

# 快看漫画大数据平台的模型思维与用户增长实践

作者：DataFun社区

阅读数：1339    2019 年 4 月 4 日

3



喜欢



收藏



评论



微信



微博



本文根据快看漫画屈世超老师在 DataFunTalk 数据主题技术沙龙活动“大数据从底层处理到数据驱动业务”中分享的《大数据平台的模型思维与用户增长实践》编辑整理而成，在未改变原意的基础上稍做修改。

## 目录

- 快看漫画发展现状
- 模型思维与数据平台搭建
- 用数据驱动亿级用户的运营和增长

3



喜欢



收藏



评论



微信



微博

今天分享的内容分为三个部分，首先介绍下快看漫画发展状况，第二个就是我们在大数据平台和大数据业务发展的过程中总结的模型思维以及平台搭建的一些经验，最后是基于我们的数据平台用数据驱动去增长用户、业务的一些实践经验。



# 快看漫画是个漫画阅读社交平台



上图是快看漫画 APP 的一些界面，基本是精品漫画加二次元社区、社交的一个平台，同时我们也生产漫画内容。快看成立于 14 年底，经过几年的发展，已经有了超过 1.3 亿的用户量，月活动量近 4000 万、日活动量 1000 万，在中国漫画 APP 中排名第一。日数据量接近百亿条，对数据平台要求比较高。

3



喜欢



收藏



评论



微信



微博

## 数据为业务发展赋能



你无法衡量的东西，  
你也无法管理

管理学大师 彼得·德鲁克

如果从事过市场推销相关业务，都明白移动 APP 发展前几年都赶上了人口红利的曝光，推广起来比较容易，用户量也是实现了比较快速的发展。但是到 17 年开始，移动网民的数量增长量很小，在推广过程中，抢的都是存量用户。随着 APP 开发推广竞争力越来越大，市场的投放、购买越来越难，成本越来越高。就目前所有 APP 开发者普遍存在的趋势就是用最小成本购买优质用户，留存一定要做高，同时通过一定的用户运营实现活跃用户的价值输出，即超级用户思维。这一切都离不开数据支撑业务增长的需求和发展。

3



喜欢



收藏



评论



微信



微博





# 模型思维无处不在

- 看到从未用过的平衡车，老司机联想到摩托车
- 架构师的核心技能
- 准确的量化模型需要长期的积累和思考



第二部分讲一下模型思维和快看数据平台搭建，模型思维的概念很抽象，但是是无处不在的。比如我们见过的新事物会潜意识进行抽象，然后特征提取、存储到我们的意识中，再发现类似的事物就会很自然联想到这个模型，通过模型来猜测这个相近的事物有什么特点、如何去使用它、如何去产生联系。举个例子如我们第一次见到平衡车，我们会想到电动车、摩托车，要驾驭它需要很好的平衡性。我们认识新事物会自动联想我们记忆中就有的模型，这就是模型思维。在技术开发过程中模型思维也是无处不在的，比如我们在有了需求以后，需要从需求抽象出一个模型，映射的模型从我们的经验记忆中联想，那些架构模型是能够解决当前的需求，再从这个大的模型不断去拆分、细化、划分模型，每个模块又是更细粒度的模型，再拆分直到我们熟悉的技术。模型思维就是我们要从需求出发，抽象出需求模型，然后映射出架构模型，然后拆分更细粒度

