

袋鼠云基于实时计算的反黄牛算法

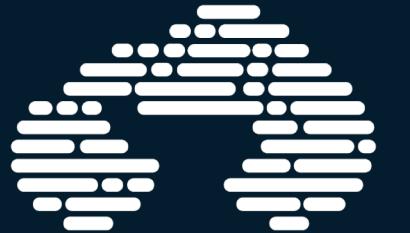
公司：袋鼠云

职位：大数据总监

演讲者：閻賽華（勒蔑）



1 业务背景



袋鼠云
DTSTACK.COM

FLINK
FORWARD

数据采集实时性、稳定性差

黄牛刷单现象严重

异常预警信息监控困难

各渠道下的流量、交易及售后
物流、服务、退款退货等环节
分析困难

实施成本高
(产品成本+开发成本)

业务
背景



3 反黄牛业务架构

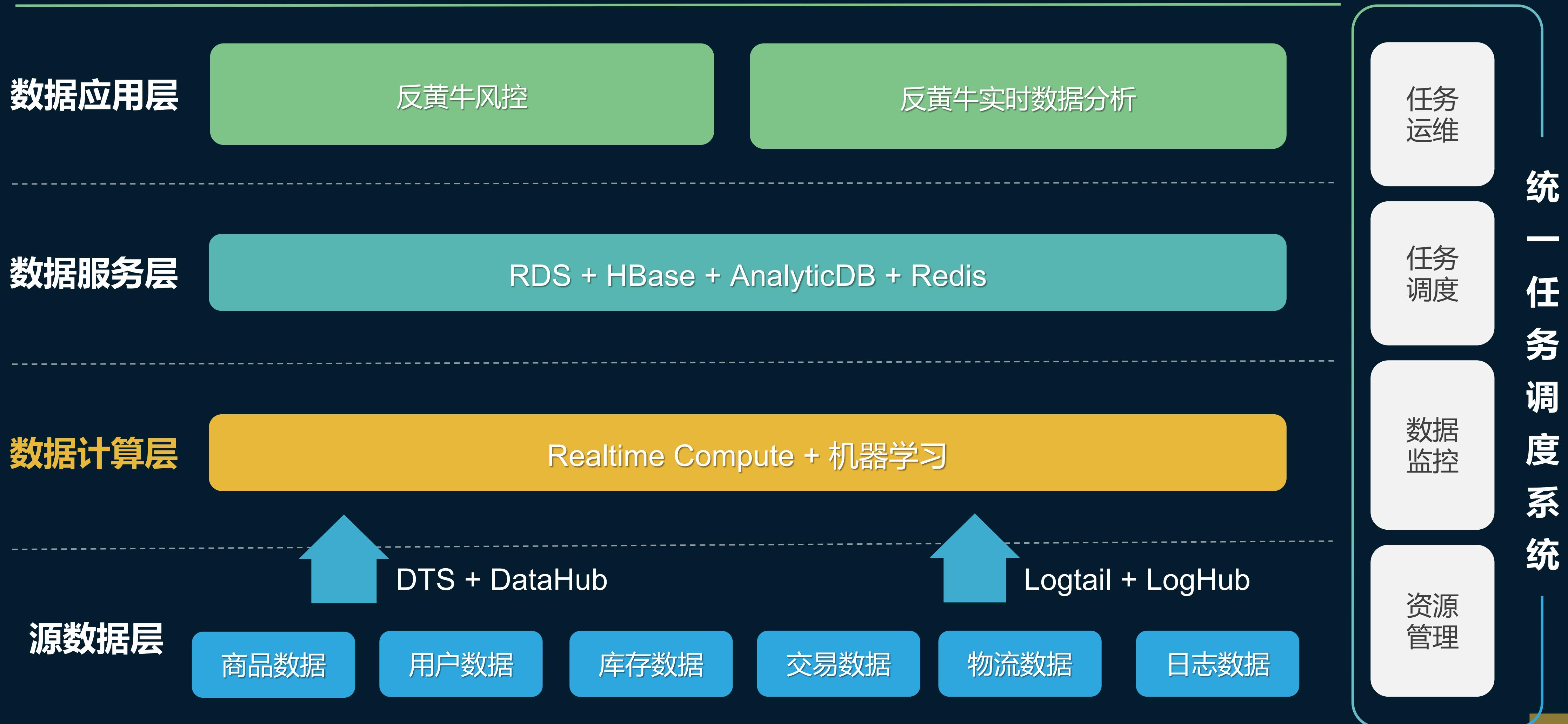


4 基于实时计算的反黄牛技术架构



袋鼠云
DTSTACK.COM

FLINK
FORWARD

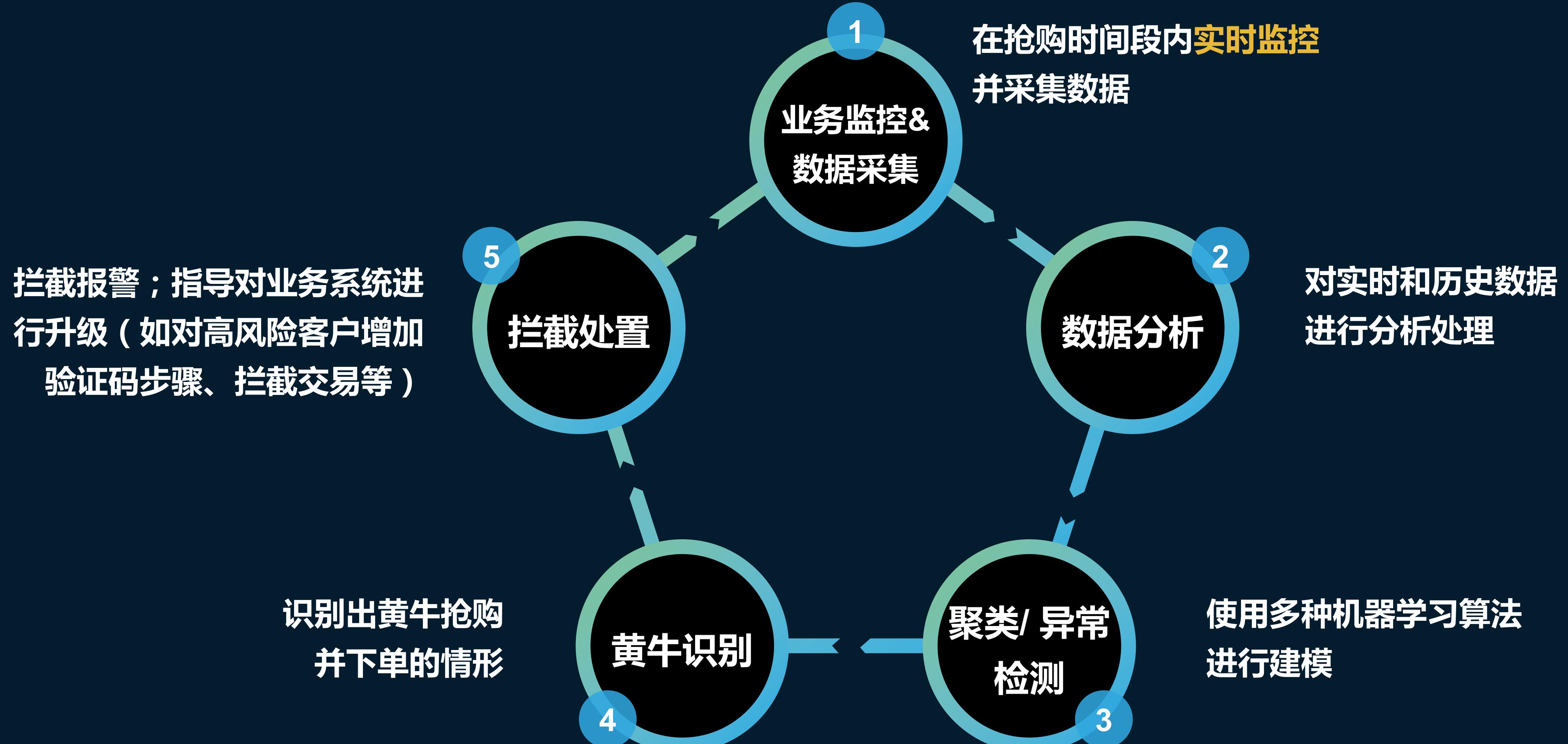


5 反黄牛业务环节

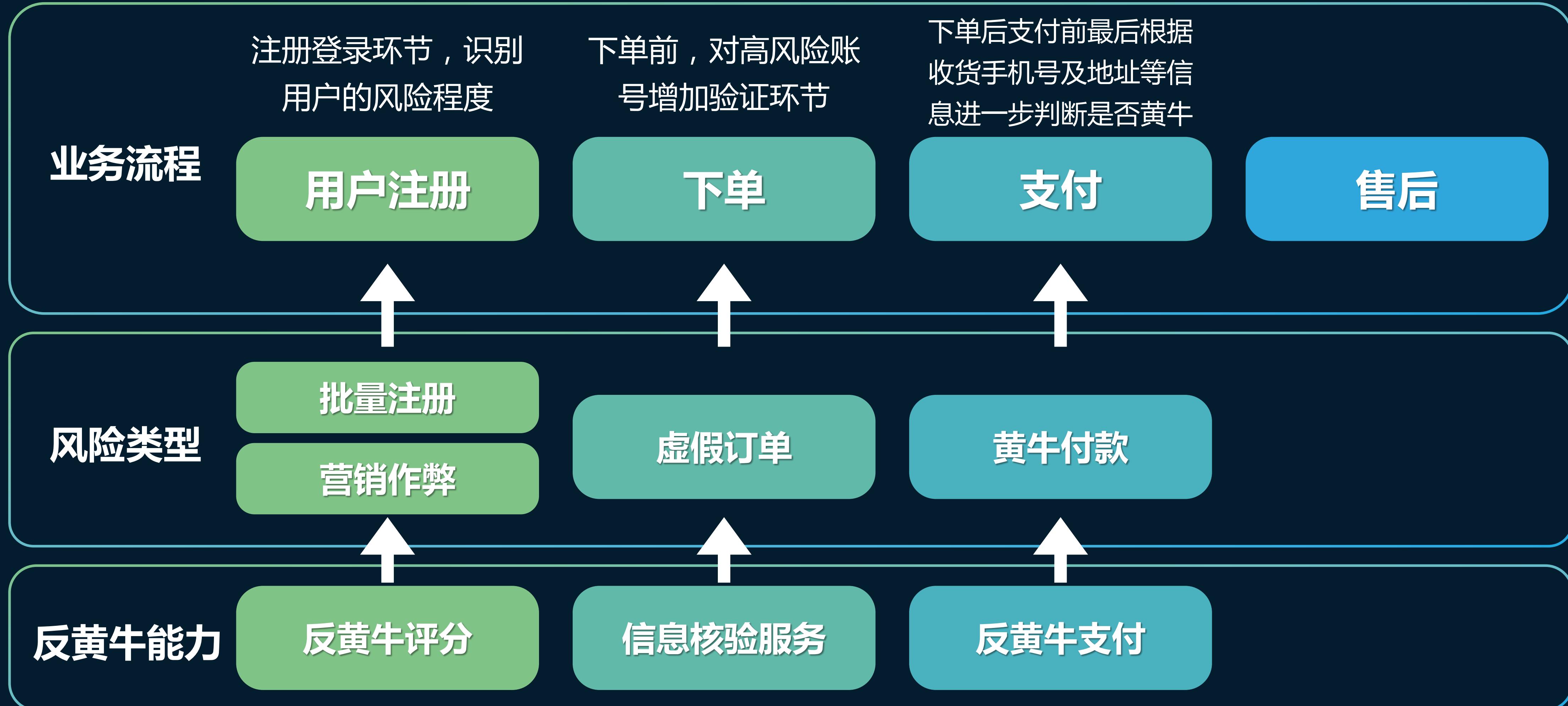


袋鼠云
DTSTACK.COM

FLINK
