

# 腾讯移动分析与服务架构

杨列昂

腾讯大数据高级工程师

# SPEAKER

---



杨列昂

腾讯大数据高级工程师

2011年加入腾讯，前后负责腾讯大数据产品、海量用户画像、腾讯移动分析(MTA)、腾讯移动推送(信鸽)等核心产品的设计研发工作，专注于产品价值挖掘，系统架构优化等方向。

# CONTENTS

---

- 移动分析工具
- 营销推广与效果监测
- 互利共生的服务生态

# 移动分析工具进化论



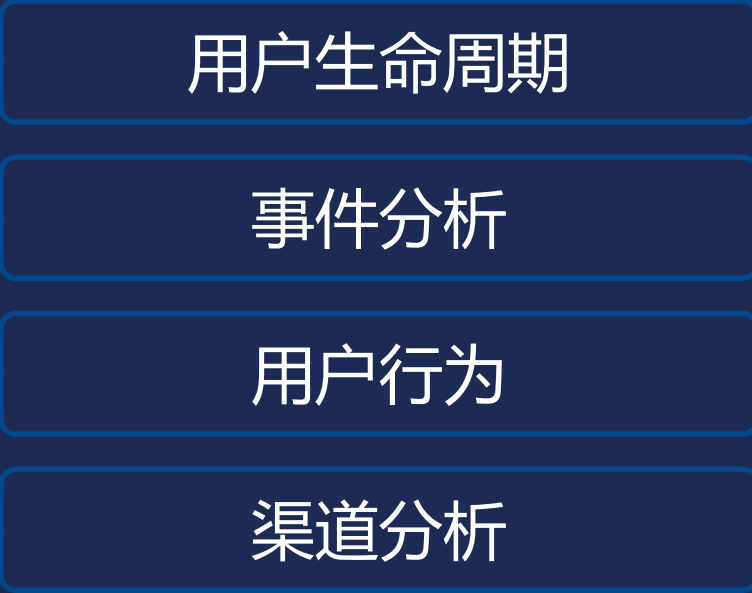
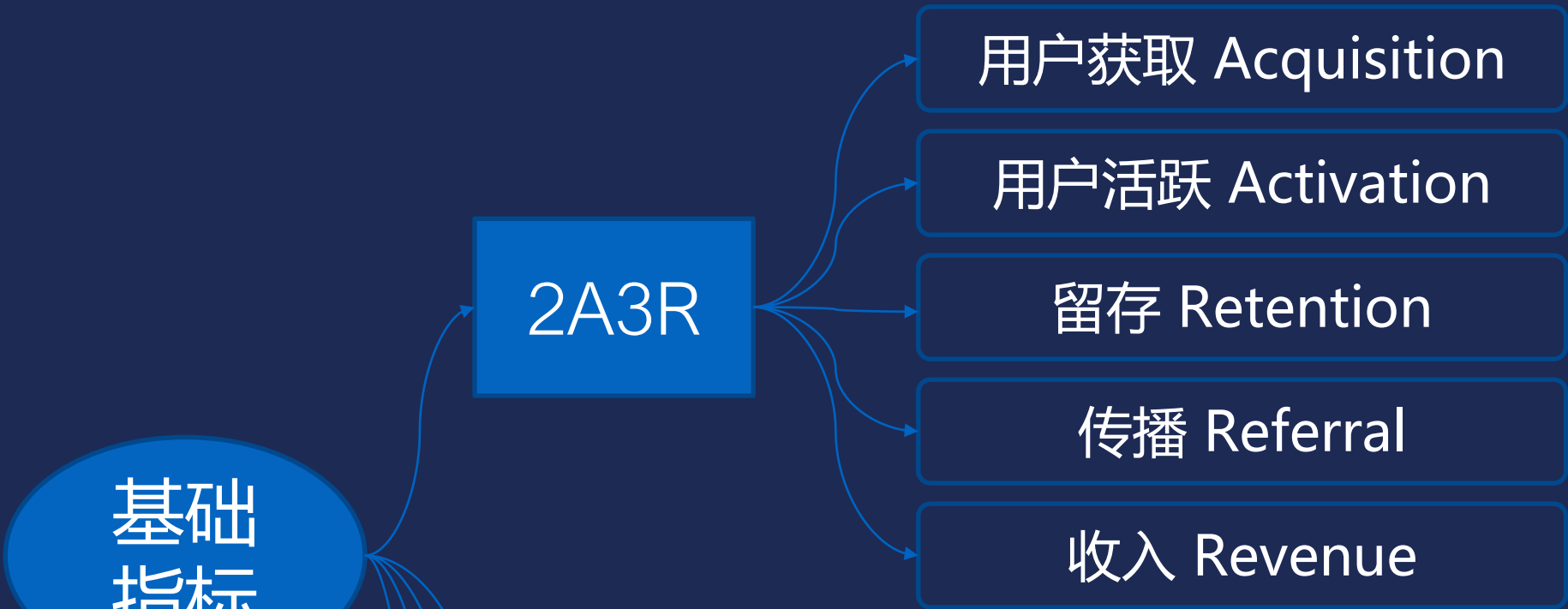




