

# 字节跳动埋点成本治理 解决方案

---

肖茜文 火山引擎大数据工程师



# 目录 CONTENT

**01** 治理背景

**03** 治理经验回顾

**02** 治理策略

**04** 规划与展望

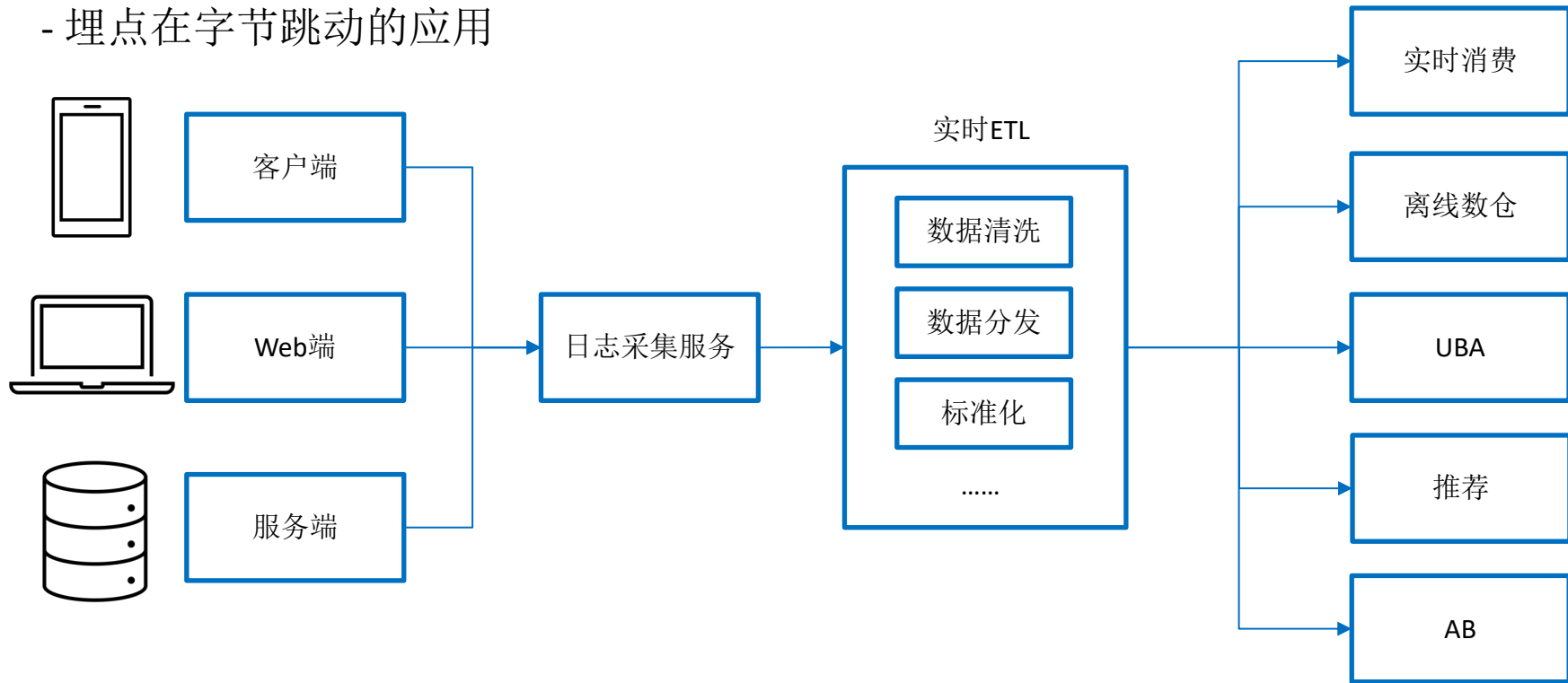
# 01

## 埋点成本治理背景



# 埋点成本治理背景

## - 埋点在字节跳动的应用



## 埋点成本治理背景

- 埋点在字节跳动的规模

1亿+/秒  
峰值流量

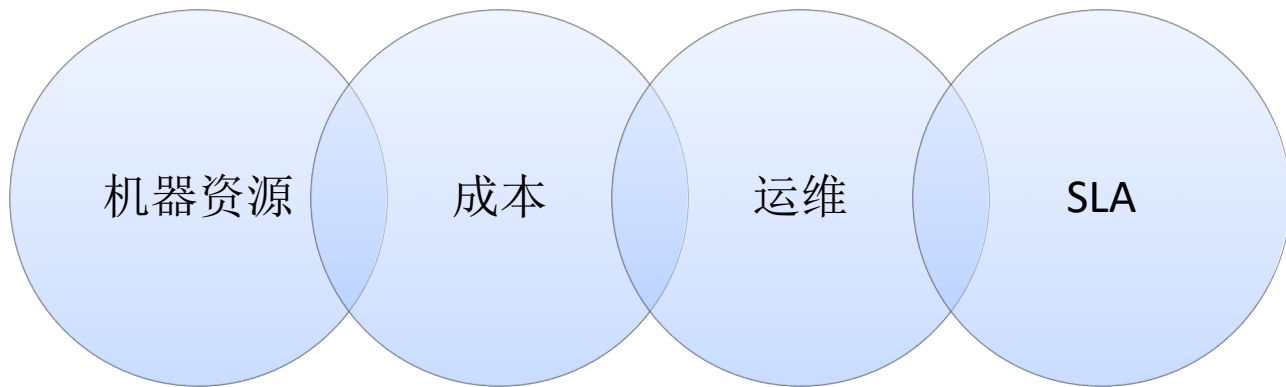
10万亿/天  
埋点量

10+PB/天  
存储增量



## 埋点成本治理背景

- 为什么要治理埋点



## 埋点成本治理背景

### - 在字节内取得的效果与收益



应用于抖音、头条等多个业务，覆盖85%业务



通过无用埋点下线，在2021年节省成本近亿元



通过埋点分级，节省100+PB HDFS存储



通过埋点采样，在2022年可节省成本 ¥3000+万/年

# 02

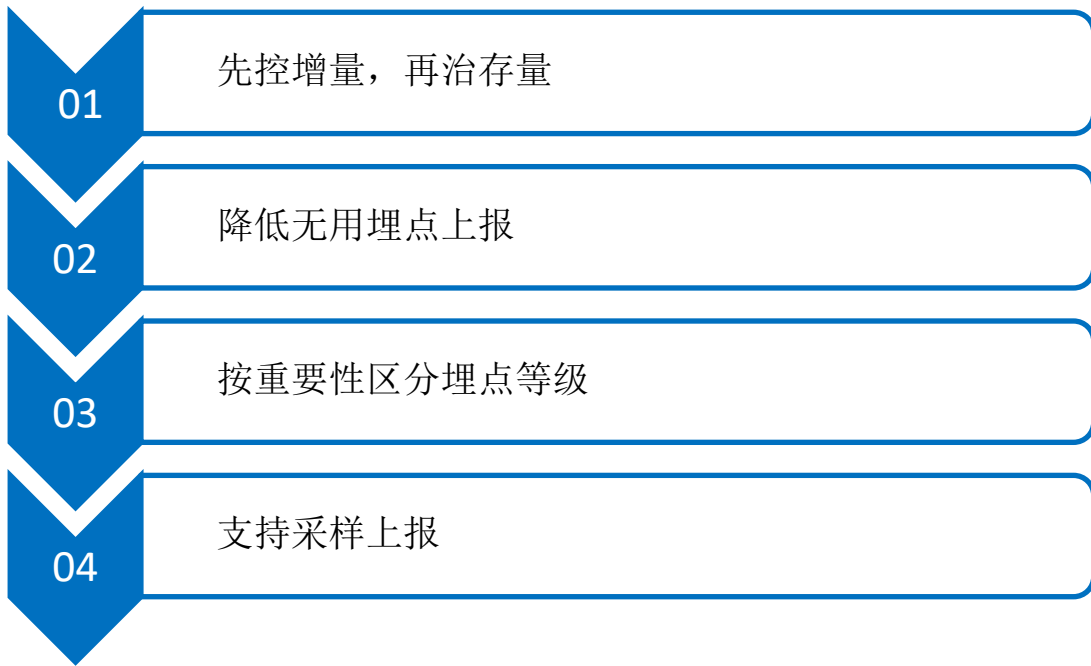
## 埋点成本治理多场景策略





# 埋点成本治理多场景策略

## - 从0到1建设治理



# 埋点成本治理多场景策略

## - 场景1:先控增量，再治存量

01

先控增量，再治存量

问题

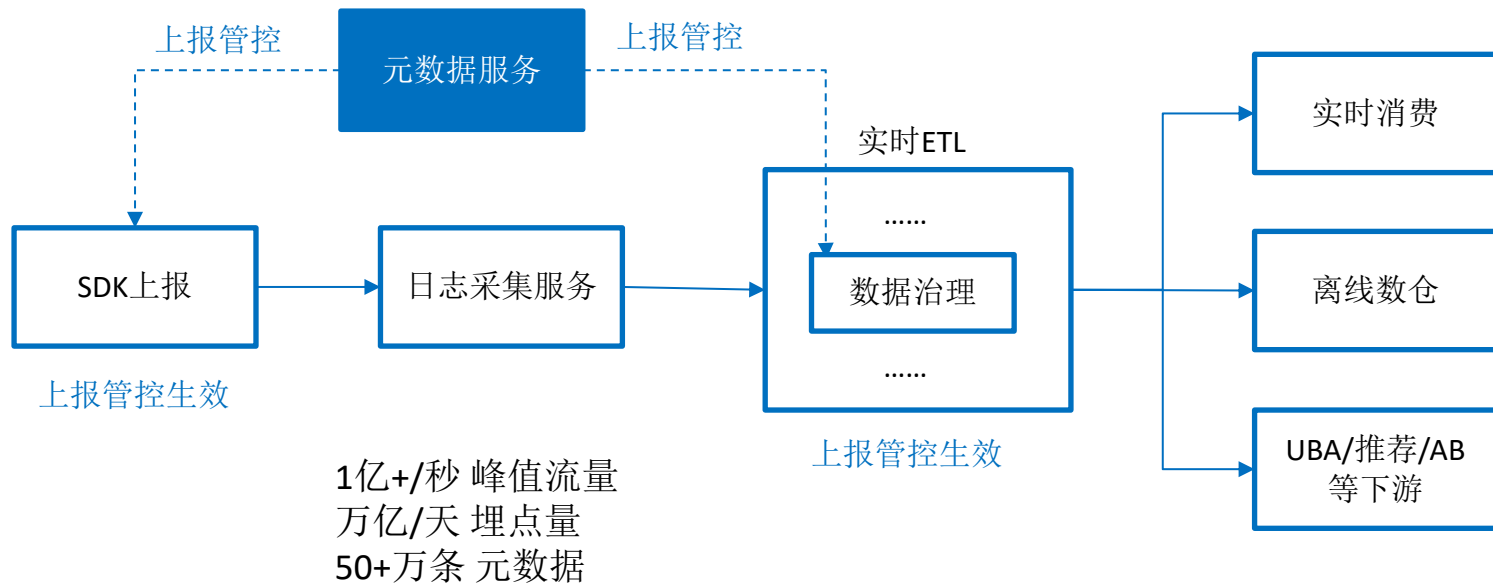
- 在业务发展过程中开始治理，既有**存量埋点**，同时也有不断增加的**新增埋点**。
- 需要先控制新增、再治理存量。

解决方案

- 埋点上报管控：在“允许上报”列表内的埋点才可上报

