

文件名称(Document Name)				
交泰金管家数据分析平台咨询方案				
文件说明(Document Description)				
交泰金管家大数据分析平台项目成型及落地，需要获得一定程度上的资源支撑。技术方案上需要获得系统及技术架构，数据库及数据库结构的支持；另外还需要获得第三方数据平台的支撑。产品需要给整个大数据平台提供信息架构及业务架构的支撑。咨询方案主要目的在对于大户数分析平台的技术、第三方支撑及采购方案。				
文档状态(status)	初稿 []	文档标识	JT-JGJ-PD.INF_PMD002	
	评审通过 []	当前版本	V0.0	
	修改 []	作者	陈志坚	
	发布 []	公司/部门	产品部	
	版本 []	完成日期		
编辑历史(History Version)				
编辑时间	版本	作者	编辑内容	标记
20190411	V0.0	陈志坚	新增	
20190412	V0.1	陈志坚	跳转目录结构 新增消息及交易分析内容	修改，增加红色字体； 调整看文档标注

目录

1. 概述.....	4
1. 数据分析目的	4
2. 数据分析内容	4
3. 系统架构支持	4
2. 数据采集	6
1. 数据埋点.....	6
2. 数据埋点位置	7
3. 数据处理	8
4. 数据分析	9
1. 用户行为分析(部分支持).....	10
2. 内容统计分析(部分支持).....	12
3. 消息统计分析(不支持)	14
4. 技术接口分析(不支持)	15
5. 用户留存活跃(部分支持).....	15
6. 转化分析(不支持)	17
7. 咨询分析(不支持)	17
8. 财务分析	18
5. 数据展现	19
6. 数据挖掘	20
7. 第三方数据解决方案及采购计划	20
1. 第三方数据采集平台	21

2. 应用发布市场计划.....	22
------------------	----

1. 概述

交泰金管家大数据分析平台的构建依赖于架构设计提供的服务支撑，保障数据稳定性，数据流畅性；数据库及数据结构，信息和内容架构为数据平台提供业务逻辑及业务架构，是数据业务的基础。大数据分析其实就是信息和数据的汇总。

为什么要开展数据分析？通过金管家大数据分析平台，我们可以解决什么问题？我们可以从哪几个地方下手，着重那几个点的分析？

1. 数据分析目的

交泰金管家 App，商业渠道、运营环境已经基本成型，产品也已经进行了多次的迭代计划，对于质量、效果及反馈，暂时没有一定的数据进行支撑。通过金管家大数据分析平台，我们可以构建一套销售、市场、运营推广及产品成效的评判标准，以及获得用户反馈，分析用户需求的通道。

2. 数据分析内容

基于交泰金管家大数据分析平台的分析目的，我们可以从用户行为、业务分布、运营活动以及会员销售情况切入交泰金管家的数据分析。我们可以从用户在 App 应用里面的行为路径、活跃留存，相关业务在产品整体环境下的效益数据，社区数据的情况，运营活动曝光，达成的收益，会员销售情况在不同业务之下的效果等等。

3. 系统架构支持

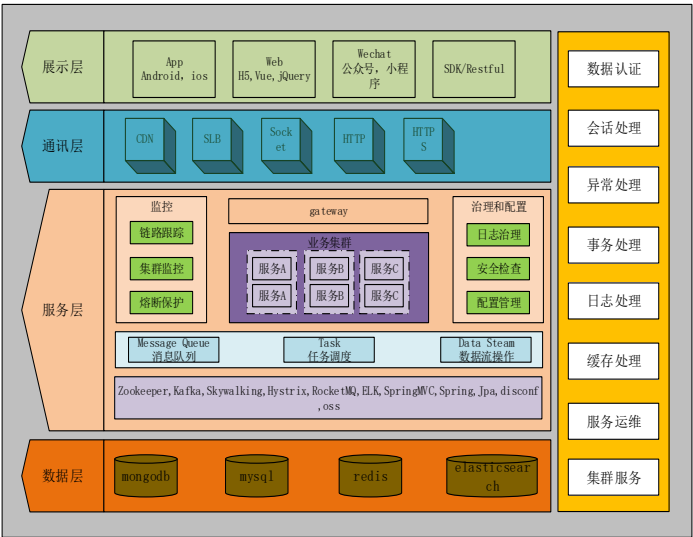
通用的架构设计是保障整个应用、服务等相关产品的高可用性、高扩展性及高并发的，架构设计的三架马车。需要通过进行部署来实现架构方案的落地，操作系统、数据库、日志、缓存、熔断及索引的构建，保障业务的输入输出、处理运行。

交泰金管家选用的架构师 SOA 架构

面向服务的架构是一种软件体系结构，应用程序的不同组件通过网络上的通信协议向其他组件提供服务。通信可以是简单的数据传递，也可以是两个或多个服务彼此协调连接。通过分布式、单独维护和部署的软件组件的集成来组成应用程序。通过技术和标准的实现使得组件能够更容易地通过网络（尤其是 IP 网络）进行通信和协作。SOA 架构中有两个主要角色：服务提供者（Provider）和服务使用者（Consumer）。

交泰金管家的产品形态，架构的展现层客户端是 App 应用界面及 web 网页页面。

在交泰金管家大数据分析平台咨询方案当中,我们需要去考虑的架构设计及支持部分包括了客户端和服务之间的通讯，服务层的监控及日志配置管理以及整个数据层。



- 通讯:客户端之间的关系、业务间的通讯关系可被采集统计。保障数据透传到数据平台的通道
- 监控:在服务断连之后，熔断机制仍能保证业务的运作及数据采集的正常运行。
- 消息队列:消息触达率的保障及维护
- 日志配置:通过日志打印及配置去给整个业务系统埋设数据采集点

技术架构的选型在前端采用的是 Html5 及应用原生代码、后台主要编码方式是 Java，数据采用的是 mySQL 语句。技术架构的选择对大数据平台的方案设计提供技术设计支持。

风险:通过和架构师的沟通，目前架构设计中的熔断机制(Hystrix)暂时还没有起到作用，当前咨询方案中，可能会存在有局部及小范围的故障风险。暂时没有在架构设计没有找到文件服务，容器部署部分的设计方案。

2. 数据采集

信息线索，嫁接着内容、数据、页面之间的逻辑分布关系，便于数据平台梳理数据事务，建立算法公式，输出数据模型结果及文案。对数据进行结构化、异构化处理，从海量动态数据中，梳理出不完整数据、冗余数据及与当前事务管理并不是特别强的稀疏数据，并找到去除数据噪音，形成正确的统计或逻辑算法公式，输出易读性、可读性更高的模型结果及文案。