# 业务指标和价值评估体系

App 自身

竞品

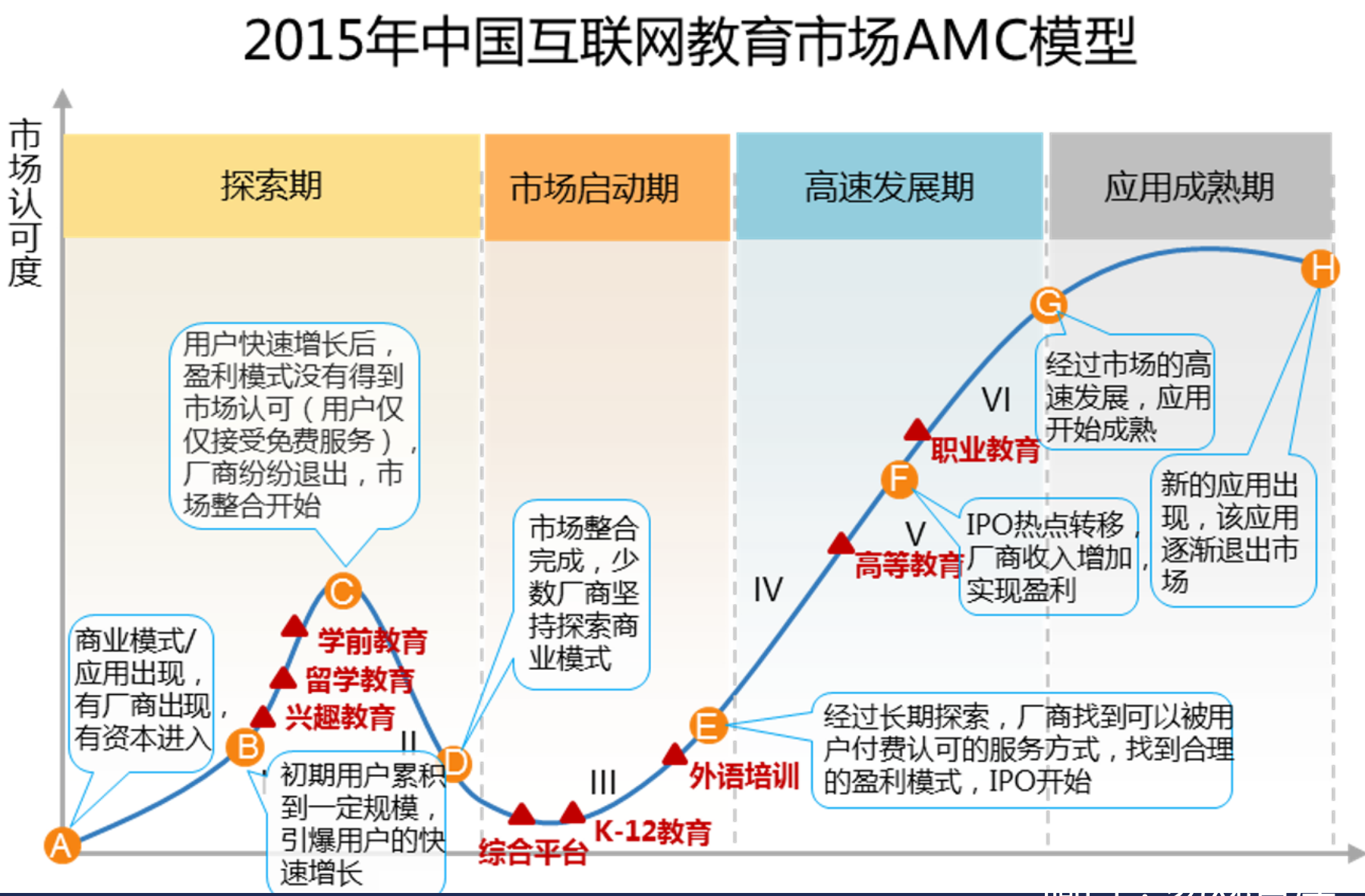
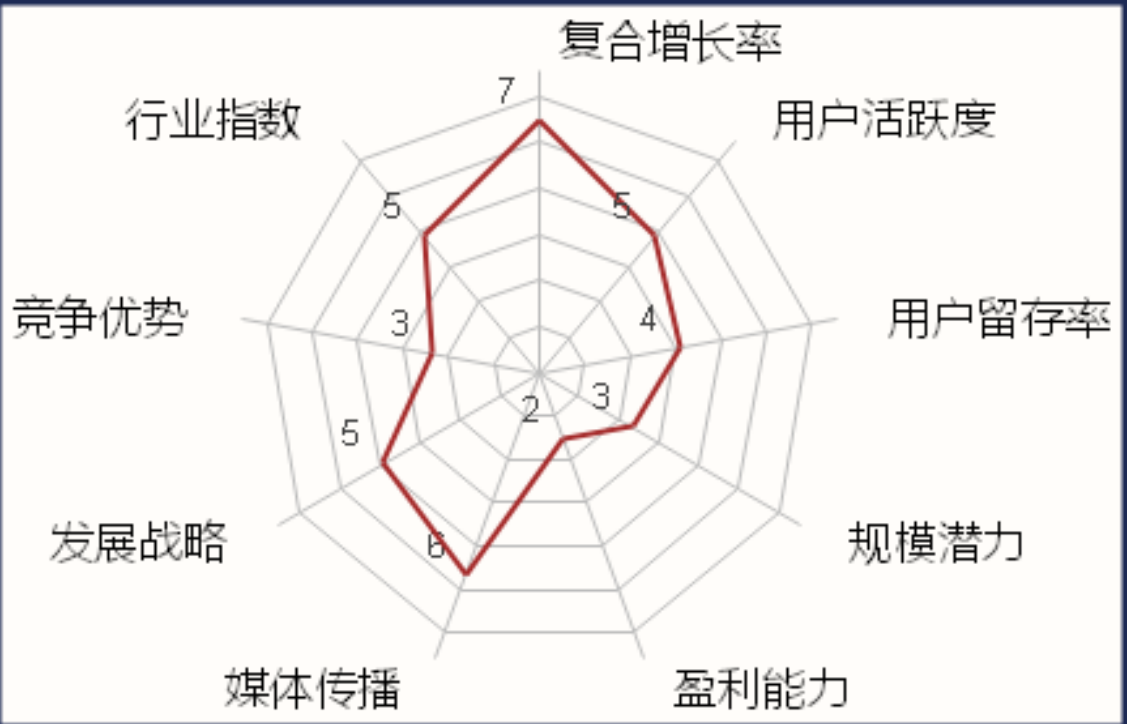
行业趋势