3



喜欢



收藏



评论



微信

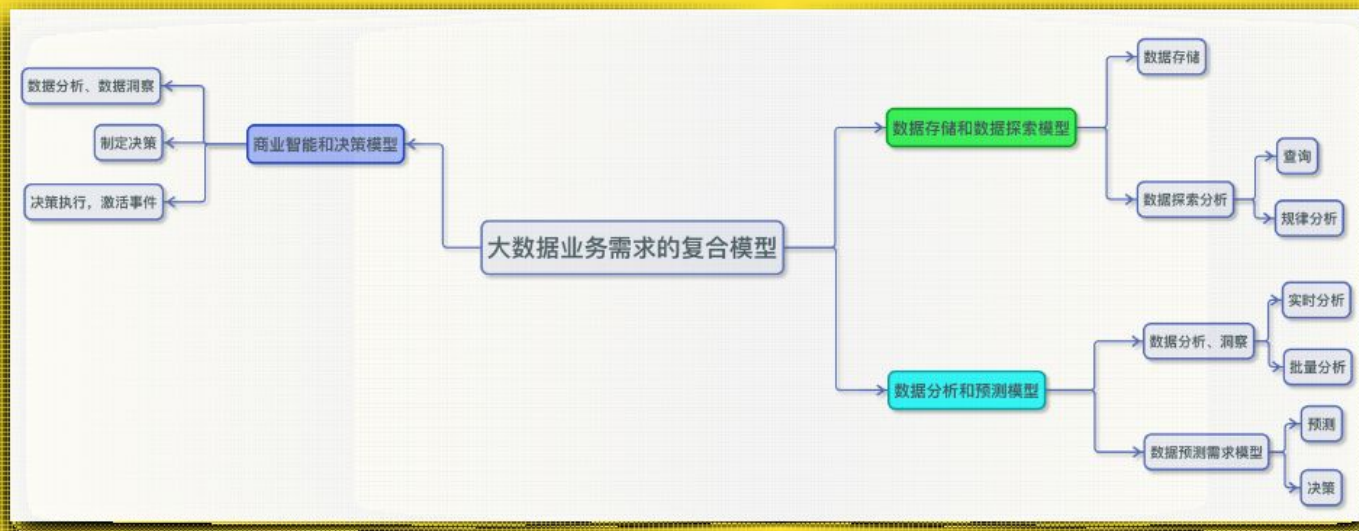


微博

模型，直到我们熟悉的技术。这其中就需要我们对多项技术的积累，多项工具、框架组合使用的经验，才能透彻理解模型思维。

快看！快

## 大数据需求模型



上图是我们大数据平台应用的总结，首先是一个需求模型的汇总，然后是架构模型的一个汇总。在架构模型中会产生更加细粒度的模型，细粒度的模型继续拆分直到不能拆分的原子模型。先从需求出发（架构设计的依据和来源），一般大数据的应用都离不开这三块，第一个就是数据存储和数据探索模型，第二部分是数据分析与预测模型，第三部分是商业智能与决策模型。比如数据存储我们需要对数据进行备份（冷备份、热备份），对数据进行查询和规律分析；数据分析和预测模型包含实时分析、批量分析，数据探查包含特征提取

