

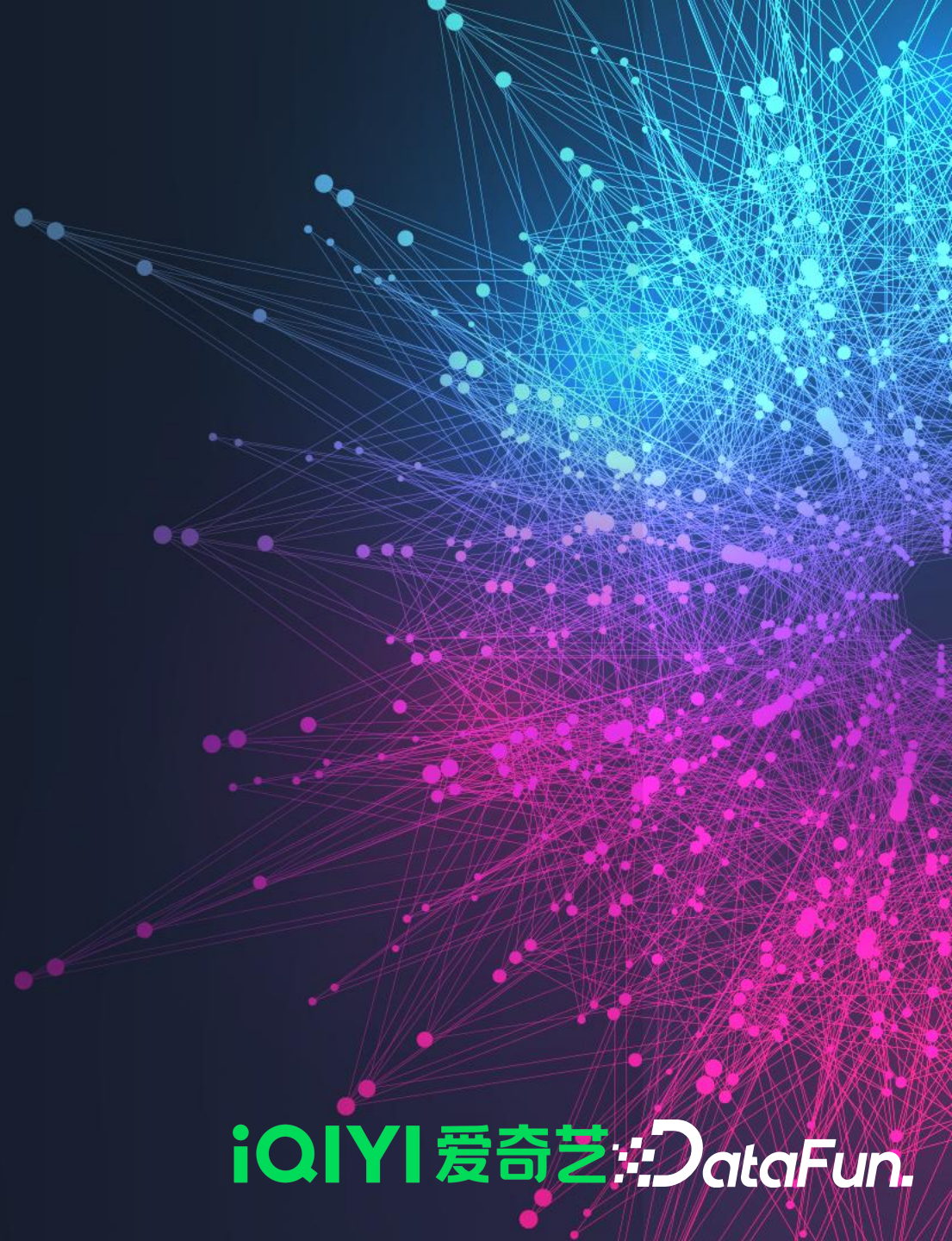
爱奇艺埋点体系

侯耀斌

pingback系统 技术负责人

DataFunSummit # 2023

iQIYI 爱奇艺 DataFun.