1. 数据埋点

数据埋点方案，主要透过三种方案完成落地实现。

其一，通过 GrowningIO(增长黑客)、友盟、神策等数据平台上，下载的数据埋点的不同端口的 SDK，对于整个应用界面，页面路径及页面层级埋设数据采集点；

其二，对于第三方数据采集平台进行服务部署方案的调研，查看业务数据最后汇总的路径是在我们自己数据平台上，还是存储在他们的平台之上。如果是的话，需要自己构建爬虫系统程序，去爬取我们的内容管理系统及数据字典构建的元数据及业务数据库，在当中去埋设数据采集点。会存在一定的遍历触达不到的风险，另外也对于系统反爬虫机制的开放，导致系统安全性，业务数据暴露等风险的存在。

其三，采用系统架构的 ETK 部署方案选型，在进行系统测试、运营及打印的时，进行日志打印，将数据埋点带入到不同的界面、业务模块下面。

通过技术和产品的沟通合计之后，我们最后选择了两种落地实现方案来实现我们数据收集，

数据埋点的方案通过第三方数据平台来进行埋点,以及对自己本身业务数据及业务系统进行日志打印的方案。

2. 数据埋点位置



埋设数据采集点的位置放在每个导航、标签以及组织导航、标签之间的业务层级,业务关系的地方进行数据埋点。将这些埋点位置的数据透传给交泰金管家大数据分析统计平台,和交泰金管家管理后台、业务数据库进行统计汇总,最后形成分析方案、图形、表格。

其次，需要在用户下载、安全、登录及注册、生成的运营活动 H5 页面站外链接及页面等内外部环境，及整个交泰金管家 App 的应用界面及 web H5 页面、页面层级、页面关系、页面链接以及页面路径进行数据埋点。方便对于用户行为、活跃留存进行数据汇总，统计分析。为产品、运营做决策提供数据支持，以及维度和方法。

3. 数据处理



对从第三方数据采集平台和(或)交泰金管家数据平台采集回来的数据进行提取，转换成数据库选型方案所需要的格式，加载到数据库之内。对数据进行清理，补缺、替换、格式化、键对值约束；合并、拆分，当中最重要和核心的是需要对数据进行颗粒度的切分，拆的越细越好，校验采集计算方案，建立事件、指标、维度评估体系。

首先，交泰金管家大数据分析平台，我们首先需要构建数据事件，产生事件行为动作。可以

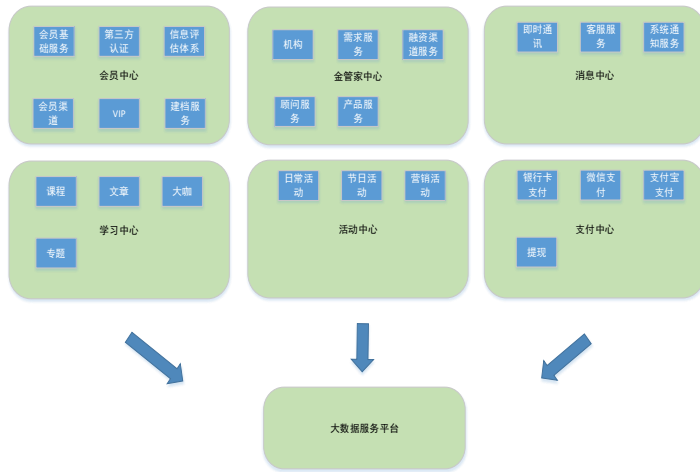
包括下载、安装、登录、注册、启动(访问)、浏览、停留、查询、分享、发布、购买等数据行为事件。

其次，我们要给这些事件赋予一定的属性，帮助我们构建去判断数据事情的维度。可以包括日期、时间、时长、层级数量、次数、人数、金额、到达率、速率、字节等等。他们在我们系统之内，可以提现为一定的单位 unit，我们也需要将数值单位定义给事件属性，时分秒(H\M\S)、年月日(Y\M\D)、小时(包括时段)、天(包括日期区间)、周(近 7 天)、月(近 30 天)、季(近 90 天)、年(近 365 天)、次(万次)、人(万人)、元(万元)、倍数、百分数、百分比、百分点、频率、比例、比例等等。

最后，我们把事件和事件属性进行结合分析，形成事件维度与指标。也需要和交泰金管家大数据分析埋点结合。下载、安装、注册、登录的增长指标(新增、净增、累计)、不同渠道及终端的增长指标；用户行为路径的活跃和留存指标、包括不同终端及渠道增长指标；内容分类、内容标签、具体内的增长及社区指标(增长维度及曝光)；消息\通讯(\业务功能分析内容)的时效、触达指标、状态、增长等指标；交易的增长、转化、成交、区域、日期时间等维度。

4. 数据分析

在架构设计方案中最底下，有对于大数据分析平台筹备方案，里面主要描述了对业务数据的采集需求，包括：



会员中心、消息中心、学习中心、活动中心、支付中心及核心业务金管家中心。

所有的采集都是针对业务数据来进行设计的，没有具体提出需要采集的科目、维度、指标及事件，也没有提出需要统计分析的事务及模型的构建。

目前的数据及架构的设计当中，有使用到第三方数据采集平台友盟的数据监控服务。

透过查看数据库字段设计及结构设计，有囊括到系统安全及数据安全之间的设计补充。账户锁定、密码错误次数机制、交易流水及对接、三方鉴权等等。

数据库设计当中有对于用户行为路径数据的采集埋点，需要产品输出模型需求设计方案。

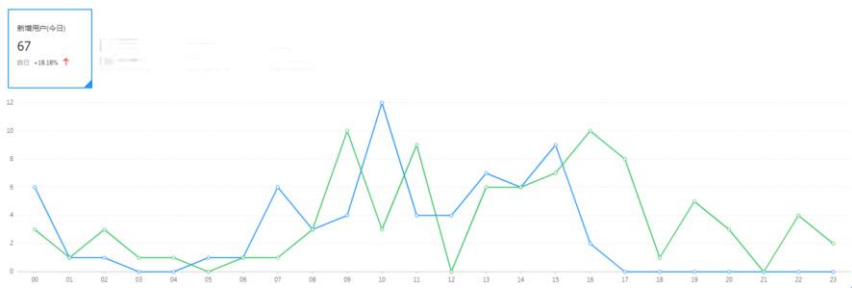
1. 用户行为分析(部分支持)

用户行为路径基础采集数据:

Field	Type	Comment
-------	------	---------

id	int(16) NOT NULL	主键
user_id	int(11) NOT NULL	用户 ID
last_login_phone_model	varchar(20) NULL	最后一次登录手机型号
last_login_phone_version	varchar(20) NULL	最后一次登录手机系统版本
last_login_device_type	varchar(10) NULL	最后一次登录设备类型 1 pc 2 h5 3 Android 4 iOS
last_login_ip	varchar(50) NULL	最后一次登录 ip
last_login_time	datetime NULL	最后一次登录时间
last_login_city_id	int(11) NULL	最后一次登录城市 ID
last_login_city_name	varchar(20) NULL	最后一次登录城市名称
last_coordinate_longitude	varchar(20) NULL	最后一次定位经度
last_coordinate_latitude	varchar(20) NULL	最后一次定位纬度
last_coordinate_time	datetime NULL	最后一次定位时间
last_coordinate_city_id	int(11) NULL	最后一次定位城市 ID
last_coordinate_city_name	varchar(20) NULL	最后一次定位城市名称
last_app_push	varchar(200) NULL	用户登陆 appPush 标识
create_time	datetime NOT NULL	创建时间
is_deleted	int(2) NULL	是否删除 0 否 1 是
update_time	timestamp NULL	修改时间