# 埋点成本治理多场景策略

## - 场景1:先控增量，再治存量



# 埋点成本治理多场景策略

## - 场景1:先控增量，再治存量

通过平台维护列表：在ByteIO平台录入埋点、且为“已上线”状态，允许埋点上报。

<div>测试</div> <div>埋点检索</div> <div>需求管理</div> <div>BTM管理</div> <div>点位管理</div> <div>来源去向</div> <div>属性管理</div> <div>属性列表</div>	<div>埋点列表</div> <div><input type="checkbox"/> 只看我负责的</div> <div>名称/描述/负责人</div> <div>筛选</div> <div>批量操作</div> <div>新建埋点</div>									
	<input type="checkbox"/> 埋点名称&描述等基本信息	操作系统	标签	埋点类型 ①	属性	埋点负责人	状态 ①	操作		
	<input type="checkbox"/> 001_33 001_33	Android, i...	视频...	行为埋点	purchase... 2 个		已上线	分析 禁用		
	<input type="checkbox"/> 001_333 001_333	Android, i...	-	行为埋点	-		已上线	分析 禁用		
	<input type="checkbox"/> 001_334 001_334	Android, i...	-	行为埋点	finance_ri... 1 个		已上线	分析 禁用		
	<input type="checkbox"/> 0623_event 测试	Android, i...	爱... ..	行为埋点	aaa、aa... 19 个		已上线	分析 禁用		

