****

**本科生毕业论文（设计）**

题 目： 广告推送后台系统设计

院 系： 数据科学与计算机学院

专 业： 计算机系网络工程

学生姓名： 陆荣志

学 号： 14348090

指导教师： 周凡教授

时间：2018 年 2 月 8 日

**表一：毕业论文（设计）开题报告**

**Form 1: Research Proposal of Graduation Thesis (Design)**

|  |
| --- |
| 论文（设计）题目  Thesis (Design) Title： |
| （简述选题的目的、思路、方法、相关支持条件及进度安排等）  （Please briefly state the research objective, research methodology, research procedure and research schedule in this part.）  Student Signature： Date: |
| 指导教师意见  Comments from Supervisor：    1.同意开题 2.修改后开题 3.重新开题  1.Approved( ) 2. Approved after Revision ( ) 3. Disapproved( )    Supervisor Signature： Date: |

**表二：毕业论文（设计）过程检查情况记录表**

**Form 2：Process Check-up Form**

|  |  |
| --- | --- |
| **指导教师分阶段检查论文的进展情况（要求过程检查记录不少于3次）**  **The supervisor should check up the working process for the thesis（design）and fill up the following check-up log. At least three times of the check-up should be done and kept on the log.**  **第1次检查（First Check-up）:**  学生总结  Student Self-summary：  指导教师意见  Comments of Supervisor：  **第2次检查（Second Check-up）：**  学生总结  Student Self-summary：  指导教师意见  Comments of Supervisor：  **第3次检查（Third Check-up）：**  学生总结  Student Self-summary：  指导教师意见  Comments of Supervisor：  **第4次检查**  **Fourth Check-up**  学生总结  Student Self-summary：  指导教师意见（Comments of Supervisor）：  学生签名（Student Signature）： 日期（Date）:  指导教师签名（Supervisor Signature）： 日期（Date）: | |
| **总体完成情况**  **（Overall Assessment）** | 指导教师意见Comments of Supervisor：  1、按计划完成，完成情况优（Excellent）： （ ）  2、按计划完成，完成情况良（Good）： （ ）  3、基本按计划完成，完成情况合格（Fair）：（ ）  4、完成情况不合格（Poor）： （ ）  指导教师签名（Supervisor Signature）：  日期（Date）: |

**表三：毕业论文（设计）答辩情况登记表**

**Form 3: Thesis Defense Performance Form**

|  |  |  |  |
| --- | --- | --- | --- |
| **答辩人 Student Name** |  | **专 业**  **Major** |  |
| **论文（设计）题目**  **Thesis（Design） Title** |  | | |
| **答辩小组成员**  **Committee Members** |  | | |
| 答辩记录  Records of Defense Performance：  记录人签名（Clerk Signature）： 日期（Date）: | | | |

**学术诚信声明**

本人所呈交的毕业论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。本毕业论文的知识产权归属于培养单位。本人完全意识到本声明的法律结果由本人承担。

本人签名： 日期：

**Statement of Academic Integrity**

I hereby acknowledge that the thesis submitted is a product of my own independent research under the supervision of my supervisor, and that all the data, statistics, pictures and materials are reliable and trustworthy, and that all the previous research and sources are appropriately marked in the thesis, and that the intellectual property of the thesis belongs to the school. I am fully aware of the legal effect of this statement.

Student Signature： Date：

**说 明**

1. 毕业论文（设计）的写作格式要求请参照《中山大学本科生毕业论文的有关规定》和《中山大学本科生毕业论文（设计）写作与印制规范》。

2. 除完成毕业论文（设计）外，还须填写三份表格：

（1）表一 毕业论文（设计）开题报告；

（2）表二 毕业论文（设计）过程检查情况记录表；

（3）表三 毕业论文（设计）答辩情况。

3. 上述表格均可从教务部主页的“下载中心”处下载，如表格篇幅不够，可另附纸。每份毕业论文（设计）定稿装订时应随同附上这三份表格。

4. 封三是毕业论文（设计）成绩评定的主要依据，请认真填写。

**Instruction**

1. Please refer to ‘*The Guidelines to Undergraduate Graduation Thesis (Design) at Sun Yat-sen University*’and ‘*The Writing and Printing Format of Undergraduate Graduation Thesis(Design) at Sun Yat-sen University* ’ for anything about the thesis format.

