

毕业设计(论文)

|  |  |
| --- | --- |
| 课题名称 | 基于B/S架构的在线数据可视化分析平台 |
|  | ——API的设计与实现 |
| 二级学院 | 计算机与软件学院 |
| 专 业 | 软件技术 |
| 班 级 | 软件1851 |
| 学 号 | 1802343134 |
| 姓 名 | 尚宇驰 |
| 指导教师 | 许莫淇 |

年 月 日

计算机与软件学院

毕业设计（论文）诚信承诺

我谨在此承诺：本人所写的毕业论文《基于B/S架构的在线数据可视化分析平台——API的设计与实现》，系本人独立完成，没有抄袭行为，凡涉及其他作者的观点和材料，均作了注释与说明，若有不实，后果由本人承担。

承诺人（签名）：

年 月 日

摘 要

伴随着信息技术与互联网的发展，信息技术对现代社会的渗入可以说是无孔不入。每天各行各业都会产生海量的信息。个人身份信息、交通航线信息、银行各项业务等多种服务都离不开信息技术的支持，其中数据可视化分析平台的建设显得尤为重要，它在整个流程中扮演了幕后英雄的角色。如果缺乏数据可视化平台的幕后支持，所有的硬件和前台界面都不能发挥到应有的功效。随着数据可视化平台的广泛推进，对我国现代各项产业的发展都能起到促进作用。

目前我国正处于数字化转型关键期，面对海量的数据的处理和挖掘这些数据的价值并发挥实际的作用是当今社会的一个重要课题。无论是传统行业还是新型行业，数据越来越被重视；无论是在商业领域还是经济行业，只有基于大量的数据和分析才能产生正确的决策。例如目前的手机应用通过收集用户行为，通过数据分析和算法运算精准为用户推送准确的信息。

本系统以Python flask作为Api端开发的技术框架，以Python作为后端开发语言，借用Python可以全方面处理数据的优势，可以更加得心应手的完成系统的功能，本系统主要完成了用户模块、数据源接入模块、数据预处理模块、数据分析/计算模块和系统维护模块。

关键字：数据可视化、Python、Python flask

Abstract

With the development of information technology and the Internet, the penetration of information technology into modern society can be said to be pervasive. Every day, all walks of life generate massive amounts of information. The development of information technology support, among them, the progress of the construction of the data visualization analysis platform is particularly important, and it plays the role of the hero behind the scenes in the entire process. The effect of. Through the extensive promotion of the data visualization platform, the development of modern modern industries can be promoted.

At present, as a whole, it is in a critical period of digital transformation. Faced with the processing of massive amounts of data, mining the value of these data and playing a practical role is an important issue in today's society. Whether in traditional industries or emerging industries, data is getting more and more attention; for example, existing mobile phone applications collect user behaviors, use data analysis and algorithmic calculations to accurately manipulate accurate information for users.

This system uses Python flask as the technical framework for Api-side development, and Python as the back-end development language. Borrowing the advantages of Python's ability to process data in all aspects, the system can be more handy to complete the functions of the system. This system mainly completes user modules and data sources. Access module, data preprocessing module, data analysis/calculation module and system maintenance module.

Keywords: data visualization, Python, Python flask

目 录

[第1章 绪论 1](#_Toc72269717)

[1.1 课题研究背景与意义 1](#_Toc72269718)

[1.1.1 课题研究背景 1](#_Toc72269719)

[1.1.2 课题研究意义 1](#_Toc72269720)

[1.2 研究目标及内容 1](#_Toc72269721)

[1.2.1 研究目标 1](#_Toc72269722)

[1.2.2 研究内容 1](#_Toc72269723)

[1.3 所做工作 1](#_Toc72269724)

[第2章 开发技术与工具 2](#_Toc72269725)

[2.1 开发技术 2](#_Toc72269726)

[2.1.1 Python Flask 2](#_Toc72269727)

[2.1.2 Python pandas 2](#_Toc72269728)

[2.1.3 PostgreSQL 2](#_Toc72269729)

[2.1.4 MySQL 2](#_Toc72269730)