透过数据库里提供的以上采集数据，当前可以直接实现的数据统计分析的模型为



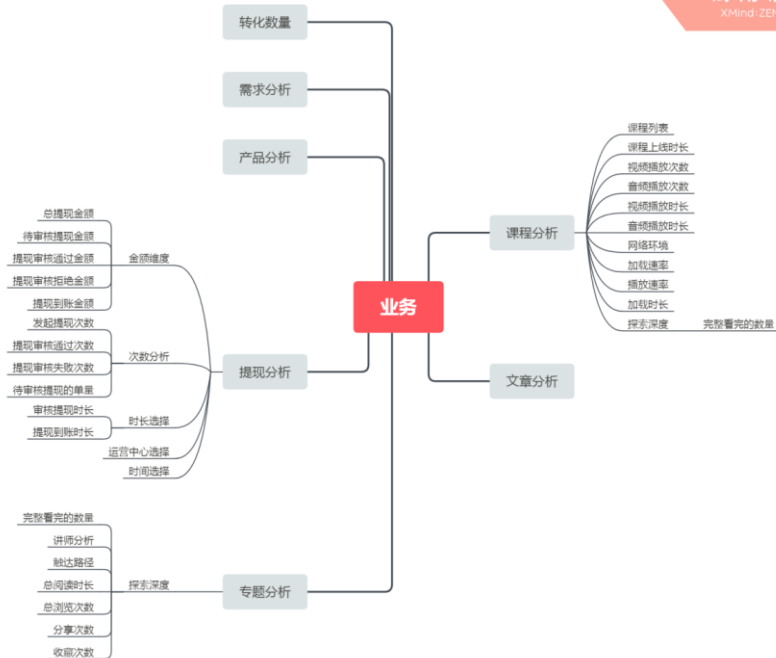
事务	数据模型	分析说明	科目&维度计算公式
		新增、累计、净增	1.累积=所有新增数量之和；净增=当日新增数量-昨日新增数量
用户行为	下载	不同终端渠道维度的指标	2.终端指标维度选择、系统\系统版本\设备品牌\设备机型
		不同日期时间维度的指标	3.每日趋势变化，按照昨天/近 7 天/近 30 天/近 90 天/近 365 天

		4.日时段变化趋势，24 小时曲线趋势及 2 时段\3 小时段\4 小时段趋势
	不同下载渠道维度的指标	5.应用市场数据接入，不同应用市场的下载量趋势 6.服务器生成链接，不同浏览器厂商及版本下载趋势
	物理位置	城市下载数量分布趋势
	下载激活率	7. (下载数量+安装数量+注册用户) /下载数量*100
安装	安装激活率	8. (下载数量+安装数量+注册用户) /安装数量*100
	增长、终端、日期时间	
	不同的网络环境维度的位置	9.4G/3G\2G\WiFi 不同网络下的用户注册数量
注册(用户)	下载激活率	10. (下载数量+安装数量+注册用户) /注册数量*100
	增长、终端、日期时间	
	单次访问用户	11.当日有注册并登陆，后续就没有登陆访问行为的用户
登录(访问)	增长、终端、日期时间、网络环境、物理位置	

当前咨询方案中，缺失下载渠道数据来源，需要提交到应用市场以及生成服务器地址链接，并进行页面路径的数据埋点布局才会有这些数据，具体的应用市场的相关采购方案，在后续文档中国会有提及。

2. 内容统计分析(部分支持)

通过对前端需求、信息架构及交互设计的分析，对于课程内容及提现，我们可以做的分析是大师、文章、专题、产品及顾问的趋势、数量增长分析；



只做页面点击、跳转等交互事件的数据采集，仅是统计相加的趋势统计分析。暂时没有办法去获取第三方分享触达率、付费、产品及需求的交易、漏斗转化、提现的到账率等暂时没有办法实现，非富媒体的内容没有办法进行浏览(停留)时长的判断。主要是当前的资源不足，无法支撑该统计分析场景的支持。这部分内容，会在[采购计划](#)提出解决方案，及供应商的选择及价格。

音频及视频的播放时长的统计，主要依赖第三方云视频服务支持，目前主流的云视频服务的厂商都支持，例如网易云视频、腾讯云视频、阿里云短视频服务等等，只需要取数据即可。

导航内容栏目的统计分析依赖具体课程所有数据的累计，输出时间趋势模型，例如昨天/近7天/近30天/近90天/近365天的分享次数、收藏次数、播放次数(视频播放次数、音频

播放次数)图形及文案结果。

学习	专题	播放 暂停 收藏 分享	次数 曝光率	次数(1-999999) 曝光率(1-100%)	点击曝光率的算法为 (总次数/次数*100%) 专题点击次数/曝光率	次数统计 相加; 曝光率基 础算术
	课程					
	大咖					
	顾问					
	文章	收藏、分享	次数、曝光率	次数(1-999999) 曝光率(1-100%)
产品	产品 详情	收藏、分享	次数、曝光率	次数(1-999999) 曝光率(1-100%)
顾问	顾问 详情	点收藏、分享	次数、曝光率	次数(1-999999) 曝光率(1-100%)

没有办法对于标签进行统计，因为一个标签可能配置了多个不同的内容详情，并且跨越了不同的内容栏目，公式模型没有办法进行统一的输出。等后续日志布局及爬虫构建之后，布局标签分析及页面分析模型。

3. 消息统计分析(不支持)

消息的统计分析主要是依赖系统架构的消息队列，核心的内容是接口设计方案，接口统计。暂时没有收到技术提供的 APIs 文档。所以没有办法进行消息触达率、发送人数、次数。而且当前的消息推送也是依赖的第三方服务支持，所有的数据也都是从那边进行获取的。也需要对第三方推送采购。会在数据采集里面推荐极光的产品，极光既包括了数据采集也包括了消息的推送机制。

4. 技术接口分析(不支持)

接口数据统计需求，调用次数、失败率、平均耗时、最大耗时的图形、表格、文案模型的输出，和业务数据分析进行拆分。直接在采购设计的 ELK 里面进行日志分析，获得结果。

5. 用户留存活跃(部分支持)

运营最重要的还是运营分析、活动分析，可以很快清楚的知道运营活动是否成功。用户增长的分析主要的数据来源是在用户的留存及活跃，用户活跃度，日活、月活等是一个产品是否成功的一个评价标准。

首先需要建立留存和活跃指标。活跃指标以单位时间内用户打开 App 的人数 or 停留时长/除以用户总人数 or 用户总停留时长为度量标准；留存指标是单位时间(D)内的访问总人数(V),V/M，单位时间内访问的总人数除以单位时间，就是单位时间内的留存率。

