

金融场景中的指标体系建设与应用

-韩秀锋-CTO



目录 CONTENT

01 金融行业指标现状

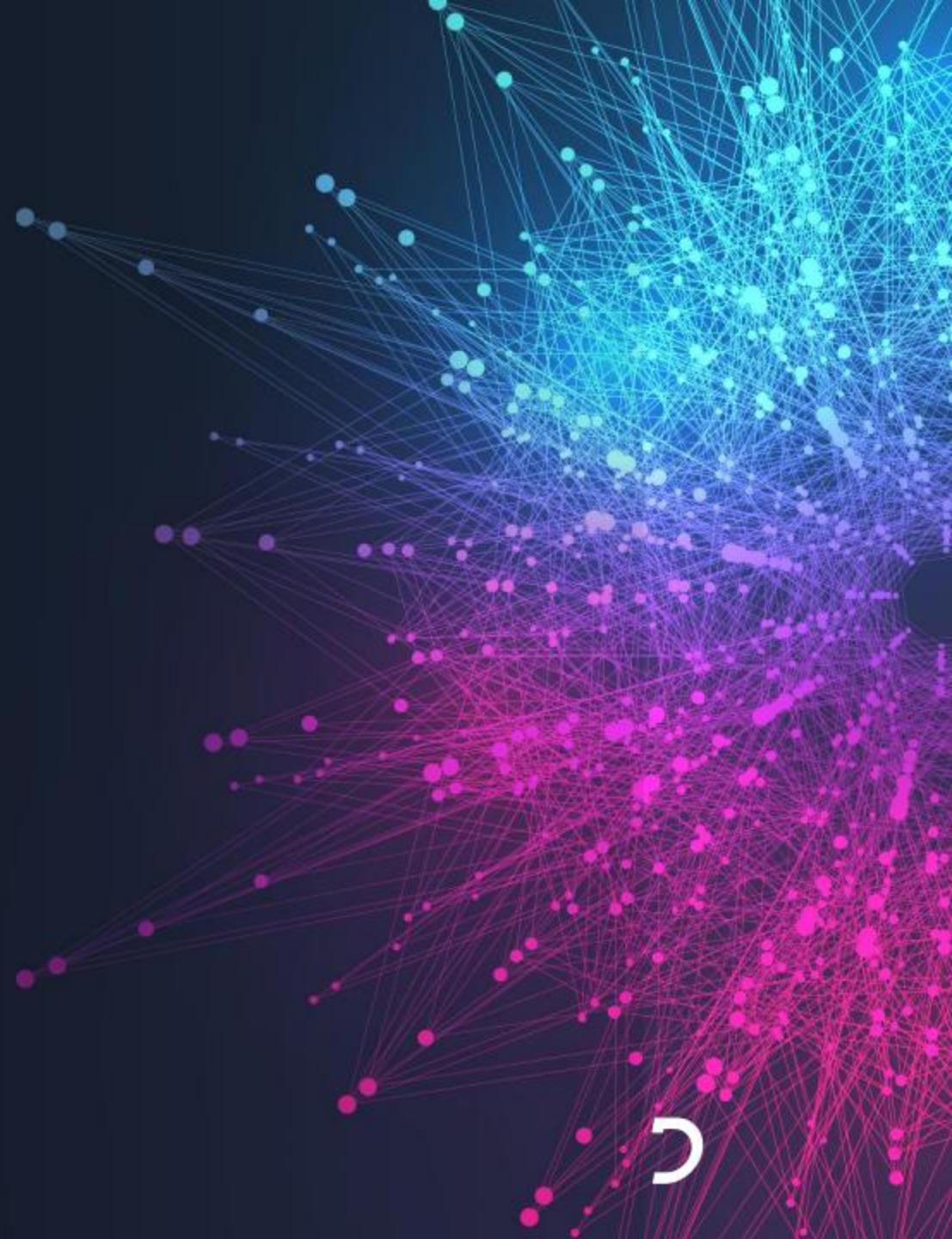
03 解法

02 痛点 -> 技术问题

04 案例

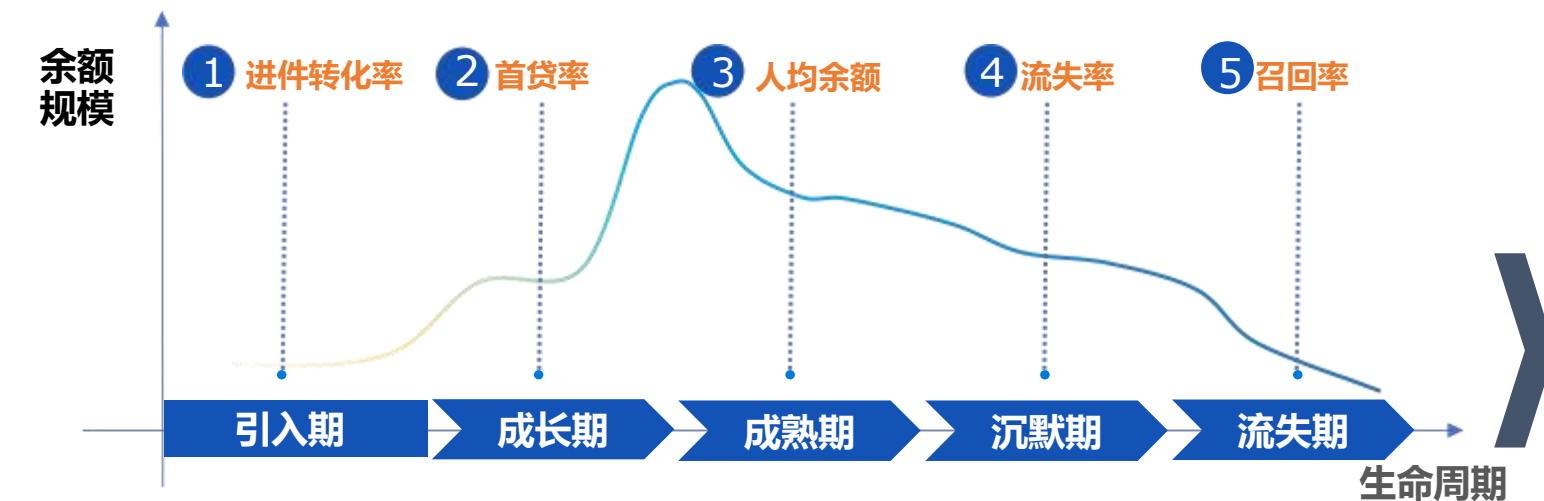
01

金融行业指标现状



指标化经营是企业数字化经营基础

存量客户数字化经营，核心是以**客户为中心**，在客户不同阶段，提供**差异化服务**，进而提升客户体验和价值。



客户可识别

获客渠道标签	授信额度	累计借贷笔数	提前还款次数	活动类型偏好
人口属性	首贷金额	参与活动次数	末次登录日期	末次借款日期

客户可分析

渠道质量分析	首贷漏斗分析	额度使用率分析	提前还款分析	召回率预测
注册至进件漏斗	首贷金额分布	活动偏好分析	客户流失概率	电销接通率预测

客户可经营

授信断点营销	首贷断点营销	老客权益推荐	提前还款干预	利益点二次唤醒
--------	--------	--------	--------	---------

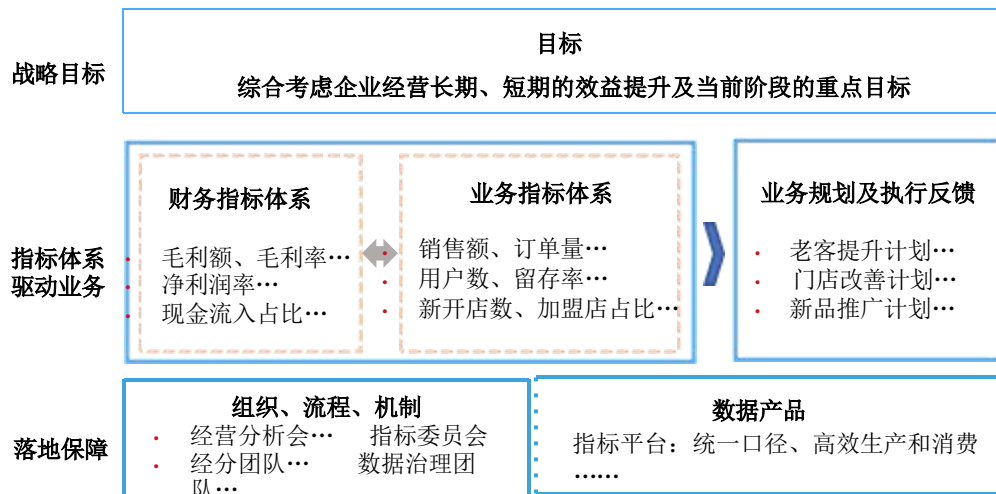
指标化经营的意义

- 指标体系可以映射业务经营逻辑，将指标拆解到可落地的业务过程指标，形成战略到执行的闭环
- 统一的指标口径，形成一致的业务水平衡量标准，提高经营决策效率
- 以面向业务的指标体系，牵引企业数据底座建设，数字化技术更好落地