3



喜欢



收藏



评论



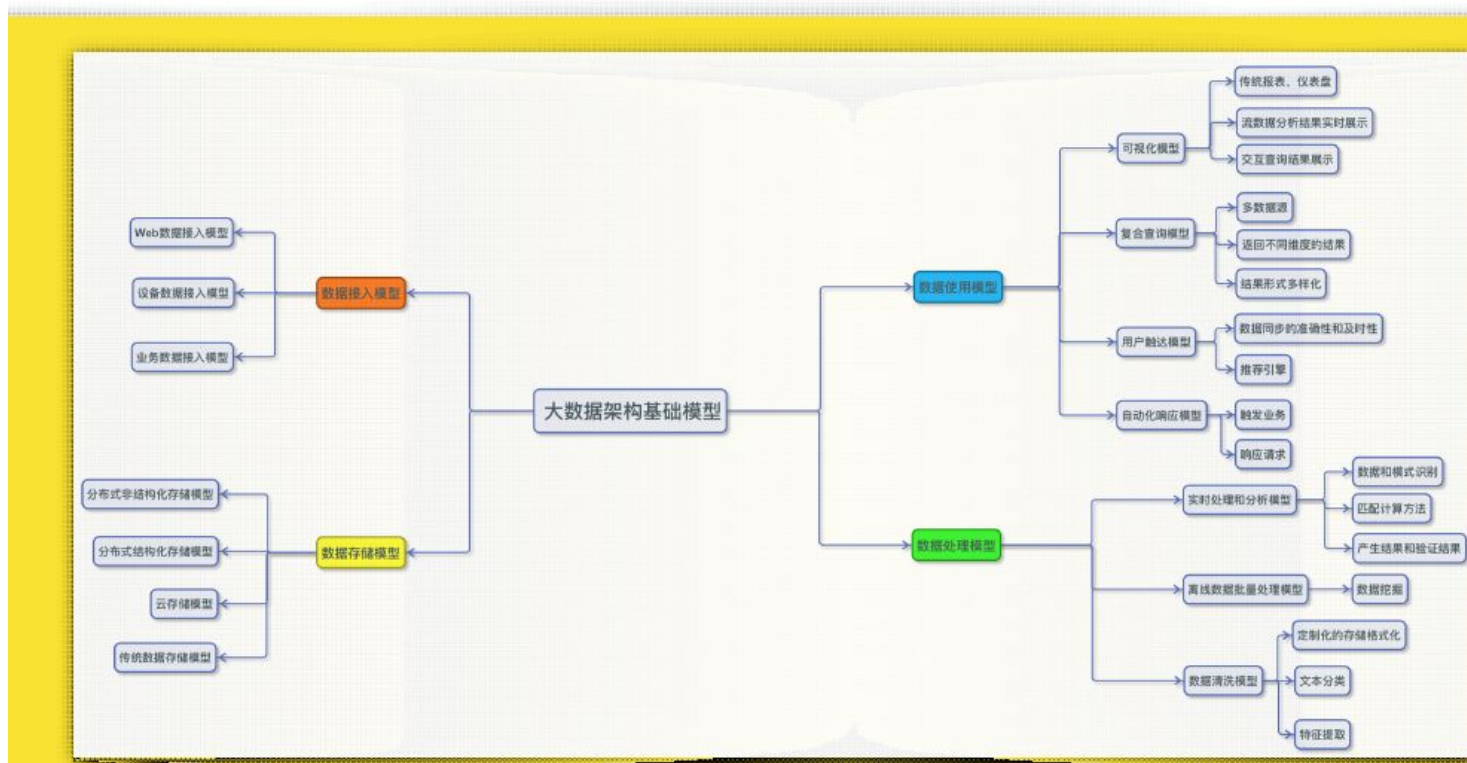
微信



微博

或个性化推荐；商业智能与决策模型在金融领域比较多，比如风控领域，对贷款人信用学习决策是否能贷款、收回贷款等。需求相关的模型基本都离不开这三部分，由需求模型映射到相应的架构模型需要经验性，依赖于需求过程中提取如何去使用这些数据、展示形式、数据源接入的方式或者根据数据源的格式或一些特点如何进行数据的清洗和处理；数据分析中用什么样方式进行洞察和分析，比如使用什么样的算法去学习个性化推荐或找到一些规律。

## 大数据架构模型



3



喜欢



收藏



评论



微信



微博

在做大数据时总会接触到一些模型，如数据接入模型，做数据采集（web 数据、设备数据、业务数据，可能有不同形式的接入方式），接入之后做数据存储，结构化与非结构化、分布式存储，云存储或存储到传统的存储模型中。然后是数据处理模型，实时、离线，还有清洗模型，不同格式、形式的数据用何种方式、工具进行处理。最后是数据使用模型，做完处理后如何进行分析、可视化，如何查询，还有用户触达，做一些智能决策，以及自动触发模型。

3



喜欢



收藏



评论



微信



微博

## 模型思维是种优秀的架构思维



参考文献: <https://www.ibm.com/developerworks/library/bd-arc4patterns4/index.html>

- 模型思维有助于业务和开发确定共同的需求目标
- 模型思维有助于开发制定关键架构决策，以及确定所需的基础组件
- 从基础组件选择到层层复合模型的实现，流程清晰
- 逐层构建的复合模型能够实现部分或者全部的原子模型替换，实现架构平滑升级

