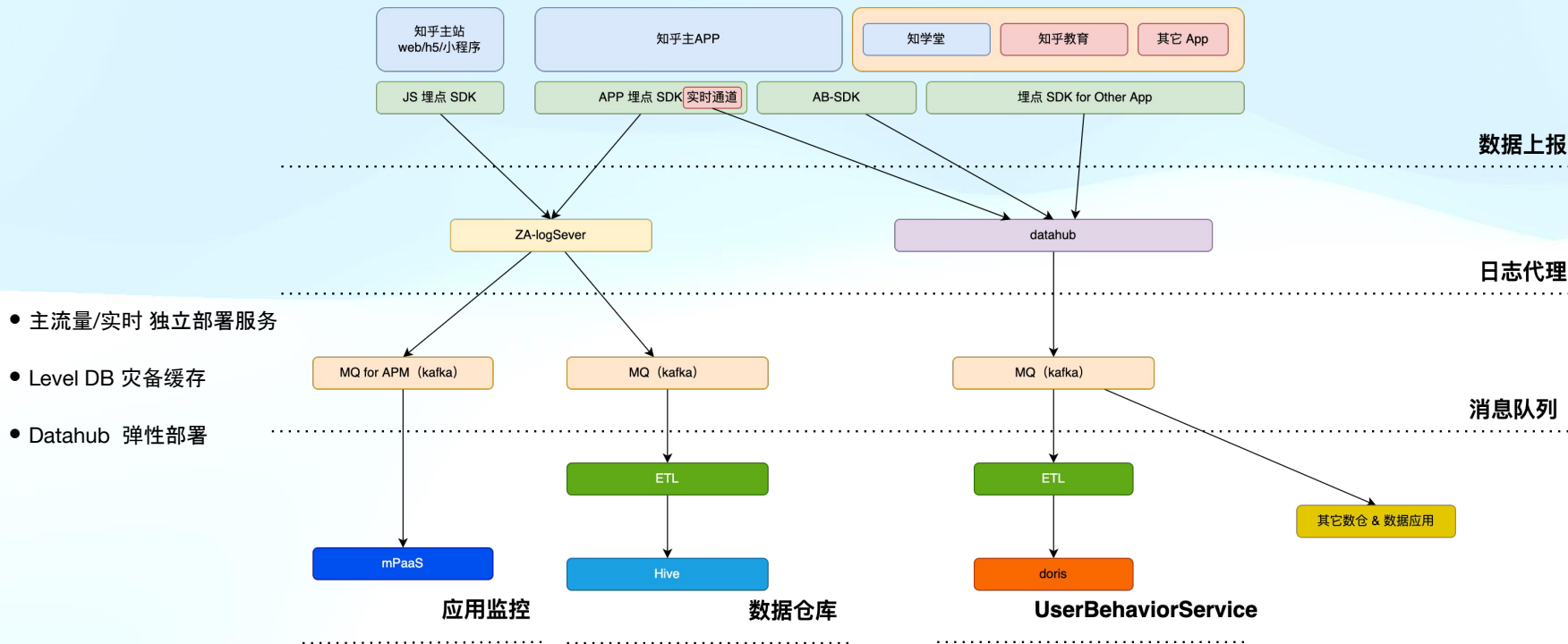


T沙龙埋点体系建设与治理

知乎大数据 张彦瑞 武蕴

整体链路介绍



埋点模型设计

- 描述用户行为的维度抽象
- 抽象资源位
- 启动归因 & 站内归因（埋点染色）

埋点模型设计

定义用户行为通用模型 - 描述用户行为的维度抽象

WHO	设备标识 用户标识	client_id/oaid/idfa/user_hash_id
WHEN	事件产生时的时间戳	client_time/send_time/server_time
WHERE	事件发生的地点	app_name/client_ip_address/ gps_info
HOW	用什么方式产生了此次行为	utm/launch_type/push_info/open_url
WHAT	用户对什么元素做了什么	channel/page/module/event_type/e xtra_info

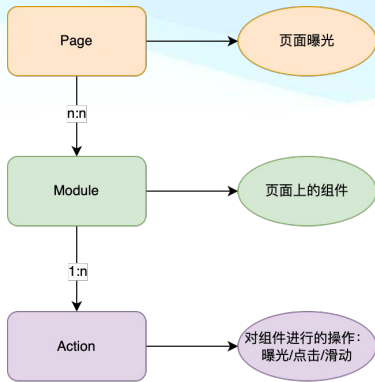
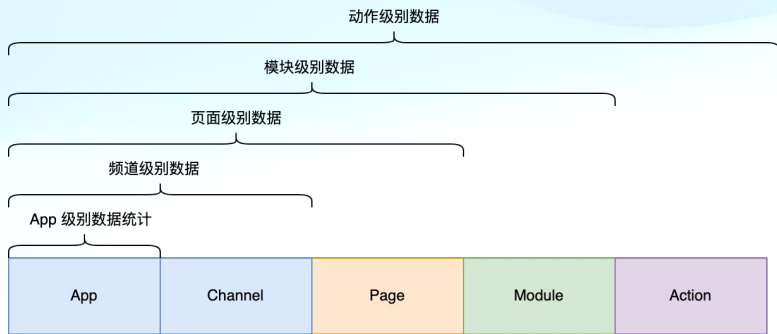
App 生命周期内相对静态