目录 CONTENT

01 埋点对数据的重要性

03 主要的投递时机和事件

02 爱奇艺埋点体系的演进

04 数据的后续加工处理

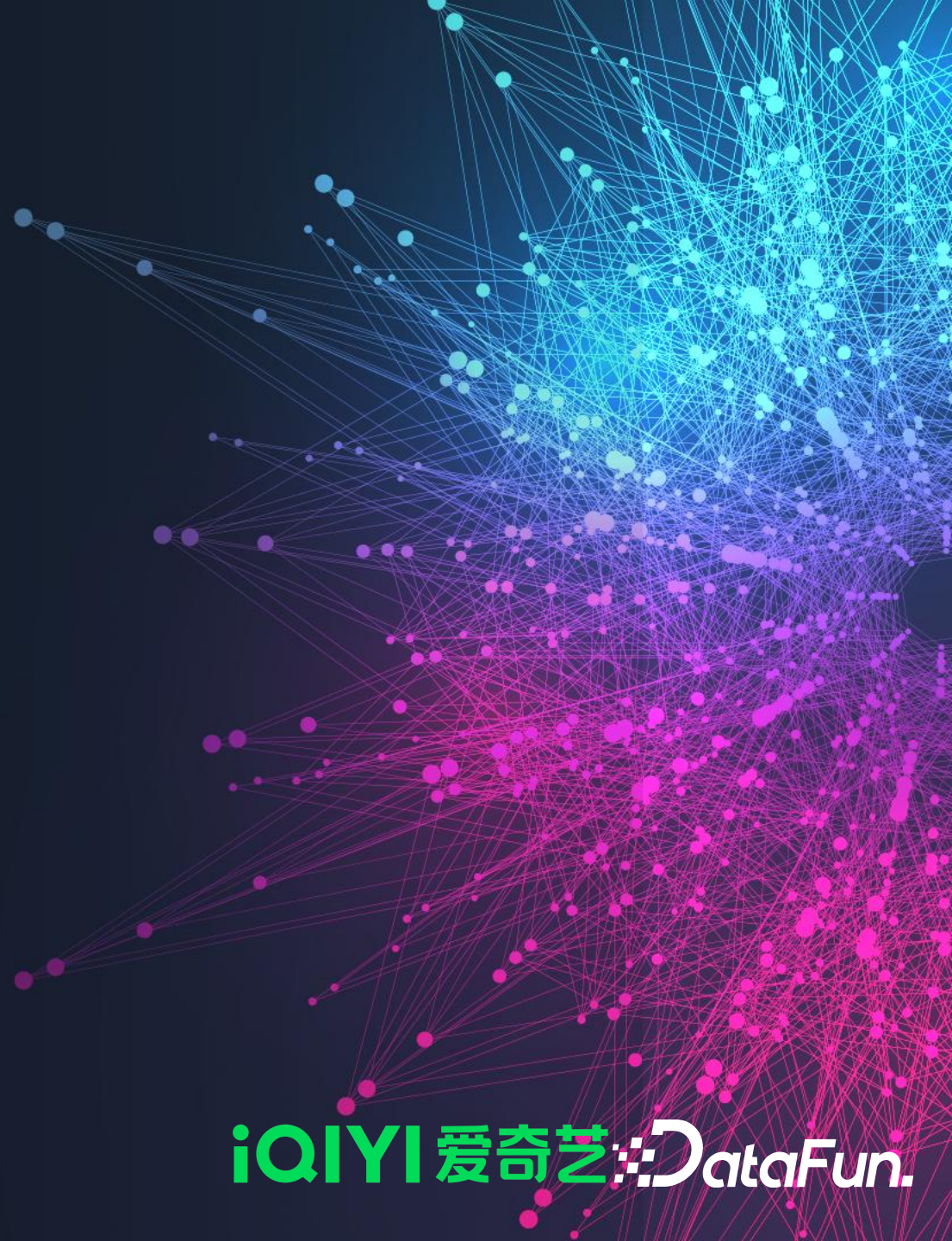
05 正在做的改进

01

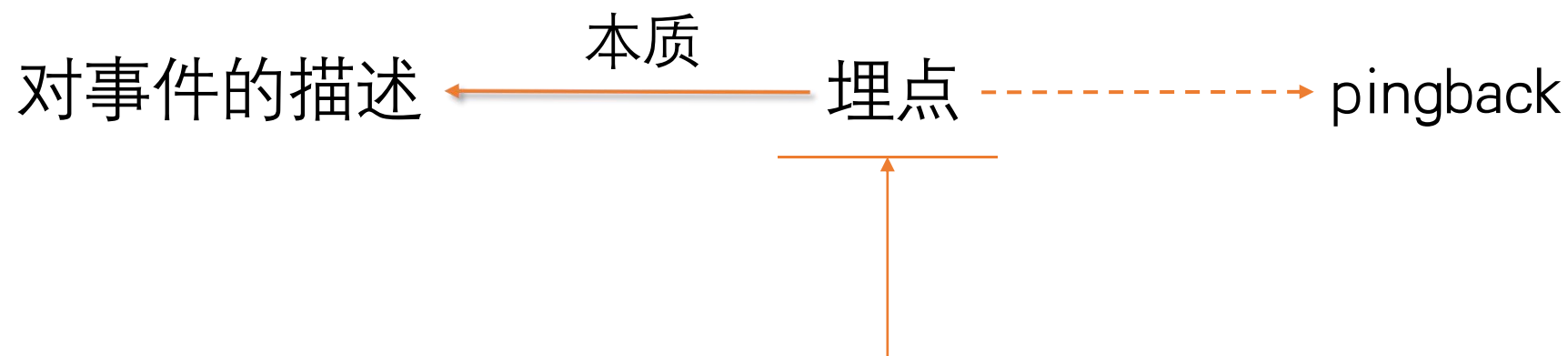
埋点对数据的重要性

DataFunSummit # 2023

iQIYI 爱奇艺 DataFun.



