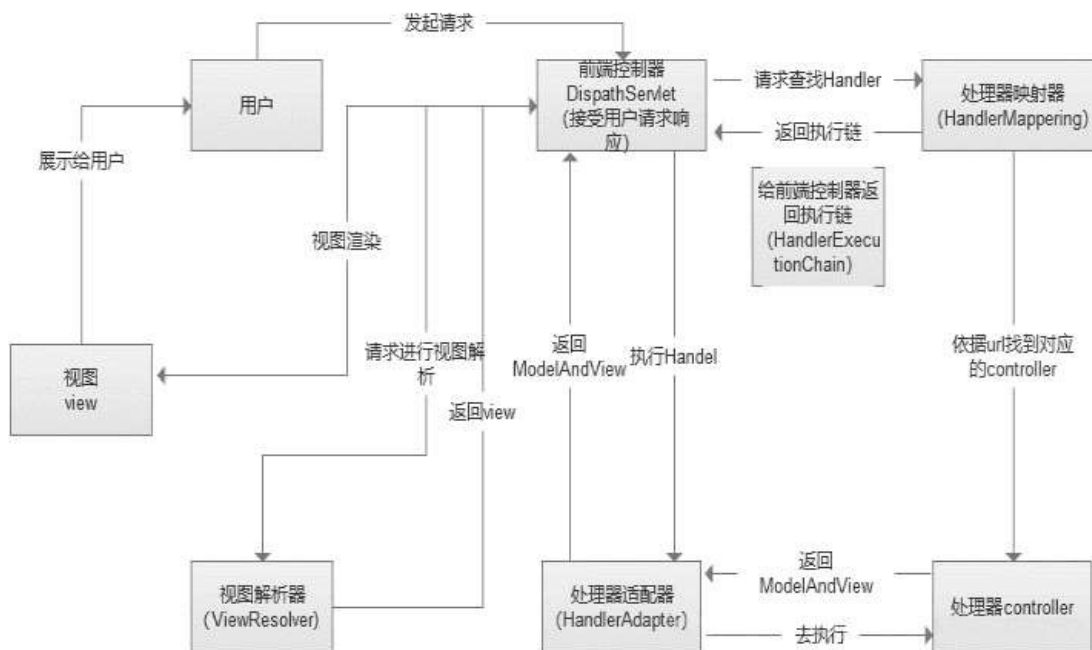


据分析预测；每种类型的用户都可以打印各类产品的价格图表方便今后直接查看。除此之外。管理员用户，负责每天的数据维护和更新还有用户权限的分配；但是因为目前该应用平台的推广和正处在成熟优化阶段，目前所有用户都免费享有会员用户的所有权限，所有历史数据也未在云端同步，各个用户也同时享有管理员权限，按天可以同步爬去所有数据。

### 3.3.2 数据流程图

前后端数据流转图：



业务流转图



图 3-3 数据流程图

基于 SSM 框架搭建的 B/S 架构 Java Web 网站，其最大的优势就是用户所需的配置尽可能地简单，对用户的隐私有着合法的保护；用户无需进行复杂繁琐的配置，只需要安装一个浏览器就能安全使用这个大数据应用平台，本网站的工作原理和前后端程序流转及业务流转都在上图 3-3 中具体体现出。

### 3.4 用例分析

本系统所拥有的所有数据都是通过互联网对特定网站进行爬取获得，所使用的技术是比较基础的 Java 常用爬虫技术，具体原理如上图 4-1 所示。首先需要前期调研进行爬取网站选取，本文所描述系统的网站是由我的导师所提供的。确定了要爬取的网站之后就要先注册该网站的账号，为了防止该网站的反爬虫机制，需要将登录账号后的 Cookie 信息添加到 HTTP 请求的头部。接着就要对该网站的 url 进行不断请求解析，然后进行对所爬取 html 页面的正则匹配。只有符合我们要求的有效规范数据才会被存储在数据库中。因为数据量比较大，所以在这里使用了多线程并发存储数据。这样不仅保证了爬虫获取数据效率的高效，同时还能有效防止所登录 Cookie 信息的过期而导致数据获取失败。

#### 4.2.2 产品未来价格预测分析

本文所描述的大数据应用平台的产品未来价格预测分析模块所应用的算法是目前市场上比较成熟的二次指数平滑算法，下面就是该算法的基本公式。

$$S_t^{(1)} = \alpha Y_t + (1 - \alpha) S_{t-1}^{(1)}$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha) S_{t-1}^{(2)}$$

指数平滑法对实际序列具有平滑效果，加权系数（平滑系数） $\alpha$  越小，平滑效果越强，但对实际数据变化的响应越迟缓<sup>[20]</sup>。在线性变化的实时序列之中，随着顺序偏差指数（阿尔法系数为平滑）的柔性加权值的增加，但当时间序列的线性趋势与第一个平滑指数有很大差异时， $N$  表示  $D$  偏差线性趋势的方向。 $e$  延迟曲线根据的是第一和第二个指数的柔性指数，并预计模型预测将是准确和客观的。

