1. 引言

1.1目的

该文档是关于数据运营系统的功能和性能的描述，重点描述了系统的功能需求，并作为系统设计阶段的主要输入。

本文档的预期读者是：需求分析人员；设计人员；开发人员；测试人员；

1.2项目背景

项目的名称：数据运营系统

项目的提出者：小组成员

开发者：东北大学大数据实验班TOP1小组（王阔、赵磊、张志浩）

用户：产品运营专员

1.3缩写说明

|  |  |
| --- | --- |
| 缩写 | 全称 |
| ODS | 操作数据存储（Operational data storage） |
| DWD | 细节数据层（data warehouse detail ） |
| DWS | 服务数据层（data warehouse service） |
| ADS | 应用程序数据存储（Application Data Store） |
| OLAP | 联机分析处理（Online Analytical Processing） |

1.4术语定义

|  |  |
| --- | --- |
| 术语 | 解释 |
| 新增用户（7日平均） | 最近7日（不含今日）每日新增用户的平均值 |
| 新用户次日留存率（7日平均） | 最近7日（不含今日/昨日）新增用户次日留存率的平均值 |
| 使用时长（7日平均） | 最近7日（不含今日）用户每日使用时长的平均值 |
| 活跃用户（7日平均） | 最近7日（不含今日）每日活跃用户的平均值 |
| 7日总活跃用户数（去重） | 最近7日（不含今日）活跃用户的总数（去重） |
| 30日总活跃用户数（去重） | 最近30日（不含今日）活跃用户的总数（去重） |
| 累计用户数 | 截止到当前时间，启动过应用的所有独立用户（去重，以设备为判断标准） |
| 总错误率 | 每日错误数/启动次数 |
| 新增用户 | 第一次启动应用的用户（以设备为判断标准） |
| 新增账号 | 第一次启动应用的账号 |
| 活跃用户 | 启动过应用的用户（去重），启动过一次的用户即视为活跃用户，包括新用户与老用户 |
| 活跃构成 | 活跃用户中新增用户的占比比例 |
| 活跃粘度 | DAU/过去7日活跃用户，DAU/过去30日活跃用户 |
| 启动次数 | 打开应用视为启动，完全退出或退至后台即视为启动结束 |
| 留存用户和留存用户率 | 某段时间内的新增用户（活跃用户），经过一段时间后，又继续使用应用的被认作是留存用户；这部分用户占当时新增用户（活跃用户）的比例即是留存率。 |
|  |  |
|  |  |

1.5参考资料

列出相关的参考资料。

1.6版本信息

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 修改编号 | 修改日期 | 修改后的版本 | 修改位置 | 修改内容概述 |
| 0 | 2020/06/02 | V1 |  |  |

2、任务概述

2.1系统定义

2.1.1项目背景

现在随着互联网的发展，好多公司的主营业务在线上(app/网站)（外卖公司、各类 app，比如：下厨房；头条；安居客；内涵段子；斗鱼；每日优鲜>、页游、唯品会、拼多多、当当、苏宁 易购、国美在线、网易严选、携程、中华会计网校、多易教育、小象学院、慕课网、尚德教育、有钱花、上哪借 ……）

这类公司，都需要针对用户的线上访问行为、消费行为、业务操作行为进行统计分析，数据挖掘！以支撑公司的业务运营、精准画像营销、个性化推荐等，来提高业务转化率，改善公司运营效果！ 这些需求，都需要通过构建一个**综合数据处理系统**来支撑！

2.1.2项目要达到的目标

这一版本我们主要针对用户的线上访问行为、业务操作行为进行统计分析，建立起从数据采集到结果展示的完整系统，以支撑公司的业务运营来提高业务转化率，改善公司运营效果。在以后的版本中可能会建立用户画像系统来支持精准画像营销和个性推荐。

