



携程技术中心

IT大咖说

知识分享平台

携程技术中心

携程技术沙龙

携程用户数据采集与分析系统

分享人 王小波



姓名 王小波

- 2014年加入携程
- 框架研发部(高级工程师)
- 用户数据采集和分析以及相关产品的设计与开发

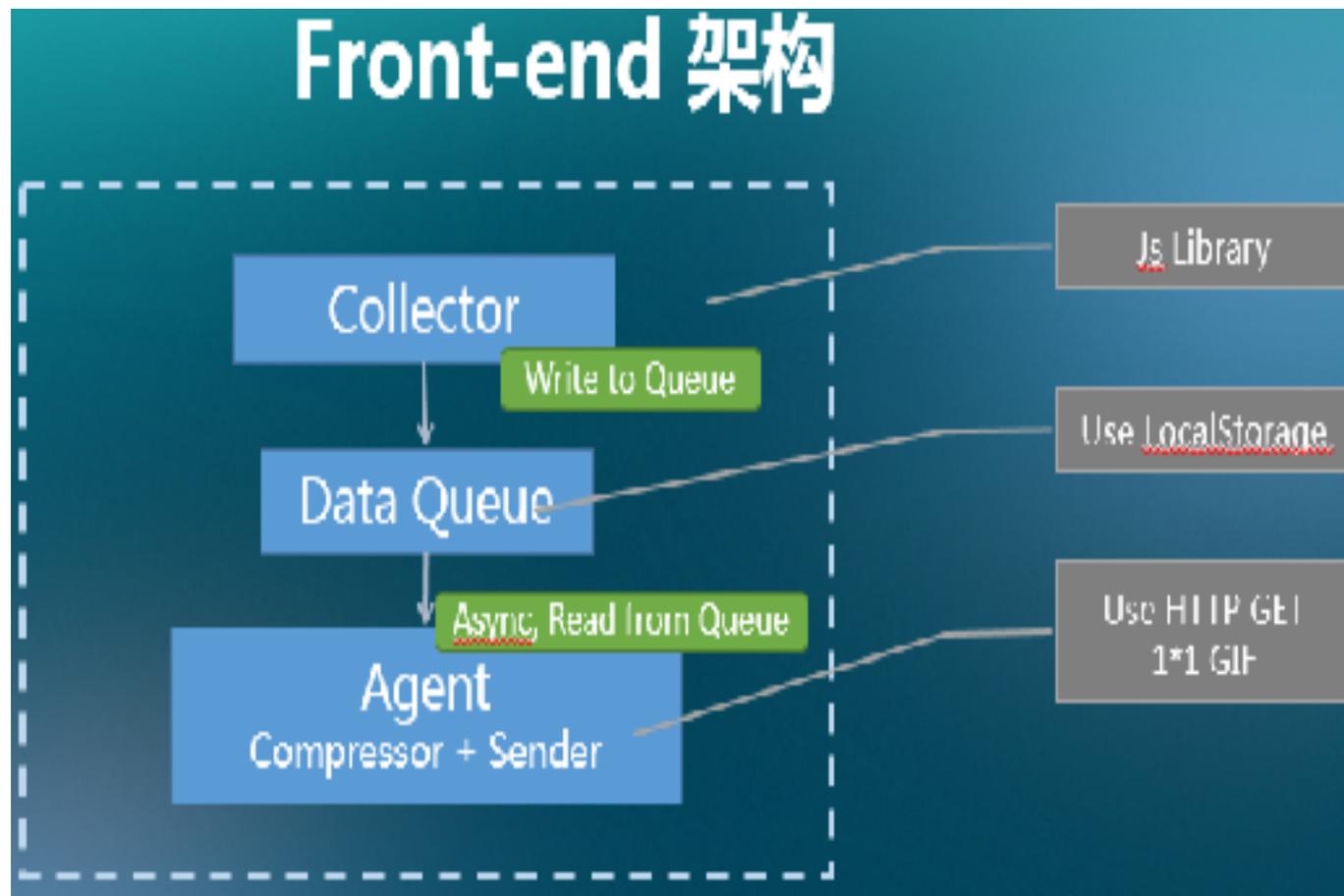
目录

CONTENTS

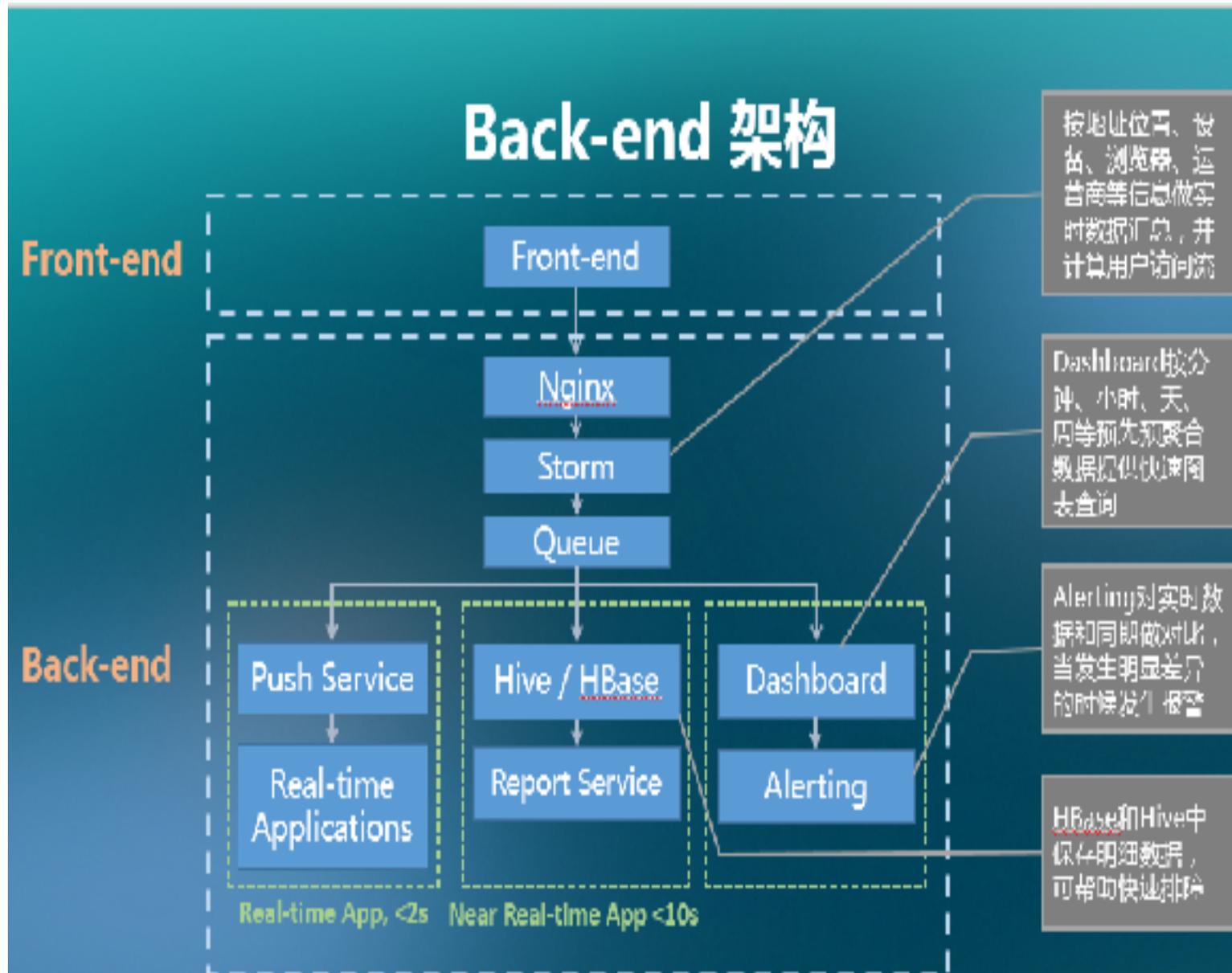
- 1 用户数据采集发展历史
- 2 前端数据采集SDK实践
- 3 后端采集服务架构实践
- 4 相关的数据分析产品
- 5 我们踩过的坑