2		腾讯视频	视频
3		QQ浏览器-WiFi新闻动漫直播	工具

应用指数

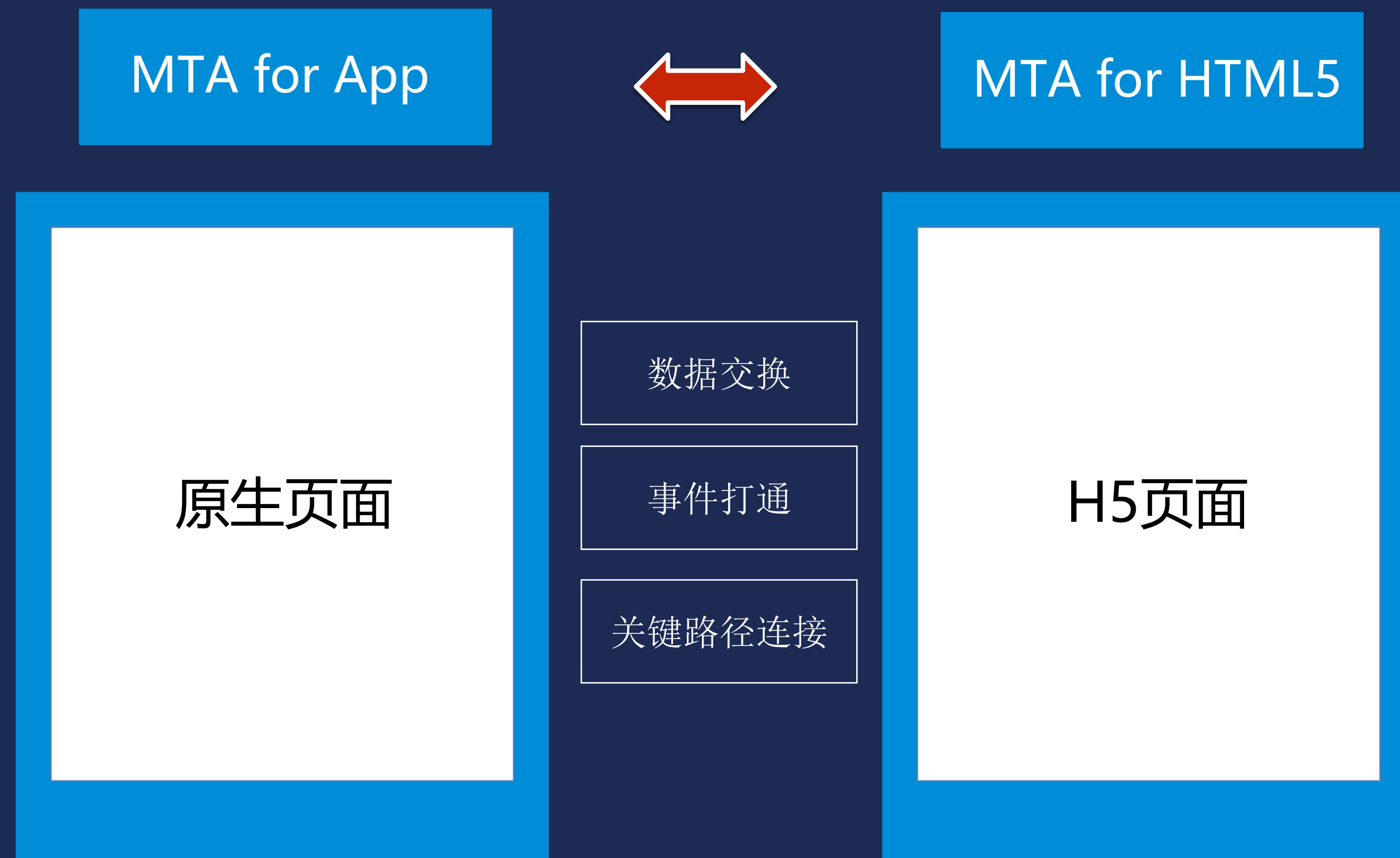
行业排名



了解所在行业潜在规模  
洞察潜在用户人群  
多种方案帮助快速增长

# 多平台接入

- Android / iOS
- HTML5 / Hybrid
- 微信小程序
- 智能硬件
- 服务器端接入



# 移动设备画像

## 账号信息

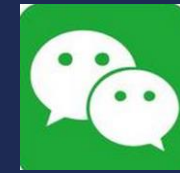
年龄, 性别, 学历...

## UGC

QZone日志, 图像, 评论...



月活跃用户超 8.4 亿  
在线人际关系链超 1000 亿



注册用户数超 10 亿  
月活跃 9.8 亿



活跃用户数超 5.6 亿  
智能终端活跃用户数 5.5 亿



移动支付超过 XX 亿笔



其他: 新闻, 视频, 音乐, 游戏, 手机管家

## 社交属性

QQ, Qzone, 好友关系链...

## 金融

理财, 消费, 信用...

## 行为偏好

新闻, 视频, 公众号...

## 商业兴趣

广告, 商品, 品牌...

## 游戏爱好

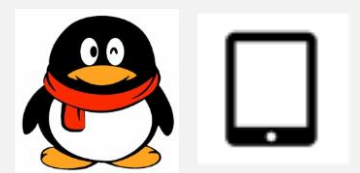
游戏时长, 类型, 付费...