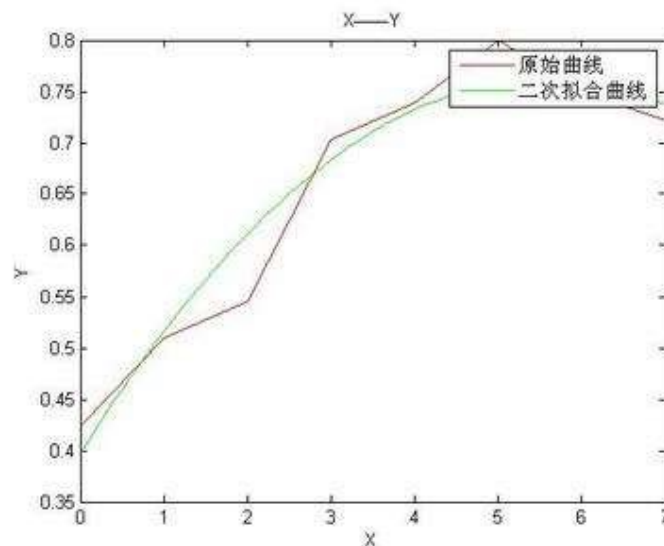


图 4-2 二次指数平滑算法原理图<sup>[20]</sup>

#### 4.2.3 大数据汇总

本文所描述的大数据应用平台的大部分功能都是对大数据的图表展示，这就离不开百度的开源项目技术——ECharts。百度开发的 ECharts（Enterprise Charts）拥有诸多特性：它提供了丰富

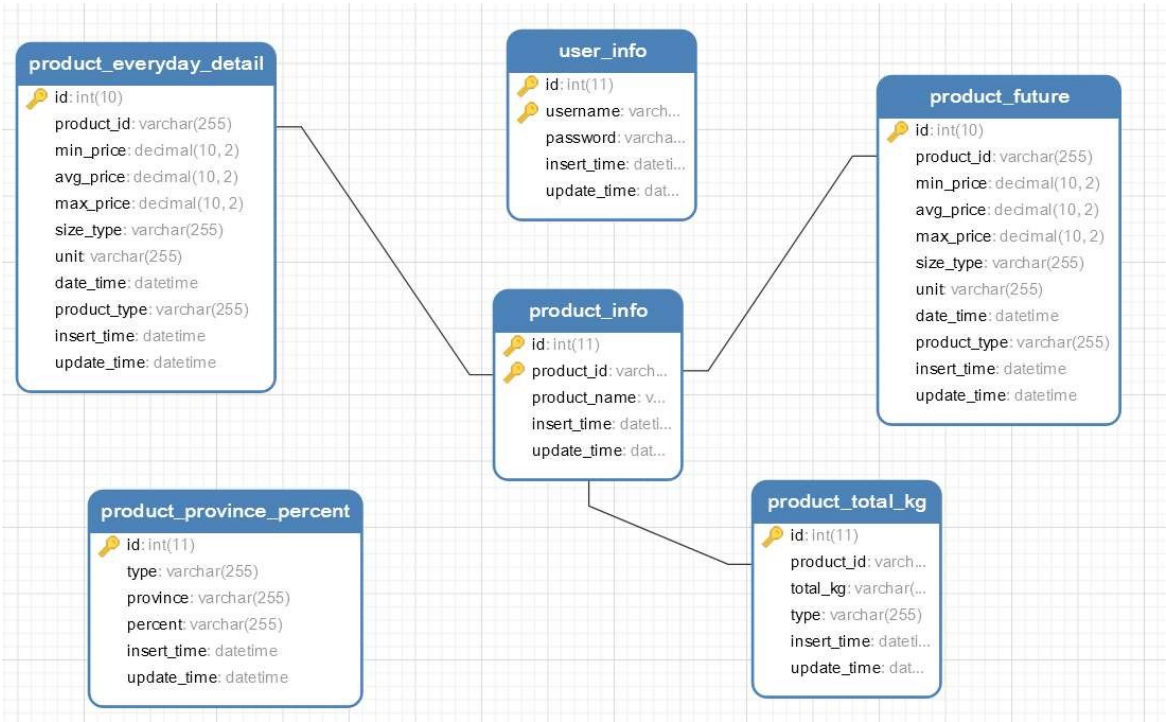


图 4-4 数据库逻辑结构图

逻辑结构图中的数据库表如表 4-1 所示。

表 4-1 数据库表

表名	备注
PRODUCT Everyday DETAIL	主要存储产品每天价格行情明细
PRODUCT FUTURE	主要存储产品未来七天的价格行情明细
PRODUCT INFO	主要存储产品的基本信息
USER INFO	主要存储用户的信息明细和权限配置
PRODUCT PROVINCE PERCENT	主要存储各类产品在各个省市的流通明细
PRODUCT TOTAL KG	主要存储各类热门产品信息明细