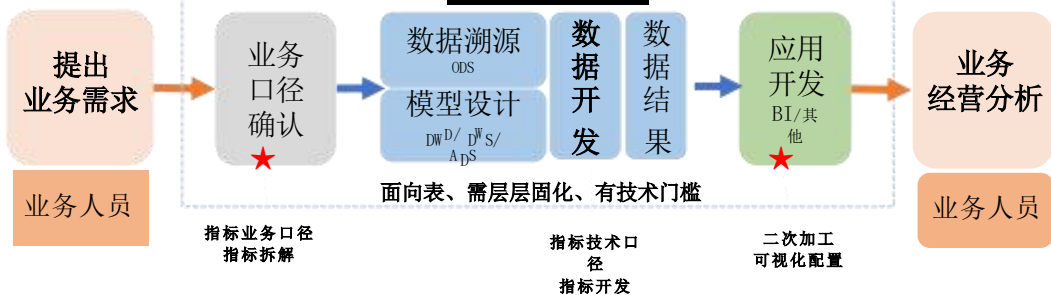
指标是实现企业战略目标的统一组织语言

企业视角

- 指标是实现企业战略目标的统一组织语言。
- 覆盖业务目标和执行过程的全域指标体系是数字化经营的核心。
- 完善数字化管理组织、流程、机制及数据产品是重要的落地保障。



数据团队



一般数据需求的开发过程

两类突出问题

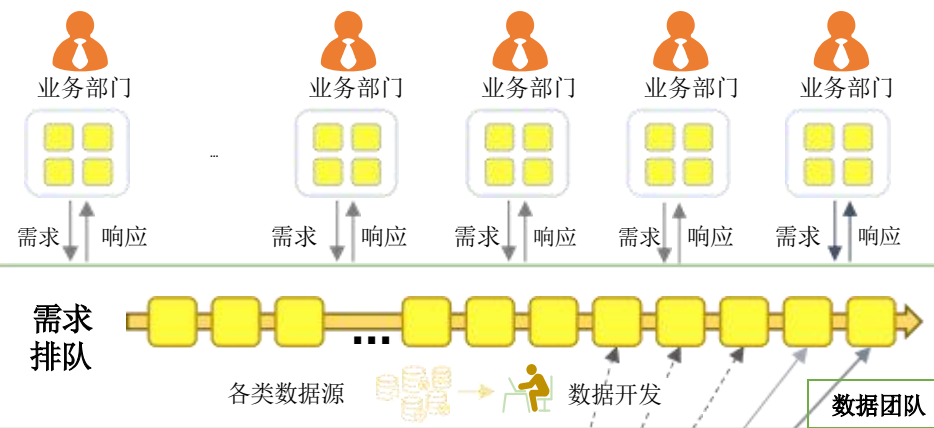
慢

- 需求响应慢
- 查询性能差
- 开发效率低
- 问题排查慢
- 业务应用建设周期长
-

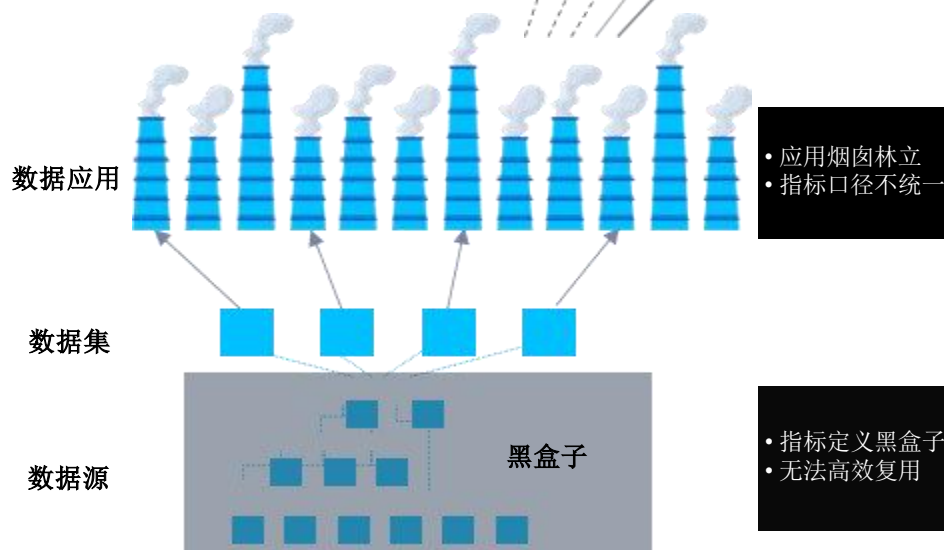
乱

- 定义乱，名/义不统一
- 口径乱，开发不规范
- 应用乱，系统烟囱多
- 血缘乱，问题难溯源
- 流程乱，权责不清晰
- 资产乱，指标无体系
-

- 大量业务数据需求涌现，开发效率受到极大挑战，难以及时响应

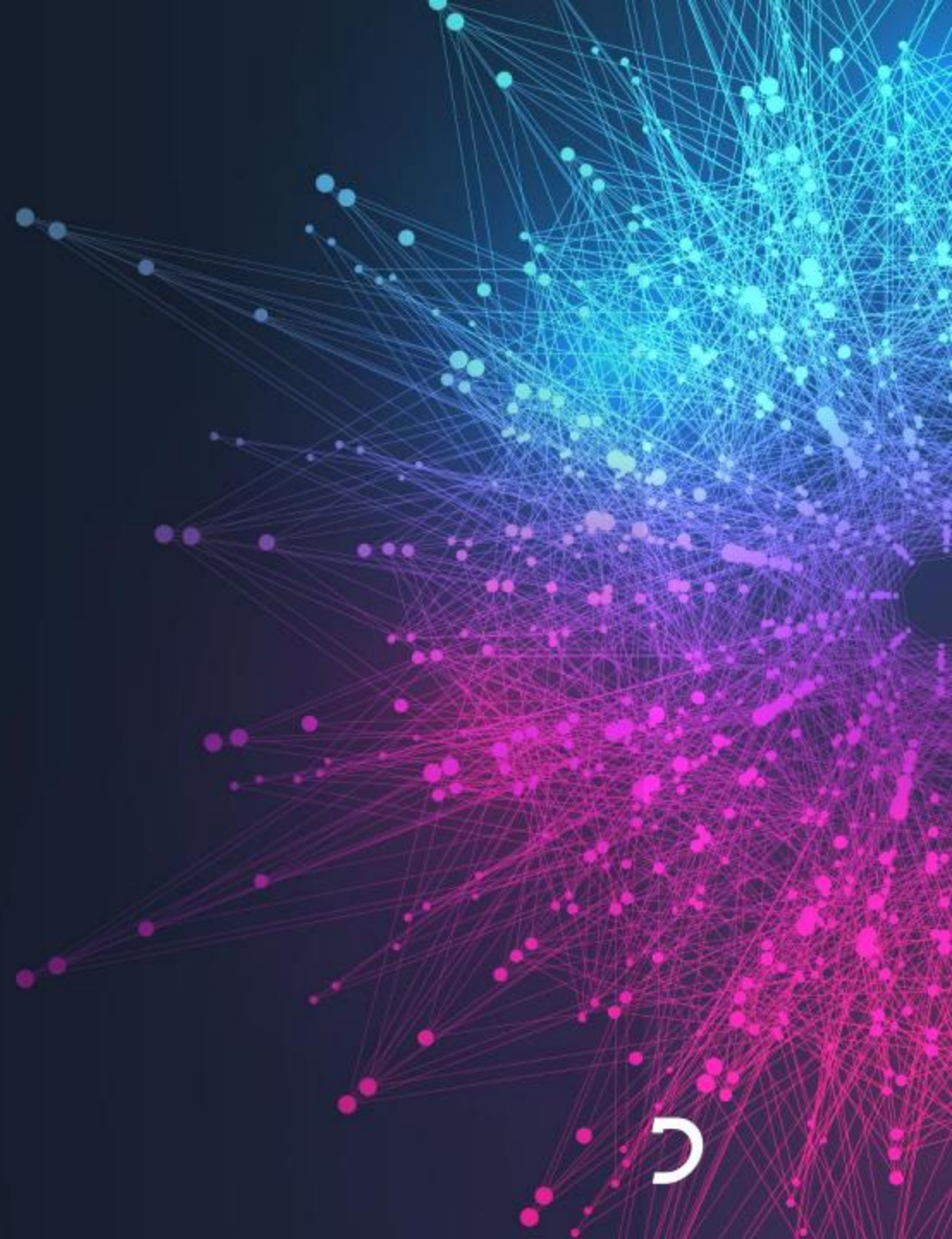


- 定义乱，指标口径不一致；数据烟囱林立；管理难、复用难；



02

痛点 -> 技术问题



指标体系建设和应用的痛点

管理团队

需要准确的指标进行**精细化管理与经营决策**，传统的数据大屏只解决了固定形式

“看数”的问题，但是一旦需要进行交互式的**问答归因**，依然是提需求给数据团队开发报表，**决策效率因此受阻**

数据团队

重复性指标开发工作多，**疲于应对临时性的报表开发需求**，很多报表开发完甚至只会被用一次就废弃了，难以真正将企业的**数据资产沉淀为“知识”**赋能给业务团队



业务团队

需要多维度的指标进行**精细化运营与自助分析**，BI工具**门槛相对较高**而且需要对底表字段非常熟悉才能上手，经常是依赖数据团队开发多维报表，自行下载到Excel里分析并人工得出结论，**得出有业务价值洞见的效率低**

