**中国矿业大学计算机学院**

**2014 级本科生课程报告**

课程名称 面向对象分析与技术

课题名称 基于WiFi探针的商业大数据分析系统

报告时间 2017.7.5

学生姓名 赵勤博

学 号 08143427

任课教师 雷小锋

任课教师评语

任课教师评语（①对课程基础理论的掌握；②对课程知识应用能力的评价；③对课程报告相关实验、作品、软件等成果的评价；④课程学习态度和上课纪律；⑤课程成果和报告工作量；⑥总体评价和成绩；⑦存在问题等）：

成 绩： 任课教师签字：

年 月 日

摘 要

数据采集、存储和数据分析技术飞速发展，大大降低了数据储存和处理的成本，使得大数据分析成为可能。本文采用面向对象分析技术，基于UP开发过程，针对基于WiFi探针的商业大数据分析系统进行初始和细化阶段的分析，UP开发中加入敏捷的思想，各个阶段的制品是可选的。

关键字： 面向对象分析；初始；细化

ABSTRACT

The rapid development of data acquisition, storage and data analysis technology greatly reduces the cost of data storage and processing, making large data analysis possible. This paper uses object-oriented analysis technology, UP development process based on the analysis of large data system of commercial WiFi probe was analyzed based on the initial and the refinement phase, with agile thinking in the development of UP products, each stage is optional.

**Keywords:**  Object-oriented analysis; initial; refinement

目 录

[1 项目概述 1](#_Toc487358236)

[1.1背景 1](#_Toc487358237)

[1.2简介 1](#_Toc487358238)

[2 可行性报告 1](#_Toc487358239)

[2.1概述 1](#_Toc487358240)

[2.2经济可行性 1](#_Toc487358241)

[2.3技术可行性 2](#_Toc487358242)

[2.4法律可行性 2](#_Toc487358243)

[3 用例与需求 2](#_Toc487358244)

[3.1用例模型 2](#_Toc487358245)

[3.1.1采集数据用例分析 3](#_Toc487358246)

[3.1.2查询入店量用例分析 3](#_Toc487358247)

[3.2需求分析 4](#_Toc487358248)

[4 细化 5](#_Toc487358249)

[4.1领域模型 5](#_Toc487358250)

[4.2新的数据类型 6](#_Toc487358251)

[4.2.1入店人次趋势值 6](#_Toc487358252)

[4.2.2驻店时长趋势值 7](#_Toc487358253)

[4.2.3活跃度统计 7](#_Toc487358254)

[4.3设计模型 7](#_Toc487358255)

[4.3.1入店量/入店率时序图 7](#_Toc487358256)

[4.3.2顾客活跃度趋势时序图 8](#_Toc487358257)

[4.4软件架构 9](#_Toc487358258)

# 1 项目概述

## 1.1背景

在近十年时间里，数据采集、存储和数据分析技术飞速发展，大大降低了数据储存和处理的成本，使得一个大数据时代逐渐展现在我们的面前。曾经，枯燥的数据统计和分析只是统计学家、数据分析师和科研学者们的专利，而大数据革新性的将海量数据处理变为可能，并且大幅降低了成本，使得越来越多跨专业学科的人投入到大数据的开发应用中来。大数据能够帮助人们更明智的制定决策、更清晰的传达理念。面对大数据瑰丽且深奥的面貌，如何才能让大型数据集变得亲切和易于理解，可视化无疑是最有效的途径。对大数据背景下的数据可视化应用。

## 1.2简介

基于WIFI探针的商业大数据分析技术主要应用于商场、景区、商业街的客流分析，结合探针系统。Wifi探针的原理是利用智能设备WIFI模块所发出的唯一MAC地址和无线广播信号进行设备的感知测量，经过大数据采集平台算法，记录访问行为和轨迹，收集客流量信息，可准确、高精度对设备范围内客流量进行统计，生成营销数据报告，完美实现客流分析、行业分析，可为商家提供运营成本的有效分配以及管理建议和业务转化率，它把传统的商业从凭经验做生意引导向了一个可以利用互联网的数据分析，信息化支持的高速道路，这也为传统商业领域提供巨大革新和，互联网化结合。