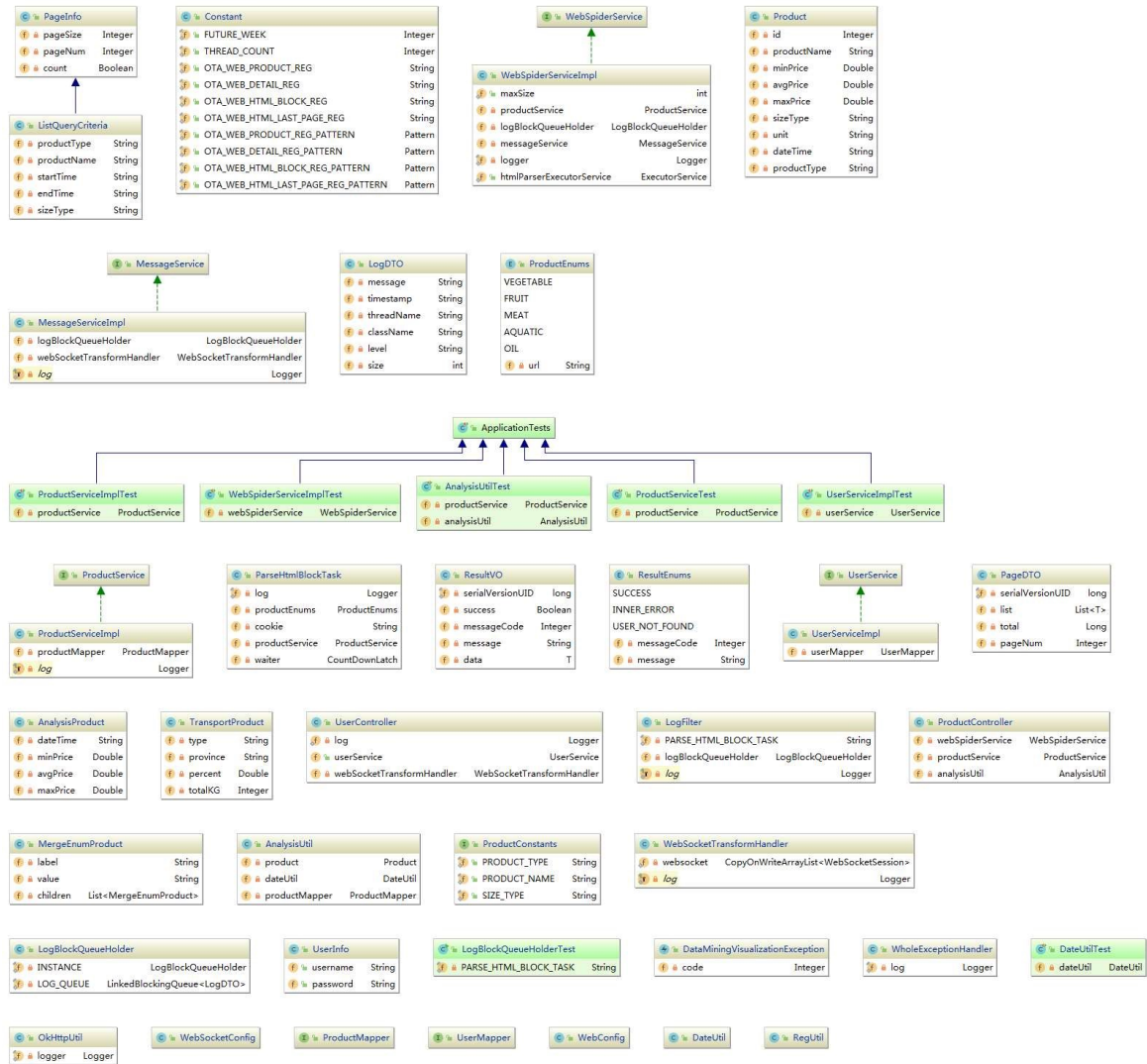


图 4-6 项目核心方法类图

## 4.6 项目文件结构

本项目采用前后端分离的技术，后端基于 Maven 进行自动化搭建，其具体的项目文件结构如图 4-7 所示。前端基于 NPM 进行 React 框架的搭建，其具体的项目文件结构如图 4-8 所示。

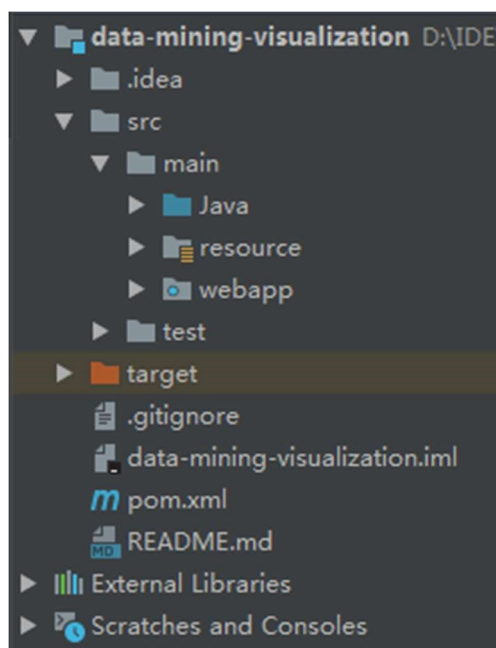


图 4-7 项目文件结构图（后端）

本系统后端的所有主要核心代码都在 Java 文件夹中，资源配置在 resource 文件夹中，所有的测试用例和方法在 test 文件夹中。使用 Maven<sup>[21]</sup>最大的好处就是不需要进行 jar 包的单独下载和引入，本系统所有的 jar 包都在 pom.xml 文件中，管理规范，简便开发步骤和提高开发时效。

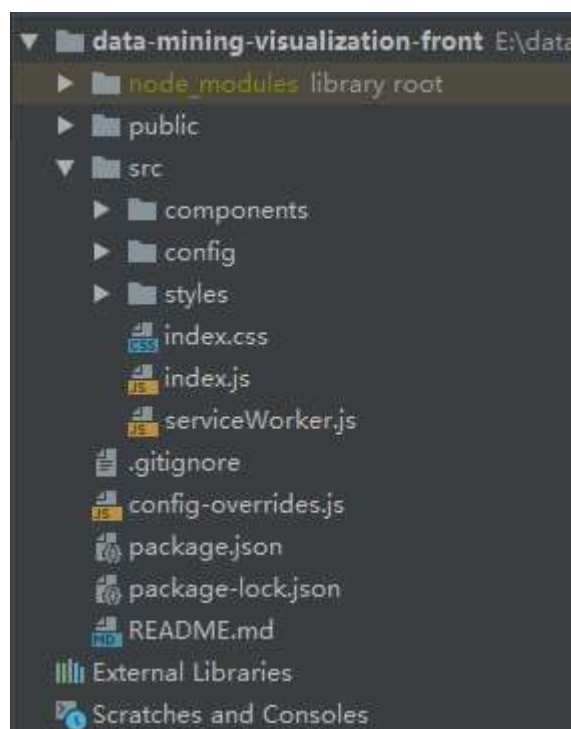


图 4-8 项目文件结构图（前端）

本系统前端的主要核心代码在 components 文件夹中，所有的依赖包都通过 NPM 进行安装引入在 node\_modules 文件夹中，package.json 文件中存储了各种依赖包的版本，统一管理方便后期版本升级等。具体项目文件结构如图 4-9 所示。