2. Three forms should be filled up before the submission of the thesis (design)：

（1）Form 1: Research Proposal of Graduation Thesis.

（2）Form 2: Process Check-up Form.

（3）Form 3: Thesis Defense Performance Form.

3. All the above forms could be downloaded on the website of the Office of Education Administration. If there is not enough space in the form, please add extra sheets. Each thesis (design) should be submitted together with the three forms.

4. The form on the inside back cover is the grading sheet. Please fill it up before submission.

**【摘要】** 本系统为广告推送后台系统，提供广告获取、广告定点投放、广告管理功能。广告推送范围根据标签系统来实现，通过与用户画像后台系统的相结合，根据用户标签进行精准定位用户。广告精准投放对提高广告效率、增加用户体验以及实现企业收入增加有着积极作用。广告内容包括但不限于文字、静态图、动态图、视频、图片组等，广告内容可伸缩拓展自行定义。广告管理包括广告增加、删除、修改，可由管理员手动进行以及系统定时清理过期广告。系统提供可视化界面供管理员使用广告管理功能，页面内可进行广告查看、修改、增加功能，则进行广告管理不需要技术背景。该系统设计到的技术有：WEB服务器搭建与维护、SSM框架、断点续传、并发控制、数据库缓存、用户画像等。该广告系统将会应用在数字家庭多业务平台中作为实践。

**【关键词】** 广告管理；广告定点投送；用户画像

目录

[1. 引言 12](#_Toc510735458)

[1.1. 背景和意义 12](#_Toc510735459)

[1.2. 设计的目的和意义 13](#_Toc510735460)

[1.3. 本文工作 14](#_Toc510735461)

[1.4. 国内外研究现状 14](#_Toc510735462)

[1.5. 论文结构简介 14](#_Toc510735463)

[2．功能需求分析 15](#_Toc510735464)

[2.1. 系统功能 15](#_Toc510735465)

[2.2. 系统运行流程 17](#_Toc510735466)

[2.2.1. 广告获取 17](#_Toc510735467)

[2.2.2. 广告管理 18](#_Toc510735468)

[3. 模块设计与技术应用 19](#_Toc510735469)

[3.1. 前端 19](#_Toc510735470)

[3.1.1. jQuery框架 19](#_Toc510735471)

[3.1.2. Ajax技术 20](#_Toc510735472)

[3.2. 后台系统 21](#_Toc510735473)

[3.2.1 MVC设计思想 21](#_Toc510735474)

[3.2.2. SSM框架 22](#_Toc510735475)

[3.2.3. 系统设计 23](#_Toc510735476)

[3.2.4. 数据库设计 25](#_Toc510735477)

[3.2.5. 文件缓存 29](#_Toc510735478)

[参考文献 29](#_Toc510735479)

# 引言

## 背景和意义

广告系统是互联网产业中一个很重要的产业链。在很多企业中广告系统提供的收入占了很重的一个比例。截至2016年12月，我国网民规模达7.31亿，普及率达到53.2%[1]。互联网的覆盖面是越来越广泛，则互联网广告的市场也随即扩大。2017年我国互联网广告市场规模一直处于高速增长的态势，预计2018年整体规模有望突破4000亿元。在这么庞大的一个市场下，设计一个有效的广告管理系统十分有意义。并且，如何做到高效率地投放广告成为企业的一个重要目标。广告的投放如果随意、泛滥，那么会影响用户体验，严重会导致用户流失。而广告精准定位，有以下好处：提升广告效率，基于大数据精准定向，能够在广告和用户体验中找到较好的平衡；对用户而言，从主动获取广告信息，变为被动需求信息的获取。

本毕业设计来源于数字家庭业务实验中心的广告投放系统设计。数字家庭业务应用是一个注重家庭智慧健康的项目，项目致力于将家庭和互联网科技联系起来，打造一个为家庭服务的数字家庭系统。项目包括一些家庭常用设备的家庭技术支持，如电视机、门口“魔镜”、移动设备（手机）。这些技术支持也就是传统意义上的客户端系统，换而言之，这些系统都是可以作为广告投放的载体的。