可行性分析

# 2 可行性报告

## 2.1概述

可行性研究报告是从事一种经济活动（投资）之前，双方要从经济、技术、生产、供销

直到社会咅种环境、法律等咅种因素进行具体调查、研究、分析，确定有利和不利的因素、

项目是否可行，估计成功率大小、经济效益和社会效果程度，为决策者和主管机关审批的上

报文件。

## 2.2经济可行性

在功能性测试阶段，只需要单服务器完成所有功能，项目运行时为节省高性能服务器的成本，采用分布式构架部署系统，需要3台及以上服务器，负载均衡器，数据库服务器集群。

人力成本需要1名较有经验的系统架构师，软件开发工程师3人以上，测试工程师1名，运维工程师1名，开发周期为6个月。

本项目投资成本适中，约40万以内，系统通用性较高，可以客观的分析因销售策略等变化引起的店铺顾客行为画像，有较高的经济价值。

## 2.3技术可行性

技术方面采用比较新的ssm框架，开源的zookeeper和negix实现集群和负载均衡，使用分布式mysql做数据库集群，使用redis作为高速缓存，bootstrap作为前端框架。技术应用都符合开源框架，且文档教程较多，易于开发实现。

## 2.4法律可行性

对收集的用户信息进行脱敏处理，不会涉及用户的敏感信息。

# 3 用例与需求

## 3.1用例模型

描述功能需求。在初始阶段，确定大部分用例名称，详细分析10%的用例。

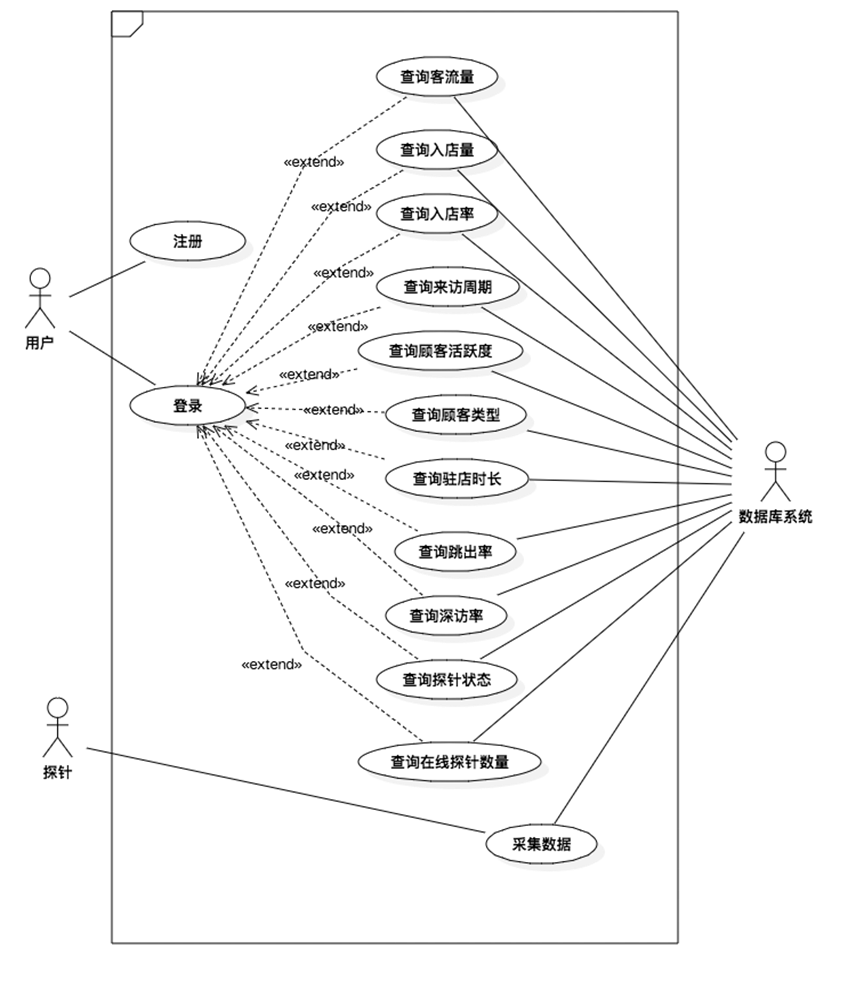


图3.1 用例模型

### 3.1.1采集数据用例分析

用例名称：采集数据

级别：子功能 采集探针数据

主要参与者：探针，数据库系统

涉众及其关注点：

-探针：希望能快速的发送数据，发送的数据项能够被解析。

-数据库系统：希望解析后的数据能够快速地、准确地、大容量的插入到数据库中，解析失败的数据不插入数据库。

前置条件：探针开启且在线，探针已经绑定到某个区域或店铺。

成功保证：存储解析数据。

主成功场景：

1.探针携带信息POST至商业大数据分析系统。

2.系统将解析后的数据发送给数据库系统。

3.数据库系统存储解析数据。

扩展：

a.POST信息未接收

1.抛弃传递失败的数据。

b.解析数据失败

1.抛弃解析失败的数据。

c.数据库系统存储失败

1.回滚当前存储的事务。

2.抛弃当前未存储的数据。

d.探针未绑定

1.提示探针未绑定信息，并停止发送数据。

e.店铺未配置

1.提示店铺未配置信息，并停止发送数据。

特殊需求：

·探针可以抓取到mac地址、距离、地理位置、时间等数据

·探针数据接收服务端最少支持1000并发

·探针设备每3秒传输一次数据

发生频率：每3秒发生一次

未解决问题：

·针对不同探针如何解决数据。

