

YOUKU 优酷

# 优酷大数据质量平台及线上质量闭环





基于实践经验总结和提炼的品牌专栏  
尽在【极客时间】



重拾极客时间，提升技术认知

通往**年薪百万**的CTO的路上，  
如何打造自己的技术**领导力**？

扫描二维码了解详情





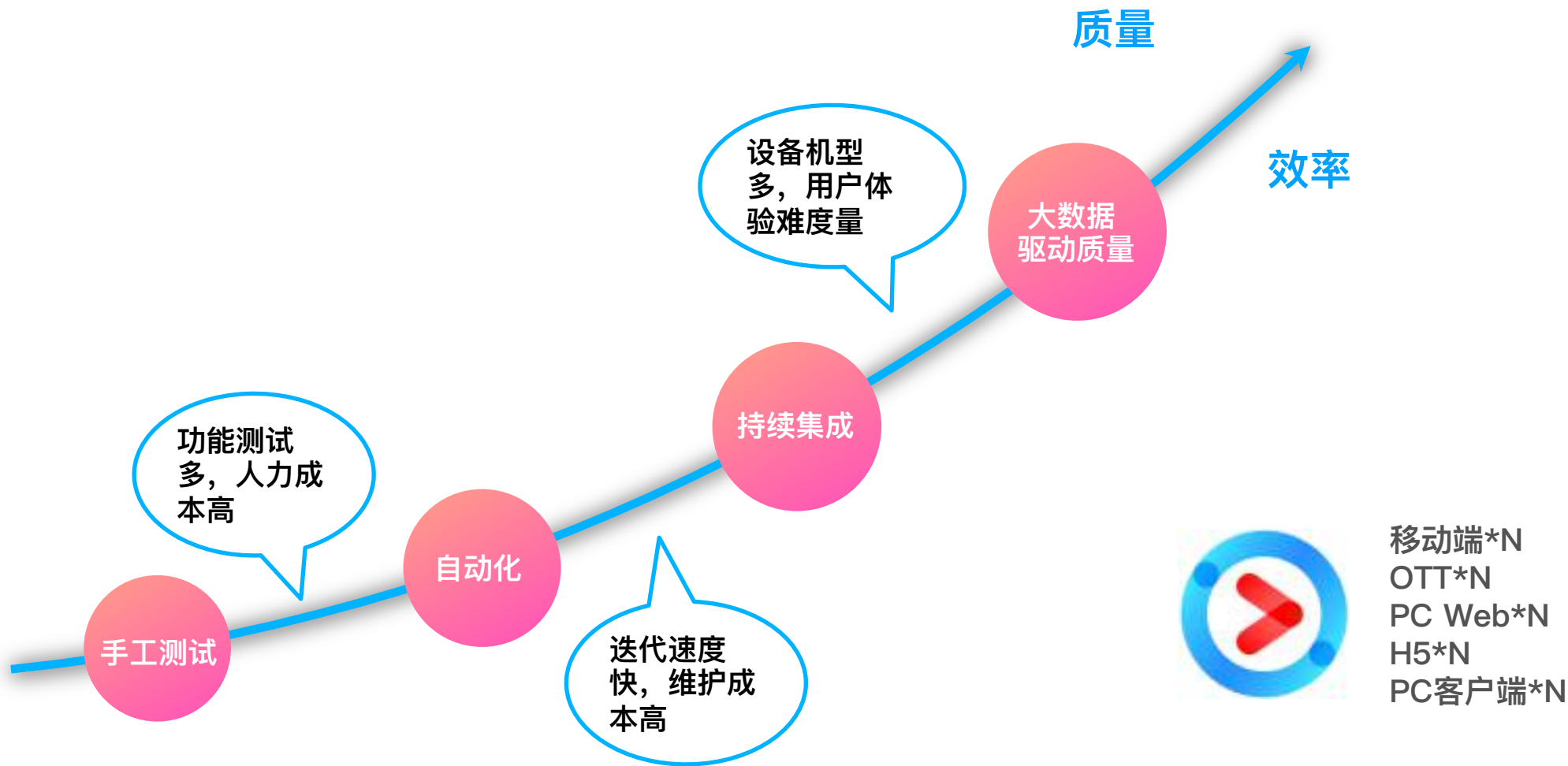
# 自我介绍

## 万传奇（花名万众）：阿里优酷质量部一技术专家





# 背景及挑战





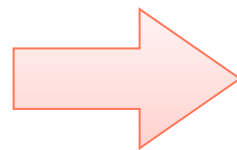
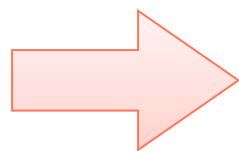
# 抛砖引玉—优酷播放体验简说

埋点字段	
通用部分	App版本、机型、OS、ISP、网络制式、省份、城市、用户信息、IP、CDN.....
维度部分	清晰度、播放方式、解码方式、媒体类型、视频协议、视频vid、广告类型、会员类型、免流类型.....
指标部分	广告播放（加载）时长、视频播放（加载）时长、帧率、码率、错误码、画面卡顿次数（时长）.....