互联网广告已经发展成一个至关重要的产业链，其带来的收益对一个互联网产品来说是不可缺少。互联网产品中，盈利的方式整体来说，主要分为下面两种方式：一个是产品应用中提供的付费服务；而另一个就是通过广告投放，从厂商中获得巨额的收益。在一些付费业务几乎不存在的互联网应用中，就必须要拥有一个可靠的广告投放管理系统，来维持该应用团队。我所在的项目组也正是属于上述类型。

互联网广告不同于传统广告。传统的广告包括海报、电视广告这些面向的用户，都是未经过筛选的，而且是未知的用户对象。互联网广告投放和传统广告的最大不同就是，它不仅拥有庞大的用户群体，而且互联网广告面向的用户几乎是透明的。这个得益于用户的互联网身份大多数都固定的，所以互联网应用中获取到的用户信息比传统的广告投放系统要多得多。根据这些用户信息，所进行的广告投放可以更加精准、有效。

总而言之，一个能够提供广告管理功能，并且容易维护、对使用者友好，并且能够精准定位到目标用户群体的广告系统，是互联网应用中不可缺少的一部分。

## 设计的目的和意义

由上一章节可知，互联网广告的特性是面向的用户群体庞大以及透明度，所以，本系统设计的目标在于设计出一个稳定可用的广告后台系统。系统的主要功能有三个：一个是广告管理功能；第二个是广告自动精准投放功能；还有就是用户客户端获取广告功能。

广告管理功能包括：增加广告、修改广告信息、删除广告、查询广告内容以及统计广告信息。广告管理功能由管理员进行，管理员的专业知识背景要求不高，为了友好性，广告管理系统提供了容易管理的界面功能。

广告自动精准投放功能是基于用户画像系统进行的。用户画像是一组根据互联网应用收集到的一些用户习惯，如用户常访问的页面等等，来使用这些用户信息进行数据分析，预测出来的用户模型。用户画像系统为广告精准投放提供了一个技术支持。建立出的用户画像准确性能够直接影响到广告投放的准确性。由于该原因，本论文设计出来的广告系统为一个与用户画像系统相关联，但是不依赖于某一个特定的用户画像系统的独立系统。广告系统提供相应接口给用户画像系统，来维护一个用户画像模型。而广告系统则根据这个模型来进行定点投放。广告系统是一个独立的系统，脱离了用户画像系统之后，仍可以进行其他两个功能：广告管理和广告获取。

广告获取功能则是一个由目标用户客户端发起的一个功能。广告系统面向的客户端具有松耦合性。则不限制于客户端的类型。无论是常见的App，又或者是我所在项目组的数字家庭应用的电视平台、网页平台、“魔镜”系统均可使用该系统。出于这个目的，广告投放设计为目标用户主动发起请求进行获取。

## 本文工作

使用Apache Tomcat搭建Web服务器，选择使用Spring + SpringMVC + Mybatis作为开发框架。数据库使用MySQL与Mybatis集成作为数据持久层。前端界面采用基于jQuery的AmazeUI框架进行开发。前端提供广告管理界面。对于精准投放的大数据处理，使用Hadoop + Spark 的分布式系统进行开发。精准投放是基于其他系统生成的用户画像来进行的。

## 1.4. 国内外研究现状

国外对精准广告投放研究有成就的有以下几位。2010年Pak Alexander N 提出一种维基百科与语境广告匹配的方法，提高了广告的投放精度。该研究是一种新的机遇维基百科的文章作为“参考点”的广告选择。2011年，Hof，Robert D提出了一种机遇社交网站Facebook的社交图谱的数据，挖掘出网络用户的兴趣，并根据其推荐相应的广告，完成精准投放。由于Facebook的大数据支持，对其进行数据挖掘，能够将广告准确定位到相对应的兴趣和位置的用户。2013年，台湾的陈艳秋和谢慧清提出的采用神经网络算法改进了O2O（Online To Offline）商务模式中广告投放的精准度。O2O指的是指将线下的商务机会与互联网结合，让互联网成为线下交易的平台。把广告精准投放给目标用户之后，才会有更多的用户在网上支付购买商品，并且在实际的线下得到对应的商品。