·探针绑定需要提供哪些信息。

### 3.1.2查询入店量用例分析

用例名称：查询入店量

级别：用户目标

主要参与者：用户，数据库系统

涉众及其关注点：

-用户：希望能够便捷快速的查询入店量信息，显示的结果易于理解。

-数据库系统：希望快速准确的计算查找相关信息。

前置条件：用户已登录。

成功保证：返回入店量的详细数量和时间段。

主成功场景：

1.用户登录。

2.用户设置查看店铺ID。

3.用户设置查看的时间区间。

4.用户请求查看入店量。

5.系统解析用户传入的参数信息，并请求数据库分析。

6.数据库将分析的结果发送给系统。

7.用户查看到指定店铺ID和时间区间的入店量信息。

扩展：

a.用户未登录

1. 提示用户未登录，没有查看相关信息的权限。

b.店铺ID不存在

1. 提示用户店铺不存在，创建店铺或重新选择。

c.时间区间不合理

1.提示用户不是合法的时间区间或该时间区间造成查询数据量过大而无法查询。

d.redis存在相关的查询数据

1.直接从存缓中返回数据结果，不使用数据库查询。

特殊需求：

·查询速度在30秒以内

发生频率：经常会发生

## 3.2需求分析

表3.1 需求分析

|  |  |
| --- | --- |
| 功能性需求 | （1）WIFI探针  探针设备可以进行服务端的相关配置（服务端IP、端口、路径、发送数据时间间隔）。  能够采集MAC地址、地理信息、与探针大概距离、采集时间等信息。  （2）数据采集  服务端主要接收探针定时发送的数据，将数据保存到数据分析平台待用。  接收数据不能有数据丢失或者数据失真。  探针每3秒发送一次数据。  数据采集并发量不得低于1000台设备。  （3）数据分析  客流量：店铺或区域整体客流及趋势  入店量：进入店铺或区域的客流及趋势  入店率：进⼊店铺或区域的客流占全部客流的比例及趋势  来访周期：进⼊店铺或区域的顾客距离上次来店的间隔  新老顾客：一定时间段内首次/两次以上进⼊店铺的顾客  顾客活跃度：按顾客距离上次来访间隔,划分为不同活跃度（高活跃度、中活跃度、低活跃度、沉睡活跃度）  驻店时长：进⼊店铺的顾客在店内的停留时长  跳出率：进⼊店铺后很快离店的顾客及占比(占总体客流)  深访率：进⼊店铺深度访问的顾客及占⽐(占总体客流)（可以根据定位轨迹或者停留时长判定）  以上数据指标支持环比和历史对比，并且可以从小时、日、周、月多维度分析。数据分析中的关于范围的界定支持自定义阈值。 |
| 非功能性需求 | 数据分析平台健壮，数据展现界面友好，探针设备配置简单、运行稳定。  能够对探针设备状态进行监控，并可展现所有接入探针的运行状态。  能够对探针设备进行在线远程控制，包括关机，重启，参数配置，升级固件，设定自动开关机时间，远程控制开关。 |

