

第三代指标平台的物化加速实践

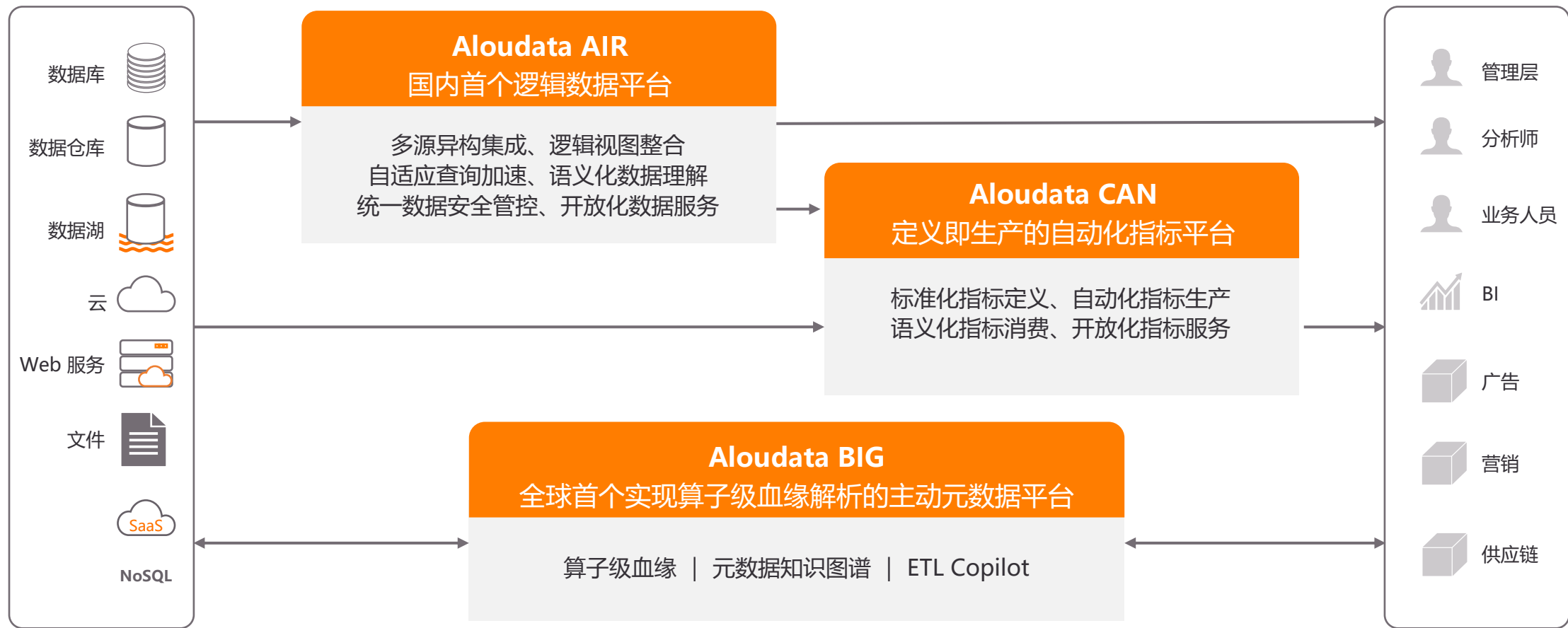
杜雪芳 | Aloudata 首席业务架构师

2024/02/29

Aloudata

NoETL 重塑数据生产力，让数据随时就绪

Aloudata 产品矩阵



创始团队：自动化数据管理的践行者和引领者



周卫林

创始人&CEO

- 20 年大数据领域经验
- 原蚂蚁研究员 (P10)，原蚂蚁数据平台部总经理，蚂蚁数据技术主要奠基人
- 曾从零构建蚂蚁 BI、AI 工具产品体系，孵化多个进入 Gartner、Forrester 领导者象限的阿里云拳头产品



周泉

CTO

- 12 年互联网和大数据平台架构经验
- 原蚂蚁数据平台全域架构师
- 曾主导金融级数据平台架构的发展和演进，先后从零构建蚂蚁实时智能决策平台、智能数据治理体系以及数据安全合规体系



肖裕洪

首席产品官

- 12 年大数据平台产品经验
- 原蚂蚁数据平台产品负责人
- 曾从零主导建设蚂蚁集团数据平台产品体系，包括数据建设与管理平台、增强数据分析平台、实时智能决策平台及 AB 实验平台等



