

网易云音乐数据研发模式 DataOps落地实践

网易云音乐/宋东辉



网易云音乐/宋东辉

资深测试开发工程师

浙江大学硕士毕业之后，加入支付宝，参与蚂蚁端稳定性 Sapienz 项目建设，数字内容治理黑鸟项目核心成员，双12大促主要质量负责人之一。

2021年加入网易云音乐，云音乐大数据质量建设和数据治理负责人，发起并牵头多方共建 DataOps 项目，提升了数据研发效率和质量水位。

目录

Contents

1、背景简介

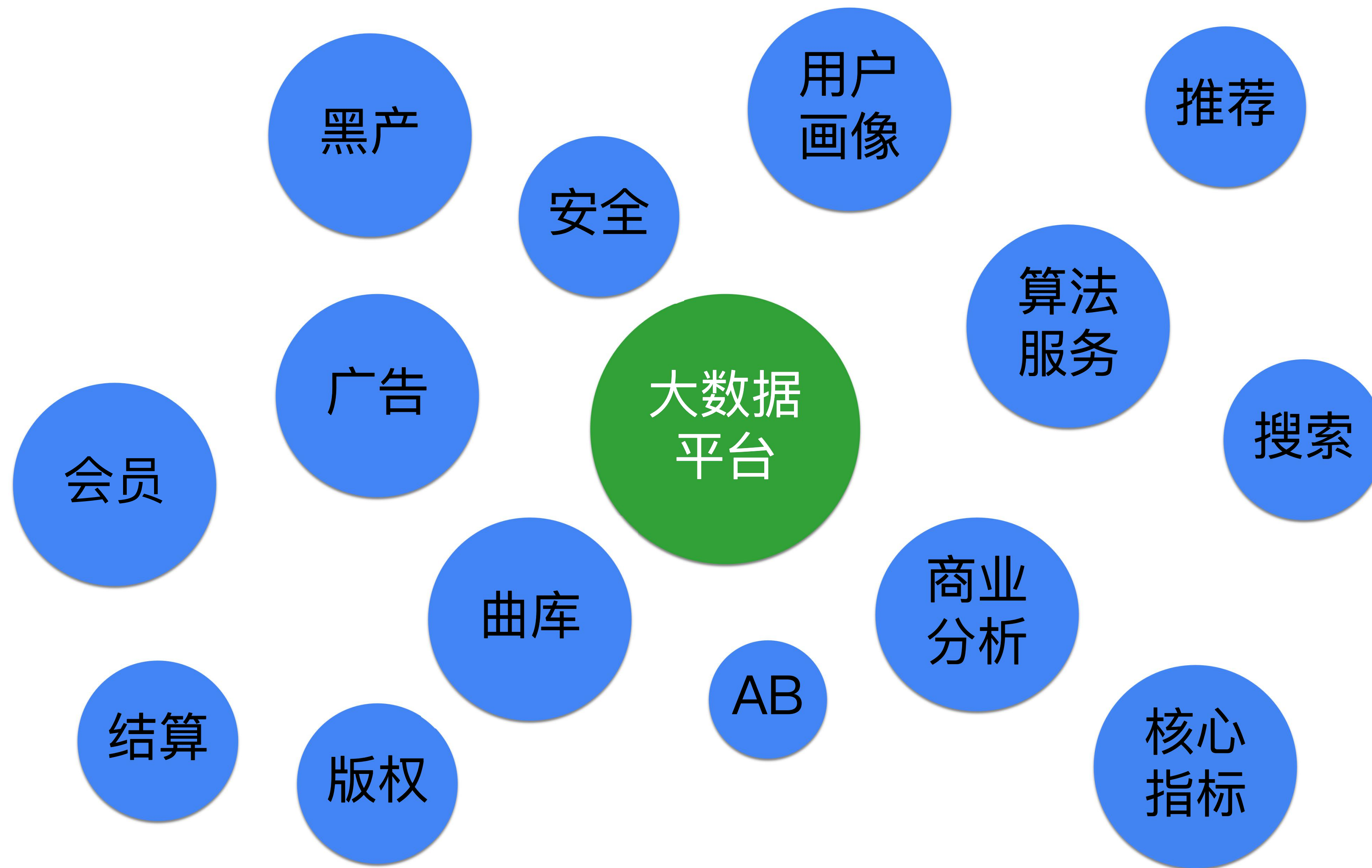
2、方案介绍

3、落地效果

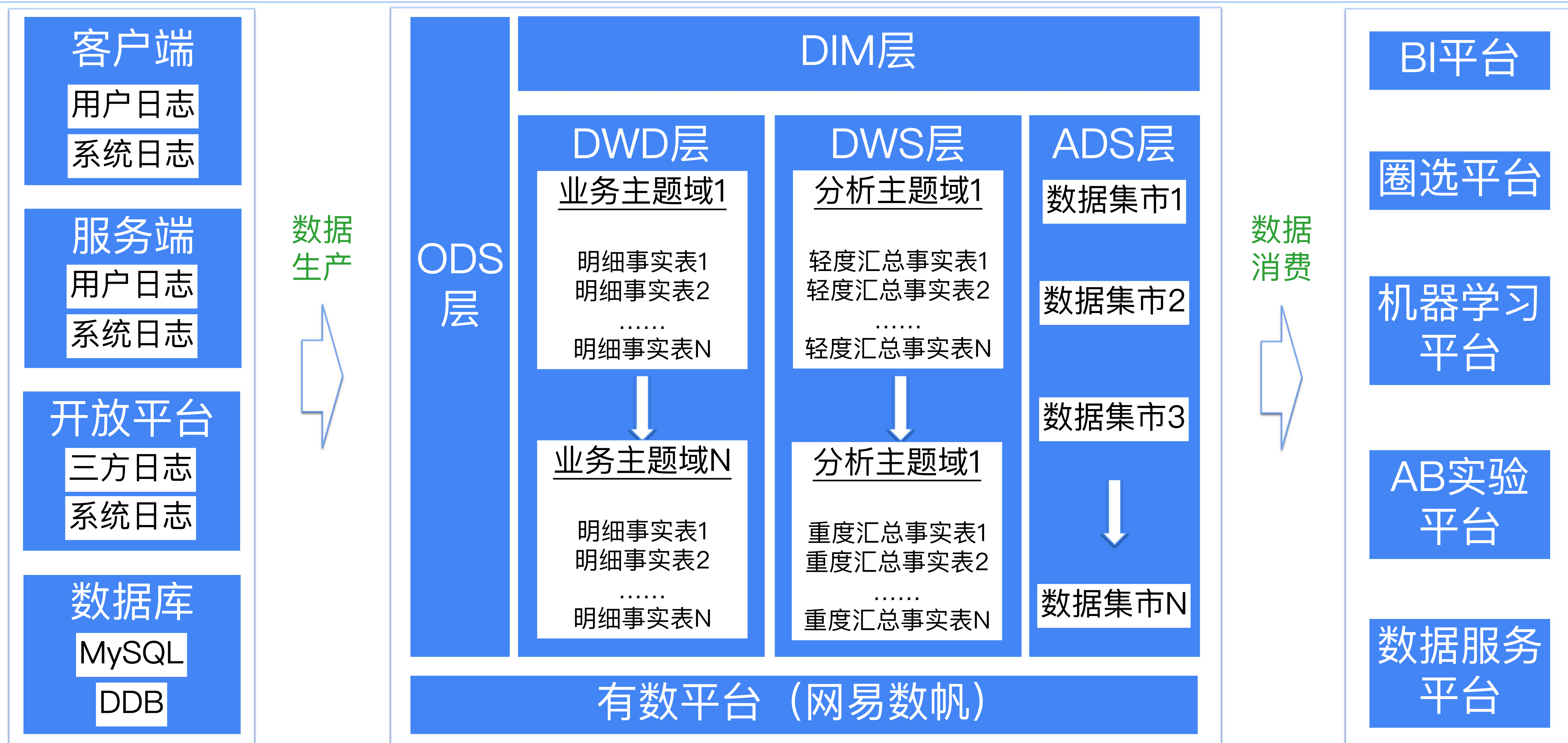
