

2019

05

08-10

北京新云南皇冠假日酒店

数据风云 十年变迁

DTCC 第十届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2019

IT168 ChinaUnix IT PUB

+

DTCC 2019

第十届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2019

苏宁数据中台架构实践

苏宁大数据中心

翟小青

01

总览

02

数据建模与指标化

03

维度管理

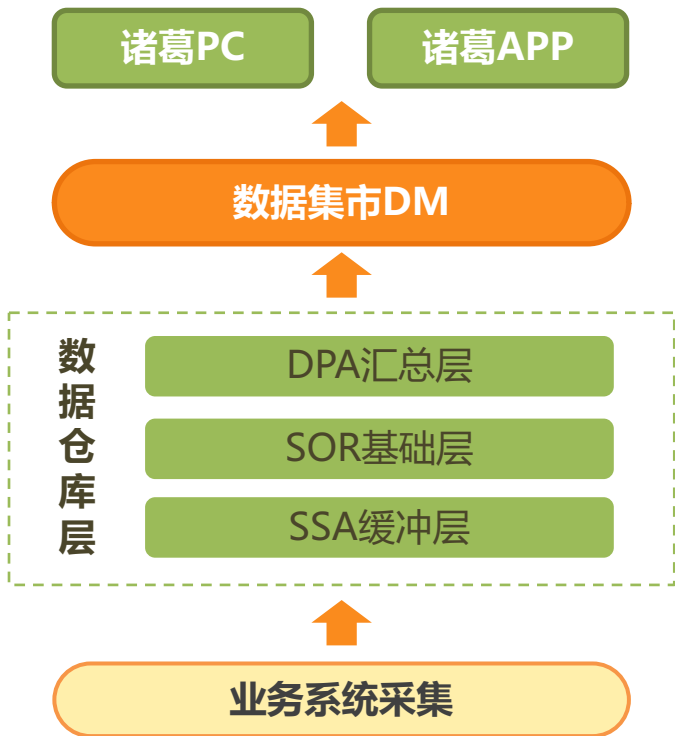
04

指标查询服务与OLAP引擎

05

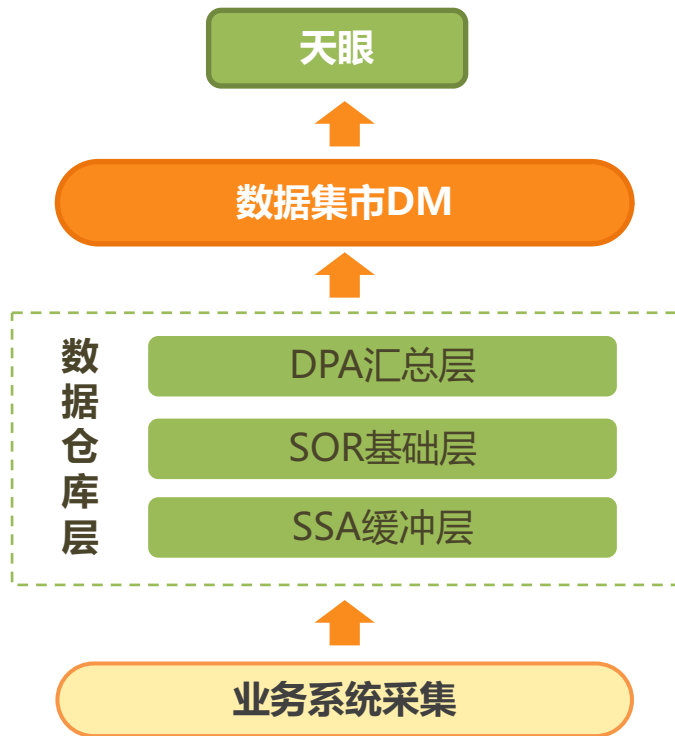
总结

销售域