什么是埋点



在研究对象中特定的流程收集一些信息

跟踪研究对象使用的状况，为后续的优化或运营策略提供数据支撑

埋点的要素



埋点的意义

- 数据的基础
- 统计分析，活跃、留存，衡量增长、运营的质量
- 用户画像、标签准确性
- 推荐效果，最终体现在用户时长

02

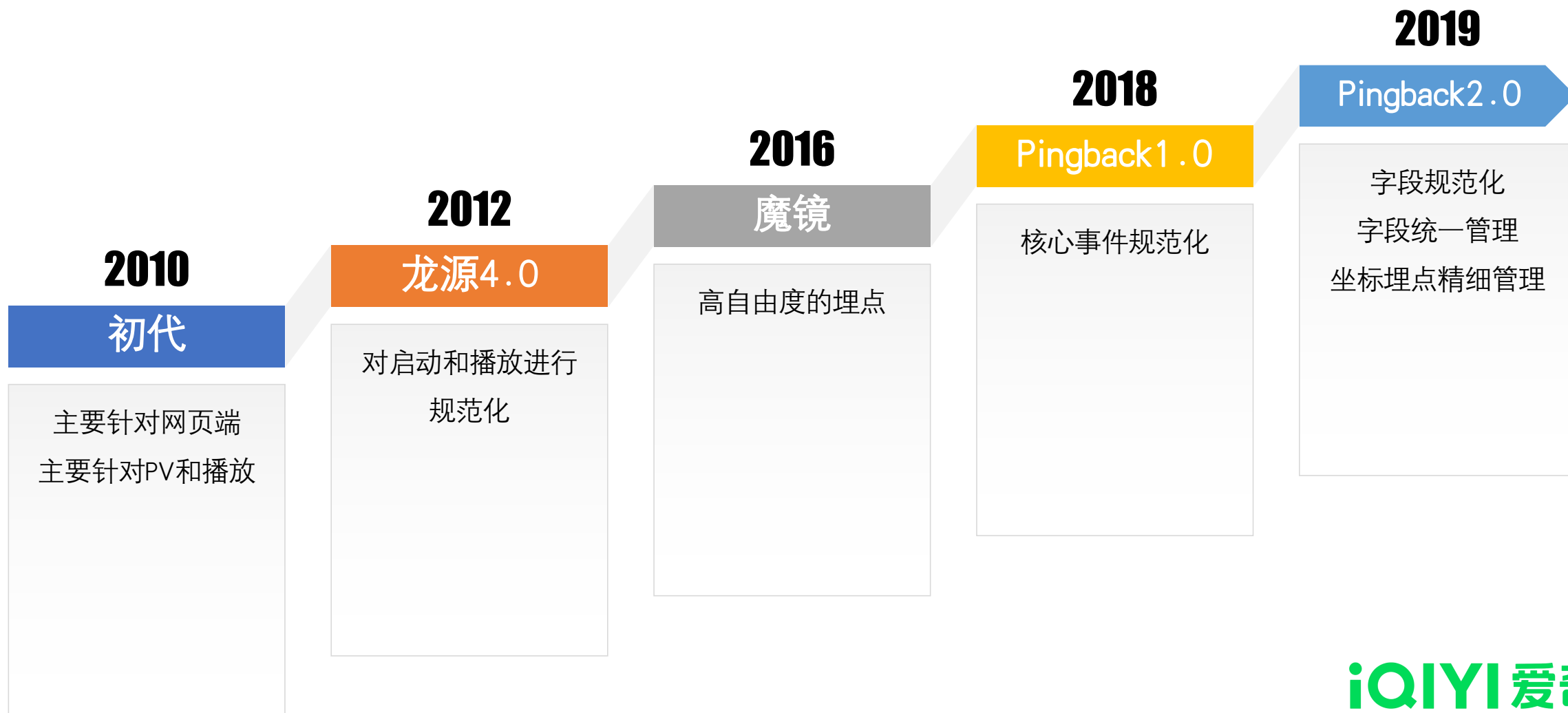
爱奇艺埋点体系的演进

DataFunSummit # 2023

iQIYI 爱奇艺 DataFun.