➤ 用户数据采集发展历史

- Web/Hybrid/H5
- 只支持手动埋点



- 基于Nginx访问日志
- 基于Storm日志解析服务
- 基于Nginx的Queue模块



- **存在的问题**
 1. 采集数据较单一
 2. 采集效率较差
 3. 数据有序性较难保证
- **新一代数据采集系统**

➤ 前端数据采集SDK实践

- 多平台支持

1. iOS、Android、Web、Hybrid、RN、小程序

- 多种采集方案

1. Tcp + Http(s)

2. 自动化埋点 + 手动埋点

- 采集数据质量保证

1. 数据上下文关联

2. 序列号连续性校验

3. 数据优先级设置

4. 数据发送策略优化

- 自动化埋点
 1. 页面PV事件
 2. 用户点击事件
 3. 应用性能数据（网络、启动时间）
 4. 劫持监控数据

➤ 后端采集服务架构实践

- 系统架构设计要求

1. 高吞吐、低延时

2. 数据的安全性、有序性

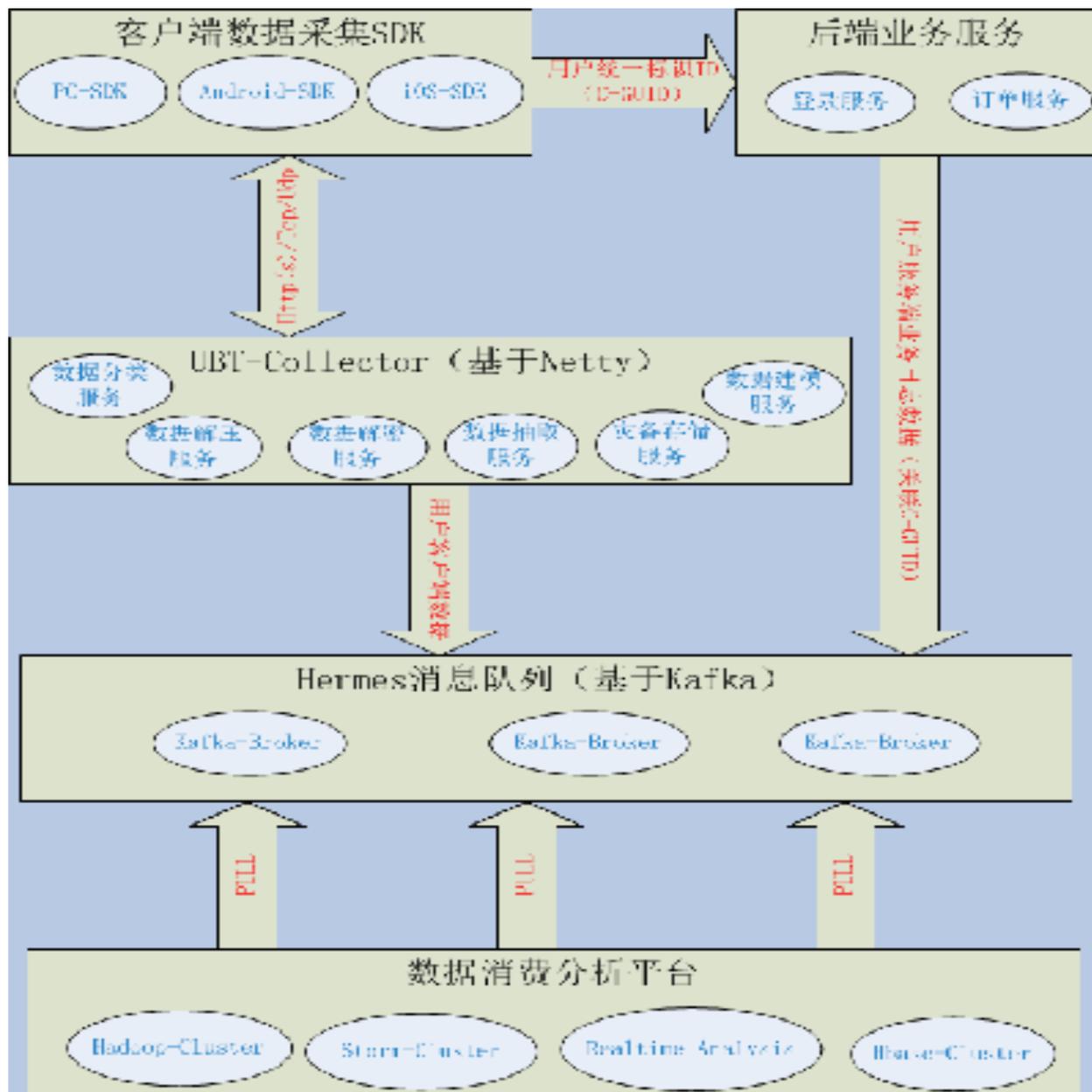
3. 数据灾备、零丢失

系统整体架构设计

1. 采集服务基于Netty

2. 存储服务基于Kafka

3. 实时分析基于JStorm



- 系统整体架构设计

4. 采集协议Tcp/Http(s)

5. 数据加密压缩传输

- 系统整体架构设计

6. 采集服务实时监控

7. 本地灾备高效存储

➤ 相关的数据分析产品

- 数据分析产品看板

报表

 KPI看板
性能报表 性能报表(国际) 性能报表(开发者)