赵尉淋

客户成功官

- 20 年数据仓库领域经验
- 原蚂蚁、菜鸟数据平台部数据架构师，菜鸟物流数据体系主要设计者和建设者。
- 曾主导全域数据架构升级，负责数字金融线的数据研发、计存治理和合规治理工作

目录 Contents

1. 指标平台的演进趋势
2. 第三代指标平台核心能力突破
3. 第三代指标平台物化加速实践

业务需求发生变化：从数字化管理到数智化运营



管理层与业务人员

- 数字化转型需要一线业务人员进行数据分析和洞察，数据的灵活性对发现新机会至关重要
- 数据民主化是所有大型组织的首要任务，数据分析的准确性，数据口径的一致性，是数据民主化的基础



数据团队

- 由于需求差异性，面向不同业务部门的需求分别进行数据开发，存在大量重复开发
- 业务在分析探索过程中，不断新增或调整需求，新增字段或指标口径变更成本高，需求响应效率低



CIOs and CDOs

- 信息孤岛和数据烟囱现象突出，业务系统各自为战，数据质量低，计算存储成本高
- 数据标准不统一，不同系统、不同业务部门定义的口径存在差异，无法快速提供准确可用的数据

业务数据需求的两种典型响应模式

数仓模式

慢

由 **ETL 工程师** 承接业务需求

完全按需开发，所有指标按场景预先聚合成宽表或汇总表

优点： 遵循数仓建设规范，集中式管理

缺点：

- **开发慢：** 需求排期等待长
- **重复建：** 同一指标不同维度开发不同表
- **字段少：** 静态报表，无法实现灵活分析

BI 模式

乱

由 **分析师或业务团队** 承接业务需求

在 BI 中通过拖拽式 ETL 或自定义 SQL 产生大量私有数据集