# 4 细化

## 4.1领域模型

领域模型是可视化字典，表示领域的重要抽象、领域词汇和领域的内容信息。

定义的概念类有：用户、店铺、顾客、探针、探针日志。关联与属性如下图所示：

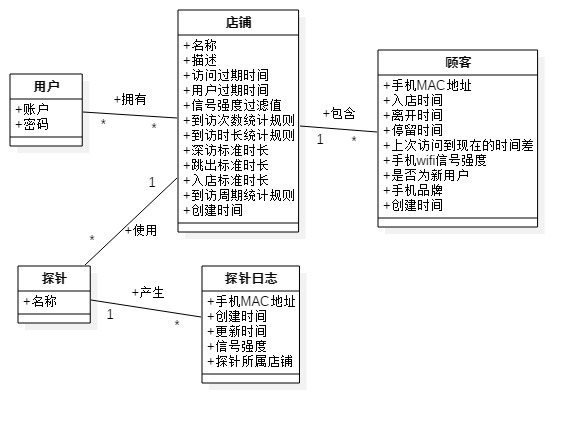


图4.1 领域模型

## 4.2新的数据类型

### 4.2.1入店人次趋势值

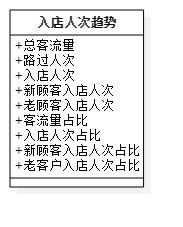


图4.2入店人次趋势值

### 4.2.2驻店时长趋势值

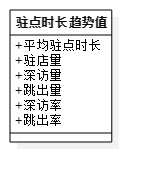


图4.3 驻点时长趋势值

### 4.2.3活跃度统计

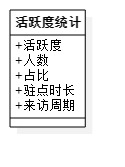


图4.4 活跃度统计

## 4.3设计模型

### 4.3.1入店量/入店率时序图

用户选择要查询的开始时间、截止时间、时间间隔，然后点击界面上入店量／入店率按钮，来发送获取入店量／入店率请求。UI界面会向系统传递三个参数： idShop、start、end、period（时间间隔，共四种：hour、day、week、month）。