指标体系建设和应用痛点的具体表述

管理

Data ≠ Insight: "我们公司花了很大精力建设了驾驶舱和各种大屏，看起来很高端，但数字并不等于见解。每次有具体问题，团队还得回去挖数据、分析指标，周报变成了无尽的等待游戏"

Speed is Everything: "每当董事会要求针对数据作出快速反应时，我都希望能够立刻拿到结论，这个数据大屏却只能给我表面的数，深入分析还得是手工在数仓里建作业，太慢了"

The Golden Why Question: "我们的驾驶舱解决了数据展示的问题，但未触及数据解释的关键。当涉及到业务背后的为什么时，它似乎没有答案"

业务

SQL学习难度高: "尽管我们有意愿深挖数据背后的真相，但SQL的复杂性让我们非技术人员望而却步。许多时间都浪费在了查询语言的学习上，而不是洞察和行动"

BI报表配置难度高: "我们依赖BI工具来提供数据的可视化，但每次都需要技术团队介入来配置数据集和报表，它的复杂程度令人沮丧"

大量时间花费在导出数据到Excel里输出结论: "分析团队的日常就像Excel的奴隶，时常花费大把时间将数据导入、整理和分析，这种重复低效的工作限制了我们对数据快速反馈的能力"

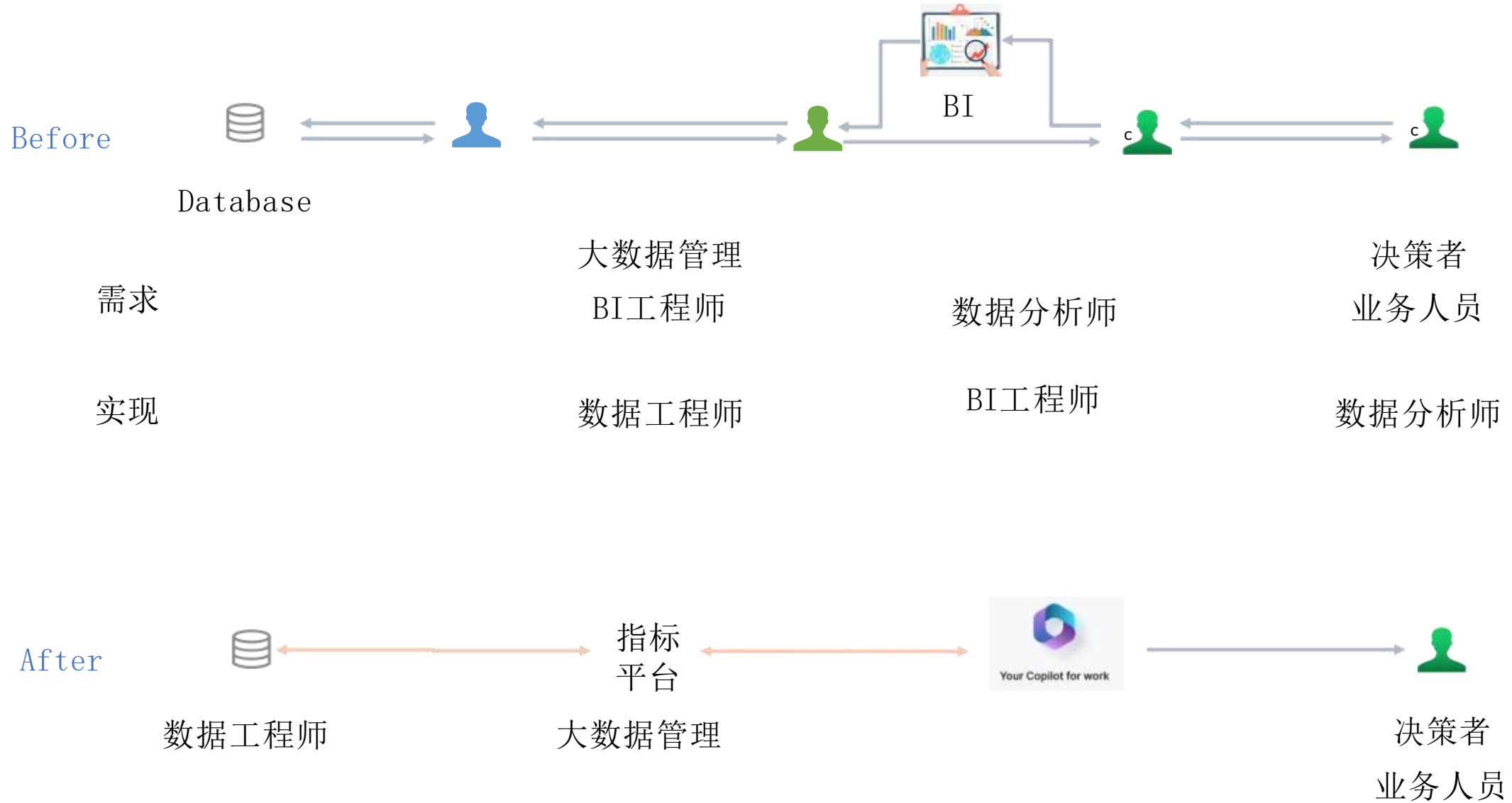
数据

报表需求灵活多变，临时表冗余: "我们数据团队大部分时间都花在了应对业务部门多变的报表需求上，每次业务流程一变，我们就得开发各种临时表、中间表。这不仅效率低，还很容易造成错误，甚至有的表用一次就废了"

指标口径不一致: "虽然公司有众多部门在使用数据，但每个团队对同一指标的定义却截然不同，没有统一的数据口径和解释标准。这种不一致性给跨部门的沟通和决策带来了混乱"

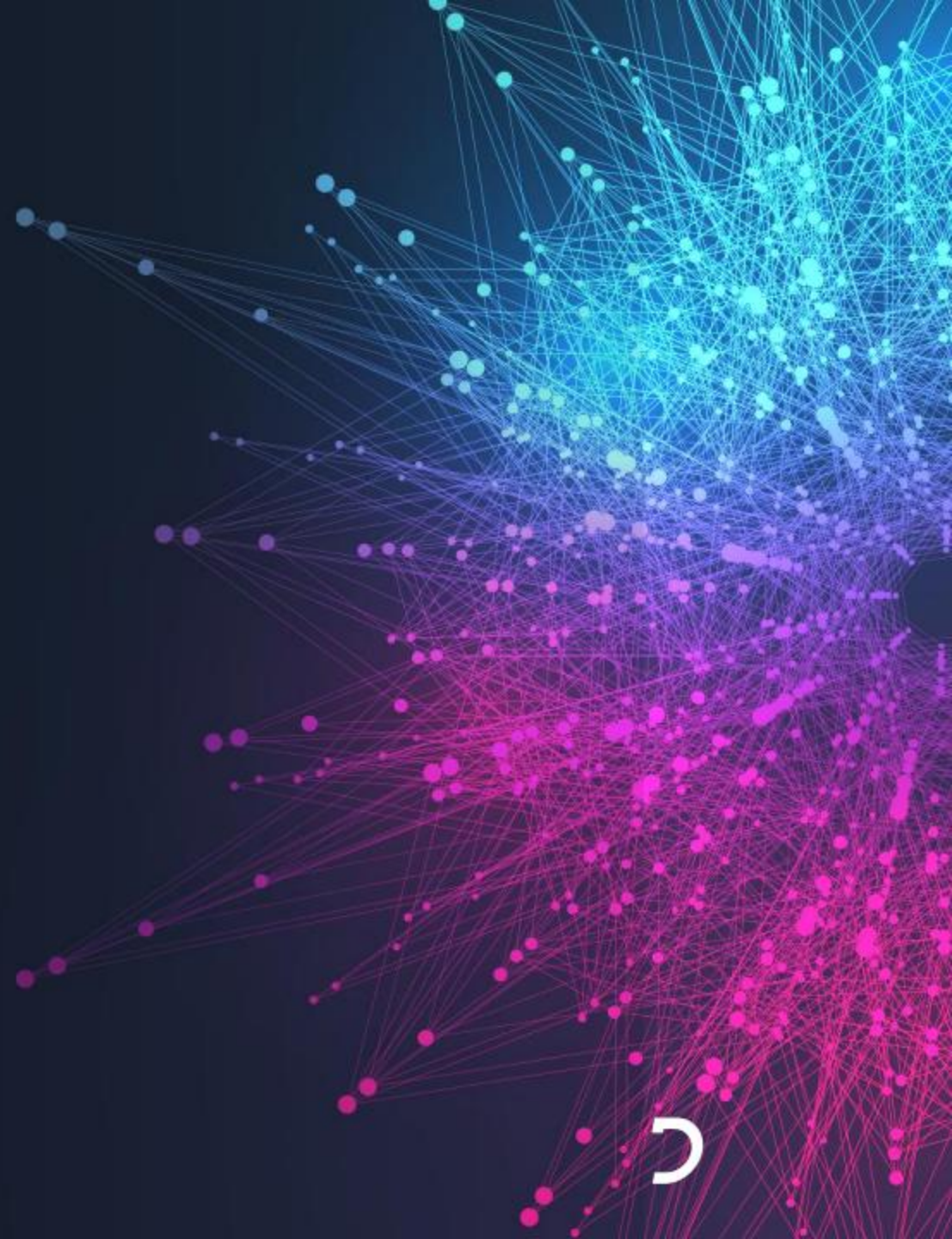
难以实时响应业务新增指标需求: "每次业务人员新增一个指标开发需求，都希望我们能半小时内提供相应的指标。现状是，虽然我们已经在数仓加班加点开发了，但还是被业务团队说反应慢，有苦说不出"