4、未来展望

1、背景简介

大数据的重要性

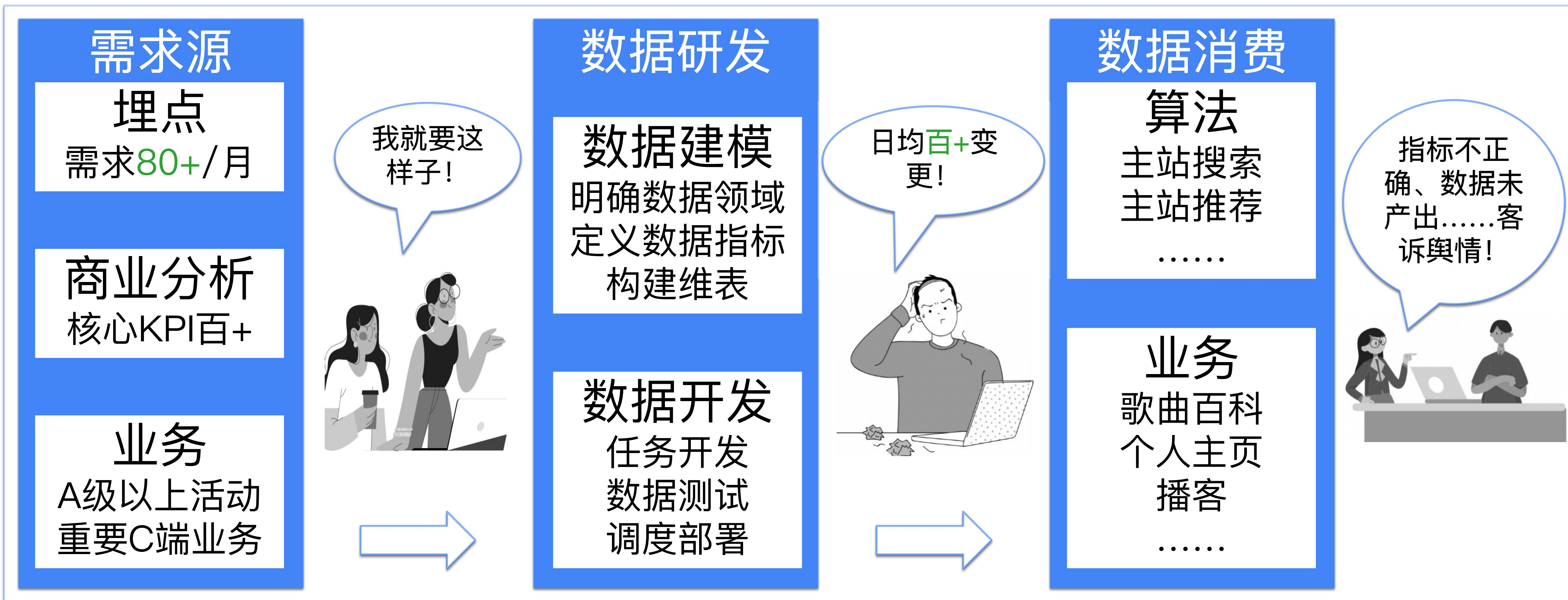


大数据的框架



大数据的痛点

现状



痛点

业务感知

- 数据线上问题平均10+个/月
- 线上问题平均解决时长高达44h
- 需求延期交付比高达50%

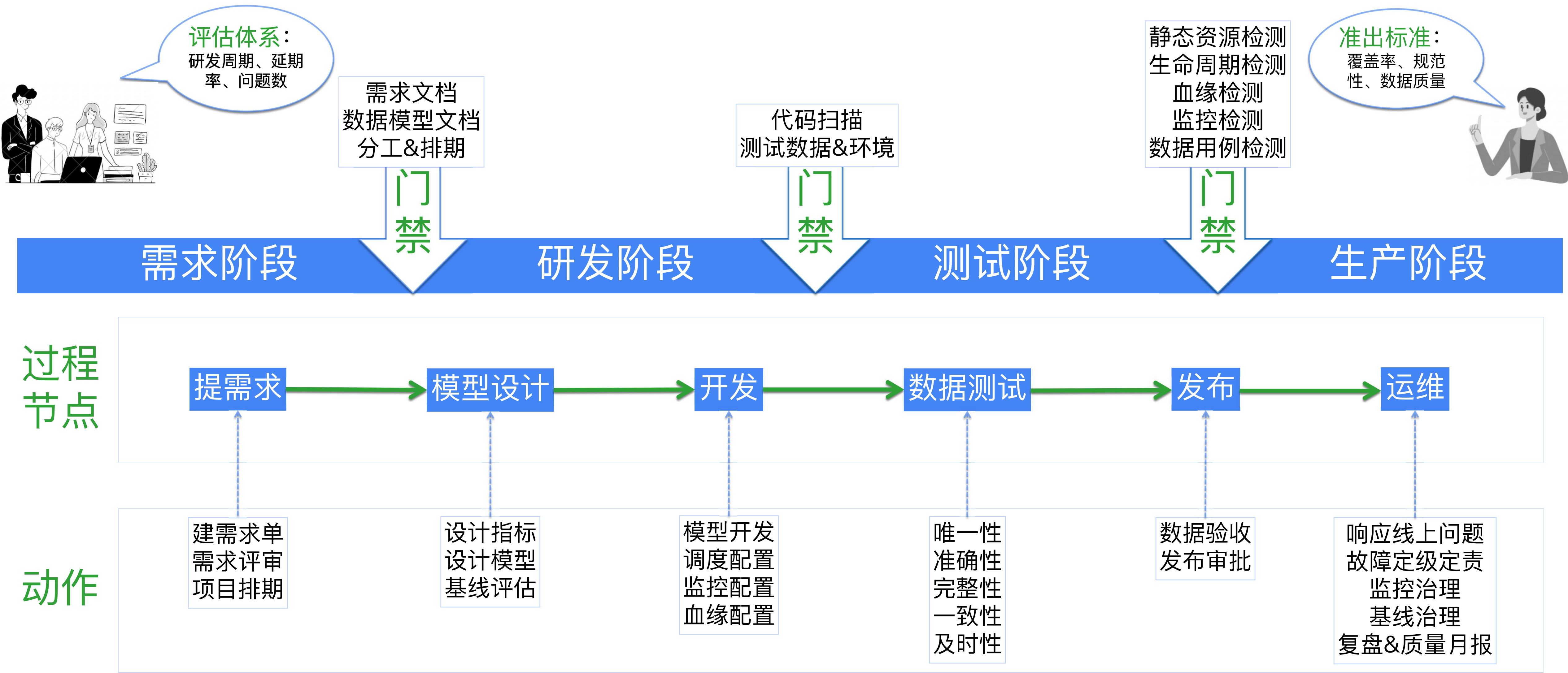
研发感知

- 回刷数据计算成本单次万元级别
- 研发自测效率低, 准出标准不清晰
- 线上问题感知滞后

2、方案介绍

2.1、数据研发标准

数据研发标准



2.2、平台化

平台化DataOps