由用户的具体行为生成

埋点模型设计

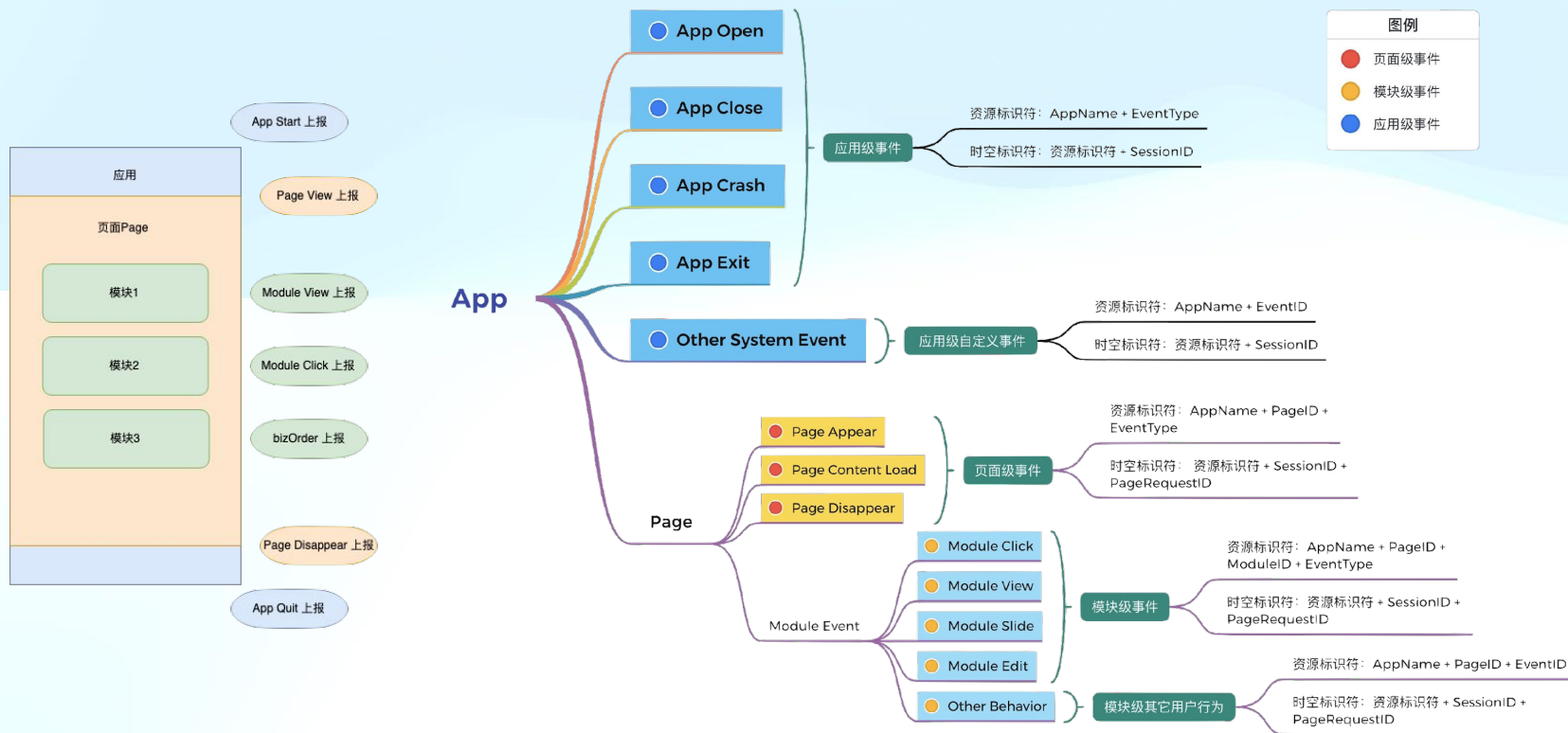
定义用户行为通用模型 - 抽象资源位

- 一个埋点 = App + Channel + Page + Module + Action
- 精准定位
- 位置运营
- 方便管理和验证
- 部门成本分摊(channel)
- 下游通过 channel 分流和自定义 ETL



埋点模型设计

定义用户行为通用模型 - 动态组合, 确定时空标识符



埋点模型设计

定义用户行为通用模型 - 事件类型概览

资源类型	资源标识符	事件类型 (EventType)	位置标识符	生命周期标识符 (runtime)	是否统一包含时长	埋点方式
App	AppName	App Open (程序活跃) App Close (程序失活) App Crash (程序崩溃) App Exit (程序进程退出)	AppName + EventType	App 级别生命周期标识符 (SDK 自动生成): Session/AppRequestID	App Close & App Exit	SDK 自动埋点
Page	PageID	Page Appear (页面展示) Page Content Load (页面内容加载完成) Page Disappear (页面消失)	AppName + Channel ID + PageID + EventType	页面级别生命周期标识符 (SDK 自动生成): PageRequestID	Page Disappear	声明式埋点 (业务参数需要手动传入, 上报时机由 SDK 统一控制)
Module	Module ID	Module Click (模块点击) Module View (模块展示) Module Slide (模块滑动) Module Edit (模块编辑) ...	AppName + Channel ID + PageID + ModuleID + EventType	依托于 App 和 Page 生命周期	否	声明式埋点 + 手动埋点
Other System Event	Event ID	非用户类的行为统计 (网络、帧率等)	AppName + Channel ID + Event ID	依托于 App 生命周期	否	手动埋点
Other Behavior	Event ID	上述类型不能覆盖的其它用户行为	AppName + Channel ID + PageID + Event ID	依托于 App 和 Page 生命周期	否	手动埋点

