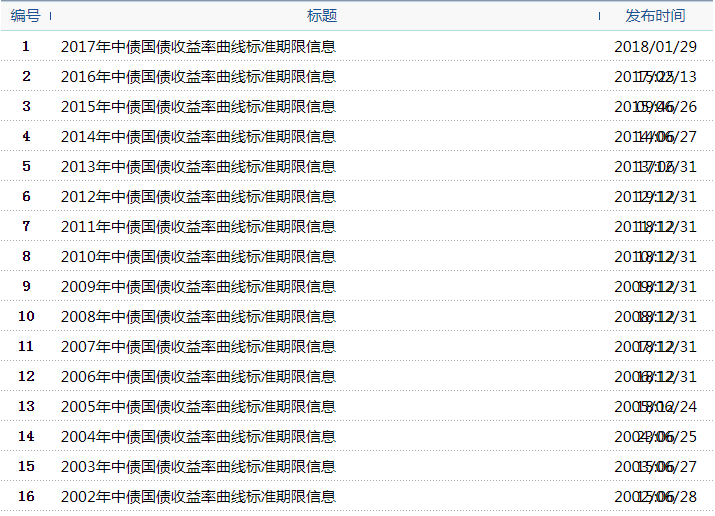
规定总体的开发流程如下：



由于业务逻辑主要在后端部分，所以先进性后端程序开发。

数据的获取依赖于Python爬虫，从爬取中国国债数据开始着手：

本次数据爬取的对象确定为中债网，中债网的中国国债日收益率以Excel文件的形式提供给用户下载：



通过查看其浏览器的控制台可以看到其HTML代码如图：



可以发现是一个ul标签，内嵌了一系列的li标签，形成了一个链接列表，其中的a标签指明了下载Excel数据文件的url和标题，所以只需要解析到这个a标签即可，然后根据对应的url下载服务器资源。

使用的模块工具是urllib.request和pyquery，其中urllib.request用于发起HTTP请求，而pyquery是一个类似jQuery的HTML DOM操作库，但是它是在服务器端对DOM进行渲染和操作，而非浏览器，可以用于对返回的HTML文件进行解析。由于pyquery是第三方模块，所以需要使用Python包管理器进行安装。整个程序只有两个函数，一个用于获取Excel下载的url，一个用于根据提供的url下载对应资源，并保存到本机指定位置。

