

中国软件杯大学生软件设计大赛

“川大 NET”

**基于 WIFI 探针的商业大数据分析技术
算法描述与系统设计书**

学校： 四川大学

队伍： 川行

指导老师： 赵辉

队员： 袁豪 王晗 李春辉

Email: maicius@outlook.com

Tel: 18996720676

目录

1	引言	4
1.1	目的	4
1.2	背景	4
1.2.1	大数据中挖掘商业价值	4
1.2.2	大数据与人工智能的结合	5
1.3	关键词	5
2	任务概述	5
2.1	目标	5
3	需求规定	5
3.1	一般性需求	5
3.1.1	WIFI 探针：	5
3.1.2	数据采集：	6
1.1.3	数据展示：	6
3.2	功能性需求	7
3.2.1	人工智能算法辅助决策	7
3.2.2	数据和信息安全	7
3.2.3	搭建分布式计算平台	8
3.2.4	功能列表	8
	从本节开始为具体功能点需求	9
3.3	基础数据维护	9
3.3.1	数据采集	9
3.3.2	数据预处理	11
3.3.3	数据分析	12
3.3.4	数据展示	18
3.3.5	商场管理	19
3.3.6	管理员管理	20
3.3.7	探针管理	21
3.3.8	属性管理	22
3.3.9	用户管理	23
3.4	智能决策	24
3.4.1	智能决策管理	24
3.5	短信管理	25
3.5.1	阈值设置	25
3.5.2	短信推送	26
3.6	自动生成报表	27
3.7	系统安全性的要求	27
3.7.1	数据存储安全	28

3.7.2	访问控制安全	28
3.7.3	网络传输安全	28
3.7.4	系统约束	28
3.7.5	其他专门要求	28
4	运行环境规定	28
4.1	运行环境	28
4.1.1	软件环境	28
4.1.2	硬件环境	28

1 引言

1.1 目的

智能设备的 WIFI 模块会一直向周围空间广播无线信号，该信号带有智能设备唯一标识符，即 MAC 地址。通过记录和分析 mac 地址出现的时间和地点，能精准的掌握商场或门店内人流量信息，精准监控客流情况，从而帮助商家检测营销效果，不仅可以为商家制定高效、便捷的运营策略提供全方位的数据参考，也可以利用人工智能算法为商家制定有用的营销策略。

1.2 背景

1.2.1 大数据中挖掘商业价值

2012 年 2 月，《纽约时报》的一篇专栏中所称，“大数据”时代已经降临，在商业、经济及其他领域中，决策将日益基于数据和分析而作出，而并非基于经验和直觉。哈佛大学社会学教授加里·金说：“这是一场革命，庞大的数据资源使得各个领域开始了量化进程，无论学术界、商界还是政府，所有领域都将开始这种进程。”

最早提出“大数据”时代到来的是全球知名咨询公司麦肯锡，麦肯锡称：“数据，已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来。”“大数据”在物理学、生物学、环境生态学等领域以及军事、金融、通讯等行业存在已有时日，却因为近年来互联网和信息行业的发展而引起人们关注。

随着云时代的来临，大数据(Big data)也吸引了越来越多的关注。著名云台的分析师团队认为，大数据(Big data)通常用来形容一个公司创造的大量非结构化和半结构化数据，这些数据在下载至关系型数据库用于分析时会花费过多时间和金钱。大数据分析常和云计算联系到一起，因为实时的大型数据集分析需要像 MapReduce 一样的框架来向数十、数百或甚至数千的电脑分配工作。

“大数据”在互联网行业指的是这样一种现象：互联网公司在日常运营中生成、累积的用户网络行为数据。这些数据的规模是如此庞大，以至于不能用 G 或 T 来衡量，大数据的起始计量单位至少是 P(1000 个 T)、E(100 万个 T)或 Z(10 亿个 T)。

大数据到底有多大?一组名为“互联网上一天”的数据告诉我们，一天之中，互联网产生的全部内容可以刻满 1.68 亿张 DVD;发出的邮件有 2940 亿封之多(相当于美国两年的纸质信件数量);发出的社区帖子达 200 万个(相当于《时代》杂志 770 年的文字量);卖出的手机为 37.8 万台，高于全球每天出生的婴儿数量 37.1 万……