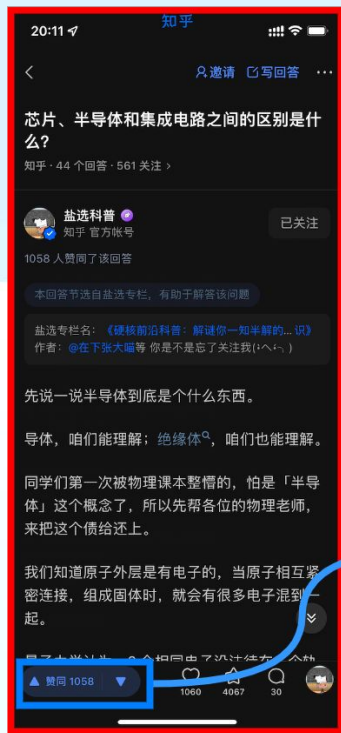
埋点模型设计

定义用户行为通用模型 - 用户行为时间线



埋点模型设计

定义用户行为通用模型 - 举例回答详情页



Page

Module

```
{
  "launch_info": {
    "launch_type": "widget",
    "session_id": "session_id_1"
  },
  "page_identify": "answer_detail",
  "page_request_id": "page_request_id_1",
  "event_type": "page_show",
  "extra": {
    "content_type": "answer",
    "content_id": "123"
  },
  "ts": 1662209316936,
  "device": { "somekey": "somevalue" },
  "user": { "somekey": "somevalue" },
  "source": {
    "app": "zhihu",
    "version": "3.12.0"
  }
}
```

```
{
  "launch_info": {
    "launch_type": "widget",
    "session_id": "session_id_1"
  },
  "page_identify": "answer_detail",
  "page_request_id": "page_request_id_1",
  "module_identify": "approve",
  "event_type": "click",
  "page_extra": {
    "content_type": "answer",
    "content_id": "123"
  },
  "extra": {
    "approve_cnt": 1058
  },
  "ts": 1662209335936,
  "device": { "somekey": "somevalue" },
  "user": { "somekey": "somevalue" },
  "source": {
    "app": "zhihu",
    "version": "3.12.0"
  }
}
```

• 业务需要填充:

- 此次 page_show 的页面标识(page_identify)
- 此次 page_show 的业务参数(extra)

• SDK 自动生成此次 page show 的 request_id, 并填充基础信息

- 基础信息包含事件发生时间、当前 APP 信息、用户启动来源、启动 session_id 等

• 当页面消失时, SDK 会自动计算页面停留时长并携带此次 request_id 上报 page_disappear 事件

• 事件处理 ETL 流程会将上述 page_disappear 事件携带的时长补充到 page_show 供离线数据直接使用

• 业务需要填充:

- 此次 click 的模块标识+动作(module_identify + action)
- 此次 click 的业务参数(extra)

• SDK 自动关联此次 click 所在页面 page show 的 request_id

• SDK 自动填充页面 extra 到 page_extra, 并填充基础信息

埋点模型设计

定义用户行为通用模型 - 启动归因 & 会话周期

启动来源

- 用户主动打开
- 推送
- 外链唤起
 - 解析 url 内的 utm 等参数

会话周期定义

- 一次会话代表一个用户的一个连续周期内的所有活动
- 用户超过一定时间未活跃，再次活跃时，我们认为是一个新的会话周期
- 一个会话周期内，用户可能会跳转到其他 App，如果再次跳转回 APP 的时候是通过 push 或者外链，我们一般会将其重置为一个新的会话周期

