用户行为的深度追踪

——用户行为分析平台

概要设计说明书

项目名称：用户行为的深度追踪

——用户行为分析平台

组 长：李冠增

成 员：易才华、李志成、李翔

完成日期：2018年3月27日

# 一、引言

1.1 编写目的

为了更加清晰的展示项目的结构，小组成员对该平台进行了概要设计。主要是基于以下目的编写此说明书。

1、对用户行为的深度追踪—用户行为分析平台概要设计的阶段任务成果形成文档，以便阶段验收、评审，最终的文档验收。

2、对需求阶段的文档再次确认过程，对前一阶段需求没有做充分或错误的提出修改。

3、明确整个系统的功能框架和数据库结构，确定后续工作的进行。

4、明确编码规范和命名规范，确定系统个环节工作任务。

1.2 定义、专业术语

UI：即User Interface（用户界面）的简称，也称人机界面。是指[用户](http://baike.baidu.com/view/200012.htm)和某些系统进行[交互](http://baike.baidu.com/view/383268.htm)方法的集合，这些系统不单指电脑程序，还包括某种特定的机器，设备，复杂的工具等。

UI设计：UI设计则是指对软件的[人机交互](http://baike.baidu.com/view/523297.htm)、操作逻辑、[界面](http://baike.baidu.com/view/240704.htm)美观的整体设计。是一个不断为最终用户设计满意视觉效果的过程。

MVC：是模型(model)－视图(view)－控制器(controller)的缩写，表示一个三层的结构体系。是一种软件设计典范，用一种业务逻辑、数据、界面显示分离的方法组织代码，将业务逻辑聚集到一个部件里面，在改进和个性化定制界面及用户交互的同时，不需要重新编写业务逻辑。

1.3.读者对象

该文档读者对象为：

详细设计人员

编码人员

1.4.参考资料

[1]《基于大数据平台的用户画像与用户行为分析》

[2] 《软件测试技术概论》 古乐 史九林编著 /清华大学出版社

# 二．范围

2.1. 系统主要目标

通过埋点采集用户在网站上的行为数据，通过flume收集的采集数据，然后从多维度分析用户使用网站的情况，数据通过前台直观的报表展示，可以给网站运营人员提供真实的数据建议，改进业务线，比如通过分析用户的活跃程度，给出了新增，活跃用户，沉默用户等等，通过这些数据，可以更加准确的判断网站运营情况。

2.2. 主要软件需求

1).搭建高可靠、高可用的实时计算框架和离线分析框架。

2).实时计算每个接口PV;

3).离线分析出用户行为和终端访问情况，并且事对件和流量情况统计。

4).对用户的请求做到毫秒级响应。

2.3. 应用对象

1).网站运营人员，通过系统的数据分析结果了解网站发展状况，发现营销产品和运营中可能存在的问题,提高网站亮点和优化用户体验。

2).需求分析人员，通过系统的分析结果，更好的了解用户需求，减小用户流失，洞察数据实现业务增长。

3).项目经理，通过系统分析结果， 实现更精细和精准的运营与营销，让产品获得更好的增长。

2.4. 应用环境

1). 操作系统 ：Windows和Linux

2).数据库： redis，MYSQL

3）web服务器：tomcat

2.5. 创意描述

1).埋点日志采集，模仿百度统计实现无埋点数据分析，简单、便捷、快速分析出用户的行为。

2).实时计算和离线分析相互结合，多维度交叉分析

3).多指标分析，全方位了解用户行为，减少用户流失

2.6. 功能简介

1).用户分析：分析用户的新增、活跃、沉默等等的情况指标，同时分析出用户的留存情况。

2).访客流量分析：分析出用户群、访问量、浏览量、访问者分析等等。

1. .终端分析：分析用户访问来源的机型、来源域名、操作系统。
2. .事件分析：分析出按钮的点击次数等等，并且支持多维度交叉分析等。
3. 报表统计： 包括日、周、月、年等多维度的时间周期各分析统计，支持报表导入导出。

2.7. 特色综述

1).分析指标齐全，分析速度快。

2).能够多维度地查询接口信息。

3).毫秒级响应时间。

**三．系统总体设计**

3.1 复审数据流，控制流

复审需求分析阶段对软件的功能需求的描述是否合理和精化。

3.2 软件体系结构

通过变换映射（如基于DFD或基于IDEF0的结构设计），并采用适当的优化准则进行软件结构的优化。改过程可在本节中分段进行描述。

已得到优化的软件体系结构用结构图进行描述。

3.2.1 系统架构



*图3-2-1-1 系统架构图*

3.2.2 技术选型

java后端

前端数据展示

数据 收集

分析

平台:Centos(数据分析和后端) + Windows 10 (前端)

