毕业设计说明书

大学生一卡通数据分析系统的设计与实现

|  |  |  |  |
| --- | --- | --- | --- |
| **班 级：** | 14140Y01 | **学 号：** | 1414010513 |
| **姓 名：** | 张丹 | | |
| **学 院：** | 软件学院 | | |
| **专 业：** | 软件工程  （云计算与大数据分析方向） | | |
| **指导教师:** | 曹辉 章永来 | | |

2018年 6 月

大学生一卡通数据分析系统的设计与实现

摘 要

本文主要介绍了大学生一卡通数据分析系统课题的背景、意义、使用技术、需求分析、数据库设计、总体设计、模块设计、实现编码与测试等。整个课题体系分为前台页面概述模块、后台页面管理模块、数据可视化模块、一卡通微信小程序模块，旨在完成利用大数据技术对一卡通数据进行处理和可视化展示，从而使用户能够更加便捷的获取一卡通信息。该课题的创新点体现在利用HDFS技术实现一卡通数据的存储、Hive和MySQL技术实现一卡通数据的统计和分析、Html5展示系统概述、移动端微信一卡通小程序模拟用户行为并与WEB端相结合、K-MEANS聚类算法分析一卡通数据进而得出学生消费水平、等距离离散化一卡通数据得出男女消费习惯、Matplotlib和Echarts进行可视化。

本系统采用WEB与移动微信小程序相结合完成大学生一卡通数据分析系统，使用户能够利用此系统清晰的了解该系统的设计思想和一卡通数据的分析情况，为生活和工作带来便利。用户能够通过WEB查看系统；各个角色的用户可以在登录界面进行登录、可以从性别维度、POS机维度、年级专业班级维度等对数据可视化、修改个人信息；系统管理员可对用户做增删改查操作；该系统实现者对一卡通数据预处理,将一卡通数据提交给Hadoop集群实现存储和分析,并通过不同维度对数据可视化展示。运用大数据和可视化技术能够提高研究一卡通消费数据情况的能力和水平。

关键词：大学生一卡通，大数据，数据分析，可视化

Design and Implementation of College Students

Card Data Analysis System  
 Summary  
 This article mainly introduces the background, significance, use technology, demand analysis, database design, overall design, module design, coding and testing of the college student card data analysis system. The entire topic system is divided into front page overview module, background page management module, data visualization module, and one-card WeChat applet module. It aims to complete the use of big data technology to process and visualize the card data, so that users can more easily obtain a card. information. The innovation of this topic is reflected in the use of HDFS technology to achieve the storage of card data, Hive and MySQL technology to achieve the statistics and analysis of card data, Html5 display system overview, mobile micro-character card program simulation user behavior and WEB-side combination, K The MEANS clustering algorithm analyzes the card data to obtain the student's consumption level, equidistantly discretizes the card data to derive the male and female consumption habits, and visualizes Matplotlib and Echarts.

This system uses WEB and mobile WeChat applets to complete the college students' card data analysis system, so that users can use this system to clearly understand the system's design ideas and analysis of card data, and bring convenience to personal life and work. Users can view the system through WEB; users of each role can log in on the login screen, and can visualize data and modify personal information from the POS dimension, gender dimension, grade professional class dimension, etc. The system administrator can add, delete, and modify the user's operations. The system implementer preprocesses the card data, submits the card data to the Hadoop cluster for storage and analysis, and visualizes the data through different dimensions. The use of big data and visualization technology can improve the ability and level of research on the status of the consumer data card.

**Keywords:** College Student Card, Big Data, Data Analysis, Visualization

目 录

[1 引言 1](#_Toc515218269)

[1．1 课题的提出 1](#_Toc515218270)

[1.1.1 课题的背景和意义 1](#_Toc515218271)

[1.1.2 课题的现状及其发展 2](#_Toc515218272)

[1．2 本课题的主要工作 2](#_Toc515218273)

[1．3 本课题技术支持 3](#_Toc515218274)

[1.3.1 Hadoop技术 3](#_Toc515218275)

[1.3.2 JAVA WEB技术 4](#_Toc515218276)

[1.3.3 K-MEANS算法 4](#_Toc515218277)

[1.3.4 等距离离散化 5](#_Toc515218278)

[1.3.5 数据可视化技术 5](#_Toc515218280)

[1.3.6 微信小程序开发技术 5](#_Toc515218281)

[2 系统需求分析 6](#_Toc515218282)

[2．1 系统业务需求 6](#_Toc515218283)

[2．2 系统功能需求 6](#_Toc515218284)

[3 大学生一卡通数据分析系统设计 7](#_Toc515218285)

[3.1 系统方案选择 7](#_Toc515218286)

[3.1.1 数据库的选择 7](#_Toc515218287)

[3.1.2 WEB框架的选择 7](#_Toc515218288)

[3.1.3 处理数据的选择 7](#_Toc515218289)

[3.1.4 展示数据的选择 7](#_Toc515218290)

[3．2 系统总体结构设计 7](#_Toc515218291)

[3.2.1 在线系统总体设计 7](#_Toc515218292)

[3.2.2 离线系统总体设计 8](#_Toc515218293)

[3．3 在线系统功能模块设计 9](#_Toc515218294)

[3.3.1 前台页面模块 9](#_Toc515218295)

[3.3.2 前台开发技术页面模块 9](#_Toc515218296)

[3.3.3 前台系统预览页面模块 10](#_Toc515218297)

[3.3.4 前台系统登录页面模块 10](#_Toc515218298)

[3.3.5 前台系统关于页面模块 12](#_Toc515218299)

[3.3.6 移动端一卡通小程序模块 12](#_Toc515218300)

[3.4.1 数据存储模块 13](#_Toc515218302)

[3.4.2 数据处理模块 15](#_Toc515218303)

[4．1 系统实现 19](#_Toc515218304)

[4.1.1 在线系统实现 19](#_Toc515218305)

[4.1.2 离线系统实现 19](#_Toc515218306)

[4.1.2 微信一卡通小程序系统实现 21](#_Toc515218307)

[4．2 测试 23](#_Toc515218308)

[4.2.1 错误的来源 23](#_Toc515218309)

[4.2.2 错误的处理 23](#_Toc515218310)

[4.2.3 测试用例 24](#_Toc515218311)

[5 结果分析 25](#_Toc515218312)

[6 结束语 29](#_Toc515218313)

[6．1 设计结论 29](#_Toc515218314)

[6．2 心得体会 29](#_Toc515218315)

[参 考 文 献 30](#_Toc515218316)

[致 谢 32](#_Toc515218318)

1 引言

1．1课题的提出

1.1.1 课题的背景和意义

随着数字化时代的发展，全国各个高校建立了属于自己的校园一卡通系统。校园一卡通具有“一卡走遍校园”的优势，一张卡可以代替师生在学校内的所有个人证件，教师和学生可以在自助终端随时充值、领款、操作个人信息等;可以使用校园卡在可刷卡区域实现餐饮、淋浴、购热水、超市购物、通过图书馆门禁、医疗挂号与缴费等。

根据各项统计与分析，全国各个高校在校大学生人数平均在1万以上，每位学生每天会使用一卡通在校园内进行各类刷卡，进而产生大量的刷卡记录。但是目前各个高校使用的校园一卡通系统只是针对师生的各种业务需求实现了各种应用，缺乏的是对刷卡记录进行有效的分析,比如学校一直以来对校内某一区域的基础建设的建设规划都是以人为的估计与预算作为一个建设基准，然而这样的决策会因为使用一卡通的师生的行为的不确定性导致一些基础建设的建设规模偏大或偏小，无法准确地满足一卡通用户的实际需求。如果实现一个校园大学生一卡通数据分析系统，对数据库系统中存储的刷卡记录进行统计处理，并从一些维度进行分析同时得到结果，就可以对所有一卡通用户的刷卡记录做轨迹分析、推荐及建议，进而为今后学校基础建设、一卡通建设等方面提供参考信息和参照标准。

