**基于Spark的股票分析系统**

**组长：何禹姗**

**组员：惠文静**

**刘俊彤**

**日期： 2024年7月26日**

1. **项目背景**

**1.市场需求**

随着全球金融市场的快速发展和复杂化，股票交易数据量呈爆炸式增长。传统的股票分析方法已难以应对如此庞大的数据量，且处理效率低下，难以实时捕捉市场动态，需要更加高效的数据处理能力。

**2.技术发展**

大数据技术，特别是分布式计算框架的兴起，为处理海量数据提供了强有力的支持。Apache Spark作为当前最流行的大数据处理框架之一，以其高效、可扩展、易用等特性，在处理大规模数据集时展现出显著优势。

**3.挑战**

股票数据来源多样，质量参差不齐，需要有效的数据清洗和校验机制。同时，由于金融市场的瞬息万变，数据的实时处理能力也十分重要，如何在保证数据安全的前提下构建一个高效准确的股票分析模型是一个巨大的挑战。

**4.预期**

利用大数据和spark技术，构建风险预警模型，提前识别潜在的市场风险，降低投资损失，此外通过深入分析历史数据和市场趋势，为投资者提供个性化的投资策略建议，提高投资回报率。

1. **项目功能描述**

系统设有注册页和登录页，用户需要先登录才可以使用查询功能。登录后首页和查询页会显示当前登录的用户名和头像，头像旁边会出现退出登录按钮；页面包含导航栏，可以快捷导航到各页面；首页写明了系统功能简介，下方四个组块介绍四种查询功能，并可以点击跳转对应的查询页；

系统共提供三个查询功能：

1、单只股票k线查询：输入开始时间、结束时间、股票名称、查询数量，点击查询。结果呈现两张图，图1为股票k线图以及均线和涨幅线，图2为成交量和成交量均线；

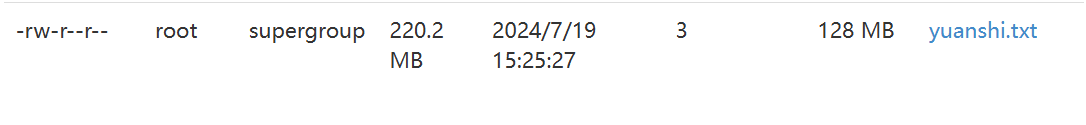
2、两只股票对比查询：输入开始时间、结束时间和两只股票名称，点击查询。结果呈现两张图，分别为两只股票的k线和均线；

3、单日所有股票查询：输入待查询日期，点击查询，结果呈现为列表，显示当天所有股票的股票名称、开盘价、收盘价、最高价、最低价、涨幅和成交量。

1. **数据**

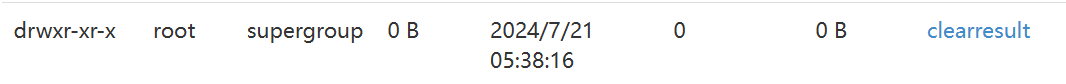
对于股票的原始数据，我们将其导入集群中进行清洗分析，具体的过程包括：

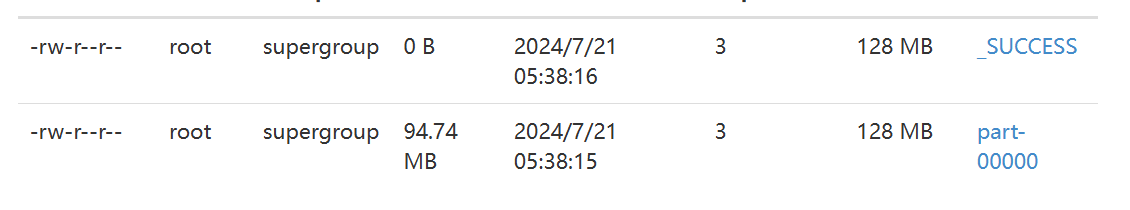
1. 导入原始数据：在SFTP中输入“put D:\yuanshi.txt”将本地文件导入集群