优点： 快速响应业务需求

缺点：

- **口径乱：** 不同数据集业务逻辑不一致
- **复用差：** 指标分散，跨平台/工具难复用
- **治理难：** 治理程度低，维护成本高

指标平台是解决指标“乱”和“慢”的最佳策略

统一语言，消歧去重



存在歧义

沟通与理解成本高

统一认知

使用同一种语言

消除业务与技术的鸿沟



表：技术语言

了解数据结构、表关系、字段含义

指标：业务语言

无需关心表和字段，熟悉的业务语义

通过指标平台实现集中式业务语义管理，平衡指标口径一致和分析灵活

指标平台的演进趋势

第一代

指标口径登记
与管理

管用分离 管研分离

第二代

依赖 ETL 开发宽表
的指标平台

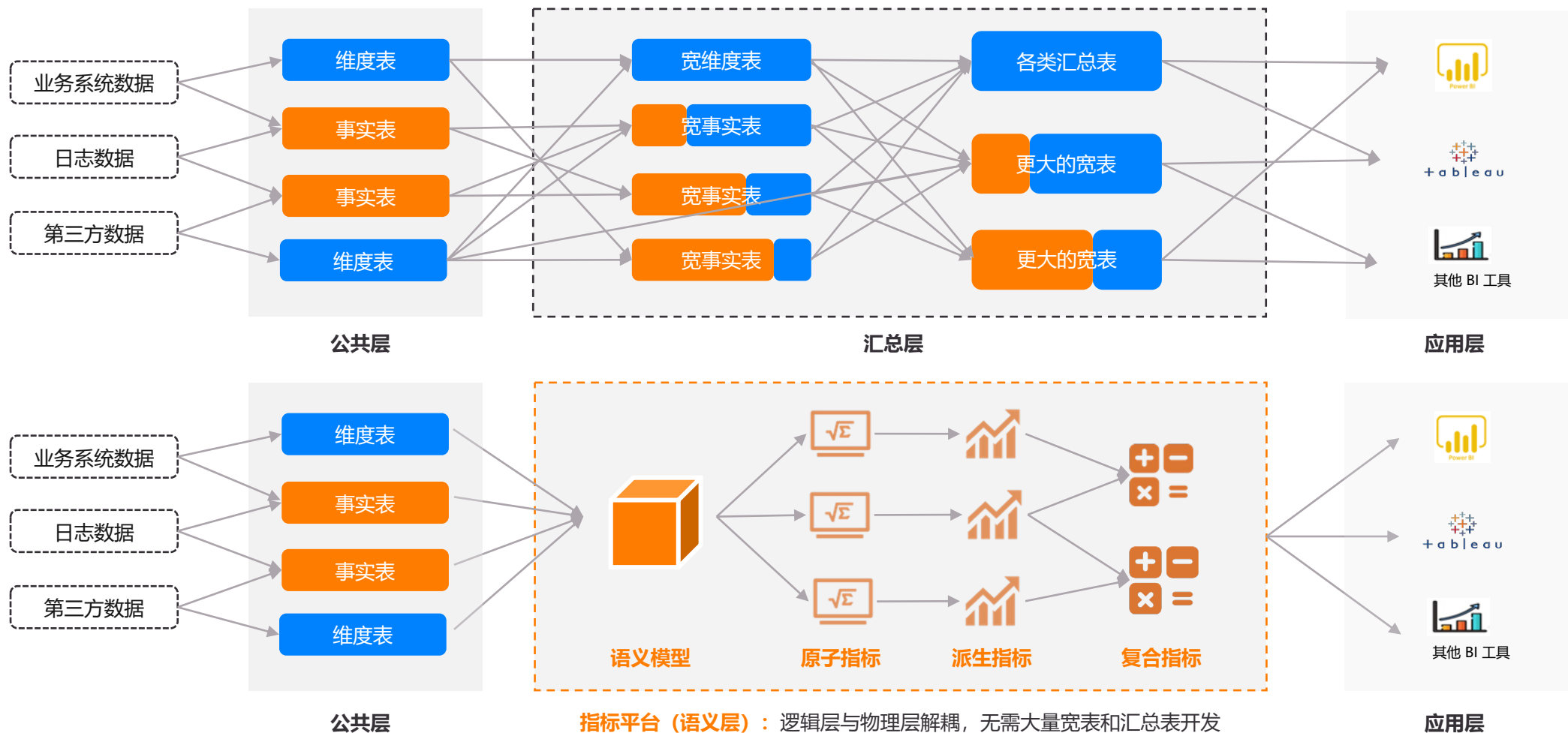
管用一体 管研分离

第三代

NoETL 自动化的
指标平台

管研用一体

什么是第三代指标平台：NoETL 自动化的指标平台



第三代指标平台：Aloudata CAN



定义即治理：100% 指标口径一致

基于强大的语义建模和指标要素配置化定义，系统自动进行“同名不同义、同义不同名”的判重校验

定义即开发：开发周期从周缩短到分钟

依托物化加速引擎，系统代持宽表与汇总表的生产，实现应用层NoETL

定义即沉淀：降低指标排查与沟通成本

指标业务语义的统一沉淀，统一业务用数语言，并提供字段级血缘和指标多版本，清晰指标加工全链路并掌握指标历史变更记录

定义即消费：一个指标，多场景消费

提供 API、JDBC 标准化接口，支持指标、维度、筛选、时间范围等灵活的接口参数，实现“一个指标，多场景消费”

