**基于WIFI探针的商业大数据分析系统**

**软件系统设计说明书**

|  |  |  |
| --- | --- | --- |
| 文件状态：  [√] 草稿  [ ] 正式发布  [ ] 正在修改 | 文件标识： | Company-Project-LJBB-DESIGN |
| 当前版本： | V 2.0 |
| 作 者： | BugKiller |
| 完成日期： | 2017.6.10 |

**文档信息及版本历史**

|  |  |  |  |
| --- | --- | --- | --- |
| 文档信息 | | | |
| 项目名称 | 基于WIFI探针的商业大数据分析系统 | | |
| 项目编号 | A1 | | |
| 文档名称 | 需求规格说明书 | | |
| 存储位置 | /Project/doc | | |
|  | | | |
| 版本 | 作者/修改者 | 日期 | 描述 |
| V1.0 | BugKiller | 2017-05-17 | 完成数据采集与书记展示模块 |
| V1.2 | BugKiller | 2017-05-24 | 完成数据模拟脚本 |
| V2.0 | BugKiller | 2017-05-27 | 增加数据图表 |
| V2.1 | BugKiller | 2017-06-03 | 修复数据预测模块BUG |
| V3.0 | BugKiller | 2017-05-07 | 优化算法 |
| V3.4 | BugKiller | 2017-06-07 | 优化算法增加预测算法 |

目 录

[第1章 系统概述 4](#_Toc27496)

[第2章 设计约束 4](#_Toc1526)

[1、本软件的开发环境： 4](#_Toc19650)

[2、运行环境： 4](#_Toc16040)

[第3章 开发、测试与运行环境 5](#_Toc14595)

[1开发环境： 5](#_Toc12541)

[2集成测试环境： 5](#_Toc20059)

[第4章 数据库设计概述 6](#_Toc30476)

[4.1 数据库环境说明 6](#_Toc2185)

[4.2 数据库命名规则 6](#_Toc18324)

[4.3 安全性设计说明 6](#_Toc11615)

[1、防止用户直接操作数据库的方法 6](#_Toc23268)

[2、用户帐号密码的加密方法 6](#_Toc7831)

[4.4 表汇总和表设计 7](#_Toc29678)

[第五章 系统详细设计概述 9](#_Toc17156)

[5.1系统结构设计及子系统划分 9](#_Toc5919)

[5.2系统功能模块详细设计 9](#_Toc2730)

[5.2.1数据接收模块 9](#_Toc1191)

[5.2.2数据处理及预测模块 10](#_Toc27717)

[5.2.3数据展示模块 11](#_Toc31322)

# 第1章 系统概述

利用探针数据的客流分析打破模式束缚，不仅仅只是提供可信的客流数据分析，同时还利用延伸的标杆管理才能，深刻洞悉并提供有助于推动实际客流量和消费者习惯行为的一系列因素。这种专业才能呈现了经济分析，社交和环境等一些超出你控制范围的因素，却对商业绩效产生主要的冲击力。

# 第2章 设计约束

## 1、本软件的开发环境：

A、硬件：内存：8GB

硬盘：250GB

处理器：2.9GHZ

B、软件：操作系统：Ubuntu

HADOOP版本：Hadoop2.0

JAVA版本： jdk1.8

开发工具：Pycharm, notepad++

数据库工具：Mysql, Hive, Sqoop

## 2、运行环境：

A、硬件：内存：8GB

硬盘：200GB

处理器：2.9GHz

B、软件：操作系统：Gnome

.NET框架：.NET4.0

# 第3章 开发、测试与运行环境

## 1开发环境：

|  |  |  |
| --- | --- | --- |
| 类别 | 标准配置 | 最低配置 |
| 计算机硬件 | CPU:2.9GHz  内存：8GB  硬盘：250GB | CPU:2.0GHz  内存：4GB  硬盘：100GB |
| 软件 | Hadoop2.0, java1.8,sqoop,hive | Hadoop2.0, java1.8,sqoop,hive |
| 网络通信 | 4MBb宽带 | 2MBb宽带 |
| 其他 |  |  |

表3-1运行环境

## 2集成测试环境：

|  |  |  |
| --- | --- | --- |
| 类别 | 标准配置 | 最低配置 |
| 计算机硬件 | CPU:2.6GHz  内存：4GB  硬盘：250GB | CPU:2.0GHz  内存：2GB  硬盘：50GB |
| 软件 | .NET4.0 | .NET4.0 |
| 网络通信 | 4MBb宽带 | 2MBb宽带 |
| 其他 |  |  |

