

如何高效进行数据采集？

GrowingIO 高级技术顾问 王汉

目录
CONTENT

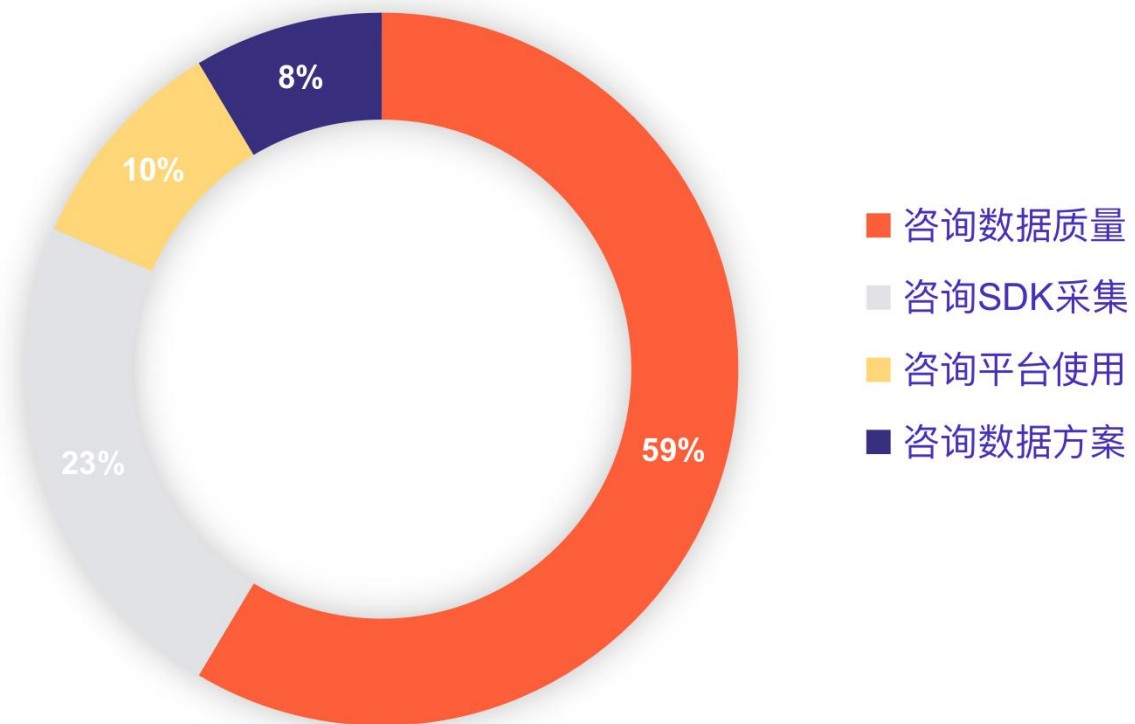
- 1 聊聊「数据采集」
- 2 不同场景下该如何选择埋点和无埋点?
- 3 完整埋点方案设计的 4 要素
- 4 1 套完整的数据采集协作流程