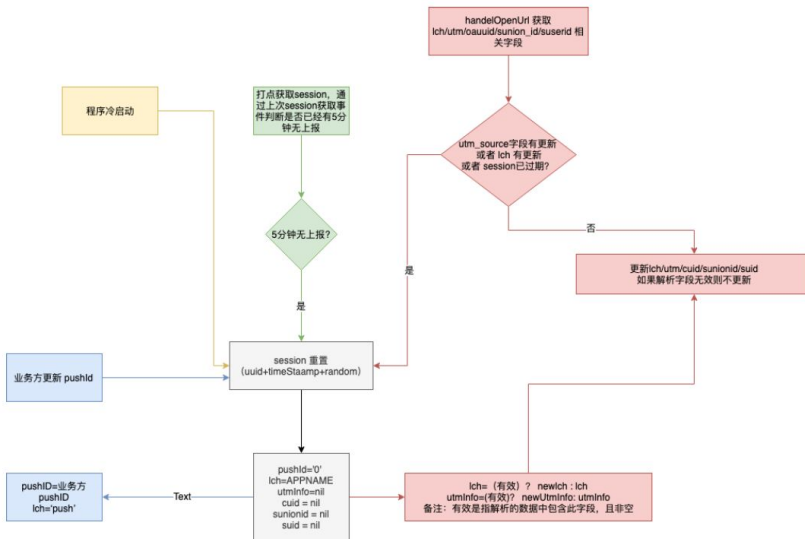
session 相关逻辑说明

黄色部分为冷启动逻辑；

红色部分为 handle Open Uri 相关逻辑流程；

绿色部分为 session id 超时逻辑流程；

蓝色部分为 push id 更新逻辑流程。

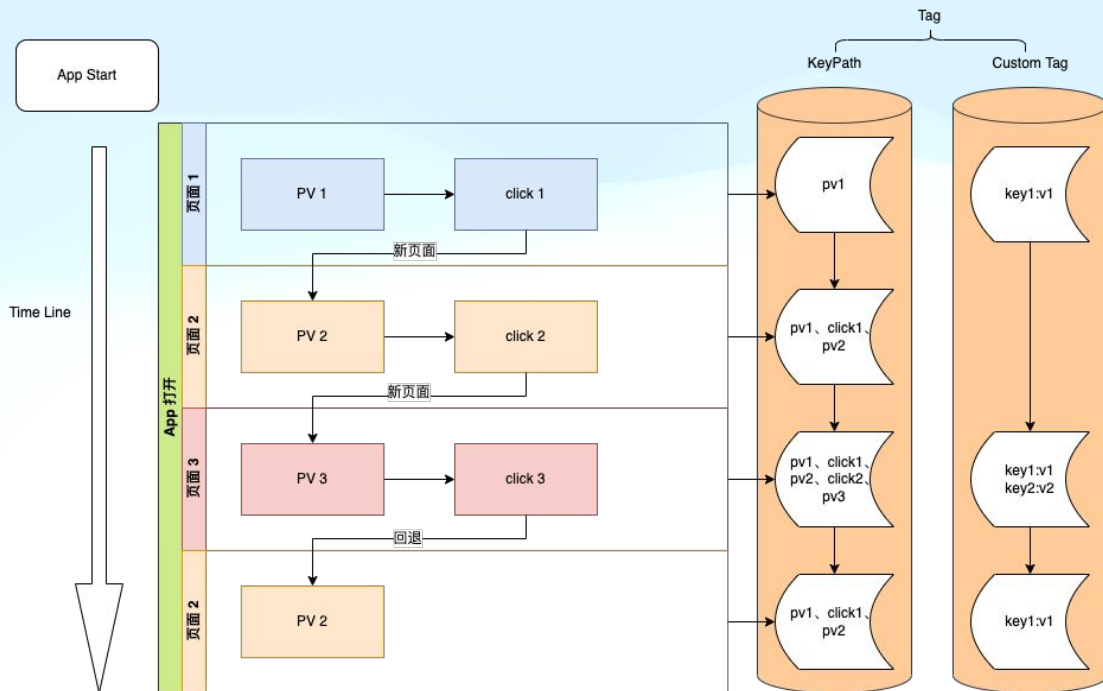


埋点模型设计

定义用户行为通用模型 - 站内归因（埋点染色）

端上染色

- SDK 会将关键路径携带到后续事件
- 关键路径生命周期与页面回退绑定
- 支持自定义染色信息
- 可支持漏斗分析、路径分析等
- 对于一些订单/支付事件，染色的意义尤其重要



埋点方案选型

常见埋点方案比较 & 知乎实践

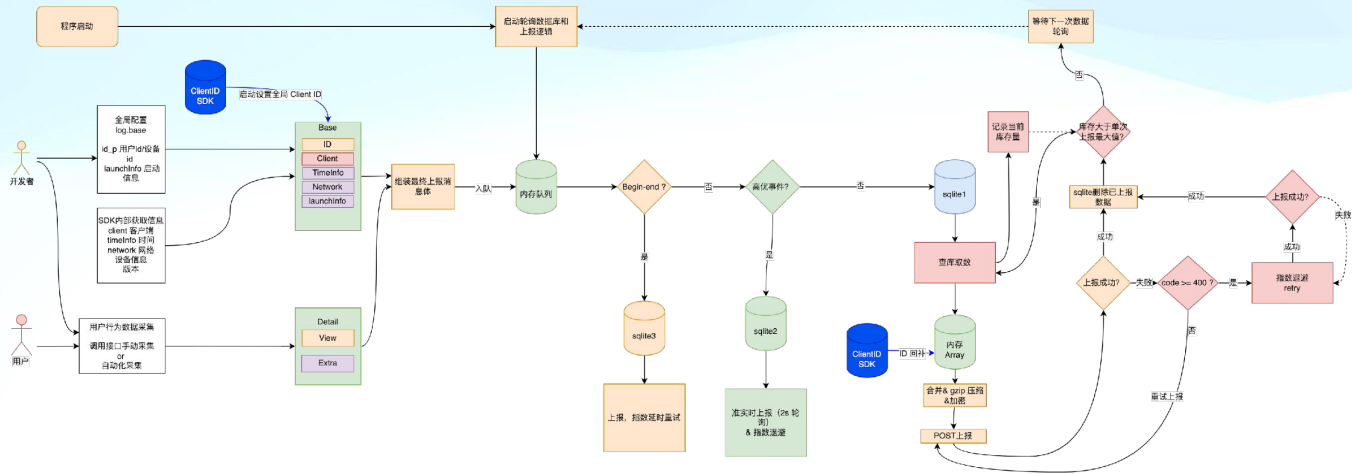
	优点	缺点
全埋点(无痕埋点)	节约开发成本, 不需要提前预置埋点	流量成本高、埋点信息不精确、遗漏重要数据、数据需要预加工、无法避免收集私密信息
圈选埋点(可视化埋点)	节约开发成本, 圈选可配置下发	圈选配置定位健壮性极低, 很容易因为发版导致定位失败
声明式埋点	开发者只需要填充内容, 不需要关心上报时机	研发成本相对较高
手动埋点	按需采集, 开发者可完全自定义上报内容和上报时机	研发成本高、业务侵入性强