Java 文件夹主要存放了所编写的后端核心代码，com.zafu.nichang 为本人的包名，config 中编写了主要的全局配置文件，比如跨域请求和 Websocket；constants 文件夹主要存放本项目的一些常量的申明；controller 中编写了本项目的接口，并进行不同 URL 请求的拦截；entity 中是一些实体对象的一些扩展，在这里统一规范了返回前端请求数据的主要结构；enums 进行全局的一些枚举值的申明；exception 对本项目的异常进行了统一规范的处理；filter 主要进行前后端日志格式的交互规范和本地后端控制台日志的保存存储；handler 中的主要模块是使用 Spring 中的 Websocket 功能，搭配后端日志进行实时前端推送消息；mapper 主要编写 Mybatis 和数据库交互的接口方法；model 申明了本项目中的所有实体类、抽象类；service 是本项目的主要逻辑层，包含各种接口方法和其实例函数；util 主要进行数据爬取模块的编写，解析网页、多线程高并发进行爬取等操作。本项目后端核心代码的文件结构如图 4-9 所示。

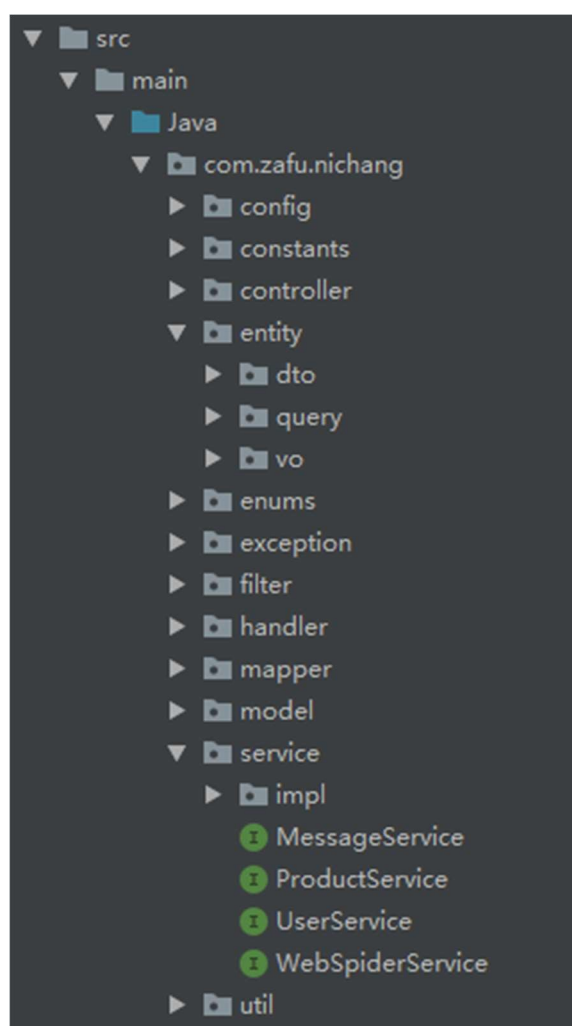


图 4-9 核心代码目录结构图（后端）

本项目的前端核心代码目录结构主要如图 4-10 所示，主要在 components 文件夹中进行编写各个界面；在 config 文件夹中主要进行全局的 IP 和 URL 配置和后端返回 data 数据格式的处理。



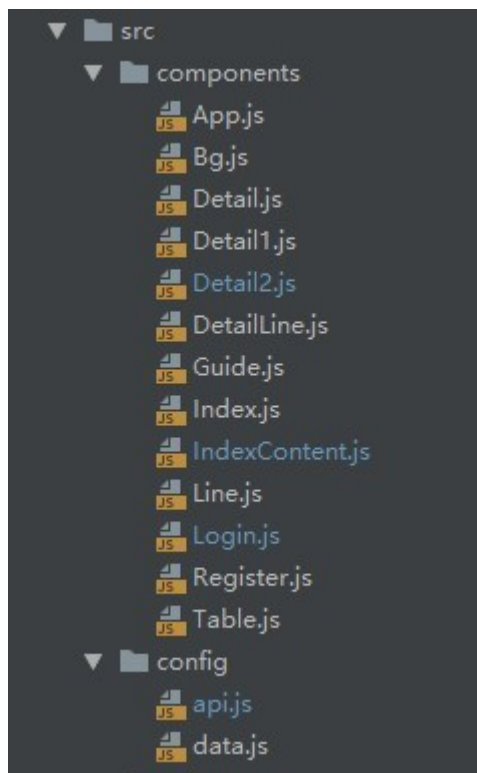


图 4-10 核心代码目录结构图（前端）

#### 4.7 项目命名规范

本项目后端代码参考《阿里巴巴 Java 开发手册》进行标准化格式规范编码<sup>[22]</sup>。前端采用标准化前端编码规范：HTML 规范、CSS 规范、JavaScript 规范以及各类文件命名规范。

## 5 系统实现

### 5.1 用户模块

本应用平台的主界面主要由三部分组成，标题部分、动态背景以及主要主体。其中动态背景是通过构图函数和时间延时函数进行组合所构建出来的。最终形成一个从左下方往右上方在一定范围内随机速度运行的动态背景壁纸。