2.1.3系统整体结构

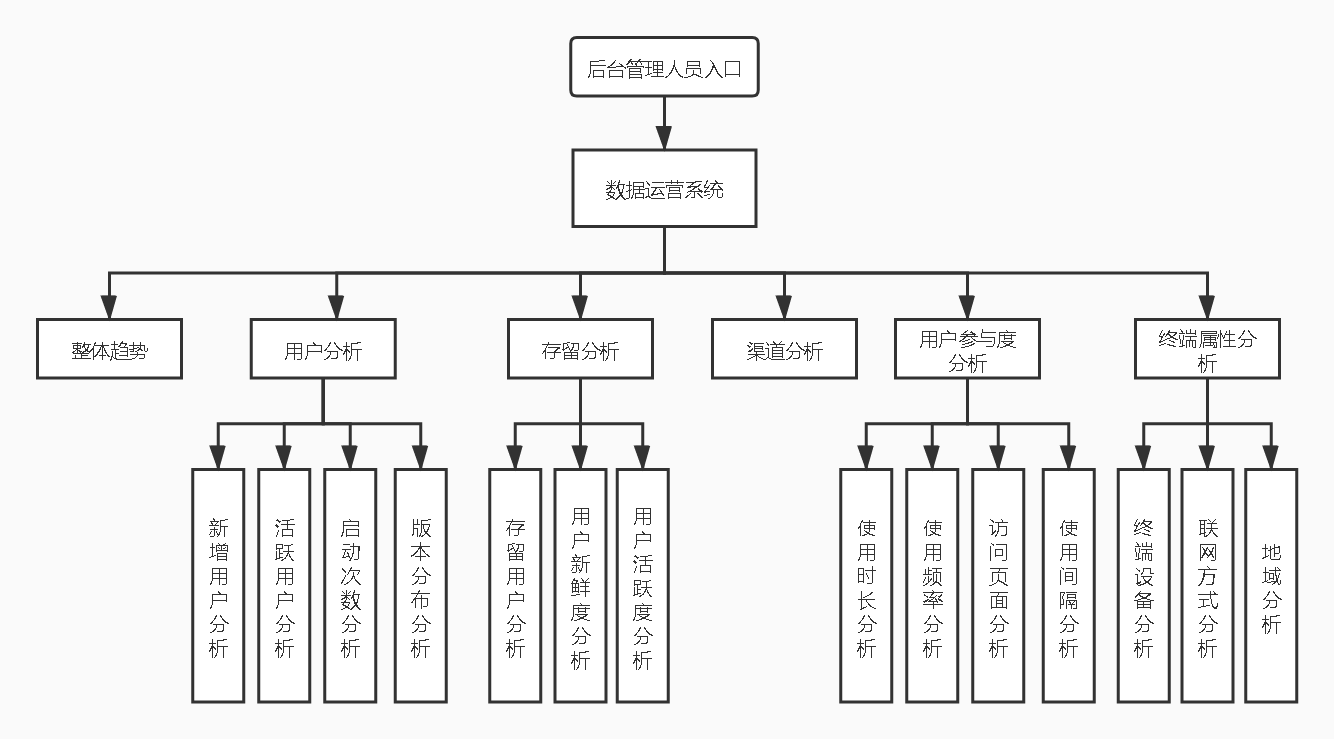


图2.1.3-1系统功能结构图

2.1.4系统各部分组成，与其他部分的关系，各部分的接口等

2.2运行环境

2.2.1设备环境

普通PC机：处理器P4以上，内存1GB以上，支持鼠标键盘工作。

2.2.2硬件环境

普通PC即可。

2.2.3软件环境

无。

2.2.4网络环境

正常的网络连接。

2.2.5操作环境

普通PC浏览器打开本站登录即可。

2.2.6应用环境

2.3条件限制

2.3.1软件开发工作的假定和约束

例如经费限制、开发期限等。

2.3.2软件的最终用户、用户的教育水平和技术专长

用户公司的管理人员针对本软件提供的用户线上访问行为、消费行为、业务操作行为的各种报表进行统计分析以提高业务转化率，改善公司运营效果，要求会看图表，会操作计算机。

2.3.3软件的预期使用频度

因为本软件是分析用户的线上行为来提高公司的运营效果，所以公司的管理人员每日都会关心昨天的用户线上行为，每日都会被使用本软件，但是每日中的使用频次无法估计。

3、数据描述

3.1静态数据

商品信息，订单信息等存储在业务系统数据库

3.2动态数据

App/Web/微信小程序端埋点返回的日志：返回的日志类似于JSON格式文件的结构，其中对象由花括号括起来的逗号分割的成员构成，成员是键值对，如：

{"name": "John Doe", "age": 18, "address": {"country" : "china", "zip-code": "10000"}}