一卡通产生的消费数据具有数据量大、种类多、产生速度快、价值低等特点，这些都符合大数据的四大特征，日益增长的海量数据基本处于“沉睡”状态， “唤醒”这些“沉睡”的一卡通数据[1-2]，对一卡通的刷卡记录进行挖掘，统计分析大学生的刷卡行为进而分析生活轨迹，结合大数据技术来进行存储处理，利用可视化技术把大数据分析结果呈现出来[3-5]，给高等院校的领导、师生、后勤管理员提供更加直观的决策及管理依据变得越来越重要[5] 。

通过实现校园一卡通数据分析系统，以高校学生的一卡通的消费数据作为该研究的数据基础，采集大量的一卡通消费数据记录，通过大数据存储和分析技术分析大学生行为，发现大学生行为存在的问题，分析其产生的原因，为规范大学生的生活和学习管理行为提供引导和对策，进一步挖掘大数据中存在的有效价值[4]；该数据分析系统对于学校学院管理员和后勤管理员而言，为其决策提供充分的科学依据，使管理更规范、更有针对性，同时提高学校各个机构调整和改善的合理性。

1.1.2 课题的现状及其发展

中国大数据研究表明，在国内很多大学已经在利用大数据技术对校内各种类型数据进行分析并尝试从中获取有价值的分析结果。最早从2012年，福州大学、复旦大学、广东师范大学、天津财经大学等多个学校开始关注和了解大数据的归一化和收集分析，并启用多种数据抽取工具和方法尝试数据分析工作[5]。广东师范大学将学校把在校生的消费记录做统计分析，结合学校的预警平台对学生的圈存记录进行分析，从而利用数据挖掘算法来预测学生的家庭情况，在班级或学院做决策时提供强有力的依据，对于相对困难家庭的学生可以给予一定的帮助。

罗华群[6]提出：将高校各个单位的建设申请和校园一卡通系统的消费数据相结合来审核学校助学金和贷款发放工作，比单纯由各院系收集学生们写的贫困申请会更加合理和高效，也会更客观[7-9];在此决策中，陈建兵[10]也提出相关观点，并利用一卡通系统数据辅助辨别贫困生[7-9];金培莉等人[11]提出从学校后勤管理等部门以餐饮人数、月均消费情况和淋浴情况3组数据分析来支撑学校决策等[7-9]。

2007-2009年Romero&Ventura和Baker&Yacef[12]提出要对10余年教育行业的数据做分析，并且使用数据分析方法归纳出五大类大数据挖掘和分析的方法:统计分析与可视化、聚类、预测、关系挖掘、文本挖掘，多种多样的新技术为基础的数据分析方法促进了大数据的不断向前发展[5,13]。

1．2 本课题的主要工作

(1)本课题的主要工作是设计与实现校园一卡通数据分析系统，对一卡通数据进行处理、存储和分析，通过查阅文献资料和相关课题的技术研究，确定系统实现的实现技术。系统的特点体现在利用HDFS技术实现一卡通数据的存储、Hive和MySQL技术实现一卡通数据的统计和分析、Html5展示系统概述、移动端微信一卡通小程序模拟用户行为并与WEB端相结合、K-MEANS聚类算法分析一卡通数据进而得出学生消费水平、等距离离散化一卡通数据得出男女消费习惯、Matplotlib和Echarts进行可视化。

1. 明确系统要实现的功能和模块。采用B/S模式、微信小程序完成校园一卡通数据分析系统，使用户能够利用此系统清晰的了解该系统的设计思想和一卡通数据的分析情况，为个人生活和工作带来便利。系统应提供如下功能：用户能够通过WEB查看系统；各个角色的用户可以在登录界面进行登录、可以从时间维度、消费类型维度、POS机维度、性别维度、消费水平维度、年级专业班级维度等对数据可视化、修改个人信息；系统管理员可对用户做增删改查操作；该系统实现者对一卡通数据预处理,将一卡通数据提交给Hadoop集群实现存储和分析,并通过不同维度对数据可视化展示。
2. 完成系统的编码工作并进行软件测试。

1．3 本课题技术支持

1.3.1 Hadoop技术

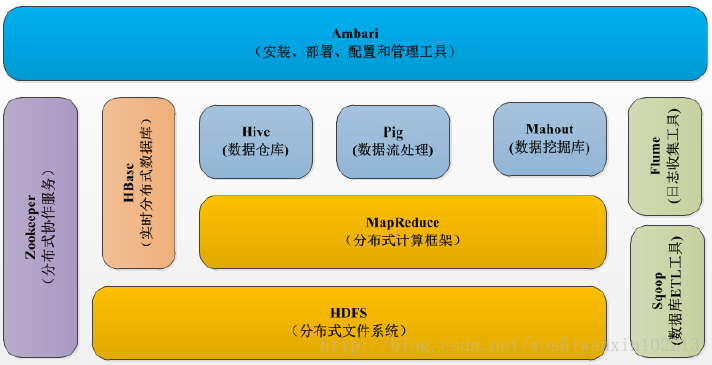


图1.1 Hadoop生态圈

HDFS（Hadoop分布式文件存储系统）:Hadoop生态体系中存储和管理数据的基础。Hadoop生态圈的文件存储系统是一个容错度很高的系统，能够检验和发现文件存储问题，可以在价格较低的普通硬件系统上运行。

MapReduce（分布式离线计算框架）:MapReuce体现了的思想是分而治之,可以进行PB级别海量数据的计算和处理,具有良好伸缩性和容错性。分为Map阶段和Reduce阶段，第一阶段对海量数据进行操作，生成键值对形式的中间结果。第二阶段对第一阶段生成中间结果操作，相同Key的所有Value进行聚合处理，得到需要的数据。

Hive（Hadoop数据仓库）:是大数据中的数据存储仓库，核心数据处理框架是MapReduce，Hive有内置数据库可以进行元数据管理,主要应用于数据的离线处理。Hive还提供了数据仓库功能:数据抽取转化加载,面向主题分析,海量数据的存储，数据查询。利用SQL查询语言，简单容易入手，一定程度上省去了书写代码的过程。

HBase（NoSQL分布式列存数据库）:传统的数据库模型已不能满足一些特定场景的需要,NoSQL的出现很好的解决了传统数据库不能解决的一些问题。HBase可以在大量数据中实现并发读写操作,HBase使存储有了伸缩性，列可以动态的增加，自动切分数据，HBase中的数据使用MapReduce离线计算框架来处理。

Sqoop（数据同步工具）:Sqoop是可以在关系型数据库和Hadoop之间传送数据的工具。例如:可以把MySQL中的数据传送到HDFS,也可以将HDFS上的数据传送到MySQL。

1.3.2 JAVA WEB技术

Servlet是Java Servlet的简称，是服务器端程序，编程语言为Java。Servlet的功能在于前台发送请求后台接受请求，可做一定操作，这一过程即交互式，生成动态Web内容。

1.3.3 K-MEANS算法

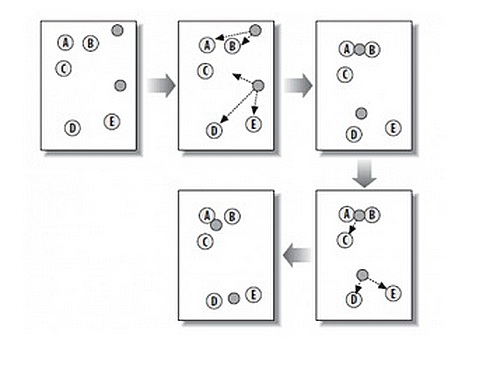
K-MEANS算法是聚类分析算法中比较简单和容易使用的算法，主要用来解决数据的聚集问题，适用于处理大量的数据并且算法高效,主要思想是在划分多个中心点，不断取距离中心点最近值的算法。算法的核心在于计算中心点的公式，不同求中心点的算法会有差异,在本系统中可以将距离中心点近的多点归为一类，可以直观的看出一卡通数据的聚类效果。