事务	数据模型	分析说明	科目&维度计算公式
用户活跃	活跃指标	时间	日时段变化趋势，24 小时曲线趋势及 2 时段\3 小时段\4 小时段趋势 及时段分布对比分析
		日期选择	每日趋势变化，按照昨天/近 7 天/近 30 天/近 90 天\近 365 天
		新增、累计、净增(访问次数/时长/人数)	累积=所有数量之和；净增=当前数量-前段数量
	用户活跃	平均访问次数/时长/人数	访问次数/访问人数/访问时长
		访问页面指标	页面路径，页面活跃度的降序排序
		日期时间	不同日期时间维度的指标
	活跃	活跃度	每日趋势变化，按照昨天/近 7 天/近 30 天/近 90 天\近 365 天\选择时间区间
		活跃	以单位时间内用户打开 App 的人数 or 停留时长/除以用户总人数 or 用户总停留时长为度量标准
		用户	融资顾问/普通用户(是否 VIP)
	渠道	终端维度	系统、系统版本、设备机型

对用户打开访问浏览的人数、次数、访问(停留)时长，以及在每个页面内的人数、次数、访

问(停留)时长进行采集统计，还需要记录每个页面的访问路径。

单位时间内，输出用户访问浏览人数、次数、访问(停留)时长的趋势模型；输出单页面的人数、次数、访问(停留)时长对比模型，及漏洞模型；需要统计分析用户最常访问路径。

用户留存，需先构建特定的用户行为，根据当前交泰金管家产品的业务场景进行定制的用户行为，【分享】【成功发布需求】【开通VIP】【提现成功到账】【成功上传征信报告】【完成单个课程所有内容】等等。可以去分析单个业务场景下的用户留存率情况。

留存率	留存指标是单位时间(D)内的访问总人数(V), V/M ，单位时间内访问的总人数除以单位时间，就是单位时间内的留存率		
日期	不同日期时间维度的指标	每日趋势变化，按照昨天/近7天/近30天/近90天/近365天\选择时间区间	
用户留存			
渠道			
特定行为			
流失率			

流失率，既有用户留存，那么肯定就会有流失。用户流失率的算法为

用户总数(T)*单位时间=M

单位时间内用户总数(N)*单位时间*(单位时间内用户总数(N)/用户总数(T))=Y

$(\text{用户总数}(T) - \text{单位时间内访问用户总人数}(V)) / (M + Y) = \text{用户流失率}$

- M是总用户生命周期
- N是单位时间内用户周期
- $(\text{单位时间内用户总数}(N) / \text{用户总数}(T))$ 是单位时间内的用户增长速率，保留两位小数点

➤ (用户总数(T)-单位时间内访问用户总人数(V))是总用户流失量

6. 转化分析(不支持)

用户在学习内容页面的时候,点击购买的次数及结算金额,用户所在物理空间的区域等分析。

需要定义占位数据, 占位数据的定义为

$N \text{ 环节占位数据} = (\text{第一环节进入人数} - N \text{ 环节进入人数}) \div 2$

$N \text{ 环节转化率} = \text{第 } N \text{ 环节进入人数} \div \text{第 } (N-1) \text{ 环节进入人数}$

$N \text{ 环节总体转化率} = \text{第 } N \text{ 环节进入人数} \div \text{第 } 1 \text{ 环节进入人数}$

在会员转化统计分析里面, 我们把学习内容主页定义为第一环节, 将专题、大咖&顾问、我

的-开通会员定义为第二环节, 将内容详情页面, 例如文章详情, 专题、大咖&顾问的详情页

均定位第三环节

学习	专题	购买	日期时间 人数 次数 金额 省市			
	课程					
	大咖					
	顾问					
	文章					
	免费体验					
	学习					
我的	开通会员					

7. 咨询分析(不支持)

用户在产品内容的时候

需求统计

普通用户发布需求, 需求发布数量、需求发布普通用户人数

融资顾问发布悬赏(需求), 需求发布数量、需求发布融资顾问人数

征信统计

征信上传数量、征信上传人数、征信上传次数、征信审核通过数量、征信审核拒绝数量

产品统计

融资顾问发布产品, 产品发布数量, 产品发布次数、产品发布融资用户人数、产品审核通过数量、产品审核拒绝数量

咨询发起统计

普通用户咨询的发起, 产品页面咨询、机构-机构详情咨询(含产品和顾问标签)、融资顾问咨询、产品-产品详情咨询、融资顾问-顾问详情咨询

融资顾问联系的发起, 首页联系、客户-客户详情联系、融资顾问悬赏/换单/甩单(需求)-详情联系、我的-我的客户

咨询状态统计

和普通用户的咨询, 和融资顾问的咨询 (同行咨询)

有意向的咨询、联系的咨询

融资顾问发布换单(订单), 发布数量、发布人数

融资顾问发布甩单(订单), 发布数量、发布人数

咨询人数、咨询订单数量

批注 [A1]: 这里存在问题, 我是跟用户在聊还是在融资顾问在聊?

批注 [A2]: 换单和甩单的业务流是怎么样的?

批注 [A3]: 是否存在有单个有多个需求的情况?

8. 财务分析

提现的审核通过率及效率、到账的统计分析。

提现的金额、提现次数、提现人数、提现审核通过数量、提现审核拒绝数量、提现未审核

数量、提现发起时间、提现审核通过时间

5. 数据展现

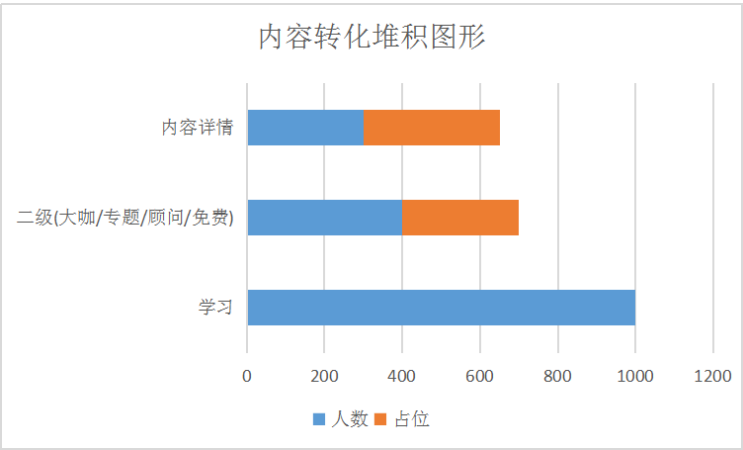
根据数据分析出来的结果，筛选和选择不同的数据事件、属性、指标，构建数据事务模型。