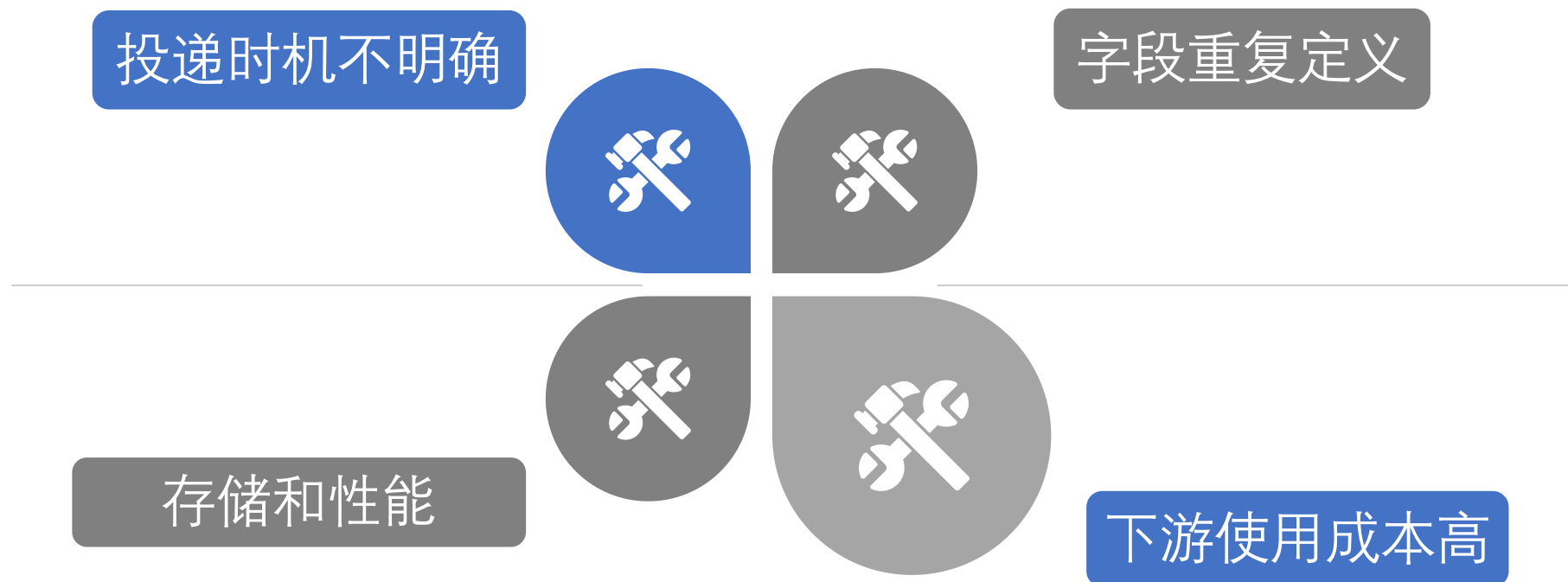
爱奇艺pingback的历史



魔镜：自助化阶段



自助化阶段出现的问题



中台阶段

二

字段库

启动、播放，展现点击
阅读、投屏、支付

二

字典值

梳理、清洗，核心字典统一
编码

二

页面坐标

页面，区块

二

统一核心事件

启动、播放，展现点击

二

QOS、自定义事件

自由定义时机，灵活使用投递

中台阶段

统一核心事件， 启动、播放， 展现点击
细化事件类型， 阅读、投屏、支付、下载

页面打开时、锁屏解锁或从后台切回前台、
返回至该页面、从其它页面返回该页、手动
或超时自动刷新时；特别的，播放页如果有
手动切集或自动联播，需要投出一次播放页
的展示

爱奇艺BI · Pingback			
规范查询			
事件列表			
事件类别/事件	t/ct	事件投递时机	详细说明
› 启动退出			
› 播放			
√ 展现点击			
页面展示	■	页面打开时、锁屏解锁或从后台切回前台后返回至该页面、从其它页面返回该页、手动或超时自动刷新时；特别的，播放页如果有手动切集或自动联播，需要投出一次播放页的展示	页面打开时、锁屏解锁或从后台切回前台后返回至该页面、从其它页面返回该页、手动或超时自动刷新时；特别的，播放页如果有手动切集或自动联播，需要投出一次播放页的展示
页面停留	■	当前页面关闭（包括锁屏或者退到后台，和页面展现时机对应）或跳转的时候投递	当前页面关闭（包括锁屏或者退到后台，和页面展现时机对应）或跳转的时候投递
区块展示	■	区块露出时、返回至该区块有刷新时、手动或超时自动刷新时，已投过的区块不再投（有页面、区块刷新除外）	区块露出时、返回至该区块有刷新时、手动或超时自动刷新时，已投过的区块不再投（有页面、区块刷新除外）
内容位展示	■	按实际露出至少一个像素的内容投递，已投过的无刷新则不需要重投（有页面刷新或切换除外）	按实际露出至少一个像素的内容投递，已投过的无刷新则不需要重投（有页面刷新或切换除外）
点击	■	用户手动点击或使用工具选择点击的时候投递	用户手动点击或使用工具选择点击的时候投递
焦点	■	电视端遥控的选择光标移动至输入框尚未点击时投递	电视端遥控的选择光标移动至输入框尚未点击时投递
› 状态事件			
› 支付			
› 阅读			
› 下载			
› 投屏			
› 互动			
› 自定义事件			
› ...			