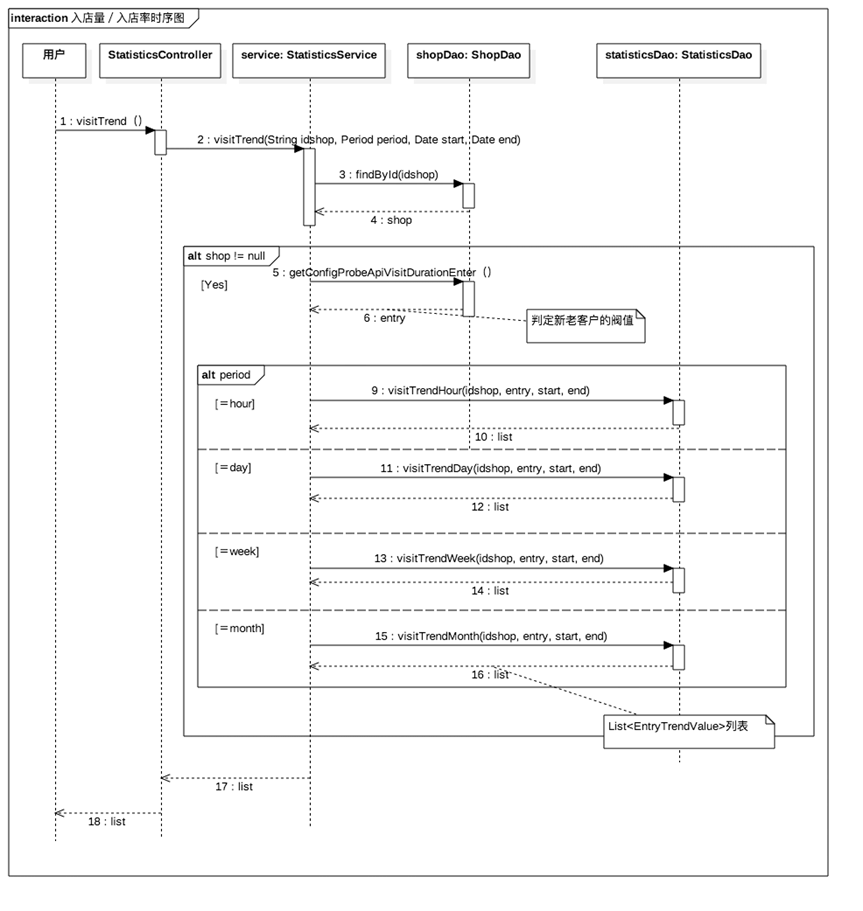


图4.5入店量/入店率时序图

### 4.3.2顾客活跃度趋势时序图

计算方法：

1） 将顾客活跃度分为5个标准： "高活跃度 0-1天", "中活跃度 1-3天", "低活跃度 3-7天", "沉睡顾客7-15天", "深度沉睡 15天-"

2） 根据查询的开始时间、截止时间、店铺ID、访客／顾客标准筛选出数据

3） 将数据按照用户选择的时间间隔（4种：hour、day、week、month）来进行分组

4） 对每个时间段进行统计，若顾客距离上次的访问时长处于该活跃度内，则计入该时间段的总数

5） 计算完毕，返回数据集

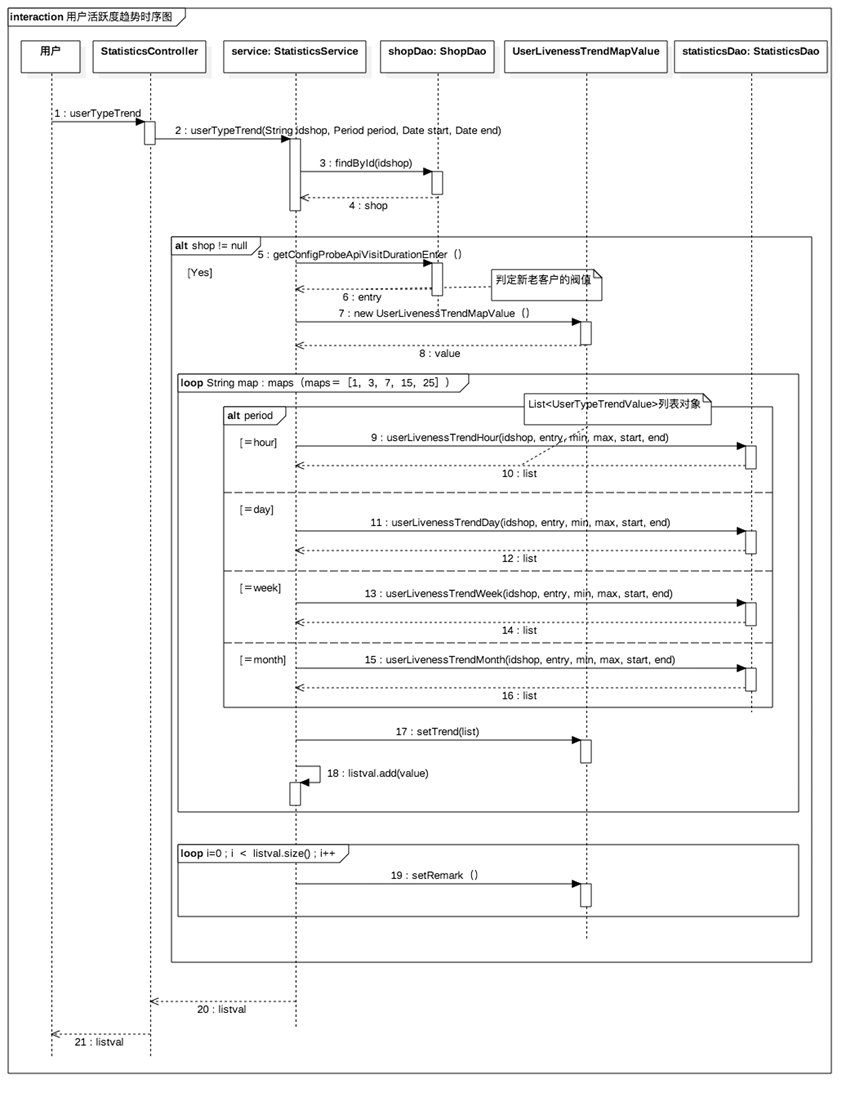


图4.6顾客活跃度趋势时序图

## 4.4软件架构

·Maven（模块化构建）

·Spring（IOC DI 声明式事务处理）

·SpringMVC（支持Restful风格）

·Hibernate Validate（参数校验）

·Mybatis（最少配置方案）

·Quartz时间调度

·Redis缓存（ProtoStuff序列化）

·Redis Sentinel主从高可用方案

·Druid（数据源配置 sql防注入 sql性能监控)