所有的架构都是由需求衍生而来，由需求产生模型，这些模型是由小的原子模型通过一定的组合、封装，逐层构建更高层次的模型，最后将技术架构实现。学习技术和架构是一样的关系，有了这种思维后，在学习技术时更有目标，在整个架构它的定位和功能是什么，与同功能的优劣对比更清楚。这种思维能够快速将需求



抽象成模型，业务方和开发方就确立共同目标，有助于开发制定关键架构决策，什么样的业务大致能知道用什么结构去解决。架构落地是由底层到高层逐层实现的，从原子模型落地，通过层层组装实现整体架构的落地。还有一个优点是有了整体架构后，能够平滑的实现架构升级、修复、替换，因为每次改动是基于原子模型或几个原子模型的组合。

## 模型思维在App发展初期的应用

3



喜欢



收藏



评论



微信



微博

- 精益求精的态度
- 低成本接入和使用第三方基础数据统计平台
- 多平台数据对比，保证数据准确性
- 关注宏观指标和业界指标分布，明白App定位



# 模型思维在App成长期的应用



前面介绍了模型思维，接下介绍快看如何发展数据平台。创业公司最开始关注的并不是用户的行为，更多的时项目是否适用于我想象的用户群体，这个群体在 APP 使用时长、活跃度、留存量。我们开始使用的是低成本接入和和使用第三方基础数据统计平台，就能满足需求，前期成本低，通过多个第三方能够保证数据准确性便于做数据相关性验证以及很多宏观指标确定 APP 的业界定位。快看发展第一年主要依据第三方数据平台做相关分析以及自己的统计任务基本能够满足需求。在成长期需求呈爆发式增长，业务线不断分散。随着业务量发展和数据增多，需求不断细化不会只看宏观基础指标，指标会细化到每一个业务，每个业务也会查看功能性指标。开始尝试做原子模型的替换，调研发现一个“二八定律”，就是如果要对用户行为进行分析，接入的第三方数据分析平台，利用 20% 的成本接入和维护能够满足各个业务线 80% 的需求。随着业务量以及业务需求增加，许多数据分析是定制化的，第三方数据平台就无法解决，就自建大数据平台。搭建原