表3-3集成测试环境

# 第4章 数据库设计概述

## 4.1 数据库环境说明

本系统采用mysql数据库作为项目的数据库系统，设计以及编程工具均采用mysql数据库，大数据数据库采用hive数据库。

## 4.2 数据库命名规则

数据库表：以T\_开头，后接有意义的单词，单词的首字母大写。如：T\_UserInfo

视图：以VW\_开头，后接有意义的单词，单词的首字母大写。如：VW\_UserInfo

存储过程：以SP\_开头，后接有意义的单词，单词的首字母大写。如：SP\_UserInfo

字段：以有意义的单词组成，开头大写。如username

## 4.3 安全性设计说明

### 1、防止用户直接操作数据库的方法

用户只能用帐号登陆到应用软件，通过应用软件访问数据库，而没有其他途径操作数据库。

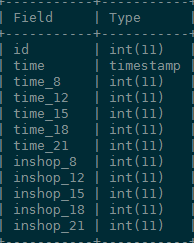
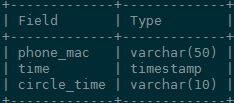
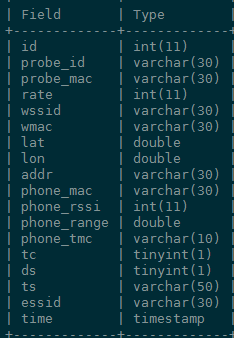
### 2、用户帐号密码的加密方法

对用户帐号的密码进行加密处理，确保在任何地方都不会出现密码的明文。

## 4.4 表汇总和表设计

## 

Data表： data\_set表： history\_count表：



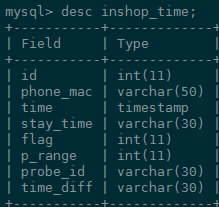
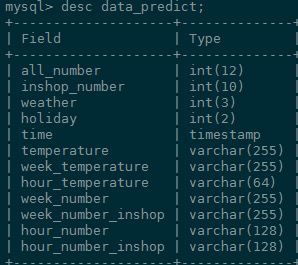


图4-1表汇总

|  |  |  |  |
| --- | --- | --- | --- |
| 表名 | Denglu | | |
| 列名 | 数据类型（精度范围） | 空/非空 | 约束条件 |
| Name | Varchar(15) | 非空 |  |
| Mima | Varchar(20) | 非空 |  |
| Id | Varchar(15) | 非空 | 主键 |
| 补充说明 |  | | |

表4-1 denglu表

|  |  |  |  |
| --- | --- | --- | --- |
| 表名 | Qiandao | | |
| 列名 | 数据类型（精度范围） | 空/非空 | 约束条件 |
| Id | Varchar(15) | 非空 | 主键 |
| Name | Narchar(50) | 非空 |  |
| Dep | Varchar(50) | 非空 |  |
| Time | Bit | 非空 |  |
|  |  |  |  |
| 补充说明 |  | | |

表4-2 qiandao表

|  |  |  |  |
| --- | --- | --- | --- |
| 表名 | Info | | |
| 列名 | 数据类型（精度范围） | 空/非空 | 约束条件 |
| Id | Varchar(15) | 非空 | 主键 |
| Name | Varchar(15) | 非空 |  |
| Info | Varchar(100) | 非空 |  |
| Adr | Bit | 非空 |  |
| Num | Datetime | 非空 |  |
| Sex | Datetime | 非空 |  |

表4-3 Info表

|  |  |  |  |
| --- | --- | --- | --- |
| 表名 | Fabu | | |
| 列名 | 数据类型（精度范围） | 空/非空 | 约束条件 |
| Ide | Char | 非空 | 主键 |
| Name | vachar | 非空 |  |
| Ifo | Char | 非空 |  |
| Power | Int | 非空 |  |