中台阶段

字段库

接近500个字段

40+个标签

字段

高级条件：无

查询

适用范围	投递字段/dwd字段	字段名称	优先级	字段类型	标签	备注
全部	/ use_duration	耗时	P0	NUMBER	耗时	记录各种事件的耗时，单位：毫秒；配合各个事件使用，如使用时长、播放时长
全部	/ high_energy_play_duration	高能片段播放时长	P1	NUMBER	耗时	高能片段主动点击播放，在结束播放事件中，tm字段即播放时长添加了高能片段的时
全部		上次使用时长	P2	NUMBER	耗时	上次冷启动到冷启动退出整个生命周期内的使用时长，即取上次冷启动生命周期的tm
全部	play_vr_duration	vr分屏模式播放时长	P2	NUMBER	播放 VR	用于记录VR模式下的实际播放时长，单位：秒，取整，四舍五入。一次播放时多次切
全部	/ kd_duration	卡顿时长	P2	NUMBER	播放 耗时	卡顿时长，用于基线gphone播放时长（t=2）中出现卡顿情况下的时长统计
全部	/ play_earphone_duration	耳机播放时长	P2	NUMBER	播放	用户使用耳机播放的播放时长，单位秒，取整，四舍五入。当一次播放有多次使用耳机
全部	barg_duration	弹幕开启播放时长	P2	NUMBER	播放	弹幕开启状态下的播放时长，一次播放多次开启关闭需要加和开启的周期的播放时长，
全部		投屏的剩余时长预估	P2	NUMBER	投屏	投屏断连时，投递当前播放视频的剩余时长。单位：毫秒
全部	play_backend_duration	后台耗时	P2	NUMBER	耗时	目前用于播放行为中后台播放的时长记录
全部	/ ee_duration	耳机增强模式播放时长	P2	NUMBER	播放	毫秒； 当次播放期间，用户使用耳机增强模式的总时长； 用户成功启用耳机增强模式

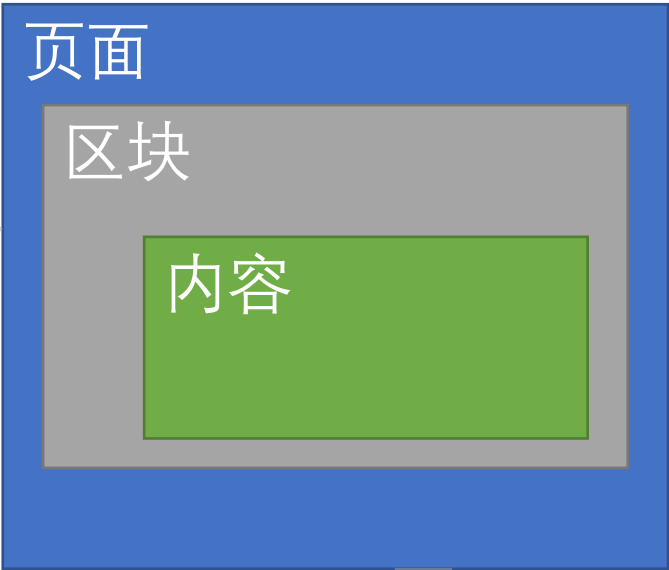
10条/页

1 2 3 4

环境、播放、耗时，会员、行为串联...

中台阶段

页面坐标管理
明确页面元素的层级关系
为数据切换提供可能性



rpage/block	名称	2.0是否通过	2.0通过版本	类型	编辑状态	
category_home.1	电影推荐页	验证通过	11.3.0	页面	已通过	查看
category_home_popup	频道全屏浮层	验证通过	12.5.5	区块	已通过	查看
F1:1000000	底feed单片大图	验证通过	11.12.0	区块	已通过	查看
F1:10000001	底feed榜单模块	验证通过	11.12.0	区块	已通过	查看
F1:10000002	底feed主题模块	验证通过	11.12.0	区块	已通过	查看
F1:10000003	电影频道底feed影人模块	验证通过	11.12.0	区块	已通过	查看
F1:1000000_1	电影频道feed流标签召回card			区块	已通过	查看
O:0200010010	电影频道底部主题feed	验证通过	11.3.0	区块	已通过	查看
O:0200010019	精选榜单	验证通过	11.3.0	区块	已通过	查看
R:13279113712	phone电影频道1010榜单切换	验证通过	11.3.0	区块	已通过	查看
R:16279888712	品牌专区主推 (有深色背景和品牌logo)	验证通过	11.3.0	区块	已通过	查看
R:19745913212	phone电影频道垂直入口	验证通过	11.3.0	区块	已通过	查看
R:201713312	phone 电影频道页card网络大电影	验证通过	11.3.0	区块	已通过	查看