# 大数据下的质量体系思路



- 确定数据与质量对应关系，并且能够通过线下数据反映功能、性能、稳定性卡口

- 小规模真实用户数据，通过与线下数据和线上基线对比，及时修复

- 海量用户数据，能够进行实时、多维度监控分析，并跟踪异常进行及时修复



# 大数据下的质量体系思路—线下测试



- 提升效率：线上线下一套处理数据处理方案，通过渠道标记位来剥离测试数据



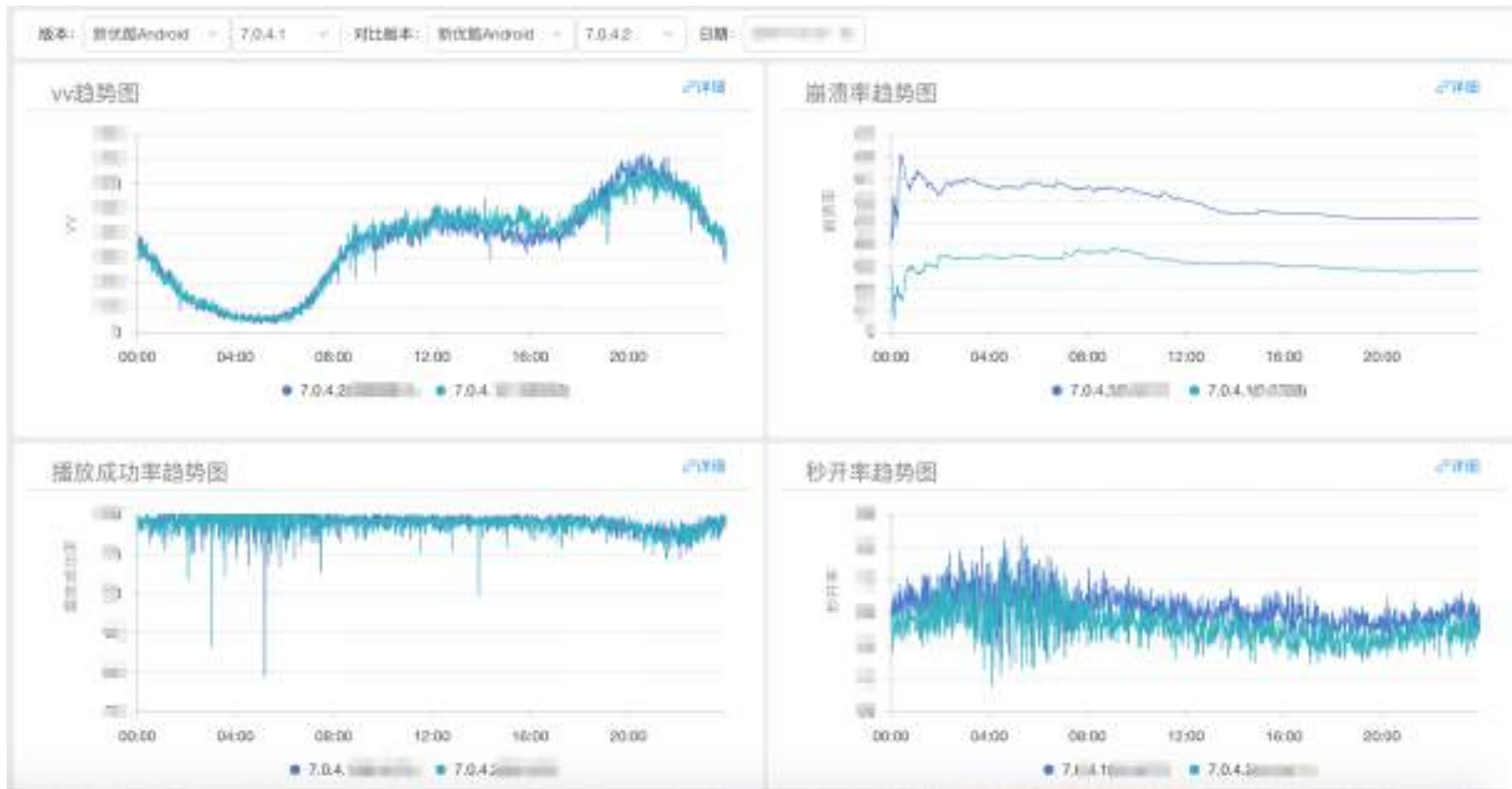
# 大数据下的质量体系思路—灰度验证

## 大数据下的灰度用户特征

机型	网络制式	APP版本
操作系统	OS版本	运营商
省份	城市	Crash设备
异常设备	活跃用户	.....

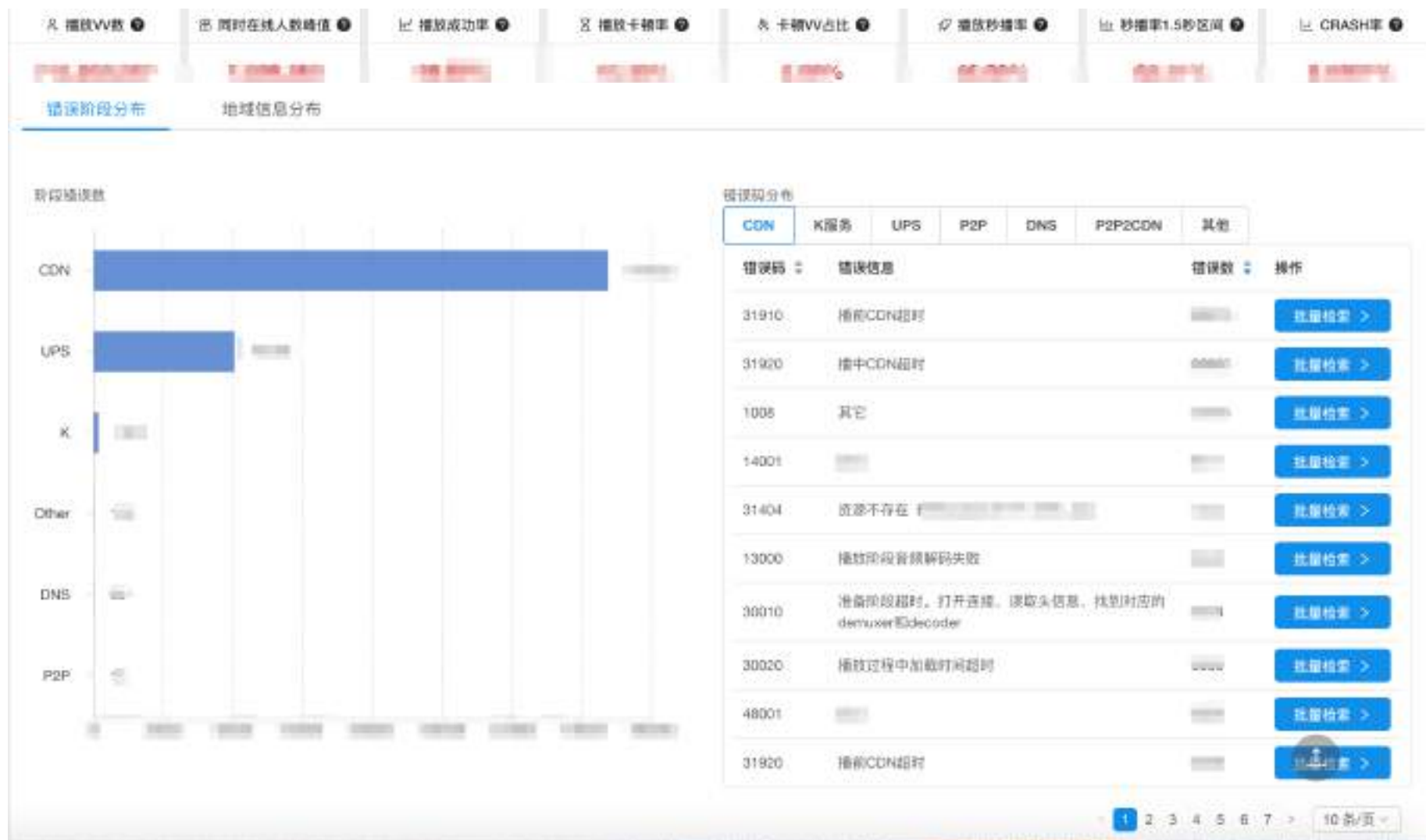
## 大数据如何提升效率

- 实时性：确保灰度数据实时展现
- 多维度分析：灰度版本与基线多维对比
- 发布Pipeline：卡口设定，无人值守



- 提升效率：保证灰度数据实时性

# 大数据下的质量体系思路—线上大数据



## 实时度量

例:

- 1、指标大盘
- 2、实时异常
- 3、监控告警
- 4、双十一晚会





# 大数据下的质量体系思路—线上大数据



## 离线分析

例：

- 1、重点性能分析：  
广告加载时长、正片加载时长、清晰度分析等
- 2、专项性能分析：  
多屏分析、编解码分析等

# 大数据下的质量体系思路—线上大数据



## 决策支持

例：同时在线人数  
决策网络带宽部署



# 大数据质量平台从0到1

## 优酷业务场景

- 数据量百亿
- 多维度多指标
- 准实时
- 秒级查询
- 数据聚合
- 全文检索

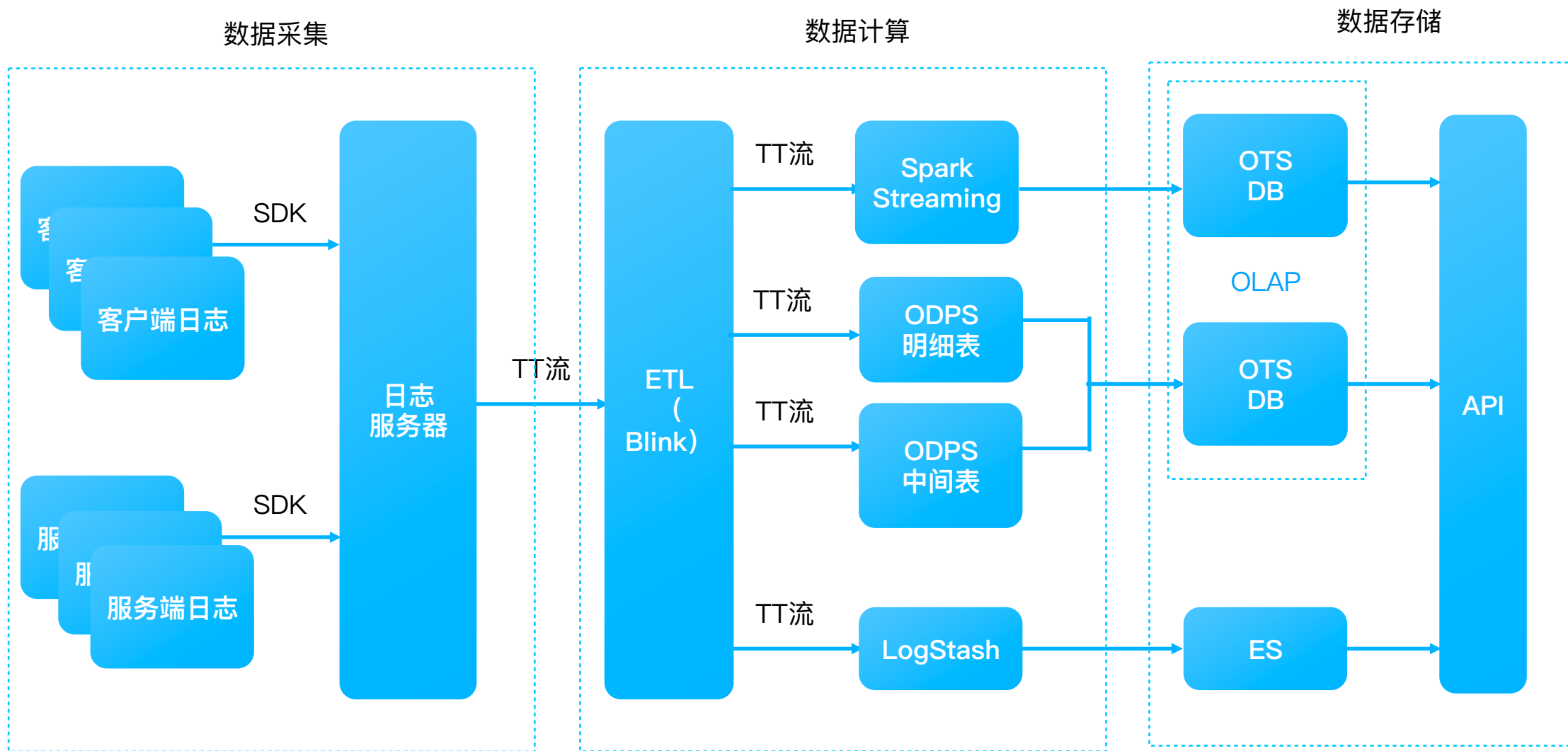


## 大数据技术

- 实时计算: Flink、Blink、Spark Streaming、Druid、Strom...
- 离线计算: Hive、Pig、ODPS、Kylin、MR...
- 数据存储: Hase、OTS、HDFS、ES、DB...
- 订阅采集: Kafka、TT、Flume...