整体框架

卡点能力

事后巡检

研发效能



目标：协调多方高效合作，
辅助研发敏捷开发！

效能
大盘



研发周期
长！



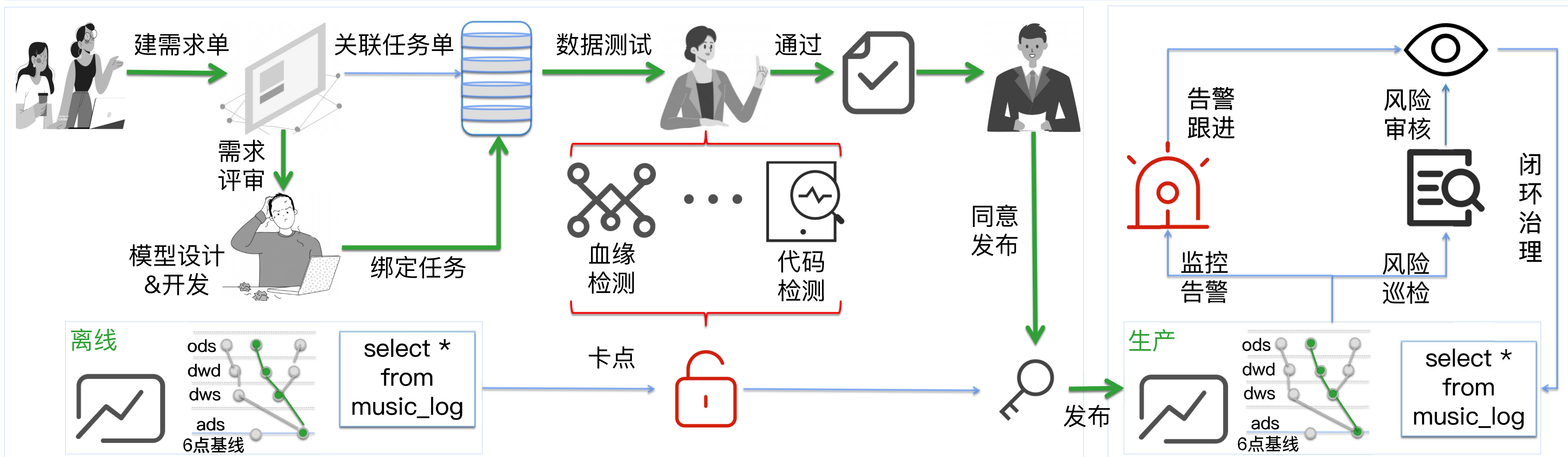
研发质量
有提升！



线上问题
增多！

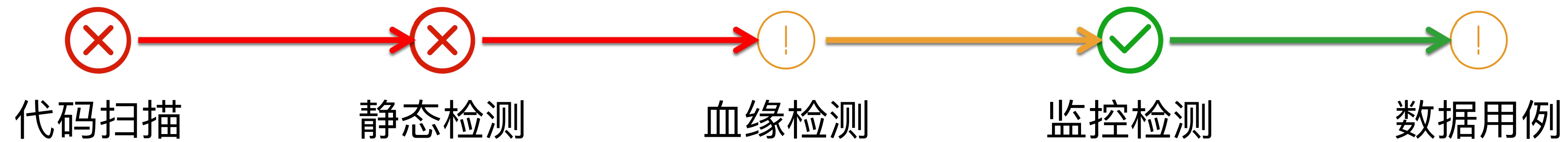


人力成本
增加！



卡点自动化

pipeline



维度

代码规范
代码质量
代码性能
变更点提示

调度配置
血缘配置
基线监控配置
表&分区配置
表&任务规范

血缘关系
链路基线风险
依赖风险
生命周期

监控覆盖率
监控误报率
告警跟进率