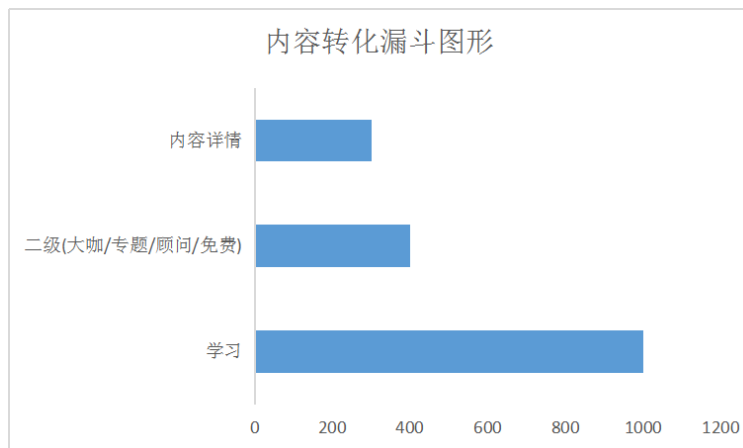
使用表格、图形及报告文案对整个整个数据采集、数据处理、数据统计分析的整个起因、过程、结果进行完整的复盘，复现。

建立良好的数据维度选择、数据表现交互方案，将数据结论使用有效、直观的方式展现出来。

通过饼图、柱状图、条形图、折线图、散点图、雷达图、金字塔图、矩阵图、漏斗图、正负太分布图等图标表达方式将表达出来。

转化





6. 数据挖掘

挖掘有意义的数据结果，对整个业务进行建议、表达数据所呈现的问题要素，并提出对应的解决方案。

7. 第三方数据解决方案及采购计划

采购计划，是第三方平台资源来补充现有的人员、技术储备，用于节省数据平台的人员、时间、技术事项的成本。目前我们现在已经在交泰金管家产品上面使用并运用了大量的第三平台支持技术的部署、实现等等。我们现在数据分析平台需要支持的技术能力包括第三方数据采集平台，另外业务需要补充和完善的能力，包括支付、分享、推送、应用发布等能力，有很多云服务平台或者第三方服务平台可能具备多种支持能力，能够为项目节省更多的成本及能够很好的集中性去解决某些缺失的能力部分。

我们作为金融知识付费平台，相对来讲，支付、推送，应该是我们具备的核心能力。分享和应用发布的这是属于互联网公司该有的基础能力。

1. 第三方数据采集平台

第三方数据采集平台，现在正在调研，或已经调研的平台包括，技术部门正在使用部分能力的友盟，他们对于友盟可能会比较熟悉，而且对于数据埋点及导入 SDK 的操作熟悉的匹配度是很好。

友盟+网站是由北京锐讯灵通科技有限公司拥有并运营，友盟+是公司自有品牌。
友盟+以“数据智能，驱动业务增长”为使命，基于卓越的技术与算法能力，结合实时更新的全域数据资源，挖掘出 40+标签大类，15,000+客群标签、100+用户/行业分析指标，通过 AI 赋能的一站式互联网数据产品与服务体系，帮助企业实现深度用户洞察、实时业务决策和持续业务增长。
截至 2018 年底，友盟+已累计为 165 万款 App、760 万家网站提供 9 年的专业数据服务。
(友盟官网的介绍)

友盟也有消息推送服务能力支撑，但他们主要的优势还是在数据采集，统计分析。另外友盟还可以提供应用市场分发的数据统计情况。

极光的主要优势在除了数据采集及统计分析之外，他还有集成其它的支持能力，例如推送可应用到交泰金管家的内容推送及系统消息推送等等、IM(即时通讯)可应用到交泰金管家咨询模块以及客服系统的搭建上面(更好的集成，例如腾讯云通信)、极光分享，可集成微信、微博、QQ 等社交平台的授权分享。(极光的短信校验不推荐，短信校验还是可考虑创蓝 253；极光的鉴权服务也不推荐，可以考虑京东万象之类的平台，可能更靠谱)

套餐	价格	门槛限制		自有系统接入
基础	39600¥/年	初始 DAU 日活	2W	不支持
		运营/增长用户 DAU 峰值	1000¥/月	

深圳市和讯华谷信息技术有限公司创立于 2011 年,其团队核心成员来自腾讯、摩根士丹利、豆瓣、Teradata 和中国移动等公司。公司总部位于深圳，在北京、上海、广州、成都均设有分支机构和办事处。

极光推送搭建起一个高度稳定、可扩展的云端架构，极大地帮助移动应用开发者节约开发和维护的成本，轻松实现毫秒级的精准推送。

我们将品牌升级为“极光”，完成了由国海证券、TCL、Mandra Capital、复星、金光集团、汉鼎宇佑等机构投资的数千万美元的 C 轮融资，同时将业务拓展至开发者服务、精准营销和数据服务三大体系。其中，开发者服务产品线增加了极光 IM (JMessage)、极光统计 (JAnalytics) 和极光分享 (JShare) 三个产品。数据服务领域开展了金融风控服务、客户洞察服务与行业研究服务，并推出移动应用监测平台“极光 iAPP”。

我们相信，品牌成长的过程，就是与客户肩并肩迈向成功的过程。我们将以市场为导向，以创新为动力，以技术为支持，不断用心努力，为每一位尊贵的客户提供极致的服务。

GrowingIO 增长黑客对于移动应用服务支持非常的友好,可以对小程序矩阵、集成公众号、应用程序的统计数据。和所有的统计分析数据平台一样都有专门的互联网金融行业的 demo, 演示。

GrowingIO 是基于用户行为的新一代数据分析产品，首推国内领先的“分析工具+运营咨询+持续增长”数据服务体系，帮助企业构建数据运营闭环。依托于快速部署无埋点技术，实时采集全量行为数据，通过搭建完整的数据监控体系，高效管控核心业务指标，帮助企业用数据驱动业务增长。

GrowingIO 专注于互联网、金融、新零售、运营商等行业，帮助企业挖掘更多商业价值。成立以来，服务超 7000 家企业级客户，获得人人贷、滴滴、陌陌、58 赶集、链家、春秋航空、Camera360、华住、如家、北森、销售易等数千家客户的青睐。

GrowingIO 用数据，驱动企业增长。

第三方数据统计分析平台的服务及价格比对

其他集成服务能力

2. 应用发布市场计划

应用市场的发布的门槛在 iOS，苹果分发市场 App store 的相关数据不好拿，导致了我们也没有办法去做分享、数据采集、推送的 Apple 手机相关的业务支持，因为这些平台上完成这些都依赖和依靠 Apple 官方给到应用开发者和(或)公司的 App ID，而且对于 AppID 的要求是企业账户。苹果开发者的账户类型分为个人、公司与企业。

<https://developer.apple.com/cn/support/compare-memberships/>

账号类型	个人	公司	企业
价格	99\$/年	99\$/年	299\$/年
App Store	yes	yes	No, 企业内部使用
最大 uuid 支持数	100	100	不限制
协作人数	1	多人	多人
邓白氏编码 (DUNS Number)	No	No	yes