聊聊数据采集

数据采集带来的数据质量问题

2019年 GrowingIO 技术支持部门收到的咨询问题



咨询数据采集带来的数据质量问题相关的问题高达58%；

而这些问题的发生正是由于数据采集阶段：

- 前期沟通业务不明确；
- 采集时机口径对不齐；
- 采集点没有统一管理；
- 版本更新等。



数据采集方式：埋点和无埋点的适用场景

不同场景下的埋点/无埋点该如何选择？

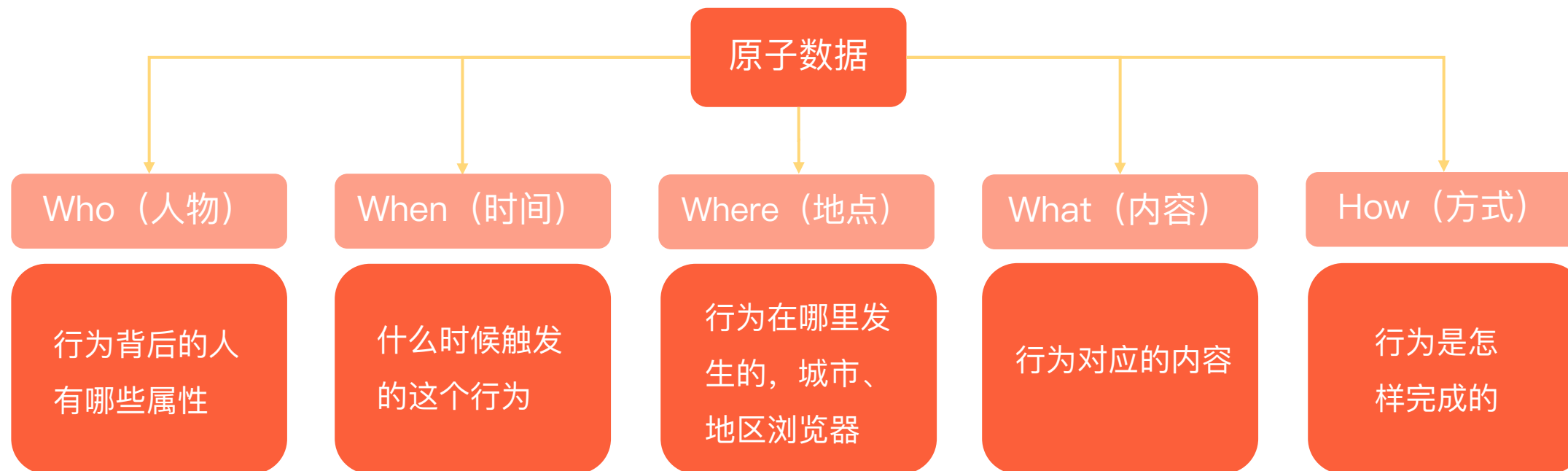
您是否遇到过这样的场景？



- 做了一场运营活动，但没有埋点的产研资源； --需要人力
- 想衡量交互细节，而需要查看的交互细节非常多； --繁琐
- 想查看一个用户在访问时的一切行为轨迹，探索用户使用产品场景；
- 产品每周发版，想要快速衡量每次发版的效果；
- 要做一个分析发现没有所需事件，来不及打点积累数据，但又要尽快出结论；
- 当新功能上线时，你突然发现有一个重要的元素忘了埋点；

.....

什么是无埋点（圈选）？



结论：无埋点即是 **人物、时间、地点、内容、方式** 的数据采集方式。与无埋点相呼应的就是可视化定义工具（圈选）可以让用户所见即所得地定义指标，查看实时数据。这就是我们 GrowingIO 提到的圈选功能。圈选的核心思想就是基于5个元数据，建立规则引擎，做数据匹配。

无理点的最佳使用场景



- ✔ 业务属性弱，交互属性强；
- ✔ 需求及时性强，要快速落地得出结论；
- ✔ 数据使用周期短，不需要长期监控；
- ✔ 相比准确性，更需要趋势稳定；
- ✔ 非核心数据，数据可及性强。

不同场景下的埋点/无埋点该如何选择？

面对不同的场景，我们需要明确目的是什么。结合各种情况综合判断，去采用更合适的数据采集方式。

数据采集方式	优势	劣势	适用场景
埋点	<ul style="list-style-type: none">数据定义清晰，稳定性高可以添加业务属性以支持维度拆解和下钻分析	<ul style="list-style-type: none">需提前规划，跨团队协作确定埋点方案历史数据不能回溯	适用于「监控与分析式」数据场景： 核心 KPI 数据 需要长期监控和存储 业务属性丰富
无埋点	<ul style="list-style-type: none">自主性高，便于灵活采集；可以回溯过去7天数据	<ul style="list-style-type: none">受制于产品开发框架和开发规范事件级维度无法拆分，且无法采集滑动等行为	适用于「探索式」数据场景： 交互属性强 突发问题快速及时分析 作为补充数据相互印证

无埋点指标特点：确定快，有历史数据，有预定义维度加持，适合快速查看某些趋势形或流程型数据

埋点特点：灵活、稳定、局限性低、精度高，适合跟踪关键节点，隐藏程序逻辑搭配业务维度观察的数据



完整埋点方案设计的四要素

准备埋点需求方案

1

确认事件
与变量

2

明确事件的
触发时机

3

规范命名

4

明确实施
优先级

1. 确认事件与变量

1

确认事件 与变量

- 将事件视为产品中的操作，例如加入购物车、支付成功等；
- 将变量视为描述事件的属性，“不同商品的加购次数”，“商品名称”就是变量；
- 创建新的事件还是新的变量？
 - A. 事件1：商品a购买；事件2：商品b购买
 - B. 事件：购买；变量：商品名称

基于数据需求，找到事件与变量搭配的最优解。

希望分析性别对于各商品品类支付成功的占比;请问,这里面什么可以是指标?维度又是那些?