第三代指标平台的核心能力

数仓应用层 NoETL，做“轻”数仓

语义化

强大的指标定义能力

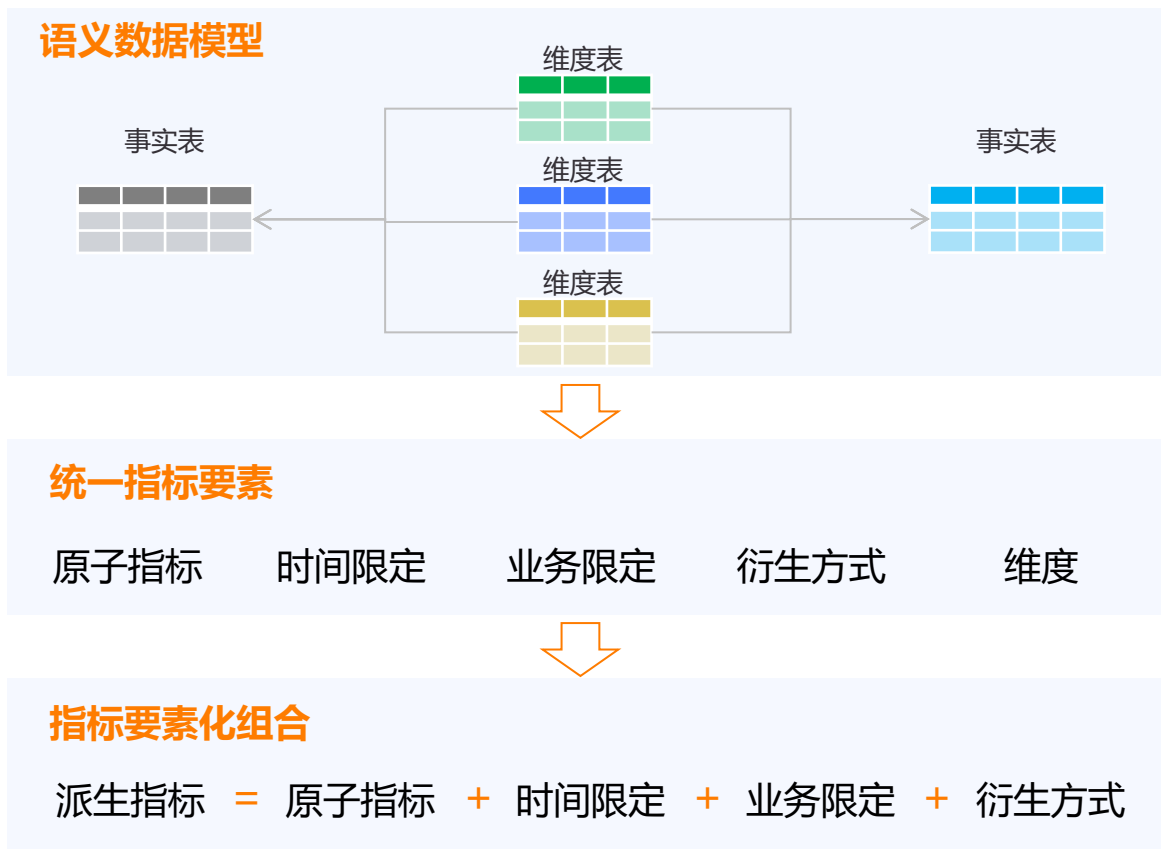
强大的语义模型和函数能力，复杂指标通过指标模板实现配置化定义，实现数据加工者与消费者隔离，做轻数仓