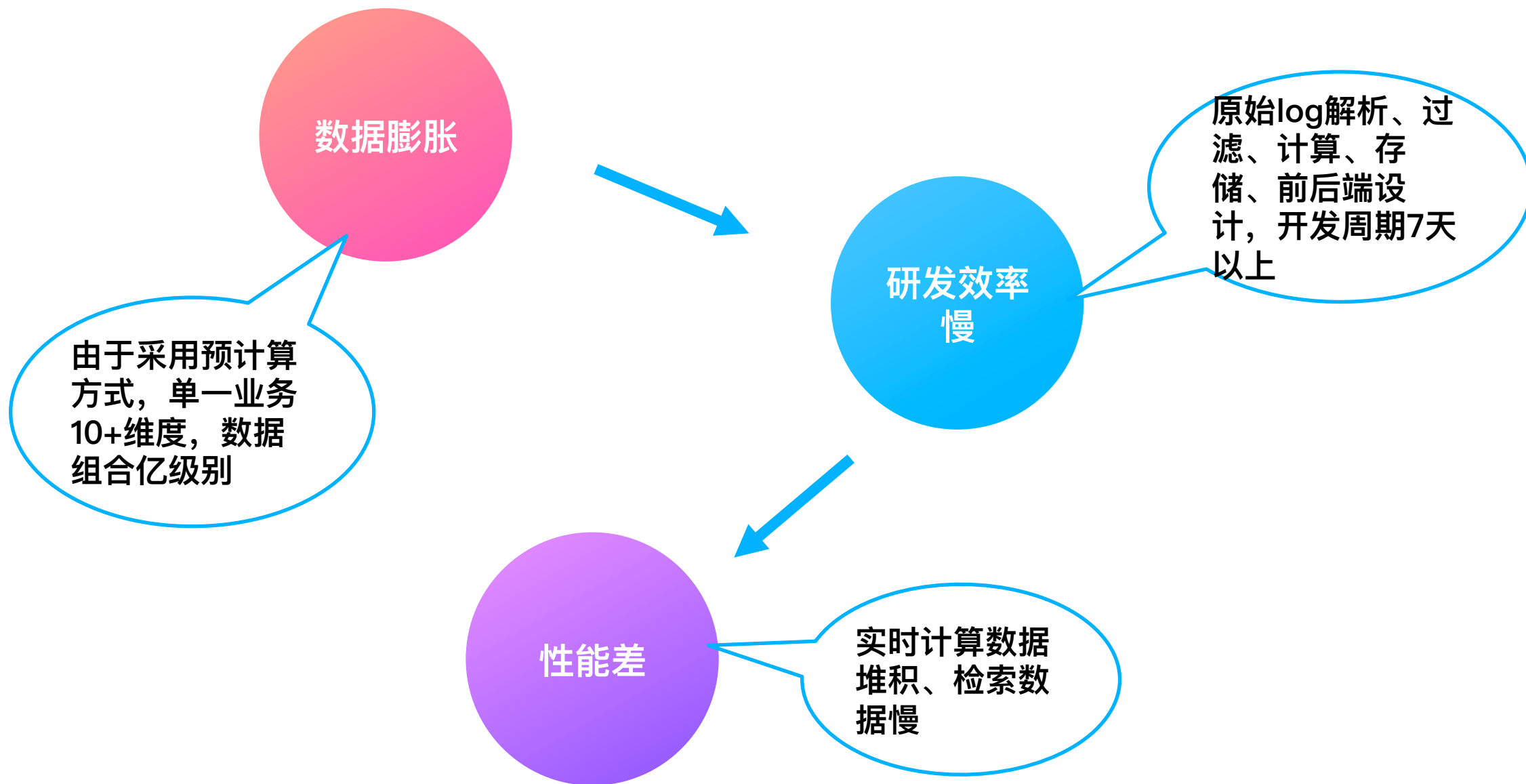
# 大数据质量平台架构



备注：TT类Kafka+Flume，Blink类Flink，OTS类Hbase，ODPS类Hive



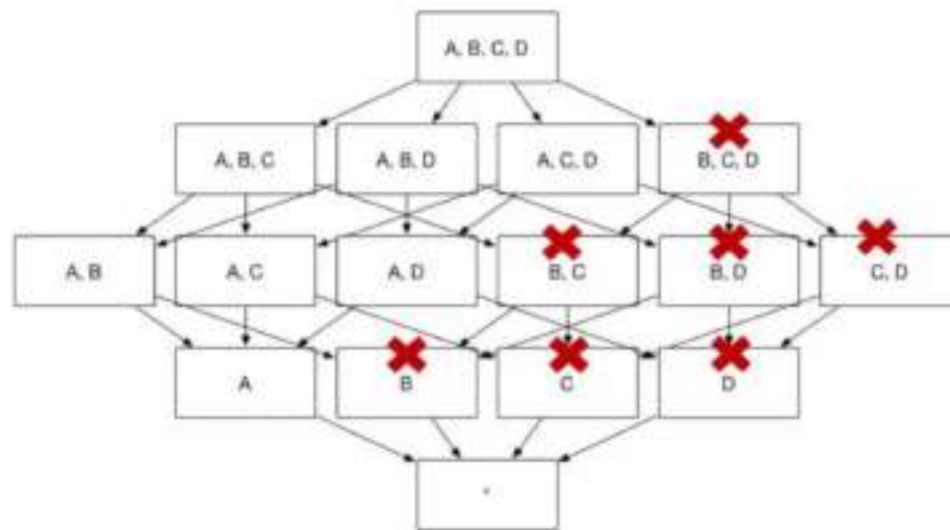
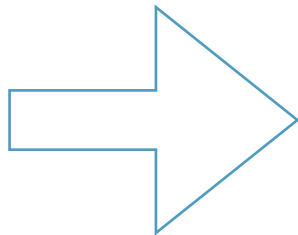
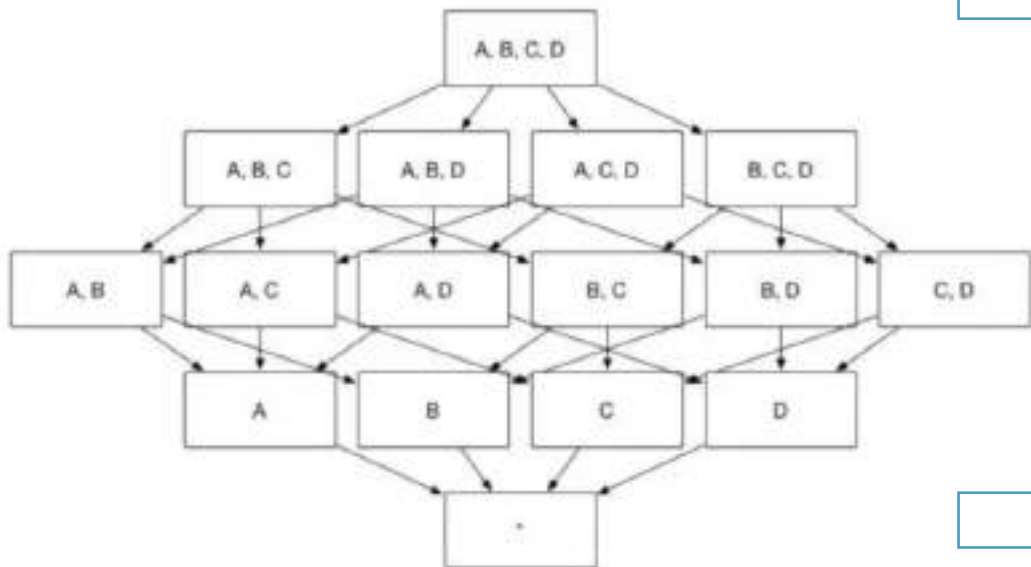
# 大数据质量平台一面临的问题



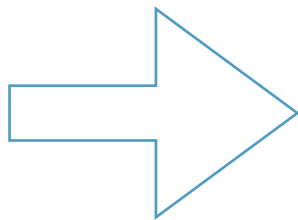
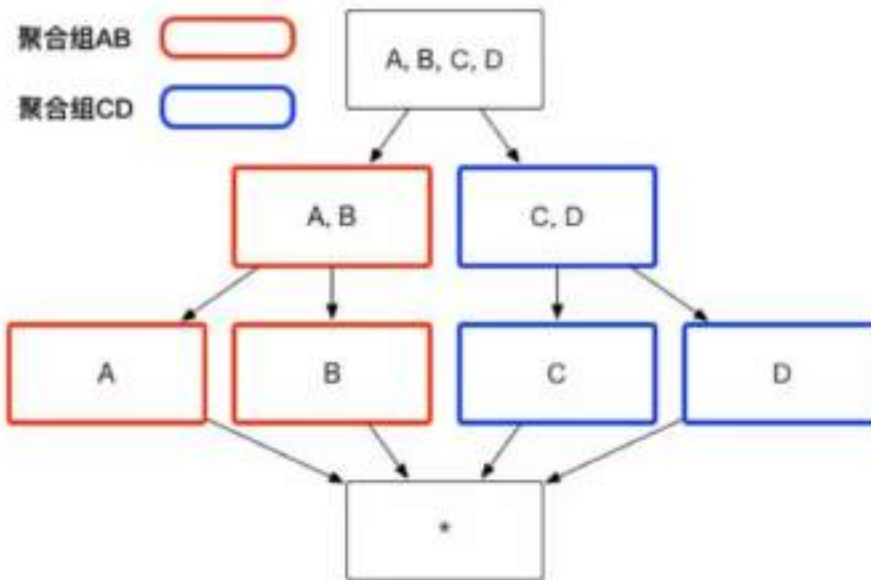