1. 接着在idea中生成两个架包，分别对原始数据进行清洗和筛选

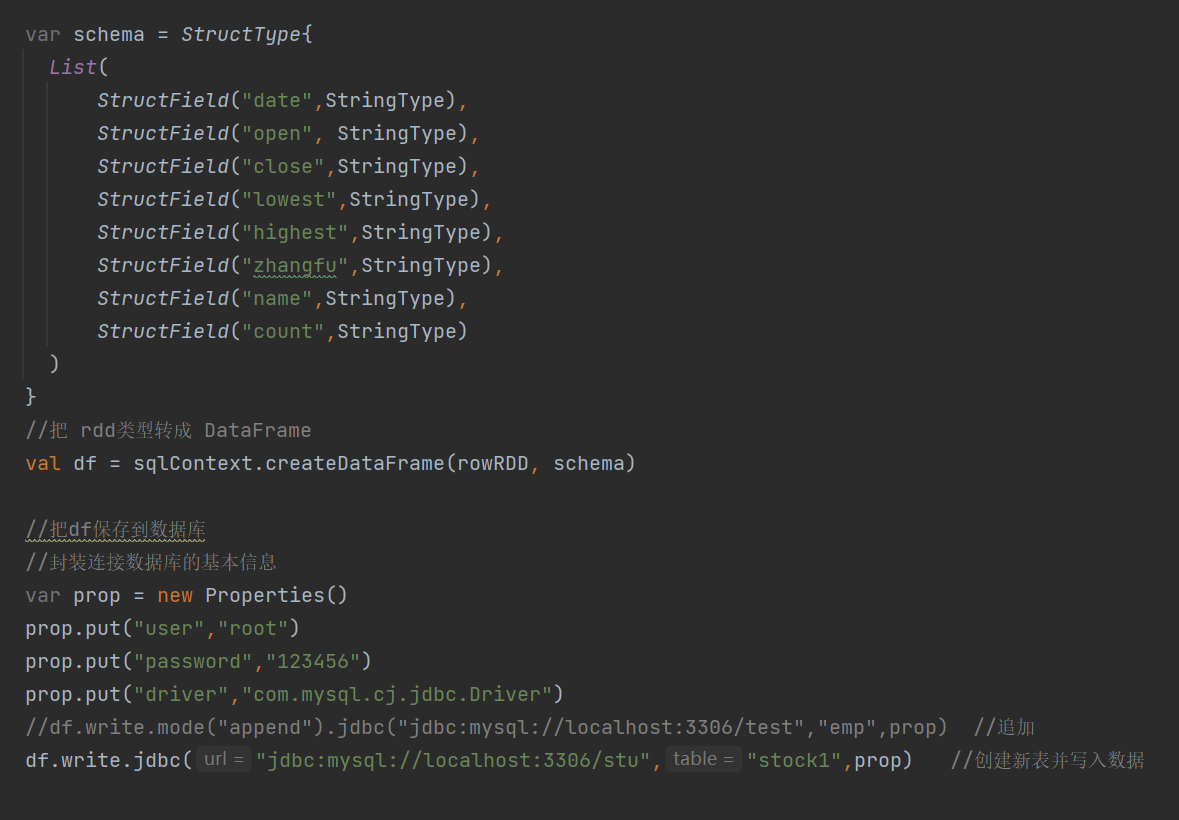


这个架包的目的是筛选原始数据中不含空白项的项目，并将其保存在” clearresult.txt”文件中，再通过”/root/apps/spark-2.2.3-bin-hadoop2.7/bin/spark-submit --class com.Aa --master spark://cent1:7077 /root/Aa.jar ”指令运行架包，在hdfs中查看结果：

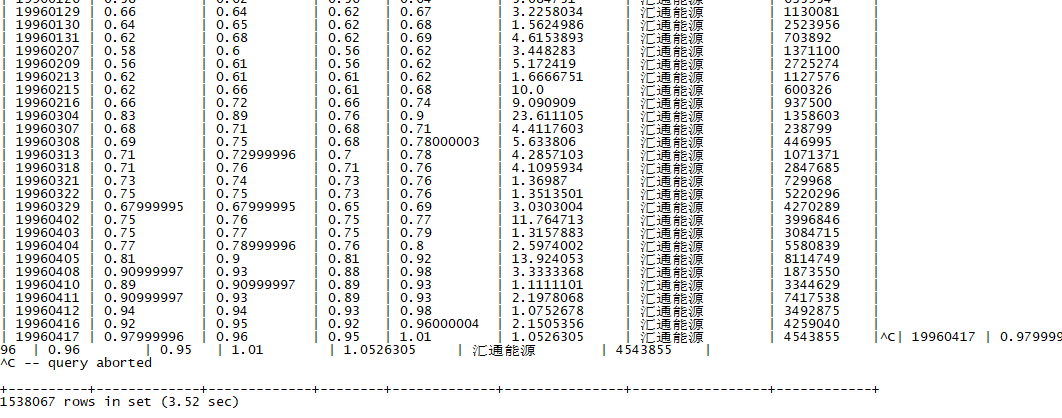


说明我们第一个架包成功运行。

接着进行第二个架包的运行，第二个架包的目的是筛选结果文件中阳线数量大于阴险数量的股票数据并将其保存到数据库中，部分代码如下：



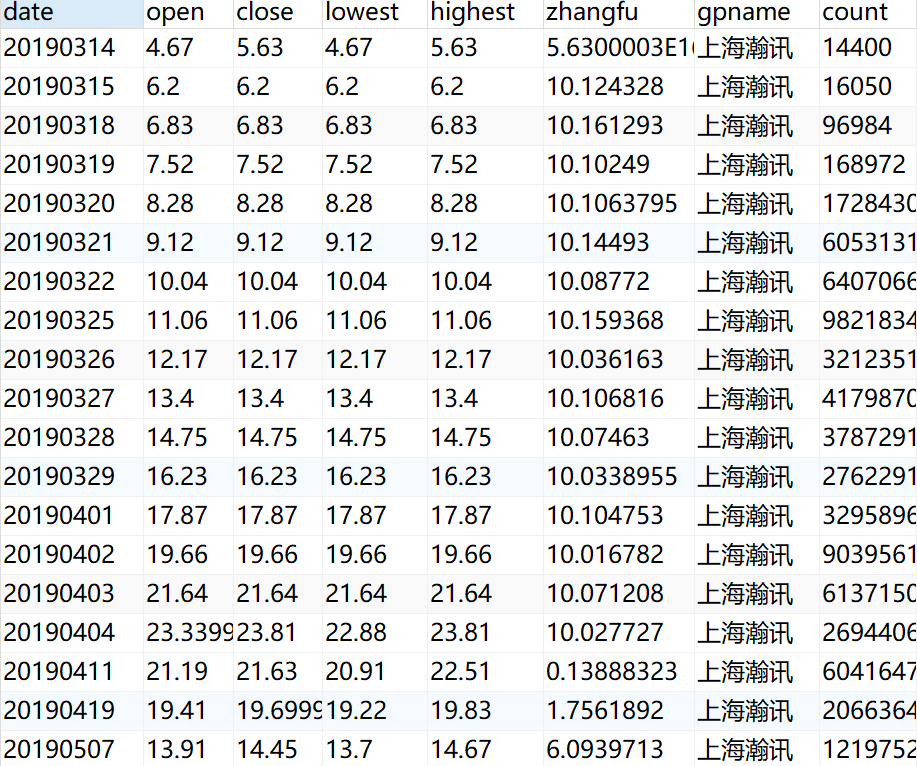
同第一个架包运行的过程相同，运行成功后，此时我们打开mysql数据库，检查数据库中是否有数据存在：