国内的研究有以下几位。2013年，郭新宇、张荣提出一种基于用户搜索行为的潜在语义的用户分割方法。2014年，张晓阳提出基于cookie的精准广告投放技术，根据用户的cookie行为信息进行精准广告投放。投放精度提高的同时，必须对用户隐私进行保护。[2]

## 1.5. 论文结构简介

本文章分为以下几个模块来进行阐述说明本人的广告后台系统：系统设计需求分析，系统各模块介绍与设计，系统效果展示，技术难点与解决方案。

1. 系统需求分析。分析系统所需要的功能，分析现有的技术，选取最适合实现这些需求的技术。
2. 系统各模块介绍。这一模块将会展示系统各个功能的实现方案、UML时序图，以及整个大系统的UML用例图、类图。介绍整个系统是怎样运行的，数据流动等。
3. 效果展示。效果展示从两个角度进行：一个是客户端的角度，在客户端上能够正确获取到目标广告；一个是管理员的角度，管理员能够进行广告管理并且得到正确的反馈；
4. 系统的验证与测试。系统的可用性的衡量有两个，一个是系统的正确性，一个是系统的可用性。正确性在于广告能够正确投放，客户端能够获取得到目标广告，广告管理能够实现其功能。可用性要求系统要稳定高可用，并且能够应对一定的高并发。在开发完成之后，会对系统做压力测试，并且把测试结果写进报告里面，并且对测试中暴露出来的问题作出改进。
5. 遇到的困难以及解决方案。该部分会列出在编码或者设计的时候遇到的难点，以及自己解决的思路和方案。
6. 总结与展望。总结系统的优缺点，提出自己的不足，以及未来进一步扩展的展望。

# 2．功能需求分析

本设计的广告管理后台系统的目标需求主要有三个：管理员对于后台中广告的管理、目标客户端对广告的主动获取以及广告的精准投放。

## 2.1. 系统功能

整个系统的功能用例图如图1所示。能够完整作为一个发布的广告模块包括两个部分：用于展示广告的目标应用的用户客户端和用于后台管理的管理员系统。

用户客户端和后台系统是松耦合的，只需要提供能够展示广告的接口，如一片广告区域来展示图片广告，又或者是展示文字亦可。广告后台系统接口具有通用性。在我所在的实验室的数字家庭应用中，客户端包括手机端App，电视端应用以及网页端应用，这些不同的客户端是可能展示广告的效果不同，但是其从后台获取广告的方式是一样的。都可以通过HTTP请求从后台服务器获取一定格式广告文件，然后再由客户端负责进行展示。



**图1 广告系统用例图**

用户客户端系统不属于项目内容。用户客户端只需要提供用户的身份标识即可得到用户的投放广告列表，然后再根据投放列表获取相应的广告文件即可。

而另一个模块——后台管理系统目的在于方便管理员高效地进行广告操作：增加、删除、查询以及修改。增加广告包括添加广告信息以及添加广告文件。广告信息包括：广告名字、点击广告后跳转的URL、广告类型、广告投放时间段、广告目标人群、广告有效时间。查询广告可以通过广告信息中某一个点来进行模糊查询。修改广告内容则可以修改广告信息和广告文件。

同时，广告后台模块拥有自管理功能，能够定期清理过期的广告，让系统更加人性化。

## 2.2. 系统运行流程

### 2.2.1. 广告获取

广告获取由客户端发起。当客户端满足以下条件之一时，发起广告更新请求：

* 广告有效时间已到
* 客户端界面切换到广告界面
* 达到定时获取广告的时间



**图2 客户端获取广告流程**

其中获取广告流程为：先根据用户身份标识给广告服务器发送HTTP请求获取广告信息列表。广告信息列表包括了后台系统经过分析之后，给该用户定制的广告的除广告文件之外的所有信息，包括广告投放时间段、广告类型、名字等；然后根据广告列表，可能客户端本地会做一些处理，来决定自己应该展示什么广告；然后客户端把自己要展示的广告ID发送给后台服务器，来下载广告文件，用于前端广告展示。整体系统的时序图如图2所示。



**图3 获取广告时序图**

### 2.2.2. 广告管理

该功能属于后台功能。广告的管理有两种：一种是管理员进行添加、删除和修改操作；另一种是系统定期检查更新，剔除过期的广告。广告管理需要进行数据持久化，来保存更新过的广告信息。