链路覆盖率
用例成功率
数据异常

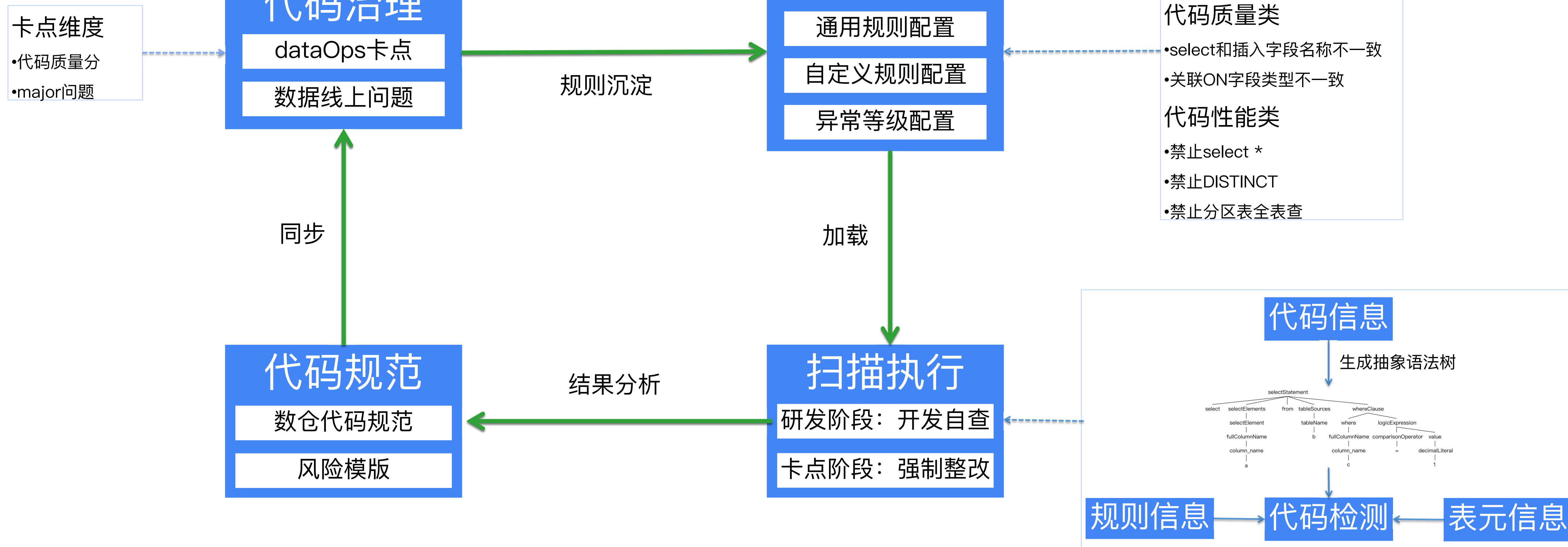
卡点自动化-代码扫描

整体框架

卡点能力

事后巡检

研发效能



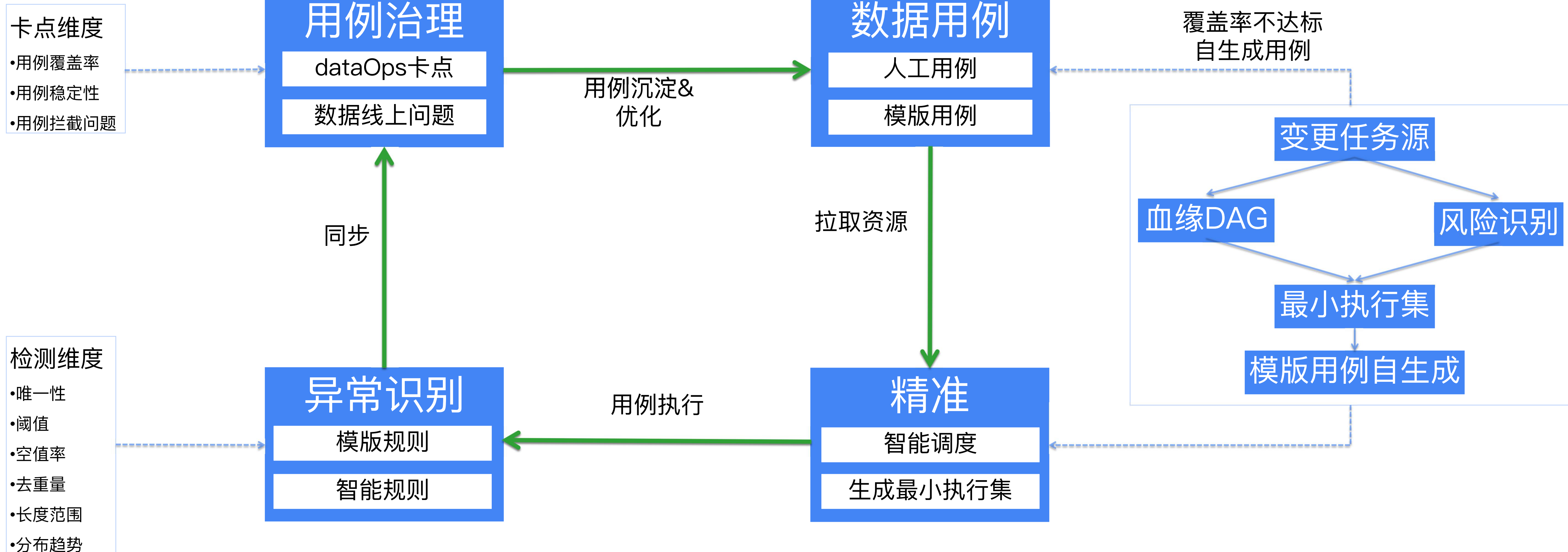
卡点自动化—数据测试用例

整体框架

卡点能力

事后巡检

研发效能



卡点自动化—数据监控治理

整体框架

卡点能力

事后巡检

研发效能

卡点维度

- 核心指标覆盖率
- 监控误报次数
- 监控周跟进率

监控治理

dataOps卡点

数据线上问题

指标梳理

核心域的表指标

核心KPI指标

C端业务依赖指标