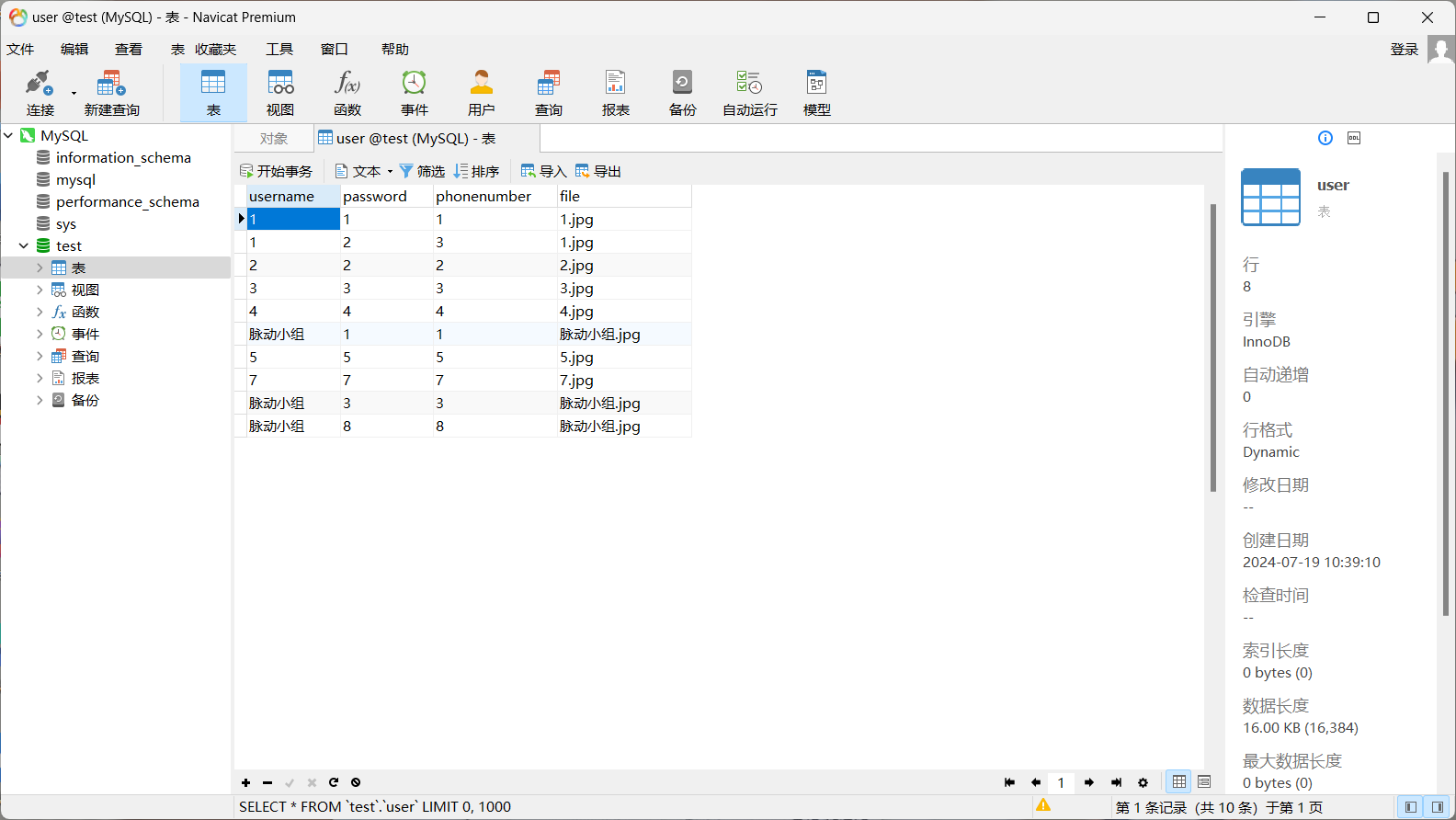
说明我们成功将清洗过的数据导入到集群中的数据库中。

1. 举例部分股票信息表数据：

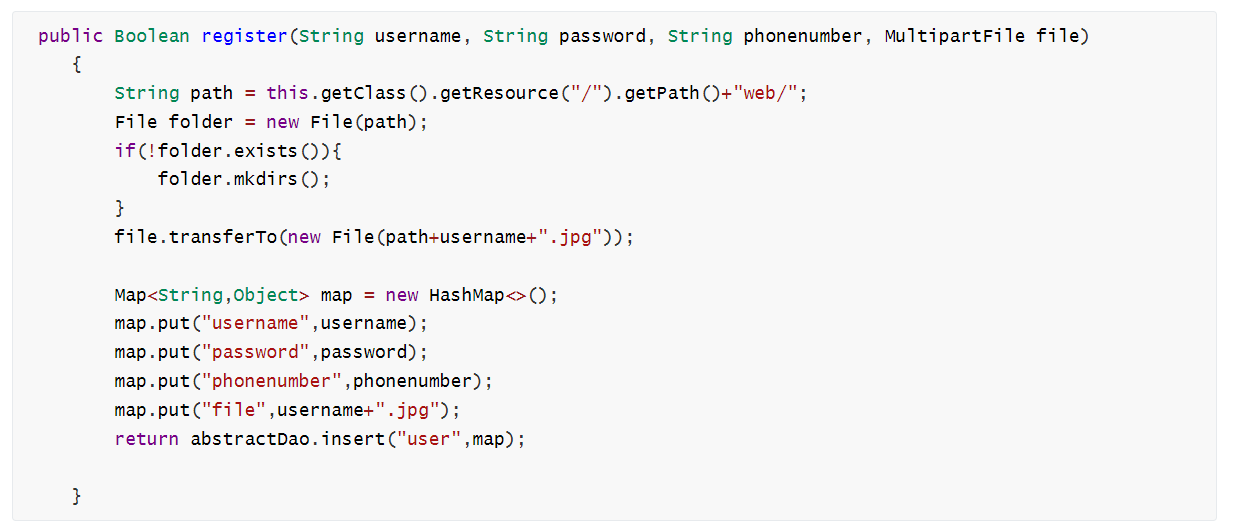
日期、开盘价、收盘价、最低价、最高价、涨幅、名称、成交量



此外，除了在集群的数据库中保存的股票信息，我们还在本地的数据库中保存用户的具体信息，我们设置了四个变量，包括用户名、用户密码、手机号和头像，其中前三者通过在注册页面中用户填写时存入本地的数据库中，用户头像则通过后端代码，线将其存在后端的文件夹中，再将文件名通过前端代码存储到用户的数据库中。