图1.2 K-MEANS算法示例图

1.3.4 等距离离散化

离散化在数据分析中是将连续型数据进行分段。等距离离散化正是分段的一种原则，此方法是将连续型数据分为距离相等的“段”，即将连续变量均匀划分成n等份并且每一份的间距相等，进而分析多个段之间的关系。

1.3.5 数据可视化技术

ECharts，一个可视化库，利用JS语言编码实现，特点是开源，在 PC端和移动设备上都可以流畅运行，兼容目前大部分浏览器（IE8/9/10/11，谷歌，火狐，苹果端Safiri等），用户可以定制满足自己需求的可视化图表。

1.3.6 微信小程序开发技术

微信用户群体的大量增加，需要微信开发来支持大量的APP应用。小程序是近些年人们热烈谈论的技术，小程序的面世解决传统APP需要下载并且占用手机内存的情况，用户可以通过扫描二维码直接进入APP应用。相比于服务号，小程序开发成本低，易于传播和简单易用。小程序可以个性化自己的服务，简洁的交互页面，提高了用户体验。

2 系统需求分析

2．1 系统业务需求

校园一卡通涉及到很多方面的事情。对学校高层决策者来说，对数据存储与管理分析，对多年采集的高达几十GB的庞大数据进行存储与管理决策提供依据；对于后勤管理员来说，对消费数据分析，使各经营单位（超市、食堂）尽量做到供求偏差大、有挤压现象减少；对消费时间分析，查看性别与消费时间点关系，调整窗口饭菜口味以提高收益；关注POS机收益情况，调整窗口卫生情况及饭菜质量；对于学院管理员来说，对消费数据分析，审核学生的家庭情况。对于餐饮消费位于中上等水平的同学，在每年国家助学金进行补助时可以不予考虑或较后考虑，而优先考虑那些平时餐饮费用处于中低消费水平的同学。对学生来说，根据一卡通消费情况的分析可以调整自己的消费时间点，避免就餐和淋浴拥挤；学生需要能及时展示当前的一卡通消费和余额情况,结合当前的一卡通状况改善消费状况；

随着移动端的快速发展,PC端的一卡通信息查询已不能满足用户的需求，人们需要实时、便捷的在移动端查询一卡通数据，系统需要增加移动端的一卡通信息查询功能。

2．2 系统功能需求

为了用户能更加便捷的使用大学生一卡通数据分析系统，系统应提供以下功能：

(1)系统介绍：在WEB前台对大学生一卡通数据分析系统和微信小程序、系统预览、系统使用和开发技术进行概述,使用户可以清晰地了解该系统。

(2)登录：学生用户、后勤管理员、学校或学院管理员、系统管理员都可以登录一卡通后台管理系统。

(3)管理功能：上述用户通过登录页面进入后台管理,可以修改个人密码，系统管理员可对普通用户做增删改查操作。

(4)数据处理功能：系统实现者能够对数据进行清洗，转换成符合系统业务格式的数据，利用大数据存储和计算框架对业务数据存储和分析。

(5)可视化展示：利用可视化技术可以直观、友好地展示数据。

(6)微信小程序：用户登录后查看近10次的消费和余额消费，模拟消费和充值。

3 大学生一卡通数据分析系统设计

3.1 系统方案选择

3.1.1 数据库的选择

本系统采用MySQL数据库。相比于ORACLE数据库，MySQL数据库对于个人是免费的,其可以处理大约80%的业务需求，可以同时处理大量的用户发送的请求，处理高达百万的数据,命令执行速度快,对开发者来说，简单有效。

3.1.2 WEB框架的选择

本系统WEB框架选择MVC设计模式。能够进行项目的快速迭代开发，开发软件的时间成本会明细降低，可维护性较高，能够适应项目需求的变动，三层架构思想中的业务层和展示层合理分离，耦合度较低，使得项目需求变动的风险降到最低，可以便捷的进行多次开发,为项目积累开发经验和技术。

3.1.3 处理数据的选择

本系统处理数据选择Hive。Hive底层使用MapReduce计算模型，不需要编写MR代码，即可对数据进行处理。**Hive主要应用于离线数据处理,原始数据能够保存在Hive上，经过Hive处理后的数据也可以直接存储的在Hive上,Hive的数据存储在HDFS文件系统上。**

3.1.4 展示数据的选择

本系统处理数据选择Echars.js。Echarts是百度开源的可视化技术,文件体积小，可以满足用户的使用需求，Echarts已为用户打包好库文件，用户下载后可自由选择需要的图表和组件。其中的图表具有良好的交互性体验,能够满足很多用户的不同需求，在移动端也有自适应效果,Echarts能够满足一卡通数据可视化的需求。

3．2 系统总体结构设计

3.2.1 在线系统总体设计

在线系统模块分为:

WEB端:

(1)系统介绍:对该系统进行概述

(2)开发技术:对该系统的开发技术进行介绍

(3)系统预览:对该系统的功能进行可视化

(4)系统登录:跳转到登录页面

(5)系统关于:对该系统的使用进行说明

移动端:

1. 学生登录个人信息页面，查看近10次消费和余额情况
2. 在线模拟消费和充值行为

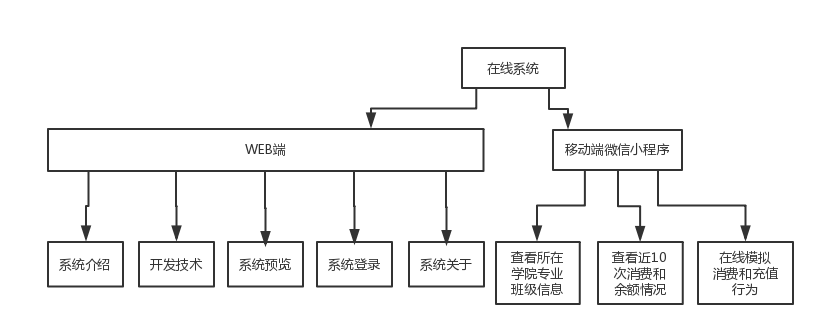


图3.1 在线系统示意图

3.2.2 离线系统总体设计

离线系统是系统实现者的操作,包含一卡通数据存储、一卡通数据预处理、一卡通数据分析、一卡通数据可视化展示等是系统最核心的部分,有一些操作在web上进行,还有一些离线操作在Mac、Linux进行。

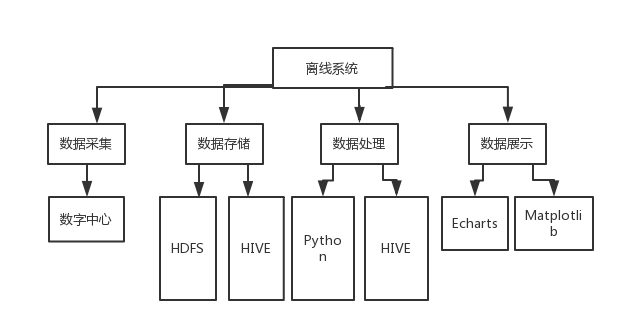


图3.2 离线系统示意图

3．3 在线系统功能模块设计

3.3.1 前台页面模块

前台设计页面包含:系统介绍、开发技术、系统登录、系统预览、系统关于。



图3.3 系统介绍页面

3.3.2 前台开发技术页面模块



图3.4 开发技术页面

3.3.3 前台系统预览页面模块

传统的一卡通信息通过一些表格,文件方式浏览一卡通信息,具有缺乏直观性、灵活性等特点。现在的数据展示方式是通过灵活生动的图表生动的展示出数据,提升了用户体验，满足不同数据的处理需求，为人们的工作和生活提供便捷。系统采用Echars.js进行可视化展示，同时也有传统的表格展示一卡通信息，两者对比，更能体会到数据可视化的优点。