[2.2 开发工具 2](#_Toc72269731)

[2.2.1 Pycharm 2](#_Toc72269732)

[2.2.2 Navicat 2](#_Toc72269733)

[2.2.3 CentOS 8 2](#_Toc72269734)

[2.2.4 GitHub 2](#_Toc72269735)

[第3章 系统需求分析 3](#_Toc72269736)

[3.1 系统总体需求概述 3](#_Toc72269737)

[3.1.1 系统稳定性分析概述 3](#_Toc72269738)

[3.1.2 产品特点分析概述 3](#_Toc72269739)

[3.2 功能性需求分析 3](#_Toc72269740)

[3.2.1 用户模块 3](#_Toc72269741)

[3.2.2 数据源接入模块 3](#_Toc72269742)

[3.2.3 数据预处理模块 3](#_Toc72269743)

[3.2.4 数据分析/计算模块 3](#_Toc72269744)

[3.2.5 系统维护模块 3](#_Toc72269745)

[第4章 系统概要设计 4](#_Toc72269746)

[4.1 数据库设计 4](#_Toc72269747)

[4.1.1 用户表设计 4](#_Toc72269748)

[4.1.2 用户行为表设计 4](#_Toc72269749)

[4.2 系统日志存储设计 4](#_Toc72269750)

[4.3 系统接口设计 4](#_Toc72269751)

[4.3.1 用户模块接口设计 4](#_Toc72269752)

[4.3.2 数据源接入模块接口设计 4](#_Toc72269753)

[4.3.3 数据预处理模块接口设计 4](#_Toc72269754)

[4.3.4 数据处理/分析模块接口设计 4](#_Toc72269755)

[第5章 系统详细设计 5](#_Toc72269756)

[5.1 用户模块 5](#_Toc72269757)

[5.1.1 GitHub第三方登录功能设计 5](#_Toc72269758)

[5.1.2 记录用户行为功能设计 5](#_Toc72269759)

[5.2 数据源接入模块 5](#_Toc72269760)

[5.2.1 PostgreSQL、MySQL数据库连接功能设计 5](#_Toc72269761)

[5.2.2 Excel、Csv多文件上传功能设计 5](#_Toc72269762)

[5.2.3 查询数据库中所有表功能设计 5](#_Toc72269763)

[5.3 数据预处理模块 5](#_Toc72269764)

[5.3.1 字段分类功能设计 5](#_Toc72269765)

[5.3.2 查询并过滤数据表中所有数据功能设计 5](#_Toc72269766)

[5.3.3 查询数据库中所有表功能设计 5](#_Toc72269767)

[5.3.4 查询指定字段数据功能设计 5](#_Toc72269768)

[5.3.5 自定义构建计算指标功能设计 5](#_Toc72269769)

[5.4 数据处理/分析模块 5](#_Toc72269770)

[5.4.1 数据处理/分析功能设计 5](#_Toc72269771)

[5.4.2 脏数据处理功能设计 5](#_Toc72269772)

[5.5 系统维护模块 5](#_Toc72269773)

[5.5.1 日志记录系统行为功能设计 5](#_Toc72269774)

[5.5.2 统一响应体设计 5](#_Toc72269775)

[第6章 系统功能的实现 1](#_Toc72269776)

[6.1 用户模块 1](#_Toc72269777)

[6.1.1 GitHub第三方登录功能实现 1](#_Toc72269778)

[6.1.2 记录用户行为功能实现 1](#_Toc72269779)

[6.2 数据源接入模块 1](#_Toc72269780)

[6.2.1 PostgreSQL、MySQL数据库连接功能实现 1](#_Toc72269781)

[6.2.2 Excel、Csv多文件上传功能实现 1](#_Toc72269782)

[6.2.3 查询所选连接数据库所有表功能实现 1](#_Toc72269783)

[6.3 数据预处理模块 1](#_Toc72269784)

[6.3.1 字段分类功能实现 1](#_Toc72269785)

[6.3.2 查询并过滤数据表中所有数据功能实现 1](#_Toc72269786)