# 埋点成本治理多场景策略

## - 场景2:降低无用埋点上报

02

降低无用埋点上报

问题

- 埋点已**不再使用**，但仍持续上报，造成资源浪费
- 案例：上报初期使用，一段时间后不再使用

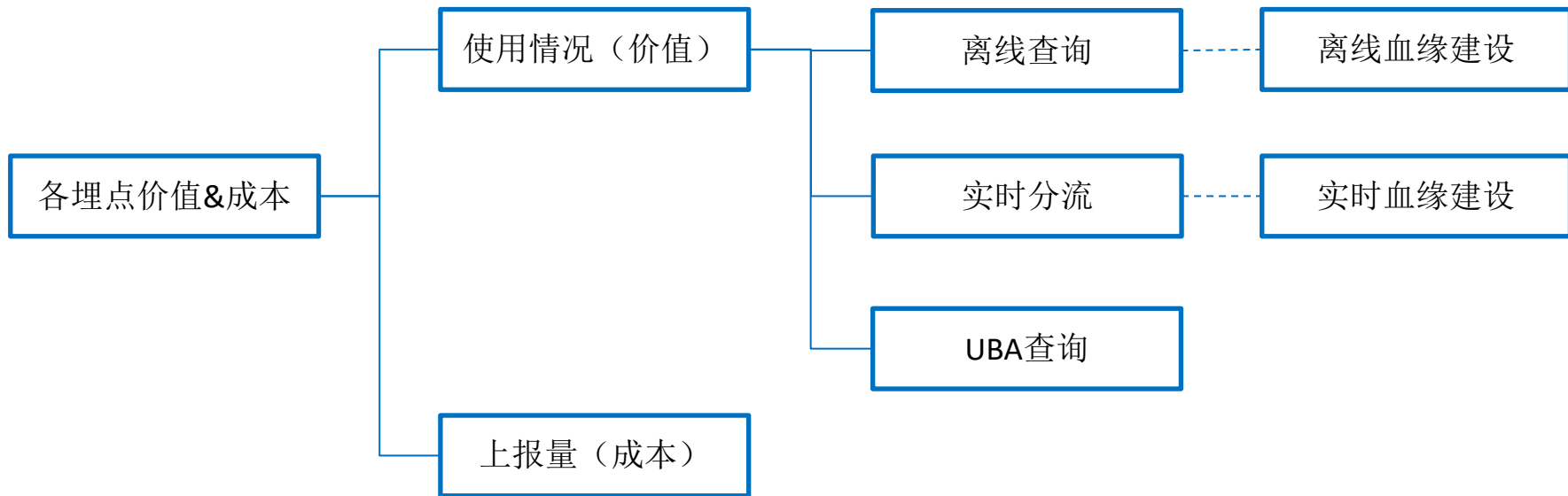
解决方案

- 无用埋点下线：筛选无用埋点，从“允许上报”列表中移除。

# 埋点成本治理多场景策略

## - 场景2:降低无用埋点上报

如何定义和筛选无用埋点：分析对比各埋点的价值&成本



## 埋点成本治理多场景策略

## - 场景2:降低无用埋点上报

1 筛选和发起

治理监督人筛选和配置相关人、关键时间节点

2 标注

埋点负责人和相关人员标注是否保留。

3 预下线确认

治理监督人确认预下线埋点名单。

4 预下线观察期

从数仓下线，如遇异常治理监督人可恢复数据。

5 最终下线确认

治理监督人确认最终下线埋点名单

6 最终下线

埋点将被禁止上报，数据不可恢复

Event

埋点名称

日志数 不低于

Event	日志数 (昨日)	流量占比	UBA使用情况	离线查询情况	实时分流情况	下线收益(每天)	是否治理
001_33	0	0.00%	0	0	否	¥0	<input type="checkbox"/>
001_333	0	0.00%	0	0	否	¥0	<input type="checkbox"/>
001_334	0	0.00%	0	0	否	¥0	<input type="checkbox"/>
0623_event	0	0.00%	0	0	否	¥0	<input type="checkbox"/>
0623hbhevent	0	0.00%	0	0	否	¥0	<input type="checkbox"/>