图3.5 系统预览页面

3.3.4 前台系统登录页面模块

登录角色分为学生、后勤（餐饮、淋浴、热水）管理员、学院管理员和系统管理员。



图3.6 登录页面

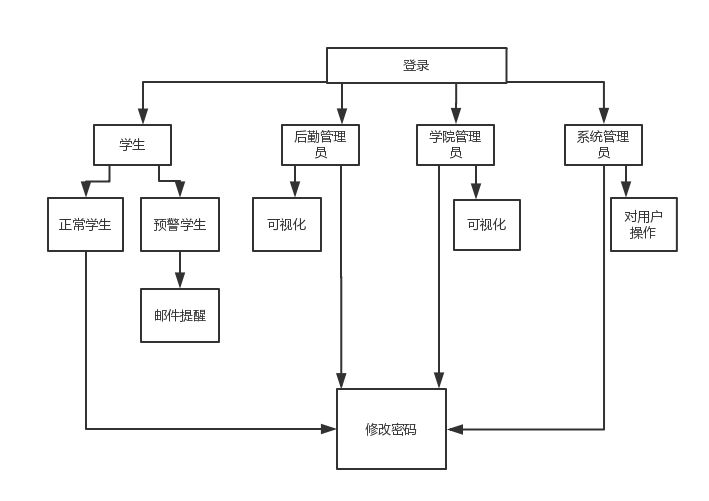


图3.7 登录模块图

3.3.5 前台系统关于页面模块

图3.8系统关于页面

3.3.6 移动端一卡通小程序模块

目前，移动端用户的数量在迅猛增加，学生需要在移动端实时查询一卡通消费和余额信息并进行充值来便利个人的生活。采用微信小程序实现移动端的一卡通模块，并使用一卡通小程序实时查询一卡通状况并对个人生活做一些改善，同时学生可以在线消费、充值。



