

QTEST

广告算法模型测试探索

移动展示广告实践

From: 廖淑波

CONTENTS

360技术嘉年华

01 *Part One*
背景简介

02 *Part Two*
测试方案

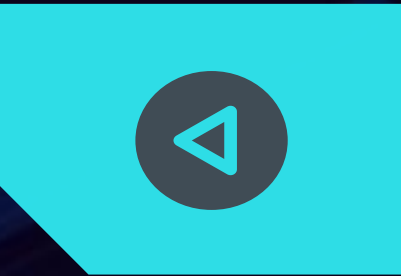
03 *Part Three*
目前效果

04 *Part Four*
未来展望

01

Part One

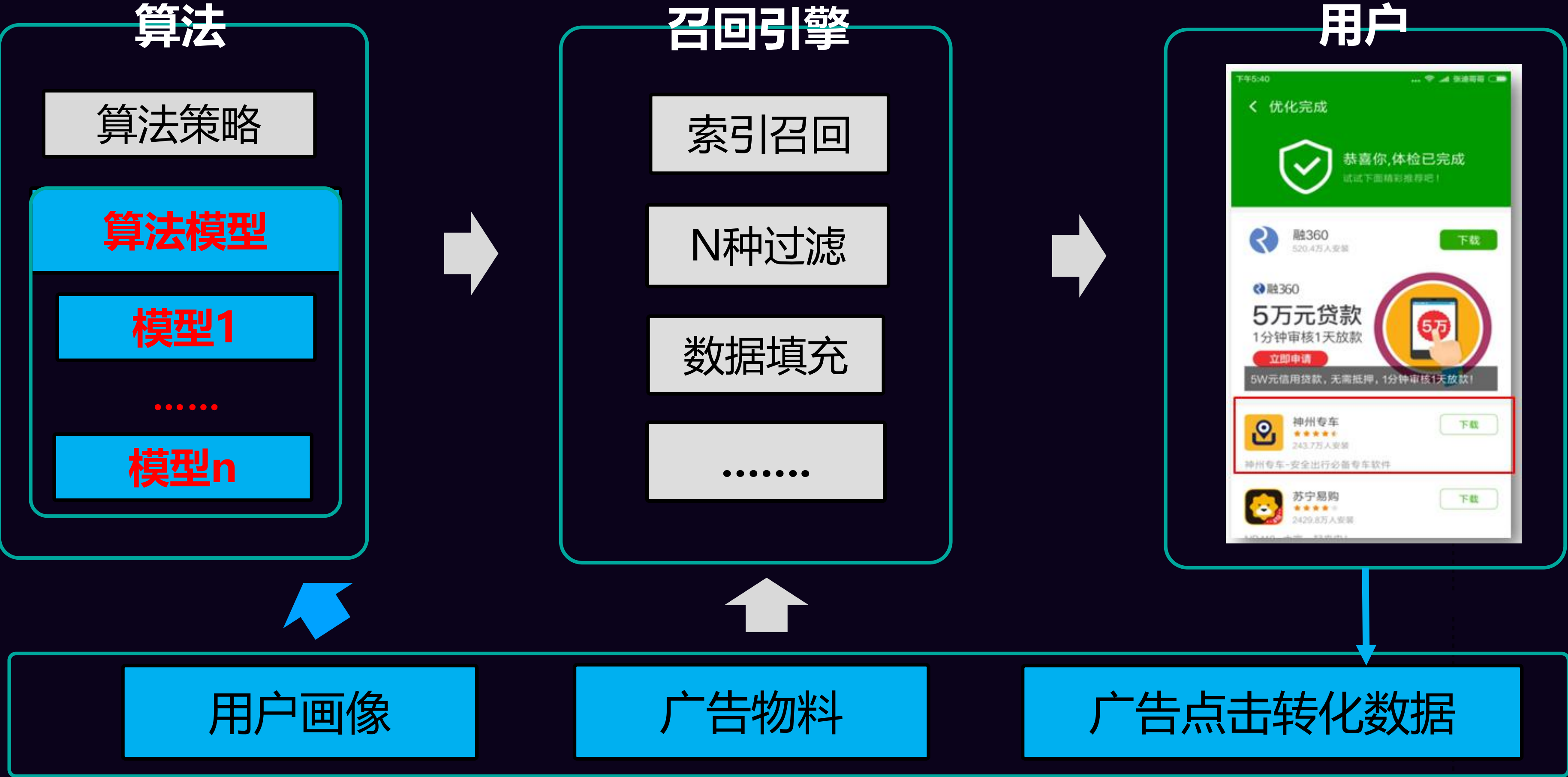
背景简介



广告召回流程&业务问题



！！业务压力，模型项目和迭代增加，各种问题暴露！！





算法模型基础概念

360技术嘉年华

- 模型？
- 样本？
- 算法？
- 特征？
- 训练？

1,4,16... () ...256...?