3



喜欢



收藏



评论



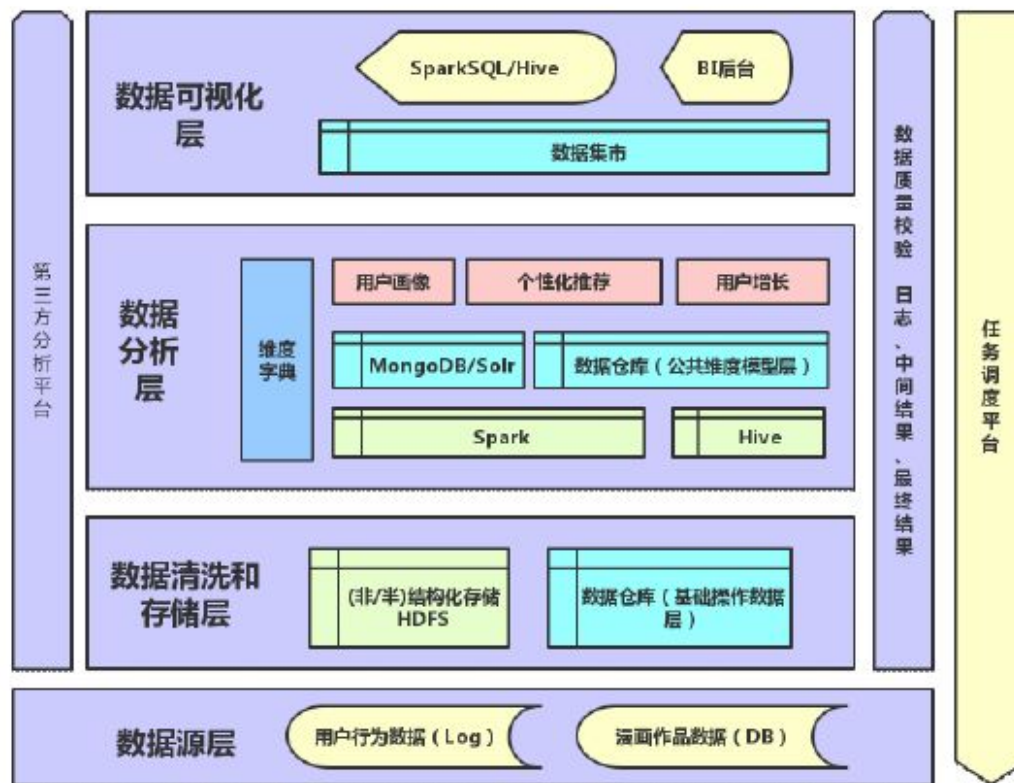
微信



微博

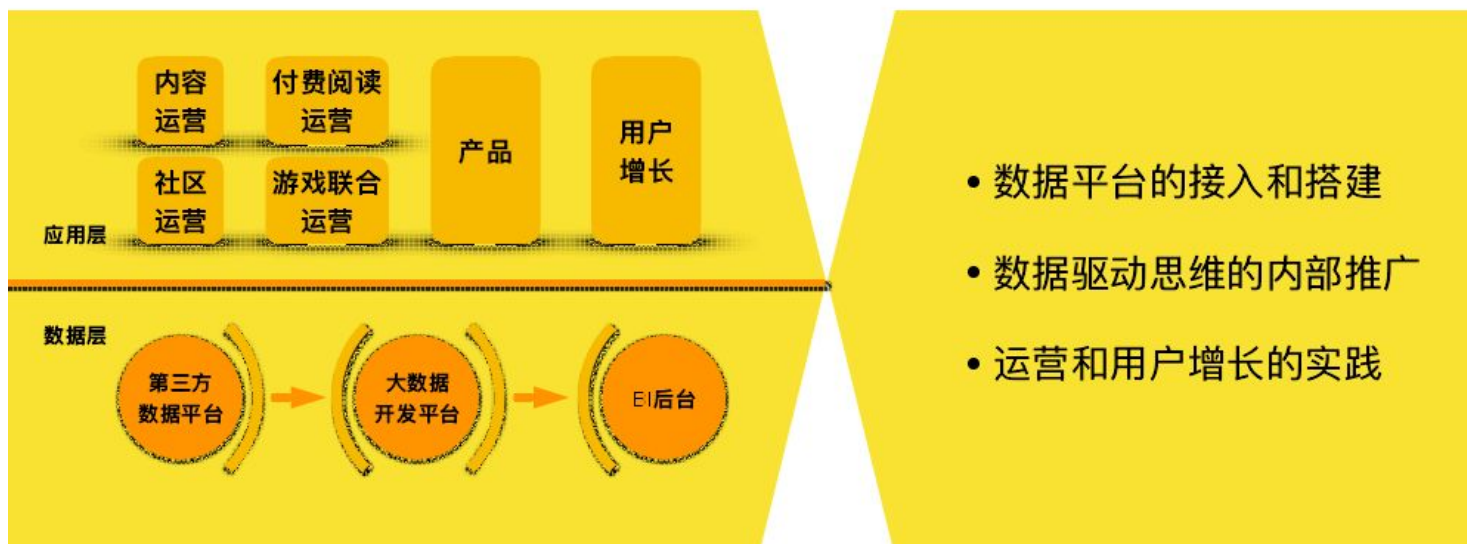
则是利用最低成本解决需求，优先使用第三方平台能解决的方案，不能解决利用大数据平台做定制性的数据洞察，比如个性化推荐，做精准的作品分发。后续数据洞察的需求越来越多，业务指标越来越定制化，数据统计的粒度会越来越细。