# 埋点成本治理多场景策略

## - 场景3:按重要性区分埋点等级

03

### 按重要性区分埋点等级

#### 问题

- 埋点需要使用，但**重要性不同**，资源无法优先保障高优数据
- 案例：RD排查问题使用的埋点，不需要TTL/SLA保障

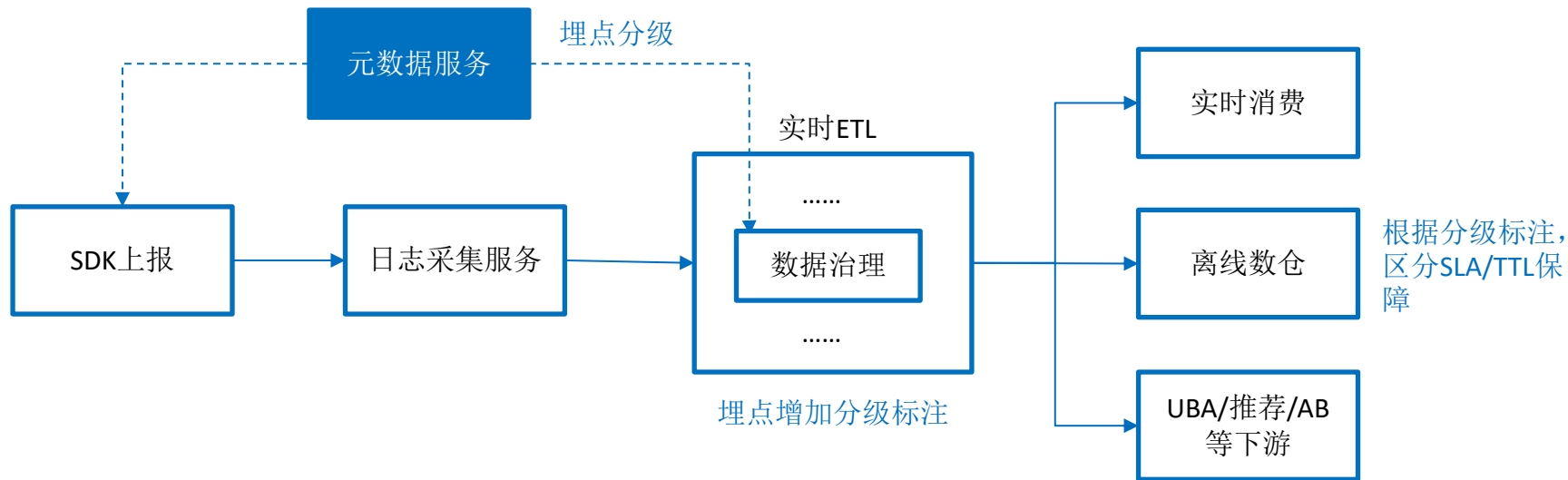
#### 解决方案

- 按重要性区分埋点等级，提供不同的运维保障



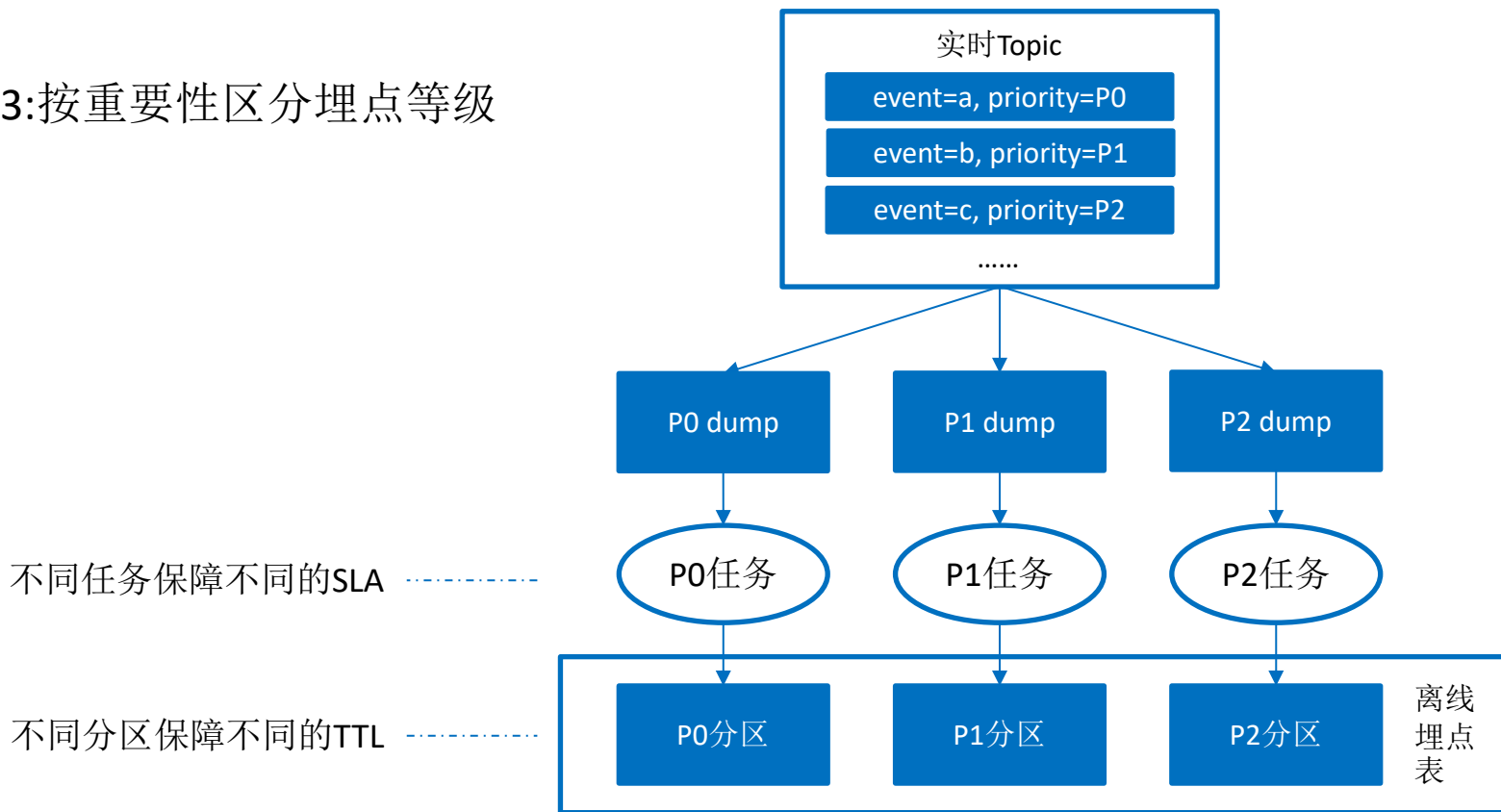
# 埋点成本治理多场景策略

## - 场景3:按重要性区分埋点等级



## 埋点成本治理多场景策略

### - 场景3:按重要性区分埋点等级



# 埋点成本治理多场景策略

## - 场景3:按重要性区分埋点等级

通过平台区分等级：在ByteIO平台对埋点区分等级，不同等级将对应不同SLA/TTL保障。

multi\_sync P2 全量 已上线 分析 禁用

埋点描述 功能演示 前日上报PV 操作系统 Android, iOS

基本信息 场景细分 变更历史 血缘关系

触发时机 - 负责人 - 标签 testAdd 编辑

事件类型 性能埋点 相关人 - 开发类型 代码埋点

全部属性 搜索名称/描述 新建属性

属性	类型	描述	Value录入值	必传	前日上报PV	状态	操作
duration	integer	参数演示	-	否	-	已上线	Value采样 编辑

共 1 行 每页行数 20

节省100+PB HDFS存储

# 埋点成本治理多场景策略

## - 场景4:支持采样上报

04

支持采样上报

问题

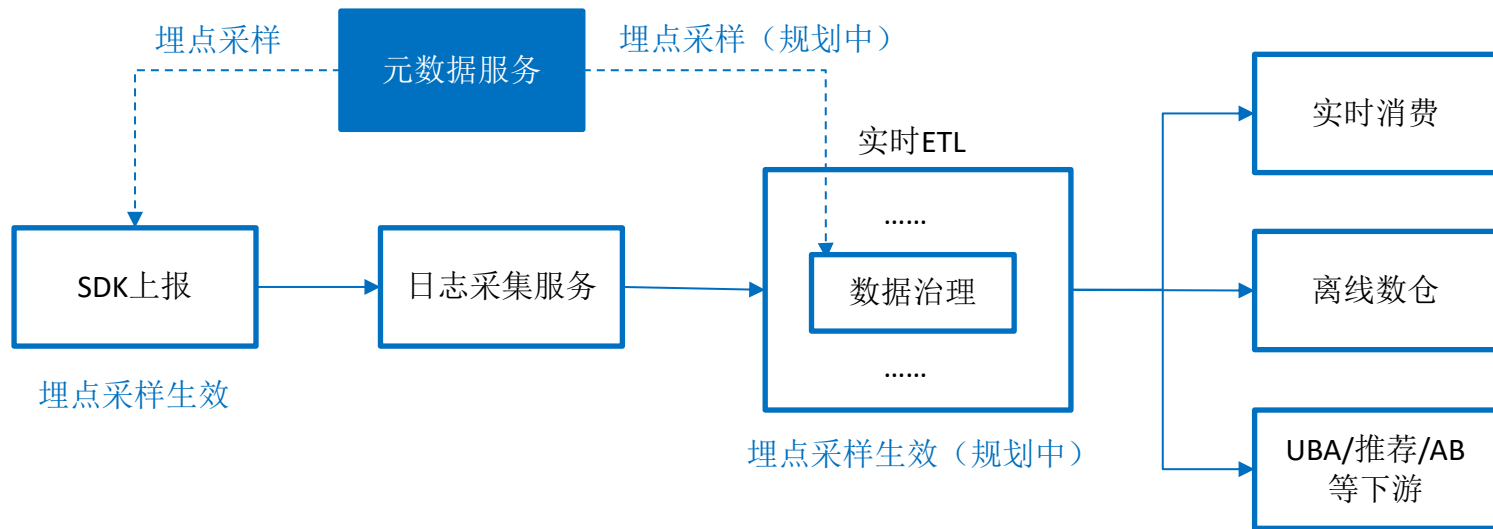
- 埋点需要使用，但不需要全量上报，有一定资源浪费
- 案例：RD排查问题使用的埋点，可采样上报