广告管理流程见图3所示。首先由管理员登录到广告后台管理页面，进行广告信息的编写与资源的上传。后台系统得到新修改的广告数据之后，先将广告信息持久化到数据库，然后根据广告的用户标签列表的变化，作出相应修改：如果广告中的用户标签有增加/删除，则在该标签的所有用户的广告列表中增加/删除该广告ID。然后将修改后的用户和广告的对应关系更新到数据库。

**图4 广告管理时序图**

# 3. 模块设计与技术应用

## 3.1. 前端

广告管理系统前端的职责是建立一个简洁、友好、扁平化的用户交互界面。并且负责一些简单的功能：数据校验如输入检验；解析后端数据并显示。

### 3.1.1. jQuery框架

广告管理界面采用网页HTML的形式来进行开发。而现有的WEB前端技术中，使用比较广泛的是jQuery。

jQuery是一个轻量级的JavaScript的函数库，整体大小只有30KB左右，并且Min版经过Gzip压缩之后，大小只有18KB。对于WEB网页来说是十分优秀的。并且jQuery作为一个2006年John Resig开创的一个开源项目，至今为止，由于越来越多的开发者的假如，jQuery的功能已经变得身份强大了。可以用很少的代码，完成很多复杂而困难的功能。jQuery目前集成了JavaScript、CSS、DOM以及Ajax，用起来可以说是方便至极。

本广告系统前端页面使用的是基于jQuery的Amaze UI。Amaze UI 是中国首个开源的HTML5跨屏前端框架。选择使用Amaze UI作为前端开发框架是基于以下优点的：

1) 开源。这也是最重要的一点。作为学生并没有经费来购买其他前端框架；

2) 轻量级，高性能。和jQuery一样，整个Amaze UI的函数库文件大小也是非常可观的；

3) 组件丰富，模块化。Amaze UI 含近 20 个 CSS 组件、20 余 JS 组件，更有多个包含不同主题的 Web 组件，可快速构建界面出色、体验优秀的跨屏页面，大幅提升开发效率。

4) 扁平化设计。简单来说，就是好看。



**图5 基于Amaze UI的广告管理前端页面**

### 3.1.2. Ajax技术

Ajax的全称为“Asynchronous Javascript And XML” ，则异步的JavaScript和XML，是一种用于创建快速动态网页的技术。

Ajax技术始于1998年，到现在几乎每个WEB网页都会使用到该技术。传统的网页如果需要更新内容，必须重新加载整个页面。而Ajax技术能够很好的解决这个问题，无需刷新即可动态更新页面。

广告系统前端采用Ajax异步请求的方式，与后端进行数据的交互。

## 3.2. 后台系统

### 3.2.1 MVC设计思想

MVC的全称为Model View Controller，则模型——视图——控制器。这是一种将逻辑、数据、界面显示分离的方法组织代码模式。

View（视图）是应用中负责界面显示的部分，也就是用户能够看得到的页面；

Controller（控制器）是应用中负责处理用户请求，与用户交互的部分；Model（模型）是应用程序中用于处理应用程序数据逻辑的部分。

View层是根据Model层来进行创建的；而Controller则负责从View层读取数据，然后把数据发送给Model层；Model层负责数据的持久化——则数据库的存取。



**图6 MVC设计模式**

采用MVC设计模式的原因主要是想要把V层和M、V层分离开来，则前后端分离。这样做的好处是，代码能够模块化，前端的功能只需修改View层，而不影响到后端的代码。同理，如果后台来了新需求也是一样，不需要影响到View层的架构。MVC设计模式能够让系统可拓展性更强，更加容易维护和更加易读。

### 3.2.2. SSM框架

SSM框架，是Spring + Spring MVC + MyBatis的缩写。这个是目前比较主流的Java EE企业级框架，也是比较主流的Web应用框架。

Spring是一个2003年就开始发行的Java开源框架，其主要的作用在于使用JavaBean来进行开发。JavaBean可以理解成一个实体类。Spring带来的技术好处有两个：控制反转和面向切片编程。面向切片编程是一种编程思想，面向对象编程也是一种编程思想。本项目并没有很多用到这个功能。而用的比较多的是控制反转。控制反转允许我们交给Spring去控制一个对象的实例化和销毁。应用的比较多的是：Spring将前端传过来的Json数据，解析转换成特定的对象，极大简化了编程。