## 快看漫画数据平台架构



上图是平台架构，分为四层。虽然有很多模块还不是很细，但是随着业务的发展，也是适合当前人力成本取舍的一个方案。这是快看创业阶段的经验，如果有在创业公司工作的，有需要可以参考下。

## 数据驱动体系在快看的搭建和应用



- 数据平台的接入和搭建
- 数据驱动思维的内部推广
- 运营和用户增长的实践

接下来讲一下我们利用数据驱动来实现用户增长的经验，业界关于类似的探索还比较少，因此做一个总结分享一下。快看数据部门对数据的定位是基于我们的平台，对公司所有业务线提供数据需求，做好数据驱动，每个业务线做好业务增长。这个业务很宽泛也很难，以为很多创业公司是从零开始发展起来，很多时候并不太注重数据，数据平台是由需求驱动发展的，但是我们想数据平台去驱动业务线的发展，我们花费一年半时间做数据内部思维推广，由于成员年轻化，很多都不是很理解数据，还不断强化他们如何使用数据、好处是什么。然后就是聚焦在运和业务增长方面。

3



喜欢



收藏



评论



微信



微博



# 自上至下的数据驱动思维



管理者对数据驱动思维的重视

前提

自上至下的执行和配合

执行

用户增长团队的定位：驱动者

动力

业务团队对数据分析的应用

落地

先讲一下数据思维内部推广，培训介绍现有数据平台、第三方数据分析平台、BI 后台以及数据字典维度等，介绍有哪些数据；然后介绍各个业务能够使用那些数据；最后就是如何将数据用起来，让 CEO、CTO 实施让员工更多的去使用数据，强化管理层对数据的重视，要求一些数据指标的统计，从上到下分配到每一个团队，产品经理和用户增长团队辅助确立业务核心指标，着重如何确立综合指标来评价你当前的工作和业务发展；数据研发团队提供数据支持，将数据抽象化、平台化、自动化。但是也要进行一定的取舍，因为会有一些不切实际的需求，或者提出的需求分析没有用等等。

3



喜欢



收藏



评论



微信



微博

# 培养业务团队的数据思维

- 核心指标的制定以及分解
- 分配指标和增长规划到每个业务团队
- 用户增长团队参与并引导业务团队
- 业务团队掌握数据和分析工具



用户增长团队后期工作是介于每一个开发线与业务之间的工作，如制定指标，针对指标做一些细化，细化为可落地，对当前业务有增长的可执行指标。引导他们去使用并做一些优化，如付费运营，每一次活动完成后分析是否达到效果，以及下次运营需要哪些工作。会监控每一个页面位置作品曝光率以及用户在每一个分发页面的参与度，比如社区运营、游戏运营也有类似工作。产品经理和用户增长团队联系密切，共同去做好数据分析。