指标体系建设和应用痛点问题的抽象



03

解法



指标体系建设和应用痛点的解法（数据右移、角色解耦）

数据右移

数据建模右移

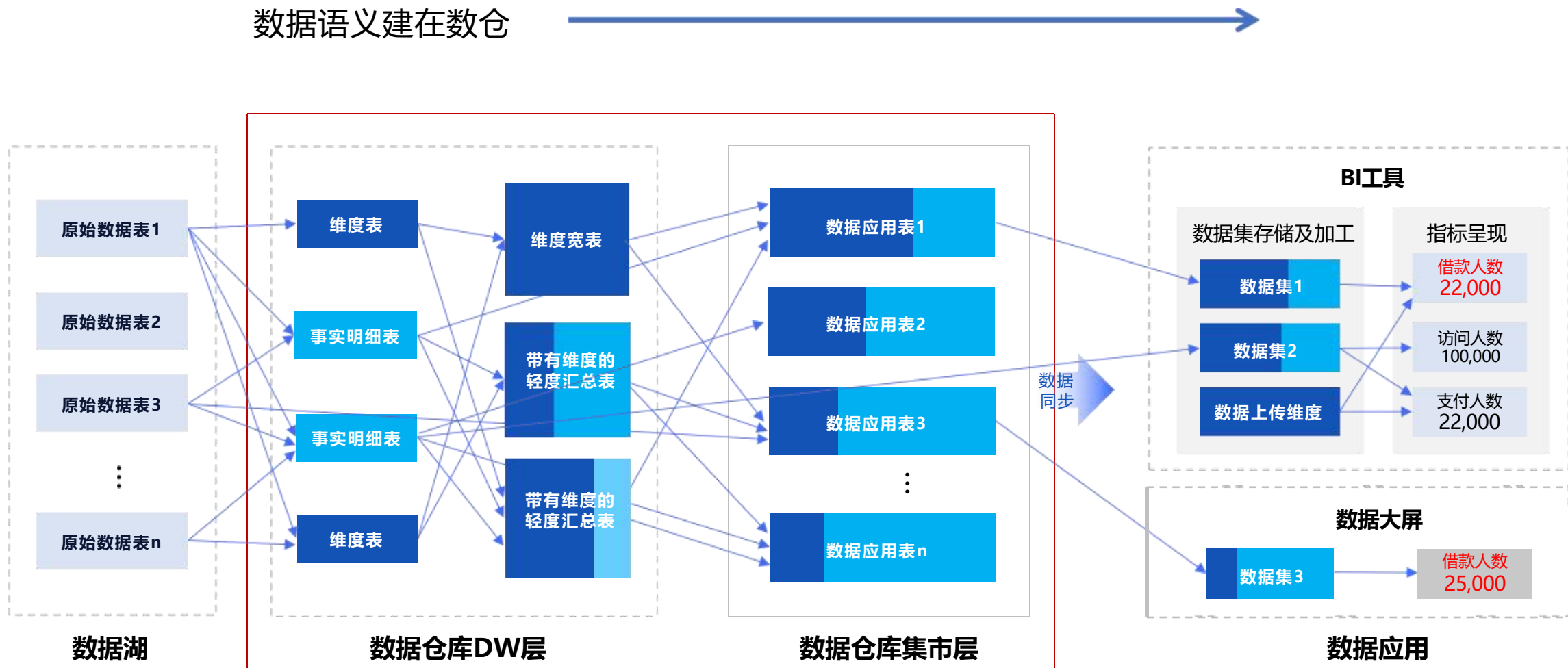
数据解读右移

角色解耦

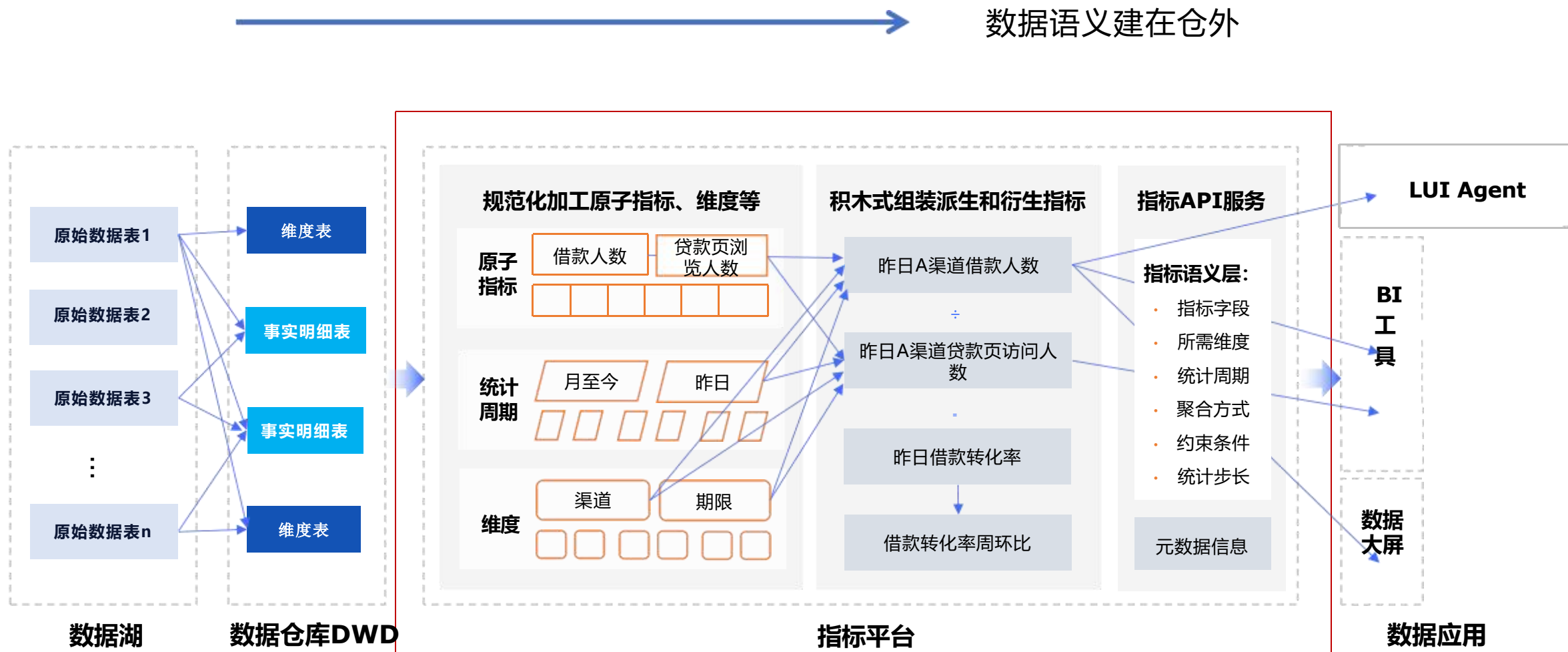
- 1 企业通过“数据右移”赋予非技术用户前所未有的洞察力。
- 2 “数据右移”揭示了数据民主化的力量，以及如何通过智能工具来简化决策过程，同时释放数据分析团队的资源，使他们能够专注于公司更高层次的战略任务。
- 3 当数据被孤立在不同的系统中时，公司会错失创收机会；当使用错误或不完整的信息时，公司会做出糟糕的决策，因此寻找企业的单一数据源将是重点。
- 4 人工智能分析的未来发展方向会是用于数据工程师的数据栈将与服务于非技术背景用户的新兴数据右移解决方案并存。