另一方面，2016 年，移动互联网流量占到互联网流量的 68%，全球已经正式进入移动互联网时代。全球智能手机普及率达到 56%，预计 2018 年将达到 63%。智能手机，已然成为人们日常生活必不可少的工具，因此，通过对智能手机的行为分析，就能掌握到用户的行为轨迹。通过对一定区域内智能手机

mac 地址等行为统计，就能获得该区域的人流量等重要信息，而这些数据，将成为各大商场、实体店制定商业策略的重要依据。

1.2.2 大数据与人工智能的结合

前不久结束的围棋人机大战中，AlphaGo 击败柯洁等一众围棋高手，成功证明了“深度学习”（Deep Learning）技术的潜力，“人机大赛是我们技术上的里程碑，展示了人类的功劳，我们要以欢迎的心态接受它。”卡耐基梅隆大学计算机科学学院教授邢波在《为人工智能装上引擎》主旨演讲中提到，人工智能与机器学习的区别在于：“人工智能是我们希望达到的一个方向，方法论就是机器学习，实际上是让计算机跟数据、模型环境接触，训练一些东西，达成的效果是希望最终的模型在某些方面具有强大的功能。”

大数据的到来为人工智能插上了翅膀，人工智能也帮助大数据有了更广阔的应用前景。人工智能与大数据的结合，能精准的掌握业务状况，为客户制定最佳的决策方案。

1.3 关键词

大数据 人工智能 探针 智能决策 机器学习

2 任务概述

2.1 目标

1. 通过大数据分析掌握商场的区域客流量、入店量、入店率等十大基础数据
2. 通过友好的界面将数据展示给用户，并提供自动生成报表等工具
3. 广泛应用人工智能算法，辅助客户进行商务决策

3 需求规定

3.1 一般性需求

3.1.1 WIFI 探针：

探针设备可以在其他渠道（淘宝）购买，也可以自行开发硬件和固件，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（不允许采集个人隐私信息，比如手机号等信息），特别针对 ANDROID6.0 和 IOS10 版本后的移动终端设备进行测试能否采集到 MAC 地址。

3.1.2 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，文件系统可以使用 HDFS 或者其他适合的分布式文件系统。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。数据分析

能够分析以下十大指标：

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问隔,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）
手机品牌	根据 MAC 地址前三个字节进行查询

3.1.3 数据展示：

1. 以友好的界面实时展示采集得到的数据
2. 提供用户查询的入口，支持用户按照年、月、日等时间范围进行查询历史数据
3. 用户切换到不同页面时检查用户 session 状态，保证用户合法性，从而保护数据
4. 提供自动生成报表的功能，方便用户得到纸质版数据
5. 所有页面状态刷新支持异步刷新

3.2 功能性需求

3.2.1 人工智能算法辅助决策

人工智能在本项目中的应用包含但不限于以下几个方面：

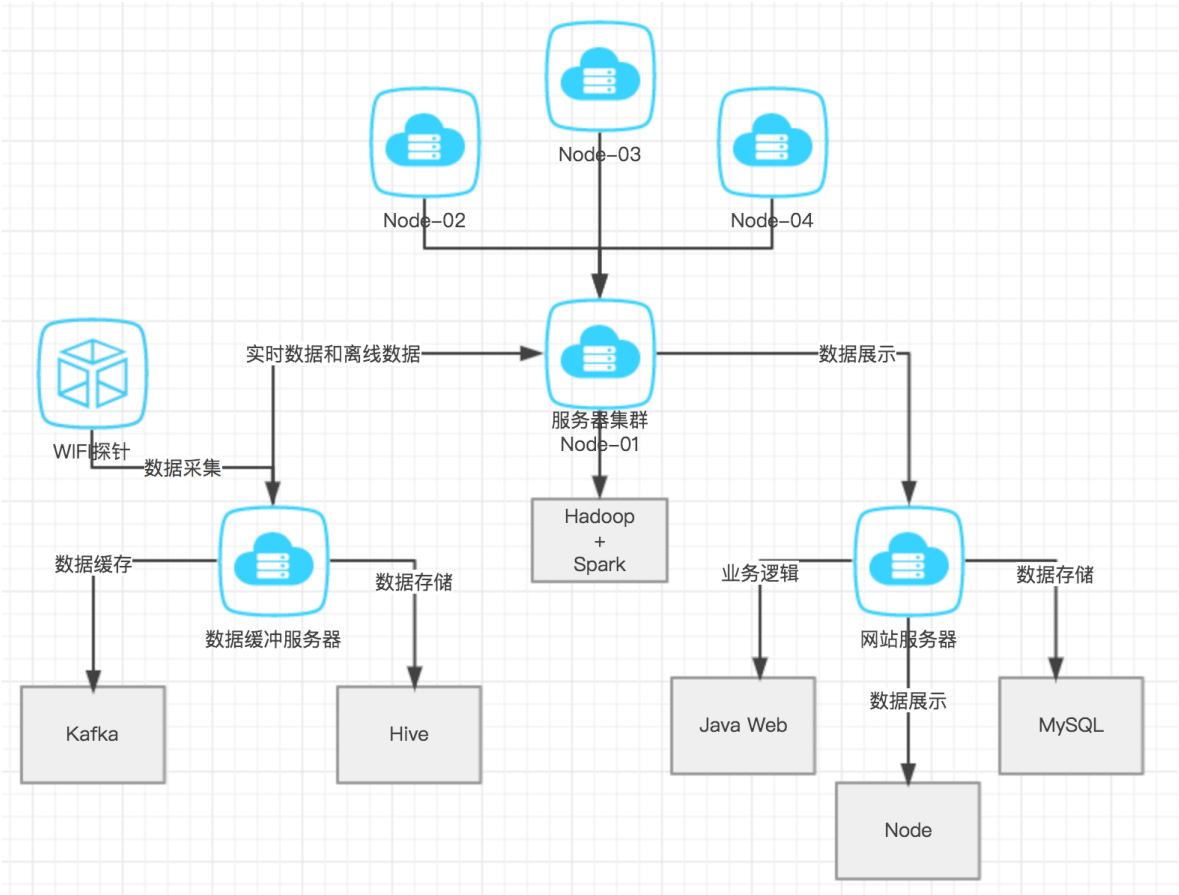
1. 对每个商场的评分使用机器学习算法归类
2. 对每个用户的活跃程度判断使用机器学习算法归类
3. 根据用户设定预测未来一定时间段的流量情况
4. 根据预测的流量峰值为用户提供合理的商务决策方案
5. 提供的商务决策方案由用户评分后进行反馈学习，以此不断增强机器学习训练效果