[6.3.3 查询数据库中所有表功能实现 1](#_Toc72269787)

[6.3.4 自定义构建计算指标功能实现 1](#_Toc72269788)

[6.3.5 指定数据表查询所有字段功能实现 1](#_Toc72269789)

[6.4 数据处理/分析模块 1](#_Toc72269790)

[6.4.1 数据处理/分析功能实现 1](#_Toc72269791)

[6.4.2 脏数据处理功能实现 1](#_Toc72269792)

[6.5 系统维护模块 1](#_Toc72269793)

[6.5.1 日志记录系统行为功能实现 1](#_Toc72269794)

[6.5.2 统一响应体实现 1](#_Toc72269795)

[第7章 系统测试 2](#_Toc72269796)

[7.1 测试环境准备 2](#_Toc72269797)

[7.2 功能性测试 2](#_Toc72269798)

[7.3 系统部署测试 2](#_Toc72269799)

[第8章 总结和展望 3](#_Toc72269800)

[8.1 总结 3](#_Toc72269801)

[8.2 展望 3](#_Toc72269802)

[致谢 4](#_Toc72269803)

[参考文献 5](#_Toc72269804)

# 绪论

近年来，研究人员对数据分析研究的越来越深入，分析的方法也越来越多样化，加之，数据可视化技术的日益成熟，可以借助图形化的方式，直观生动的展现数据的变化趋势，进而分析出数据的发展走向；所以数据可视化平台是当今社会发展必不可少的助力器。本章将先介绍课题研究背景和研究意义，然后明确研究目标叙述研究的内容，最后陈述在准备论文时所准备的工作。

## 课题研究背景与意义

### 课题研究背景

### 课题研究意义

## 研究目标及内容

### 研究目标

### 研究内容

## 所做工作

# 开发技术与工具

调研近年来数据可视化“Data visualization”方向的资料，我们发现Python、R、Ruby是主要流行语言，从学习成本和快捷开发的角度分析，我们选择了Python来实现此项目，并且在众多框架中选择了轻量级的Web框架——flask，来实现搭建，大量数据处理则选用pandas库来完成。相较与传统的MySQL关系型数据库，PostgreSQL标榜自己是世界上最先的开源数据库，它对于集群的支持更加友好；本章将讲述系统所使用的开发技术和开发工具。

## 开发技术

### Python Flask

Python Flask是一个轻量级的Web框架，它的特色为”micro”即“微型”。“微型”并不意味着整个Web应用程序必须适合单个Python文件，也并非Flask缺少功能。微框架意味着Flask核心保持简单但可扩展。它主要在于可以通过添加你所需要的功能对应的库，来为应用程序提供支持，而不是直接给一个大而全的全家桶框架，不管你是否需要。Flask提供许多扩展库，可以满足你的所有需求。

### Python pandas

Python pandas 是一个数据分析处理能力优秀的开源库。它致力于成为应用Python语言，实际的、真实的解决数据分析问题的基础高级构建块。DataFrame对象是一个快速高效的对象，用于带有集成索引的数据操作，包括对数据进行索引、切割、合并、排序、分类聚合等等操作，是pandas处理数据方便快捷的主要优势。它还提供内存数据结构和不同格式之间读取和写入数据的工具，支持CSV和文本文件，Microsoft Excel，SQL数据库和快速HDF5格式。

### PostgreSQL

PostgreSQL是一个功能强大的、开源的、跨平台的对象-关系型数据库系统，它使用和扩展了SQL语言，用于安全地存储数据，和结合了许多功能，可以扩展最复杂的数据工作负载，并且在众多数据库系统中，拥有更强的稳定性。它具有更多细节功能支持，使用起来舒适度更高，更便捷，比如：text字段可以支持所有大小，支持OVER子句，可以存储array和json，并且在之上建立索引，在复制的特性上做到同步，异步，半同步等等。