- 知乎对一些系统级别的事件做了自动化埋点
 - SDK 会保障统一的在屏时长
- 对页面、各类控件分别做了扩展和封装, 用来支持声明式埋点
 - SDK 会保障统一的页面时长
 - SDK 统一接管曝光、点击时机, 研发填充业务参数
- 最佳实践: 声明式埋点+手动埋点+特定场景自动化
- 参考: [mixpanel为什么放弃无痕埋点](#)

埋点 SDK 上报质量保障 - SDK 上报通道质量保障

- 高优(App 级别)
- 普通(强交互类型日志)
- 低优(模块曝光、APM 等其他日志)

- 指数退避
- 积压数据连续上报
- 串行上报队列



埋点质量保障

- 埋点 SDK 上报质量保障
- 知乎埋点链路
- SDK 上报通道质量监控
- 埋点流程管理
- 埋点管理系统
- 语义埋点实施
- 埋点验证

埋点质量保障

埋点 SDK 上报质量保障 - SDK 上报通道质量监控

日志监控体系

- 上报成功率
- 上报延时率
- 关键字段空值率
- 发码质量 DQC

计算逻辑

成功率

去重的有效上报量
/sum[每次
session(max_seq -
min_seq)]

延迟率

$RTT = server_time -$
 $send_time$
积压时间 = send_time -
log_time

多发码

一个 shumeng_id /
oaid/IDFA 对应多个
client_id

少发码

一个 client_id 对应多
个 shumeng_id /
oaid/IDFA

空值率

各个关键字段空值率
占比
如 session/request_id/
设备信息

客户端

sequence 管
理

打点时间戳

NTP时间校
准

服务端发码

数盟
/Oaid/IDFA

规则名称/id

规则类型

库&表名

Android_client_id少发码监控
660033

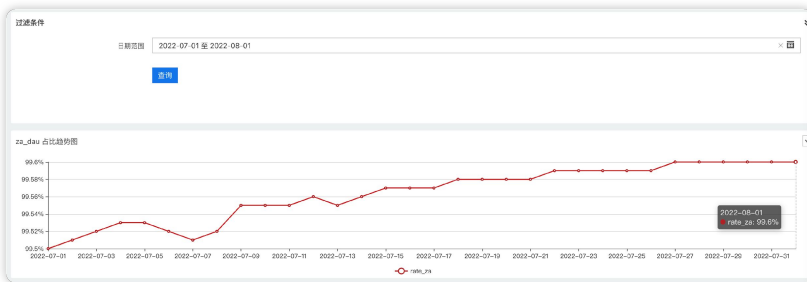
波动性检测

dws_user
dws_device_clientid_aggr_pd

Android_client_id多发码监控
570032

波动性检测

dws_user
dws_device_clientid_aggr_pd



• SQL 举例

```
select
  'ios',
  sum(
    case
      when send_time - client_timestamp > 3600000 then 1
      else 0
    end
  ) as more_than_1_h,
  sum(
    case
      when send_time - client_timestamp <= 3600000
      and send_time - client_timestamp > 60000 then 1
      else 0
    end
  ) as one_min_to_1_h,
  sum(
    case
      when send_time - client_timestamp <= 60000
      and send_time - client_timestamp > 20000 then 1
      else 0
    end
  ) as _20s_to_one_min,
  sum(
    case
      when send_time - client_timestamp <= 20000
      and send_time - client_timestamp >= 0 then 1
      else 0
    end
  ) as less_than_20s,
  count(*) as all_cnt
from
  ods_flow.ods_ ph
where
  p_date = '2022-09-27'
  and client_platform in ('iPhone', 'iPad')
  and client_version_name > '11.0'
```