3.2.2 数据和信息安全

1. 使用 SSL 协议加密传输数据，禁止使用 Http 和 ws，分别改为 Https 和 wss
SSL, 全称 Secure Sockets Layer, 安全套接层，它可以加密传输数据，以防止数据在传输中被窃取。
2. 加密存储重要数据(比如用户密码和商场信息)
使用 sha1 加密算法
3. 记录网站访问日志和服务器日志
用于服务器故障时快速分析问题原因
4. 数据库添加防注入模块
防止数据库被爆破
5. 登陆拦截设置
除登陆注册外所有界面禁止非登陆访问，要求从前端和后端都实现拦截。

3.2.3 搭建分布式计算平台

系统架构图如下：



其中包括：

- 1.服务器集群：搭建 Hadoop 和 Spark，进行大数据分析
- 2.数据缓冲服务器：进行数据预处理和数据缓存
- 3.探针：采集和上传数据
- 4.网站服务器：利用 MySQL 存储分析之后的数据，同时负责与用户的交互和处理业务逻辑。

3.2.4 功能列表

功能编号	功能名称	备注	裁剪说明
AF001	数据采集	通过 WIFI 探针采集 mac 地址等信息	不可裁剪
AF002	数据预处理	对数据进行预处理，格式化数据	不可裁剪
AF003	数据缓存	对数据进行缓存，减轻服务器集群的负担	不可裁剪

AF004	数据分析	根据采集到的数据分析得到客流量、入店率等十大数据	不可裁剪
AF005	商场管理	管理员查看各商场的基本情况，支持统计和分析	不可裁剪
AF006	探针管理	远程配置探针参数，查看探针设备状态	不可裁剪
AF007	用户管理	用户的注册、登陆、个人信息维护等。	不可裁剪
AF008	数据对比	用户可以根据年、月、日，自由选择时间范围进行各种数据的对比	不可裁剪
AF009	自动生成报表	自动生成商场数据报表(excel格式)	不可裁剪
AF010	辅助决策	根据数据利用智能算法进行离线训练，帮助用户进行商务决策	不可裁剪
AF011	Mac 轨迹追踪	根据 mac 地址到附近三个探针的距离,计算出 mac 的位置(经纬度)并显示在地图上	可裁剪
AF012	短信提醒	当客流量巨大时发送短信提醒用户注意	不可裁剪
AF013	日志管理	记录服务器日志	可裁剪