中台阶段

字典值，梳理、清洗，核心字典统一编码

核心字段，保证统一

用于统计生产，其他数据平台，提升可读性

字段字典 (启动类型)

字典类型：

全局

分产品

取值/含义/备注

取值	含义
0	客户端普通开启
1	
2	
3	
4	
5	

字段字典 (码流)

字典类型：

全局

分产品

字典来源：

人工维护

第三方系统维护

添加取值

批量操作

取值/含义/备注

20条/页

取值	含义	备注	状态	操作
-1	智能码流	客户端根据网速自动切换码流	通过	<div>编辑</div> <div>删除</div>
1	流畅		通过	<div>编辑</div> <div>删除</div>
2	高清		通过	<div>编辑</div> <div>删除</div>
3	超清		通过	<div>编辑</div> <div>删除</div>
4	720p		通过	<div>编辑</div> <div>删除</div>
5	1080p		通过	<div>编辑</div> <div>删除</div>
10	4K		通过	<div>编辑</div> <div>删除</div>
27	1080P50帧		通过	<div>编辑</div> <div>删除</div>
32	4KMAX		通过	<div>编辑</div> <div>删除</div>
33	蓝光1080p		通过	<div>编辑</div> <div>删除</div>
34	超清720p		通过	<div>编辑</div> <div>删除</div>
35	高清480p		通过	<div>编辑</div> <div>删除</div>
36	流畅360p		通过	<div>编辑</div> <div>删除</div>
37	省流		通过	<div>编辑</div> <div>删除</div>
38	标清		通过	<div>编辑</div> <div>删除</div>
39	2k		通过	<div>编辑</div> <div>删除</div>
40	蓝光		通过	<div>编辑</div> <div>删除</div>
41	8k		通过	<div>编辑</div> <div>删除</div>
42	6k		通过	<div>编辑</div> <div>删除</div>
43	xx1080p		通过	<div>编辑</div> <div>删除</div>

<

1

2

>

新旧埋点双投

01

双倍的生产压力和资源消耗

验证关键指标

02

PV, UV, 人均指标

分析指标差异 业务修复

03

多投、漏投

数仓生产切换

04

新旧数据合并

新旧埋点切换方案

埋点迭代的阻碍

01

业务切换成本高

开发量大，陈旧代码维护困难

02

不能及时看到收益

时间跨度长，涉及多个版本

03

数据验证难度大

深入业务，下钻分析

04

长时间的数据生产合并

旧版本数据，自然消亡时间长

03

主要的投递时机和事件

DataFunSummit # 2023

iQIYI 爱奇艺 DataFun.