Spring MVC则是一个基于MVC思想的WEB框架。Spring MVC很好的把M层、V层、C层从WEB项目中抽离出来。



**图7 Spring MVC框架[3]**

MyBatis一个基于Java的MySQL数据库持久层框架。MyBatis几乎消除了所有JDBC（Java DataBase Connectivity，java数据库连接）的代码以及参数的手工设置以及结果集的检索。在MyBatis中使用简单的XML可以将数据库查询结果映射为Java对象返回给调用者。

整体的基于SSM的WEB框架如图8所示。



**图8 基于WEB的SSM框架**

### 3.2.3. 精准投放算法

广告的精准投放功能是基于用户画像(User Profile)系统的。用户画像[4]则用户标签，是一种用标签来表示用户的方法，也就是使用标签来为用户建模。标签是一个描述用户的角度，举个例子一个人收入高、坐办公室可以用“白领”标签来表示。



**图9 用户画像的建立**

用户画像的建立是通过收集用户的日常行为、习惯，然后将收集到的数据放到大数据平台进行分析，建模而得到的。用户画像数据的采集，一般是从很细的行为开始，如常用软件的时间，甚至使用App时手指滑动的速度，这些都能成为有贡献的数据。用户画像的生成如图9所示。

有了用户画像系统的支持，广告系统的定点投放就变得简单了。用户画像将用户表示成如：（学生：0.8，追星：0.5……），这样的数字模型。我们只需要在广告管理中，也为广告添加这样的目标用户范围，然后求出用户与广告的目标用户的吻合度，就可以进行定点投放。前提是广告管理者要对添加的广告熟悉，知道自己的广告的面向群体，这个问题不大。

广告与用户吻合度匹配算法方面。最简单的就是直接判断用户的画像是不是广告目标用户范围的子集。如果是，则把广告加入到用户的广告列表。这是我一开始使用的算法，虽然对于精准投放有一定的影响，但并不能作为一个好方法。

其存在的缺点有：第一，不能完全利用好用户画像标签，只存在于符合目标用户和不符合这两个选项，而实际上用户标签是有权重的，上述算法并没有用到用户画像的权重；第二个缺点是，对于一些很“大众”的，用户标签类型不突出的用户，广告的投放的精准度不够。直接匹配的算法，无法更好的分辨出用户之间的细微差别；第三个缺点是，用户广告列表没有一个很好的数量限制，会出现某些用户广告列表上百上千的情况。

为了解决以上缺点，对算法进行改进。上述缺点的原因一个是在于用户画像的权重没有得到很好的利用，另一个是在于用户的广告列表没有一个明确的限制。

为了利用好用户画像，可以对广告的目标用户标签与用户的用户标签通过余弦相似度求出其相似度权重，然后将权重作为精准投放的标准。余弦相似度是一个广泛用于推荐系统的算法。余弦相似度反映的是两个对象之间的吻合度。

而对于用户广告列表的限制，可以对相似度权重进行排序，优先选择权重较大的广告进行投放。用户在拉取广告的时候，后台先对广告进行截取，根据策略不同对广告列表进行筛选，如只选择与用户相似度大于0.5的前20个广告作为投放对象等等。

### 3.2.3. 系统设计

除了外部提供的用户画像系统之外，整个广告系统的服务器分为4个部分。分别是广告管理服务器、大数据处理服务器、通信服务器以及数据库服务器。



**图9 服务器架构**

用户画像服务器为外部服务器，主要提供用户标签的更新。而广告系统的用户标签的持久化和维护，需要用户画像服务器调用广告系统的接口来进行。

广告管理服务器为WEB服务器，也是主要的服务器。负责管理员的广告操作（增加、删除、修改、查询），以及与广告用户的交互，用户广告获取。并且负责后台功能中的广告信息持久化，用户标签持久化，用户广告相似度更新请求的发起。其他服务器都是为了负载均衡的需求才搭建的，所有服务器都是为主服务器而服务。