指标:支付成功事件
维度:性别/商品品类



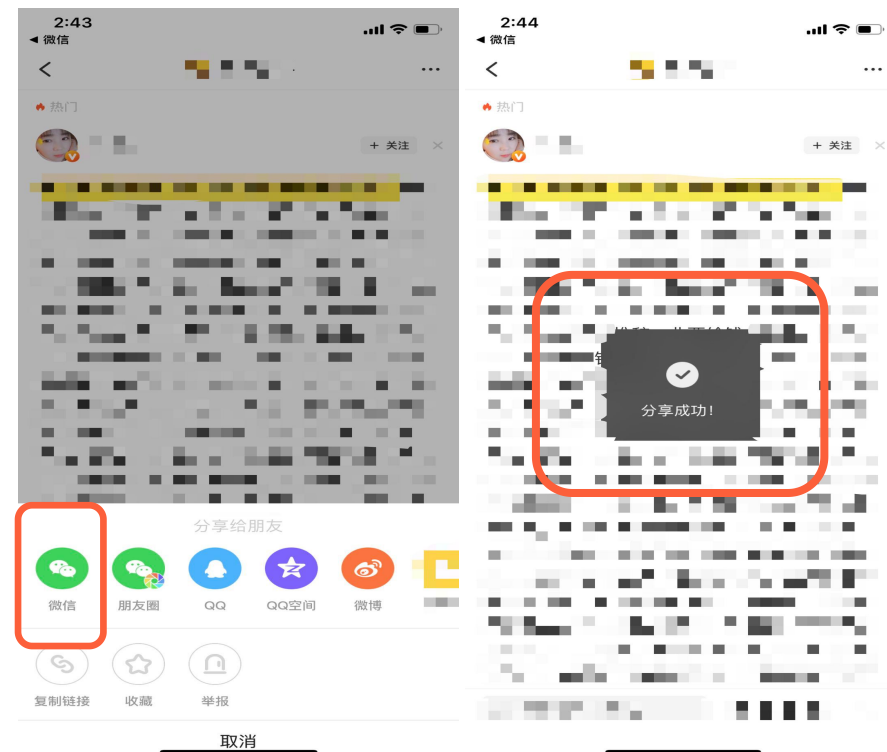
2. 明确事件的触发时机

2

明确事件的触发时机

- 何时需要记录这个事件
 - 以“分享成功”事件为例：不同的时机，带来不同的“分享成功率”
- 所有数据使用者需要明确触发时机

不同的触发时机 = 不同的数据口径



3. 规范命名

3

规范命名

- 动词+名词 or 名词+动词
 - Eg: 加入购物车 — addToCart
 - Eg: 商品点击 — produceClick
- 确保事件命名规范一致
- 驼峰法命名

统一命名规范有助于提高数据的实用性

4. 明确实施优先级

4

明确实施 优先级

- 基于业务目标，优先实施产品关键路径的所有事件；
e.g. 以电商为例，购买流程的关键事件应当优先实施
- 业务目标优先级相同需要参考技术实现成本
- 同样技术实现难度成本参考业务数据价值更高的

通过明确优先级，专注于产品中需要跟踪的真正重要事件；
实现价值的持续交付

埋点方案文档格式样例

基于上述四要素来完成埋点方案设计，不仅可以提升需求方与开发团队的协作效率，更能为后期的数据提供质量保障。

模块/场景	优先级	是否埋点?	是否验证?	标识符	名称	触发时机	类型	描述	事件级变量 (属性)			
									事件级变量名称	事件级变量标识符	类型	取值示例
核心购买转化 商品/品类运营 优惠券	中	是	验证通过	productClick	商品点击	点击商品时触发	计数器	衡量商品点击次数	商品ID	productId_var	字符串	
									商品名称	productName_var	字符串	
									一级品类	firstCat_var	字符串	
									二级品类	sndCat_var	字符串	
									三级品类	thirdCat_var	字符串	
									所属页面	productPage_var	字符串	首页 商品列表页 商品详情页
	高	是	验证未通过	productDetailView	浏览商品详情页	打开商品详情页时触发	计数器	衡量商品详情页打开次数	楼层	floor_var	字符串	
									坑位	position_var	字符串	
									商品ID	productId_var	字符串	
									商品名称	productName_var	字符串	
									一级品类	firstCat_var	字符串	
									二级品类	sndCat_var	字符串	
	高	是	未验证	addToCart	加入购物车	加入购物车成功时触发	计数器	衡量加入购物车成功次数	三级品类	thirdCat_var	字符串	
									商品ID	productId_var	字符串	
									商品名称	productName_var	字符串	
									一级品类	firstCat_var	字符串	
二级品类									sndCat_var	字符串		
三级品类									thirdCat_var	字符串		
高	否	未验证	createOrder	生成订单	生成订单时触发	计数器	衡量订单生成次数	购买数量	quantity_var	整数		
								订单ID	orderId_var	字符串		
								实际支付金额	payAmount_var	小数		
								是否使用优惠券	ifCouponUsed_var	字符串		
								优惠券类型	couponType_var	字符串		
								优惠券名称	couponName_var	字符串		