从本节开始为具体功能点需求

3.3 基础数据维护

3.3.1 数据采集

使用探针设备进行数据采集，采集的 JSON 数据结构样例如下：

```
{
  "id": "0010f377", //嗅探器设备 id
  "mmac": "5e:cf:7f:10:f3:77", //嗅探器设备自身 WiFi mac
  "rate": "1", //发送频率
```

```
"wssid": "kaituo", //嗅探器设备连接的 WIFI 的 ssid
"wmac": "a8:57:4e:c0:d4:8c", //嗅探器设备连接的 WIFI 的 mac 地址
"time": "Sat Jun 04 22:45:28 2016", //时间戳，采集到这些 mac 的时间
"lat": "30.748093", //北半球，纬度
"lon": "103.973083", //经度
"addr": "江苏省南京市玄武大道 699-22 号", //地址信息
"data": [{
    "mac": "9a:21:6a:7b:62:6a", //采集到的手机 mac 地址
    "rssi": "-30", //rssi，手机的信号强度，如 rssi=-75dbm
    "range": "1.0", //手机距离嗅探器的测距距离字段，单位米
    "ts": "hello", //目标 ssid，手机连接的 WIFI 的 ssid
    "tmc": "00:01:02:03:04:05", //目标设备的 mac 地址，手机连接的 WIFI 的 mac 地址
    "tc": "Y", //是否与路由器相连
    "ds": "N", //手机是否睡眠
    "essid0": "七天连锁_wifi", //手机用户 9a:21:6a:7b:62:6a 曾经连接过的 WIFI 的 SSID
    "essid1": "工商银行", //手机用户 9a:21:6a:7b:62:6a 曾经连接过的 WIFI 的 SSID
    "essid2": "东方明珠",
    "essid3": "home",
    "essid4": "abcd",
    "essid5": "xiong",
    "essid6": "XX 会馆"
}, {
    "mac": "1c:31:72:5c:83:6b",
    "rssi": "-69",
    "range": "14.0",
    "ts": "world",
    "tmc": "00:01:02:03:04:06",
    "tc": "Y",
    "ds": "Y",
    "essid0": "七天连锁_wifi", //手机用户 9a:21:6a:7b:62:6a 曾经连接过的 WIFI 的 SSID
    "essid1": "工商银行", //手机用户 9a:21:6a:7b:62:6a 曾经连接过的 WIFI 的 SSID
    "essid2": "东方明珠",
    "essid3": "home",
    "essid4": "abcd",
    "essid5": "xiong",
    "essid6": "XX 会馆"
```

```
}  
    }  
}
```

3.3.2 数据预处理

该部分主要用于对数据进行预处理，格式化数据以及缓存数据。

3.3.2.1 需求描述

功 能 需 求	
功能名称	数据预处理
需求编号	UN001
优先级	高
参与者	系统
业务背景	探针发送的数据可能格式不完整，需要对其格式化
基本需求	接收探针发送的数据 格式化数据
可选需求	无
约束条件	格式化后的数据必须是 json 格式
裁剪说明	不可裁剪

3.3.2.2 数据缓存

功 能 需 求	
功能名称	数据缓存
需求编号	UN002
优先级	高
参与者	系统
业务背景	数据量过大时为避免数据丢失，将数据缓存到 Kafka 中，以供大数据分析服务器实时提取。同时，将所有原始数据保存到 Hive 中，方便数据分析服务器提取离线数据，进行机器学习。
基本需求	将格式化后到数据存储到 Kafka 中 将格式化后到数据存储到 Hive 中
可选需求	无
约束条件	无
裁剪说明	不可裁剪

3.3.3 数据分析

该部分主要用来分析和处理数据

3.3.3.1 区域客流量

功 能 需 求	
功能名称	区域客流量
需求编号	UN003
优先级	高
参与者	系统
业务背景	该功能主要用来维护区域客流量相关的信息。区域客流量指当前时间、当前范围内的 mac 数量
基本需求	、 统计当前时间内的所有探针发送过来的数据中的 mac 地址数量 、 去除重复的 mac 地址
可选需求	无
约束条件	Mac 地址唯一且有效
相关查询	查询区域客流量信息
其他需求	无
裁剪说明	不可裁剪

3.3.3.2 入店客流量

功 能 需 求	
功能名称	入店客流量
需求编号	UN004
优先级	高
参与者	系统
业务背景	该功能主要用来维护入店客流量相关的信息。入店客流量指当前时间、当前范围内的进入店面或商场的 mac
基本需求	1.统计当前时间内的所有探针发送过来的数据中的 mac 地址数量 2.去除重复的 mac 地址 3.读取用户设置的范围参数 4.根据用户设置的参数统计范围小于该数值的 mac 地址
可选需求	根据 mac 到不同探针的距离和探针对应的经纬度，以及用户设置的地图，精确判断顾客是否进入店铺
约束条件	Mac 地址唯一且有效