## 连接

## 分析

## 标准化

账号



PGM



DNN



LPA



结构化

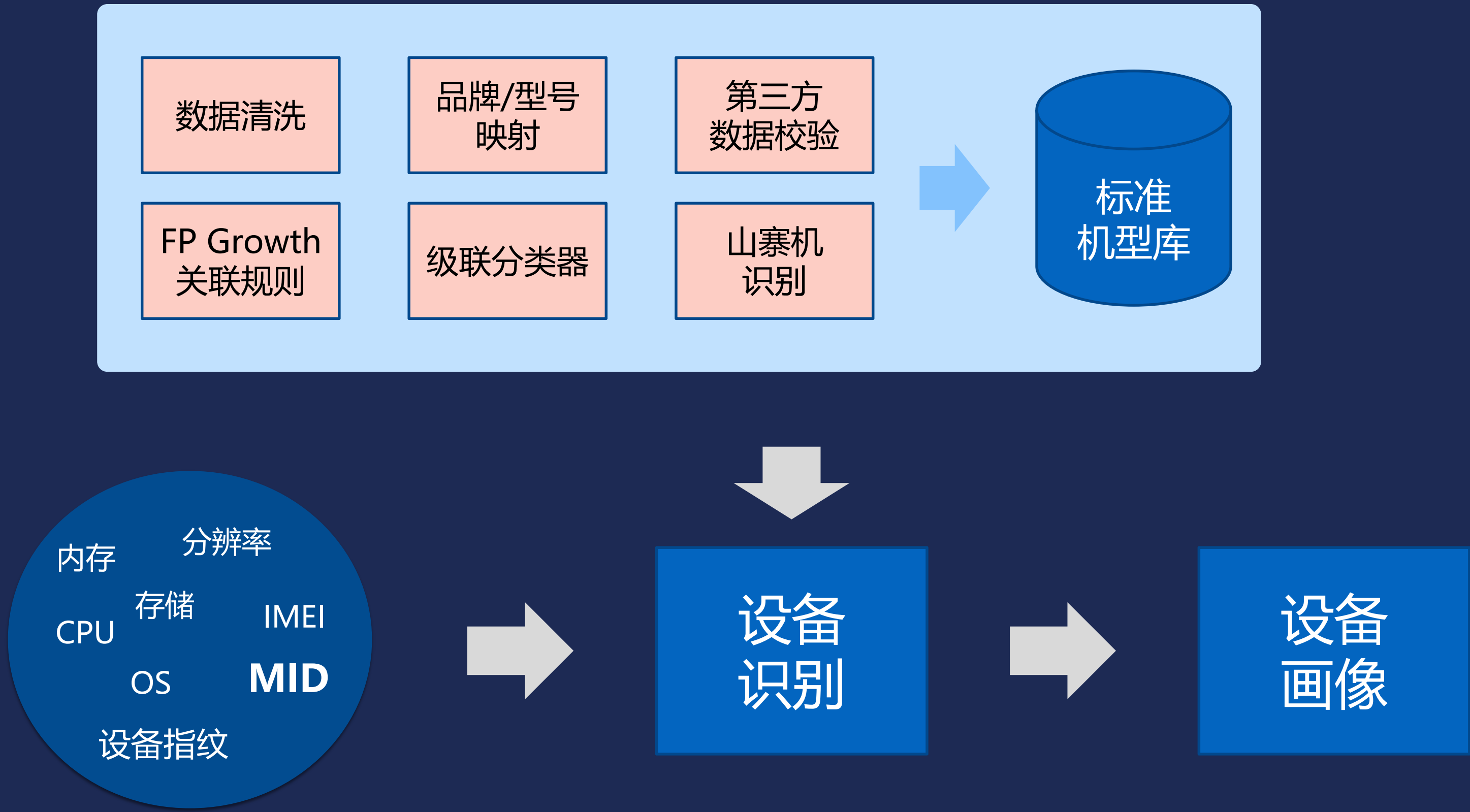


非结构



# 设备识别

- 山寨机
- 模拟器
- 篡改设备号
- 伪造请求
- 系统权限控制



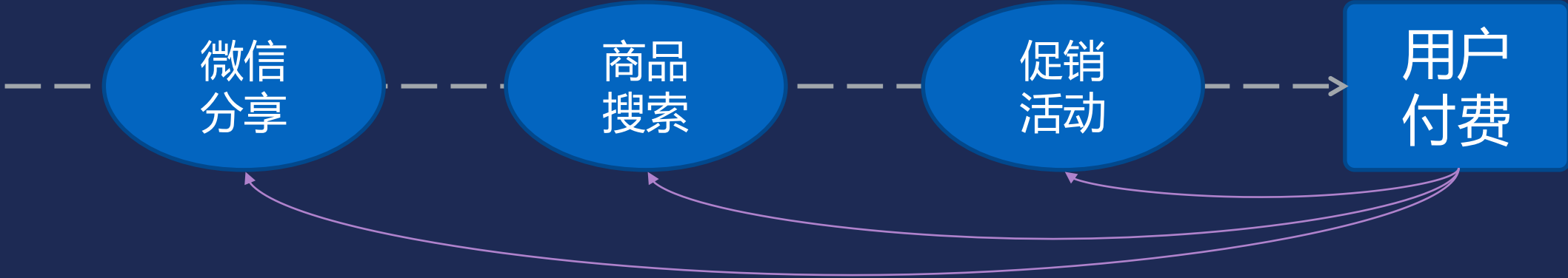


# 归因分析



# 归因分析

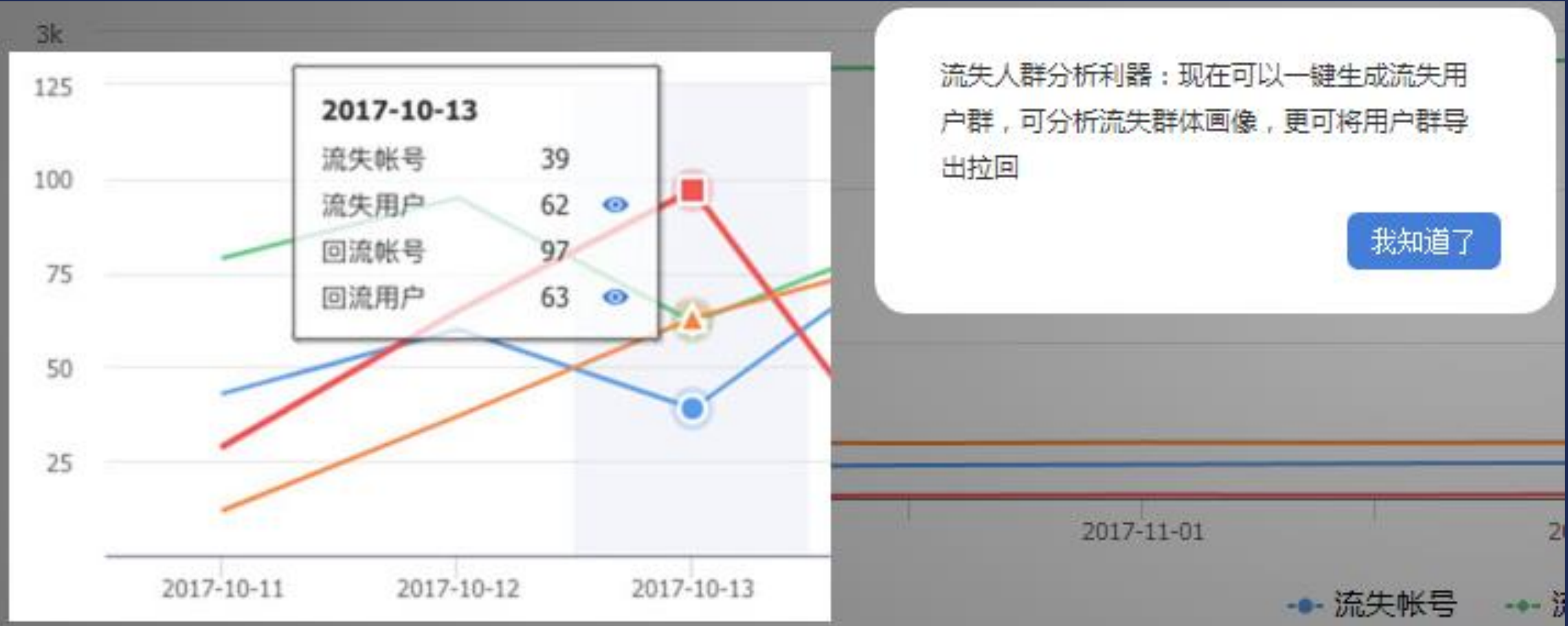
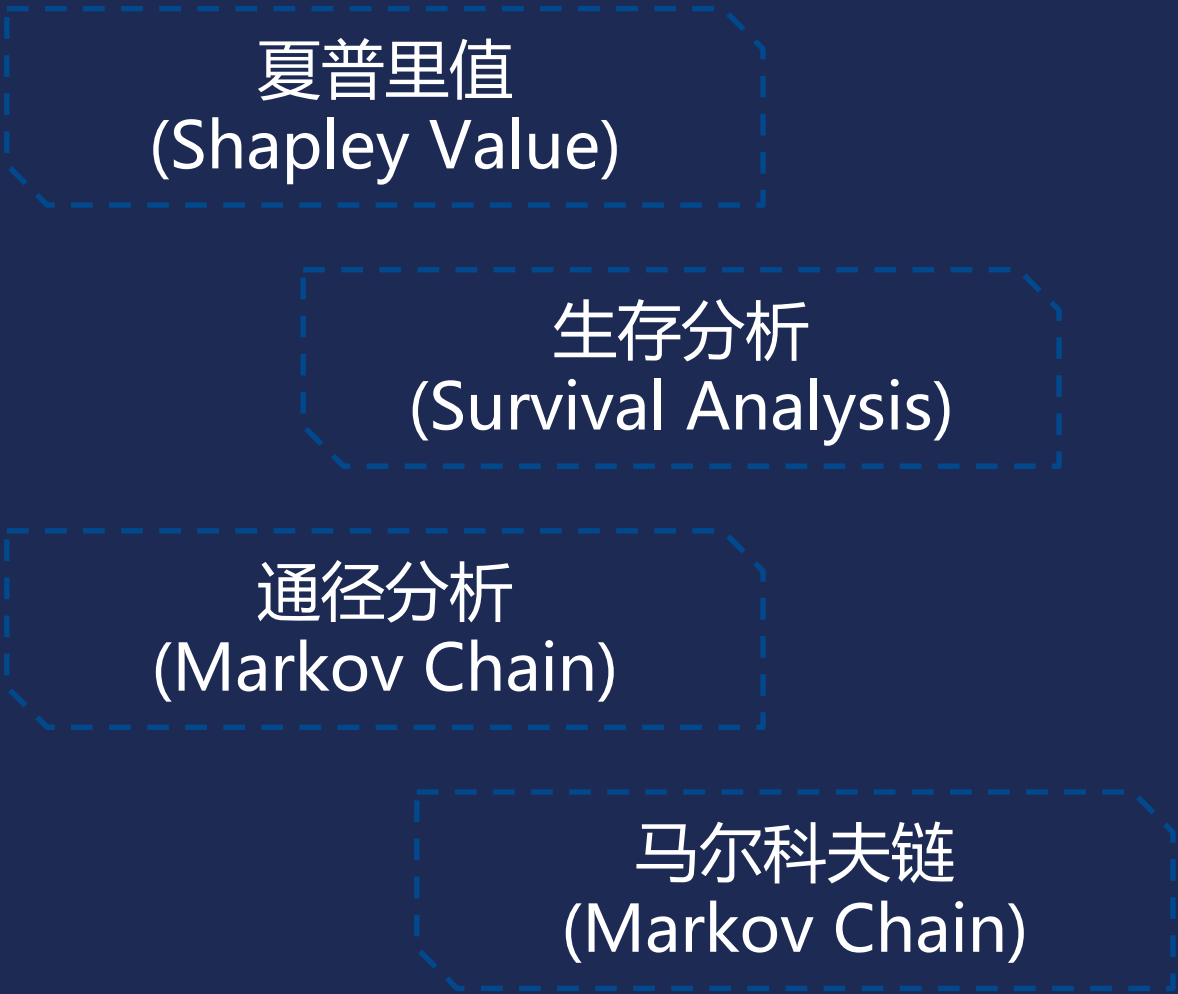
## Multi-Touch Attribution Model



## 跨渠道归因分析

线性回归模型  
时间衰减模型

首次互动归因  
首次营销互动归因  
最后互动归因  
最后营销互动归因  
按位置加权归因



	43.76% AccountingActivity	32.10% 测试地址名称	14.04% AccountingActivity	10.30% 测试地址名称
	6.18% BillActivity	8.23% ClassificationActivi...	1.98% BillActivity	2.64% ClassificationActivi...
	2.80% 起始	0.01% 起始	0.90% 起始	0.00% 起始