图3.9 一卡通小程序个人信息页面

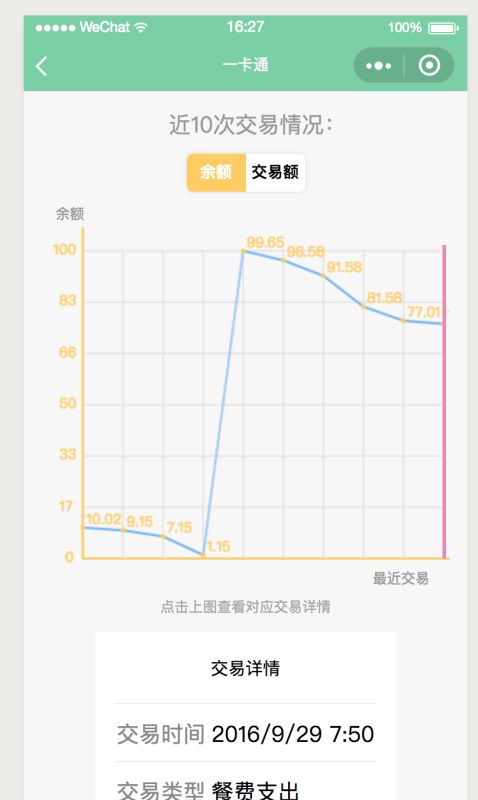


图3.10 一卡通小程序余额和交易详情页面

3．4 离线系统功能模块设计

3.4.1 数据存储模块

一卡通数据存储在Mac本地的MySQL,Linux中的Hive和HDFS中,数据的导入是通过Sqoop将一卡通数据导入Hive,以便一卡通信息的处理和分析。

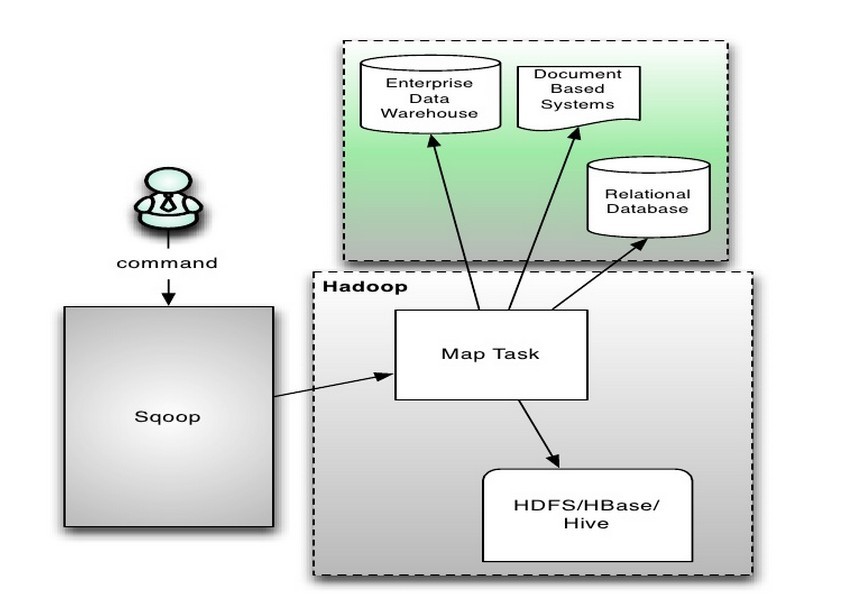


图3.11 Sqoop架构图

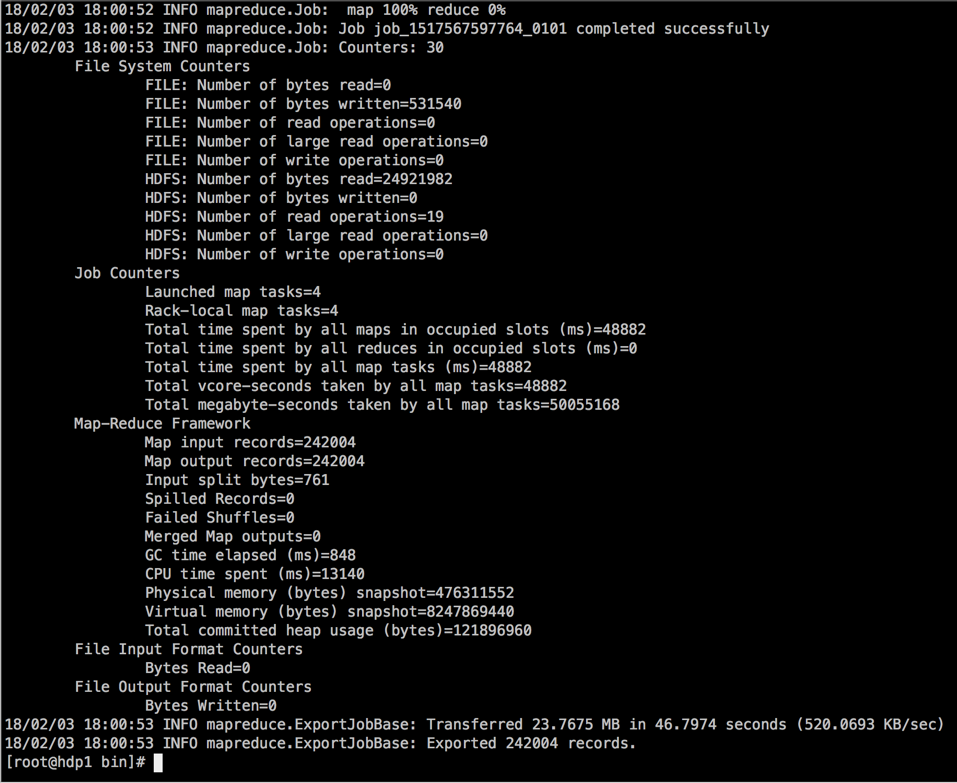


图3.12 Sqoop工具将HIVE数据导入MySQL成功图

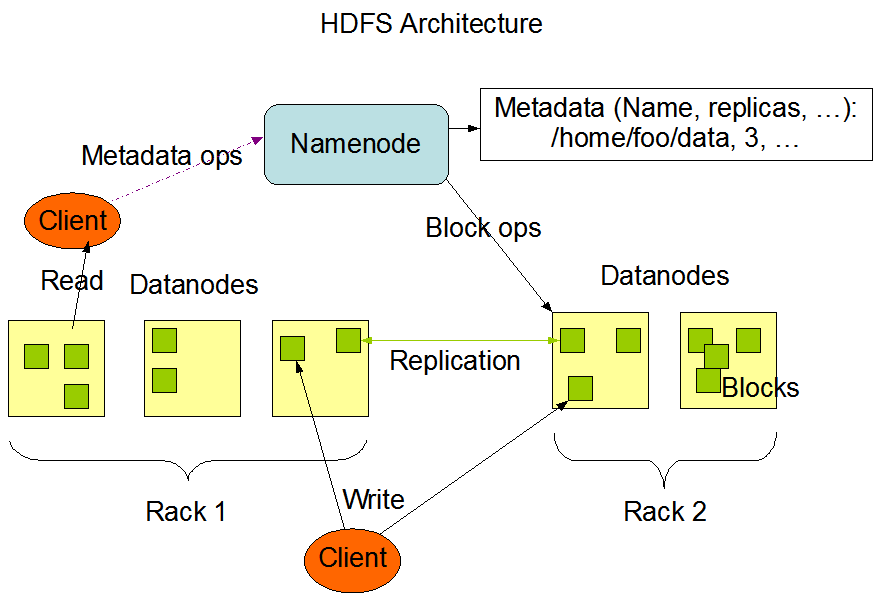


图3.13 HDFS架构图

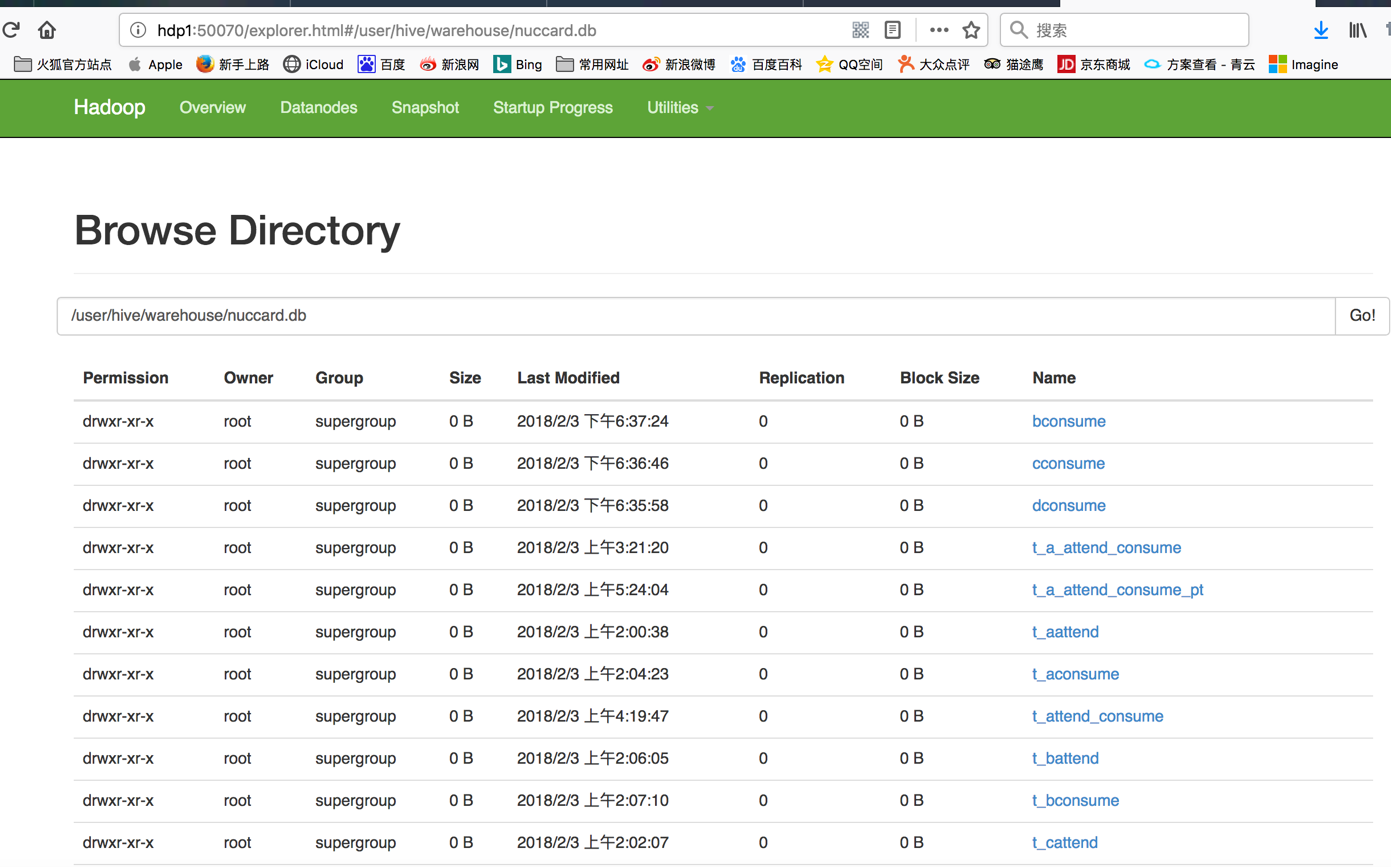


图3.14 HDFS存储结果

3.4.2 数据处理模块

系统使用MapReduce、Java、Python、MySQL、Hive处理数据。从学校数据中心获取的数据不是直接可用的信息，需要经过处理转化为有价值的数据。本系统处理数据的方式有三种:

1. 使用HDFS的API通过java代码将本地一卡通数据导入HDFS。
2. 使用K-means聚类算法分析一卡通数据,使用Python的Matplotlib库进行可视化展示。
3. 使用等距离离散化算法对消费时间处理，分析消费数量与消费性别的关系。
4. 使用Hive的HQL语言对HDFS中存储的一卡通数据进行统计分析，根据需求利用Sqoop工具将数据在Hive和MySQL互相传输，通过Echarts技术进行可视化展示。