我们接下来以app端日志的某一行来解释数据

{

"eventid":"addShowEvent",

"event":{"adId":"8","pgId":"599","adPosition":"","adType":"","adTitle":"","adSource":"","adResourceID":""},

"user":{"uid":"508288","account":"","email":"","phoneNbr":"18771250825","birthday":"","isRegistered":"","isLogin":"","addr":"","gender":"",

"phone":{"imei":"9065623596907112","mac":"d1-b8-a5-cb-49-23-68","imsi":"5701516260659471","osName":"windows","osVer":"10.0","androidId":"","resolution":"800\*600","deviceType":"HUAWEI\_RY6","deviceId":"u5L6ES","uuid":"aXhLQEbks0RGc4nS"},

"app":{"appid":"com.51doit.mall","appVer":"2.2.8","release\_ch":"应用超市","promotion\_ch":"10"},

"loc":{"areacode":620702400,"longtitude":100.67998669005996,"latitude":38.924405643707416,"carrier":"ISP03","netType":"WIFI","cid\_sn":"439981747513","ip":"104.86.115.114"},

"sessionId":"sid-6b37a9bd-1072-4ee9-a092-eb3ccab3c828"},

"timestamp":"1575533430000"

}

|  |  |
| --- | --- |
| Key | 解释 |
| Eventid | 事件id |
| Event | 事件 |
| Pgid | 页面id |
| adPosition | 广告位置 |
| adType | 广告类型 |
| adTitle | 广告标题 |
| adSource | 广告源 |
| adResourceID | 广告源id |
| User | 用户 |
| uid | 用户id |
| Account | 账户 |
| email | 邮箱 |
| phoneNbr | 电话号码 |
| birthday | 出生日期 |
| isRegistered | 是否注册 |
| isLogin | 是否登陆 |
| addr | 地址 |
| gender | 性别 |
| phone | 电话 |
| imei | 国际移动设备识别码 |
| mac | MAC地址 |
| imsi | 国际移动用户识别码 |
| osName | 操作系统名字 |
| osVer | 操作系统版本 |
| androidId | [Android设备唯一标识](https://www.baidu.com/link?url=ifREgME3tRBs6rvxLpT1H1LtySXuqV5bLWKtRWQIM4_07N32HcwtPOZ9h_9Tszxd&wd=&eqid=9a30e1760001a551000000035ed70d09) |
| resolution | 屏幕分辨率 |
| deviceType | 设备类型 |
| deviceId | 设备id |
| uuid | 通用唯一识别码（Universally Unique Identifier） |
| app | 手机软件 |
| appid | 软件id |
| appVer | 软件版本 |
| release\_ch | 发行渠道 |
| promotion\_ch | 升级渠道 |
| loc | 位置 |
| areacode | 地区代码 |
| longtitude | 经度 |
| latitude | 纬度 |
| carrier | 运营商 |
| netType | 网络类型 |
| cid\_sn |  |
| ip | 网际互连协议（Internet Protocol） |
| sessionId | 会话id |
| timestamp | 时间戳 |