机器学习

算法分类：

分类算法: 决策树、贝叶斯、神经网络...

聚类算法: K-means、K-mediods...

模型分类：

分类模型、回归模型、预测模型、 聚类模型



测试问题 & 挑战

360技术嘉年华

超大规模样本

超大规模特征

数据构造难

预估值难解释

黑盒性

难解释

可借鉴少

数据复杂

测试极难

机制复杂

多不完整

离线训练

实时预估

问题定位难

人工干预成本高

不同的模型区别

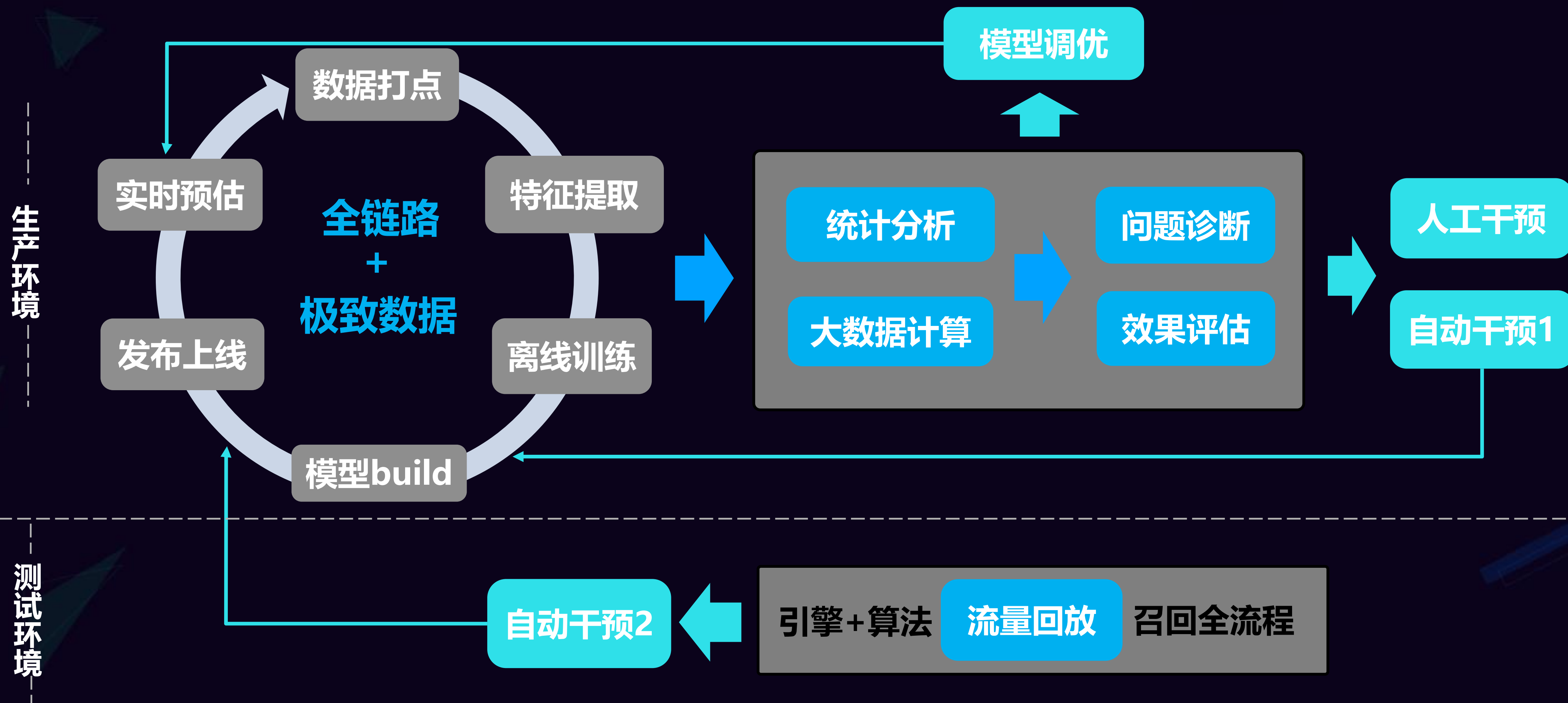
数据维度不统一

02 *Part two* 测试方案



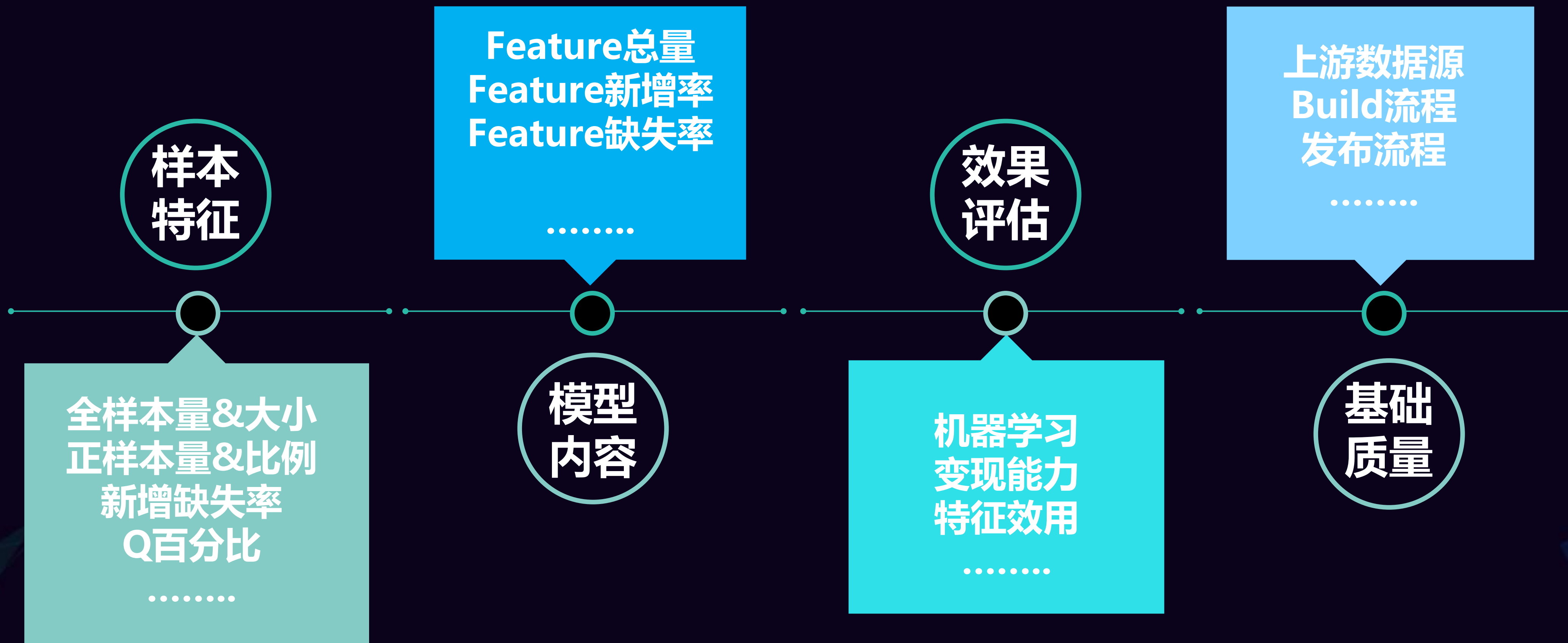
整体测试方案

360技术嘉年华





数据监控分类



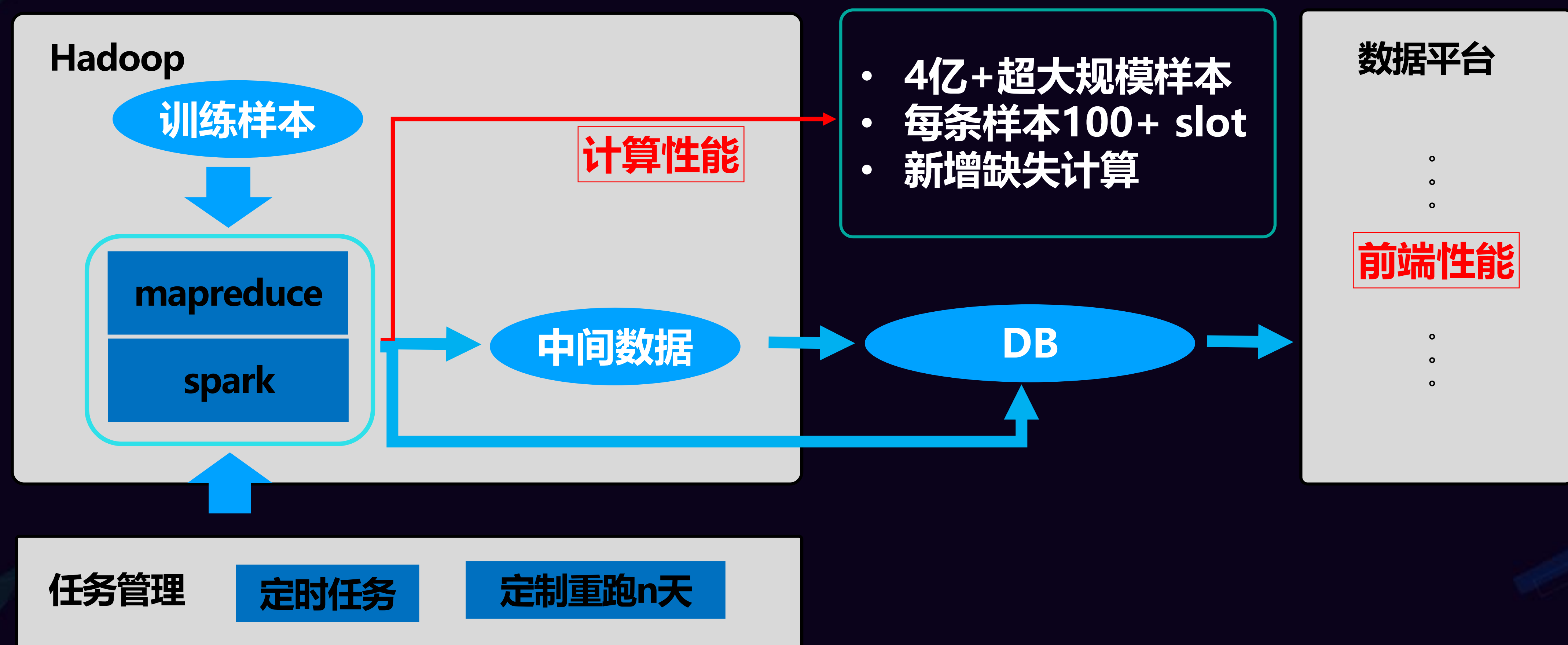


样本特征&模型内容



- 天/小时级
- 单/多日
- 同环比
- 分版本对比