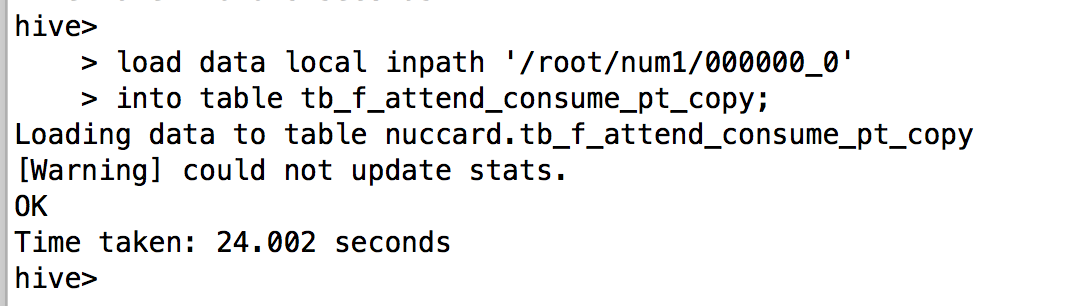


图3.15 Hive加载数据图

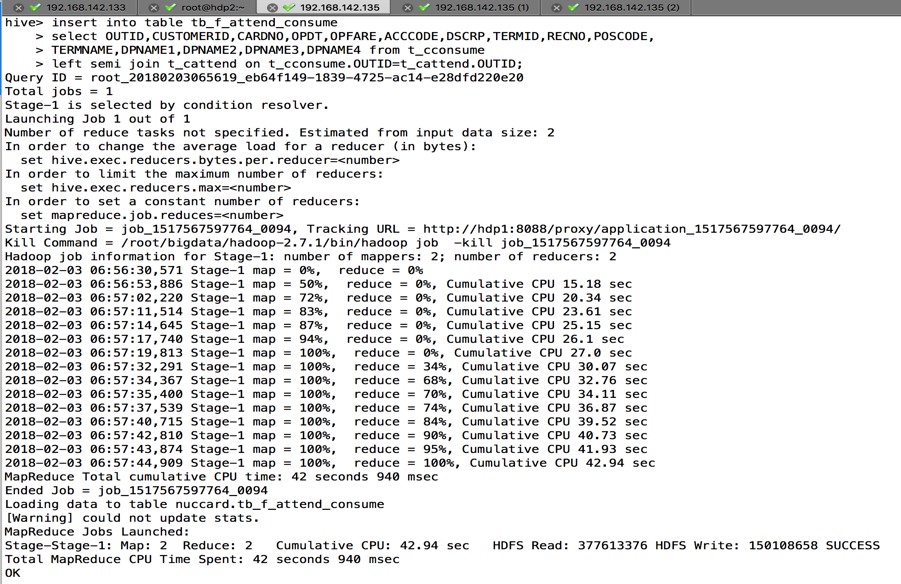


图3.16 Hive处理数据过程图

3．5 数据库设计

数据库设计了用户(card\_user)、角色(card\_role)、聚类分析(k\_means)、

正常学生一学期消费(tb\_n\_attend\_consume\_date)、

微信用户消费(tb\_n\_attend\_consume\_date\_wx)等数据表。

表3.1 数据库card\_user表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名称 | 属性类型 | 额外设置 | 是否空 | 描述 |
| id | bigint(20) | auto\_increment | NO | 主键ID |
| nickName | varchar(255) | | YES | 微信名 |
| userCode | varchar(15) |  | YES | 用户编码 |
| userName | varchar(15) |  | YES | 用户名称 |
| userPassword | varchar(15) |  | YES | 用户密码 |
| status | int(11) |  | YES | 登录状态 |
| gender | int(10) |  | YES | 性别 |
| address | varchar(30) |  | YES | 地址 |
| userRole | bigint(20) |  | YES | 用户角色 |
| createdBy | bigint(20) |  | YES | 创建者（userId） |
| creationDate | datetime |  | YES | 创建时间 |
| modifyBy | bigint(20) |  | YES | 更新者（userId） |
| modifyDate | datetime |  | YES | 更新时间 |

表3.2 数据库k\_means表

|  |  |  |  |
| --- | --- | --- | --- |
| 属性名称 | 属性类型 | 是否为空 | 描述 |
| outid | int(11) | NO | 学号ID |
| count | int(11) | YES | 消费总次数 |
| OPFARE\_SUM | double(11,0) | YES | 消费总金额 |

表3.3 数据库card\_role表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 属性名称 | 属性类型 | 额外设置 | 是否为空 | 描述 |
| id | bigint(20) | auto\_increment | NO | 主键ID |
| roleCode | varchar(15) | | YES | 角色编码 |
| roleName | varchar(15) | | YES | 角色名称 |
| createdBy | bigint(20) |  | YES | 创建者 |
| creationDate | datetime |  | YES | 创建时间 |
| modifyBy | bigint(20) |  | YES | 修改者 |
| modifyDate | datetime |  | YES | 修改时间 |

表3.4数据库tb\_n\_attend\_consume\_date表

|  |  |  |  |
| --- | --- | --- | --- |
| 属性名称 | 属性类型 | 是否为空 | 描述 |
| OUTID | int(11) | YES | 学号ID |
| OPDT | varchar(255) | YES | 消费日期 |
| ACCCODE | int(11) | YES | 消费类型编码 |
| DSCRP | varchar(255) | YES | 消费类型 |
| TERMNAME | varchar(11) | YES | 消费POS机编码 |
| OPFAR | float | YES | 消费金额 |
| DPNAME1 | varchar(255) | YES | 年级 |
| DPNAME2 | varchar(255) | YES | 学院 |
| DPNAME3 | varchar(255) | YES | 专业 |
| DPNAME4 | varchar(255) | YES | 班级 |

s

表3.5 数据库tb\_n\_attend\_consume\_date\_wx表

|  |  |  |  |
| --- | --- | --- | --- |
| 属性名称 | 属性类型 | 是否为空 | 描述 |
| OUTID | int(11) | YES | 学号ID |
| OPEDT | varchar(255) | YES | 消费日期 |
| ACCCODE | int(11) | YES | 消费类型编码 |
| DSCRP | varchar(255) | YES | 消费类型 |
| TERMNAME | varchar(11) | YES | 消费POS机编码 |
| income | float | YES | 充值金额 |
| OPFAR | float | YES | 消费金额 |
| balance | float | YES | 余额 |
| DPNAME1 | varchar(255) | YES | 年级 |
| DPNAME2 | varchar(255) | YES | 学院 |
| DPNAME3 | varchar(255) | YES | 专业 |
| DPNAME4 | varchar(255) | YES | 班级 |

4 系统实现及测试

4．1 系统实现

4.1.1 在线系统实现

在线系统实现WEB端前台和后台的编码工作，下面是在线系统的核心实现代码,分块实现了在线系统首页的功能。

<div class="menu">

<ul class="nav navbar">

<li><a href="#work" class="scroll">Analyze</a></li>

<li><a href="#projects" class="scroll">Visualize</a></li>

</ul>

</div>

<div class="col-md-4 portfolio-grids">

<a href="images/g8.jpg" class="swipebox" title=" ">

<img src="images/g8.jpg" class="img-responsive" />

<div class="figcaption">

<h3>放大<span> 查看</span></h3>

<p>个人每人消费K图展示.</p>

</div>

</a>

</div>

4.1.2 离线系统实现

离线系统实现数据获取，数据存储，数据处理的编码工作，下面是离线系统的核心实现代码,HDFS的API实现存储，HIVE实现了一卡通数据处理的功能，Echarts实现可视化，Java的Mail实现邮件预警功能。

// HDFS的API实现存储，跟HDFS建立上连接，要知道NameNode的地址即可

@Test

//打开本地文件系统的一个文件作为输入流

Configuration conf = new Configuration();

conf.set("fs.defaultFS", "hdfs://hdp1:9000");

FileSystem fileSystem = FileSystem.get(conf);

InputStream in = new FileInputStream("/Users/Zd/Desktop/20180205大数据/aAttend.txt");

//使用hdfs的fileSystem打开一个输出流

FSDataOutputStream out = fileSystem.create(new Path("/local\_hdfs"));

IOUtils.copyBytes(in, out, 1024, true);

//关闭fileSystem连接

fileSystem.close();