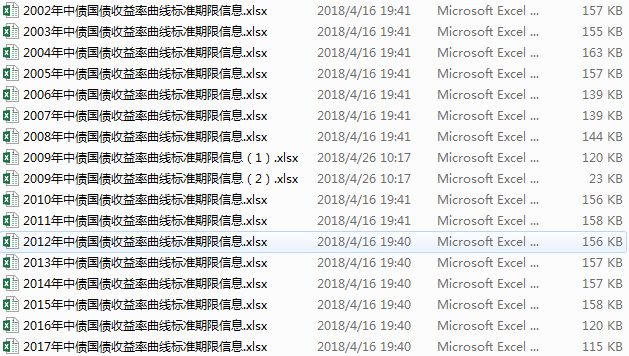
获取url的函数代码如图：



根据url下载对应Excel数据文件的函数代码如图：



依次执行以上两个函数，所有Excel文件会被下载到指定的路径如图：



以上就是爬取中国国债日收益率历史数据的方法，爬取其他数据的爬虫开发与此类似。

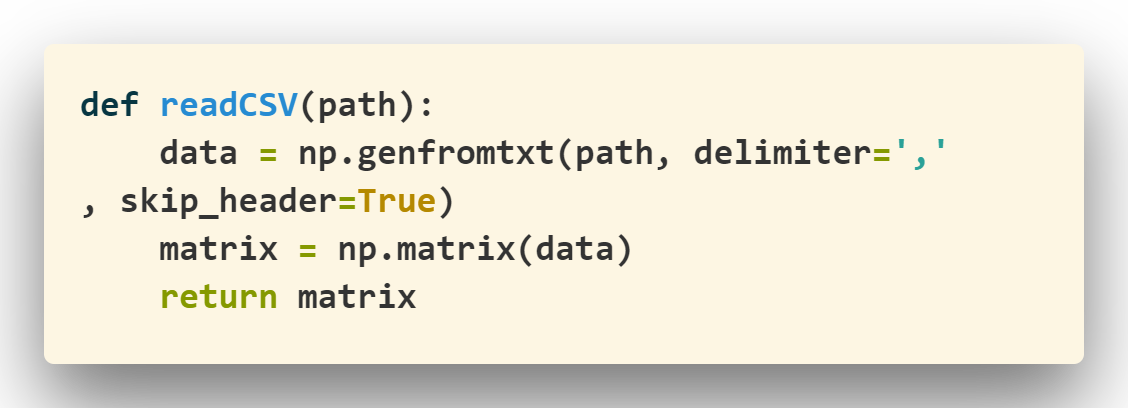
爬取到数据之后，需要对数据进行分析，这就需要开发Python脚本。

在上一阶段中已经研究了PCA算法的理论，并使用SPSS统计学软件进行了计算分析，接下来使用Python程序进行PCA主因子提取。Node程序调用Python进行PCA主因子提取时，需要将CSV数据的地址以及所要提取的主成分个数传入，Python使用传入的数据进行PCA主因子提取，提取出指定个数的主成分之后，将各个