	1.75% ClassificationActivi...	0.00% BillActivity	0.56% ClassificationActivi...	0.00% BillActivity
	1.02% SettingActivity	0.00% LoginActivity	0.33% SettingActivity	0.00% LoginActivity
	0.92% WallpaperActivity		0.29% WallpaperActivity	
	0.73% LoginActivity		0.23% LoginActivity	
	0.27% ExchangeRateActivity		0.09% ExchangeRateActivity	
	42.57% 离开应用	3.41% 离开应用	13.66% 离开应用	1.10% 离开应用

测试地址名称	访问次数	访问人数
	75,528	14,729

# CONTENTS

---

- 移动分析工具
- 营销推广与效果监测
- 互利共生的服务生态

# Growth Hacking

自传播(Refer)

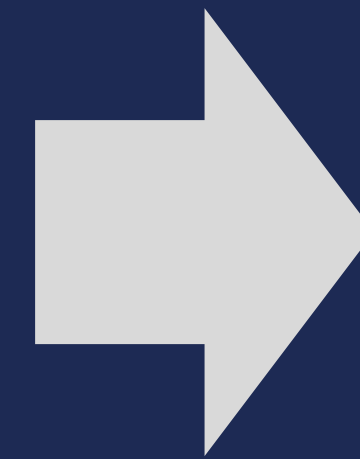
获取收入(Revenue)

提升留存率(Retention)

提高活跃度(Activation)

获取用户(Acquisition)

洞察 ( Awareness )