- 特征项可筛选
- 精度可筛选
- 对比历史天数可筛选
- 表格详情
- 特征值变化趋势

样本特征&模型内容





样本特征&模型内容

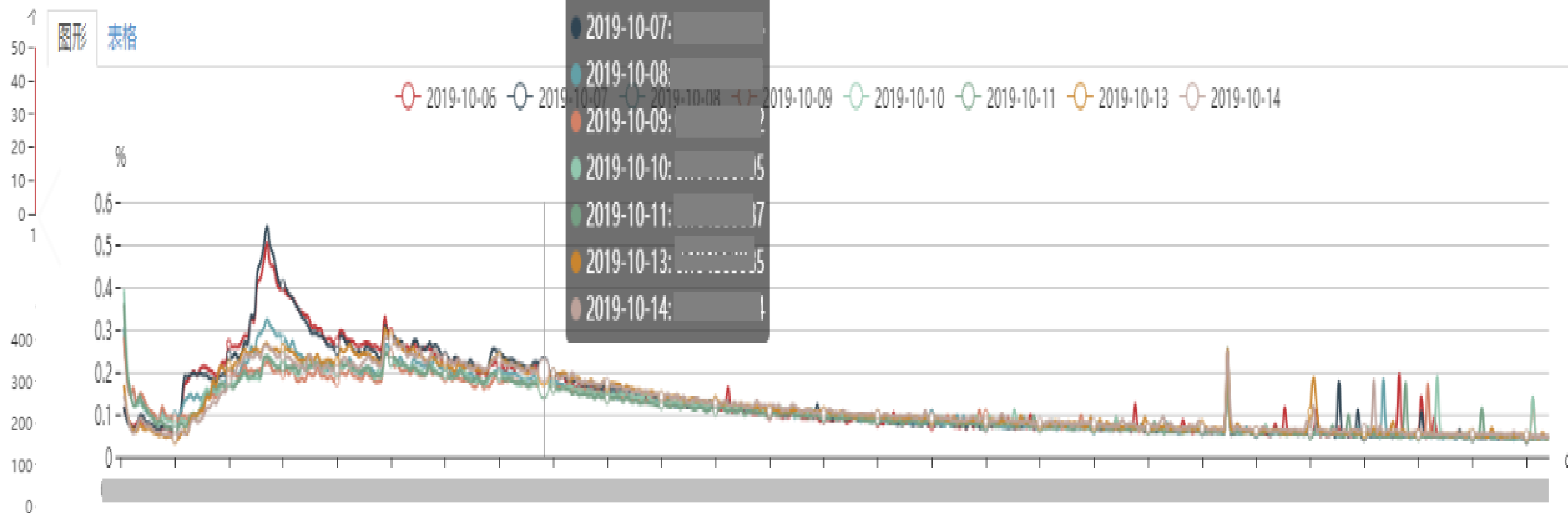
360技术嘉年华

特征值统计ctr/cvr(每个特征项下取top50) [版本v2.0] ?

ctr/cvr范围 0 ~ 100 % 日期 2019-10-14 查询

图形 表 q分布百分比[版本v4.0] ?