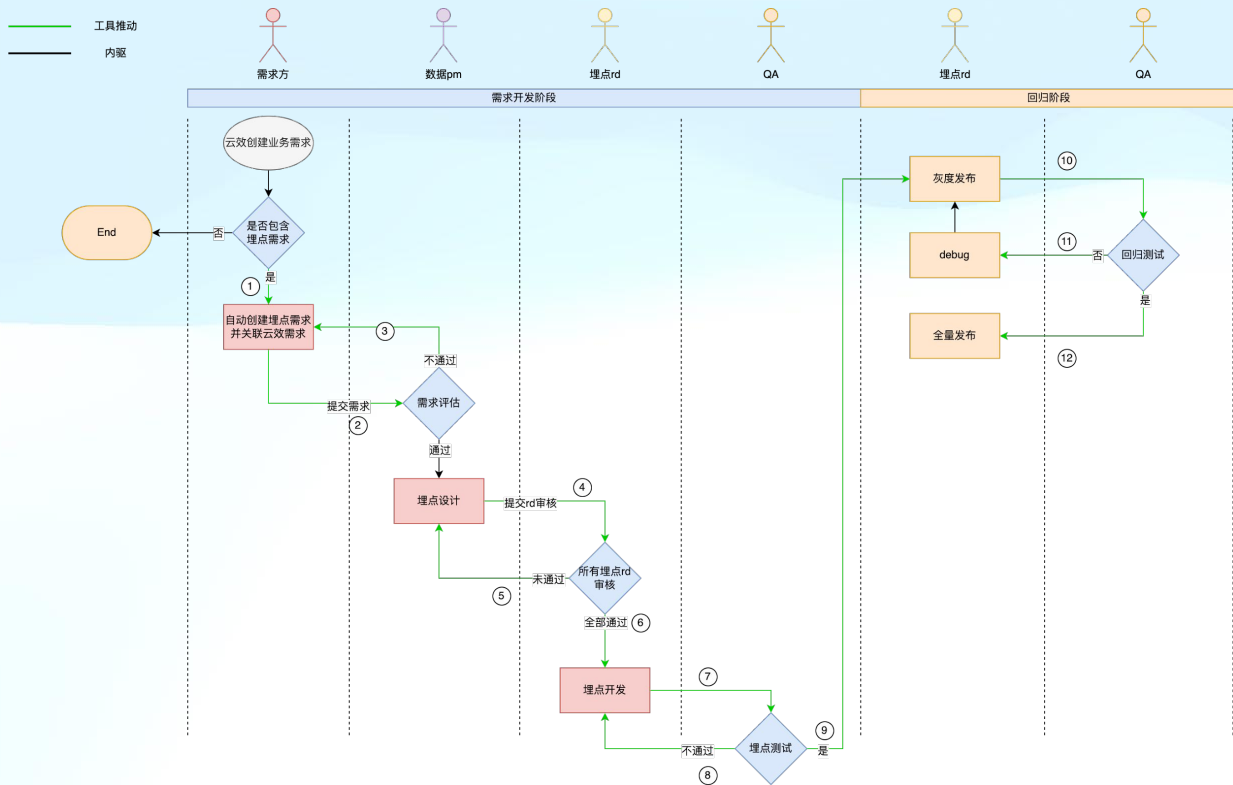
```
select
  p_date,
  sum(
    case
      when locationtype = 'IP' then 1
      else 0
    end
  ) as ip_client_id_cnt,
  sum(
    case
      when locationtype = 'GPS' then 1
      else 0
    end
  ) as gps_client_id_cnt
from
  (
    SELECT
      p_date,
      locationtype,
      client_id
    FROM
      ods_flow.ods_flow_xxx_location_ph
    WHERE
      p_date BETWEEN '2022-09-20' and '2022-09-27'
    GROUP BY
      p_date,
      locationtype,
      client_id
  ) a
group by
  p_date
```


发码监控

```
SELECT
    device_idfa,
    count(*) as client_id_cnt
from
    (
        SELECT
            device_idfa,
            client_id
        from
            ods_flow.ods_flow_v3_ph
        WHERE
            p_date = '2022-08-29'
            and p_platform in ('iphone', 'ipad')
            and client_id IS NOT NULL
            and device_idfa IS NOT NULL
        GROUP BY
            device_idfa,
            client_id
    ) a
GROUP BY
    device_idfa
HAVING
    count(client_id) > 1
```

埋点质量保障

埋点 SDK 上报质量保障 - 埋点流程管理



埋点质量保障

埋点 SDK 上报质量保障 - 埋点管理系统

定义埋点语义

- 页面语义：
 - App + 频道 + 页面 + 类型(PV / PD)
- 动作语义：
 - App + 频道 + 页面 + 模块 + 动作
- 上述组合可以确定唯一一个事件
- 语义包含各事件粒度的所有元数据

知乎 埋点平台	埋点管理	测试工具	使用说明 提交反馈	张彦奇	管理
需求管理	页面管理	模块管理	语义管理	审批管理	权限管理
base.client.product	产品	ZhihuET			
detail.view.element_location.page.page_id	页面ID	11307			
detail.view.url	页面URL	https://et.in.zhihu.com/manage/page/{page_id}			
detail.view.element_location.module_id	模块id	et_export_btn			
log_type	日志类型	Event(业务行为)			
detail.view.event_type	用户实际动作类型	Click			
detail.view.action	业务行为类型	Download			
detail.view.element_location.type	UI基础组件	Button			
config_map:					
et_trend_type		*			
et_element_type		page,module			
et_data_type		*			

埋点质量保障

埋点 SDK 上报质量保障 - 语义埋点实施

埋点平台

流量分析

埋点管理

测试工具

使用说明 | 提交反馈 | 武彦 | 普通用户

需求管理

页面管理

模块管理

语义管理

参数管理

审批管理

关联信息

关联页面: 付费咨询广场页

关联模块:

字段信息

表格模式

代码模式

Android

IOS

复制代码

```
[ZASDK :^(ZALog3Entry * _Nonnull log) {
    log.zaLogEntryNew.detail.view.url = @"https://www.zhihu.com/consult";
    log.zaLogEntryNew.detail.view.elementLocation.page.pageId = @" ";
    log.zaLogEntryNew.detail.view.elementLocation.text = @"推荐答主";
    //该行可能存在多个值, 请自行选择
    log.zaLogEntryNew.detail.view.elementLocation.content.type = proto3ContentType_Type_Conversation;
    log.zaLogEntryNew.extra.link.url = @"https://www.zhihu.com/consult/people/id";
    log.zaLogEntryNew.detail.view.elementLocation.block.blockText = @"consultation_homePage_answerTab_hotAnswerCard";
}];
```

© 2022 埋点平台

埋点质量保障

埋点 SDK 上报质量保障 - 埋点验证

- 选中一批待测语义
- 扫码连接测试设备
- 操作手机生产日志, 验证服务自动关联语义, 并校验参数情况
- 测试同学手动确定待测语义测试结果
- 生成对应的测试报告