FORWARD

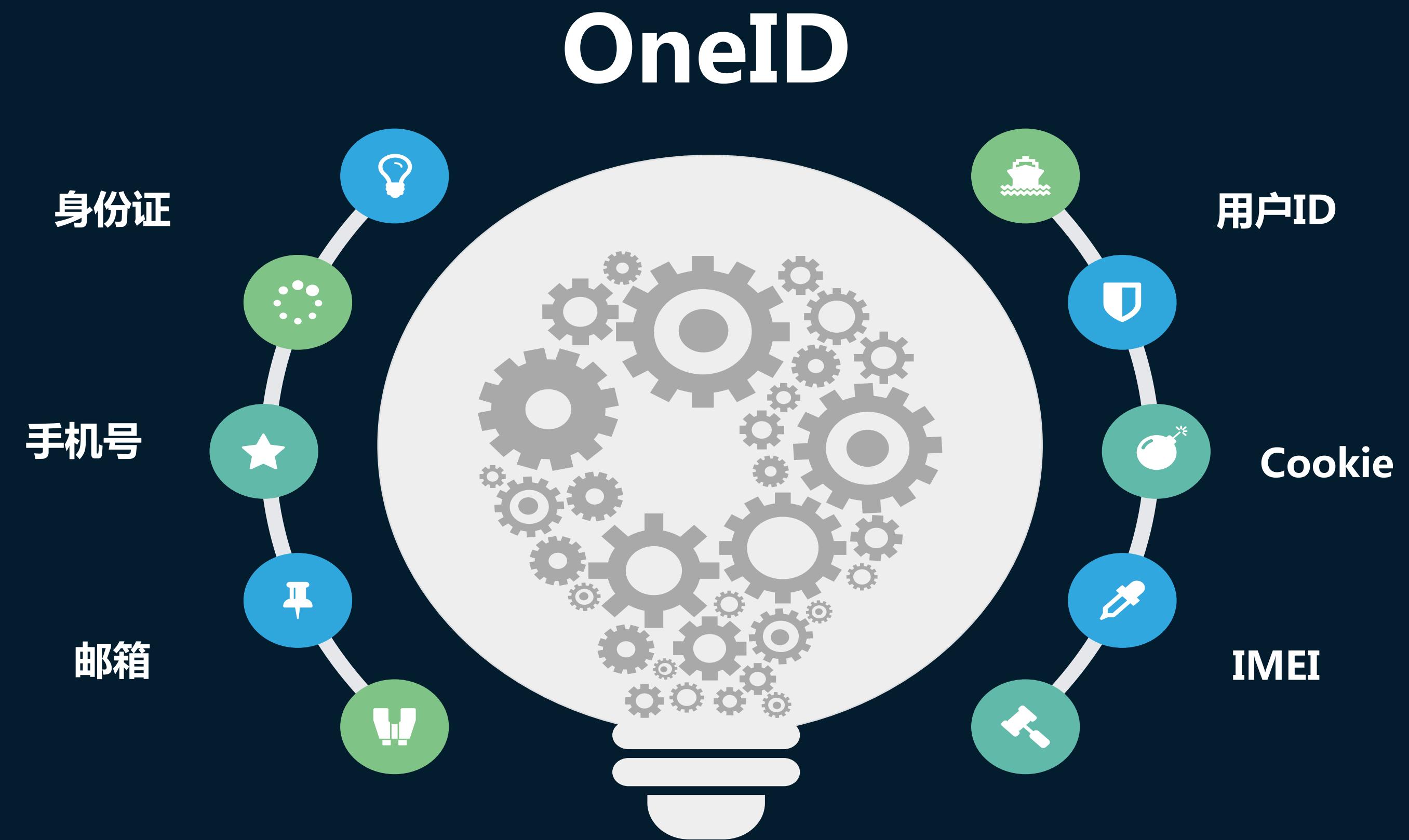


6 反黄牛风控流程



7 黄牛用户识别模型

采用OneID技术，基于IDMapping打通多端、多种形式下的用户识别体系，360度还原黄牛身份

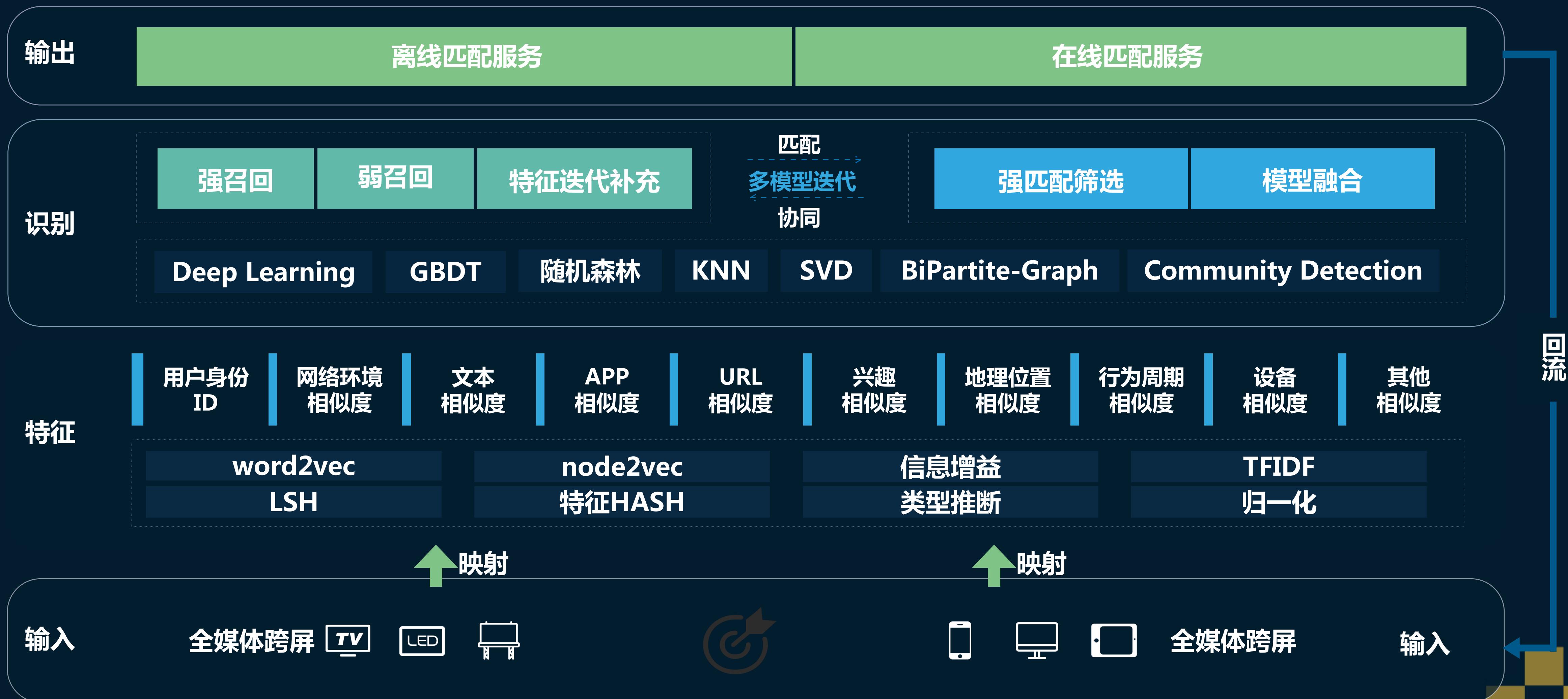


8 黄牛用户识别技术-IDMapping

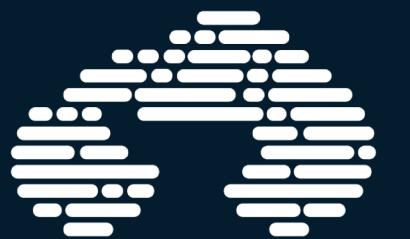