解决方案

- 支持埋点采样上报

# 埋点成本治理多场景策略

## - 场景4:支持采样上报



# 埋点成本治理多场景策略

## - 场景4:支持采样上报

通过平台区分采样：在ByteIO平台对埋点设置采样比例，SDK将按照采样比例上报埋点。



multi\_sync | P2 | 5%抽样 | 已上线

埋点描述 功能演示 前日上报PV - 操作系统 Android, iOS

基本信息 场景细分 变更历史 血缘关系

触发时机 - 负责人 - 标签 testAdd 编辑

事件类型 性能埋点 相关人 - 开发类型 代码埋点

全部属性 搜索名称/描述 新建属性

属性	类型	描述	Value录入值	必传	前日上报PV	状态	操作
duration	integer	参数演示	-	否	-	已上线	Value采样 编辑

共 1 行 每页行数 20

在2022年节省成本 ¥3000+万/年

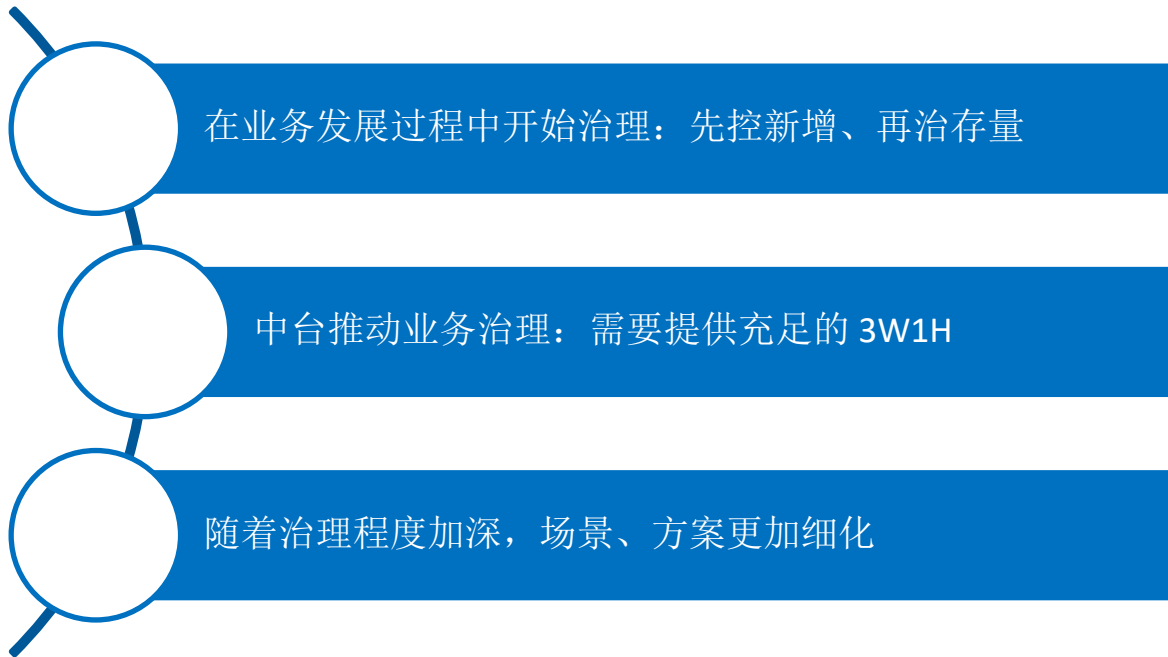