<h1>个人每日支出分析K图演示</h1>

$.ajax({

type: 'GET',

url: '${pageContext.request.contextPath}/nordaydyshow',

dataType: 'json',

success: function (data) {

mapOnlyKey8.length = 0;

mapKeyValue8.length = 0;

for (var i = 0; i < data.length; i++) {

mapOnlyKey8.push(data[i].dt);

var mapOnlyValue8 = [data[i].min, data[i].max, data[i].ave, data[i].opfare];

mapKeyValue8.push(mapOnlyValue8);}

tooltip: {

trigger: 'axis',

formatter: function (p) {

var res = p [0].seriesName + ' ' + p [0].name;

res += '<br/> 最低 : ' + p [0].value[0];

res += '<br/> 最高 : ' + p [0].value[1];

res += '<br/> 平均 : ' + p [0].value[2];

res += '<br/> 总价 : ' + p [0].value[3];

return res;}

},

series: [{

name: 'opfare',

type: 'k',

data: mapKeyValue8},

]};

myChart8.setOption(option8);

var lastIndex = 0;

var len = option8.series[0].data.length;

clearInterval(timeTicket);

timeTicket=setInterval(function () {

lastIndex += 1;

myChart8.addData([

[0, // 系列索引

option8.series[0].data[lastIndex % len], // 新增数据

false, // 新增数据是否从队列头部插入

false,

option8.xAxis[0].data[lastIndex % len]]

]);}, 2000);}});

</script>

4.1.2 微信一卡通小程序系统实现

微信一卡通实现后台查询消费和余额信息，并实现在线模拟消费和充值行为。下面是微信一卡通小程序的核心实现代码,scan方法实现了一卡通充值行为。

scan: function () {

var \_this=this;

wx.scanCode({

success: function (res) {

console.log(res)

var opfare = res.result;

console.log(opfare);

wx.request({

url: "http://www.ocpro.cc/NUC-CARD/add\_opfare?opfare=" + opfare,

data: {

income: \_this.data.income,

userCode: \_this.data.userCode,

time: new Date(),

openid: getApp().globalData.openid, },

method: "POST",

success: function (res) {

console.log(res.data[0].balance)

var rs=res.data[0].balance-opfare

var t=res.data[0].opedt

if (rs>0||rs==0){

wx.showModal({

title: '消费成功',

content: +opfare + "元",}),

\_this.setData({

balance:\_this.data.balance-opfare,

last\_time:t}),

getApp().globalData.balance = \_this.data.balance;}

if(rs<0){wx.showModal({

title: '失败',

content: '余额不足！',})

}},

});

}})},

4．2 测试

软件测试是软件开发过程中一个必不可少的组成部分，测试的目的是为了确保程序的正常运行并能及时发现程序运行问题的过程，软件测试可以编写测试用例和报告，发现编码开发中没有找到的错误以及性能问题，提升开发质量。完成测试后需要进行实际上线调试，检验其是否能够适应运营的要求，在上线后还需要根据软件的运行状况,不断的去发现软件可能出现的问题,保证项目的正常运行和问题的防范。

4.2.1 错误的来源

错误的来源有多种，每种错误所产生的影响的大小也各不相同，常见错误来源如下：

(1)数据缺失：可能是在后台数据查询时，数据并没有返回，或者数据发生错误而没有显示。

(2)数据不符合要求：在一些需要输入信息的页面,字段不符合要求，可能会发生错误。

(3)编码错误：不同系统，可能对编码的要求是不一样的，页面和后台的编码不同，会发生乱码。

(4)权限错误：如果没有考虑系统的权限问题，不具备相关授权的用户访问未授权的资源，会给系统造成安全隐患。

(5)浏览器兼容:浏览器内核对不同的显示页面效果支持是不一样的,系统完成编码，还需要再不同浏览器上测试。

4.2.2 错误的处理

(1)注意数据展示的格式和时间类型,有显示格式不正确的,找到正确的格式。

(2)在登录、主页、后台管理页面使用Js客户端校验，规定输入数据格式。

(3)把所有的页面,文件，数据库格式设置为统一的编码格式。

(4)设置使用系统的用户能够有相应的权限访问对应的资源，不具备条件的权限不能访问，可以使用权限框架处理或者流程控制来实现。

(5)页面在不同的浏览中测试,出现显示不正确的页面，调整为自适应,使系统可以在各个浏览器友好显示页面。

4.2.3 测试用例

表4.1 微信一卡通小程序测试用例

|  |  |
| --- | --- |
| 测试用例名称 | 小程序个人信息查询测试 |
| 测试流程 | 1,输入学号  2,输入密码 |
| 预期结果 | 1,信息输入正确,且与个人微信名匹配，可以查询个人信息  2,信息输入学号或密码错误或与绑定微信不一致,需要重新输入 |
| 实际结果 | 可以查询个人信息 |
| 结论 | 实际结果与预期结果一致,测试通过 |

表4.2 数据库乱码测试用例

|  |  |
| --- | --- |
| 测试用例名称 | 数据库乱码测试 |
| 测试流程 | 输入带有中文的添加用户信息 |
| 预期结果 | 1,数据库用户信息没有出现乱码  2,数据库用户信息出现乱码 |
| 实际结果 | 数据库没有出现中文乱码 |
| 结论 | 实际结果与预期结果一致,测试通过 |

5 结果分析



图5.1　系统管理员查看所有用户信息

图5.1中的数据是一卡通用户的数据。系统管理员可以对用户做增删改查操作。查询框可以根据用户角色查询用户信息。



图5.2　用户操作个人信息

图5.2中的信息是一卡通用户可以修改的信息。系统用户可以对个人的登录密码做修改操作提高安全性。

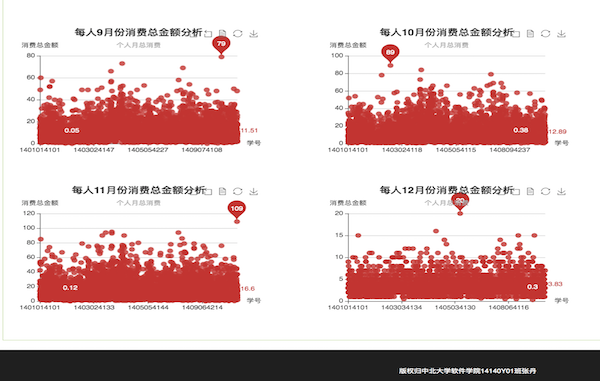


图5.3 14级所有学生9-12月份淋浴收益可视化散点图

图5.3中的数据是14级每位同学在9-12月份的淋浴总消费，使用Hive对原始数据按学号、月进行统计分析，可以分析得出9-12月每位同学消费的总金额、最低金额、最高金额、月平均金额。

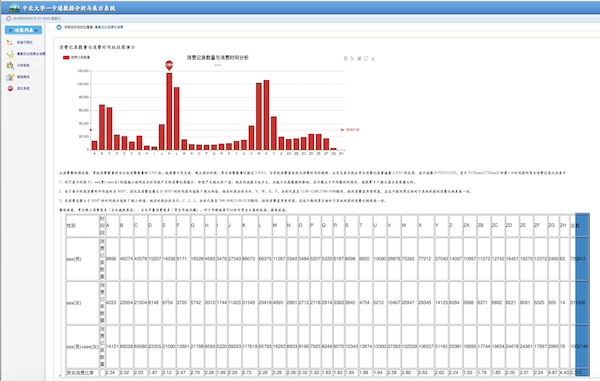


图5.4 14级所有学生消费性别与消费数量可视化柱状图

图5.4中的数据是14级每位同学在9-12月份的淋浴总消费，使用Hive对原始数据按性别、消费数量进行统计分析，可以分析得出整体来看，男生晚上消费更多，女生早晨消费更多。