精度节点范围 0.0001 ~ 0.045 查询





效果评估

机器学习能力

- AUC
- COPC
- LOGLOGSS
- PCTR
- 准确率
- 召回率
- ROC曲线
-

变现能力

- 收入
- CTR
- CVR
- CPM
-

特征效用

- IV
- WOE
- 特征变现变化
-

- 不同阶段对比
- 阶段一致性
- 时间小时/天
- 版本对比
- 分桶
- 分adx

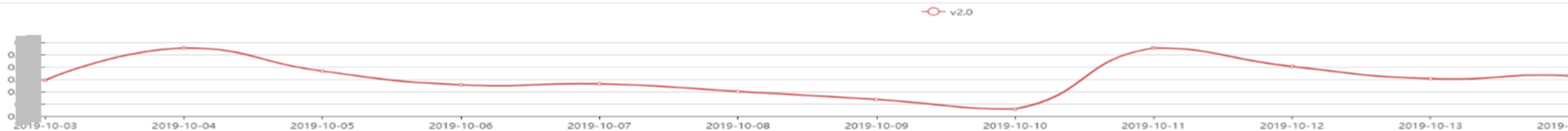


效果评估

360技术嘉年华

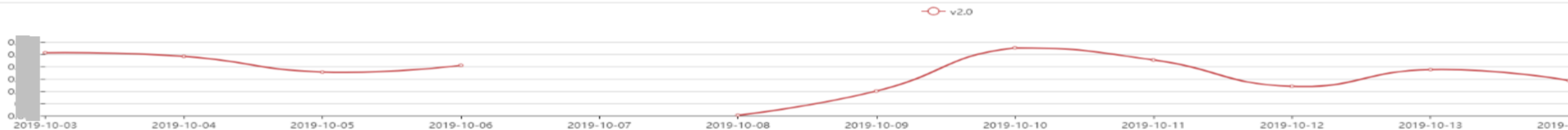
AUC_train

图形 表格



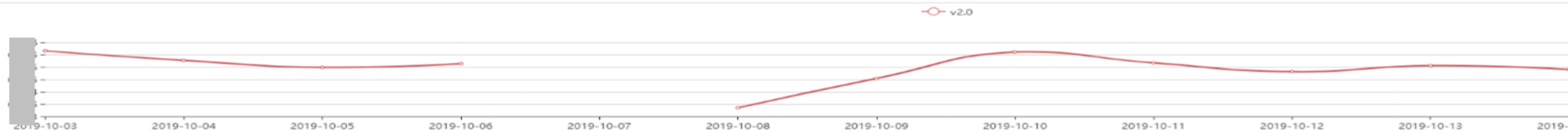
AUC_online

图形 表格



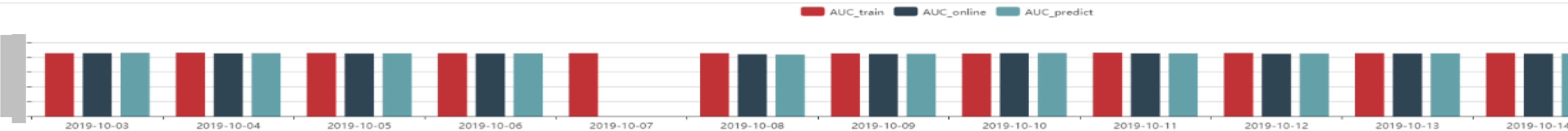
AUC_predict

图形 表格



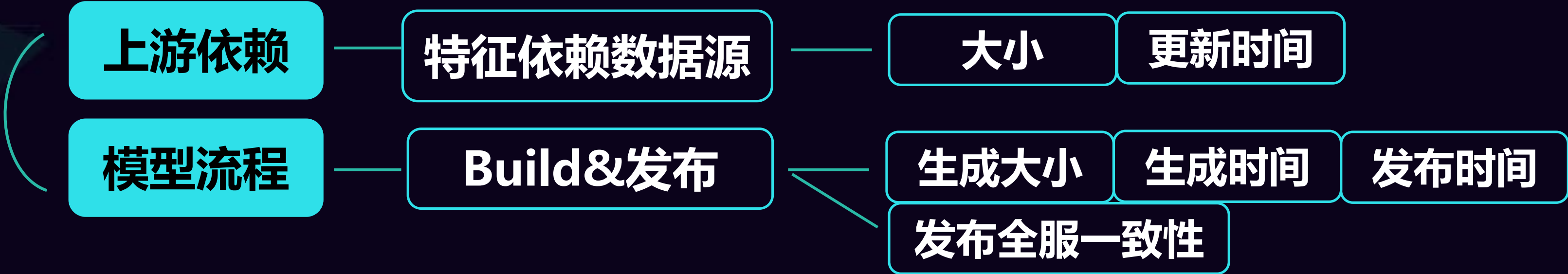
AUC一致性(版本: v2.0)

图形 表格





基础质量



- 动态阈值
- 追踪分析工具

模型-线上全服一致性?

图形

模型-线上全服一致性?