# 03

## 埋点成本治理经验 回顾



## 埋点成本治理经验总结

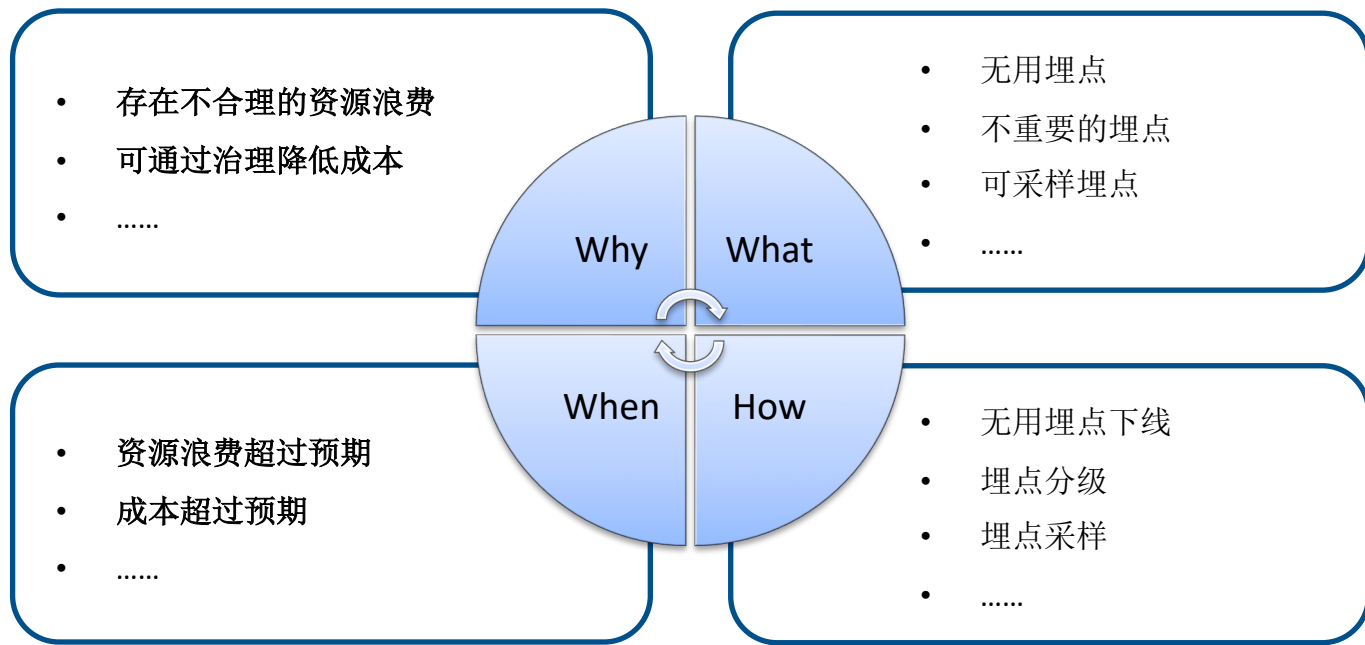
### - 推动治理过程中的一些问题





# 埋点成本治理经验总结

## - 推动业务治理：3W1H



# 埋点成本治理经验总结

## - 推动业务治理：提供观测指标



埋点上报总量：上报  $a$  万亿条/天



埋点成本：花费  $b$  万元/天



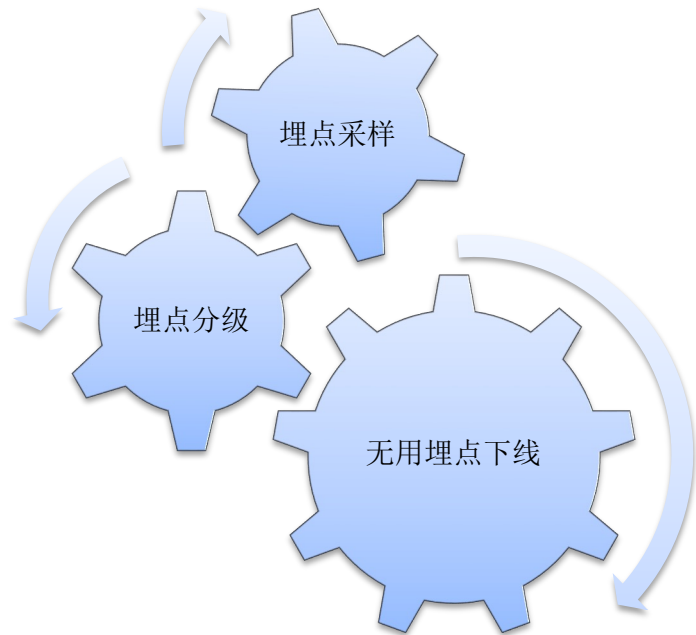
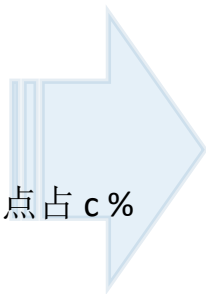
无用埋点占比：在上报总量中，无用埋点占  $c\%$



埋点密度：平均上报  $d$  条/分钟

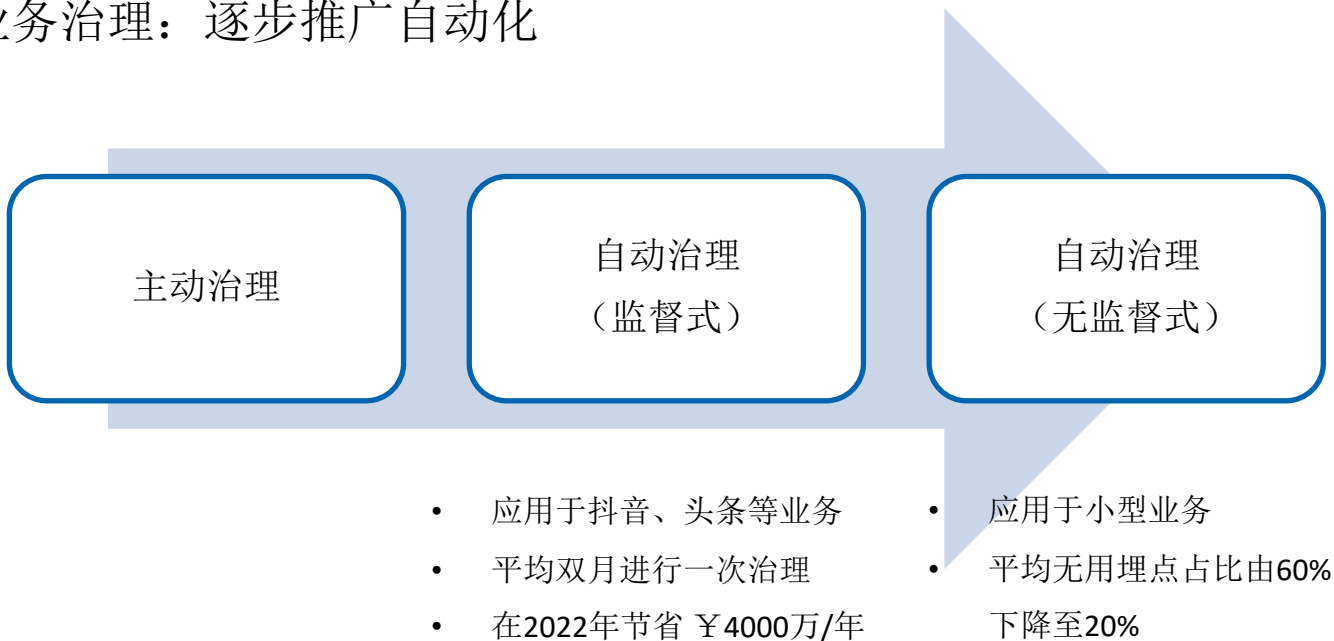
埋点上报总量  $\div$  用户活跃总时长

...