指标体系建设和应用痛点的解法（数据右移、角色解耦）

数据语义建在数仓



指标体系建设和应用痛点的解法（数据右移、角色解耦）



数据建模右移 - 指标平台、HeadlessBI

OLAP

HME

HeadlessBI

产品力

10+客户成功交付

首发《指标平台白皮书》

首家 by信通院认证

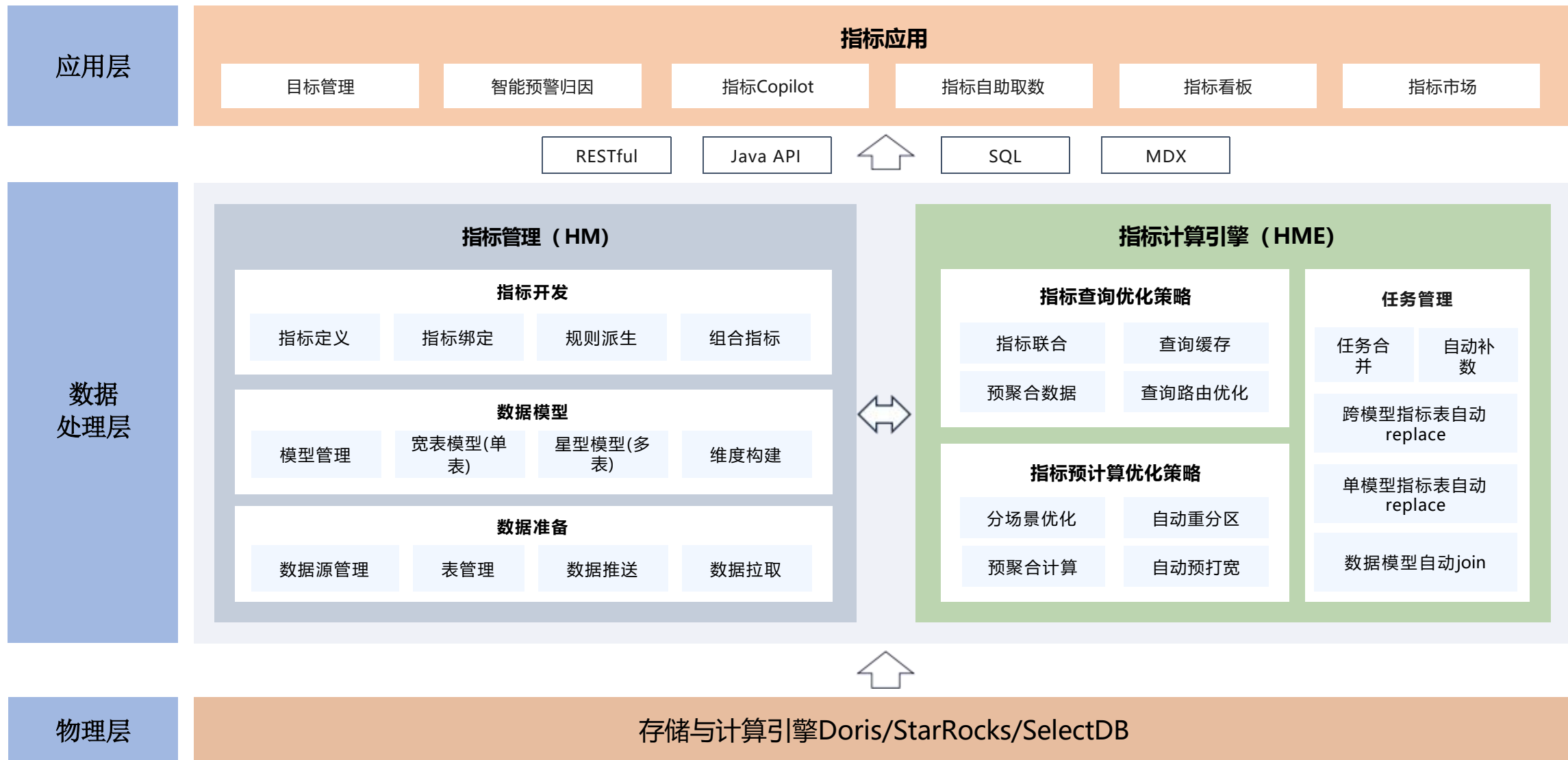
灵活性

性能

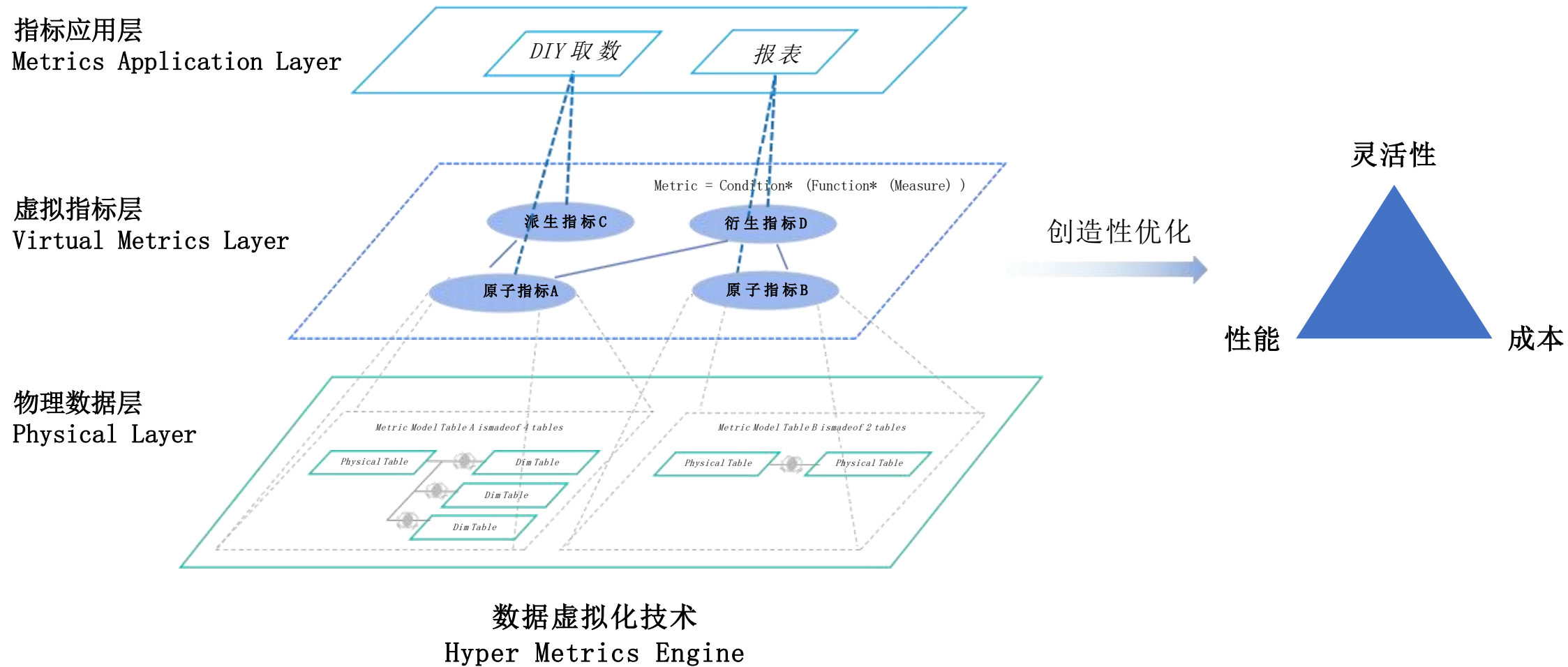
成本



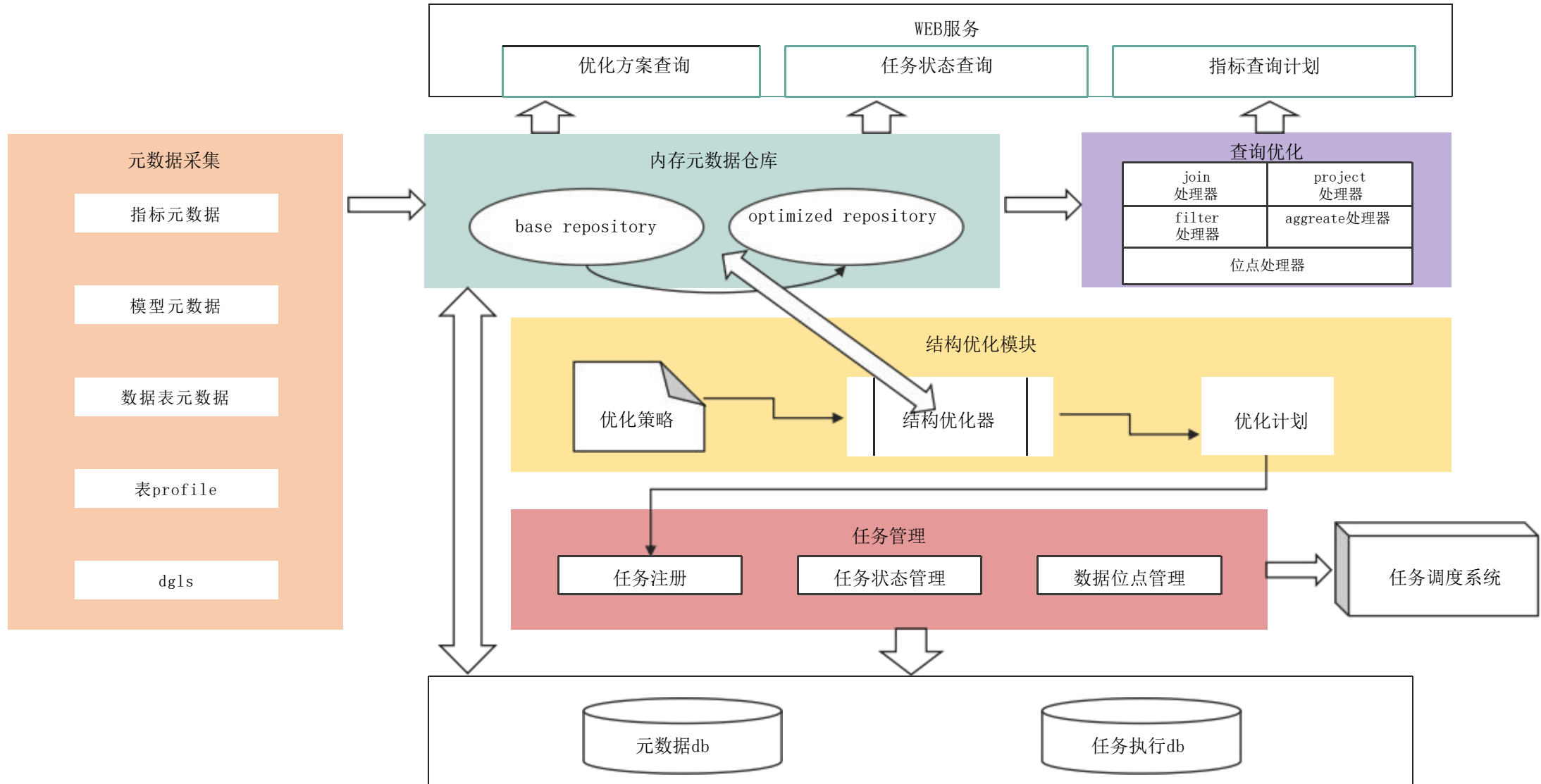
数据建模右移 - HeadlessBI 架构



数据建模右移 - 高性能虚拟计算引擎V3.0



数据建模右移 - HMEV3.0 基于ROI智能自迭代加速



数据解读右移 - SwiftAgent



数据分析师

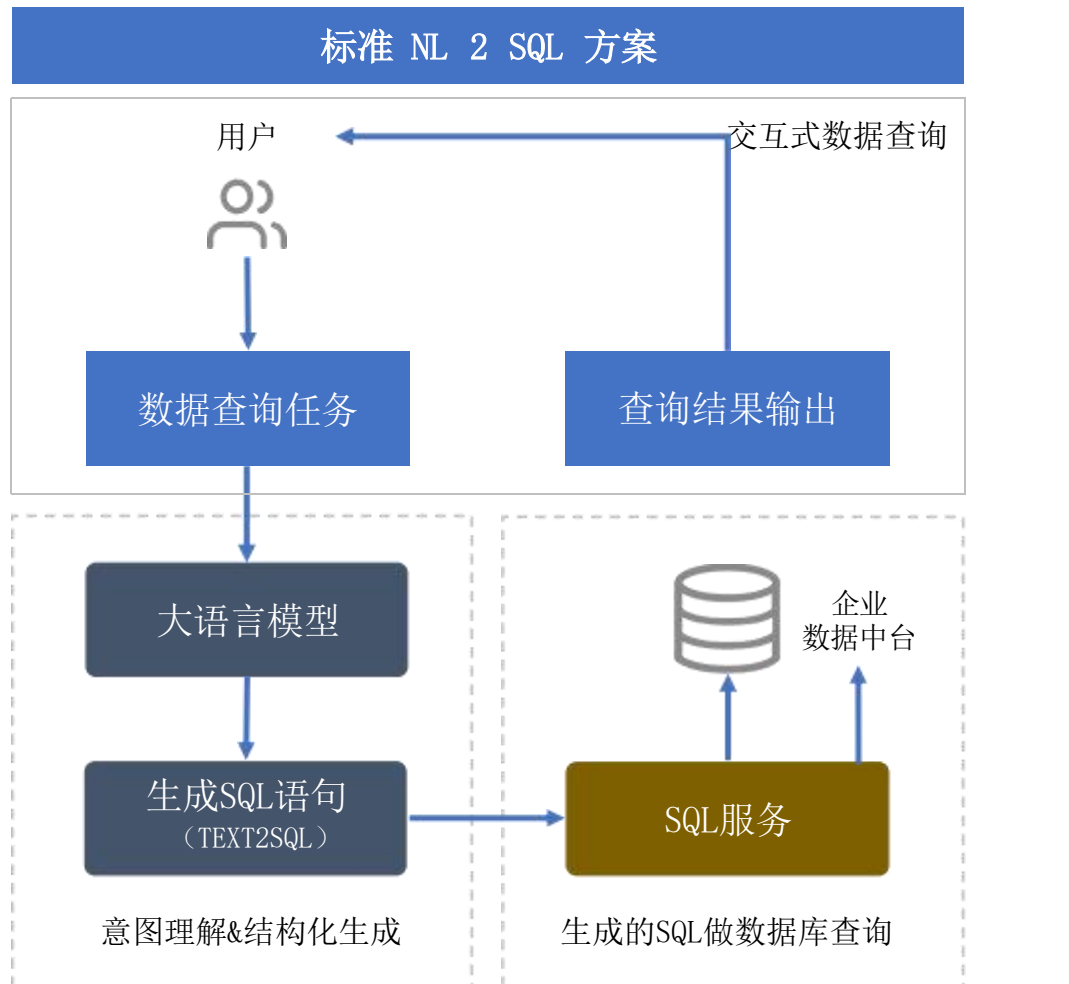


数据消费者

? 不会用BI

? 不会问问题

数据解读右移 - NL2SQL存在效果瓶颈和性能风险



① 准确率低

- 企业级（千/万张表）数据查询 **准确率约60-70%**
- 即使是GPT也难有一个直接可用的准确效果

② 性能存在风险

- 大模型写出的SQL未经优化，可能导致长查询的占比较多，性能不好甚至导致查询服务异常，多表关联的真实查询 **性能不可控**