# 大数据质量平台——如何节省资源

完全立方体



冰山立方体

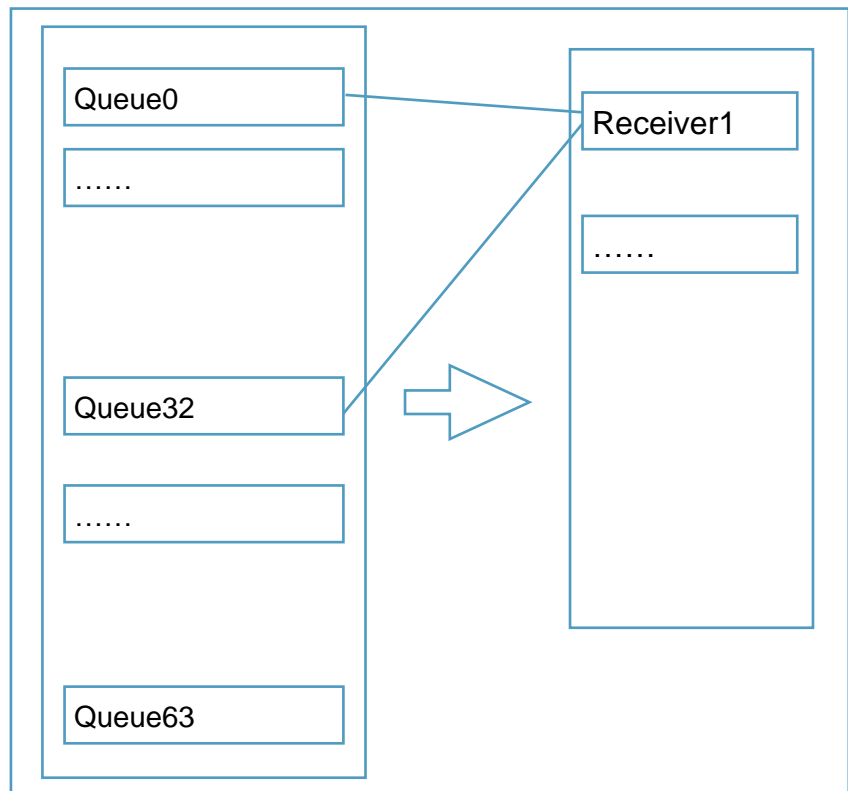




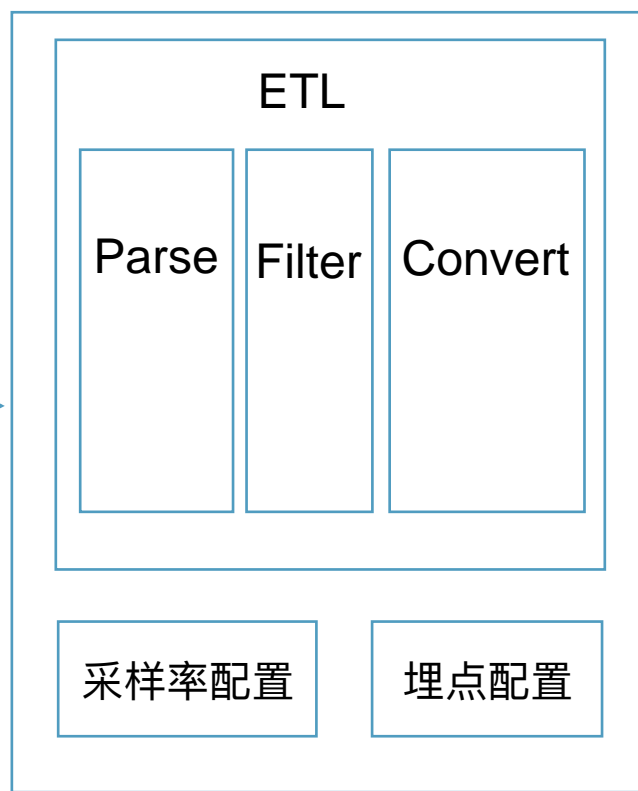


# 大数据质量平台——如何提高开发效率

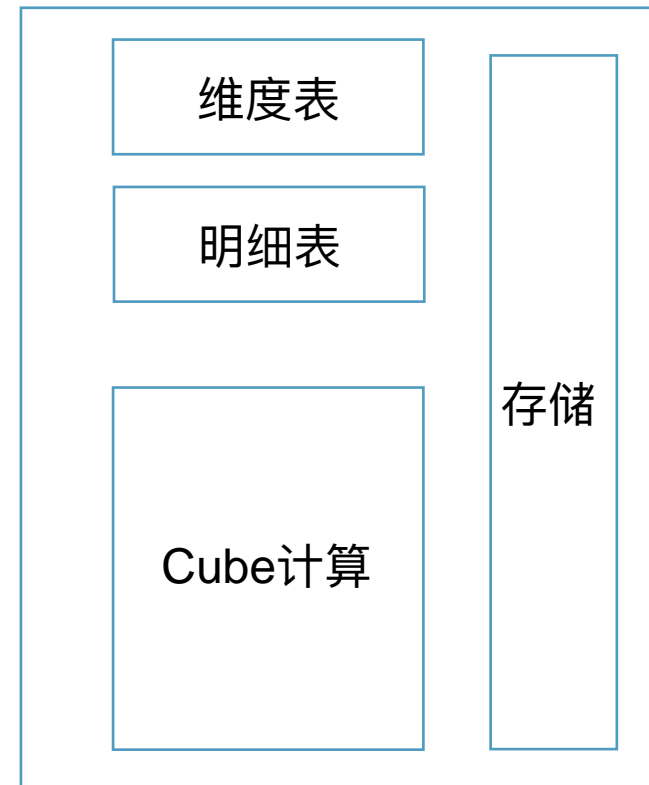
资源通用配置



业务解析通用配置



计算通用配置





# 大数据质量平台——如何提高性能

PK0: (PK1+PK2).hashCode().toHexString()

PK1: \${serviceId}+\${date}

PK2: \${dim1}+\${dim2}+.....