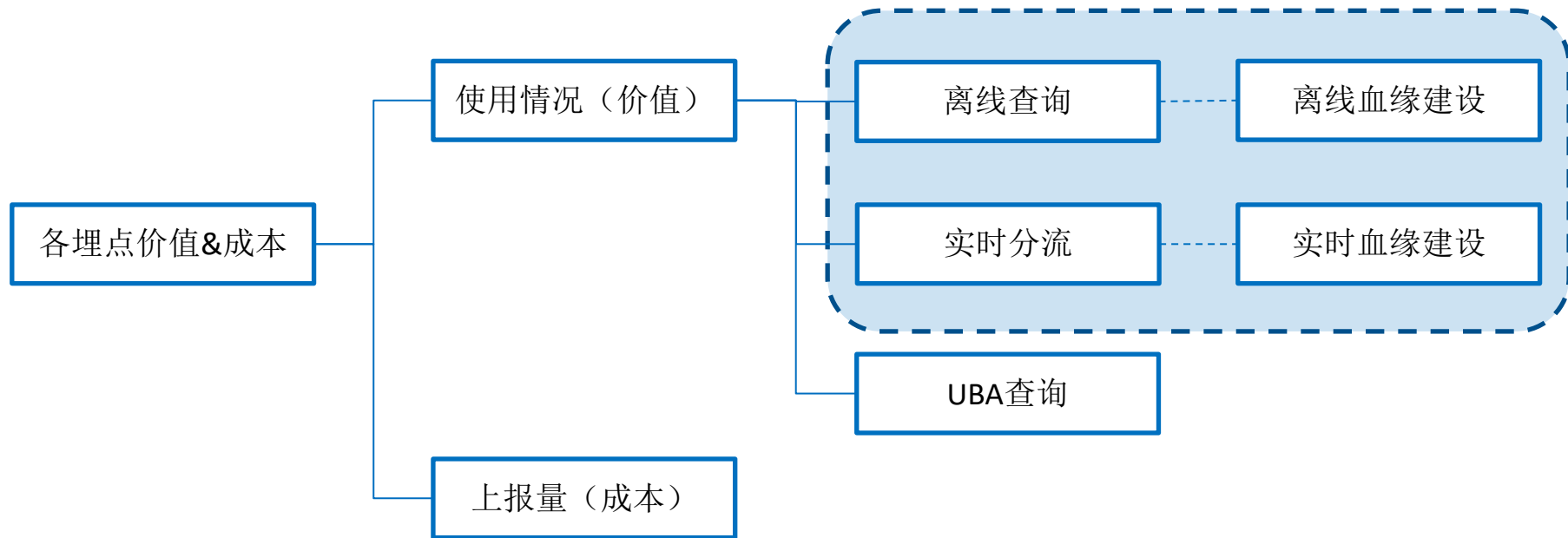
## 埋点成本治理经验总结

- 推动业务治理：逐步推广自动化



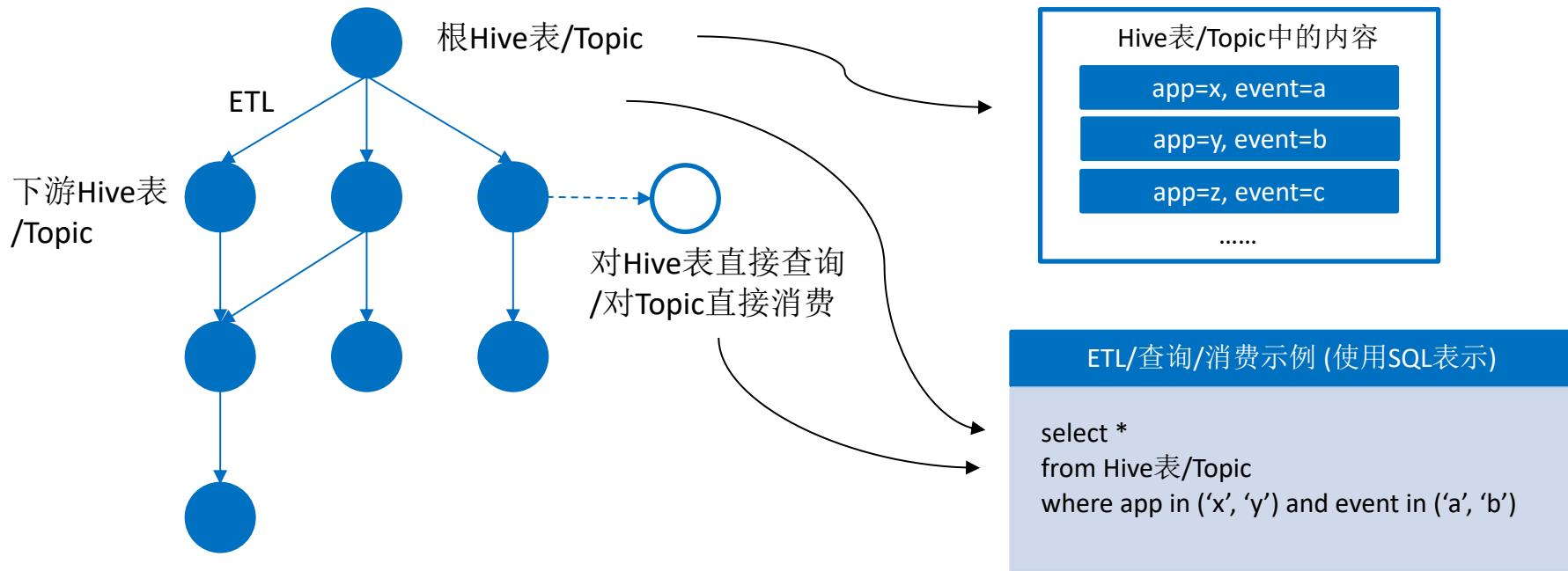
# 埋点成本治理经验总结

## - 埋点使用情况分析



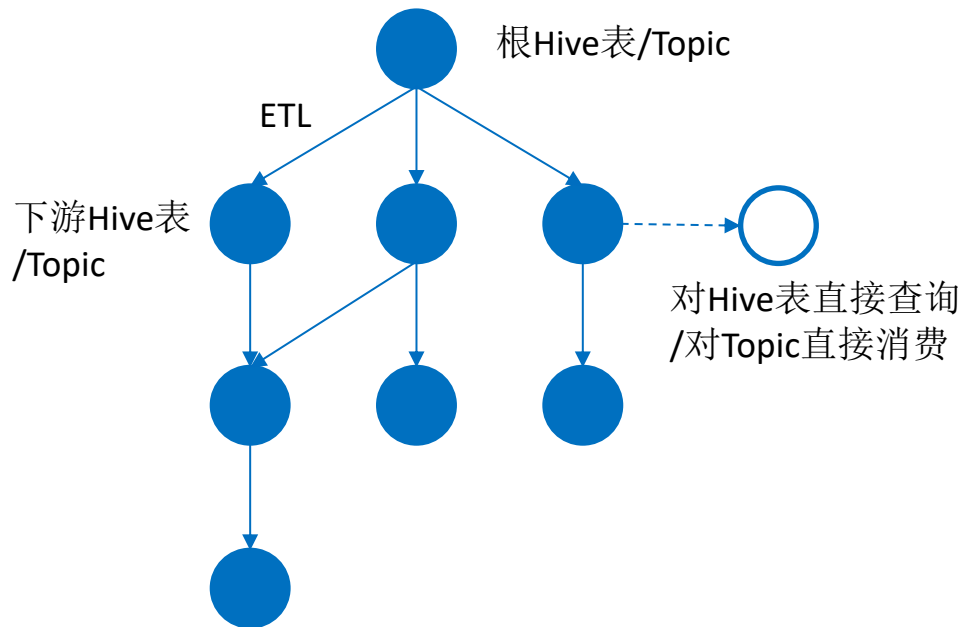
# 埋点成本治理经验总结

## - 埋点使用情况分析：行血缘



# 埋点成本治理经验总结

## - 埋点使用情况分析：行血缘



### 目标

- 各层级Hive表/Topic包含埋点
- 各层级Hive表/Topic被查询、消费的埋点

### 方法

- 解析ETL/查询/消费内的埋点相关逻辑