启动

01



广义的启动

拉起
后台切前台

02



平台之间的差异

安卓和IOS
app和web的差异

03



重要性

统计新增、日活、版本覆盖

页面展现和点击

02

区块展现

区块首次露出时，返回至该区块且有刷新时，手动或超时自动刷新时；页面、区块刷新除外，已投过的区块不再投。

04

点击

用户的任何操作，都可以认为是点击

01

页面展现

页面打开，刷新

03

内容展现

更细粒度的展示

05

页面停留

当前页面关闭（包括锁屏或者退到后台，和页面展现时机对应）或跳走的时候投递

播放

开始播放

开播第一帧画面

正片播放

真实视频开播第一帧

播放计时

持续播放时，间隔一段时间记录播放时钟计时

结束播放

正常播完、退出播放器、
离开播放页或退出浏览器、应用

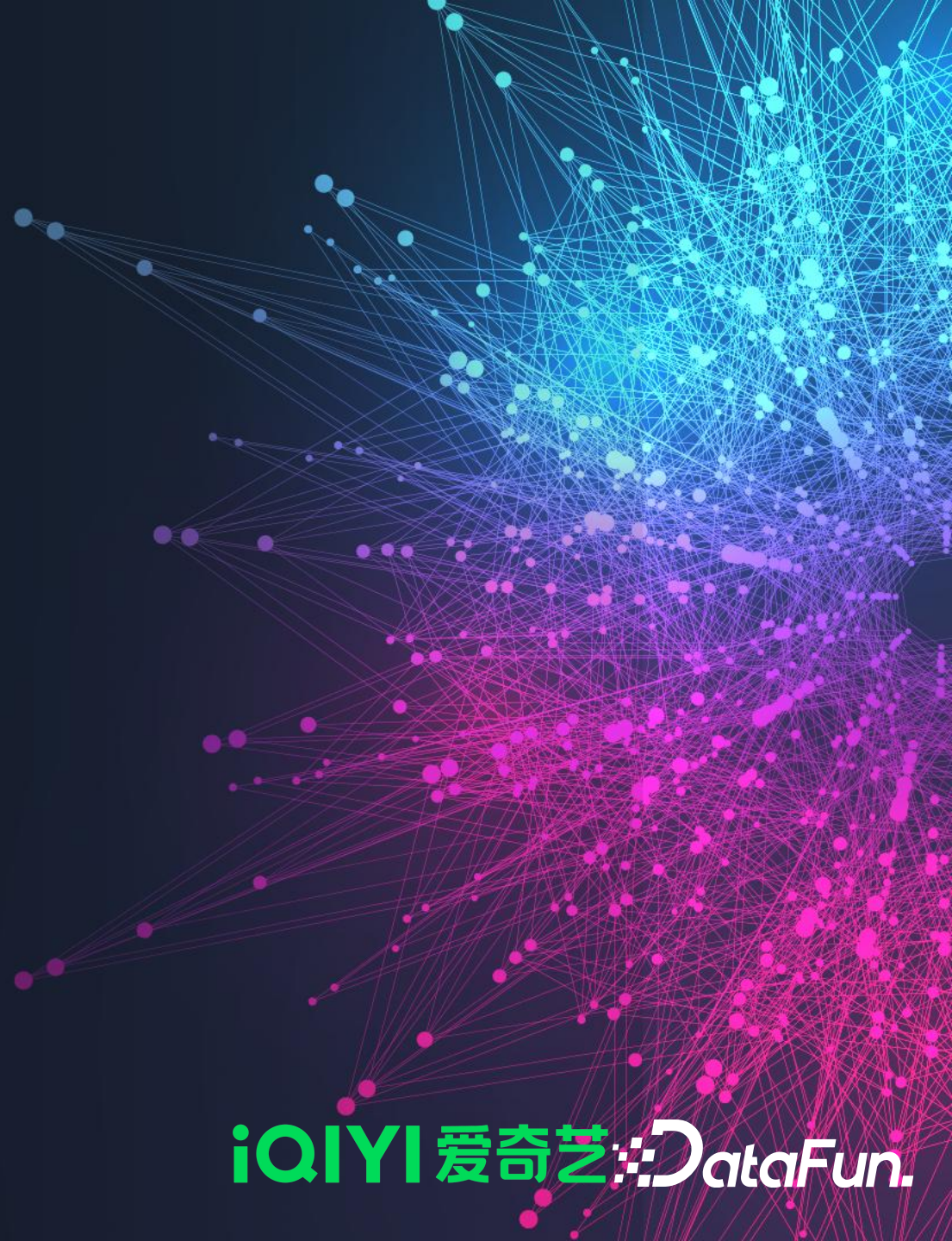


04

数据的后续加工处理

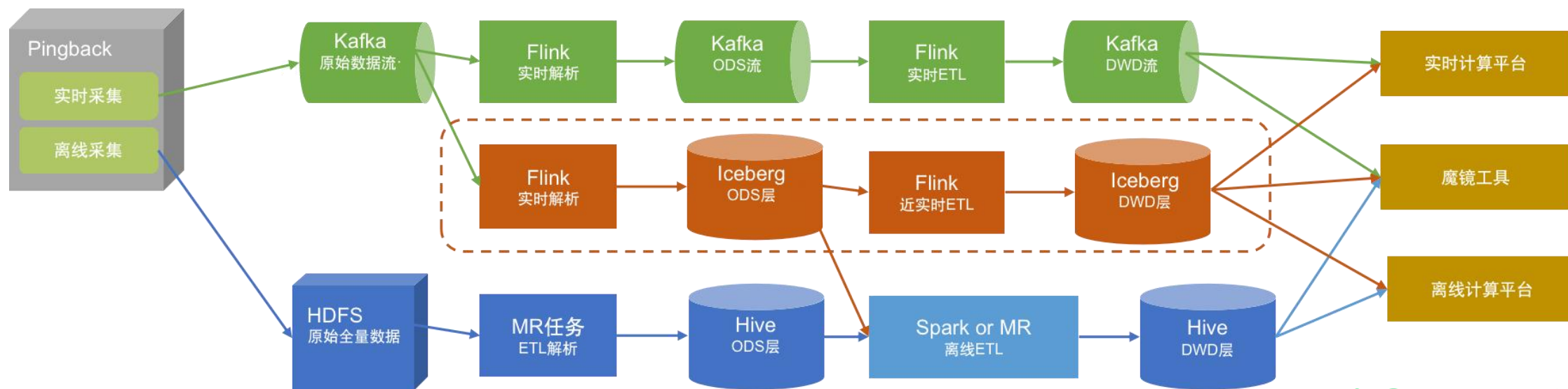
DataFunSummit # 2023

iQIYI 爱奇艺 DataFun.