自动化

自动的指标开发

基于自研的物化加速引擎，IT 无需开发宽表与汇总表，系统自动编排、物化与回收数据

语义化：基于数据模型的标准化指标定义



数据模型：逻辑关系而非物理关联

只需表达数据集之间的关联关系，无需物理关联打宽或汇总计算，保留模型的灵活性，实现指标灵活定义与下钻分析

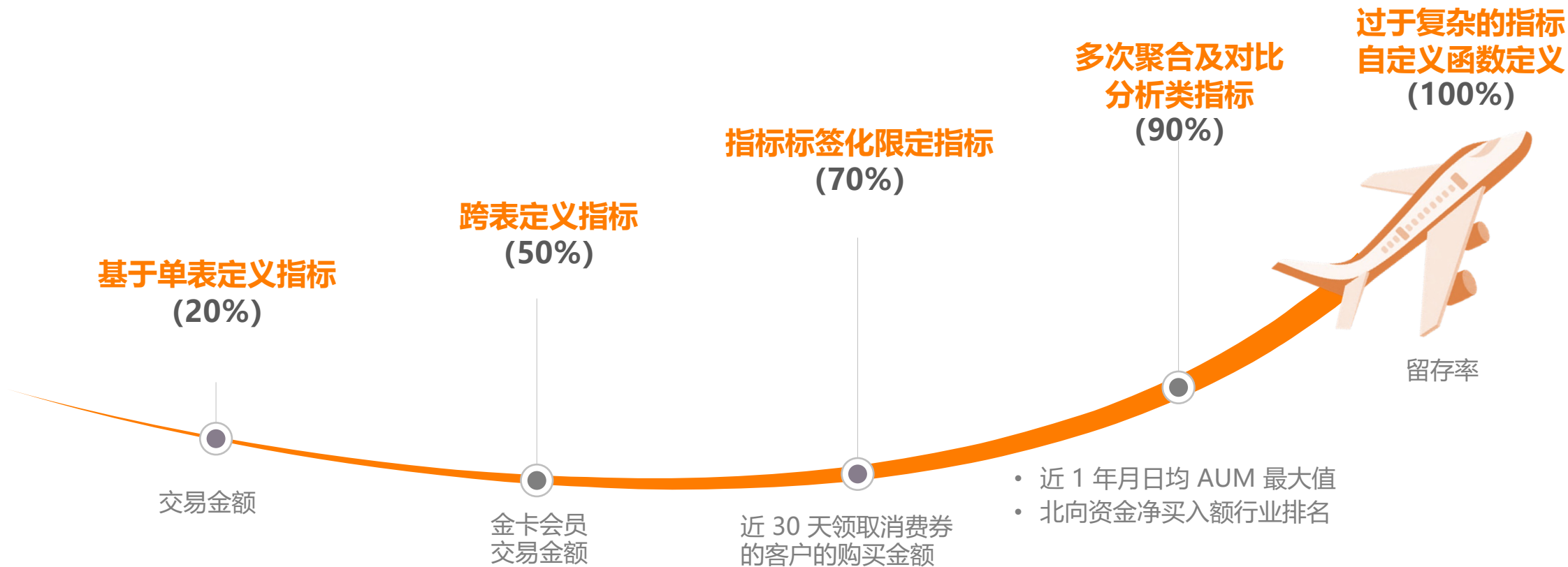
4 大指标要素：标准化指标定义与管理

指标要素绑定唯一物理口径，实现逻辑口径及物理口径的标准化

基于指标要素组合快速定义指标

基于指标要素背后绑定的口径，自动生成 SQL 计算逻辑，并自动物化加速，实现自动化研发，大幅提升指标开发效率

语义化：复杂指标配置化定义，降低指标开发门槛



语义化：复杂指标配置化定义示例

多次聚合类指标

复杂衍生方式

创建派生指标

编辑

基础信息

* 指标中文名

门店销售订单量在大区内排名

* 指标英文名

shop_order_cnt_rank

业务口径

门店销售订单量在大区内的排名

* 单位

其他

* 数据负责人

aloudata_zy

* 业务负责人

aloudata_zy

类目

零售

计算逻辑

* 基础/复合指标

demo_订单量

时间限定

业务限定

衍生方式

衍生方法

同环比

☒ 排名

占比

自定义聚合

排名范围

地区 × + 添加

排名维度

shop_name × + 添加

排名方法

Rank (1,1,3,4)

降序 (从高到低)

「统计周期」、「业务限定」或「衍生方式」必填一项

帮助手册

保存后发布上线

取消

保存

创建后发布上线

取消

创建

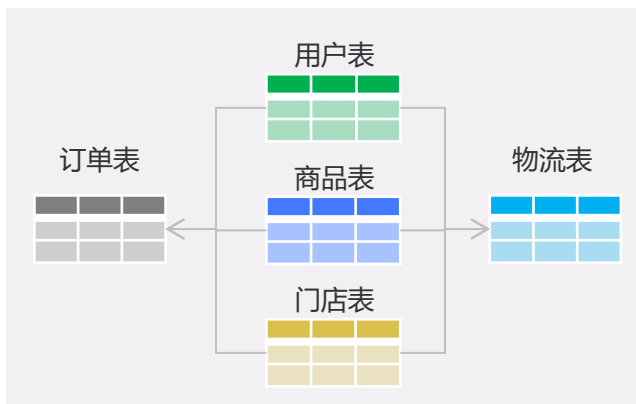
自动化：基于自研的物化加速引擎，实现“定义即研发”



- **底层选择 StarRocks 作为执行引擎：**大数据量即时计算、跨表关联查询、物化视图能力突出
- **集成最佳数据工程实践的物化加速引擎：**基于声明式的指标物化加速方案，系统自动化进行物化代码生成、编排调度和智能查询路由，实现指标“定义即研发”，做轻数仓

自动化：集成最佳数据工程实践的物化加速策略

1 冗余维度打宽



预打宽：根据语义数据模型，将常用的维度与明细事实表进行关联打宽

订单宽表

物流宽表

2 同事实同实体合并计算

物化表 T1

维度：日、地区、城市...

指标：订单笔数、订单人数

物化表 T2

维度：日、商品ID、品类....

指标：订单笔数、订单人数

物化表 T3

维度：日、商品ID、品类....