图5.5 每个学院一学期三大消费类型可视化堆积面积图

图5.5中的数据是学校每个学院每个专业餐饮消费、淋浴消费、购热水消费的数据，使用Hive对原始数据按学院和消费类型进行统计分析，可以分析得出每个学院每个专业一学期中每一消费类型的最高消费、最低消费。

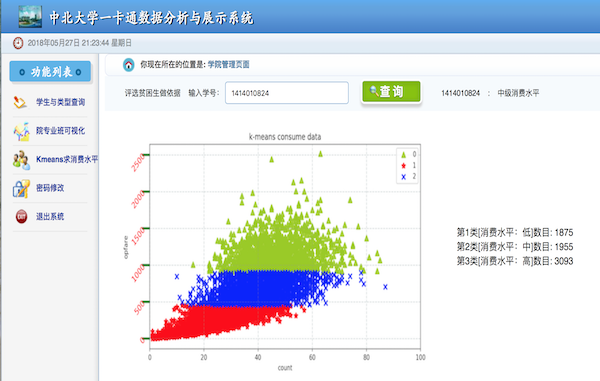


图5.6 14级学生消费水平聚类分析图

图5.6中的数据是14级学生一学期总消费的数据，使用K-means聚类算法对原始数据按学生、总消费和总消费次数进行聚类分析，可以分析得出三个消费水平，同时可以得出每位学生具体所处的消费水平。



图5.7 学院管理员查看预警学生信息展示图

图5.7中的数据是学院管理员查询预警学生的个人信息图，进而通过和学生交流关注学生动态、提高工作效率，。

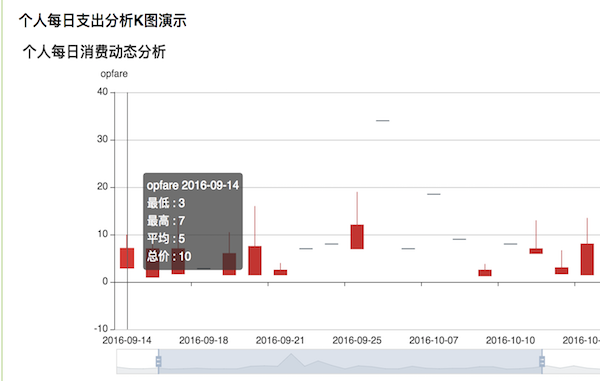


图5.8 个人每日消费情况可视化K线图

图5.8中的数据是每位同学一个学期每日消费的数据，使用Hive对原始数据按日进行统计分析，利用K线图可以分析得出每位同学每天的最高消费金额、最低消费金额、平均消费金额、总消费金额。

6 结束语

6．1 设计结论

大学生一卡通数据分析系统完成了从前端页面设计到后台开发、后台用户的管理、数据采集、存储、处理、可视化到微信一卡通小程序，具备一定的应用性。

由于时间以及技术等各方面因素，该系统还存在一些不足。其中数据处理部分由于没有得到全面数据，比如说性别，使部分分析不准确；在进行可视化展示时，大部分可视化效果采用静态数据的方式展示,少许采用了动态方式展示，系统灵活性不高;在实现编码的过程中,有一定量的代码冗余,没有全部优化代码，易修改性不高;把部分消费数据放在服务器上时，在数据加载时还存在一定的延时问题；以上不足在今后开发中值得进一步提升和改进。

6．2 心得体会

《大学生一卡通数据分析系统的设计与实现》毕业设计即将完成，我在做毕业设计期间学习和收获了很多知识。这次的毕业设计是在保研学校实习期间完成的，在此实习期间，不断学习系统使用的技术和经验，我掌握了前端开发技术、WEB开发技术和大数据技术，在实现该系统的过程也更深入地了解了实际软件开发的工作流程和软件工程思想。

在搭建Hadoop集群分布式的环境中遇到很多问题，比如Mac系统配置静态IP使项目开发有些停滞，之后通过CSDN、博客园等博客上看到别人的配置过程,慢慢地配置好网络模式，在大三学过Linux的基础上，快速搭好了集群分布式环境,安装了Hive、Sqoop、HBase、Spark等框架；在学习前端开发时也了解到页面的灵活性，在大二JSP课程中学习过Html、Css基础技术，这一学习过程更加了解了JSP标准标签库JSTL的使用；此外， WEB开发技术SSH和SSM框架也花费了一定时间，在大三时，学习过Strus2技术，MVC思想对于学习SSM框架是很有帮助的。毕业设计对我来说是一次提高自身技术实力和书写文档规范的机会，能够积极主动的去学习毕业设计用到的技术,遇到问题迎难而上,使自己的编程能力有了很大的提升。

通过这次毕业设计我学习到了很多编程技术和经验，培养了自己处理项目问题的能力，学习编程技术应该从原理、架构、性能方面着手，同时要与同学、老师多交流,查阅论文和文献资料,不断地充实和提升自己。

**参 考 文 献**

[1]朱光婷. 大数据环境下 C2C 模式中消费者行为研究[D]. 上海工程技术大学，2015:11-15.

[2] 志程,杨鹏,王宝会. 面向大数据的数据管理平台架构研究与应用[J]. 网络新媒体技术，2015,4(2):22-93.

[3]刘志龙. 校园一卡通数据分析系统的设计与实现[D]. 华东师范大学, 2007.

[4]段群, 弋改珍, 韩丽娜. 利用大数据分析大学生消费行为的方案设计[J]. 现代计算机, 2017(12):64-66.

[5]王国琼. 大数据可视化对某高校学生行为分析的呈现[D]. 山东大学, 2016.

[6]杨媛媛. 社交网络中的关系构建和亲密度分析[J ]. 石家庄学院学报, 2009, 11(6): 91-96. YANG Y Y. An analysis of relationship formation and intimacy in social network[J]. Journal of Shijiazhuang University, 2009, 11(6): 91-96.

[7]金培莉, 王晓震, JINPei-li,等. 校园一卡通系统决策支持实例分[J]. 华东师范大学学报(自然科学版), 2015, 2015(s1):525-529.

[8]金培莉, 王晓震. 校园一卡通系统决策支持实例分析[C]// 中国高等教育学会教育信息化分会第十二次学术年会论文集. 2014.

[9]罗华群. 校园一卡通数据的挖掘与应用[D]. 华东师范大学, 2009.

[10]吴昊, 刘东苏. 社交网络中的好友推荐方法研究[J]. 现代图书情报技术, 2015, 31(1): 59-65. WU H, LIU D S. Research on the method of friends’recommendation in social network[J]. New Technology of Library and Information Service, 2015, 31(1): 59-65.

[11] 张淑娟. 大学生心理健康状况与应对措施[J]. 中医药管理杂志, 2016(3): 30-31. ZHANG S J. The mental health of college students and the countermeasures[J]. Journal of Traditional Chinese Medicine Management, 2016(3): 30-31.

[12]Romero，C.，Ventura,S.，&Garcia,E.(2005).Data mining illcourse management systems:Moodlecasestudyandtutorial.Computers&Education,51(1)，36 8—384.

[13]程玉霞. 数据挖掘在学习成绩预测中的应用研究[D]. 江南大学, 2016.

**致 谢**

感谢我的导师章永来老师，在完成毕业设计和毕业论文的过程中,章老师给予了我很多指导，章老师辛勤操劳、耐心细致的辅导学生的精神深深地感染着我，在章老师的帮助下，我收获和学习到了很多东西。同时因为有导师辛勤的工作本毕业设计和论文才能顺利进行，在这里我向章老师表示衷心的感谢！

感谢所有给予我支持的老师和同学！感谢14140Y01班！

感谢使我不断成长的中北大学和软件学院！