·Dubbo+Zookeeper分布式服务框架

·合理的分布式服务划分（common+api+service+web）

·资源调度和治理中心(SOA)（dubbo-admin）

·服务监控方案(dubbo-monitor)

·分布式系统中的异常处理

·分布式事务

·分布式锁

·dubbo服务集群、负载均衡策略

·Redis集群高可用方案

·Zookeeper集群高可用方案

·消息中间件 ActiveMQ 的引入

·ActiveMQ 集群高可用方案

·分布式文件系统（FastDFS）的引入

·FastDFS集群高可用方案

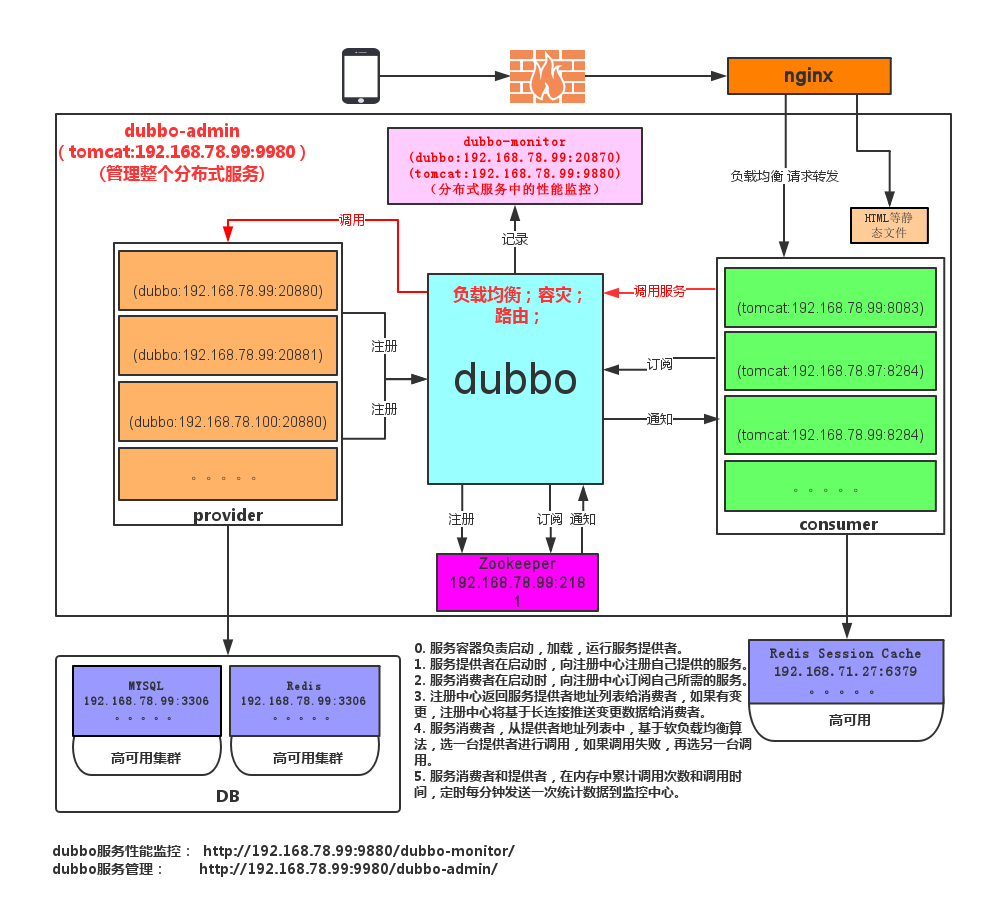


图4.5 架构图