指标：发货订单数

3 长周期依赖短周期

物化表 T4

维度：日、地区、城市

指标：近1日订单笔数

物化表 T5

维度：月、地区、城市

指标：当月订单笔数

4 细粒度上卷聚合计算

物化表 T6

维度：a、b、c、d

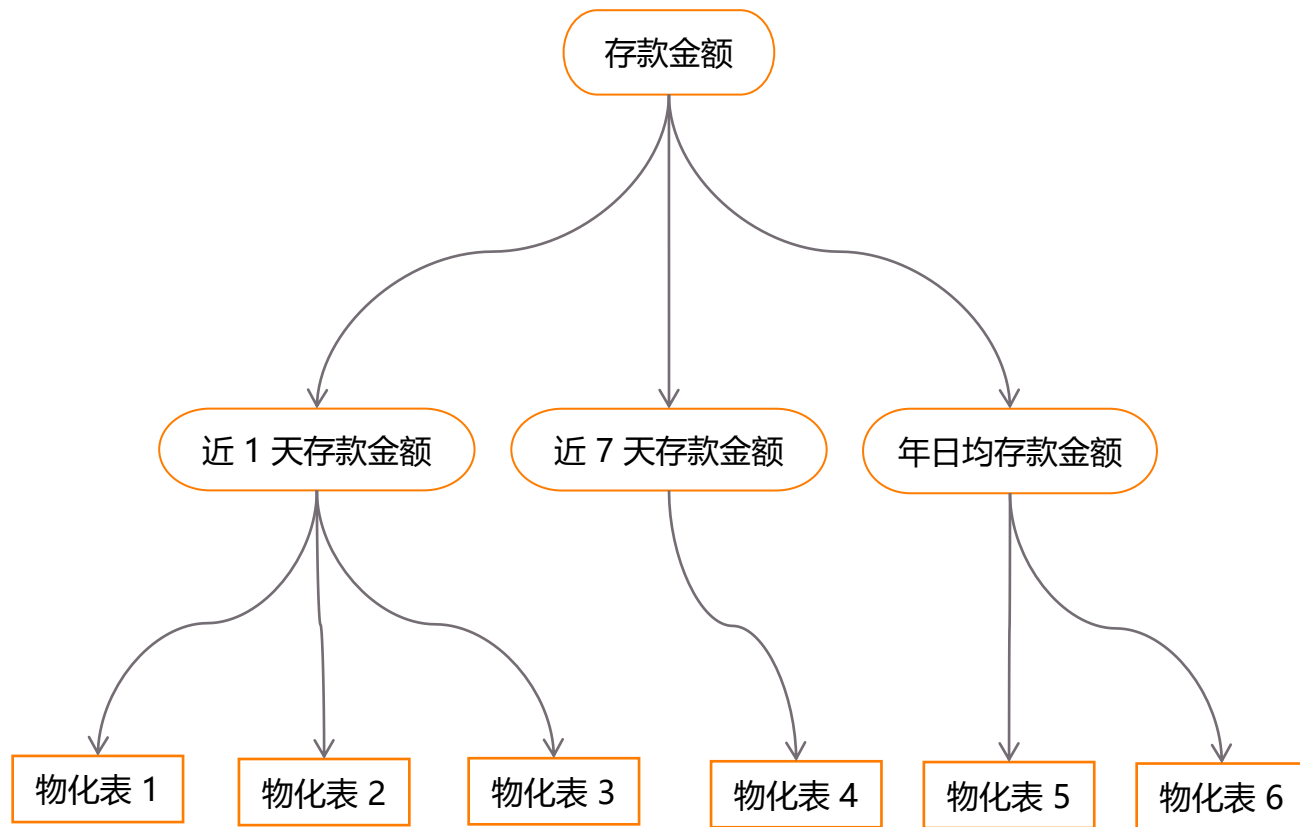
指标：近1日订单笔数

物化表 T7

维度：a、b

指标：近1日订单笔数

自动化：集成最佳数据工程实践的自动变更刷新



自动识别变更影响

当数据口径或指标口径变更时，自动识别出受影响的作业脚本和数据



自动更新作业脚本

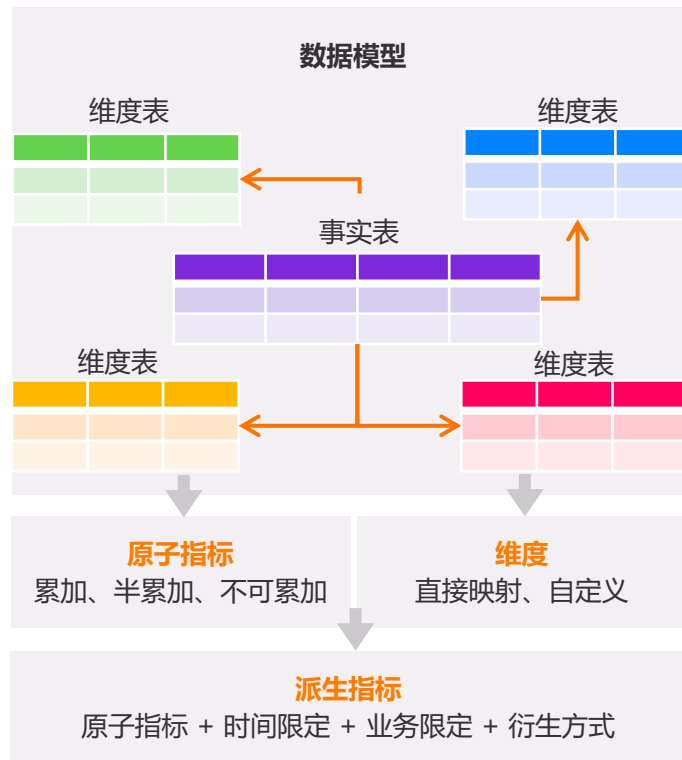
根据指标新的口径，自动生成并变更作业脚本



自动化指标回刷

根据用户配置的指标刷新范围，自动化生成刷新脚本并执行刷新

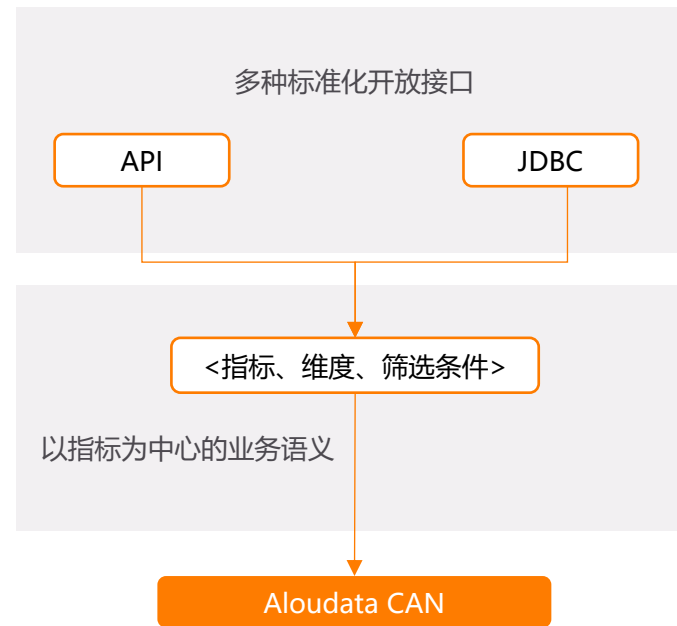
Aloudata CAN 整体运作机制



1 规范化指标定义



2 自动化指标生产