PK3: \${time(HH:mm)}

**大数据工程师 or 调参工程师?**  
网络? 计算? 存储? IO?

- hash避免写热点
- 通用性, 通过serviceid区分业务
- 支持分钟级别的范围查询



# 大数据质量平台的纵向发力

思考：

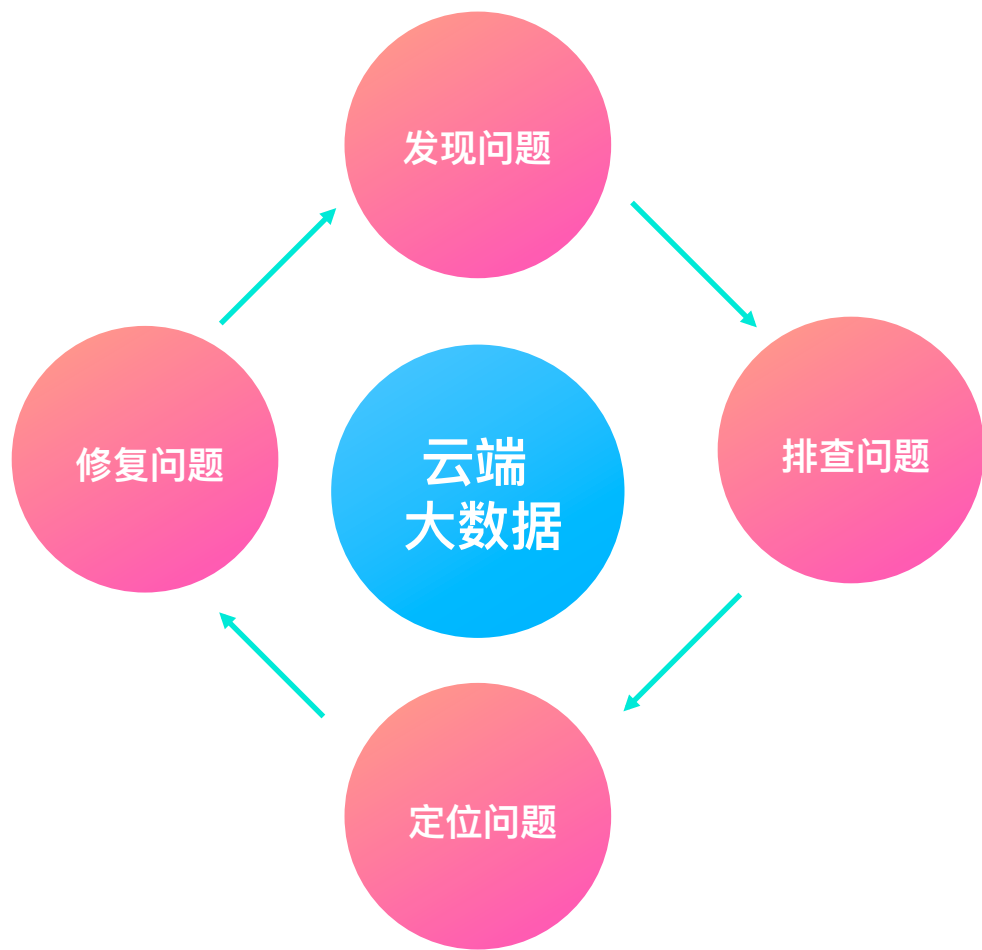
大数据技术能够在哪些业务单点发挥作用？

如何将这些单点连成质量闭环？

如何高效解决线上问题？



# 线上质量闭环概述



发现问题：实时性、准确性

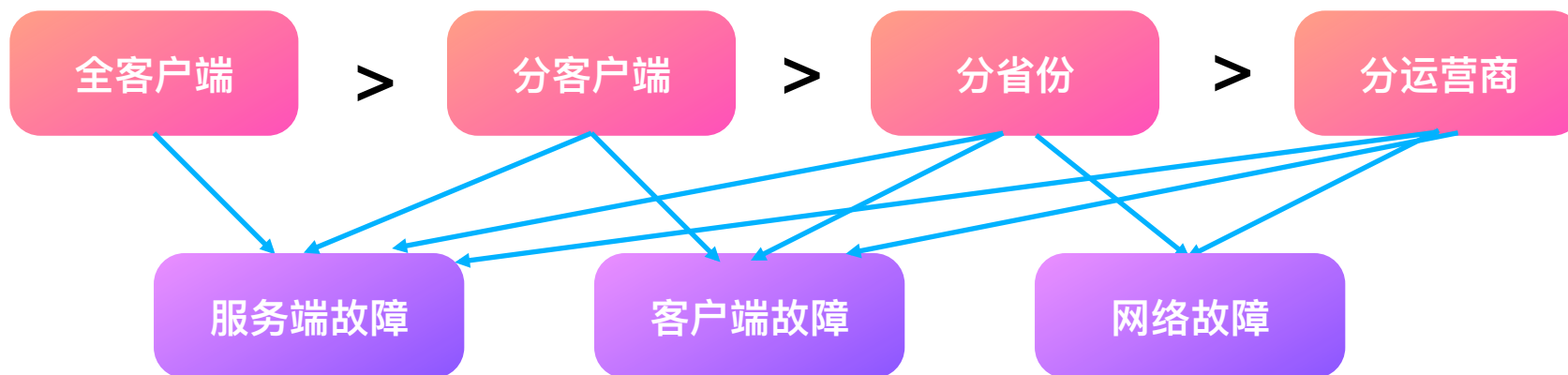
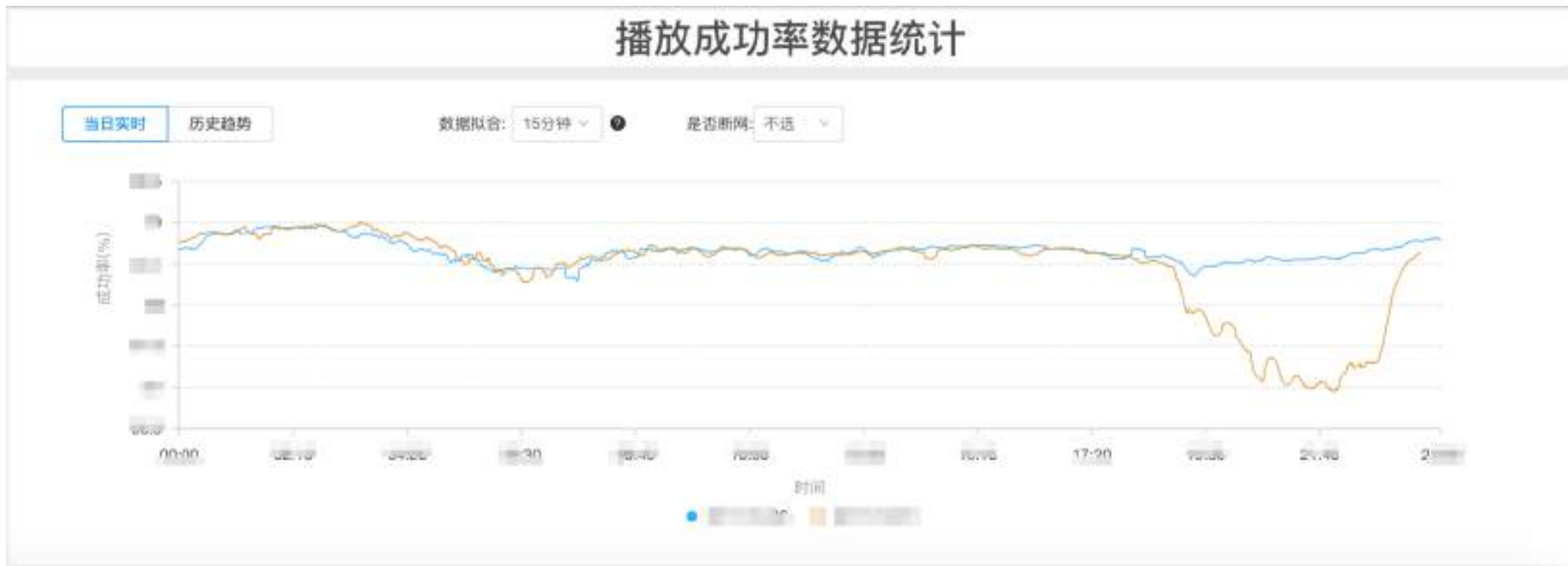
排查问题：错误分类细化、多维度