## 常见运营方式

付费广告

社交媒体推广

联盟营销

A/B Test

免费增值服务



# 广告效果监测

## 人群选取

- 用户分群
- 相似人群拓展

## 广告投放

- 一键导出到WIN
- 一键导出到广点通

## 效果监测

- 对接标准广告渠道
- 对接定制化渠道

## 流量清洗

- 异常流量识别

# 用户分群

## 用户分群

### 基于规则

活跃、非活跃用户

App打开次数、时间

漏斗模型

访问页面

自定义事件

用户行为事件、参数

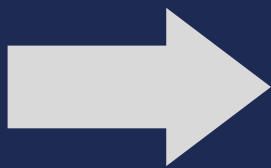
用户、设备画像

机型、年龄、性别

### 基于算法

预测流失用户

高潜力用户



## 广告平台

广点通

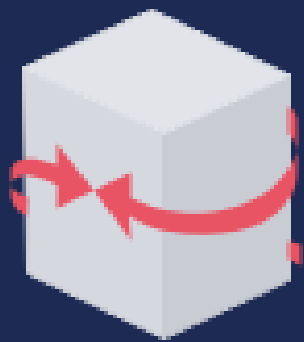
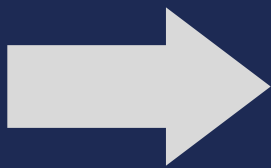
腾讯效果推广

win  
智赢销

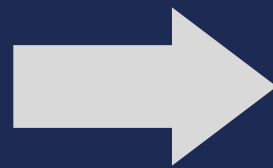
用户洞察

定向广告投放

LookAlike 扩散



数据开放网关



## 业务自建平台

运营分析

数据资产

# 相似人群拓展

腾讯数据应用 —— 某人群Lookalike - QQ群分类数据 - 按TGI排序TOP20

	群类目	种子用户 各类目下的用户数	总用户数	种子	大盘	类目下 覆盖人数	TGI
	中档车	280***	1274***	2.20%	0.40%	1638***	556
	改装与保养	428***	1274***	3.36%	0.80%	3332***	418
	低档车	120***	1274***	0.94%	0.27%	1110***	351
	二手车	77***	1274***	0.60%	0.20%	8108***	309
	豪华车	8***	1274***	0.06%	0.02%	1035***	251
	高档车	11***	1274***	0.09%	0.03%	1449***	247
	租车	60***	1274***	0.47%	0.21%	8546***	228
	保险	29***	1274***	0.23%	0.17%	6859***	137
	水上运动	4***	1274***	0.03%	0.02%	1008***	129
★	家装建材	291***	1274***	2.28%	1.77%	7354***	129
	精品男装	2***	1274***	0.02%	0.01%	5063***	128
★	建筑工程	28***	1274***	2.20%	1.79%	7401***	123
★	户外探险	22***	1274***	1.77%	1.45%	6027***	121
	个人租房	58***	1274***	0.46%	0.38%	1575***	120
★	机械器材	15***	1274***	1.22%	1.08%	4465***	113
★	国内游	37***	1274***	2.96%	2.70%	1117***	110
	主题旅游	5***	1274***	0.04%	0.04%	1542***	105
	二手房	9***	1274***	0.07%	0.07%	2895***	101

\* TGI : 即Target Group Index ( 目标群体指数 ) 反映目标群体在特定研究范围内的强势或弱势

# 数据开放 一键导出

## 用户群列表 (每个应用至多创建20个用户群)

排列方式：

修改时间

名称

启用状态：

不限

已启用

已停用

已过期

画像状态：

全部

?

新建用户群

11091529

人群规模

----

最近修改时间：2017-11-09 03:29:45

覆盖条件

1种活跃条件

1个事件

开始洞察

编辑

导出

11091528

人群规模

----

覆盖条件

1种活跃条件

1个事件

开始洞察

编辑

导出

## 人群导出

10月活

用户群名称：

请输入

创建人：

请输入

导出方式：

全部

创建日期：

2017-10-12 至 2017-11-11

搜索

重置

+ 创建导出任务

任务创建时间	用户群名称	导出方式	创建人	状态 ?	操作
2017-11-10 17:20:28	10月活跃用户	WIN	<div></div>	<div></div>	删除
2017-11-10 17:19:54	10月活跃用户	广点通	<div></div> 90468)	<div></div> 已推送至广点通	删除



# 定制化渠道对接方案



# MMA标准：异常流量识别

## 常规无效流量General Invalid Traffic , 简称GIVT

- 非浏览器用户代理头或其它形式的未知浏览器带来的流量
- 来自声明的机器人爬虫的流量
- 依据投放项目维度的参数设定，超出合理访问频度、访问时间间隔等目标设定的流量
- 已知的来自数据中心的流量执行了预加载、且没有指定加载后触发时间的流量
- 媒体方发起的对广告主提供的点击链接的定期例行检测流量
- 来自已知高危作弊来源的流量，依据预设列表判断
- 带有非法参数及字符的流量，包括通过伪造等非正常手段生成的唯一标识等关键参数
- 基本信息缺失或信息矛盾的流量（例如缺失用户代理信息）

## 复杂无效流量Sophisticated Invalid Traffic , 简称SIVT

- 高度模拟真人访客的机器人和爬虫流量
- 广告插件、恶意软件产生的流量
- 被劫持的设备产生的流量
- 虚拟化设备中产生的流量
- 被劫持的广告代码产生的流量
- 广告伺服时故意隐藏、叠加，掩盖、自动刷新或使过程模糊
- 通过作弊代理服务器产生的流量
- 以金钱补偿为动机的操纵测量数据行为
- 伪造的可见性曝光判定属性参数的流量
- 篡改或重复使用窃取获得的Cookie信息的流量