3



喜欢



收藏



评论



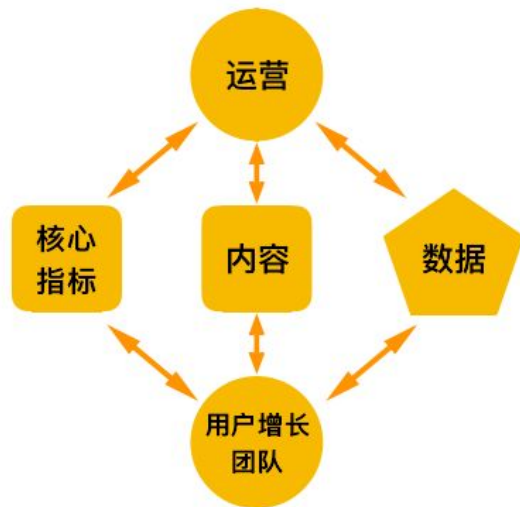
微信



微博



# 数据发现用户对内容的需求



- 漫画平台的核心功能：满足用户对于内容的需求
- 核心点：内容变化趋势，新内容的开发
- “领域知识 + 用户理解 + 数据支撑”，探索内容开发方向

快看! 快看漫画

3



喜欢



收藏



评论



微信



微博

# 用数据平台监控内容质量



用户搜索关键词分布

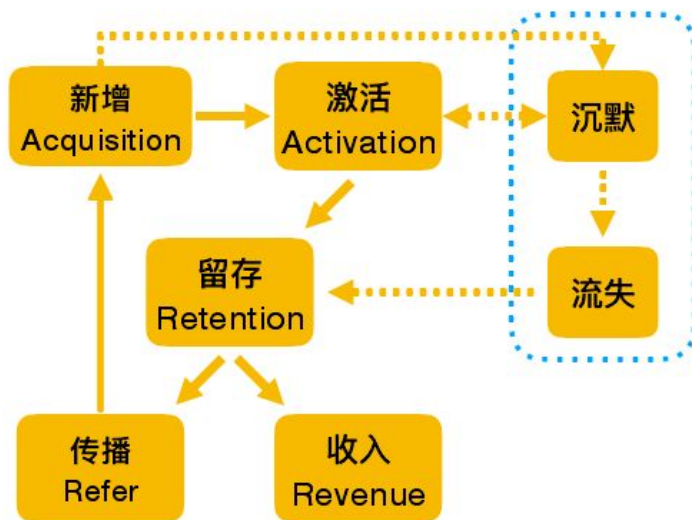
关键词	搜索次数
关键词A	102,345
关键词B	98,235
关键词C	97,235
关键词D	87,601
关键词E	67,313
关键词F	65,839
关键词G	61,097
关键词H	59,023
关键词I	45,980

接下来讲一下我们在这方面的例子，快看作为漫画平台核心功能就是满足用户对于内容的阅读需求，核心点就是把控作品内容变化趋势，同时根据用户的需求、喜好开发一些新内容，这些是内容模块的指标。让内容开发者基于对用户的理解，利用数据去支撑他们的猜测，探索内容开发方向。具体方式就是会分析每一个作品的周留存、日留存和月留存，以及用户阅读次数，看一下用户对作品持续关注度。然后通过用户搜索关键词把控新作品开发方向，这样能够把控作品质量，同时每个用户活跃度的变化能够反映作品对用户的吸引力。

- 3
- 喜欢
- 收藏
- 评论
- 微信
- 微博



# 生命周期转化追踪与运营



- AARRR（海盗模型）定义生命周期
- 实现近实时追踪每个用户生命周期
- 以用户生命周期为依托的个性化运营

上图是用户运营的思维，拉新越来越难，因此要对现有用户留存，利用的是海盗模型去定义用户的生命周期，从激活到留存还有沉默到流失，以及内容传播与收入等都有实践。指导思想就是对每个用户实时跟踪，实时搜集用户行为，标记用户生命周期，根据生命周期标签做相关推荐，如新用户进行冷启动学习进行推荐，沉默会做拉活尝试，流失会做用户召回活动。留存也会做优化，但更多是基于用户画像进行个性化作品分发以及个性化 push。

3



喜欢



收藏



评论



微信



微博



## 市场推广运营效果评估

- 渠道多，接入方式多，定制化的BI后台
- 用户拉新准确归因，精准统计
- 用户拉活，广告素材个性化投放
- 付费推广反作弊



运营这一块很重要，每年花费在市场推广成本很高，如何评价和吸引高质量的用户难点很多，主要有对接的渠道很多，接入方式多种多样，借助第三方平台实现高质量推荐不现实。因此定制化 BI 后台，尝试对用户拉新精准归因，然后做精准统计。平台应用中，发现利用自己的平台能节省 20% 的运营成本。在做用户拉活和广告素材方面会做个性化投放，通过用户个性化标签生成对应的广告素材做精准拉活和召回投放。

本文来自屈世超在 DataFun 社区的演讲，由 DataFun 编辑整理。

大数据 架构 最佳实践

3



喜欢



收藏



## 评论



微信



微博