定位问题：全链路日志聚合、解析

修复问题：问题直达、结果反馈



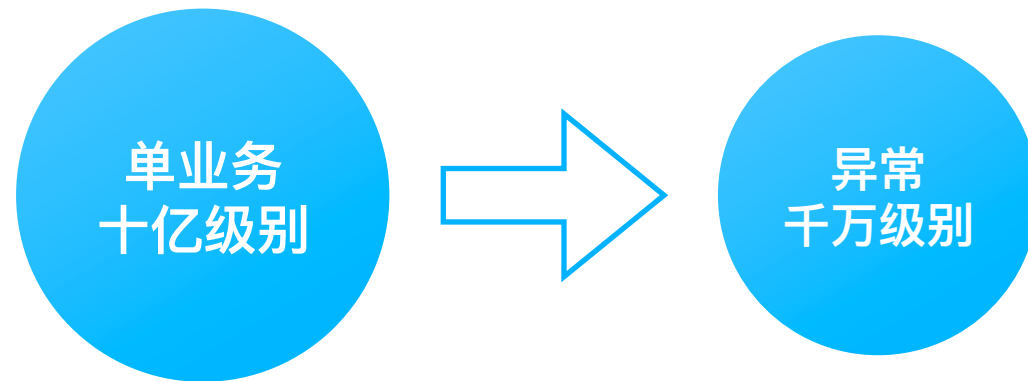
# 业务场景—发现问题





# 业务场景—排查问题

聚合维度	范围锁定
CDN	网络负责人
用户设备ID	误报
用户IP	地方运营商
错误码	功能开发负责人
APP版本	发布运维
机型	适配兼容问题
操作系统	适配兼容问题



ElasticSearch: Aggregations

- 多次单索引聚合
- 保存1-3天
- 只存入异常数据
- 只有维度概念

多维度聚合：快速找到错误特征

# 排查问题事例：一个典型误报

快捷查询: [appId](#) [appVersion](#) [city](#) [province](#) [cdnIPGeoInfo.isp](#) [clientIpGeoInfo.isp](#) [videoErrorCode](#) [utdid](#) [impairmentOrder](#) [移动三端](#) [OTT四端](#)

排查种类: [播放错误](#) 查询条件:

&province=福建省&clientIpGeoInfo.isp=移动  
&appId=23570660@android,23569910@iphoneos,23570470@ipad

日期: 2018-03-31 11:58 - 2018-03-31 12:03

查询

按查询条件生成URL

聚合维度:  UTDID  版本  省份  城市  客户端  clientIpGeoInfo.isp  cdnIPGeoInfo.isp  cdnIP  错误码

版本错误表

当前查询到错误数: 767

版本	错误数
[23570470@ipad] 4.9.7	546
[23570660@android] 7.2.1	88
[23569910@iphoneos] 7.2.1	14
[23570660@android] 7.2.0	23
[23569910@iphoneos] 6.12.0	11
[23570660@android] 6.12.0	1
[23570470@ipad] 5.0.0	10
[23570660@android] 7.1.5	8
[23569910@iphoneos] 6.6.2	5
[23570660@android] 6.6.2	1

utdid错误表

当前查询到错误数: 765

utdid	错误数
V2UFV8ywFD8DAGrwdfM+2jJz	546
WGkKRFDCeADANikNRG6wrIZ	14
V074Y6ShYPkDAH8bZkTx0x4j	5
W0EH41uET08DAPFdM+flgD7O	5
VKQd1W6SUNKDAEJuT+wMC/aW	4
VoVRj5FV/sDAMhCHZdvyAbr	3
VzNTz2G1vGkDAOZKNjYjpxsE	3
WHzUk4+cO18DAF3mPaAA1ocA	3
WSIKqaPv82MDAjlwsAoGT86O	3
WZSqq0Jz38EDAMe9ouiJ41lr	3



# 业务场景—批量查询

查询类型: 播放错误 终端: 新优酷iPhone 时间: [日期选择器]

错误阶段: 不选 错误码: 不选 解码方式: 不选 设备型号: 设备型号可模糊查询例如: MI 操作系统版本: 操作系统版本可模糊查询例如: 6.0

用户昵称: 用户昵称可模糊查询例如: 优酷用户150 APP版本号: 不选 国家或地区: 请选择国家或地区例如: 中国 省份: 不选 城市: 请选择城市例如: 北京市

设备运营商: 不选 网络: 不选 清晰度: 不选 播放类型: 不选 视频id: 请填写视频id例如: XMjccOTGzNDCCMA==

CDN运营商: 不选 cdnIP: 请填写cdnIP例如: 58.49.225.141 UTDID: 请填写utdid 设备是否在线:

[查询](#) [重置](#)

设备ID	APP版本	设备型号	操作系统	分辨率	地域信息	网络	用户昵称	视频信息	错误码和错误信息	CDN信息	监控点	客户端时间	服务端时间	操作
[模糊]	7.2.	Apple iPhone	iOS 11.2.6	2208	中国上海市上海市	WiFi	优酷用户31106	点播 超清XMz UwNzE2MTY1Mg==	错误码: 30020 错误信息: 播放过程中加载时间超时 解码方式: 2	内网IP 255.255.255.5.255	plasyin	2018-04-02 09:29:03	2018-04-02 09:32:50	<a href="#">查看日志 &gt;</a>
[模糊]	6.1	Apple iPhone	iOS 11.2.6	2208	中国贵州省贵阳市	WiFi	[模糊]15	点播 省流XMz Q1MTExMjMg==	错误码: 20101 错误信息: 请求接口超时未返回 解码方式: 2	-	before	2018-04-02 09:31:43	2018-04-02 09:32:59	<a href="#">查看日志 &gt;</a>
[模糊]	6.7.	Apple iPhone	iOS	2208	中国江苏省苏州市	WiFi	[模糊]64	点播 超清	错误码: 30020 错误信息: 播放过程中加载时间	-	plasyin	2018-04-02 09:31:43	2018-04-02 09:32:59	<a href="#">查看日志 &gt;</a>