表4-4 Fabu表

# 第五章 系统详细设计概述

## 5.1系统结构设计及子系统划分

该系统的主要功能为数据接收、数据处理及预测和数据展示，为了使系统的设计更加的清晰，我们将系统分为三个模块，分别是数据接收模块、数据处理及预测模块和数据展示模块。

数据接收模块的主要功能为接收WiFi探针发送的数据，通过实时计算来计算入店量和驻店时长并存入数据库。

数据处理及预测模块的主要功能为在每日接收数据结束之后定时的对数据进行处理(包括计算当日来访顾客中老顾客的来访周期、通过爬虫获取当日的每小时的气温和天气状况以及是否是节假日并存入数据库，计算每小时的入店量和客流量并存入数据库，从数据库客流量和入店量表中取样作为历史查询的数据保存在数据库中，将当日所有的数据导入到hive数据库中)和预测。

数据展示模块的主要功能是将数据接收模块正在接收的数据以图表的形式展示出来并能实时更新，可以查询历史数据并将预测的结果以建议的形式展示出来供使用者使用。

## 5.2系统功能模块详细设计

### 5.2.1数据接收模块

**实现功能：**接收WiFi探针发送的数据，将数据实时处理并存入数据库。

**输入：**WiFi探针发送数据的json串。

**输出：**经过处理后的数据

**算法及逻辑设计：**该模块共有两个功能，接收数据和对数据实时处理。数据接收是通过flask的request方法接收WiFi探针发送的数据，并将数据数据存入data表。与此同时，对接收的数据进行处理，通过使用operation中的time\_stay方法对接收到的数据进行分析，判断客户的驻店时长。驻店时长计算算法如下：inshop\_time表存储所有进入过店铺的mac,当接收一个新mac，查询inshop\_time表是否有flag为1的该mac的记录,如果表中没有该mac并且该mac在店铺内而且发送该数据的WiFi探针是指定的探针,将mac和time存入表,flag设为1,如果表中有该mac并且该mac在店铺外而且发送该数据的WiFi探针是指定的探针,查询表中该mac对应的时间,并与接收到的mac时间做差,将查询到的mac更新,flag设为0,并将对应的时间差的代号存入.（该算法要求将一个WiFi探针放在靠近店铺中部的位置，并且计算该探针距离店铺边缘的距离，作为算法中判断mac地址是否是在店铺内部的依据，同时也需要记录该探针的ID，作为判断使用）

### 5.2.2数据处理及预测模块

**实现功能：**在每日数据接收完毕之后计算当天的顾客来访周期将结果存入data\_set表，获取每日的客流量、入店量、天气和当天是否是节假日并存入数据库data\_predict表,用于数据预测，根据数据库中data\_predict表里的数据作为预测的源数据，通过爬虫获取第二天的天气状况和每小时的气温来预测第二天的数据。根据实时计算出的数据来获取每日特定时间的客流量和入店量作为历史查询数据。将每日接收的数据和处理后的数据导入到hive中。

**输入：**data表、inshop\_time表中的数据，爬虫爬取的每小时的气温，天气状况和当天是否是节假日。

**输出：**顾客来访周期，数据预测的基础数据，预测出的第二天的数据，

**算法及逻辑设计：**计算顾客来访周期的方法：查询inshop\_time中当日flag为0的数据,每个mac地址查询一次。根据传入的mac地址来判断data\_set中是否有该mac,若没有则将mac和时间插入data\_set,若有则根据新老时间来计算来访周期,删除旧数据,将mac,时间和来访周期存入表中.

爬虫获取数据：爬虫主要获取三部分数据，每小时的温度，当天的天气状况和当天是否是节假日。爬虫需要获取当地的数据，依据就是WiFi探针发送数据中的经纬度，爬虫先根据经纬度获取当地的省市信息，再获取当地的气温和天气状况。

数据预测：根据当日获取的客流量和入店量以及，数据预测分为两个部分，计算当日的总客流量和入店量以及当日的每小时的客流量和入店量。

### 5.2.3数据展示模块

**实现功能：**将客流量、入店量、入店率、来访周期、新老顾客、顾客活跃度、驻店时长、跳出率、深访率、历史数据和预测建议等展示到前端页面。

**输入：**无

**输出：**无

**算法及逻辑设计：**该部分数据展示包括实时更新的数据和静态的数据，实时更新的数据都是用的highchart的实时更新的折线图来展示的。Js脚本每三秒向服务器请求一次数据，服务器返回这三秒内接收到的数据的信息。静态数据展示使用的是静态折线图和饼形图。历史数据只展示客流量和入店量。因为其余的数据展示历史并没有很大的意义，所以我们只选择这两个可能会对使用者有帮助的部分进行展示。展示的数据全部都从数据库获取，根据数据的时间列来筛选。