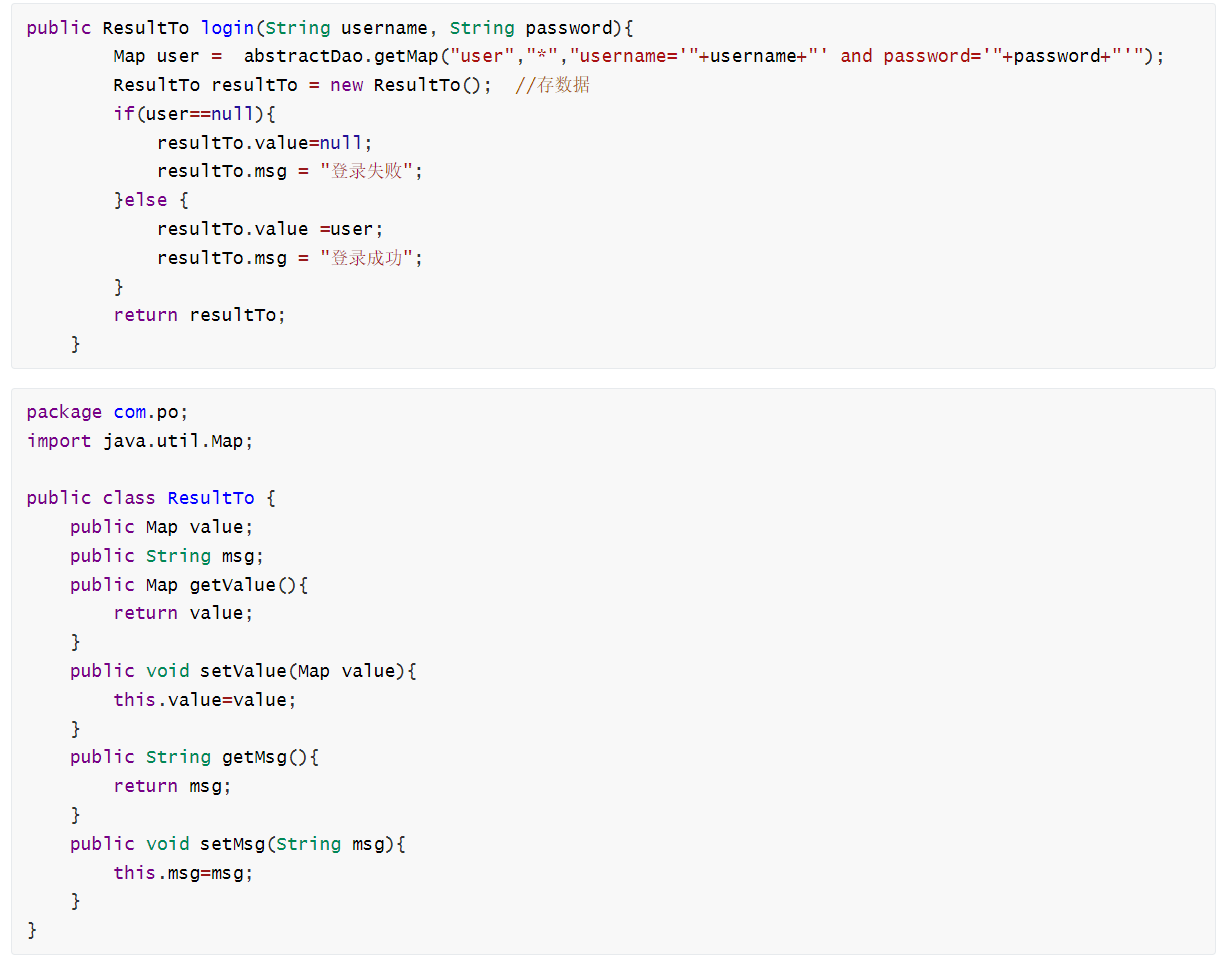


1. **功能实现**
2. **注册页面**





1. **登录页面**



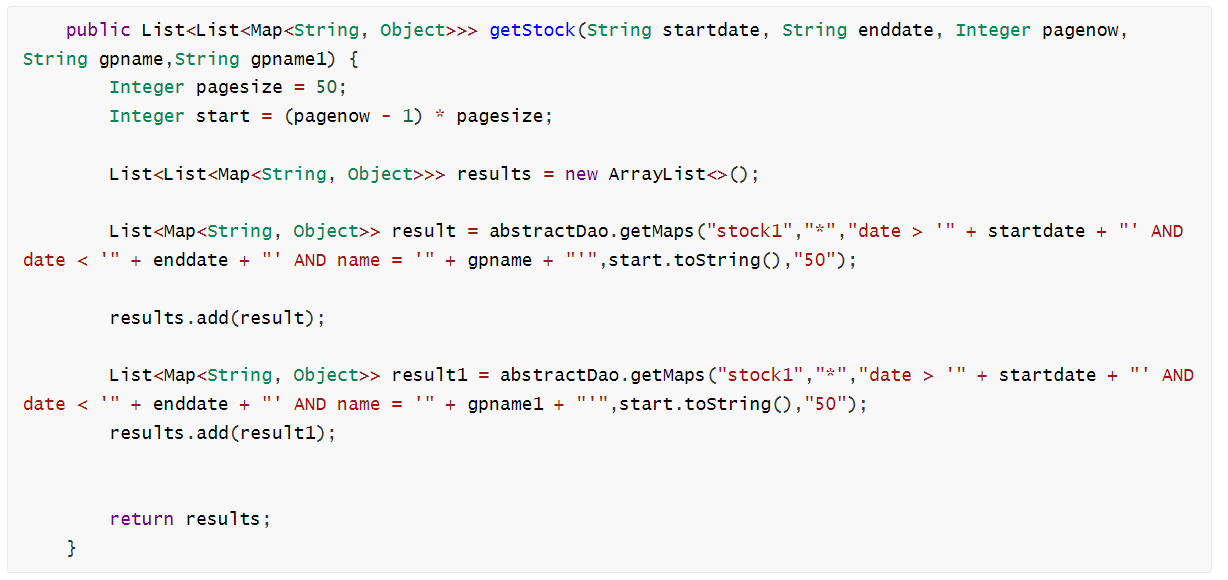


1. **k线查询**



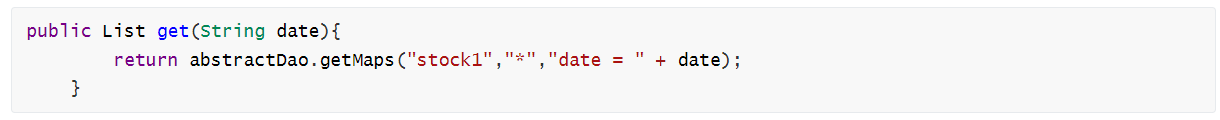


1. **对比查询**





1. **单日查询**





1. **页面展示**

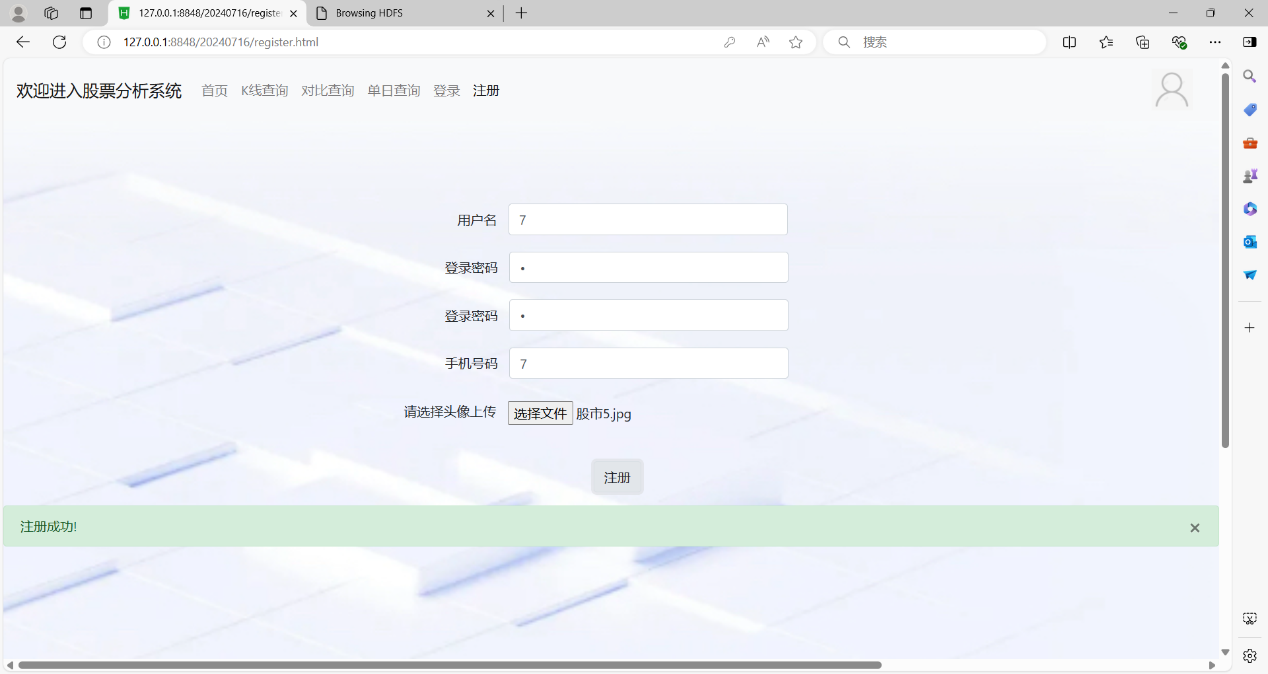