The screenshot displays the '埋点验证' (埋点 SDK 上报质量保障) interface. The top navigation bar includes tabs for '日志序号', 'local_increment_id', '日志类型', '元素类型', '手势', '业务操作', '页面', '模块', '内容类型', 'block_text', 'location_text', '平台', '关联语义', '缩略图', '测试结果', and '操作'. The main content area is divided into two panels. The left panel, titled '语义需求' (Semantic Requirements), shows a table of test results for the log entry '5281769-带有热评的内容类蓝卡片曝光'. The table has columns for '必填参数' (Required Parameters), '需求预期值' (Required Expected Value), '实际上报值' (Actual Reported Value), and '标记' (Mark). The right panel, titled '日志原文' (Log Original Text), displays a JSON log snippet. The log entry is a root object with 8 keys, including 'local_increment_id', 'log_type', 'base', 'id', 'client', 'device', 'time', 'network', 'launch', 'detail', 'view', 'uri', 'referrer_uri', 'element_location', 'type', 'content', 'page', 'page_id', 'page_level', 'page_request_id', 'block', 'card', 'pagemeshow_trans', 'extra', 'is_sdk_auto_log', 'module_increment_id', and 'attached_info'.

必填参数	需求预期值	实际上报值	标记
detail.view.element_location.page.page_id	45	45	符合预期
base.client.product	Zhihu	Zhihu	符合预期
log_type	Show	Show	符合预期
detail.view.uri	fakeuri://billboard	fakeuri://billboard	符合预期
detail.view.element_location.type	Card	Card	符合预期

其他错误

错误类型	备注	操作

暂无数据

添加错误类型

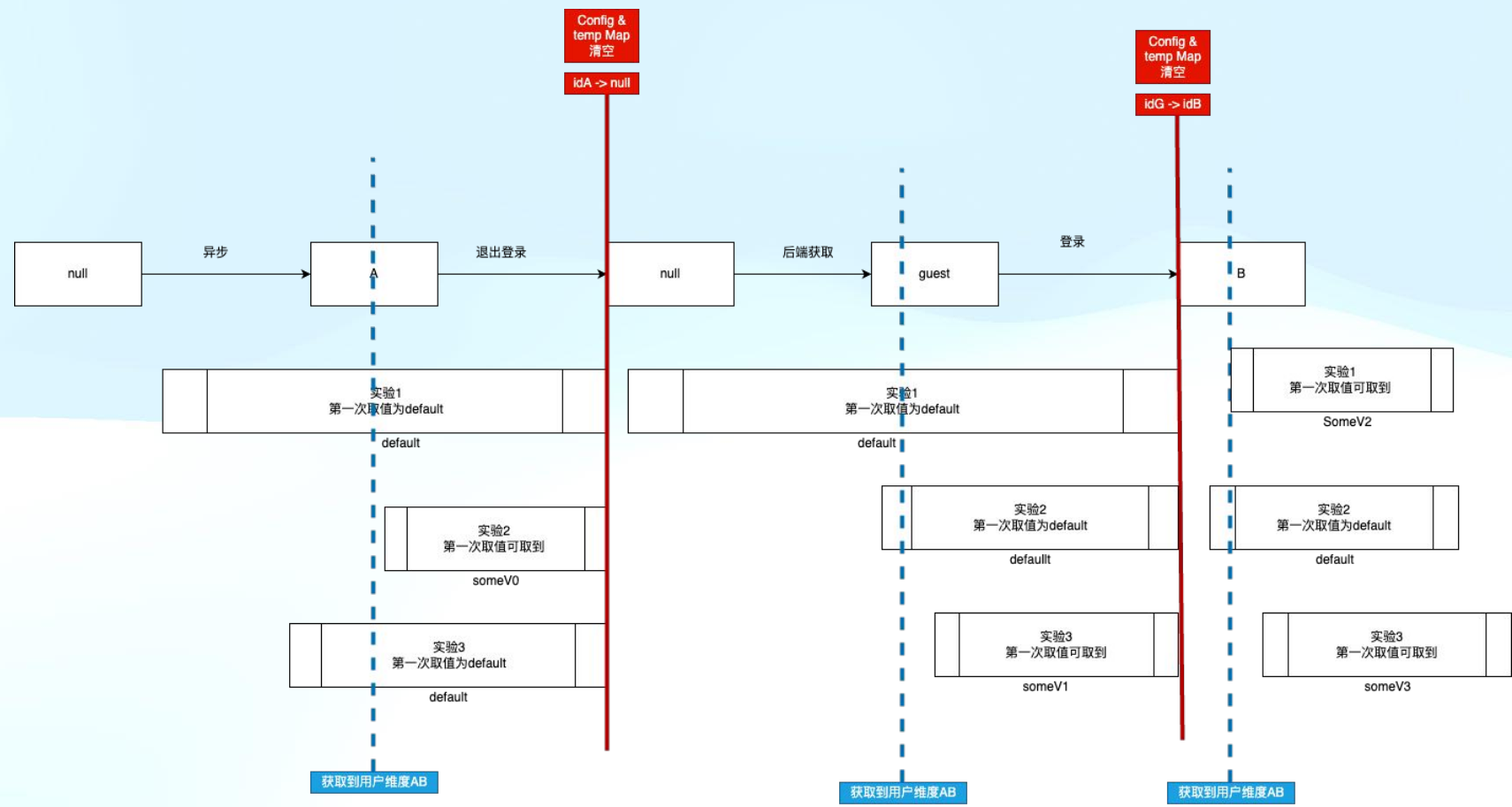
```
{
  "root": {
    "local_increment_id": 261,
    "log_type": "Show",
    "base": {
      "id": {
        "client": {
          "product": "Zhihu",
          "platform": "iPhone",
          "version_name": "8.32.0",
          "version_code": 10912
        },
        "device": {
          "time": {
            "network": {
              "launch": {
                "detail": {
                  "view": {
                    "uri": "fakeuri://billboard",
                    "referrer_uri": "fakeuri://personal_info",
                    "element_location": {
                      "type": "Card",
                      "content": {
                        "page": {
                          "page_id": "45",
                          "page_level": 1,
                          "page_request_id": "353CAA11F35740FD8BCBEE9A249D36A"
                        },
                        "block": {
                          "card": {
                            "pagemeshow_trans": {
                              "refer_page_id": "132",
                              "refer_page_level": 1
                            },
                            "extra": {
                              "is_sdk_auto_log": true,
                              "module_increment_id": 246
                            },
                            "attached_info": {