③ 学习成本高

- 预训练大模型未掌握企业客户数据与数据格式定义，当数据源过大时，大模型对表的结构 **学习成本高**（增量预训练、长窗口、functioncall等）

④ 数据安全存在风险

- 大模型直接对接数据查询，缺少数据权限管控，**隐私与安全性风险较高**，用户可通过问询获取非用户权限内的数据

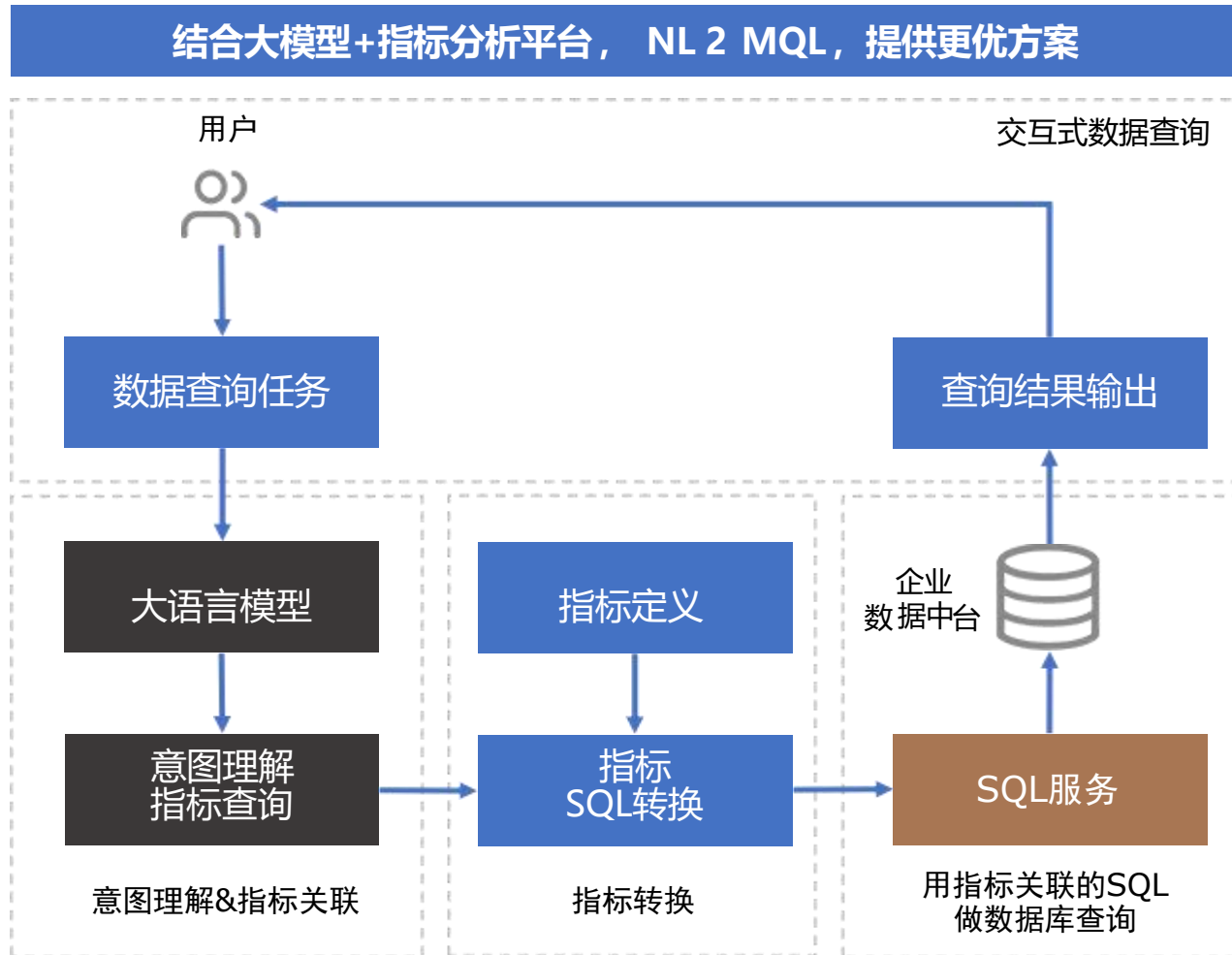
⑤ 能力单一

- 很多高级的分析问题，单纯靠大模型生成SQL **无法解决**

数据解读右移 - 数势NL2MQL之路相比NL2SQL优势

场景数据分析诉求	任务难度系数	难点	NL2SQL	NL2MQL+Agent
1 最近7天客户的 每日总资产 是多少?	★	枚举值转维度	可实现	可实现
2 XXX基金今年 累计卖了 多少?	★ ★	枚举值转维度 指标语义模糊	可实现但不准	可实现且准确
3 今年XX基金的 申购额 与 赎回额 分别是多少?	★ ★ ★	指标语义模糊 跨表多指标查询 日维度上卷年维度	难实现	可实现且准确
4 XX基金公司 最近3个月 销量最好的 Top3基金 是哪些? 这三个分别的 保有量 是多少? 帮我解读下	★ ★ ★ ★	多任务规划 指标取数 结果排序 衍生计算	可实现但不准	可实现且准确
5 华东区XX基金的 申购金额 周环比 为什么下降了?	★ ★ ★ ★ ★	枚举值转维度 多维归因 同环比计算 报告解读	难实现	可实现且易理解