**1. 注册页面**

未登录状态导航栏会显示注册页和登录页，用户头像为默认灰色头像。

注册信息未填写完整，点击注册则会注册失败，弹出信息为空的提示框，如图所示：

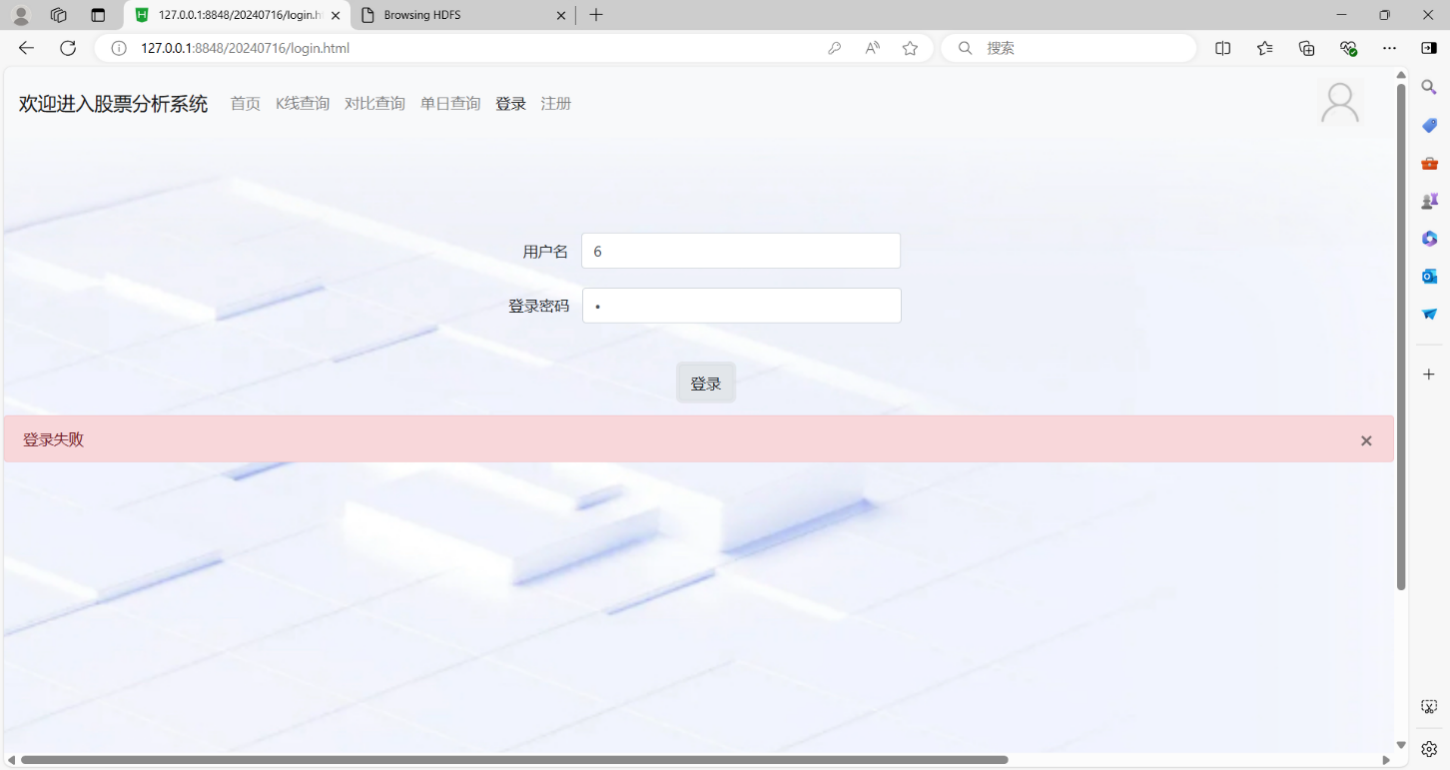


完整填写注册信息后点击注册，则会弹出注册成功的提示框，两秒之后，跳转登录页。

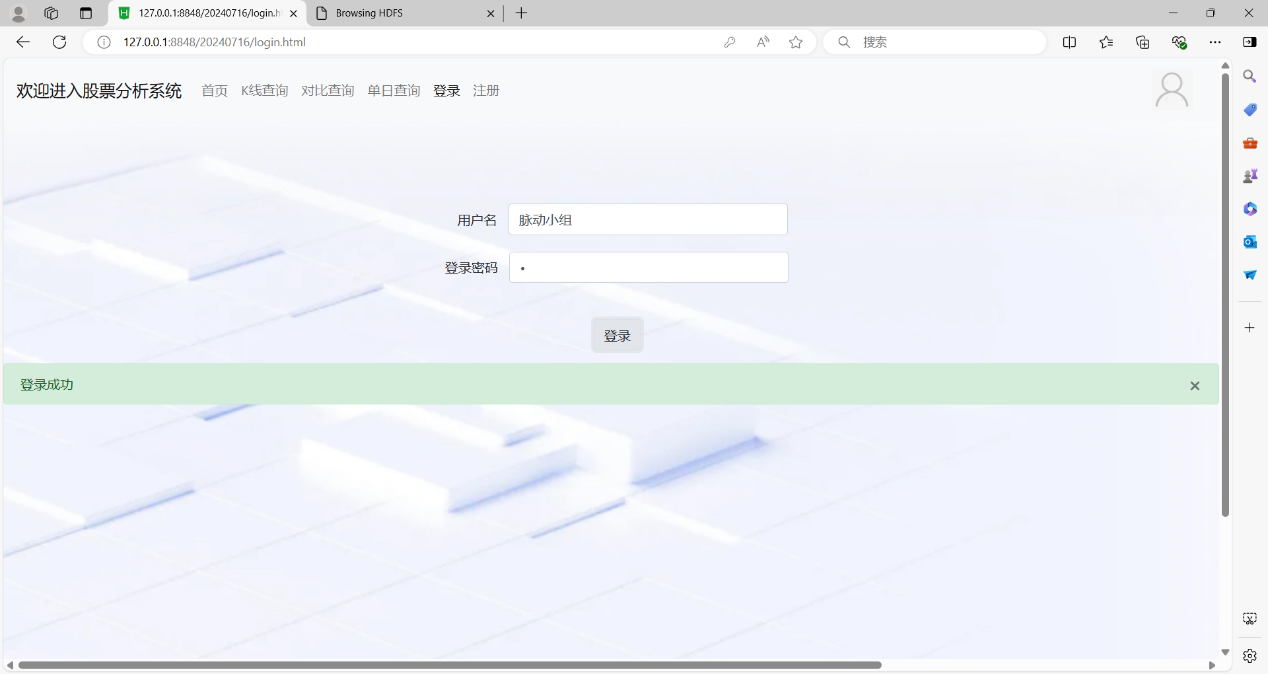


**2.登陆页面**

未登录状态，登录页面，输入用户名和密码进行登录，若输入的用户信息不存在，则会弹出“登录失败”提示框。