当前业务需求需要支持内部测试使用不受限制，也需要支持发布到应用市场，每年的支出金

额大概为 $299+99=398$ \$/年

文件名称(Document Name)				
交泰金管家数据分析平台项目方案				
文件说明(Document Description)				
交泰金管家 App 应用的前台用户行为(含功能及路径、热点)、运营活动推广、内容及产品分类标签, 布局数据埋点、业务数据字典逻辑构建、数据采集, 汇总到交泰管理后台的数据分析平台模块, 对数据梳理清洗, 科目事务及统计维度整理, 输出数据分析结果模型。				
文档状态(status)	初稿 []	文档标识	JT-JGJ-PD.INF_PMD001	
	评审通过 []	当前版本	V0.0	
	修改 []	作者	陈志坚	
	发布 []	公司/部门	产品部	
	版本 []	完成日期		
编辑历史(History Version)				
编辑时间	版本	作者	编辑内容	标记
20190410	V0.0	陈志坚	新增	

目录

- 1. 概述.....26
- 2. 构建信息架构思维26
 - 1. 信息架构描述26
 - 2. 信息线索数据埋点.....29
- 3. 搭建数据架构思路31
 - 1. 数据核心路径31
 - 2. 数据框架选型32
- 5. 交泰金管家数据平台项目规划33

1. 概述

2. 构建信息架构思维

信息架构是由前后端和(或)前后台的 web service 网页服务/App service 应用服务构建起来的信息架构, 前后端完成的是用户看的见的, 所见即所得的信息服务, 后台是用户看不见的, 业务及功能逻辑的数据服务。整个信息架构的信息服务四个部分组合而成, 主要由组织系统、导航系统、标签系统、以及搜索系统构建而成的信息服务; 而数据服务主要是元数据、受控词表, 词序表, 数据间的层级及关系。

1. 信息架构描述

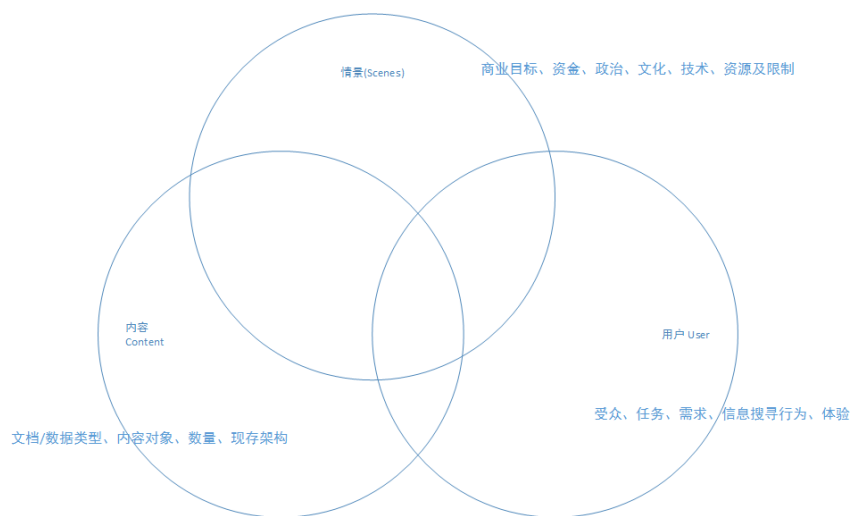
交泰金管家数据平台信息, 产品、技术、运营人员需要先搭建整套的信息架构布局, 便于后续的数据架构当中, 能够清楚明白的知道要采集哪些数据, 如何对采集的数据进行颗粒度切分? 如何进行数据的转换? 如何建立数据的事务, 数据关联之间的建模。

数据颗粒度是指数据的细化程度。数据粒度越小、信息越细、数据量越大、算法公式模型越复杂, 相对的输出结果模型越精准。

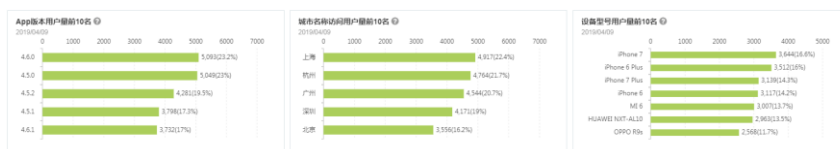
产品需对前后端的信息及前后台的数据进行信息架构的梳理, 对信息组织形式梳理, 通过导航系统的一二层级划分, 输出系统业务逻辑及流程。



我们当前的前端 App 应用的一级导航系统是【我的】【首页】【学习】【咨询】【发布需求】；二级导航系统是【产品】【机构】【信用卡】【讲师】【专题】【活动】【消息】【用户】。我们需要对这些导航进行业务数据的统计分析,例如从讲师列表或者专题列表的触达,购买 VIP, 转化成会员的用户数量有多少, 形成漏斗分析模型方案。

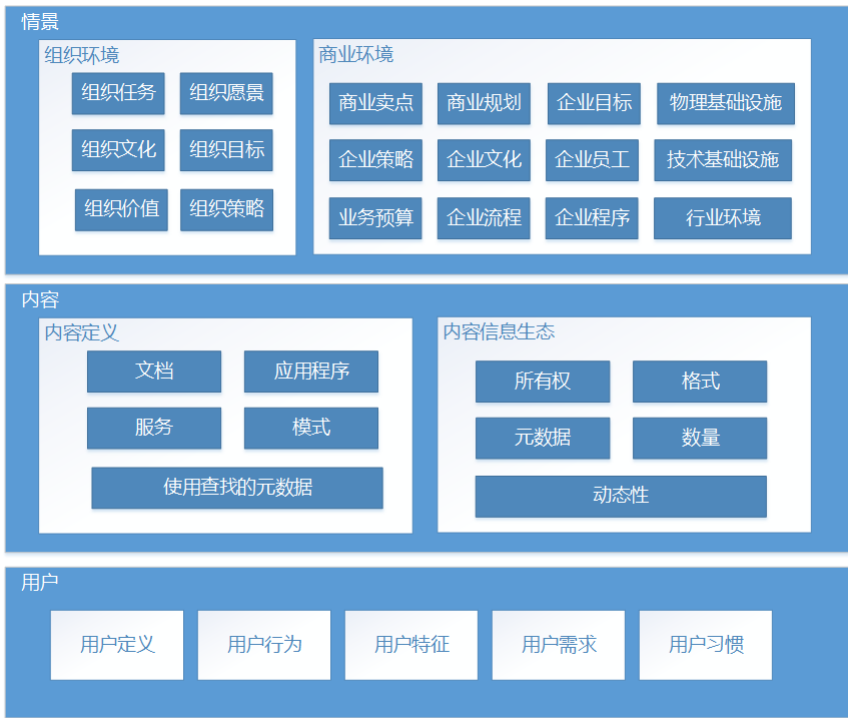


标签系统的主要组织形式是分类标签、状态标签、任务路径、顺序索引。标签分布在各个导航所承载的页面里面。例如【产品分类】【课程分类】【融资类型】【消息状态】、【选择身份】【完善信息】【时间正序】【时间倒序】等等。我们需要对于这些标签读用户行为以及偏好进行统计分析,便于我们去调整运营策略及活动指标, 优化我们的元数据信息结构。例如,采集用客户的页面点击事件热点分布图,我们可以调整优化标签位置以及活动的位置,把她放到用户关注的位置上面去。