扫码添加「增长顾问」
获取埋点方案文档





1 套完整的埋点协作流程

团队协作，是完成埋点的关键因素

如何又快又准确认埋点需求？

快

- 需要需求方（产品经理、运营）、数据规划师（数据产品、数据分析师）、开发三个团队有序协作

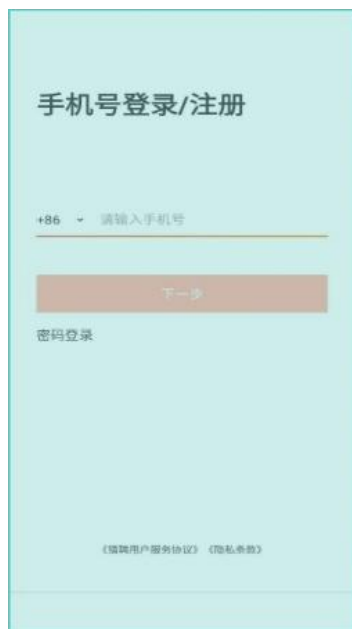
准

- 需要确保数据的「业务含义」和「数据质量」

完整的埋点协作流程



快速上手：注册场景流程拆解



注册首页



注册验证



注册信息A



注册信息B



注册信息C



APP首页

场景分析 ①

需求：希望快速分析现有注册流各个步骤间的转化率，从而找到流失较大的环节进行优化。

指标定义：各个步骤间的页面



注册首页



注册验证



注册信息A



注册信息B



注册信息C



APP首页

思路：单纯关心该流程间步骤的转化流程，那么我们关心的是用户的浏览行为动作。

简单拆解：登录动作从登录首页到进入登录后的首页共6步，且我们关注角度都在预定义维度中，比如机型、地区、国家等不参与业务范畴，这就很符合我们无埋点指标的定义规则，所以我们可以快速定义出6个浏览页面指标即可完成对于数据的分析。

无埋点指标快速定义——看数快速，简单方便



漏斗步骤

+ 选择漏斗步骤

- 1 浏览 | GIO注册流|注册首页
- 2 浏览 | GIO注册流|短信验证页
- 3 浏览 | GIO注册流|信息页A
- 4 浏览 | GIO注册流|B页面
- 5 浏览 | GIO注册流|C页面
- 6 浏览 | GIO首页|职位