维度的贡献度、累计贡献度以及因子载荷矩阵数据存储到指定路径下的CSV文件中，Node程序通过读取CSV文件中的数据向用户返回分析结果数据。

该程序主要使用第三方模块numpy进行主成分分析，需要先使用pip进行安装，安装之后在文件开头通过import引入。为了防止生成的CSV文件重名，此处使用uuid模块生成全球独一无二的标志，最终作为文件名保存。

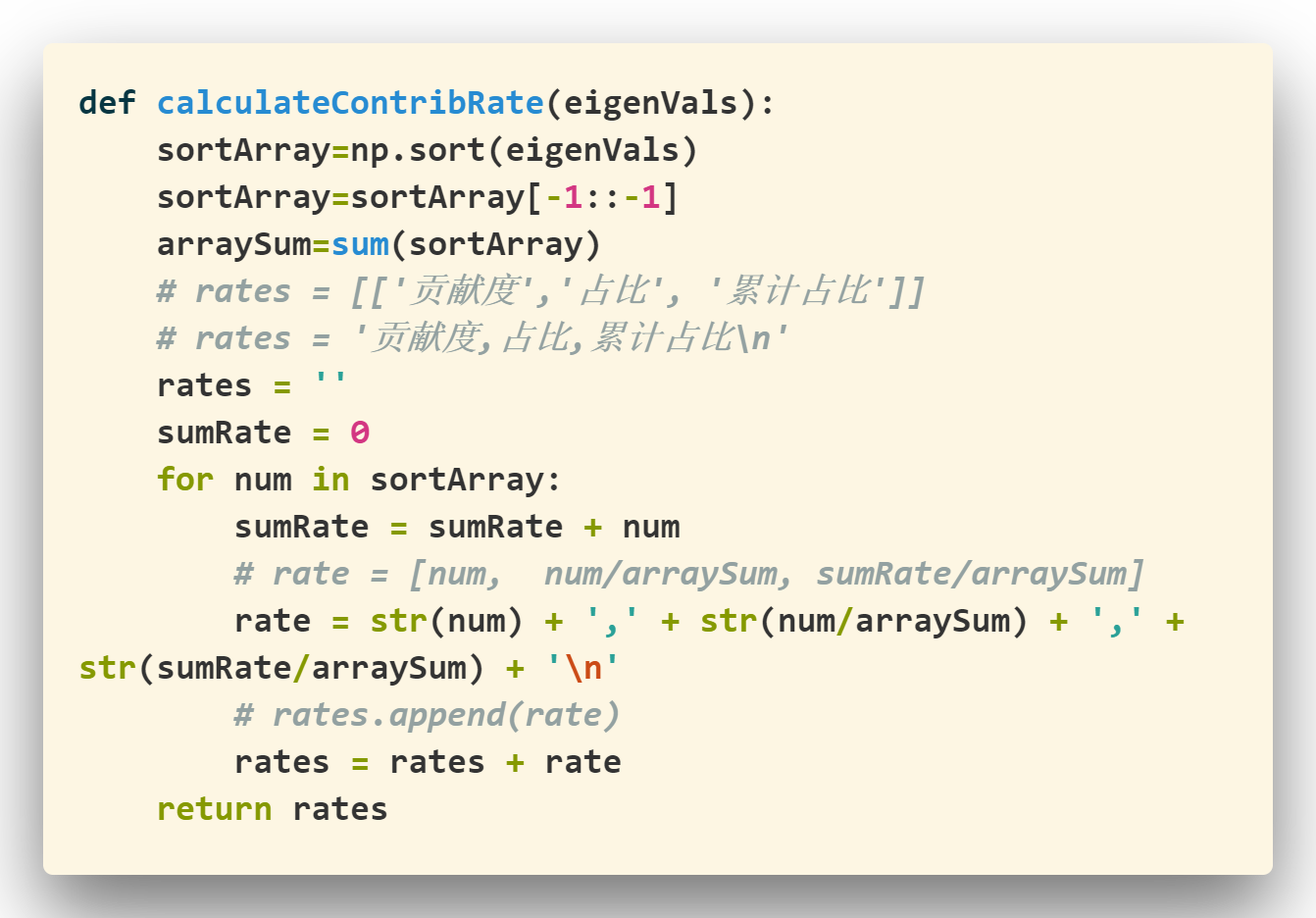
这个程序分为四个主要函数：readCSV函数负责读取指定路径的CSV文件中的数据，返回一个numpy matrix矩阵结构变量；calculateContribRate函数负责计算个维度累计贡献度；D2Matrix2CSV函数负责将numpy matrix变量写入CSV文件；最重要的PCA方法实现PCA分析，传入多维数据，返回特征向量和因子载荷矩阵。