信息及数据的整个组织系统, 我们需要通过 CMS 内容管理平台以及数据字典 ep 数据业务来进行的统一的管理, 便于我们维护信息架构的语义表达的一致性, 整个信息架构的可扩展,

可扩容，便于数据平台的建模算法公式的构建。



搜索系统是针对整个内容管理系统以及元数据的遍历查询，可以通过数据平台构建，便于系统敏锐觉察到用户对于信息的关注点，对于组织系统的探索深度，以及页面触达率的计算等等。

技术接入第三方数据采集统计分析平台，将数据模型结果前端界面接入到交泰管理后台的数据分析模块。遍历及索引导航、标签、搜索所触达的界面，业务逻辑模块，进行数据埋点；透过数据埋点采集数据平台所需的所有科目、事件及维度的各类数据。在数据平台构建数据事务及数据公式模型，最后梳理出模型结果及文案，呈现在后端界面上。



运营，需规划推广，拉新，增长、爆点活动指标，并在每次活动发布时，对于活动最后的效果结果提供成功失败的参考标准。需要产品和技术支持的资源等等。

2. 信息线索数据埋点

信息线索，嫁接着内容、数据、页面之间的逻辑分布关系，便于数据平台梳理数据事务，建立算法公式，输出数据模型结果及文案。对数据进行结构化、异构化处理，从海量动态数据中，梳理出不完整数据、冗余数据及与当前事务管理并不是特别强的稀疏数据，并找到去除数据噪音，形成正确的统计或逻辑算法公式，输出易读性、可读性更高的模型结果及文案。

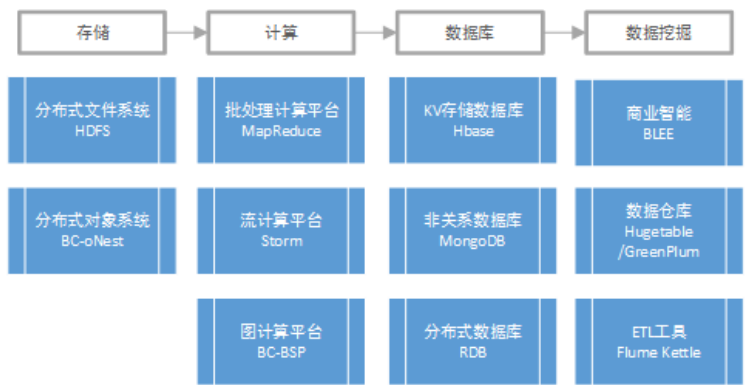
数据埋点，主要透过两种方案来进行推进。其一，通过 GrowingIO(增长黑客)、友盟、神策等数据平台上，下载的数据埋点的不同端口的 SDK，对于整个应用界面，页面路径及页面层级埋设数据采集点；其二，对于第三方数据采集平台进行服务部署方案的调研，查看业务数据最后汇总的路径是在我们自己的数据平台上，还是存储在他们的平台之上。如果是的话，需要自己构建爬虫系统程序，去爬取我们的内容管理系统及数据字典构建的元数据及业务数据库，在当中去埋设数据采集点。



对于数据库技术选型，采集、存储、统计模型的输出，可以考虑传统的 MySQL 或者是阿里的 oceanbase，或者是其它大厂且工程师熟悉的数据库选型。MySQL 数据库对于构建关系型数据业务有强大的支撑能力，但现在的数据业务方向都在趋向于分布式的数据框架及结构，数据业务的关系是趋向复杂化的。Oceanbase 既支持关系型数据，也对分布框架有很好的支持。

3. 搭建数据架构思路

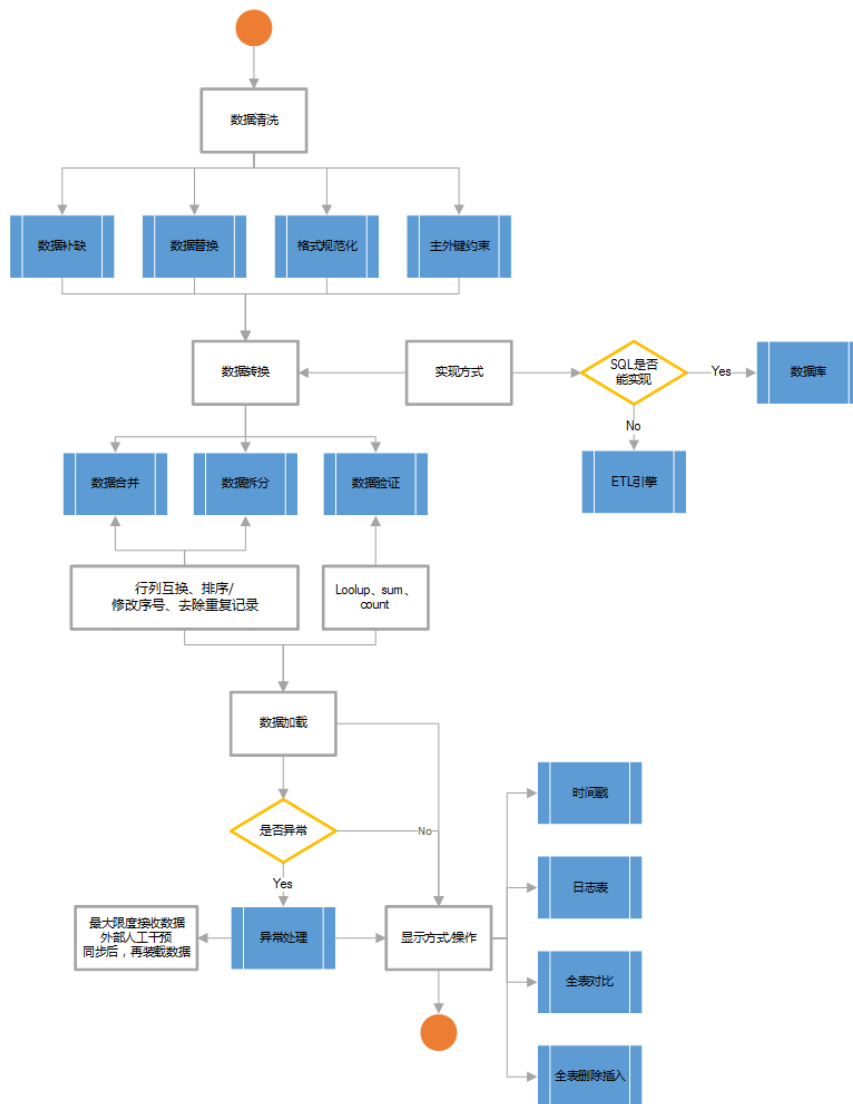
大数据平台的架构由数据采集、数据存储方案、数据管理、数据分析、数据挖掘等几部分服务所构建而成。



每块单独的服务机构均有不同的流程引擎、排序引擎进行驱动。交泰金管家数据平台的需求中包含了数据采集、存储、管理及分析的需求，后续有了一定的数据基础及业务模型之后，需要数据挖掘服务，为整个信息架构及业务系统服务。