图形

表格

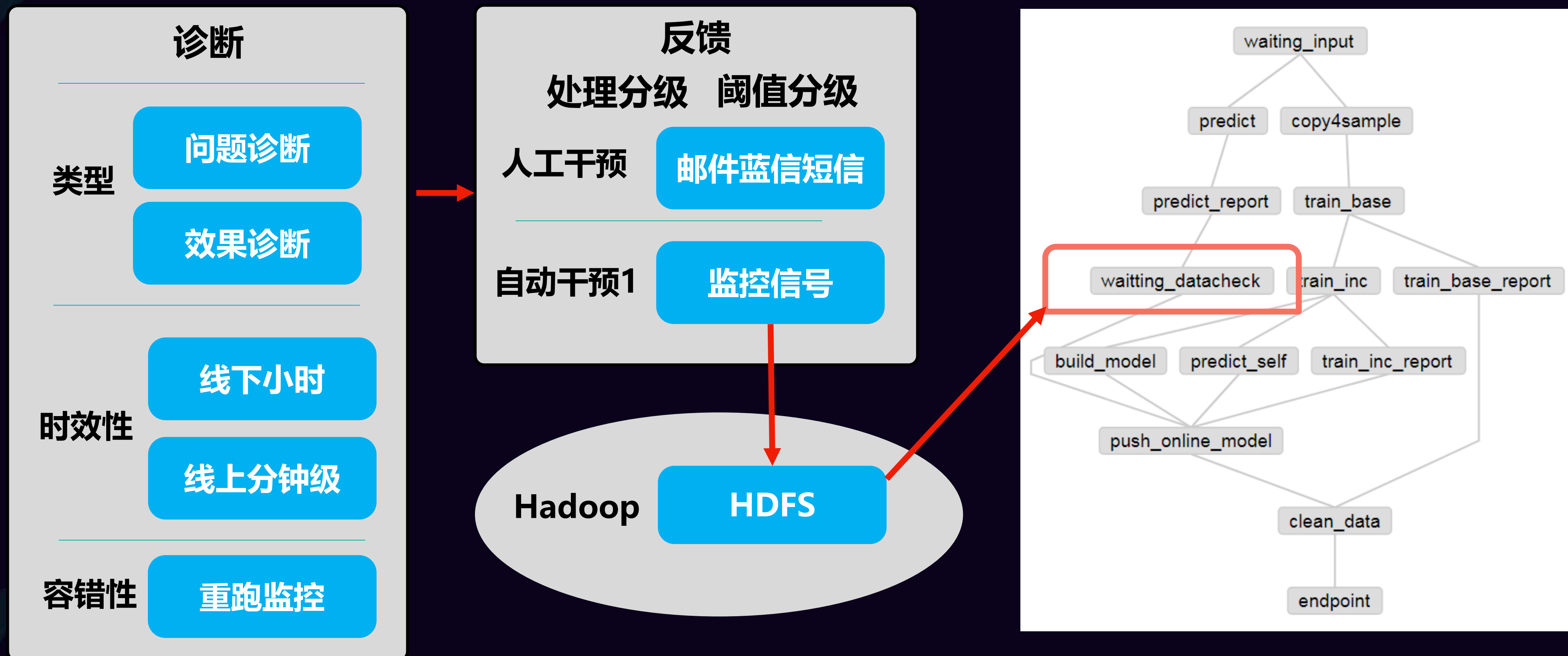
查询

不-

不一致时间点	不一致情况
2019-10-16 12:00	6e274aa1422a7d3292ce6de6db118b71(共 1 台机器): 192.168.1.1.qihoo.net, 192.168.1.2.qihoo.net, 192.168.1.3.qihoo.net, ... 7e0f9576118260f9007b04f3f8fa53ae(共 1 台机器): 192.168.1.4.qihoo.net, 192.168.1.5.qihoo.net, 192.168.1.6.qihoo.net, ...