# 流量清洗

## 实时检测

动机洞察 | 羊毛党

流量作弊

渠道作弊

## 离线清洗

多维交叉验证

先验数据分布

## 评估

渠道信用评估

流量评估

第三方验证

## 规则

运营指标

规则指标

异常指标：IP、地域、IMEI

## 模型

XGBoost

Outlier检测

关系图谱

FP Growth

K-means 聚类

策略提取

## 对抗网络

GANs

强化学习

## 特征

LBS

账号

时间

跨应用归因漏斗

设备指纹

行为轨迹

渠道特征

信用等级

## 黑白名单

设备标识

账号

案例库

样本库

## 黑产

模拟器

真机墙

假激活

僵尸网络

ROM 后门



# CONTENTS

---

- 移动分析工具
- 营销推广与效果监测
- 互利共生的服务生态

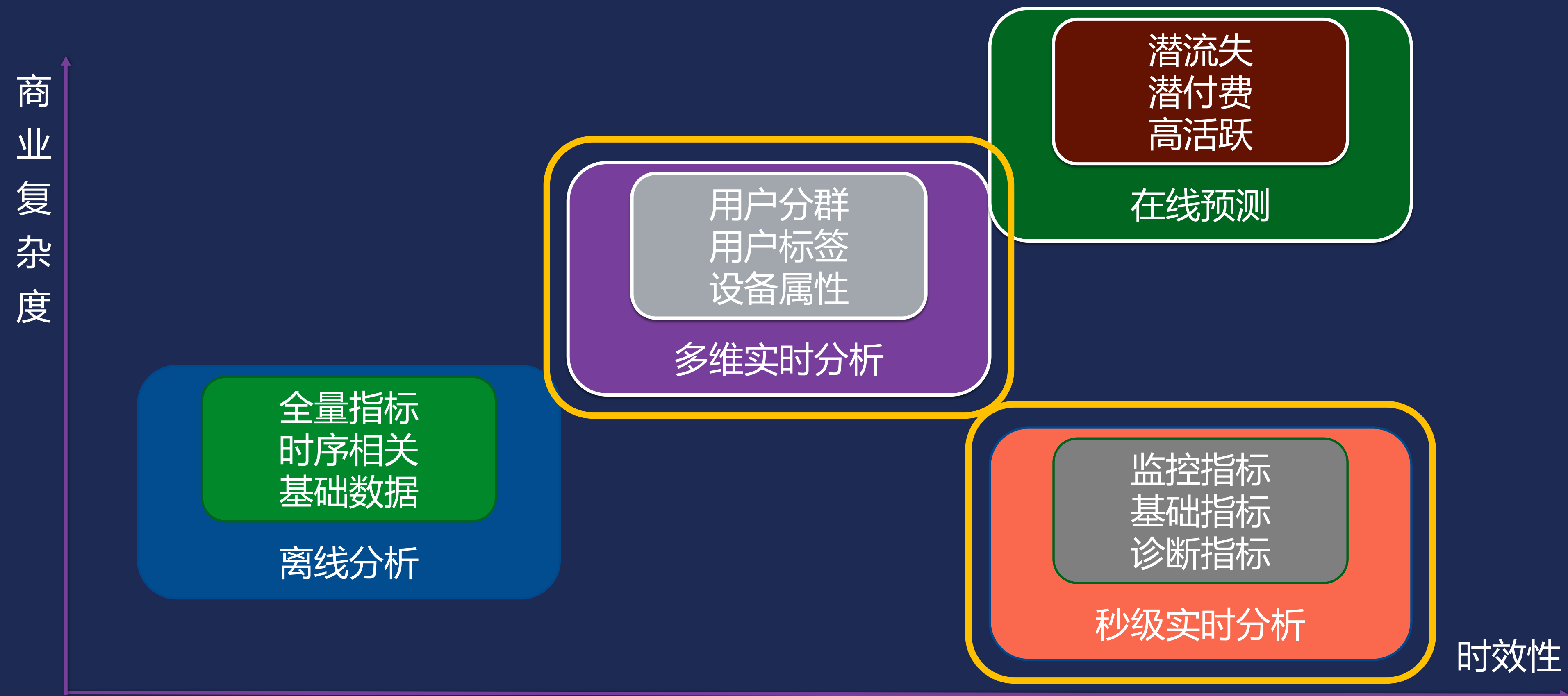


# 数据驱动产品运营

挑战：数据采集、平台建设、算法模型 .....