### MySQL

MySQL数据库是当下较为流行的对象-关系型数据库之一，是开源的 SQL数据库管理系统。数据库是一种根据数据结构来存储和管理数据的仓库，是数据的结构化集合。[1] 想要更好的应用数据库就需要一个合适的数据库管理系统即MySQL，它的数据库是关系型的，区别在于并非将所有数据随意放在大存储室中，而是把数据存在单独的表中，具有对象结构：数据库、表、视图、行和列。SQL则是“结构化查询语言”，是访问数据库的工具，通过SQL语言对数据库中的数据进行一些列操作。

## 开发工具

### Pycharm

Pycharm 是一个专门用于Python语言的、跨平台的、计算机编程集成开发环境，由捷克JetBrains公司开发。它具备代码智能感应、动态检查和快速修复功能，提供一些主流Web框架（Flask、Django）的强大支持，拥有大量的开箱即用工具集，为编程创造了舒适便捷的环境。

### Navicat

Navicat是一套多连接数据库开发工具。它拥有连接七种数据库的能力，致力于让用户专注在数据库设计上，简化其他一切创造数据模型、数据库的步骤，做到了为专业人士提供适合其特殊需要的复杂功能，同时对于新手上手简单。提供图标编辑器，可以有效的创建和设计数据模型，支持多种图标符号，比如UML、Crow’s Foot和IDEF1x 。

### CentOS 8

CentOS 8是CentOS的第8个发行版本，CentOS是Linux的众多发行版本之一，是从RHEL源代码编译的社区重新发布版。

### GitHub

GitHub是一个代码托管云平台。用户可以上传代码、提交版本来管理自己的代码，更出色的是团队开发时，多人协作管理团队代码，通过提交代码、与旧版本代码合并、更新代码库等操作，实现团队代码管理。

# 系统需求分析

对于大量级数据进行处理，会考虑到处理效率和处理时间的问题，另外，数据分类聚合也是一个挑战点，对于以上问题，本章将从系统总体需求和功能性需求两方面展开分析，首先对数据可视化平台的总体需求进行概述，分析产品的特点和系统对于稳定性的需求；其次，对平台所需要的功能性需求进行模块划分，细化每个模块所实现的具体功能。

## 系统总体需求概述

### 系统稳定性分析概述

### 产品特点分析概述

## 功能性需求分析

### 用户模块

1. 用户登录。采用GitHub第三方授权的方式进行登录。
2. 记录用户行为。将用户上传文件、添加数据库表生成记录，写入数据库。

### 数据源接入模块

1. 多数据库连接。支持PostgreSQL和MySQL两种数据库的连接。
2. 多文件上传。支持多文件上传。
3. 多类型文件上传。支持Excel和Csv两种类型文件的多问加上传。
4. 查询数据库中所有表。查询指定库中所有的表名。

### 数据预处理模块

1. 字段分类。将字段分为三类：数值、日期、其他。
2. 查询指定字段所有数据。查询指定字段所有数据。根据传入的连接对象查询所需字段的数据，连接对象中包括数据表和该表中需要查询数据的字段。
3. 自定义构建计算指标。自定义构建计算指标。根据传入需要计算的算式，首先查询算式需要的字段数据，然后根据算式需求和算术符号的优先级，返回算数结果。

### 数据分析/计算模块

1. 数据分类聚合功能。将数据以维度分类，按指标聚合。
2. 脏数据处理。将表中不能解析的空值处理为可解析状态。

### 系统维护模块

1. 记录日志。生成日志，记录系统行为。

# 系统概要设计

根据上文中对系统的需求分析，本章要对系统有整体的概要设计，首先是数据库的设计其中包括系统实现需要使用的数据库表，其次是系统的日志存储，日志存储可以详细的记录出系统运行打印的信息，是日后系统出错时检查问题的重要线索，最后涉及的是系统各模块的接口设计。