 SiteMon APP
CDataAPP CData海外

访问

 访问量
pageID访问量 CRN访问量 Native访问量 访问者概况

 访问流和转化率
单用户浏览跟踪 转化率 用户流

行为

 点击统计
点击查看工具 点击报表 控件管理

 交互
表单填写跟踪

异常

 代码异常
JS脚本错误 CRN脚本错误

 网络异常
流量劫持 网络诊断

性能

 页面加载
多页性能 单页性能

 服务可用
Native服务成功率 Native服务耗时 H5/Hybrid Restful耗时

 地理定位
Native定位成功率 Native定位耗时 H5/Hybrid定位耗时

 流量
Native页面总流量 Hybrid页面总流量 H5/Hybrid Restful流量

 网络
CDN质量监测

工具

 分析
UVisual实时分析 UBT JS检测工具 App埋点调试

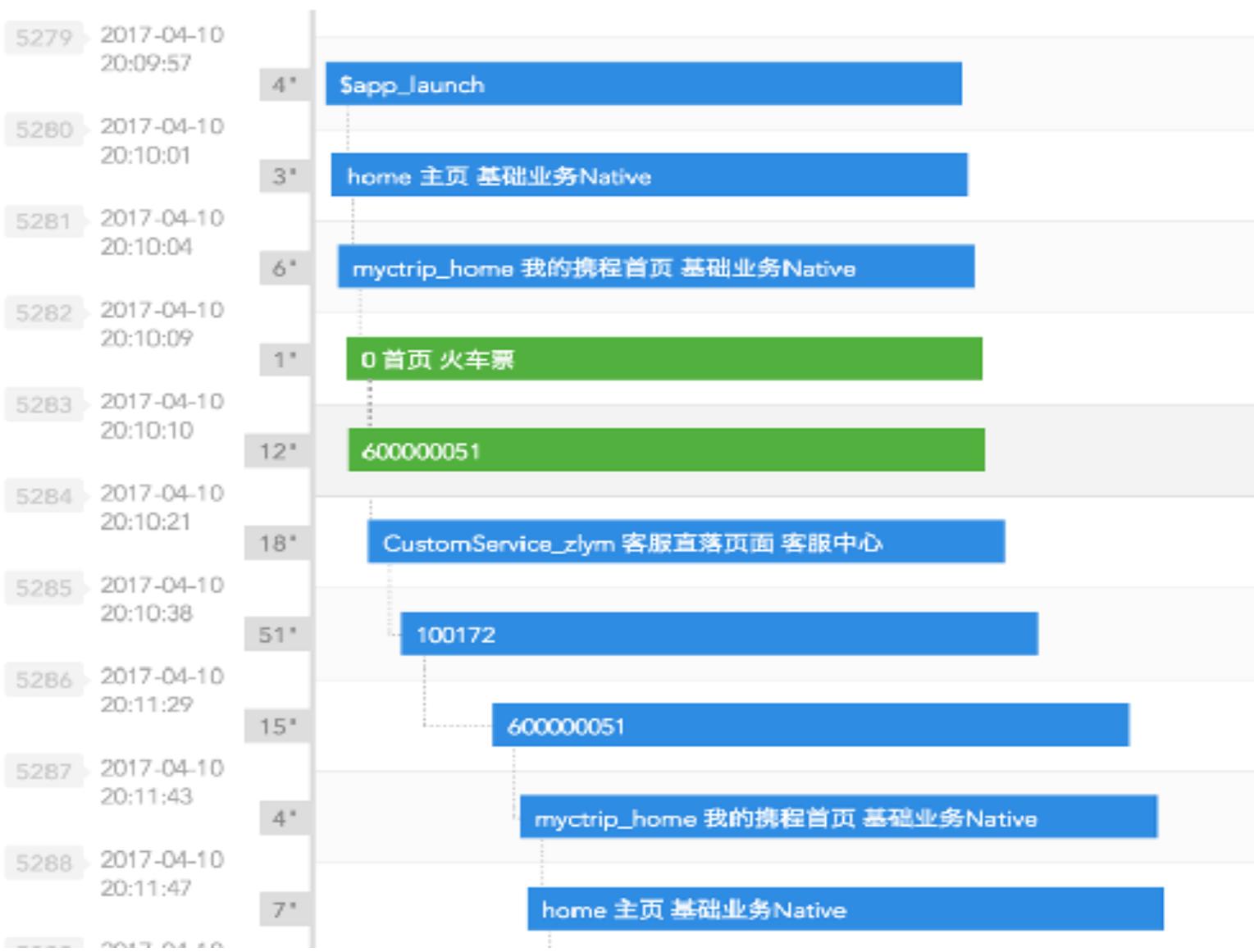
 查询
Spaceport查询 组合查询

定制

 数据流
自定义埋点注册

 配置
Metric配置 Report配置

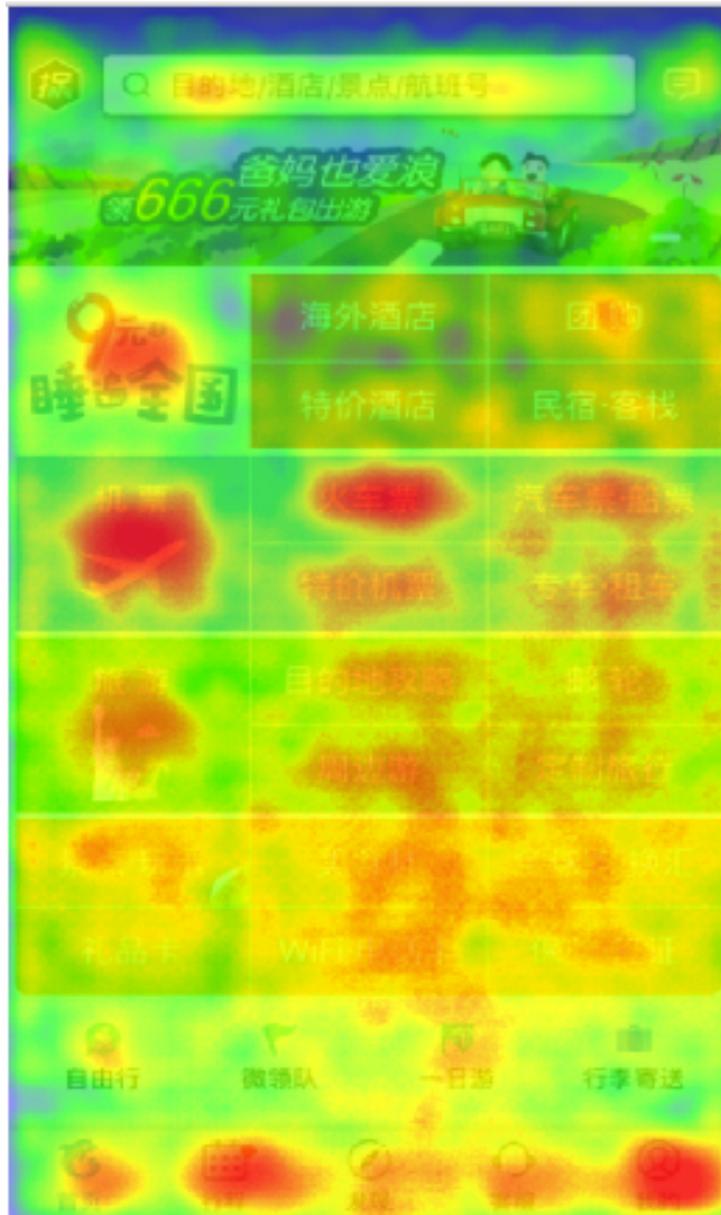
单用户浏览跟踪



• 用户访问流



- 点击热力图



采集数据调试

SCAN

开启

还原

开启(6.*)



History