登录成功后，弹出“登录成功”提示框，两秒后跳转系统首页。

****

**3.首页**

用户登录后各页面导航栏显示首页以及三个查询页，显示欢迎用户字样以及用户头像。

首页上半部分为系统的简介，下半部分4个信息框分别对应了4种查询方式。进行了查询功能简介以及按钮，点击可跳转对应的查询页。

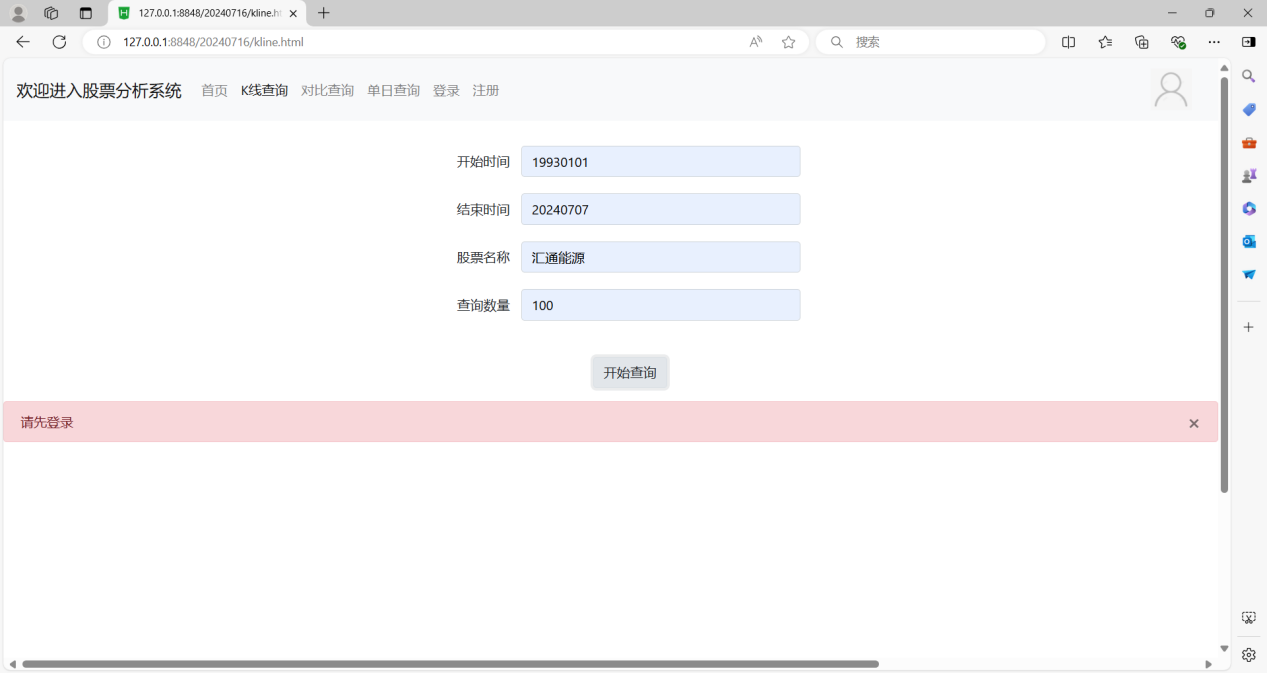
该项目中，我们共实现了三个查询功能，分别为单只股票的k线查询、两只股票的对比查询以及单日所有股票信息查询。