数据湖和实时流

- 自定义、非核心的投递直接入湖，ODS数据入湖有限开放
- 提升消费的实时性，小时级到近实时
- 实时流



数仓生产

01

字段重命名

提升可读性

02

脏值处理

脏值、异常值过滤或修正

03

维度修正

关联维度表
兼容投递上的错误

04

新旧规范的数据合并

按版本切换

05

聚合表生产

06

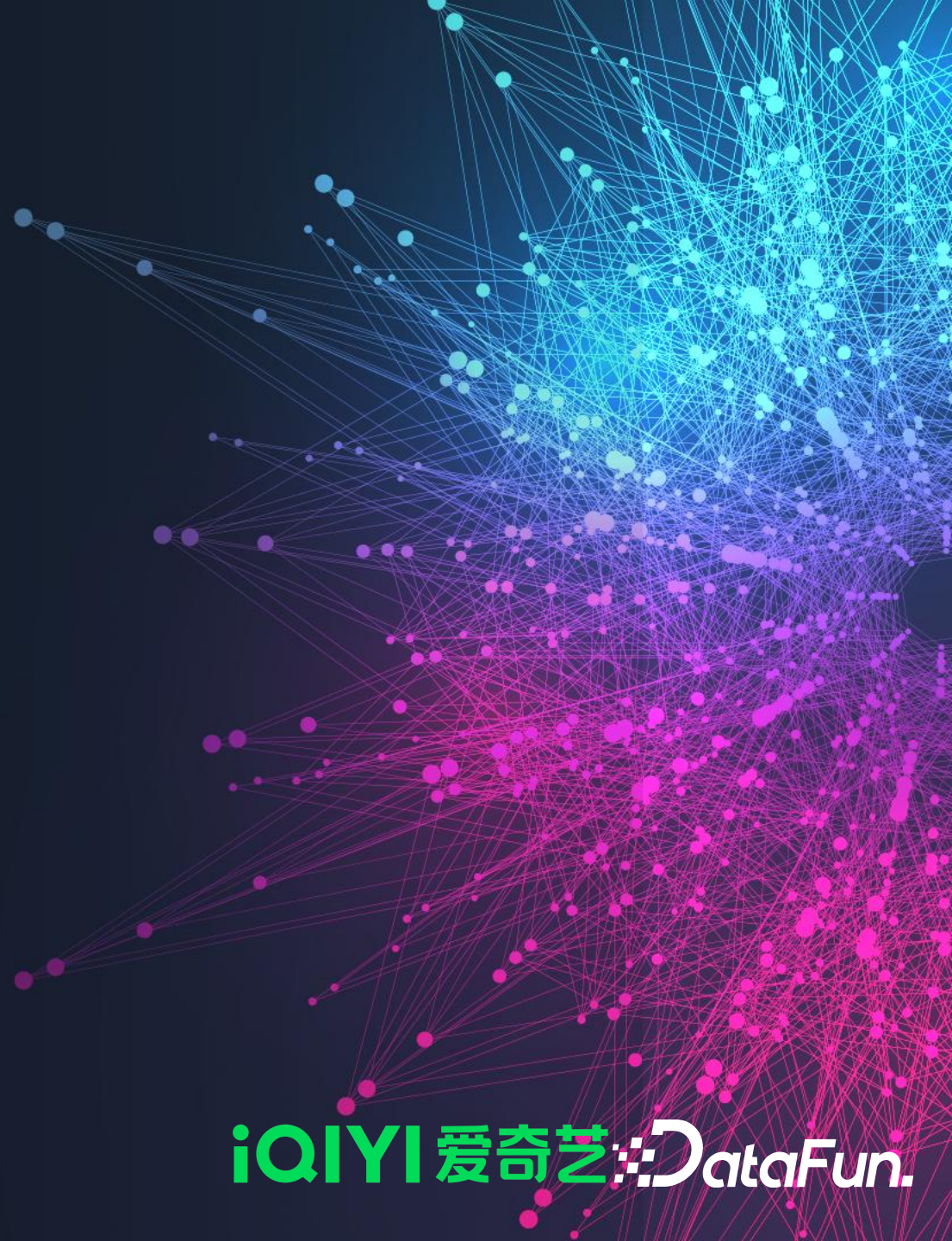
兼容兜底

05

正在做的改进

DataFunSummit # 2023

iQIYI 爱奇艺 DataFun.



正在做的改进

- 结合数据开发平台，分析生产逻辑，建立数据的上下游关系。
- 生产变更通知
- 是否有下游消费，停止生产，降低资源消耗
- 字段血缘，通知统计口径的变更



感谢观看