目标用户

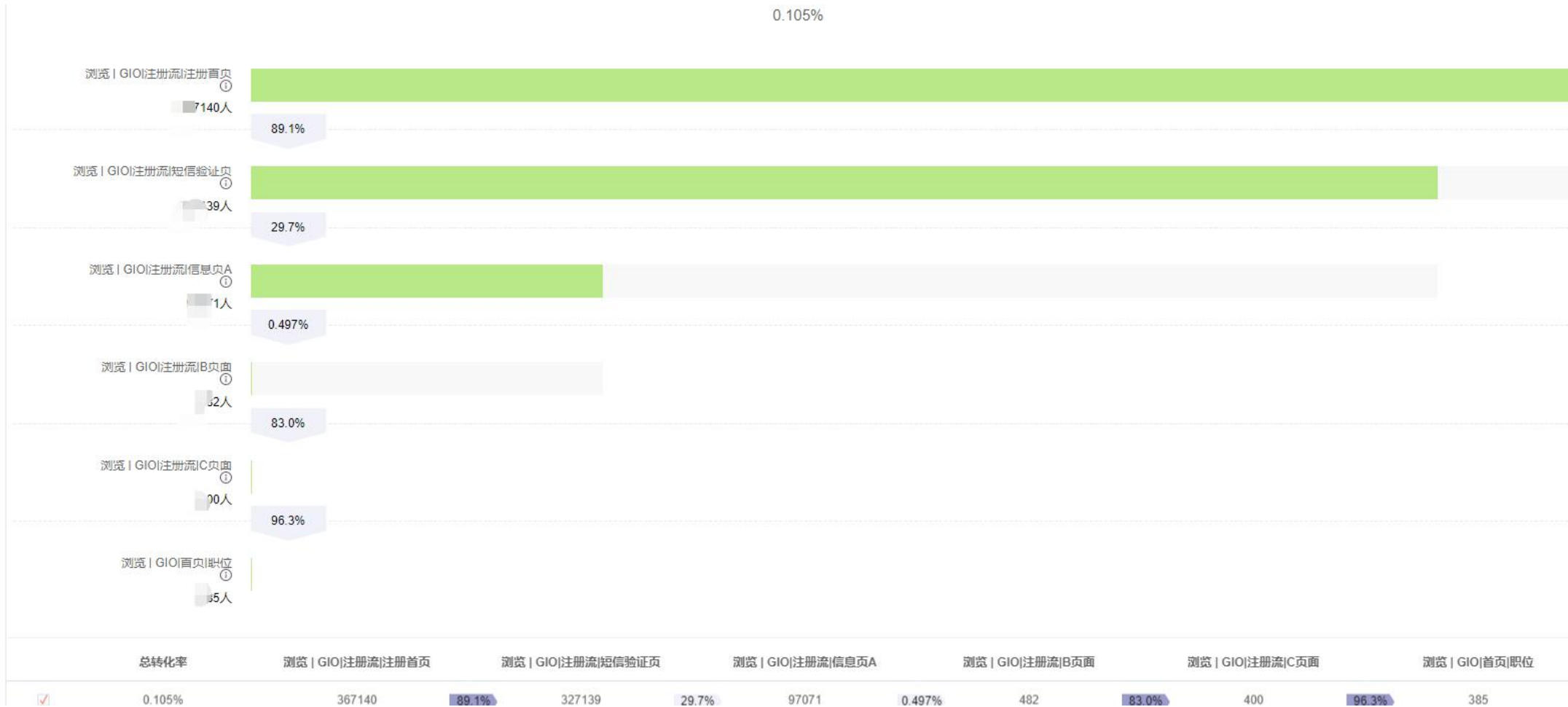
全部访问用户

全局过滤条件

+ 选择过滤条件

维度对比 用户对比

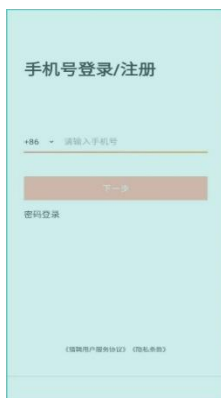
+ 选择维度



场景分析 ②

需求：希望查看完成注册的用户中，填写实习行业和性别的分布情况。

关键指标：注册完成、实习行业



注册首页



注册验证



注册信息A



注册信息B



注册信息C



APP首页

根据埋点选择，以及 4 要素的思路顺序逐一确认：

① 谁来做事件，谁来做维度？ ② 采用无埋点还是埋点指标？ ③ 埋点的触发时机是什么？

④ 确认命名：

事件命名：registerSuccess 注册完成； 维度命名：practiceVocation 实习行业、sex 性别

⑤ 确认优先级：根据数据价值决定

事件页面	是否埋点	标识符	事件名称	类型	埋点方式	事件级变量名称	事件级变量标识符	类型	iOS	Andro	版本	取值示例	触发时机	备注
注册	1	是	registSuccess	注册成功	计数器	实习行业	practiceVocation	字符串	是	是	1.X.X	服装/外贸等	onRegisterSuccess回调	咱不统计服务器异常/或网络超时等带来的注册回调
						性别	sex	字符串	是	是	1.X.X	男/女	onRegisterSuccess回调	咱不统计服务器异常/或网络超时等带来的注册回调

数据采集-最终的确认

校验关键点:

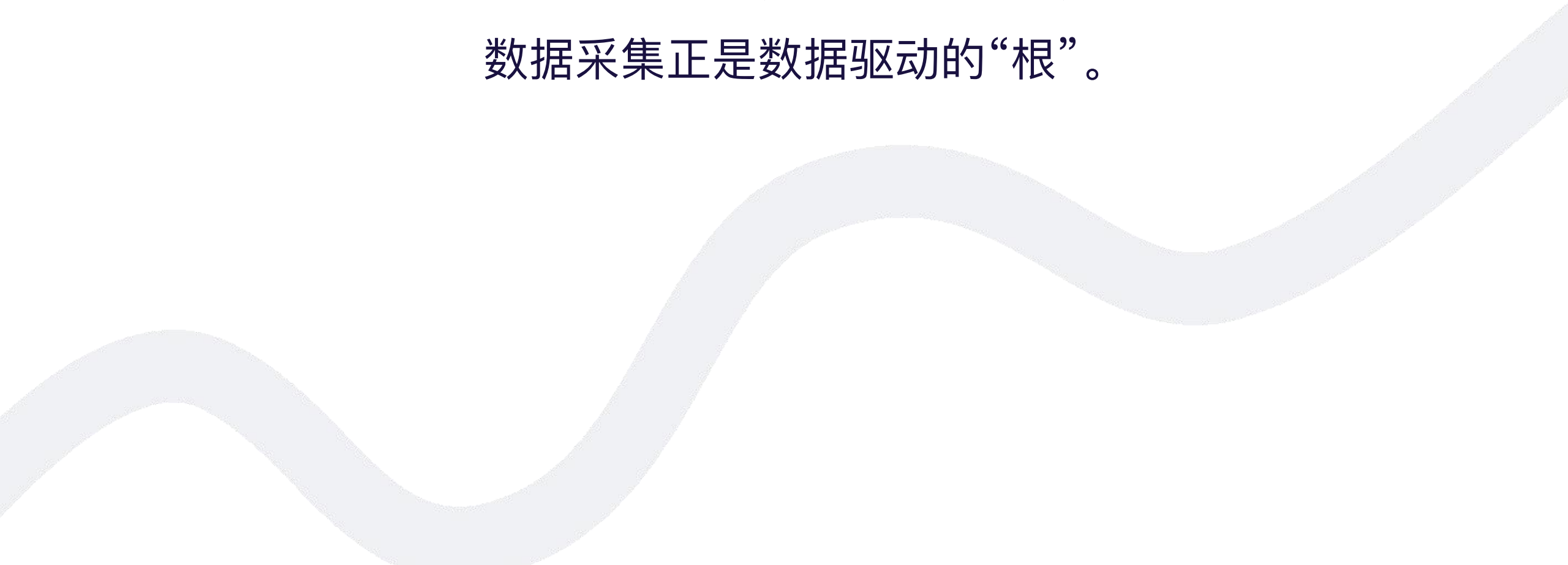
- ① 触发时机校验
- ② 无埋点+埋点动作实时显示
- ③ 埋点字段内容验证

The screenshot displays the GrowingIO data management interface. On the left, a sidebar shows navigation options like '行为信息' (Behavior Information) and '访问开始' (Access Start). The main area shows a list of events with columns for event name, time, and status. A red box highlights an event at 15:47:14 with a '错误' (Error) status. On the right, a detailed view of this event is shown, including the event name, identifier, type, and time. A red box highlights the '错误描述: 事件标识符不匹配' (Error description: Event identifier mismatch) in the '校验结果' (Check result) section. Below the error description, the '期望结果' (Expected result) is 'gio_touch_message_change' and the '实际结果' (Actual result) is 'gio_touch_message_change'. The '解决方法' (Solution) section suggests that the event identifier should be uploaded as is, and refers to a help document.

行为信息	选择事件/变量名称	事件详情
页面浏览 页面标题: GrowingIOText 路径: 测试中文	15:47:14	事件名称: 标识符: gio_touch_message_cha...
自定义事件 名称:	15:47:14 错误	类型: — 时间: 2020/05/11 15:47:14
自定义事件 名称:	15:47:14	变量名称: — 变量标识符: description 类型: —
访问开始 15:47:14		校验结果
页面浏览 页面标题: GrowingIOText 路径: CstmFunActivity	15:47:12	错误描述: 事件标识符不匹配
点击 点击内容: cstm 事件(自定义埋点事件)	15:47:12	期望结果: 实际结果: gio_touch_message_change
页面浏览 页面标题: GrowingIOText 路径: 测试中文	15:47:10	解决方法: 该事件的标识符应上传为, 接口原型请参考 帮助文档
输入 输入内容:	15:47:06	
输入 输入内容: 个	15:47:00	
页面浏览 页面标题: GrowingIOText 路径: ChngFunActivity	15:46:54	

GrowingIO 校验工具, 方便快速定位。

“柢固则生长，根深则视久”，
数据采集正是数据驱动的“根”。



提问环节



THANKS