****

**4.查询页**

在未登录状态进入查询页，如下图所示，导航栏会显示登录和注册页面索引，不会显示欢迎用户字样，头像也为默认的灰色头像。

进行查询，点击开始查询按钮，会弹出提示“请先登录”，两秒后跳转登录页面。用户需要先登录才能使用系统的查询功能。

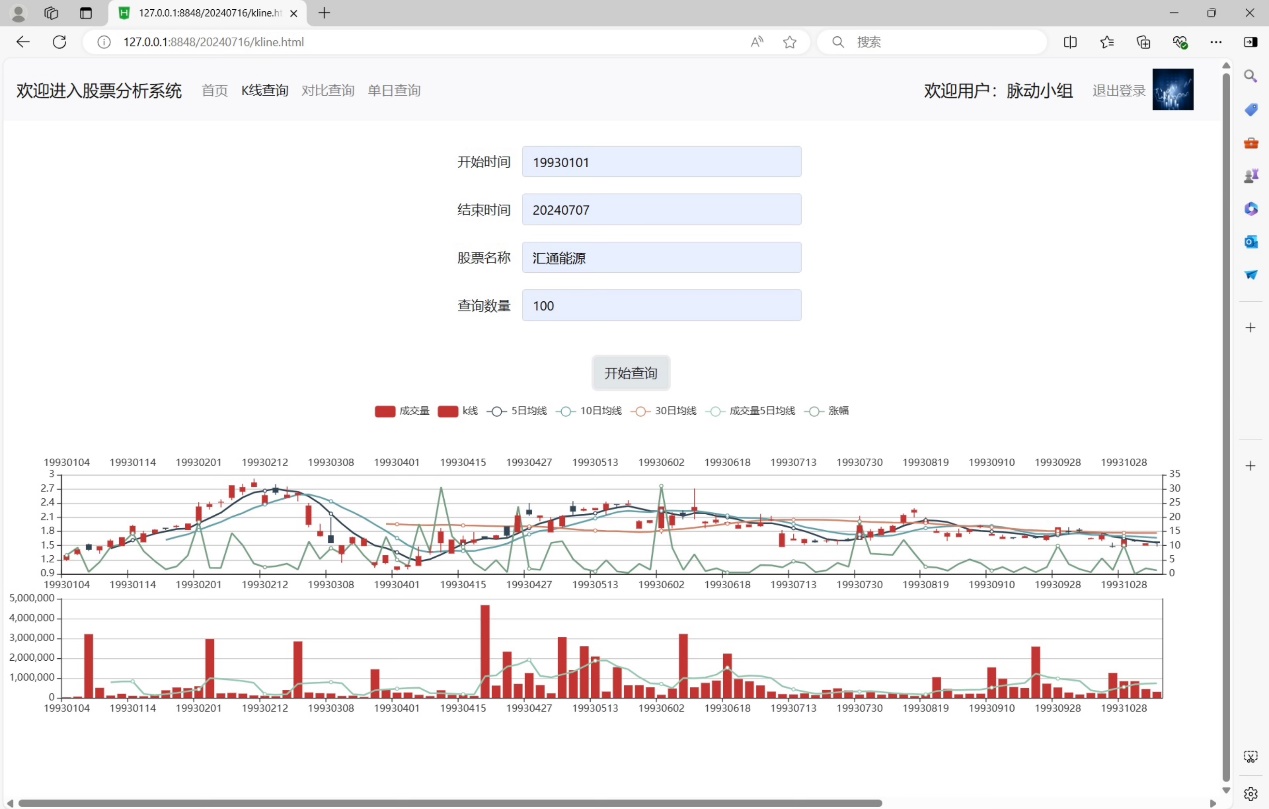


**4.1单只股票的k线查询显示**

用户在输入框输入需要查询股票的开始时间、结束时间、股票名称以及查询数量，点击查询，下方显示两张结果图表：

图1显示该只股票的k线、5日均线、10日均线、30日均线以及涨幅线；

图2显示该只股票的成交量和成交量5日均线。