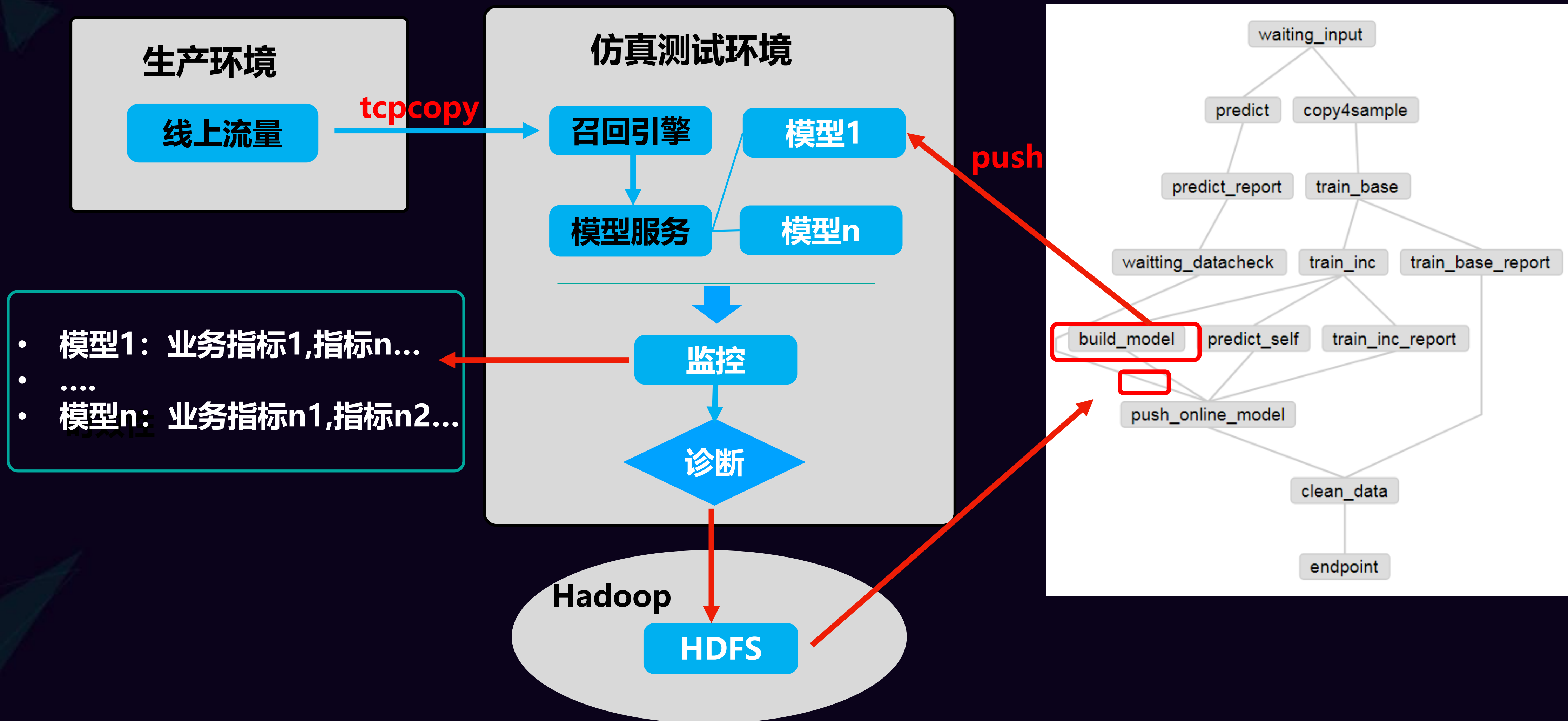
诊断&干预1





流量回放&干预2-ING

360技术嘉年华





平台化

360技术嘉年华



查询

场景自定义查询
配套复杂表格
线上部署查询



对比

同环比
不同模型对比
不同版本对比



配置

基础属性配置
特征项管理
报警阈值配置
监控信号阈值配置



用户体验

前端性能
行为监测



平台化

360技术嘉年华

🏠 主页

我的位置: MODEL / 模型配置 / 编辑模型

MODEL

🏠 主页

我的位置: MODEL / 模型配置 / slot_id管理

MODEL

📁 模型配置

▶ 模型列表管理

▶ slot_id管理

📈 模型监控

▶ 基础数据

▶ 样本分析

▶ 特征分析

▶ 模型分析

▶ 效果评估

▶ hdfs相关

📊 词典监控

模型:

精排-CTR-标准

版本:

v2.0

请输入该版本下新增特征项

新增

注: 若需添加多个特征项时, 下同

☒100 ☒101 ☒102 ☒103 ☒104 ☒105 ☒106

☒200 ☒201 ☒202 ☒203 ☒204 ☒205 ☒206 ☒251 ☒252 ☒253 ☒254 ☒255 ☒256 ☐

☒2001 ☒2002 ☒2003 ☒2004 ☒2005 ☒2006 ☒2007 ☒2008 ☒2009 ☒2010 ☒2011