大数据服务器负责处理广告相似度的计算。这是使用Spark+Hadoop搭建的分布式服务器。大数据服务器收到更新广告相似度请求之后，会从数据库服务器中获取数据，进行计算后，把结果再写到数据库服务器。

数据库服务器则负责数据的持久化，作为所有服务器的数据来源。

通信服务器负责服务器之间的通信。kafka是一个支持分布式的消息队列组件。服务器之间通信采用消息队列方法，对比与采用HTTP来说，更加安全可靠，并且响应速度快。HTTP请求伴随着外网攻击的风险。zookeeper服务器则用于kafka服务的注册。

整个广告系统的逻辑集中在主服务器——广告WEB服务器种。为了让设计出来的广告后台系统可读性强，并且功能方面容易扩展，能够满足后续的其他功能需求，代码设计采用面向对象的思想。整体框架采用MVC框架，把整个大的系统按照功能分为一个个模块，做到各个模块各尽其责，各个模块只负责单一任务。解耦的好处在于，当遇到bug或者新需求的时候，能够快速地定位到问题所在。某一模块的改变，不会影响到其他模块。人的大脑和电脑硬件有相似的地方，当程序员看代码的时候，会把代码装载进大脑，系统解耦得越好，则遇到问题的时候大脑需要考虑的地方就越少，效率就越高。

根据MVC思想以及SSM框架，我将广告后台系统分为四层：View层，Controller层、Service层以及Dao层。

View层负责网页的前端展示。其主要载体为jsp文件；Controller层负责接收用户请求，然后分发到Service层；Service层负责业务逻辑的处理，不同的业务有不同的Service；而Dao层则负责数据的持久化和读取。

整个业务响应流程为：View层负责界面的显示以及用户的交互，View把接收到的用户请求交由给Controller层进行处理，Controller根据用户请求区分成不同的业务调用Service层（业务层），Service层则可以调用Model层的Dao进行数据的持久化，则保存到数据库。然后把处理过后的信息再逆向通过View层反馈给用户。

**图10 后台系统分层模块**

如图所示。一个Controller可以调用多个Service，一个Service也可以调用多个Dao。Dao设计的原则是尽量一个表格一个Dao，尽量不要存在多个Dao同时对应到一个表格的操作。事实上，这个很难避免。我在设计的数据库的时候，插入数据操作经常会用到联表查询。数据库这样设计的好处在于，一表一个Dao容易进行数据缓存管理。数据库的设计将会在后面详细讲解。注意到的是广告文件并不保存在MySQL数据库中，因为如果这样做的话数据库的负担会很大，企业一般也不会这样做。

我根据功能的划分，将Controller分为5个。分别是单纯用于与View层界面交互的MainController；用于管理员登录管理的AdministratorController；用于外部用户画像系统更新用户画像的TagController；用于广告信息管理以及广告信息获取的AdvController；以及用于广告文件下载的DownController。

根据不同的特定业务，建立了3个Service。分别是用于管理和获取用户标签的TagService；用于管理员登录密码判断、登录状态维护以及管理员密码修改的AdministratorService；还有就是负责处理与广告信息相关的AdvService。其中最为重要的就是AdvService，在这里面负责了广告的添加、删除、查询、修改，以及负责与广告匹配大数据服务器的通信。

### 3.2.4. 数据库设计

数据库表格分为4个模块。用于存广告信息的广告表格；用于保存用户画像的用户画像表格；用于保存用户广告列表的用户广告表格；以及用于保存管理员个人信息的表格。



**图11 广告表格**

广告表格中，目标广告与目标用户标签是一对多的关系，则一个广告拥有多个目标用户标签。表格中id字段为主键唯一标识；name为广告名称；type为广告的类型，用来表示广告是图片还是视频或者是文字；is\_valid字段表示该广告是否处于有效的状态，如果值为0则表示无效，用户将不会获取得到该广告；homepage字段为广告跳转的URL；display\_detail字段用于保存广告的投放时间段，用json格式保存便于拓展；start\_date为广告开始的有效时间，end\_date为广告截止的有效时间；tag\_id对应用户画像系统的标签唯一标识。