袋鼠云
DTSTACK.COM

FLINK
FORWARD



9 机器学习建模



袋鼠云
DTSTACK.COM

FLINK
FORWARD

- 利用数据提取多维度特征，对于无标签数据，建立异常检测和聚类模型
- 对于有标签数据，建立有监督分类模型
- 建立人工和半自动化反馈打标机制，使用半监督模型

输入接口

历史数据

实时在线数据

数据建模

算法平台

分箱

等频

等宽

自动

特征工程

离散化

WOE

线性模型

逻辑回归

线性回归

输出在线API

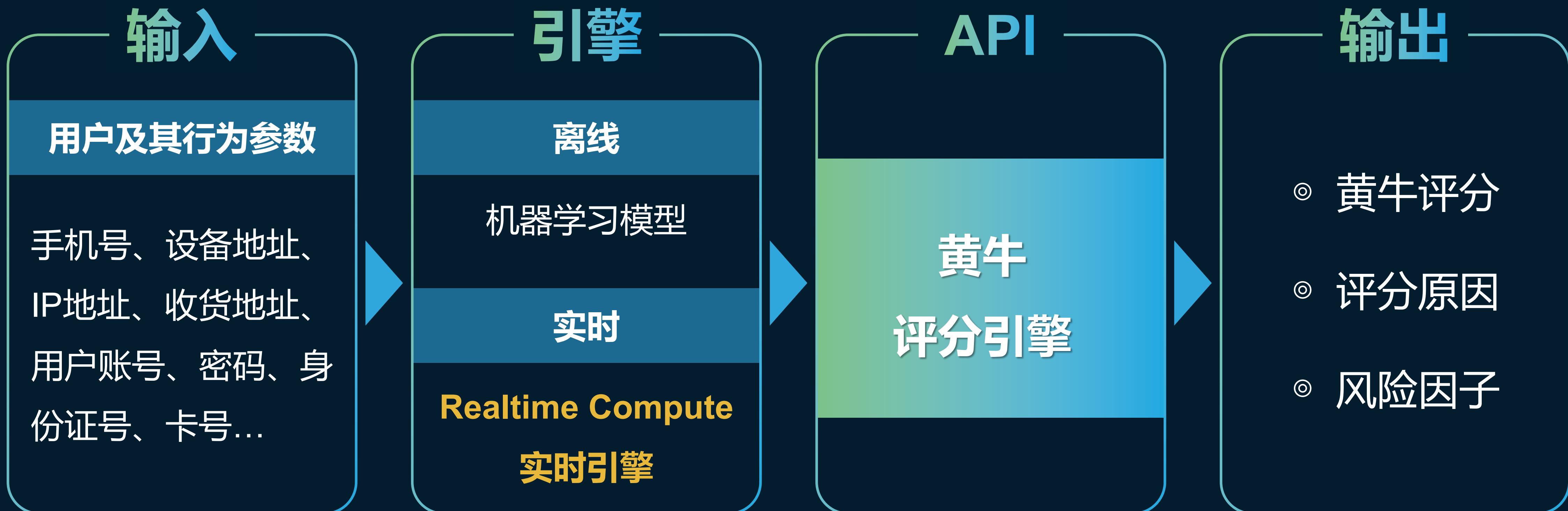
风险等级和评分

异常行为判定

风险原因

10 反黄牛API输出

黄牛风险评分是基于用户在注册、登录、浏览、收藏、下单、支付等流程的基本信息和历史行为，运用数据挖掘技术构建的评分引擎。通过对当前信息进行全面分析，给出量化的黄牛风险评分；并给出评分原因以及黄牛风险相对应的风险因子；最终API输出并嵌入到交易系统。