工具: Idea + WebStorm

开发工具包：JDK(Java集成开发工具包)

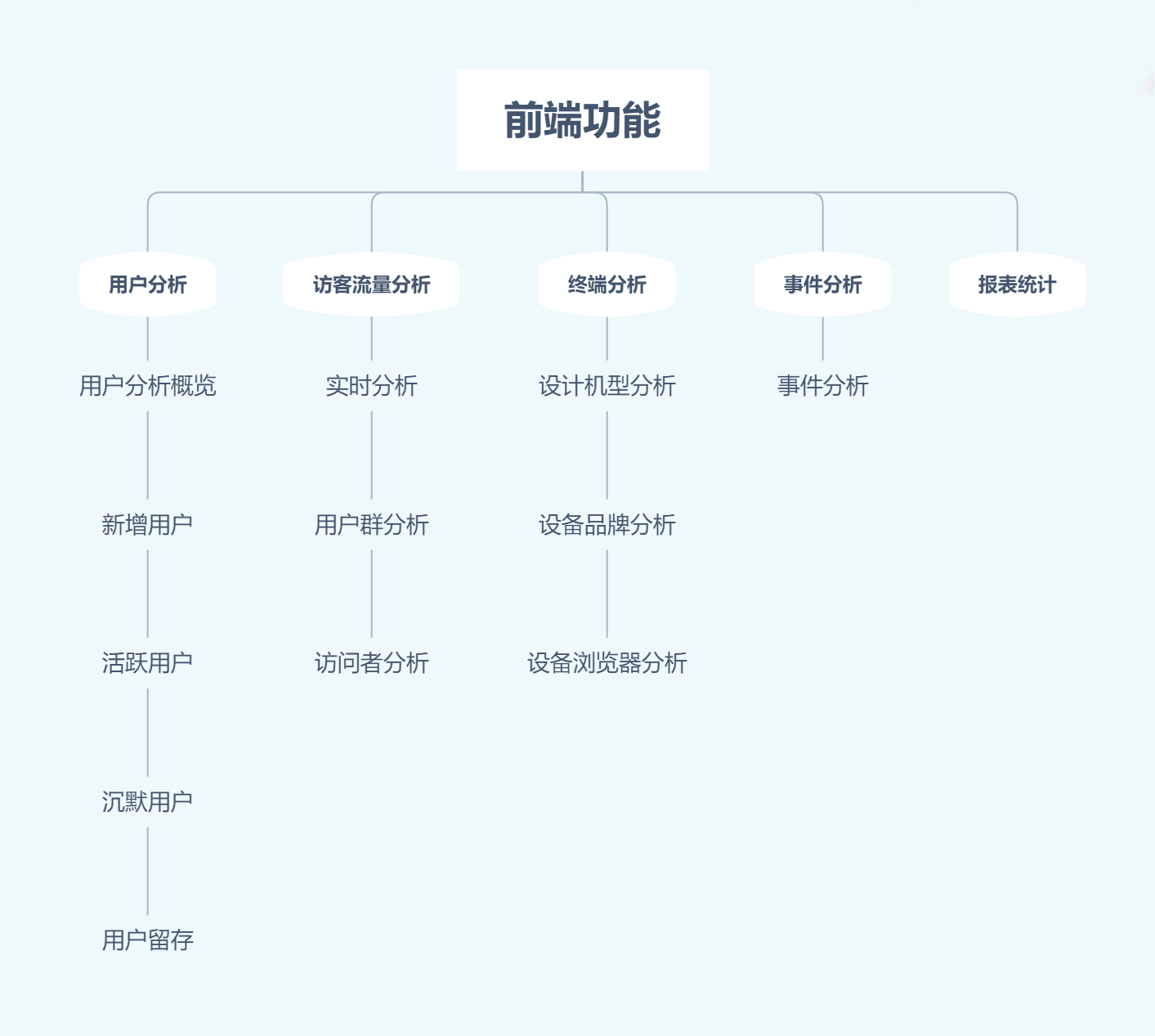
技术:

数据采集分析: flume数据采集,kafka消息队列,hdfs分布式文件系统,sparkstrom实时分析，hive数据仓库工具

后端处理: springboot, mybatis持久化层，redis分布式缓存,mysql数据库

前端展示: angular, echarts图表

3.2.3 前端功能结构



*图3-2-3-1 功能结构图*

3.2.4 模块命名规则

FXXYY00

a).F 1位，代表功能；

b).XX 2位，代表子系统编号；

c).YY 2位，代表子系统下的一级模块编号；

d).00 2位，代表二级及二级以下模块的流水编号；

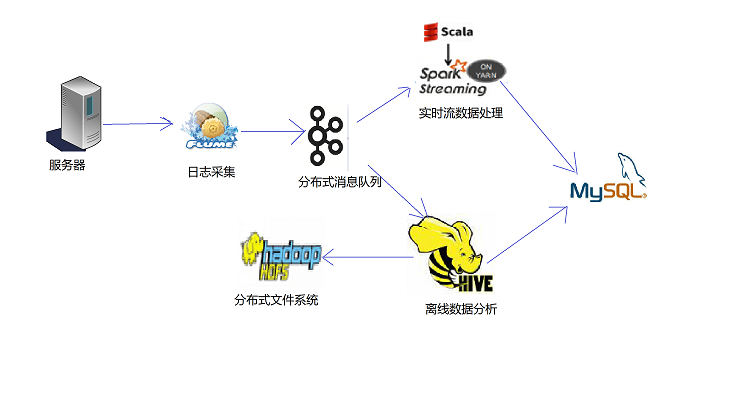
3. 3 基本设计概念和处理流程

3.3.1 WEB前后端基本设计概念和处理流程



*图3-3-1-2 web端流程图*

3.3.2 大数据基本设计概念和处理流程



*图3-3-2-1 数据采集分析流程图*

3. 4 复用策略

说明软件可复用的数据、模块和接口。复用指可利用的现有构件资源和软件中可抽取的新建构件。

**四、 数据库设计**

4. 1 数据字典复审

对需求分析阶段形成的数据字典进行复审和精化。

4. 2 数据项

数据库表：

**Mysql 中**

用户群分析表(tb\_area\_map)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| area | 地区 | VarChar | 10 | 否 |  |
| browse\_num | 浏览量 | bigint | 20 | 否 |  |
| access\_num | 访客量 | bigint | 20 | 否 |  |

事件分析表(tb\_event)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| event | 事件 | varchar | 50 | 否 |  |
| count | 数量 | bigint | 20 | 否 |  |

实时用户访客分析表(tb\_real\_time)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| browse\_num | 浏览量 | bigint | 20 | 否 |  |
| access\_num | 访客量 | bigint | 20 | 否 |  |

终端品牌分析表(tb\_terminal\_brand)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| brand | 品牌 | varchar | 50 | 否 |  |
| browse\_num | 浏览量 | bigint | 20 | 否 |  |
| access\_num | 访客量 | bigint | 20 | 否 |  |
| avg\_stay\_time | 平均访问时长 | bigint | 20 | 否 |  |

终端浏览器分析表(tb\_terminal\_browser)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| browser | 浏览器 | varchar | 50 | 否 |  |
| browse\_num | 浏览量 | bigint | 20 | 否 |  |
| access\_num | 访客量 | bigint | 20 | 否 |  |
| avg\_stay\_time | 平均访问时长 | bigint | 20 | 否 |  |

终端类型分析表(tb\_terminal\_system)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| system | 系统类型 | varchar | 50 | 否 |  |
| browse\_num | 浏览量 | bigint | 20 | 否 |  |
| access\_num | 访客量 | bigint | 20 | 否 |  |
| avg\_stay\_time | 平均访问时长 | bigint | 20 | 否 |  |

用户分析表(tb\_user\_analysis)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| new\_user\_num | 新增用户 | bigint | 20 | 否 |  |
| active\_user\_num | 活跃用户 | bigint | 20 | 否 |  |
| silent\_user\_num | 沉默用户 | bigint | 20 | 否 |  |
| retain\_user\_num | 留存用户 | bigint | 20 | 否 |  |

访问者年龄分析表(tb\_visitor\_age)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| age | 年龄 | tinyint | 4 | 否 |  |
| count | 数量 | bigint | 20 | 否 |  |

访问者性别分析表(tb\_visitor\_gender)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| gender | 性别 | tinyint | 4 | 否 |  |
| count | 数量 | bigint | 20 | 否 |  |

访问者职业分析表(tb\_visitor\_job)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 | 可否为空 | 是否为主外键 |
| time | 日期 | bigint | 20 | 否 |  |
| job | 职业 | varchar | 50 | 否 |  |
| count | 数量 | bigint | 20 | 否 |  |

**Redis中**

USER存储表如下（user）

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 |
| key | 用户编号 | varchar | 100 |
| value | 用户信息 | json | 255 |

实时分析存储表如下（realtime）

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 |
| key | 主键 | varchar | 100 |
| k | 日期 | varchar | 100 |
| value | 实时数据 | json | 255 |

数据缓存表如下（data）

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 描述 | 类型 | 长度 |
| key | 日期 | varchar | 50 |
| k | 类型 | varchar | 50 |
| value | 数据 | varchar | 255 |