内容

版权

用户

指标体系

直播

会员

社区

同步

监控优化

覆盖率治理

平台化

监控平台

告警通知

告警跟进流程

告警质量大盘&月报

监控配置

模版规则

自定义规则

字段类

- 完整性校验
- 有效性校验
- 唯一性
- 范围校验

跨字段类

- 关联性校验

跨表类

- 一致性校验
- 总量校验

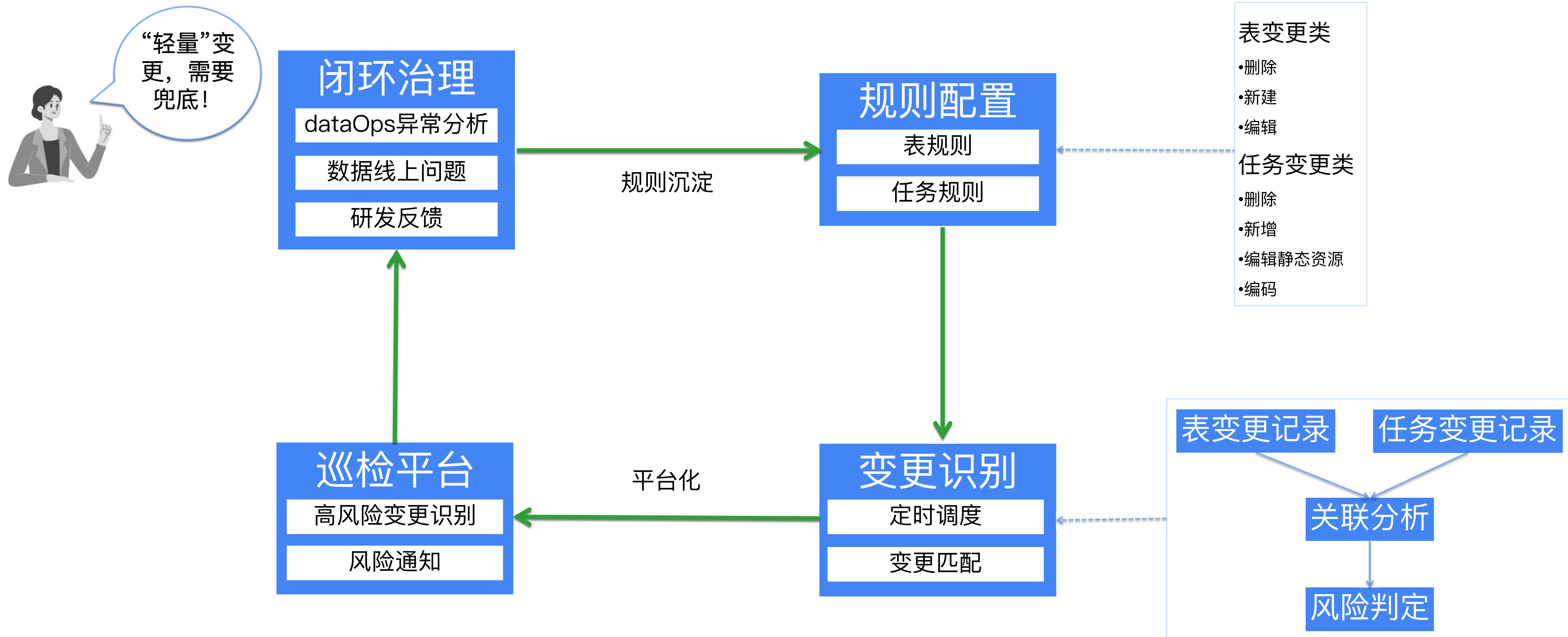
变更风险巡检

整体框架

卡点能力

事后巡检

研发效能



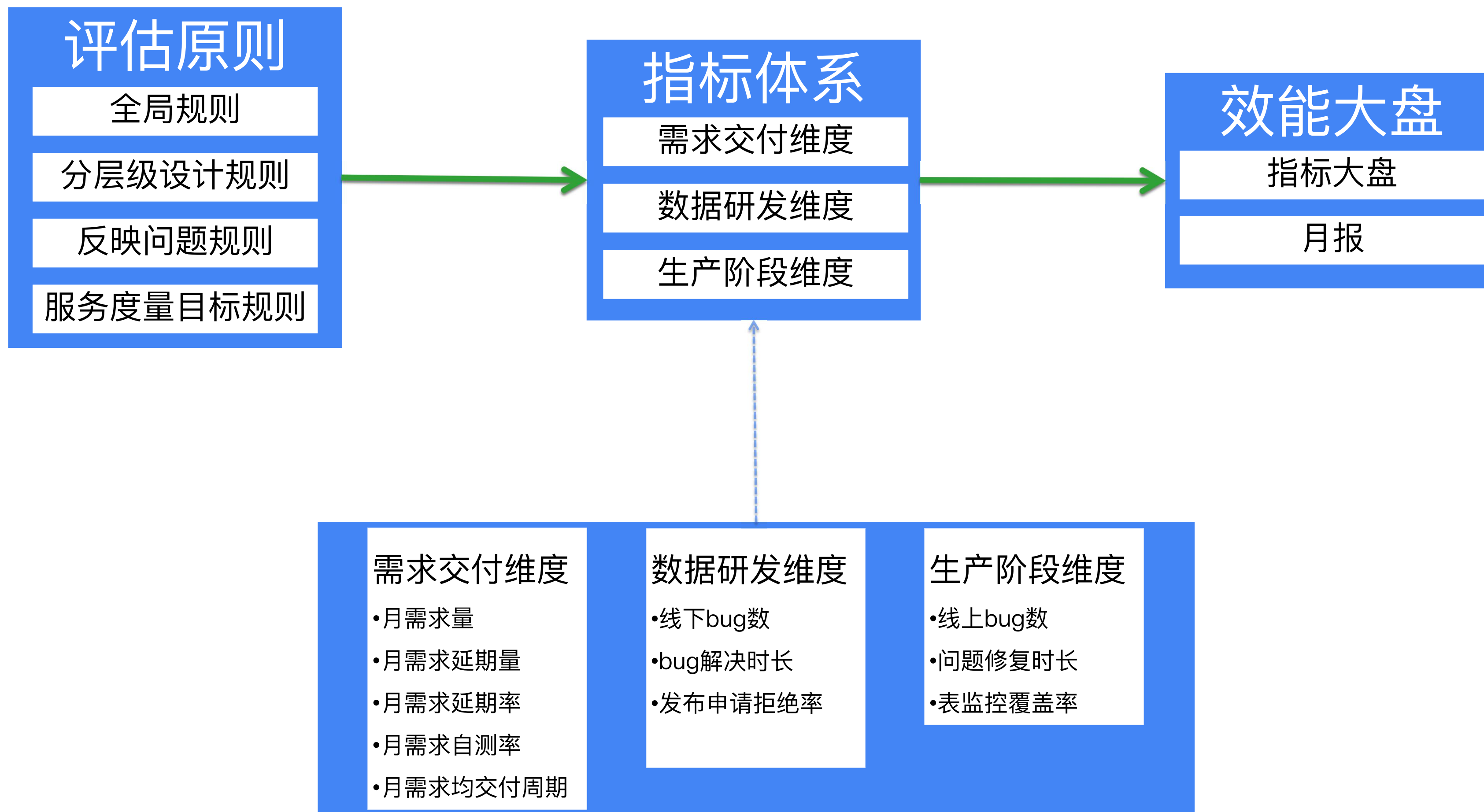
研发效能大盘

整体框架

卡点能力