#### 5.1.1 登录界面

系统的登录界面的主体是两个输入框、保持登录按钮、忘记密码链接和登录注册按钮。输入正确的用户名和密码，点击“登录”按钮后，后台根据登录用户的用户名密码进行登录权限认证，认证成功后跳转到该用户所具有的功能模块，否则登录失败，显示登录失败具体信息。目前，为了应用平台的推广和使用，后台数据库默认添加所有新注册用户均为管理员用户。具体如图 5-1 所示。

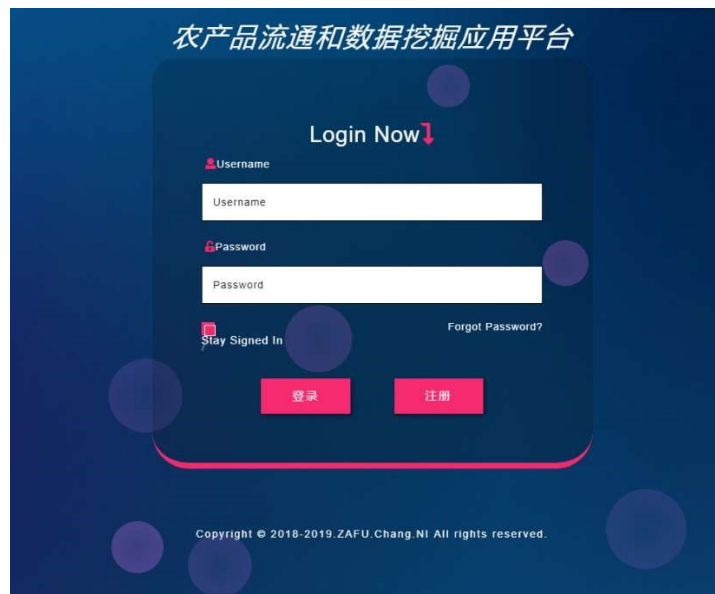


图 5-1 用户登录界面图

#### 5.1.2 注册界面

在登录界面点击“注册”按钮即可跳转到新用户注册界面。用户输入自己设定的用户名密码后进行前端的非法字符判定，判定为合法后请求到后端进行用户名的合法性判定，这里主要进行用户名是否被占用等。验证通过后跳转到登录界面，否则让该用户重新注册。具体界面详情见图 5-2 所示。



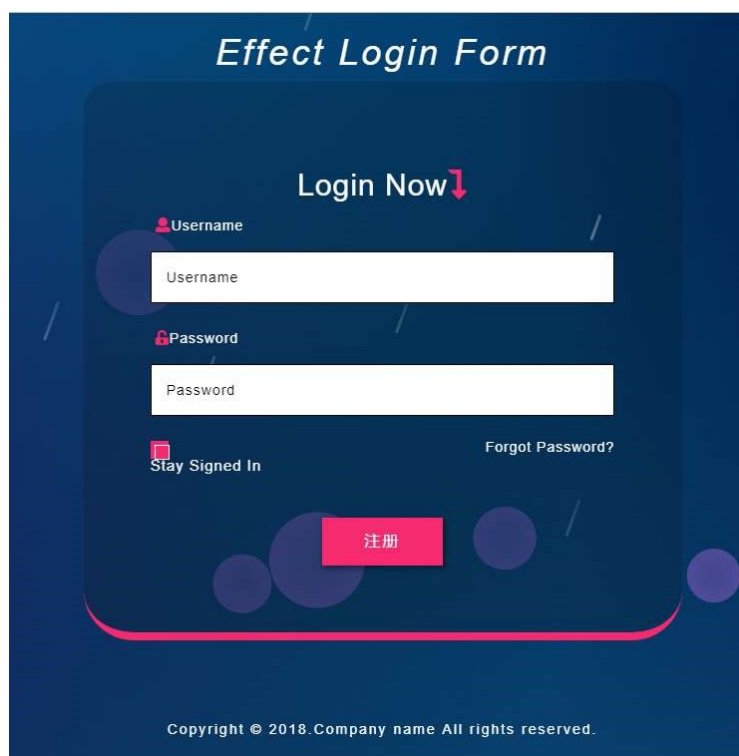


图 5-2 用户注册界面图

### 5.1.3 导航栏

在用户成功登录系统后，在本系统的左侧有一个可以伸缩的导航栏。这个可伸缩导航栏既能满足全屏首页界面的美观，又能满足用户正常随意跳转本系统的任一功能模块。具体界面详情见图 5-3 所示。



图 5-3 应用平台导航栏界面图

## 5.2 首页模块

本模块页面如图 5-4 所示。屏幕正上方可清晰看到本项目的名字《农产品流通可视化和数据挖掘平台》，首页主要由 8 个小模块组成。屏幕正中间是一张可以缩放的中国地图，其主要数据展示的是上个月以北京为例的农产品分销全国在各地的明细占比，弹框每隔一定时间进行跳转，能让用户更为直观地看出各地的供给需求占比。其余七个模块将在下方详细介绍。