全文检索: 错误数据全量存入ES, 支持20+维度数据检索

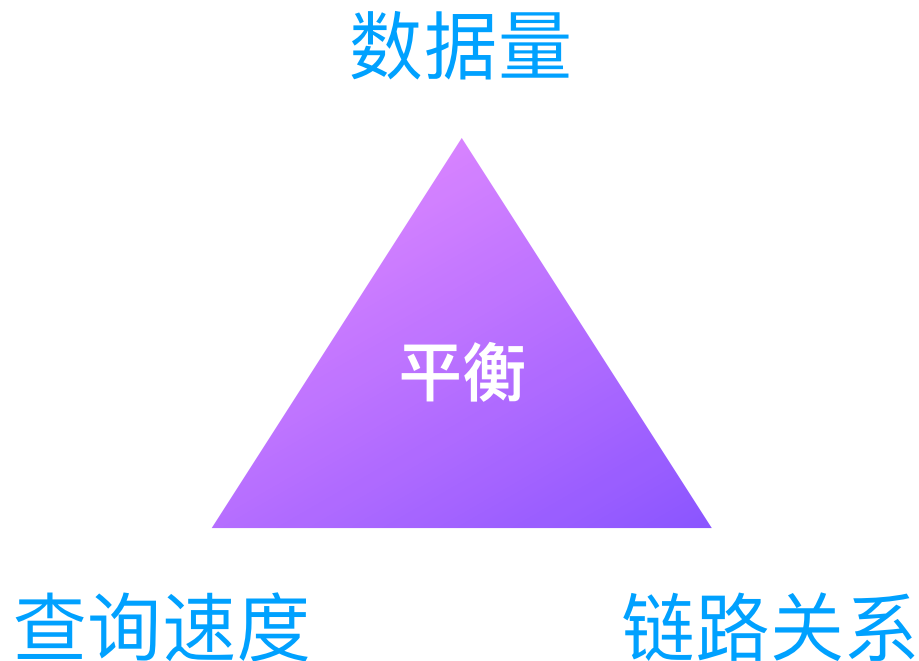




# 业务场景—单点定位

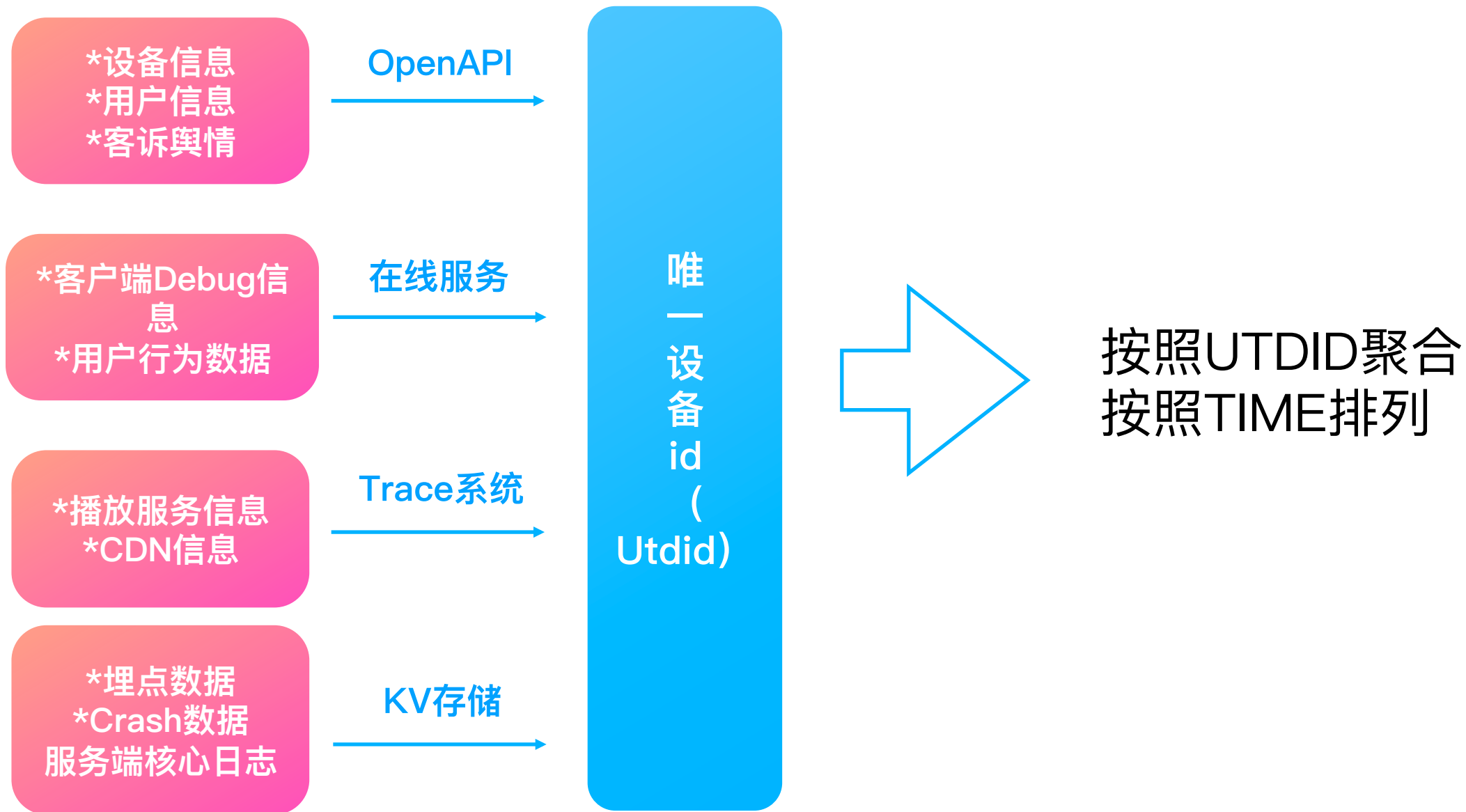
定位一个播放问题需要  
哪些线索？

设备信息  
用户信息  
行为日志  
埋点日志  
客户端Debug日志  
播放服务日志  
CDN日志  
Crash日志  
用户描述信息  
.....





## 业务场景—单点定位



# 业务场景—单点定位

如何利用这些线索

日志关联

日志解析

问题修复

The screenshot displays a user profile and a log analysis interface. The user profile includes fields for YTIM, 优酷昵称, 手机号, 邮箱, UTID, APP ID, 机型 (iPhone10.3), 设备名称, 省份 (陕西省), ISP (联通), and 语言 (zh-Hans-CN). The log analysis section shows a list of log categories with their respective counts and details:

- 日志类型: 用户意见反馈 | 日志概况: 时间段内用户提出意见反馈数量为 2 | 展开查看详情
- 日志类型: VPM埋点 | 日志概况: 时间段内VPM(beforePlay,onePlay,playing)日志数量为 2 | 展开查看详情
- 日志类型: UPS日志 | 日志概况: 时间段内UPS日志数量为 2 | 展开查看详情
- 日志类型: 用户观看视频信息 | 日志概况: 时间段内用户总计观看视频数量为 1 | 展开查看详情
- 日志类型: CRASHE日志 | 日志概况: 时间段内crash日志数量为 3 | 展开查看详情

Highlighted log entries include:

- 消息-1: 十万个冷笑话2 查看详情
- 消息-2: 十万个冷笑话2已更新,但提示因版权无法播放 查看详情
- UPS日志错误详情共计 2 条 详情

The detailed log entry at the bottom shows technical data such as access type (WiFi), client IP, client time (2018-04-02 09:10:42), device ID, device model (iPhone10.3), and various system parameters.



# 总结

交付体系：线下、灰度、线上

大数据架构：实时、离线、检索

技术点：Spark Streaming、**Blink(Flink)**、ODPS (**Hive**)、**Kylin**、OTS (**Hbase**)、ES

线上质量闭环：发现、排查、定位、修复

# GMITC 2018

## 全球大前端技术大会

—— 大前端的下一站 ——



<<扫码了解更多详情>>

关注 ArchSummit 公众号  
获取国内外一线架构设计  
了解上千名知名架构师的实践动向



Apple • Google • Microsoft • Facebook • Amazon 腾讯 • 阿里 • 百度 • 京东 • 小米 • 网易 • 微博

深圳站：2018年7月6-9日 北京站：2018年12月7-10日

# QCon

全球软件开发大会【2018】

# 上海站

2018年10月18-20日

# 7折

预售中, 现在报名立减2040元

团购享更多优惠, 截至2018年7月1日



极客邦科技  
企业培训与咨询

Geekbang

扫码关注  
获取更多培训信息







Q&A

**Thank You!**