数据解读右移 - 指标平台+大模型，实现数据价值的端到端



① 数据可信，准确率明显提升

- 预设数据指标的定义与管理，避免业务理解对不齐
- 借助思维链分析与歧义反问，提升泛化性，避免直接从文本到SQL

② 性能提升且稳定

- 基于自研的数据查询加速引擎，智能优化查询语句
- Top95 可实现 从检索到回答的 **秒级出数**

③ 学习成本低

- 指标的数据样本可通过大模型辅助生成
- 全流程白盒，企业客户用业务语言描述查询过程，方便快速排查

④ 数据安全可靠保障

- 利用指标分析平台的**权限管理能力**，结合RBAC基础，对数据与指标进行精细化的权限管控，**实现数据查询的安全可控**

⑤ 能力覆盖更全

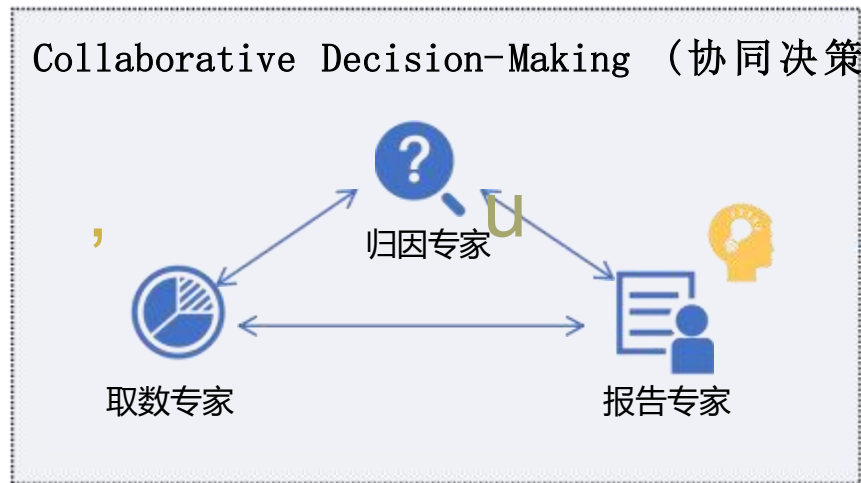
- 高级数据分析问题，可通过精准的指标进行关联与展示，实现单项数据可查、报表可展示、总结报告可生成

数据解读右移 - “Agent” 架构运转流程

N轮循环

Goal (目标)

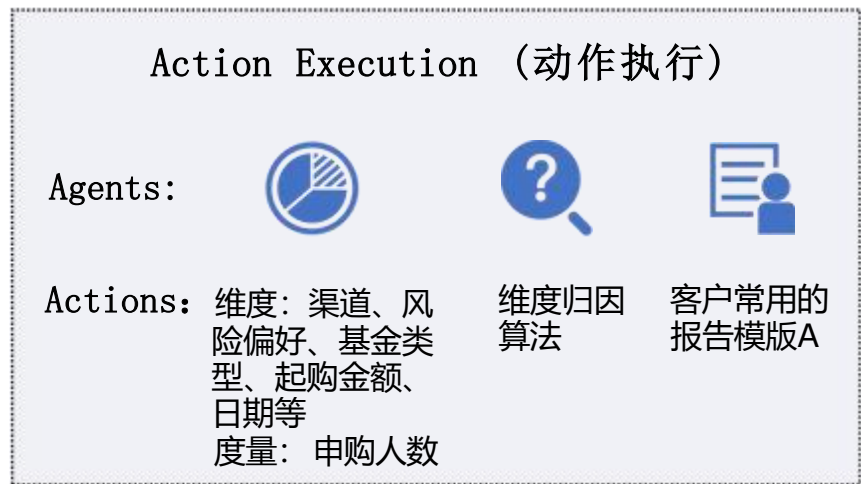
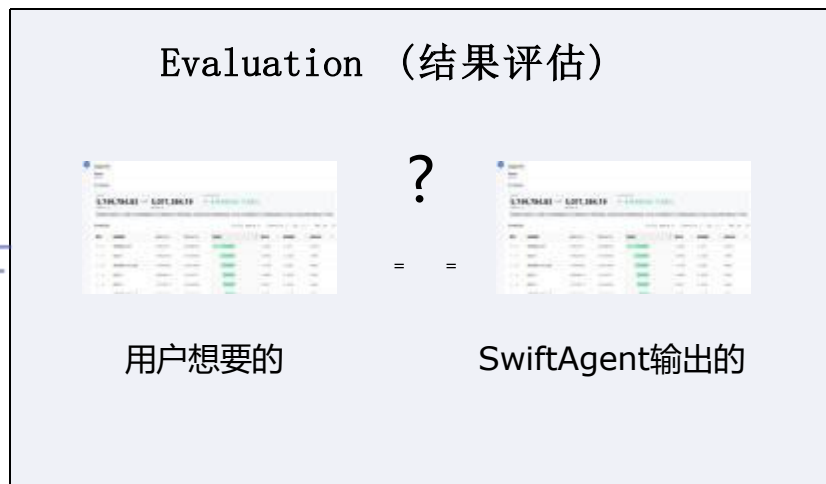
看下近7天基金申购人数，哪个渠道跌得最多，做个总结



Reward Feedback (正负反馈)

Outcome (结果)

有数据、有逻辑有业务Sense的分析报告



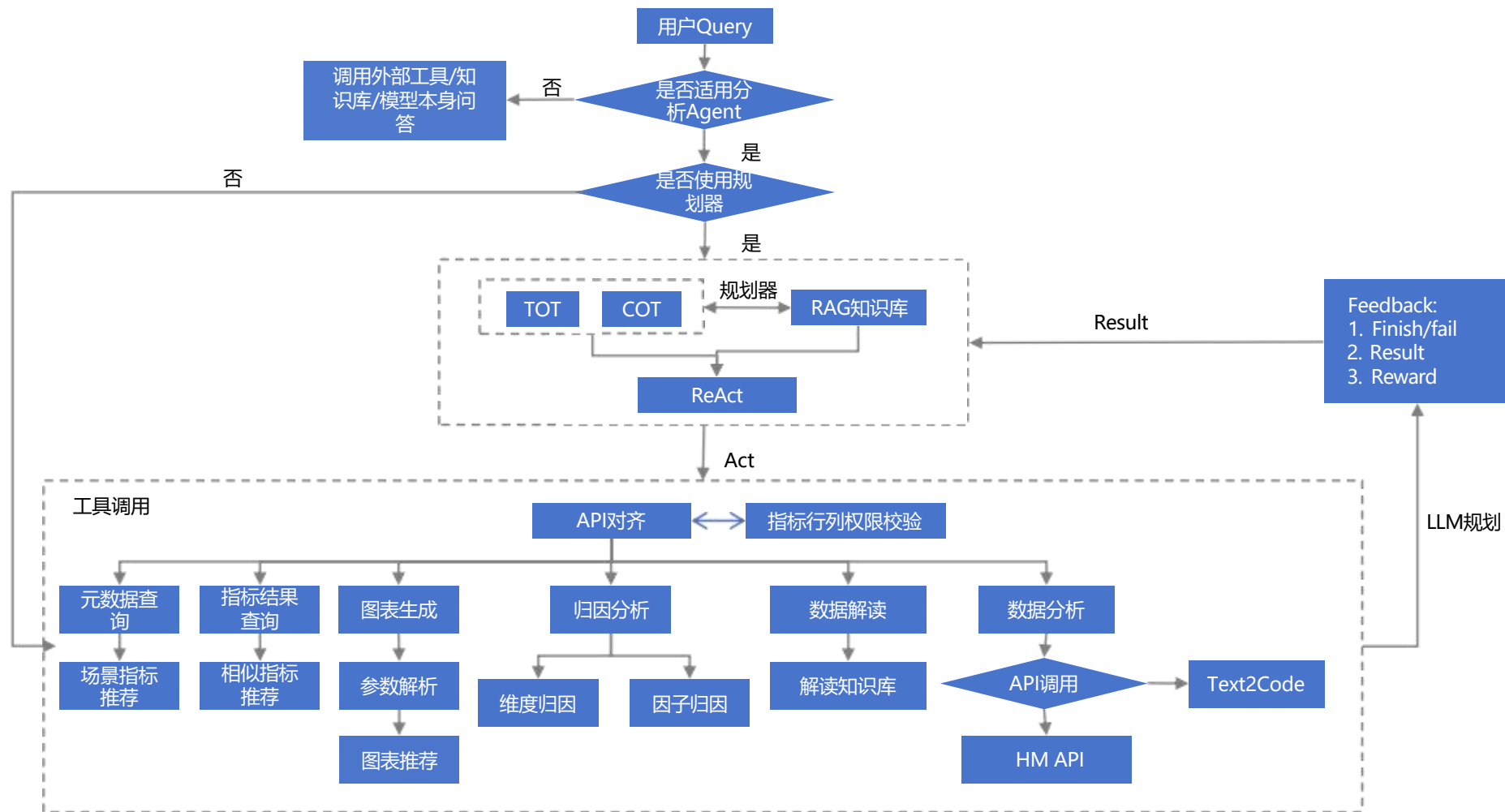
数据解读右移 - SwiftAgent流程图

基于大模型的理解和思考能力，实现基于**自然语言的数据分析**，并针对获取的结果数据进行**智能解读与归因**，提升经营分析效率

1. 任务规划能力——将复杂任务化繁为简，逐步拆解让大模型自动执行。

2. 数据理解能力——依托指标和标签的语义特性，让大模型更懂数据。

3. 高效计算能力——将计算任务通过自研计算引擎HME提交执行，极大优化了计算效率。



数据解读右移 - SwiftAgent 创造性解决5大挑战

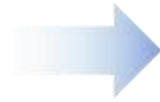
1. 分析准确性问题

统一语义层的构建 (Unified Data Semantics)