3.3数据库描述

MySQL ：是一种开放源代码的关系型数据库管理系统（RDBMS）。

3.4数据字典

1) 地理位置字典（spark 程序、高德地图 api）、

2) 页面信息字典（爬虫、NLP-自然语言处理技术-分词-关键词抽取算法 TF-IDF）、 页面 id，url，标题，频道/栏目/品类，关键词

3) GUID 字典（spark graphx）

3.5数据采集

系统运行时需要不断对业务系统数据库进行读写操作，并实时读取埋点日志服务器中的日志。

4、功能需求

4.1功能划分

4.1.1系统功能组成

主要功能模块：

1）整体趋势

2）用户分析

3）存留分析

4）渠道分析

5）用户参与度分析

6）终端属性分析

4.1.2功能编号和优先级

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 名称 | 优先级 | 主要发起者 |
| 0101 | 整体趋势 | 重要 | 用户 |
| 0201 | 新增用户 | 重要 | 用户 |
| 0202 | 活跃用户 | 重要 | 用户 |
| 0203 | 启动次数 | 重要 | 用户 |
| 0204 | 版本分布 | 重要 | 用户 |
| 0301 | 留存用户 | 重要 | 用户 |
| 0302 | 用户新鲜度 | 重要 | 用户 |
| 0303 | 用户活跃度 | 重要 | 用户 |
| 0401 | 渠道列表 | 重要 | 用户 |
| 0501 | 使用时长 | 重要 | 用户 |
| 0502 | 使用频率 | 重要 | 用户 |
| 0503 | 访问页面 | 重要 | 用户 |
| 0504 | 使用间隔 | 重要 | 用户 |
| 0601 | 设备终端 | 重要 | 用户 |
| 0602 | 网络及运营商 | 重要 | 用户 |
| 0603 | 地域 | 重要 | 用户 |

4.2功能描述

0101整体趋势：

基础统计数值：包括7日平均新增用户、活跃用户等，一周内、一个月内的统计量总数，总用户数。

指定时间区间内（默认30天内每日）：

新增用户、活跃用户、启动次数、累计用户

折线图，明细数据列表

Top版本环形图：新增用户、活跃用户、累计用户每个版本占比

Top渠道环形图：新增用户、活跃用户、累计用户每个渠道占比

0201新增用户

指定时间段、指定渠道和版本，每日新增用户折线图、明细数据表格

次日存留率折线图、明细

0202活跃用户：

指定时间段、指定渠道和版本下，

活跃趋势、活跃构成、活跃粘度、分时活跃用户、周，月活跃度

0203启动次数：

指定时间段、指定渠道和版本下，每（小时、天、周、月）的启动次数

0204版本分布：

指定时间段、版本下， 每天新增用户、活跃用户、启动次数折线图

今日、昨日截至今日版本累计用户(%)，新增用户，活跃用户(%)，启动次数

0301留存用户

指定时间段、指定渠道和版本下，每一个时间段（天、周、月）新用户和活跃用户数在接下来一段时间（天、周、月）的留存数/留存率，以表格形式和折线形式呈现。

0302用户新鲜度：

报表展示每天活跃用户的成分构成，并提供用户成分分析控件做进一步的分析。某日的活跃用户来源于当天新增用户、1天前新增用户...30天前新增用户、30+天前新增用户。

0303用户活跃度：

报表展现每个天级时间点的当日活跃用户的活跃程度。

0401渠道列表：

指定时间段、指定版本，各渠道新增用户、活跃用户、启动次数

0501使用时长：

指定某一天、指定渠道和版本下，单次使用时长分布柱形图，明细表格。

该天每个活跃用户使用时长分布柱形图、明细表格。

0502使用频率：

指定日期，指定版本、渠道，当日、上周、上个月使用次数分布柱形图，明细表格。

0503访问页面：

指定时间段（一天、一周、一个月）、指定渠道和版本下，访问页面分布柱形图，明细表格。

0504使用间隔：

查看任意30天内用户相邻两次启动间隔的分布情况，并可以进行版本、渠道及分群的筛选。以柱形图、明细表格形式展示。

0601设备终端：

指定时间段（一天、一周、一个月）、指定渠道和版本下 ，top10机型、分辨率、操作系统的新增用户/启动次数柱状图、明细表格。

0602网络及运营商：

指定时间段（一天、一周、一个月）、指定渠道和版本下 ，各种联网方式的新增用户/启动次数柱状图、明细表格。

0603地域：

指定时间段（一天、一周、一个月）、指定渠道和版本下 ，各省市的新增用户/活跃用户/启动次数柱状图（top10省）、明细表格（省市）。

5、性能需求

5.1数据精确度

严格验证每个数据的录入保证数据录入的无误，显示用户数据要和原始数据一样无差别。

5.2时间特性

在网络通常的情况下用户点击网页后5s内必须出结果

5.3适应性

网页支持主流浏览器的正常加载显示，控制可视化图表和报表显示正常。

6、运行需求

6.1用户界面

6.1.1界面风格

6.1.2界面描述和样式

6.2硬件接口

6.3软件接口

6.4故障处理

列出可能的软件、硬件故障以及对各项性能而言所产生的后果，对故障处理的要求。

7、其他要求

7.1检测或验收标准

列出故障率、出错率等验收标准。

7.2可用性、可维护性、可靠性、可移植性要求

7.3安全保密性要求

7.4开发要求