## 数据库设计

### 用户表设计

1. 用户实体。用户唯一标识、头像、昵称、角色、是否登录等。
2. 实体属性图如图xx所示。
3. 数据表如图xx所示。

### 用户行为表设计

1. 用户行为实体。行为唯一标识、名称、用户唯一标识、类型、父级编码等。
2. 实体属性图如图xx所示。
3. 数据表如图xx所示。

## 系统日志存储设计

## 系统接口设计

### 用户模块接口设计

### 数据源接入模块接口设计

### 数据预处理模块接口设计

### 数据处理/分析模块接口设计

# 系统详细设计

完成了概要设计就可以进行系统的详细设计，首先是对用户模块的设计，其中包含了用户使用系统需要进行的操作；其次是数据源接入，数据源是整个系统的灵魂，没有数据源整个系统就是一个空壳，正确的接入数据源整个平台才能正常的运作；接着就是数据预处理模块，包括数据的展示、拆分、聚合、构建计算指标等；然后是数据处/分析理模块对数据进行用户指定的处理和分析；最后是系统维护模块，为整个系统的正常运行保驾护航。

## 用户模块

### GitHub第三方登录功能设计

首先在GitHub上配置Home地址和回调地址（当前接口地址），前端先请求GitHub接口，GitHub会返回给后端接口一个code，再将code拼接到GitHub第三方的请求用户授权的接口尾部进行请求，请求成功后，会拿到一个openId，将openId拼接到请求用户信息的接口进行请求，成功后GitHub会返回用户信息。

### 记录用户行为功能设计

将用户创建数据包，添加表行为写入数据库，写入模型如下表所示。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 用户添加表类型 | 行为唯一标识 | 用户唯一标识 | 用户添加表类型 | 名称 | 类型 | 父级编码 |
| 数据包 | 1 | 用户id | 数据包 | 数据包名称 | 1 | 0 |
| 数据库表 | 2 | 用户id | 数据库表 | 数据库表名称 | 2 | 对应数据包的行为唯一标识 |
| 上传文件表 | 3 | 用户id | 上传文件表 | 上传文件表名 | 3 | 对应数据包的行为唯一标识 |
| 自定义表 | 4 | 用户id | 自定义表 | 自定义表名 | 4 | 对应数据包的行为唯一标识 |