事后巡检

研发效能



3、落地效果

落地效果

研发效率

- 需求平均研发周期从39d->33d ↓
- 需求交付逾期率降低到30% ↓

研发质量

- 数据线上问题降低37% ↓
- 研发规范原因引起的线上问题量降低50% ↓
- 线上问题平均解决时长从44h->9h ↓

数据监控

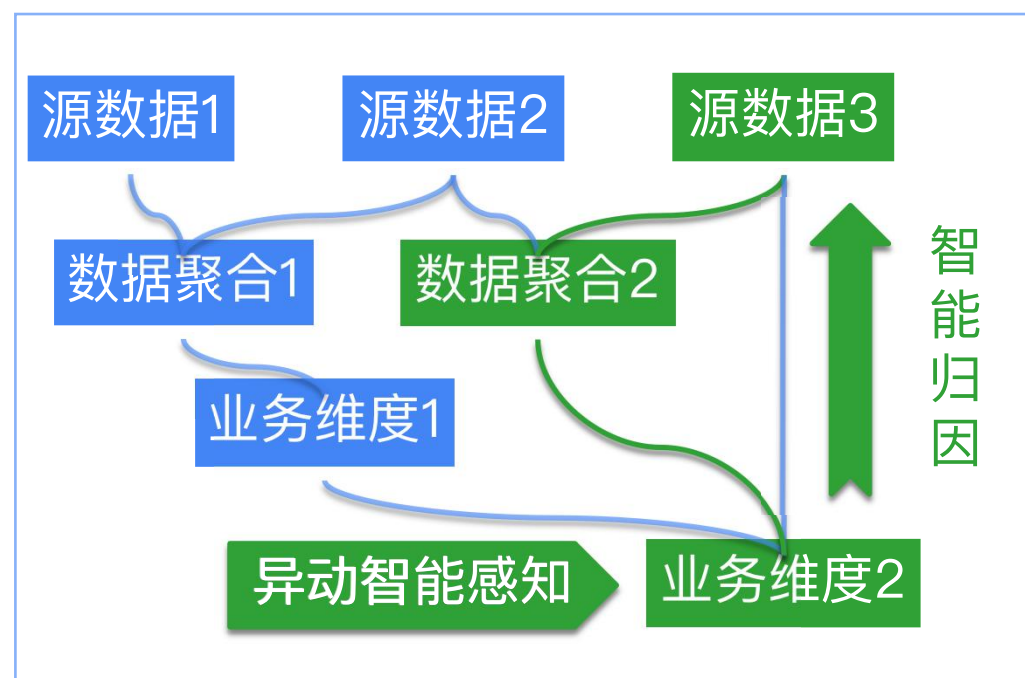
- 新增监控规则600+
- 整体监控月度误报量下降30% ↓
- 拦截有效数据异常32例
- 线上问题召回率约70%

业务意识

- 业务质量意识提升
- 研发规范意识提升
- 服务态度改善

4、未来展望

未来展望



智能监控

难点

- 提升覆盖率工作量大
- 人工规则监控误报量大
- 业务变动频繁
- 告警定位难

思路

- 构建智能监控模型
- 构建告警指标级归因

oneCode

难点

- 业务逻辑复杂
- 人工code效率低

思路

- 表&指标规范
- ChatGPT的L2级辅助编码

ChatGPT

宋

生成满足如下条件的SQL代码：
从音乐维度表中选择包括歌曲ID、专辑艺术家ID和专辑艺术家名称的数据，其中选择条件包括日期为2023年3月31日、在线状态不为-1或7、类型标签不为'宗教'、'有声音频'、'实验'、'其他'或'待运营确认'，并且类型标签不为空。

宋

```
SELECT
    song_id,
    album_artist_id,
    album_artist_name
FROM
    music_dimension.dim_itm_pgc_song_base_dd
WHERE
    dt = '2023-03-31'
    AND online_status NOT IN ('-1','7')
    AND genre_tag_name NOT IN ('宗教','有声音频','实验','其他','待运营确认')
    AND genre_tag_name IS NOT NULL;
```

致谢

致谢



想一想，我该如何把这些 技术应用在工作实践中？

THANKS