readCSV函数代码如图：

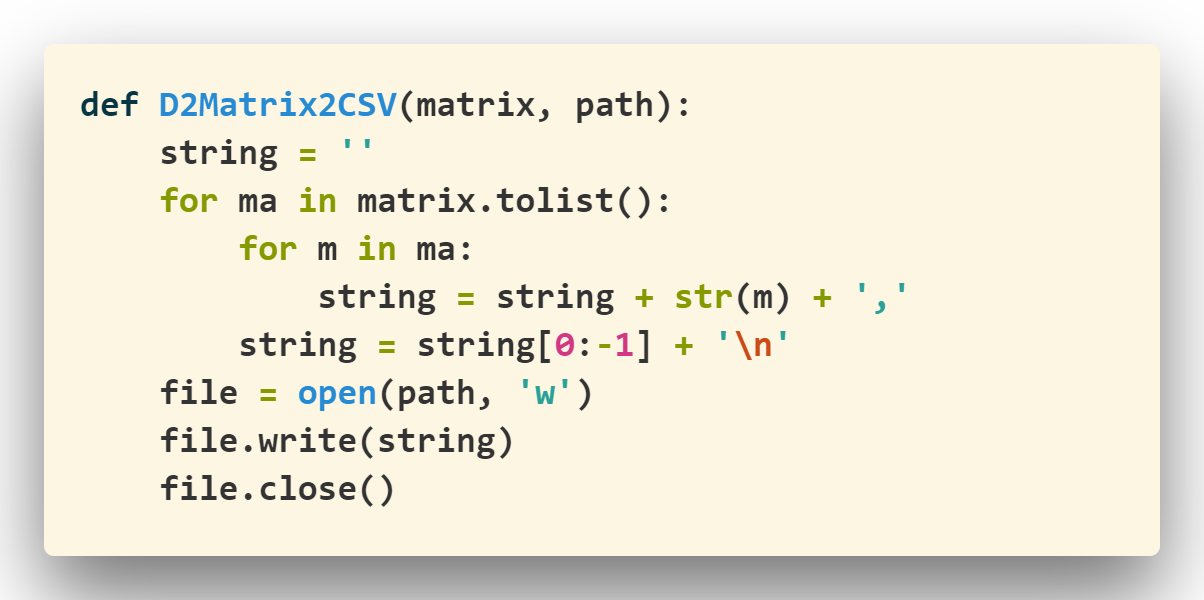


此处使用到的是numpy模块的genfromtxt方法读取CSV文件中的数据。

calculateContribRate函数代码如图：



D2Matrix2CSV函数代码如图：



这里主要是使用matrix矩阵转化为list变量，然后通过for循环遍历其中的元素，构造一个符合CSV格式的字符串，最终保存到指定路径的文件中即可。

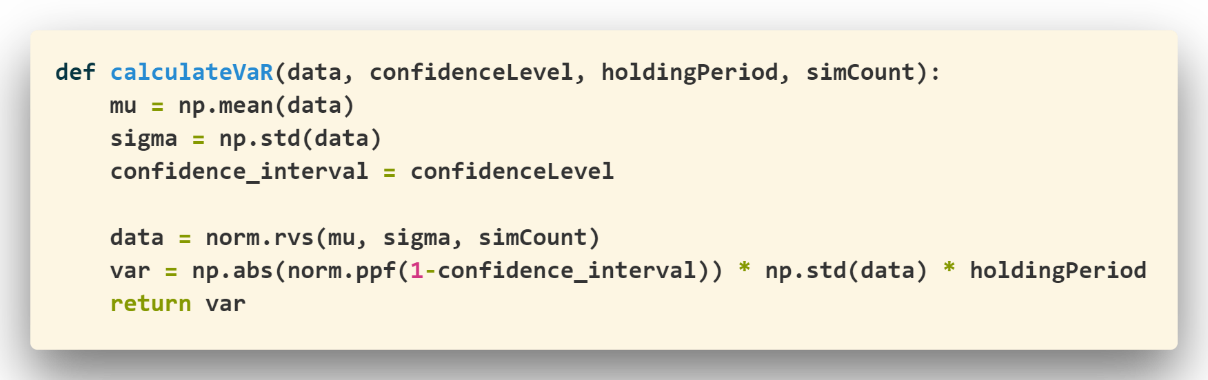
进行PCA分析的Python脚本代码如下：



曲线拟合代码如下：



计算在线价值的代码如下：



Node程序通过Express框架搭建HTTP服务器，方便快捷。接下来进行后台API开发;