输入的参数越多，识别效果越准确

11 可视化实时监控数据架构



开发效率高

- ① 云上全部基于阿里云成熟产品
- ② 开发效率高

使用的云产品

DTS

- ① 实时数据采集,类比OGG、Otter

DataHub

- ① 实时数据订阅服务

MaxCompute

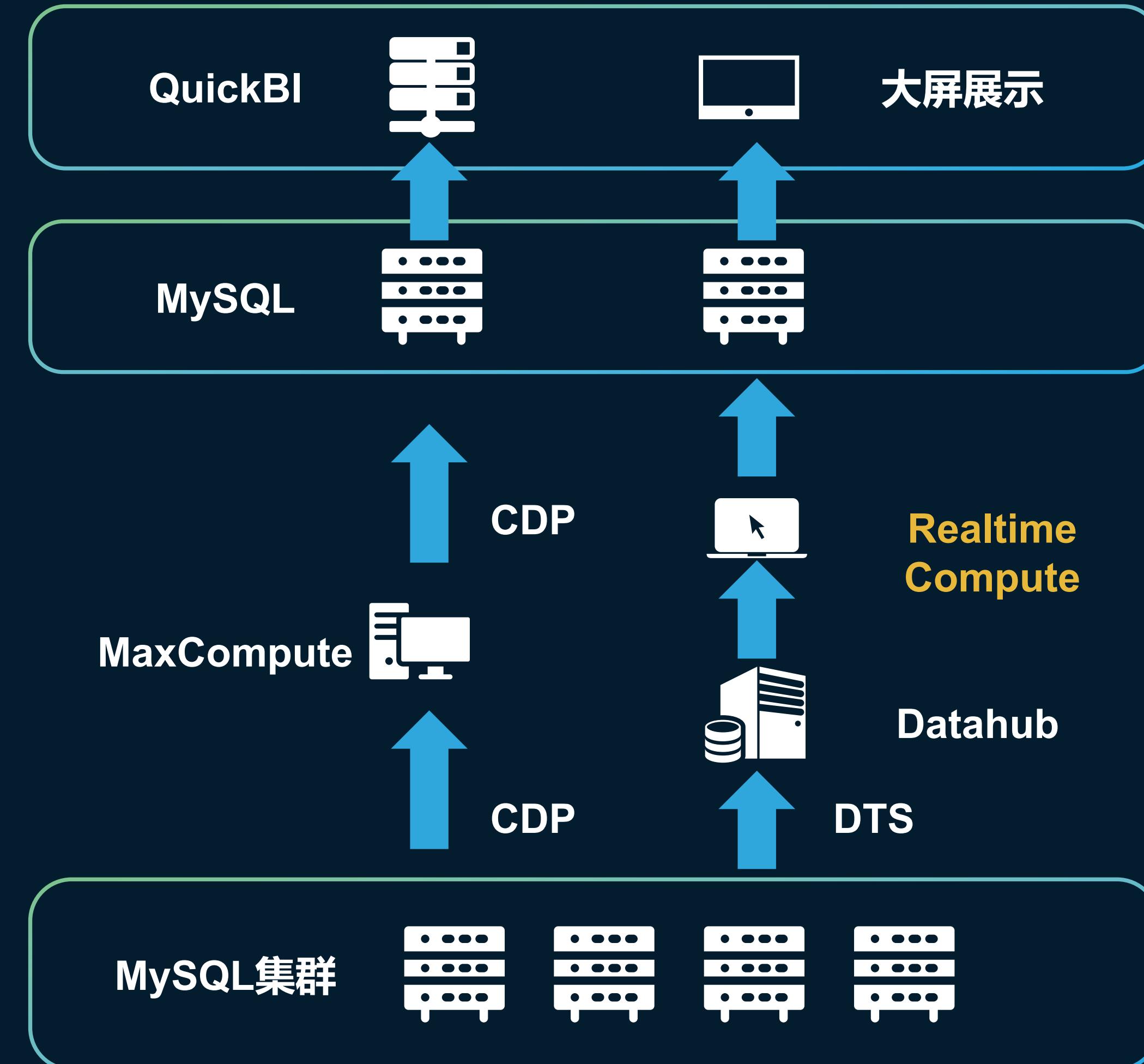
- ① 离线数据计算

Realtime Compute

- ① 实时数据计算

QuickBI+可视化大屏

- ① 输出业务分析