 **xbwang-iphone [iPhone7,2]** 🔍 🗑️

VID : 068D6B6F144B4C88B0C202C00ED8A62C

CID : 12001036710016594758

 **HUAWEI [KIW-AL10]** 🔍 🗑️

VID : BAE52690F74611E6928F216825000452

CID : 32001165110000021935

USER IDENTIFIER

请输入用户标示(限非生产环境)

Visitor ID

Client ID

Login Name

Others

其他

Value

START 🔍



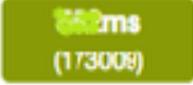
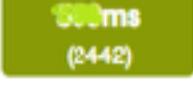
🔍 当前筛选

已配置 (支持正则表达式)

Pageview

P	home	2017-05-05 16:30:31.525	home[2018]	50072	app_version : 7.3.2 bundleid : ctrip.com page_id : home country_no : CN view_time : 2017-05-05 16:30:31 usage_tracking_enabled : 1 country_language : zh-Hans-CN idfa : 10317641-26A2-487E-8389-3A
P	flight_intent_inquire	2017-05-05 16:30:32.912	flight_intent_inquire [1919]	50101	app_version : 7.3.2 bundleid : ctrip.com page_id : flight_intent_inquire country_no : CN view_time : 2017-05-05 16:30:32 usage_tracking_enabled : 1

性能看板

36	[攻略] 6.13 搜索POI 22017501	 5924ms (5924)	
37	[机票] 国内航班列表查询V2 (自选模式) 服务 10400102	 173009ms (173009)	
38	[攻略] 目的地评论列表查询服务 22006901	 2442ms (2442)	
39	[酒店] 海外直搜酒店列表 15101001	 62453ms (62453)	
40	[攻略] 6.10开始, 搜索页里目的地的列表信息 22016401	 110035ms (110035)	
41	[攻略] 用户行程列表查询 30300101	 11059ms (11059)	
42	[攻略] 6.16 更新旅行计划 22018801	 4437ms (4437)	

➤ 我们踩过的坑

- **Android版数据采集SDK代码混淆导致序列化/反序列化错误**
- **Android-SDK在多进程环境下数据共享**
- **Android6.0以上版本WifiManager API获取Mac地址**