相关查询	实时展示 根据用户设定的时间查询入店客流量信息 商场管理界面查询入店率信息
其他需求	无
裁剪说明	不可裁剪

3.3.3.3 入店率

功 能 需 求	
功能名称	入店率
需求编号	UN005
优先级	高
参与者	系统
业务背景	该功能主要用来维护入店率相关的信息。入店率=入店客流量 / 区域客流量
基本需求	1.获得当前时间的入店客流量和区域客流量 2.计算入店率
可选需求	无
约束条件	Mac 地址唯一且有效
相关查询	1.实时展示 2.根据用户设定的时间查询入店率信息 3.商场管理界面查询入店率信息
其他需求	无
裁剪说明	不可裁剪

3.3.3.4 来访周期

功 能 需 求	
功能名称	来访周期
需求编号	UN006
优先级	高
参与者	系统
业务背景	统计同一个 mac 地址的来访周期
基本需求	1.获得当前时间的入店客流量 2.根据入店客流量中的每一个 mac 地址查询上一次来访时间 3.计算来访周期
可选需求	无

约束条件	1.Mac 地址唯一且有效 2.永久数据存储使用 Hive
相关查询	1.实时展示 2.根据用户设定的时间查询来访周期信息 3.商场管理界面查询来访周期信息
其他需求	无
裁剪说明	不可裁剪

3.3.3.5 新老顾客

功 能 需 求	
功能名称	新老顾客
需求编号	UN007
优先级	高
参与者	系统
业务背景	统计同一个 mac 地址入店的次数和时间，根据用户设置的值判断为新老顾客
基本需求	1.获得当前时间的入店客流量 2.计算每个 mac 的来访周期 3.读取用户设置的参数 4.判断当前 mac 为新顾客还是老顾客
可选需求	分类算法使用 KNN
约束条件	1.Mac 地址唯一且有效 2.永久数据存储使用 Hive
相关查询	1.实时展示 2.根据用户设定的时间查询新老顾客信息 3.商场管理界面查询新老顾客信息
其他需求	无
裁剪说明	不可裁剪

3.3.3.6 顾客活跃度

功 能 需 求	
功能名称	顾客活跃度
需求编号	UN008

优先级	高
参与者	系统
业务背景	根据顾客的入店次数和来访周期，为顾客活跃度分级
基本需求	获取当前顾客的入店次数 获取当前顾客的来访周期 获取当前顾客的驻店时长 使用机器学习为顾客活跃度分类
可选需求	顾客活跃度分类由用户指定 分类算法使用 Navie Bayes 活跃度划分：高活跃度、中活跃度、低活跃度、沉睡活跃度
约束条件	1.Mac 地址唯一且有效 2.永久数据存储使用 Hive 3.机器学习的数据模拟生成
相关查询	1.实时展示 2.根据用户设定的时间查询顾客信息 3.商场管理界面查询顾客信息
其他需求	无
裁剪说明	不可裁剪

3.3.3.7 驻店时长

功 能 需 求	
功能名称	驻店时长
需求编号	UN009
优先级	高
参与者	系统
业务背景	根据顾客 mac 地址距离探针的范围来判断驻店时长
基本需求	计算当前 mac 距探针的距离，当小于用户设定值时记录为入店； 当该距离大于用户指定值时判断为离店； 计算时长 读取数据库里当累计时长，计算累计驻店时长
可选需求	无
约束条件	1.Mac 地址唯一且有效
相关查询	1.实时展示 2.根据用户设定的时间查询该信息 3.商场管理界面查询该信息

其他需求	无
裁剪说明	不可裁剪

3.3.3.8 跳出率

功 能 需 求	
功能名称	跳出率
需求编号	UN010
优先级	高
参与者	系统
业务背景	跳出率=（入店客流量 - 深访量） / 入店客流量
基本需求	获取客流量 获取深访量 计算跳出率
可选需求	无
约束条件	1.Mac 地址唯一且有效
相关查询	1.实时展示 2.根据用户设定的时间查询该信息 3.商场管理界面查询该信息
其他需求	无
裁剪说明	不可裁剪

3.3.3.9 深访率

功 能 需 求	
功能名称	深访率
需求编号	UN011
优先级	高
参与者	系统
业务背景	深访率=深访量 / 入店量
基本需求	获取入店量 获取用户设置的判断深访参数 计算深访量 计算深访率
可选需求	无
约束条件	1.Mac 地址唯一且有效