12 可视化实时监控大屏展示

持续可视化监控抢购下单数量频次和预警

监控购买用户区域分布以及与历史用户分布的差异

展示异常购买用户的地域分布变化，以及分析结果

监控不同销售渠道的黄牛数量和对比分析结果

多维度风险请求用户识别量



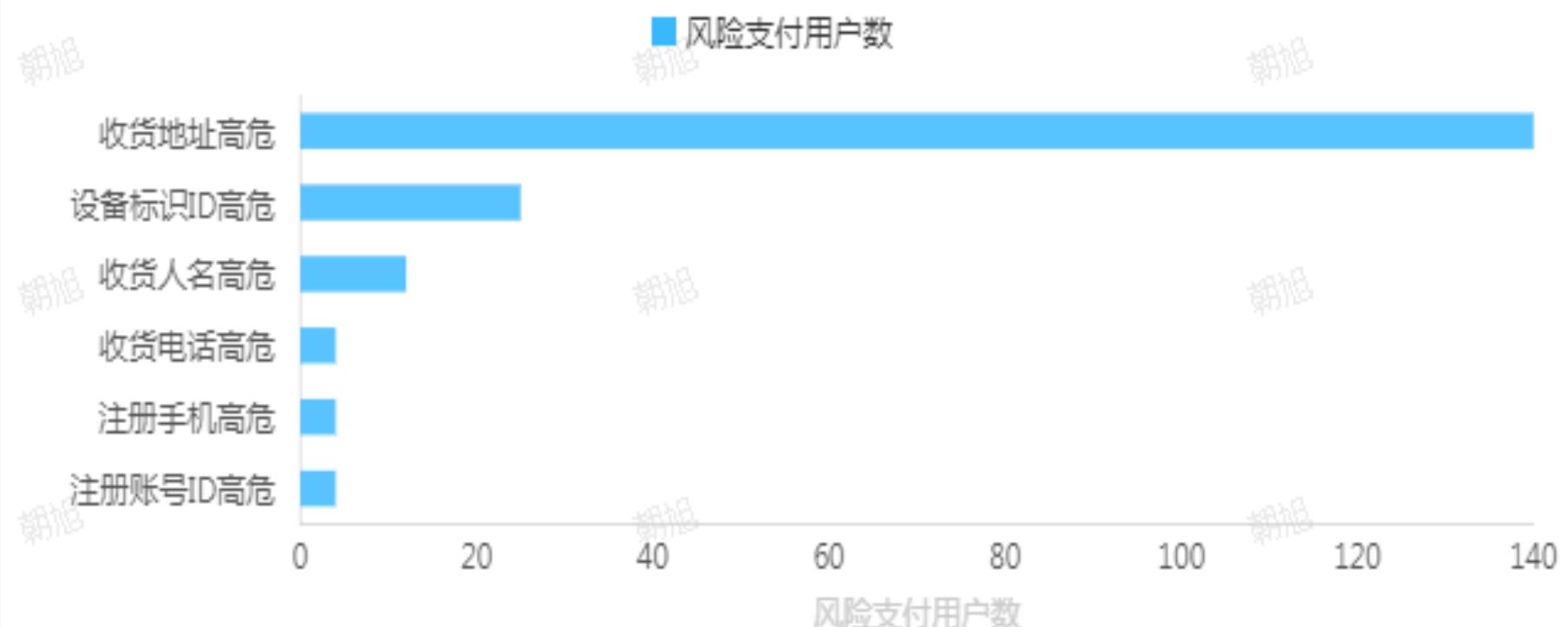
多维度风险支付用户识别量



唯一性请求风险排序



唯一性支付风险排序



13 一站式开发SaaS平台



一站式托管平台，涵盖实时数据开发全流程

作业列表 搜索作业名称...