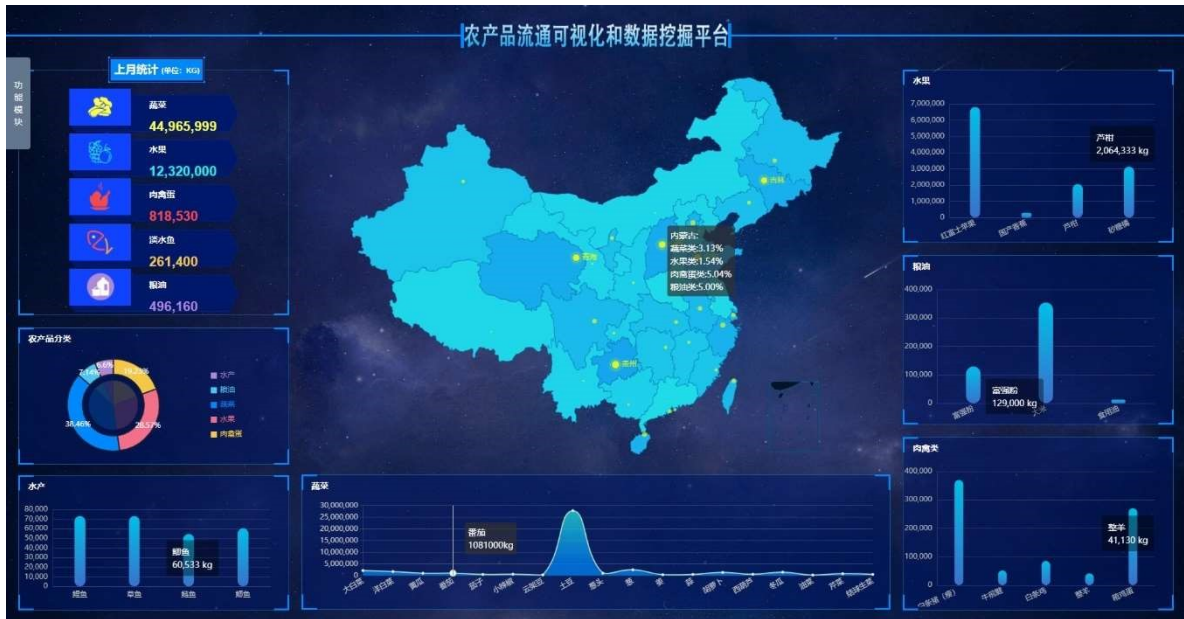


图 5-4 农产品大数据应用平台首页界面图

### 5.2.1 上月统计

左上方导航栏右侧是上月销量总计图表(注:本项目中的数据都是以北京为产地的试点数据,并不是全国总数据)。主要展示蔬菜类、水果类、肉禽蛋类、淡水鱼类和粮油类的上月销量。详情模块如图 5-5 所示。



图 5-5 各类产品上月销量图

### 5.2.2 农产品分类统计

在“上月统计”模块的下方是“农产品分类统计”模块:各类农产品的销量比重扇形图。鼠标经过扇形图的具体数据区域会变大突出,并弹出提示框进行数值可视化展示。具体模块效果图参见图 5-6。



图 5-6 农产品分类扇形图

### 5.2.3 农产品明细统计

“农产品明细统计”模块由 5 个小模块组成，分别为 5 类产品的具体明细数据可视化表达。每隔明细数据模块的提示框和“流通模块”的提示框一样，每隔一定时间自动切换，方便用户查看。具体模块如图 5-7 所示（以肉禽蛋类数据为例）。

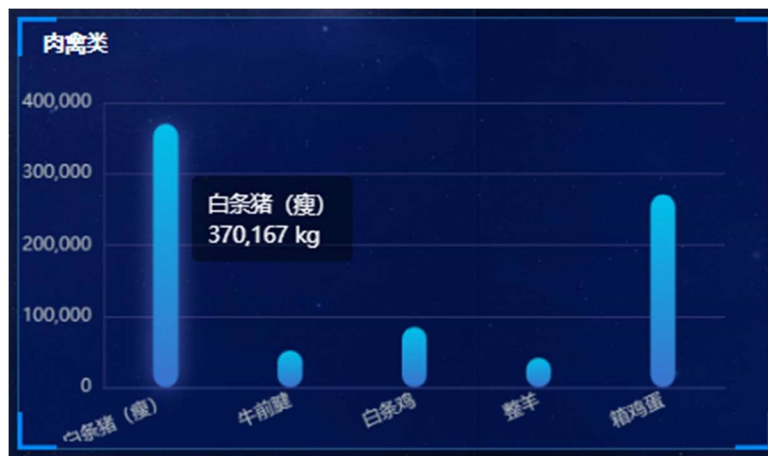


图 5-7 热门产品分类显示模块图（肉禽蛋）

### 5.3 产品价格详情模块

进一步通过点击伸缩式导航栏进行交互，选择“产品价格详情”即可进入该模块。该模块主要进行蔬菜、水果、肉禽蛋、水产和粮油等 5 大产品的每天价格历史明细。界面整洁简单，生动形象，适合绝大多数用户便捷操作。具体界面如图 5-8 显示。



图 5-8 产品价格详情页

### 5.3.1 图片按钮分类

这一模块用到了阿里巴巴开源的 Ant Design，图片按钮背景透明，简单形象，方便不识字的小孩子群体和文化水平程度低的老年人使用。点击对应图即可切换下方图表中的数据可视化展示。具体界面如图 5-9 所示。



图 5-9 分类按钮模块

### 5.3.2 价格比较图

本文所有的可视化图表大部分都采用百度开源的 ECharts 控件<sup>[23]</sup>，能比较直观的显示每种产品的最高价，最低价和平均价，方便用户挑选性价比最高的产品。具体图表情况见图 5-10。



图 5-10 价格比较图

5.3.3 分页表格

“分页表格”模块详情记录了每种产品的历史价格，按照日期降序分布，下方展示每种数据的总数据量，并能支持用户根据自己喜好每页展示多少条数据，下方表格中的数据发生变化后，为了方便用户直观体现表格中数据的变化，上方的价格比较图也将实时进行数据展示。具体的模块展示图如图 5-11 所示。