**图12 用户画像表格**

用户画像表格中，用户和用户画像也是一对多的关系。剩下的管理员表格就是一个id和密码字段，没必要单独说明。

值得一提的是，数据库中有一个表格不属于以上三个持久化模块，属于一个业务功能的表格产物。在保存广告文件的时候，我的设计是广告文件的名字是广告id+文件后缀。所以在保存广告文件的时候，必须获得广告id。MySQL中提供了自增id字段，以及last\_insert\_id()函数，可以获得最后插入的数据的id值。

但是用这个方法来获取新插入广告的id有两个缺点：

第一，获得广告id必须要在广告信息插入到数据库之后，这个和广告后台系统的需求冲突了。广告文件的保存必须要在保存广告信息之前。原因是，如果先保存广告信息到数据库，然后保存广告文件失败，就会出现数据库中的广告信息的错误的，无法得到正确的广告文件，想要解决这种情况唯有数据库的回滚。其开发成本是很高的；

第二个缺点，必须保证插入操作和last\_insert\_id()函数的调用是一组原子操作。在多线程的环境之下，插入操作同时会进行多次。要保证原子性，就要加锁。因为数据库的操作通常是很耗时的，尤其是插入操作。在这里加锁很导致高并发的时候的线程阻塞。

为了解决这一情况，我写了一个用于获取不重复id的模块。

**图14 id生成模块类图**

大致的算法是，通过调用IdMgr.genId(String name) 这个方法来获取id。IdMgr为一个单例。整个算法的思路是：IdMgr创建的时候，会访问一次数据库，读取数据库中id\_generator表格的所有记录，根据读取的内容生成IdGenerator，保存到以字段name为主键的Map中。其中current\_id字段记录了上一次的未使用的id，step字段为每个current\_id的间隔，则每个current\_id的值都能整除step字段。



**图13 id\_generator数据库表格**

生成id时，调用IdGenerator.genId()方法获取。IdGenerator.genId()的代码如下：

public Long genId() {

synchronized (this.idGenPojo) {

long maxId = this.idGenPojo.getCurrentId();

this.curId += 1;

if (this.curId >= curId) {

idGenPojo.nextStep();

UpdateDb();

}

return this.curId;

}

}

算法思路为：保存一个Long curId的变量，每次调用genId()方法的时候让其加1。然后判断this.curId是否大于数据库中的current\_id字段，如果是则让current\_id = current\_id + step，然后更新数据库中的current\_id字段。最后返回curId。

### 3.2.5. 文件缓存

之前提到过，广告文件不保存在数据库中以减轻数据库的压力。而保存在磁盘的文件，在读取的时候会耗费大量的时间。而广告文件的获取是十分频繁的。为了解决上述情况，需要对广告文件进行缓存。

在FileDao模块中，采用Google提供的Gauva中LoadingCache作为缓存中间件。选用Gauva的原因是其组件是线程安全的，并且容易使用和管理。缓存的流程为：读取广告文件的时候，先访问缓存中是否有该文件，如果有则直接从缓存获取，无则从本地磁盘读取文件。有了缓存，只需要在第一次读取的时候访问磁盘，大大提高效率。在保存或更新广告的操作中，需要先清除缓存中的对应的广告文件，然后再保存到本地中，以避免脏数据。

# 参考文献

[1]. 第39次《中国互联网络发展状况统计报告》[J].中国经济报告,2017(04):7.

[2] 高杰. 基于用户行为的精准广告投放研究[D].武汉工程大学,2016.

[3] <https://blog.csdn.net/dairuiqing/article/details/67637179>

[4] 产品研发中用户画像的数据模建\_从具象到抽象\_余孟杰

[1] 徐国建. 广告管理系统设计与实现[D].电子科技大学,2009.

**毕业论文（设计）成绩评定记录**

Grading Sheet of the Graduation Thesis (Design)

|  |
| --- |
| 指导教师评语  Comments of Supervisor：  成绩评定  Grade：    指导教师签名  Supervisor Signature ： Date: |
| 答辩小组或专业负责人意见  Comments of the Defense Committee：  成绩评定  Grade：    签名： Date:  Signatures of Committee Members |
| 院系负责人意见  Comments of the Academic Chief of School：  成绩评定  Grade：  签名 院系盖章  Signature： Stamp: Date: |