作业运维 / ads_jcmh_real_flow_td | • 运行 | 启动位点：2018-12-13 10:08 | 业务延时：1.64 s | 暂停 | 停止 | 监控

作业名称	运行状态	运行信息	数据曲线	FailOver	CheckPoints	JobManager	TaskExecutor	血缘关系	属性参数
ads_jcmh_real_flow_td	• 运行	Task状态 创建 : 0 运行 : 5 失败 : 0 完成 : 0 调度 : 0 取消中 : 0 已取消 : 0	输入TPS 1.8 Block/s	输入RPS 41.37 条/s	输出RPS 0.23 条/s	输入BPS 53003.9 B/s	消耗CU 1.97 CU	启动时间 2018-12-13 11:11:56	运行时长 8 小时 49 分钟 8 秒
ads_llfx_real_page_ty...	• 运行								
ads_yymh_real_flow_td	• 运行								
ads_llfx_real_core_ind...	• 运行								
ads_llfx_area_pv_uv_c...	• 运行								
ads_llfx_total_flow_co...	• 运行								
ads_app_member_tot...	• 运行								
ads_jcmh_real_zy_ord...	• 运行								
ads_jcmh_real_rfd_flo...	• 运行								

Vertex拓扑

切换视图 | 列表模式 | 放大 | 缩小

The diagram illustrates a Flink vertex topology with four key groups. The first key group (ID 0) has 1 parallelism, TPS of 1.80, and a delay of 4,935.00 ms. The second key group (ID 1) has 1 parallelism, TPS of 0.40, and no delay. The third key group (ID 2) has 1 parallelism, TPS of 13.90, and no delay. The fourth key group (ID 3) has 1 parallelism, TPS of 0.40, and no delay. Arrows labeled 'FORWARD' indicate the flow of data from key group 0 to 1, 2, and 3, and from key group 1 to 3.

ID	资源健康分	PARALLEL	TPS	DELAY	IN_Q	OUT_Q
0	20 分	1	1.80	4,935.00 ms	0.00 %	0.00 %
1	100 分	1	0.40	0.00 ms	0.00 %	0.00 %
2	100 分	1	13.90	0.00 ms	0.00 %	0.00 %
3	100 分	1	0.40	0.00 ms	0.00 %	0.00 %

< 1 2 3 > ID Status InQ max OutQ max RecCnt sum SendCnt sum TPS sum Task

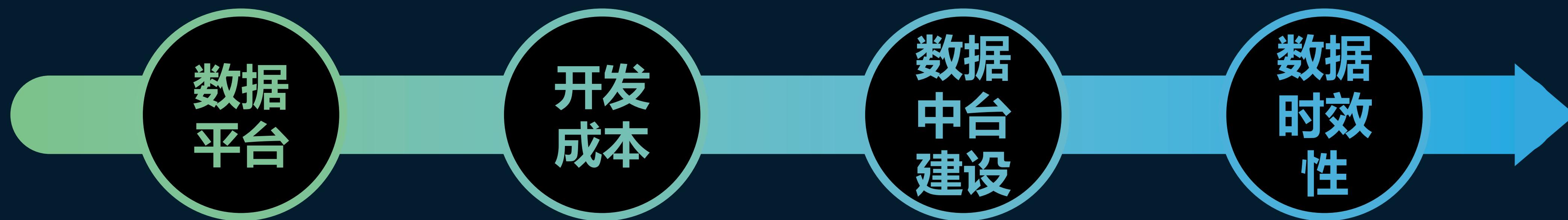
14 对企业产生的价值



采用阿里云一站式SaaS实时计算平台，提升企业运营和决策效率。

建立从数据上云，数据采集，数
据传输，数据计算，数据存储，
数据展现等全链路大数据平台

建立实时数据中台体系，
满足业务需求及迭代



- ◎ SQL开发
- ◎ 开箱即用
- ◎ 可视化运维
- ◎ 开发周期短

数据链路秒级延迟，提升
数据获取，数据决策的时
效性





关注“袋鼠云技术团队”

THANKS

回复“Flink”下载课件