1. 数据核心路径

数据核心路径主要是数据埋点及采集的方向。对从第三方数据采集平台和(或)交泰金管家数据平台采集回来的数据进行提取,转换成数据库选型方案所需要的格式,加载到数据库之内。对数据进行清理,补缺、替换、格式化、键对值约束;合并、拆分,当中最重要和核心的是需要对数据进行颗粒度的切分,拆的越细越好,校验采集计算方案,建立事件、指标、维度评估体系。

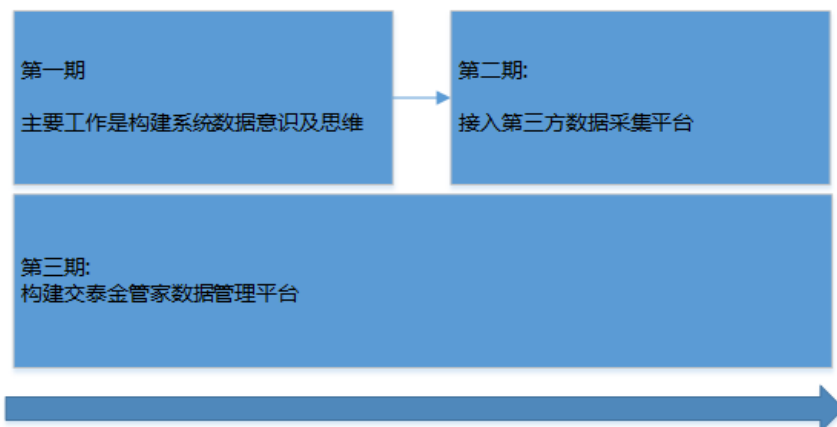


2. 数据框架选型

可供选择的架构在市面上有很多,常见的有非常老牌的,历史悠久的,阿里雅虎的 Hadoop、Google 的大数据架构 mapReduce、以及对于电商业务、交易数据支持更加友好的淘宝大数据等等。

5. 交泰金管家数据平台项目规划

整个项目方案顺序渐进，不断迭代的规划方案设计。



第三期的工作可以与第一期和第二期并行进行

第一期:主要工作是构建系统数据意识及思维(两到三个周)

1. 建设技术文档规范:

产品输出:元数据字典(.xls)、业务需求架构(.xmind&.vsdx), 新产品、新业务及业务调整

技术输出:系统架构(.vsdx,部署方案选型)、业务数据架构(关系数据库格式)需导出.xls、技术架构(.vsdx, 技术框架选型)、概要设计文档(.doc)、接口设计 Apis 文档(.doc)、技术方案文档(.doc)

2. 数据库、架构、框架, 算法及渠道引擎选型

产品:数据采集及模型的基础需求输出

技术:已知有数据库 mySQL 方案, 其它根据后续了解及产品, 运营需求进行选中

产品、技术、运营、商务、财务:采购第三方数据采集平台供应商

第二期:接入第三方数据采集平台(一到三周时间, 后续持续)

1. 前端数据埋点, 导入代码及接入 SDK

产品、运营、财务:注册第三方数据采集平台并对服务进行付费

技术:熟悉技术方案, 写入代码, 不仅要可以从第三方数据采集平台上能够看到数据, 也需要将第三方数据呈现页面转移到交泰管理后台的数据分析模块, 信息组织形式和结构和第三方数据采集平台一致

产品、测试、技术:对结果进行白盒测试校验数据准确性, 构建测试模型; 黑盒测试体验数据查阅的可读性及易读性

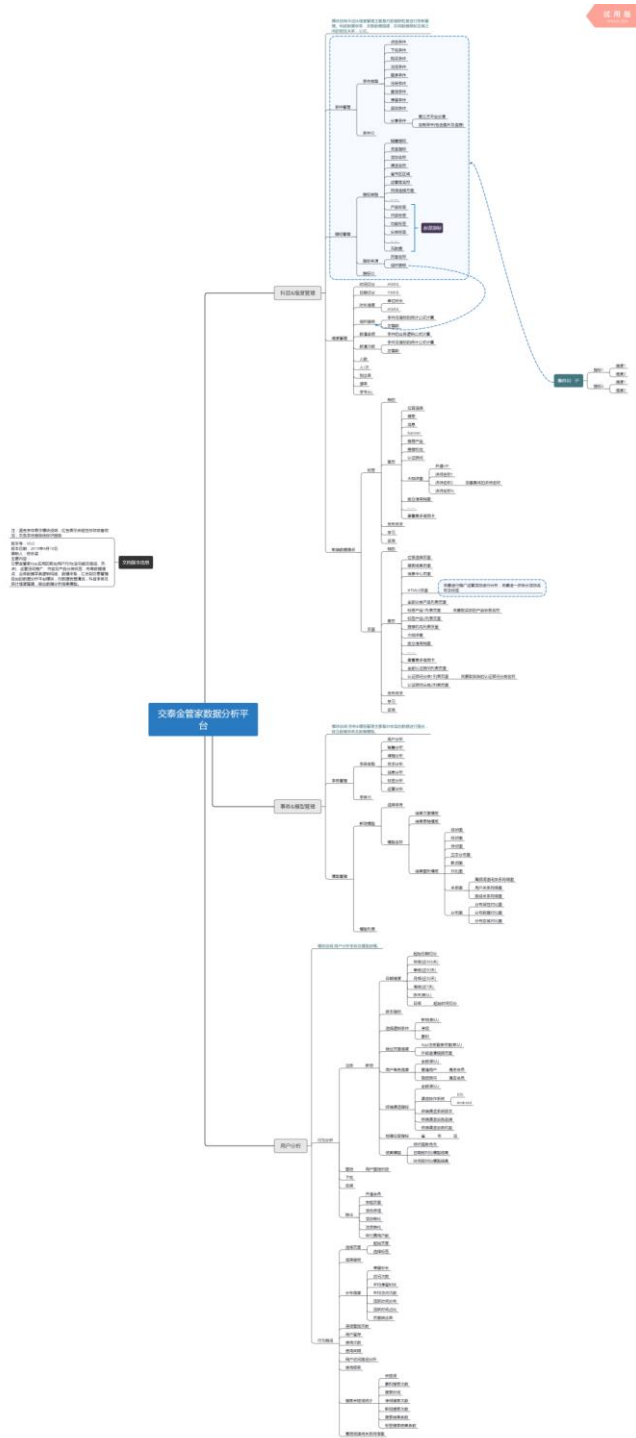
2. 数据反馈及统计, 运营可根据用户行为路径及画像进行针对的增长活动运营

产品:根据用户热点及触达路径, 调整信息架构的组织形式

运营:通过运营活动的数据反馈, 调整运营策略

第三期:构建交泰金管家数据管理平台(之前的经验是一个项目周期大概在三个月?)

1. 数据平台信息架构构建



2.技术对于整个数据平台进行概要设计方案输出