```

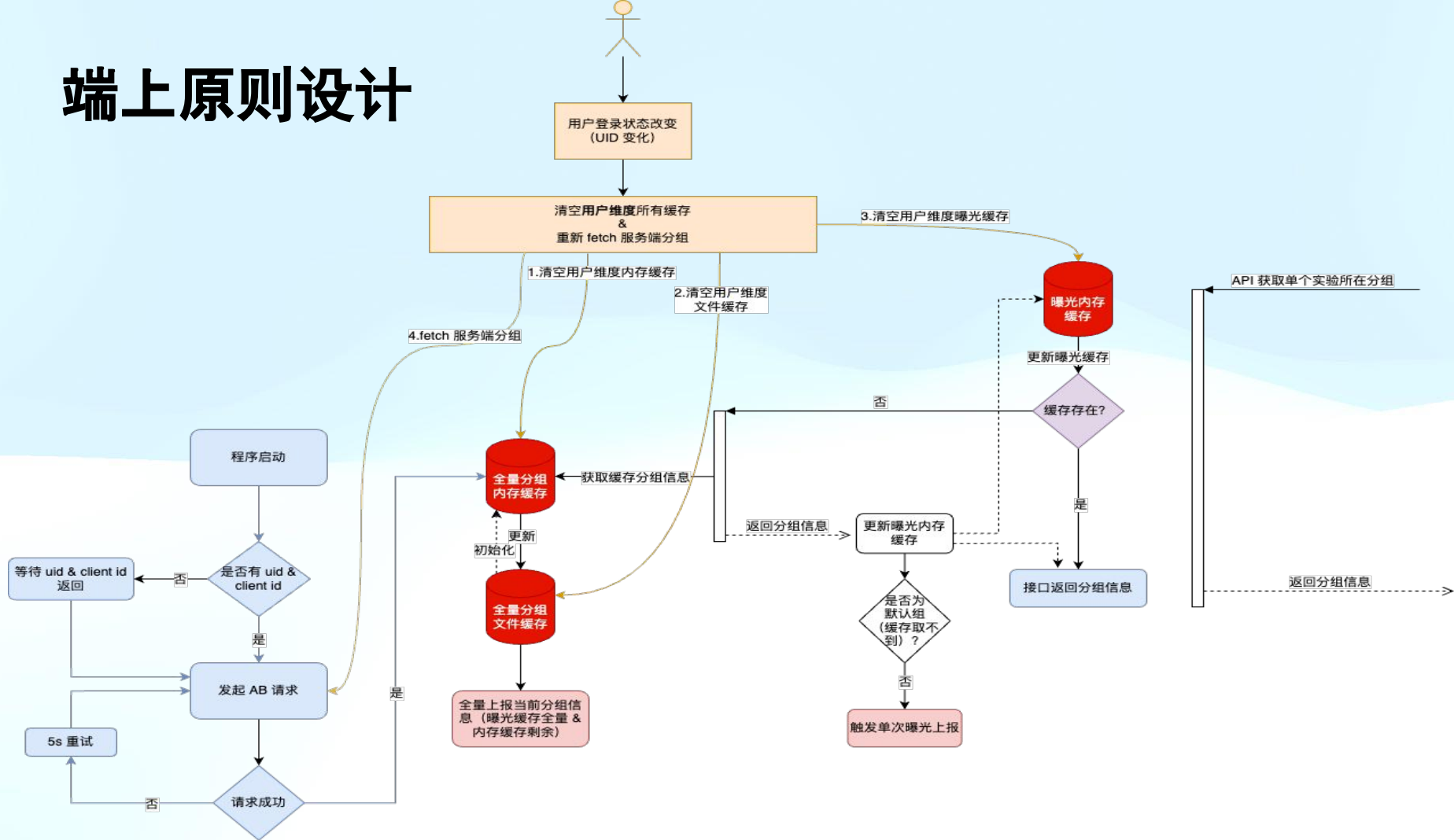
AB 实验系统设计

设计准则

实验值的稳定性保证
默认组不上报



端上原则设计



分流上报

- Etag 请求优化, 流量成本优化
- 实验轮询合并上报, 请求频次节流
- 实验结果上报前存库, 减少丢失率

实验如何实践

- 1.空转期 —— 实验上线前分组然后空跑, 查看分组是否合理
- 2.北极星指标 —— 公司认定的核心指标, 如DAU或收益

QA