相关查询	1.实时展示 2.根据用户设定的时间查询该信息 3.商场管理界面查询该信息
其他需求	无
裁剪说明	不可裁剪

3.3.3.10 手机品牌

功 能 需 求	
功能名称	手机品牌
需求编号	UN012
优先级	高
参与者	系统
业务背景	根据 IEEE 标准，根据每个 mac 地址的前三个字节可以查询到对应的设备商，以此来判断手机品牌
基本需求	将 mac 地址与所有手机厂商的品牌进行解析并存储到数据库里 获取 mac 地址信息 去除重复 mac 地址 根据 mac 地址进行查询
可选需求	无
约束条件	1.Mac 地址唯一且有效
相关查询	1.根据用户设定的时间查询该信息 2.商场管理界面查询该信息
其他需求	无
裁剪说明	不可裁剪

3.3.3.11 数据说明(user_visit 表)

用于存储上述信息的数据库表(user_visit)结构

名称	描述	是否必添
商场 ID	联合主键/int	是
探针 mac	联合主键/varchar	是
时间	联合主键/Timestamp	是
区域客流量	Int	是
入店率	double	是
入店量	Int	是
来访周期	Int	是

新老顾客	int	是
顾客活跃度	int	是
驻店时长	double	是
跳出率	double	是
深访率	double	是

3.3.3.12 数据描述(vendor_macs 表)

用于存储上述信息中的手机品牌与 mac 地址等对应关系的结构表

名称	描述	是否必填
Mac 前缀	主键/Mac 地址的前三个字节 (varchar)	是
品牌名称	手机品牌名称(varchar)	是

3.3.4 数据展示

该部分主要用来维护数据展示数据相关信息。

3.3.4.1 实时展示数据

功 能 需 求	
功能名称	实时展示数据
需求编号	UN013
优先级	高
参与者	系统
业务背景	大数据服务器分析好之后的数据交由前端网页进行实时渲染，要求提供清晰明确、对用户友好的展示方式
基本需求	<p>、 实时展示的数据包括区域客流量、入店客流量、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、手机品牌等数据</p> <p>、 页面刷新方式为异步刷新，建议使用 WebSocket 双全工通信，不要使用 ajax 轮询。</p>
可选需求	<p>使用 React 框架</p> <p>使用 React-redux 做前后端的数据交互管理</p> <p>使用 echarts 3 绘制图标</p> <p>使用 Ant Design 设计 UI</p>
约束条件	浏览器需支持 WebSocket
其他需求	无
裁剪说明	不可裁剪

3.3.4.2 数据查询

功 能 需 求	
功能名称	数据查询
需求编号	UN014
优先级	高
参与者	系统
业务背景	大数据服务器分析好之后的数据交由前端网页进行实时渲染，要求提供清晰明确、对用户友好的展示方式，同时要求用户可以根据需求进行历史数据查询
基本需求	1. 能查询到的数据包括区域客流量、入店客流量、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、手机品牌等数据 2. 以点击日历的方式选择时间，避免用户键盘输入 3. 数据表格支持用户排序、筛选 4. 支持用户搜索特定数据
可选需求	支持用户模糊搜索特定数据
约束条件	浏览器需支持 WebSocket
其他需求	无
裁剪说明	不可裁剪

3.3.5 商场管理

该部分主要用来维护商场相关信息。

3.3.5.1 需求描述

功 能 需 求	
功能名称	商场管理
需求编号	UN015
优先级	高
参与者	用户
业务背景	管理员可以看到旗下各商场的的数据情况并进行对比，方便总部的统一调配，也可以查询每个商场里所有探针的数据情况
基本需求	生成商场列表，包含每个商场的主要信息和数据信息，详细的数据库表结构信息见下文 列表支持排序和检索 支持对各商场进行评分
可选需求	支持用户模糊搜索特定数据

其他需求	无
裁剪说明	不可裁剪

3.3.5.2 数据描述

相应的数据格式如下：

名称	描述	是否必添
商场 ID	用来唯一区别一个商场，可以自动生成（int）	是
负责人 ID	商场的真正拥有者，可以是组织或个人（int）	是
商场名称	商场名称（varchar）	是
地址	商场地址（varchar）	是
联系人	商场具体管理人员人，经理（varchar）	否
电话	商场负责人电话（varchar）	否
评分	分为 1-5 个等级（int）	否

3.3.6 管理员管理

该部分主要用来维护用户相关信息。

3.3.6.1 需求描述

功 能 需 求	
功能名称	管理员管理
需求编号	UN016
优先级	高
参与者	用户
业务背景	该功能主要用来维护用户相关的信息，包括账号管理，用户旗下的商场管理、权限管理
基本需求	登陆 注册 添加商场信息 添加探针 查询数据 权限检查 短信服务
可选需求	通过录入的商场管理人员信息，管理员可以通过短信群发通知
约束条件	管理员的账户必须是手机号 登陆时必须发送短信验证