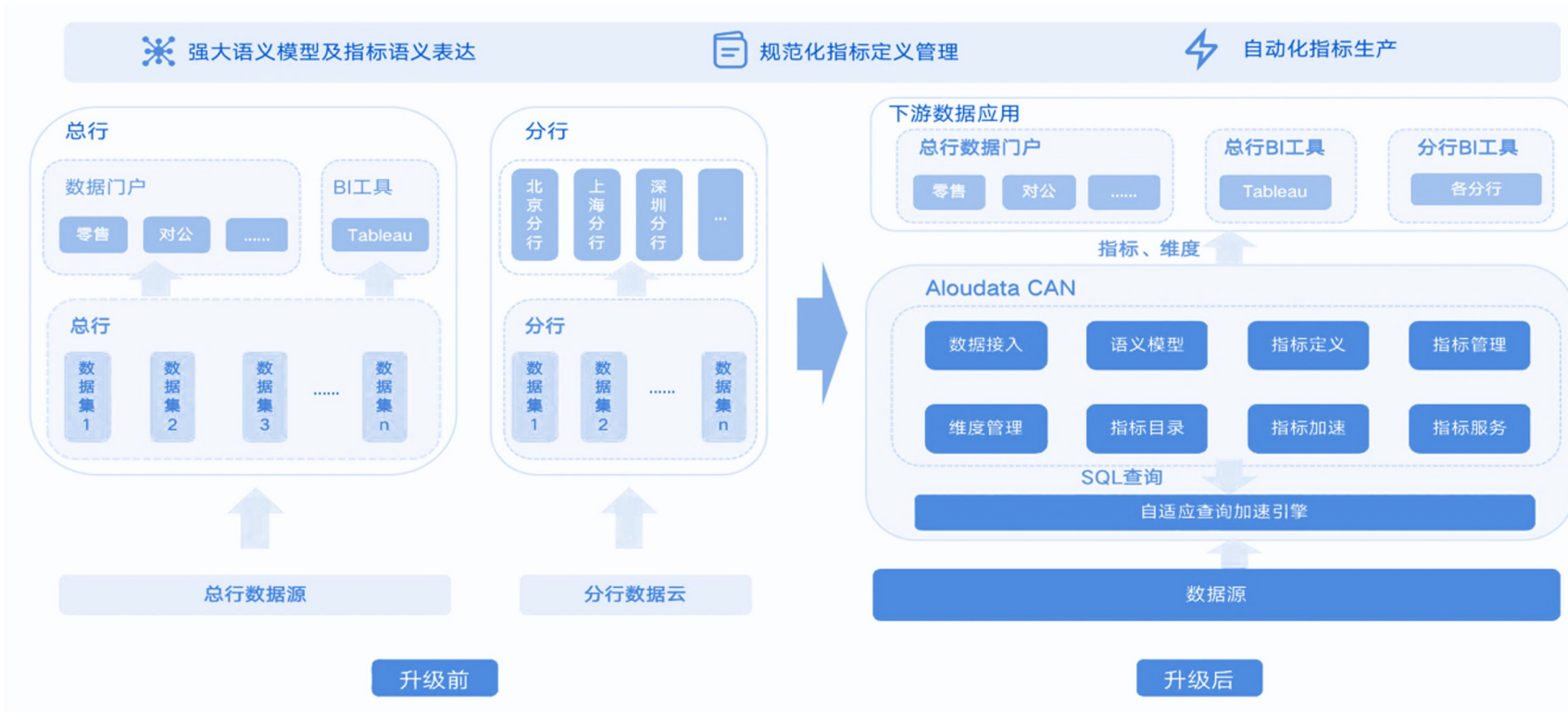
3 开放化指标服务

总结：Aloudata CAN 与第二代指标平台的核心差异

第二代指标平台	vs	Aloudata CAN (第三代)
基于单表定义指标，且定义能力弱，仅支持 sum、count、countd	指标定义	基于多表关联的语义模型定义指标，且定义能力强，支持二次聚合、自定义表达式等复杂指标定义
<ul style="list-style-type: none">- 基于数仓宽表和汇总表自动生产- 仅支持同环比有限衍生方式	指标加工	<ul style="list-style-type: none">- 基于公共层或集市层明细表自动生产- 支持同环比、均值/最值、排名、占比、累加等复杂衍生方式
<ul style="list-style-type: none">- API 接口参数简单- 一个接口仅支持查询一个指标，灵活性差	指标服务	<ul style="list-style-type: none">- API 支持灵活的接口参数- 一个接口支持多指标多维度（共同维度）
有限维度（指标定义中选择的维度）与指标组合分析或下钻分析	指标分析	任意维度（只需要数据集之间有关联关系）与指标组合分析或下钻分析
同一个指标不同维度需要在数仓加工多次，指标口径一致性难保障	指标管理	同一个指标不同维度只需要定义一次，指标共享复用高，100% 消除二义性

客户案例：助力某股份制银行业务用数效率 10 倍提升

改变该行指标开发模式，从原来 ETL 人员开发宽表与汇总表交付指标，升级为**业务定义指标自动化生产**



10 倍指标开发效率提升

标准化指标定义和自动化指标生产，实现指标开发模式的创新，指标交付效率从2周以上缩短到分钟级

10 倍业务用数效率提升

支持用户从任意维度、任意颗粒度灵活分析，无需等待 IT 排期

总分机构指标统一复用

将BI工具中定义的指标下沉至指标平台，实现跨 BI 工具、跨总分机构、跨业务条线的复用统一

Aloudata CAN
产品体验试用



Thank you.

fangxue.du@aloudata.com



扫码关注
Aloudata 公众号