品名	最低价	平均价	最高价	规格	单位	发布日期
箱装菠萝（上六下六）	1	1.15	1.3	其他类	斤	2019-01-31
圣女果（千禧）	4.1	5.35	6.6	其他类	斤	2019-01-31
黑枣	3	5.5	8	其他类	斤	2019-01-31
冬枣	15	22.5	30	其他类	斤	2019-01-31
红颜草莓	10	12.5	15	其他类	斤	2019-01-31
丹东九九草莓	14	16	18	其他类	斤	2019-01-31
甜宝(奶油)草莓	13	15.5	18	其他类	斤	2019-01-31
白雪公主草莓	30	32.5	35	其他类	斤	2019-01-31
巧克力草莓	18.5	19.75	21	其他类	斤	2019-01-31
甜查理草莓	11	11.5	12	其他类	斤	2019-01-31

当前第 1 到 10 条，总共 42246 条记录

< 1 2 3 4 5 ... 4225 > 10 跳至 页

图 5-11 产品分页展示图

5.4 未来数据挖掘模块

继续点击导航栏，就能进入本项目最为核心的模块——“未来数据预测”模块。具体的模块页面如图 5-12 所示。其核心的算法在本文的 4.2.2 中已经进行了详细阐述，在此不再赘述。下文会介绍具体的功能。

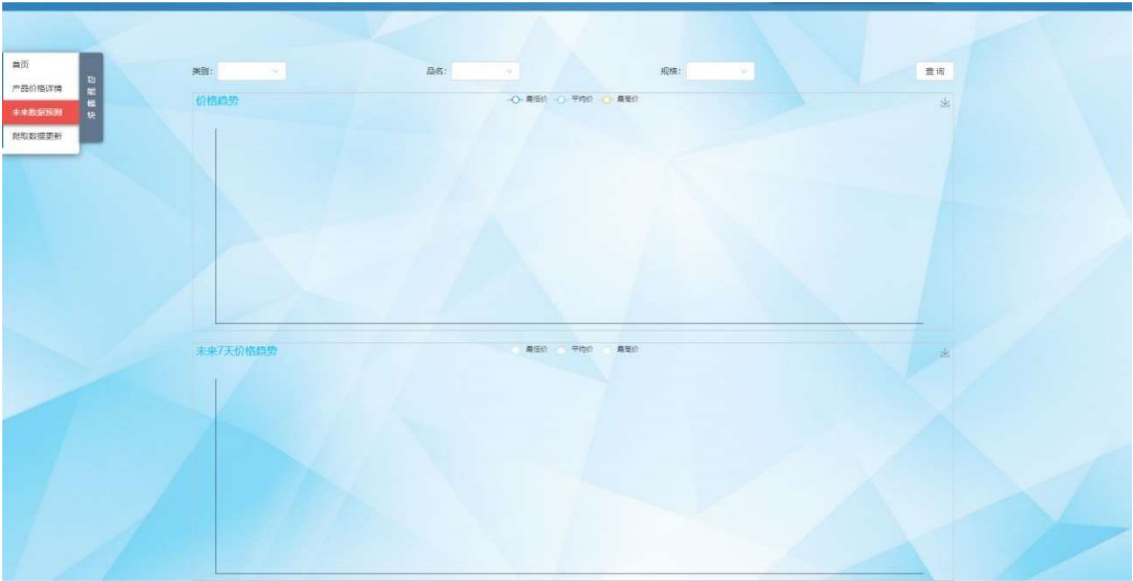


图 5-12 未来数据预测页面总览图

### 5.4.1 树状索引搜索查询

本页面上方的三个产品选择框之间有着密不可分的联动关系。三个选择框有着依次递进的层级关系，当用户首先选择了产品的类别（如图 5-13 所示）才能进行选择产品的具体名称。



图 5-13 产品类别选择列表

当用户选择了产品的类别，产品名称的列表将会显示具体的名称。以“肉类”为例，用户选择了肉类后，产品名称列表将显示具体的肉类名称，而不会显示其他类别产品下的具体产品名称。详情见图 5-14 所示。

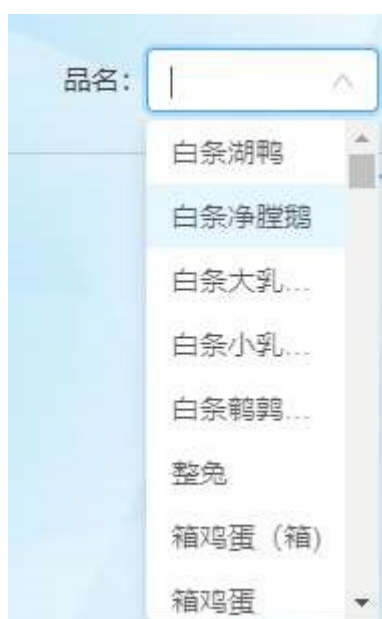


图 5-14 产品名称选择列表

同理，产品规格类别列表也是如此，详情见图 5-15 所示。这样做的好处可以精确筛选出具体的某一种产品。避免出现因产品名称相同但规格不同而导致产品的未来价格预测分析出现极大误差。比如（品名：鲫鱼，规格：<500g；品名：鲫鱼，规格：500-750g），去过菜场或者农贸市场的用户都知道，显然，两者的价格肯定不一致。通过联动树状索引的查询，可以完美解决这一问题。





图 5-15 产品规格选择列表

#### 5.4.2 历史价格走势与未来数据挖掘

本模块经过网站后端的实时运算，在第一张图表中展示该产品在过去 90 天的历史价格数据的可视化展示。之所以选择过去 90 天，是因为农产品不像是一般商品，它有着很强的季节性变化，所以分析数据的原始数据样本不宜超过一个季度，但是数据量太少也会产生数据分析的偶然性，所以本人就选择一个季度作为原始的数据分析样本数据。在第二张图表中展示该产品的未来 7 天价格趋势变化。这两张图表都包含某一产品每天的最高价格、最低价格和平均价格，方便用户了解市场行情。具体详情见图 5-16 所示。