# 指标分级治理



# 基础技术架构

分级数据接入

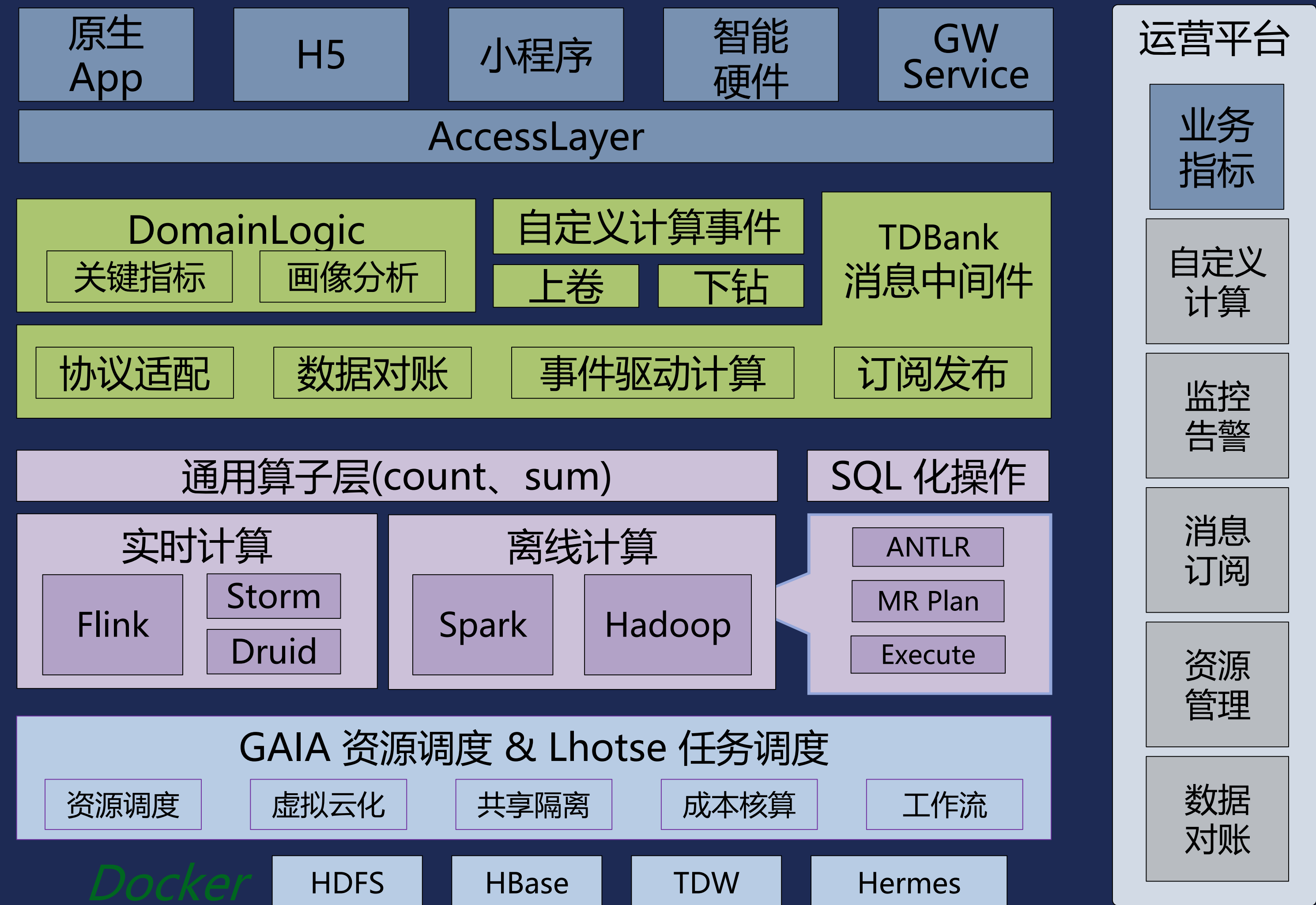
指标报表呈现

自定义计算事件

资源任务调度

数据存储、计算

用户、设备画像



# 大数据服务组件

## 数据实时

TDBank  
(实时接入)

TRC  
(实时计算)

Hermes  
(实时检索)

Spark Streaming  
(流式计算)

EasyCount  
(实时SQL)

## 平台

IDE & ideX  
(数据分析)

LZ  
(调度)

TeslaML  
(机器学习)

## 计算

SQL

Graph

ML & Deep Learning

Hive

SparkSQL

GraphX

MLlib

Tensorflow

Caffe

Torch

MR

Spark

Angel

## 调度

Gaia

Docker on Gaia

## 存储

非结构化

结构化

KV

对象

关系型

图

HDFS

Hive表

HBase

Ceph

tPG

S2Graph

## 数据应用

黄金眼  
(报表引擎)

IDEA  
(自助提取)

FACE  
(用户画像)

ADS  
(数据交易)

## 数据安全

UPS  
(权限中心)

META  
(元数据管理)



# 赋能企业大数据能力



1

可视化操作，降低门槛

2

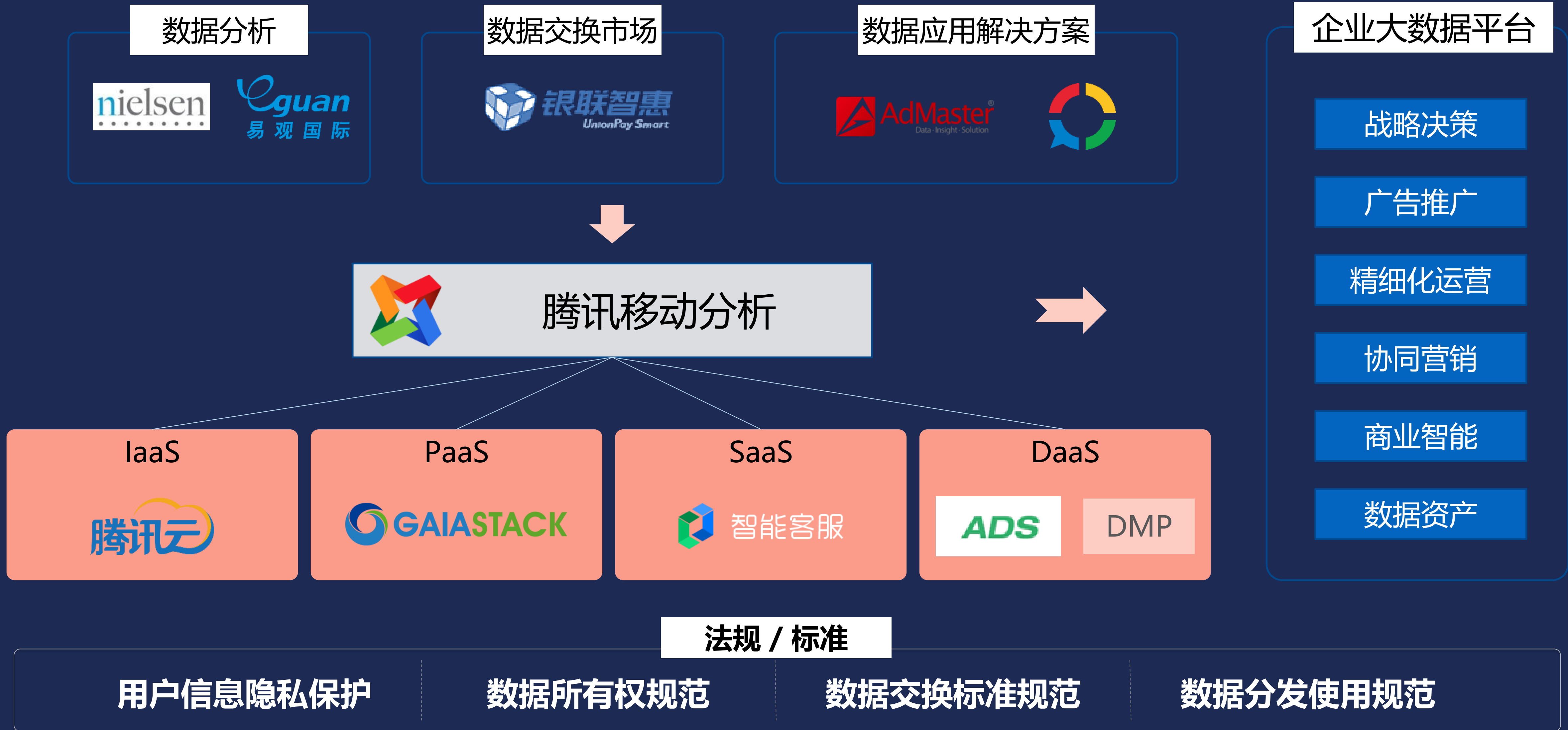
平台开放，协作或自建

# 平台开放、协作或自建



大数据套件								
自建	平台规划	平台运营	数据采集	流式计算	批处理	算法平台	资源调度	
	安装部署	定制方案	分布式存储	机器学习	应用系统	检索分析	资产管理	

# 共生的数据服务生态





# Thanks!



腾讯移动分析

联合主办方:  腾讯云 | 

直播支持:  腾讯课堂  
KE.QQ.COM 学习成就梦想