其他需求	无
裁剪说明	不可裁剪

3.3.6.2 数据描述

为避免数据冗余，查询该管理员旗下商场信息时要使用本表与商场信息表联合查询

名称	描述	是否必添
管理员 ID	用来唯一区别一个管理员，必须为手机号 (varchar)	是
管理员名称	用户称呼管理员	是
管理员密码	使用 sha1 加密存储 (varchar)	是

3.3.7 探针管理

该部分主要用来维护探针相关信息，请参考相关文档。

3.3.7.1 需求描述

功 能 需 求	
功能名称	探针管理
需求编号	UN017
优先级	高
参与者	管理员
业务背景	该功能主要用来维护探针相关的信息。
基本需求	、 设置探针基本参数 、 根据不同商场查看不同探针信息 、 探针列表支持用户检索和排序
可选需求	
约束条件	系统要求区段编号唯一。
相关查询	查询旅程区段信息
其他需求	无
裁剪说明	不可裁剪

3.3.7.2 数据描述

1.探针配置信息

名称	描述	是否必添
探针 ID	用来唯一区别一个探针，可以自动生成 (in t)	是
上传路径	数据上传的路径 (varchar)	是
端口号	数据上传的端口号(int)	是

商场 ID	探针归属的商场(int)	是
发送频率	探针发送数据的频率(int)	是
工作状态	判断探针工作状态 (varchar)	是
累计发送数据	计算探针累计发送的数据量 (int)	是
累计工作时长	计算探针累计工作时长(double)	是
开始工作时间	记录探针第一次上传数据时间(varchar)	是

3.3.8 属性管理

该部分主要用来维护参数相关信息

3.3.8.1 需求描述

功 能 需 求	
功能名称	属性管理
需求编号	UN018
优先级	高
参与者	管理员
业务背景	不同探针在不同商场环境下，判断入店和离店的条件不一样。不同的商家对于新老顾客和顾客活跃度的判断也不一样，这些参数都需要提供给管理员设置的接口
基本需求	<ul style="list-style-type: none"> 、 修改入店范围 、 查询入店范围设置 、 修改新老顾客的来访周期 、 查询新老顾客的来访周期设置 、 设置活跃程度 、 查询活跃程度 、 设置深度访问的时间限定 、 查询深度访问的时间限定
可选需求	提供友好的输入方式，比如滑动轨 分别提供对单个商场和单个探针的设置 制作新手导航，教会用户设置方法
其他需求	无
裁剪说明	不可裁剪

3.3.8.2 数据描述

参数设置信息

名称	描述	是否必添
参数 ID	用来唯一区别一个组参数设置，可以自动生成	是
来访周期	新老顾客的判断标准（int）	是
入店范围	当 mac 探针距离小于该值时判断为入店(int)	是
信号强度	当信号强度过低时忽略该值（int）	是
活跃程度	根据用户出现的次数进行活跃程度的划分(int)	是
深访时间	当用户入店时间大于该值时判断为入店(int)	是

3.3.9 用户管理

3.3.9.1 需求描述

功 能 需 求	
功能名称	用户管理
需求编号	UN019
优先级	高
参与者	管理员
业务背景	每一个 mac 地址对应一个普通用户。
基本需求	根据不同 mac 生成不同的用户 查询用户的手机品牌 记录用户每次到商场的的时间 判断用户来访周期 判断用户累计驻店时长 判断用户活跃等级 用户列表支持排序和检索
可选需求	无
其他需求	无
裁剪说明	不可裁剪

3.3.9.2 数据描述

用户表

名称	描述	是否必添
用户 ID	用来唯一区别一个组用户，可以自动生成（int）	是

Mac 地址	Mac 地址（varchar）	是
平均来访周期	平均来访周期（int）	是
手机品牌	根据 mac 地址前三个字节查询（int）	是
初次到店时间	用户第一次到店时间（int）	是
活跃等级	用户活跃等级（int）	是

用户到店时间表

名称	描述	是否必添
Id	用来唯一区别一个组用户，可以自动生成（int）	是
用户 ID	用户表中的外键（int）	是
时间	到店时间	是

3.4 智能决策

智能决策模块用来为管理员提供商务上的决策，运用人工智能算法对已有数据进行训练和学习，再结合管理员需求生成解决方案。

之前的所有数据展示和分析都是致力于为用户决策提供参考，而本模块将直接为用户提供商业方案。

3.4.1 智能决策管理

该部分主要用来维护用户智能决策相关信息。

3.4.1.1 需求描述

功 能 需 求	
功能名称	数据导入管理
需求编号	UN020
优先级	高
参与者	管理员
业务背景	机器学习、深度学习的飞速发展以及大数据分析的应用为智能决策奠定了基础
基本需求	提供触发决策的接口 接受用户输入数据类型 预测未来一段时间内的流量峰值 生成详细的数据图表 提供详细的决策建议
可选需求	用户自定义决策要求，提供决策约束