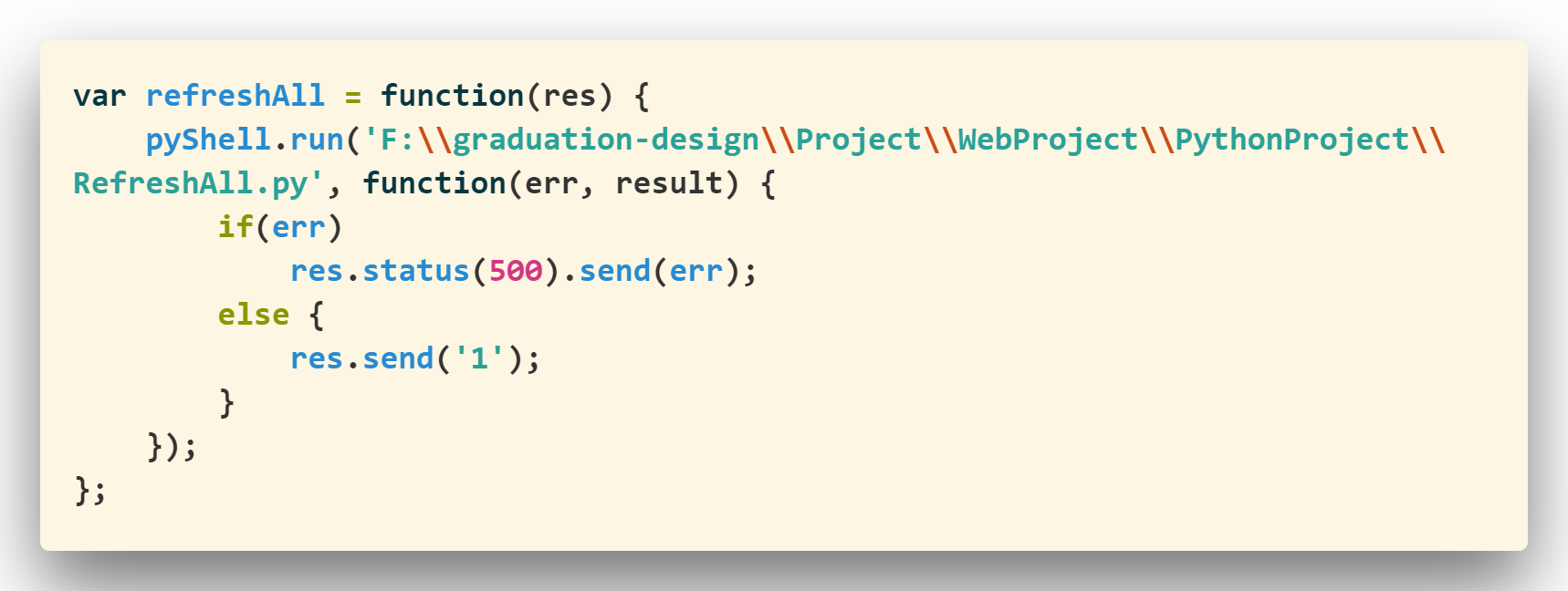
接口文档如下：

1. GET “/News” 获取新闻公告数据，无需传递参数。
2. POST “/BondsData” 返回指定年份指定类型的所有收益率数据及数据库中该类型数据的所有年份；参数为数据类型type（int型参数，1代表中债、2代表美债、3代表SHIBOR、4代表LIBOR），数据年份year（比如2018， 如果为空则返回数据库中最新年份的数据）。
3. GET “/Download” 返回指定年份指定类型的Excel文件数据，参数为数据类型type（字符传型数据，‘1’代表中债、‘2’代表美债、‘3’代表SHIBOR、‘SZ’代表深交所债券数据、‘SH’代表上交所债券数据、‘GZ’代表国家发行的债券数据）和year（数据年份）。
4. GET “/DetailData” 返回指定类型的所有最新债券数据，参数为数据类型type（字符串类型，比如‘SZ’、‘SH’和‘GZ’）。
5. GET “/YieldCurve” 返回指定日期与类型的债券的拟合收益率曲线数据。参数为date（数据日期，比如‘2018/05/07’）和type（数据类型，比如‘SZ’）。
6. GET “/ZZValuation” 返回所有中债估值数据，按时间降序排列。
7. GET “/ZZVaR” 返回所有中债VaR数据，按时间降序排列。
8. POST “/CalculateVaR” 返回指定置信区间、持有期和模拟次数的数据的在险价值。参数为数据上传方式type（1代表手写字符串数据、2为Excel或CSV文件数据上传），手写字符串数据data（使用文件上传方式时data可为空），上传文件的文件名file（比如‘data.csv’）、置信区间confidenceLevel（比如0.95），持有期holdingPeriod（比如1），模拟次数simCount（比如10000）。
9. POST “/Admin” 验证账号密码登录，返回值为登录码（①-1代表账户不存在、②0代表密码错误、③1代表登录成功）参数为账户名userName、密码password和是否记住密码的标志量remember（比如true）。
10. POST “/CalculatePCA” 计算PCA主成分的接口，返回各维度的累计贡献度以及因子载荷矩阵。参数为数据上传方式type（1代表手写字符串数据、2为CSV文件数据上传），手写字符串数据data（使用文件上传方式时data可为空），上传CSV文件的文件名file（比如‘data.csv’）、所要提取主成分的个数componentCount。
11. POST “/Search” 根据搜索条件搜索数据。参数为债券代码code（比如127326）、债券简称name（比如‘15国网05’）、债券类型type（0代表全部、1代表国债、2代表沪企债、3代表深企债、4代表中债VaR、5代表中债估值）。

查询数据代码如下：



Python运行Node程序的代码如下：



其中，pyShell是从python-shell模块中引入的一个对象。

数据库采用NoSQL的非关系型数据库——MongoDB，关于数据库数据表的设计没有什么多说的。

前端开发部分，主要采用了React框架开发视图层，使用AJAX封装库axios进行前后端交互，采用脚手架create-react-app（Facebook开发）进行代码打包、压缩等等。