2. 数据全面性问题

多源异构数据链接 (Diverse Data Connection)

3. 人机融合问题



用户可干预 (Human in the Loop)

4. 持续学习与迭代问题

持续反思学习 (Continuous Reflective Learning)

5. 计算查询效率及性能问题

数据计算加速引擎 (Hyper Computing Acceleration)

04

案例



指标产品合作案例—某头部证券公司基于指标平台的经营分析平台建设

产品激发需求爆增
研发效能成为短板

需求沟通成本高

指标开发成本高

报表上线成本高

缺乏指标管理工具
数据使用渠道有限

多口径是正常的，
但来源必须清晰

缺乏纳管所有指标的
统一平台

不能分享和导出

固化报表千人一面

自助BI 使用门槛高

固化报表不可修改

不能创建个人报表

指标产品合作案例—80%的新增指标不依赖数据人员

80% 的指标基于基础指标，使用零代码的配置工具进行派生。

举例：线上新客的股票资产

基本信息

* 指标名称: 线上新客的股票资产 * 英文名称: online_new_customer_asset_n

业务信息

* 业务含义: 当年开卡的客户视为新客。关注这个客群在当期股票账户的资产情况。

技术信息

* 可用维度: 资产类型 * 统计粒度: 月

资产指标: 股票

统计周期: 当月

快速配置: 最近3天 最近7天 最近30天 最近1周 最近2周 最近1个

业务范围: 生命周期 新客户

资产范围: 品种 股票

渠道: 线上

20% 的基础指标由数据人员开发。基础指标包含最明细的度量和维度。

以资产为例，具体到每个客户每一天在每一个账户下每一只持仓标的的资产。



举例：线上新客的人均股票资产

基本信息

* 指标名称: 线上新客的人均股票资产 * 英文名称: avg_stock_asset_of_new_customer

业务信息

* 业务含义: 当年线上新开卡客户的人均股票资产金额

技术信息

* 可用维度: 资产类型 * 统计粒度: 月

业务范围: 新客户

资产范围: 品种 股票

渠道: 线上

指标产品合作案例—面向业务人员的BI分析产品，随时随地修改卡片



配置分析指标

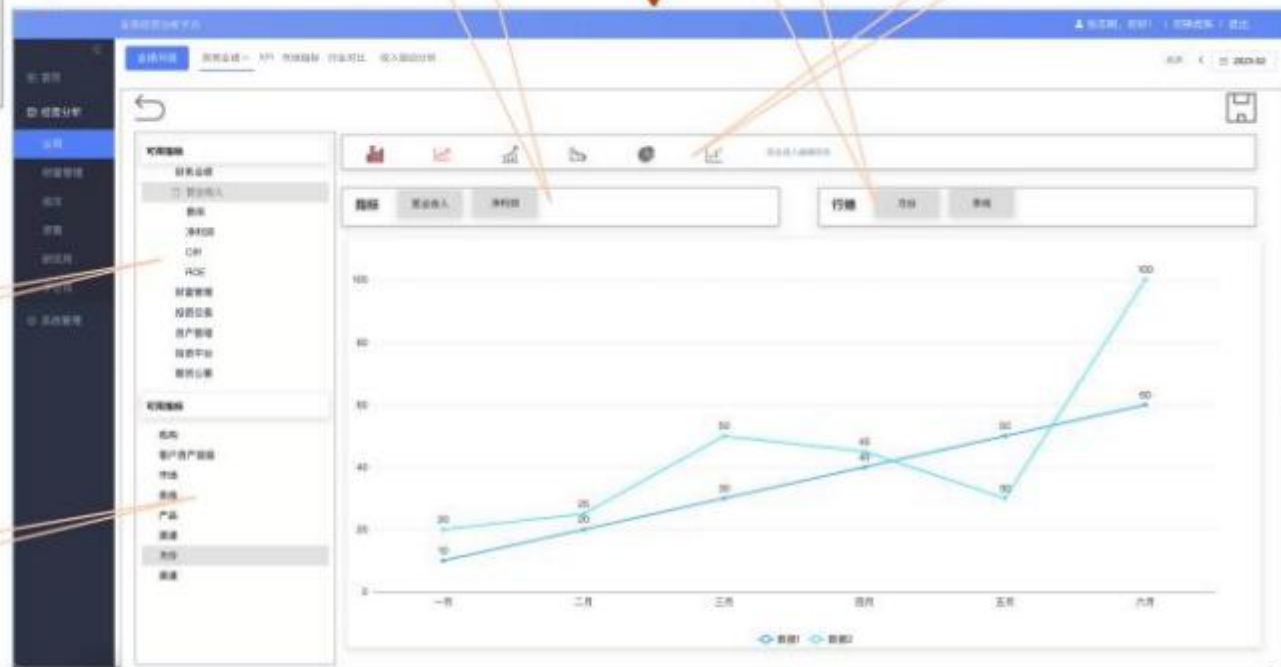
配置下钻维度

自选组件

即时自助分析

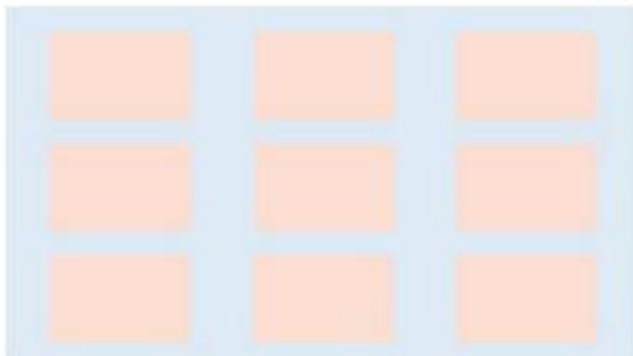
可用指标

可选维度



指标产品合作案例—面向业务场景，低门槛灵活组装个性化看板

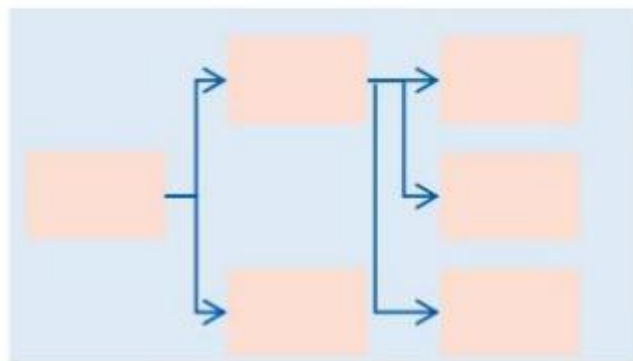
网格模板



会议模板



关系模板



选模板和布局

选卡片

自由布局

我的卡片页 选卡片尝试看板

卡片看板模式

卡片布局

- 功能测试 (11)
- 自助分析卡1
- 净利润和收入趋势
- 积分系统数据
- 积分系统数据_按部门
- 积分系统_个人活跃度排名
- 活跃部门下钻
- 部门个人活跃度排名
- 积分系统整体收入
- 营业收入
- 商品交易费收入
- 物流服务评价收入
- 普通用户佣金收入
- 信用用户普通佣金收入
- 保证金利息收入
- 运营活动效果
- 积分测试卡片1
- 净利润趋势
- 运营活动效果-小卡

积分系统历史活跃

积分系统数据

请插入卡片/元素

指标产品合作案例—支持业务同学对话式进行数据分析，更好的智能化交互体验

解决痛点

数据解读靠手工

缺少分析思路

制作报告效率低

产品效果

指标分析助手为用户自动化提供贴合专业分析师思维的报告，并将**数据概览、图表、高级计算（同环比/聚合/占比/排序/相关性/预测）与文字结论**有机地整合在一起



数据分析报告

1. 数据概览

根据查询结果，我们获得了2023年9月至11月坦克300不同车款的国内销量数据。以下是数据的一些基本洞察：

数据跨度：从2023年9月至11月
最大销量：城市版在2023年10月达到了33,415辆
最小销量：越野版在2023年11月达到了30,850辆



2. 数据详情分析

2.1 销量趋势分析

通过观察数据，我们可以看出不同车款的销量在这三个月内有所波动。城市版车款的销量从9月到10月有所增长，但在11月有轻微下降。游侠版车款的销量在这三个月内相对稳定，而越野版车款的销量呈现逐月下降的趋势。

2.2 平均每月销量分析

为了回答用户的问题，我们需要计算每个车款的平均每月销量。

城市版：平均每月销量约为32,460辆
游侠版：平均每月销量约为32,837辆
越野版：平均每月销量约为31,173辆



根据数据结果，游侠版的平均每月销量最高，约为32,837辆。

谢谢观看