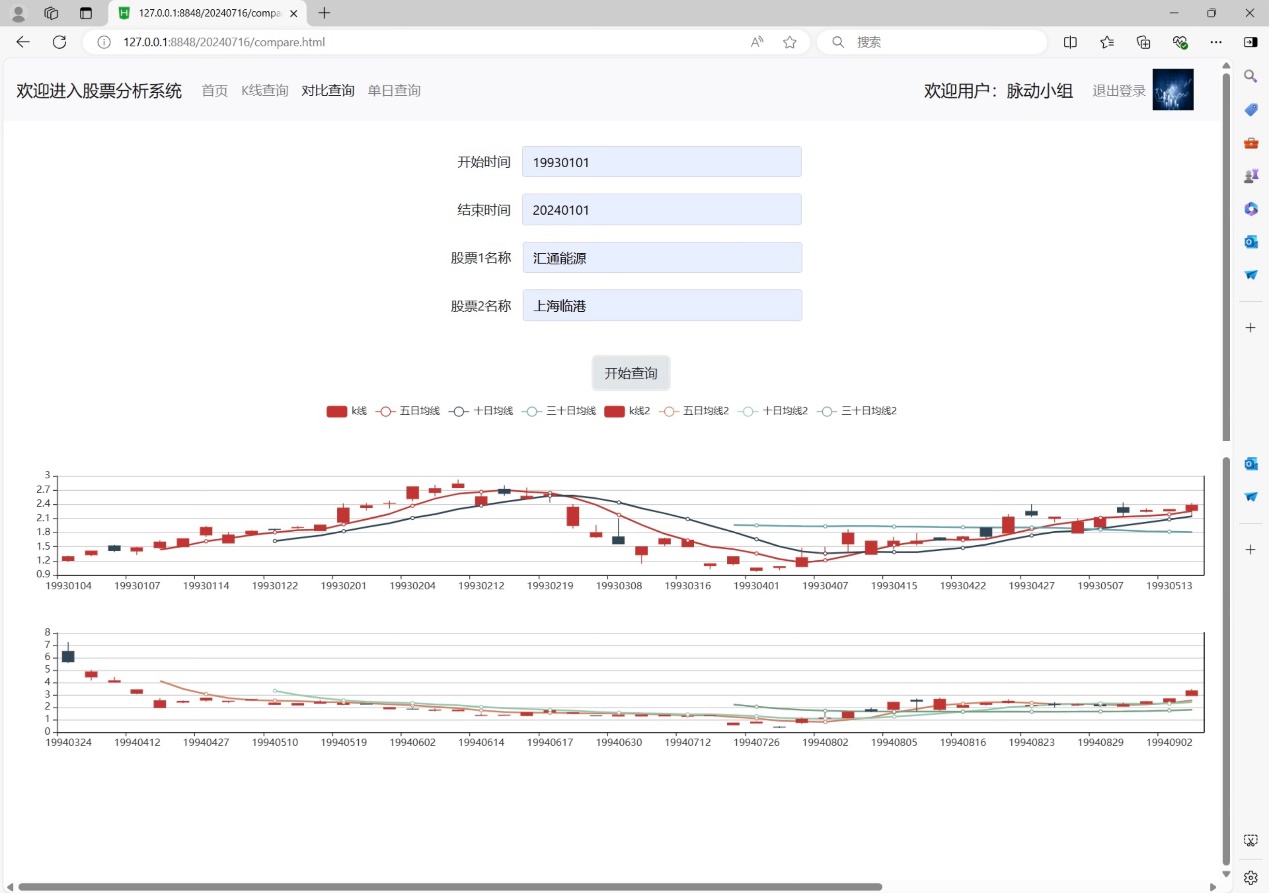
****

**4.2.两只股票的对比查询显示**

用户输入待查询股票的开始时间、结束时间以及两只股票名称，点击查询，下方显示两张结果图表：

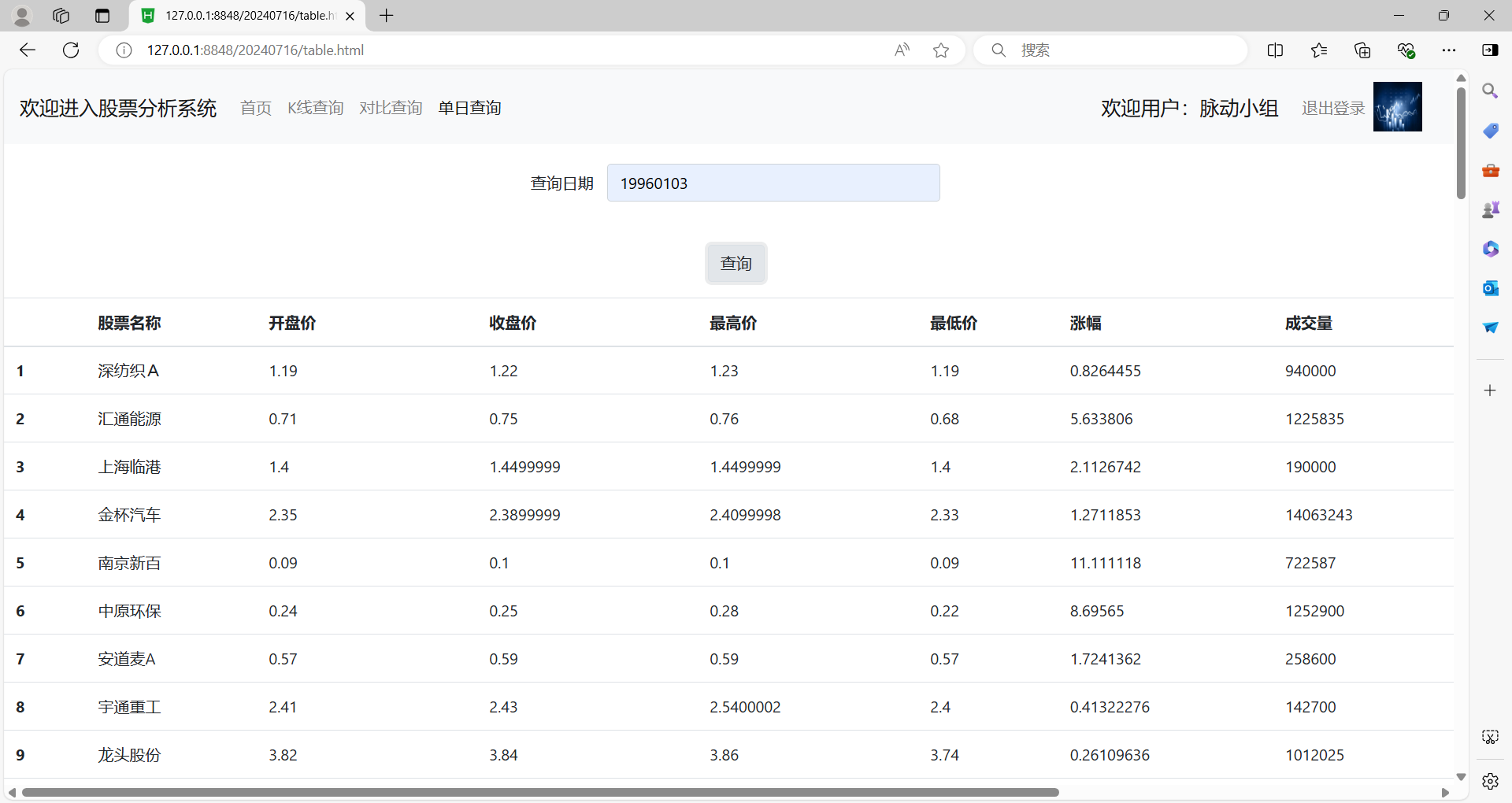
图1为第一只股票的k线、5日均线、10日均线和30日均线；

图2为第二只股票的k线、5日均线、10日均线和30日均线；

****

**4.3单日所有股票查询显示**

用户输入待查询的日期，点击查询按钮，会显示一张结果图表，显示该日的所有股票信息，包括股票名称、开盘价、收盘价、最高价、最低价、涨幅和成交量。

****