图 5-16 未来数据分析预测图

#### 5.5 爬取数据更新模块

此模块是为了管理员进行数据的定期维护和数据更新。具体的详情页面如图 5-17 所示。

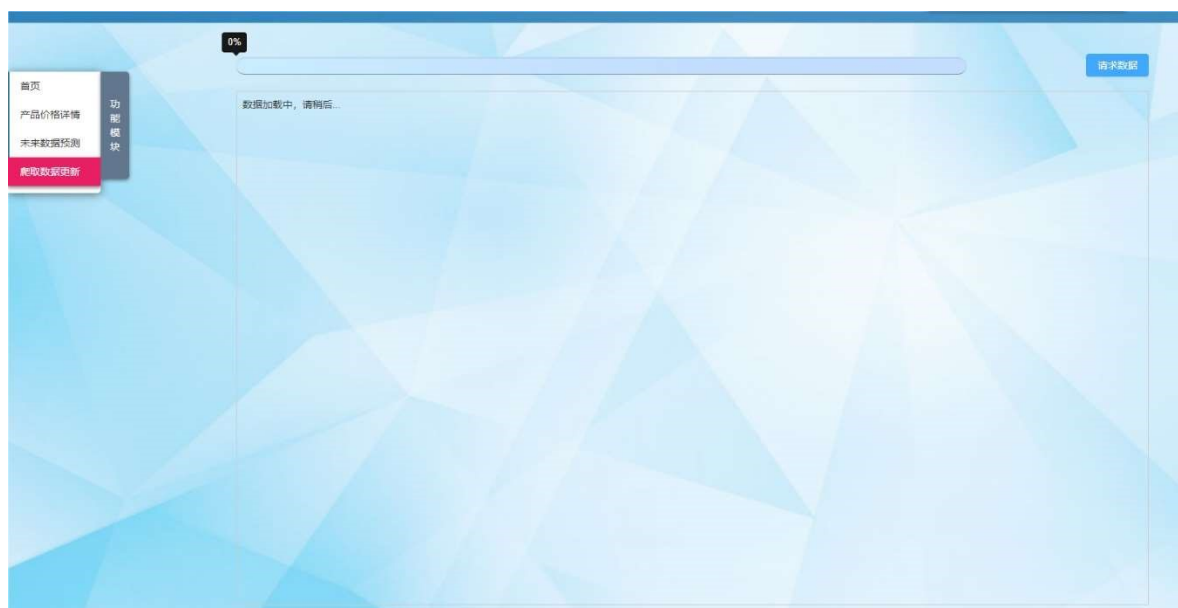


图 5-17 数据爬取模块总览图

### 5.5.1 爬取进度交互

此模块为了让用户（主要指管理员）进行数据更新的可视化展示。按下进度条后的“请求数据”按钮即可开始爬取数据，当数据爬取更新完成后进度条会增长至“100%”。具体如图 5-18 所示。



图 5-18 爬取进度交互

### 5.5.2 爬取信息交互

本模块主要通过 Websocket 进行实时交互，当数据爬取开始时，显示“数据加载中，请稍后...”开始提示语。此时页面会静止一段时间进行后台数据爬取的准备。每条语句会显示具体爬取的页面信息，比如：页面总页数、当前爬取页数、是否需要过滤爬取当前页等等。爬去完毕后会显示“数据请求完毕！”结束提示语。具体的页面如图 5-19 所示。

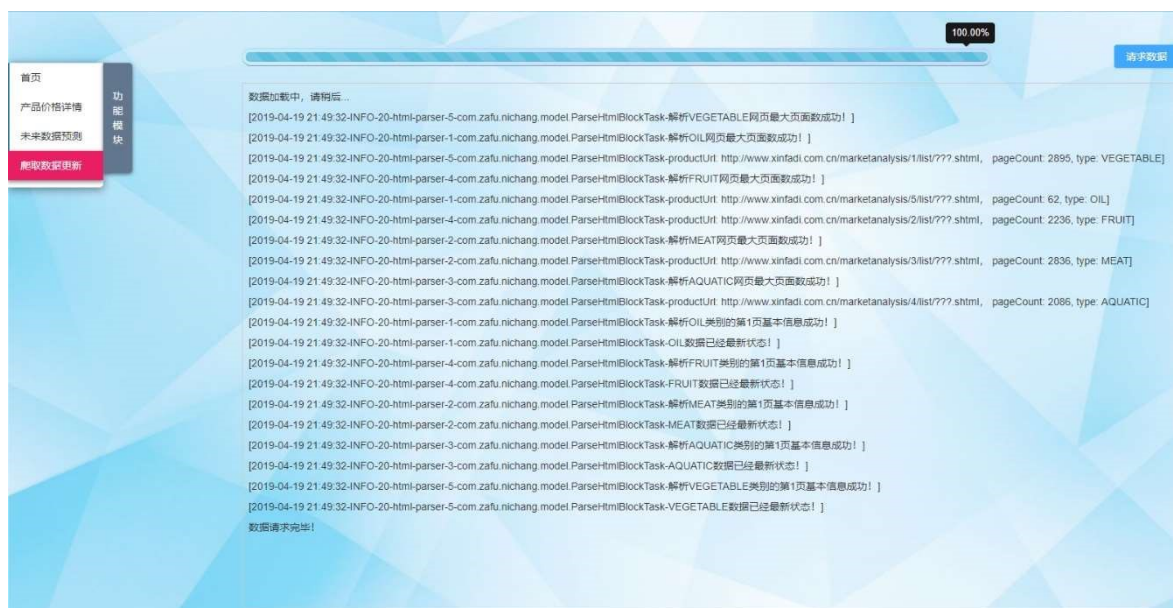


图 5-19 爬取信息交互