☒300 ☒301 ☒302 ☒303 ☒304 ☒305 ☒307 ☒308 ☒309 ☒310

☒400 ☒401 ☒402 ☒403

☒500 ☒501 ☒502 ☒503 ☒504 ☒505 ☒506

☒600 ☒601 ☒602

☒800 ☒801 ☒802 ☒803 ☒804 ☒805

☒900 ☒901 ☒902 ☒903 ☒904 ☒905 ☒906 ☒907 ☒908 ☒909 ☒910 ☒911 ☒912 ☒

修改

today | yesterday | beforeyesterday

上一步

修改

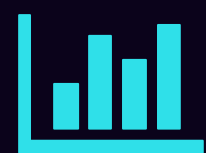
取消

03

Part two
目前效果



效果



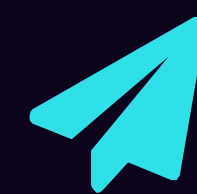
数据规范

统一模型流程数据规范
排查问题效率明显提高



报警全面

全流程, >6种问题



监控信号

已有效拦截上线多次
挽回收入损失/信任度



模型稳定性

Pipeline稳定性提高
相关事故降为0



新需求不断

每次发版都有新需求
正在新接入2个业务



模型调优

Rd基于平台进行调优
使用频率较高

04

Part two
未来展望



未来展望

360技术嘉年华



全面的健康报告

提供更全面的智能综合分析报告，指导快速发现问题



模型调优

调优跟踪与平台化、效率化，直导正向收入



交互式评测

通过测试样本，定制的流量回放等进行离线评估