约束条件	用户输入的类型包括但不限于区域客流量、入店客流量、入店率、深访率、浅访率等十大数据类型
其他需求	无
裁剪说明	不可裁剪

3.4.1.2 数据描述

预测的流量数据

名称	描述	是否必填
ID	用来唯一区别一个组，可以自动生成（int）	是
类型	预测数据类型（int）	是
时间	预测数据的时间（Timestamp）	是
数据流量	预测数据类型的流量（int）	是
评级	对该数据的评级（int）	是

智能决策

名称	描述	是否必填
ID	用来唯一区别一组决策，可以自动生成（int）	是
类型	决策类型（int）	是
内容	决策内容（varchar）	是

3.5 短信管理

根据用户设置的参数，当某类型流量达到参数值短信提醒用户

3.5.1 阈值设置

该部分主要用来设置监控的参数。

3.5.1.1 需求描述

功 能 需 求	
功能名称	阈值设置
需求编号	UN021
优先级	中
参与者	管理员
业务背景	该部分主要用来设置监控的参数。

基本需求	、 提供参数类型 、 提供设置参数值 、 可添加多组监控
可选需求	提供友好的参数设置方式
约束条件	无
相关查询	查询设置信息
其他需求	无
裁剪说明	不可裁剪

3.5.1.2 数据描述

名称	描述	是否必添
ID	用来唯一区别一组参数，可以自动生成（int）	是
监控类型	监控的流量类型（int）	是
阈值	决策内容（varchar）	是

3.5.2 短信推送

3.5.2.1 需求描述

功 能 需 求	
功能名称	短信推送
需求编号	UN022
优先级	中
参与者	系统
业务背景	当流量达到用户用户设定阈值时系统推送流量提醒短信
基本需求	读取用户存储的手机号（账号） 获取阈值 设置短信内容 获得激活信号 发生短信
可选需求	使用阿里大鱼 API 发送短信
约束条件	无
相关查询	查询短信内容
其他需求	无
裁剪说明	不可裁剪

3.5.2.2 数据描述

名称	描述	是否必添
----	----	------

ID	用来唯一区别一组设置，可以自动生成（int）	是
短信内容	短信提示内容，要有默认值（varchar）	是
手机号	用户 ID(varchar)	是

3.6 自动生成报表

3.6.1.1 需求描述

功 能 需 求	
功能名称	自动生成报表
需求编号	UN023
优先级	中
参与者	管理员
业务背景	管理员可以根据每天或每月的流量信息生成每家商场的报表，方便用户进行管理
基本需求	自由选择时间 自由选择需要导出的流量类型 生成 Excel 报表
可选需求	生成报表使用使用 Apache POI
其他需求	无
裁剪说明	不可裁剪

3.7 日志管理

功 能 需 求	
功能名称	日志管理
需求编号	UN024
优先级	中
参与者	系统
业务背景	日志记录服务器每时每刻的运行情况，在服务器出错时的可以很好的定位错误
基本需求	1.记录日志 2.日志查询
可选需求	使用 log4j 记录日志
其他需求	无

3.8 系统安全性的要求

3.8.1 数据存储安全

管理员密码使用 sha1 算法加密存储

3.8.2 访问控制安全

1. 使用 SSL 加密协议，所有 http 的访问强制重定向到 https
2. 用户 Session 失效时间为两小时，每次访问后台都要检查 Session
3. 数据库管理员账户禁止使用 root
4. 用户 session 失效时所有访问重定向到登陆页面
5. 使用预编译防止 SQL 注入

3.8.3 网络传输安全

1. 使用 SSL 加密协议传输
2. 登陆密码再前端使用 sha1 算法加密后再发送到后端，经后端加盐处理后再保存到数据库

3.8.4 系统约束

1. 必须使用分布式系统，不能是单机环境或者伪分布环境
2. 服务器操作系统使用 CentOS 7

3.8.5 其他专门要求

实时分析的服务器接收数据并发量必须达到 1300 以上

4 运行环境规定

4.1 运行环境

4.1.1 软件环境

数据库：mysql 5.6 及以上

中间件：tomcat 7.5 及以上

JDK：1.5 版本及以上

其它软件环境不作要求

4.1.2 硬件环境

服务器环境：

4 个节点，每个节点 1 核 2G 内存 80GB 硬盘