- 结合Hive表/Topic上下游关系，逐层传播

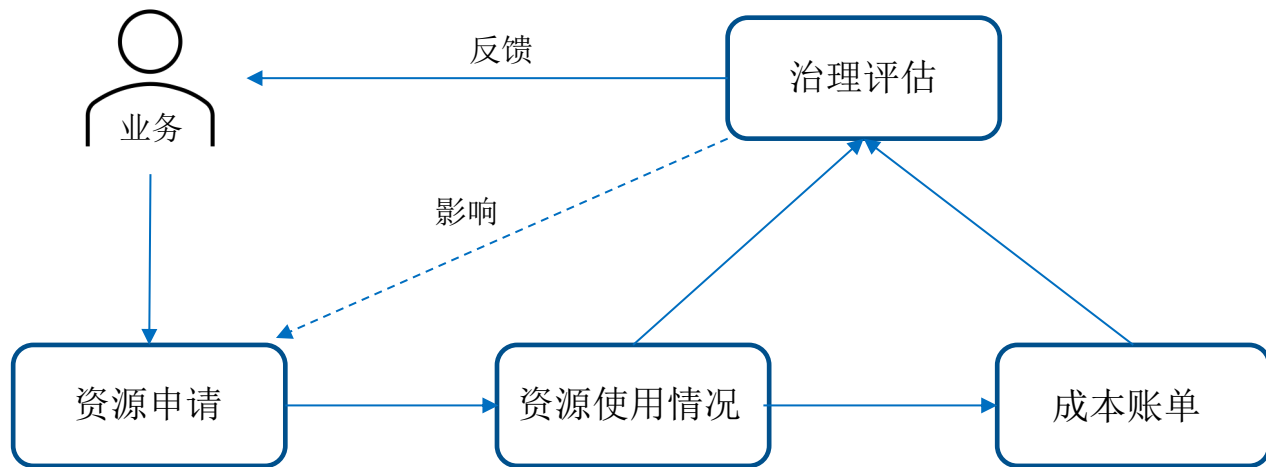
# 04

## 规划与展望



## 规划与展望

### - 打通成本与资源





## 规划与展望

- 根据不同业务现状，个性化推荐治理方案



无用埋点占比高?

埋点下线

HDFS成本高?

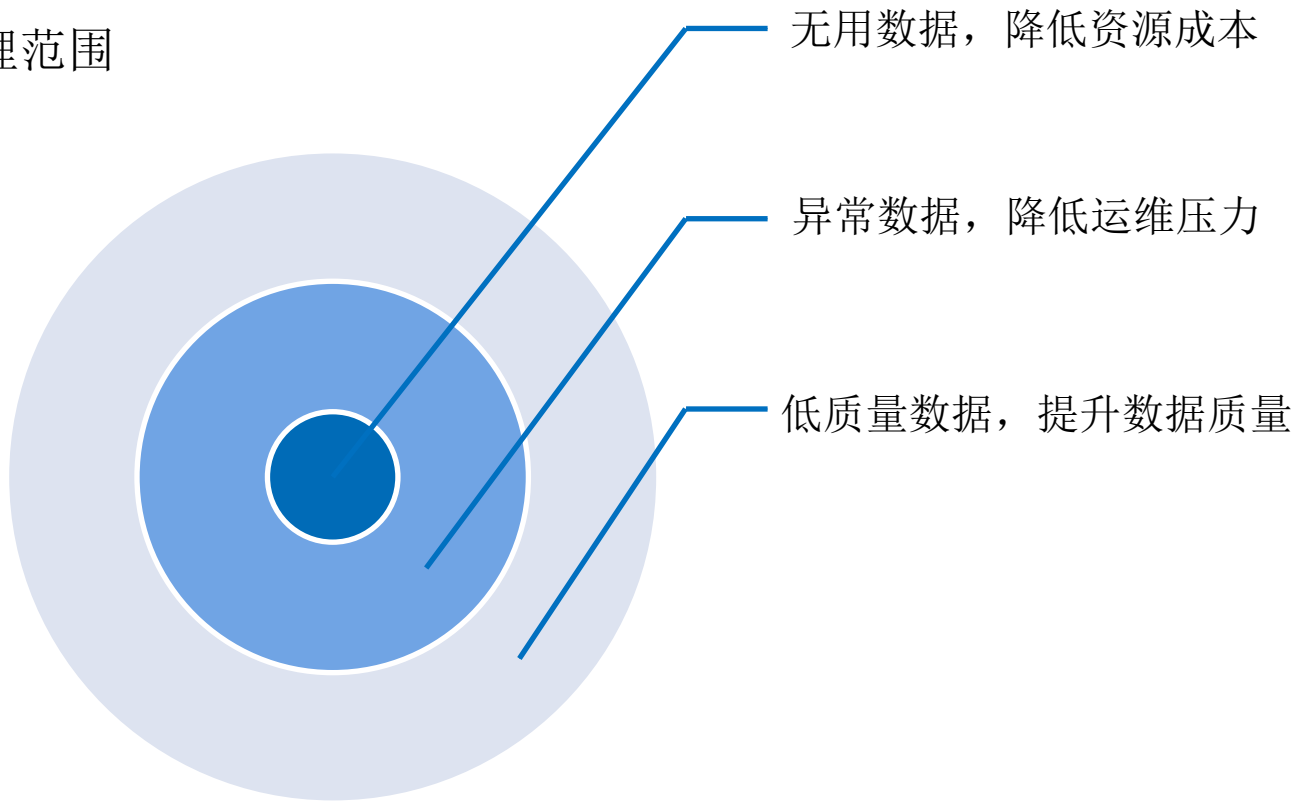
埋点分级

RD性能埋点过多?

埋点采样

## 规划与展望

### - 扩展治理范围



欢迎联系我们



扫码关注

“字节跳动数据平台” 微信公众号



扫码添加小助手

进入 “字节跳动数据平台” 官方交流群

# 非常感谢您的观看

---



火山引擎



| DataFun.