## 数据源接入模块

### PostgreSQL、MySQL数据库连接功能设计

### Excel、Csv多文件上传功能设计

### 查询数据库中所有表功能设计

## 数据预处理模块

### 字段分类功能设计

### 查询并过滤数据表中所有数据功能设计

### 查询数据库中所有表功能设计

### 查询指定字段数据功能设计

### 自定义构建计算指标功能设计

## 数据处理/分析模块

### 数据处理/分析功能设计

### 脏数据处理功能设计

表中的空值不能被解析，将空值替换为“null”字符串。

## 系统维护模块

### 日志记录系统行为功能设计

### 统一响应体设计

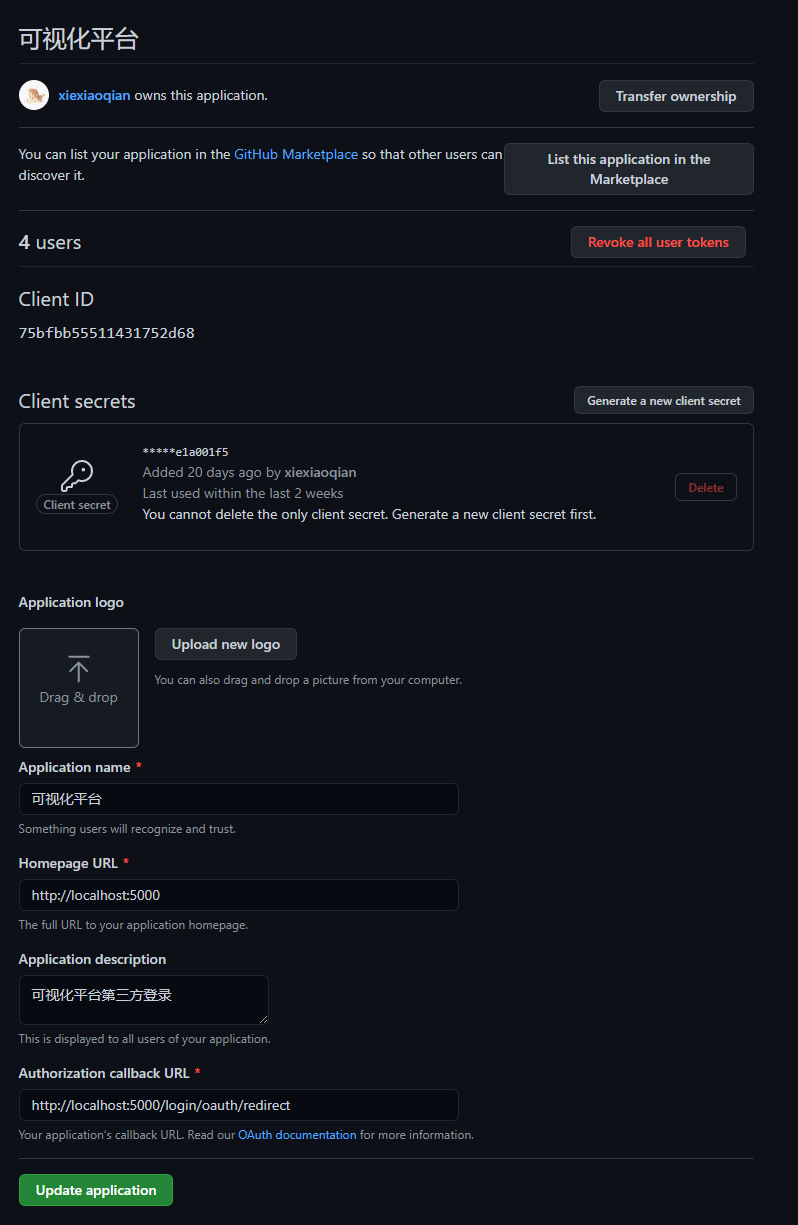
# 系统功能的实现

进行完详细的设计，就要着手实现，本章论述的就是各个模块详细功能的具体实现。

## 用户模块

### GitHub第三方登录功能实现

1. 首先在GitHub上配置平台开放授权，如图xx所示。



1. Home地址：后端Home地址，<http://localhost:5000>。
2. 平台名称：可视化平台。
3. 回调地址：接收openId的接口地址，<http://localhost:5000/login/oauth/redirect>。

配置成功后会生成client\_id和client\_secret。

1. 接口部分
   1. 用request.args.get('code')方法接收到GitHub返回的code
   2. 将client\_id和client\_secret拼接到获取token接口的尾部，作为完整地址即token\_url，设置json请求头，使用requests.post(token\_url, headers=header)方法，请求获取token的接口，方法将会返回请求结果，用res变量接收。
   3. 使用res.json()方法序列化请求结果，以字典的形式拿出access\_token。
   4. 将access\_token作为请求头，使用requests.get(user\_url, headers=headers)方法，请求获取用户信息的接口即user\_url，用res变量接收方法返回的请求结果。
   5. 使用res.json()方法序列化请求结果通过字典拿出用户信息，比如：id、昵称、头像等。

### 记录用户行为功能实现

待实现

## 数据源接入模块

### PostgreSQL、MySQL数据库连接功能实现

### Excel、Csv多文件上传功能实现

### 查询所选连接数据库所有表功能实现

## 数据预处理模块

### 字段分类功能实现

### 查询并过滤数据表中所有数据功能实现

### 查询数据库中所有表功能实现

### 自定义构建计算指标功能实现

### 指定数据表查询所有字段功能实现

## 数据处理/分析模块

### 数据处理/分析功能实现

### 脏数据处理功能实现

## 系统维护模块

### 日志记录系统行为功能实现

### 统一响应体实现

# 系统测试

系统测试是一个系统完成不可或缺的部分，本章从测试环境准备开始论述，然后功能性测试结果的展现，最后是将系统部署到远程服务器上进行测试，只有测试出正常的测试结果，系统的使用的稳定性和安全性才更有保障。

## 测试环境准备

## 功能性测试

## 系统部署测试

# 总结和展望

完成系统测试，整个论文也就接近尾声，本章是对系统的总结和对系统未来发展方向的展望。

## 总结

## 展望

# 致谢

参 考 文 献

[1] 电力公司电力标准化作业管理系统的设计与实现\_马小越[J].