数据中台出现之前

物流域



数据平台 VS 数据中台

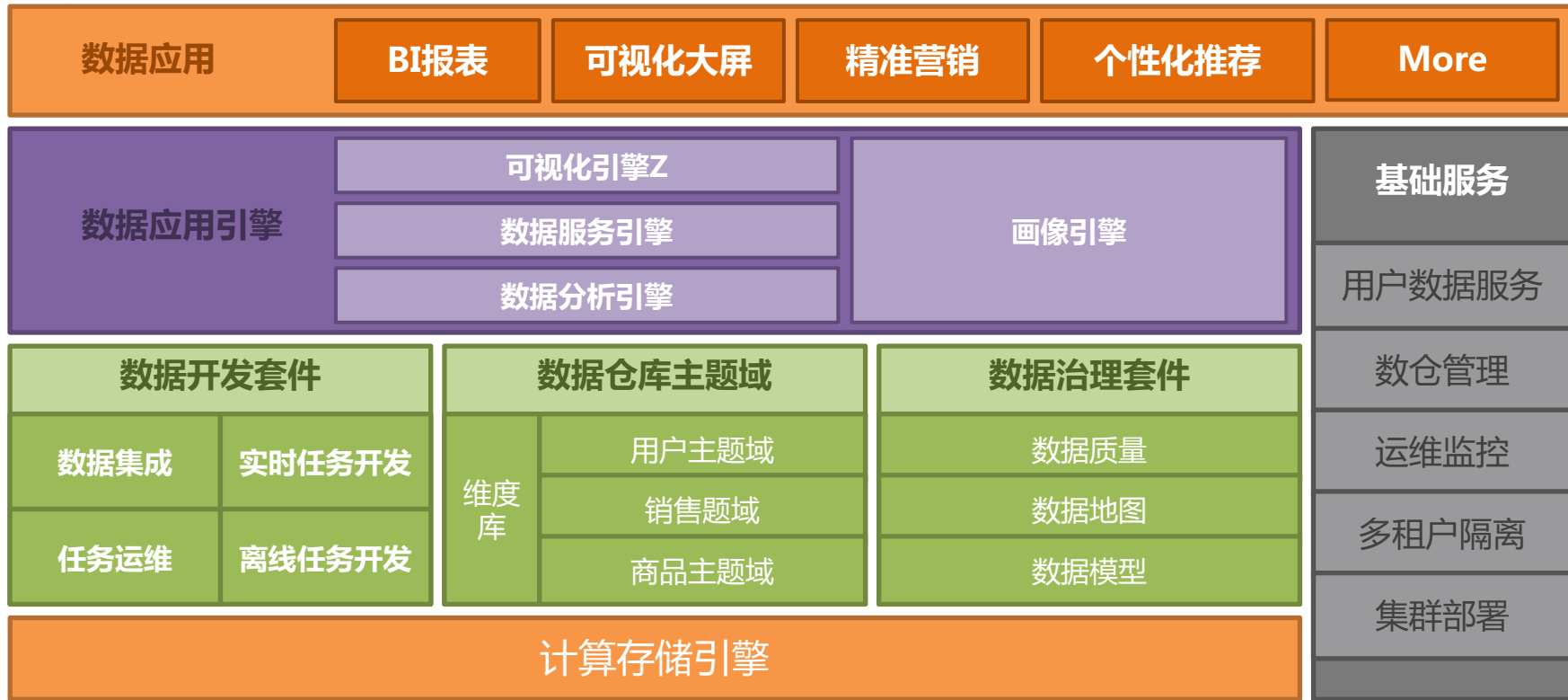
数据平台

- 有完整的数据模型设计，但偏重设计和技术，在执行过程中，很难保证数据的全，数据应用一般不跨过数据中心
- 初期数据发展快，效率高，快速体现业务价值，但是随着数仓的建设，数据量急速叠加，整体成本居高不下，导致数据混乱、灾难。

数据中台

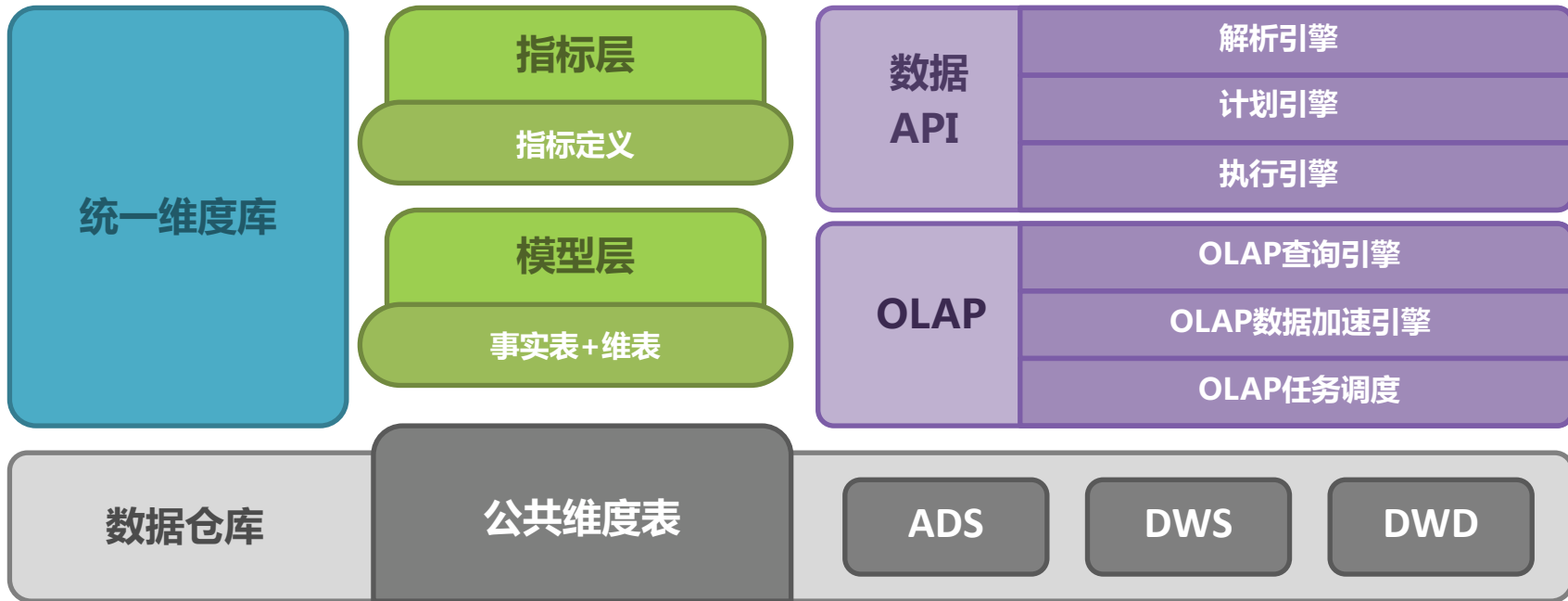
- 数据中台的基本理念是：将所有数据汇聚到数据中台，每个数据应用都以数据中台为唯一数据来源。
- 苏宁数据中台的目标是为苏宁的数据战略提供有力的支撑，从企业全局进行统一规划，统一建设，强调数据的“全”，从设计、组织、建设、流程角度保障了模式的落地。
- 数据中台的建设减低了数据使用门槛

数据中台系统架构

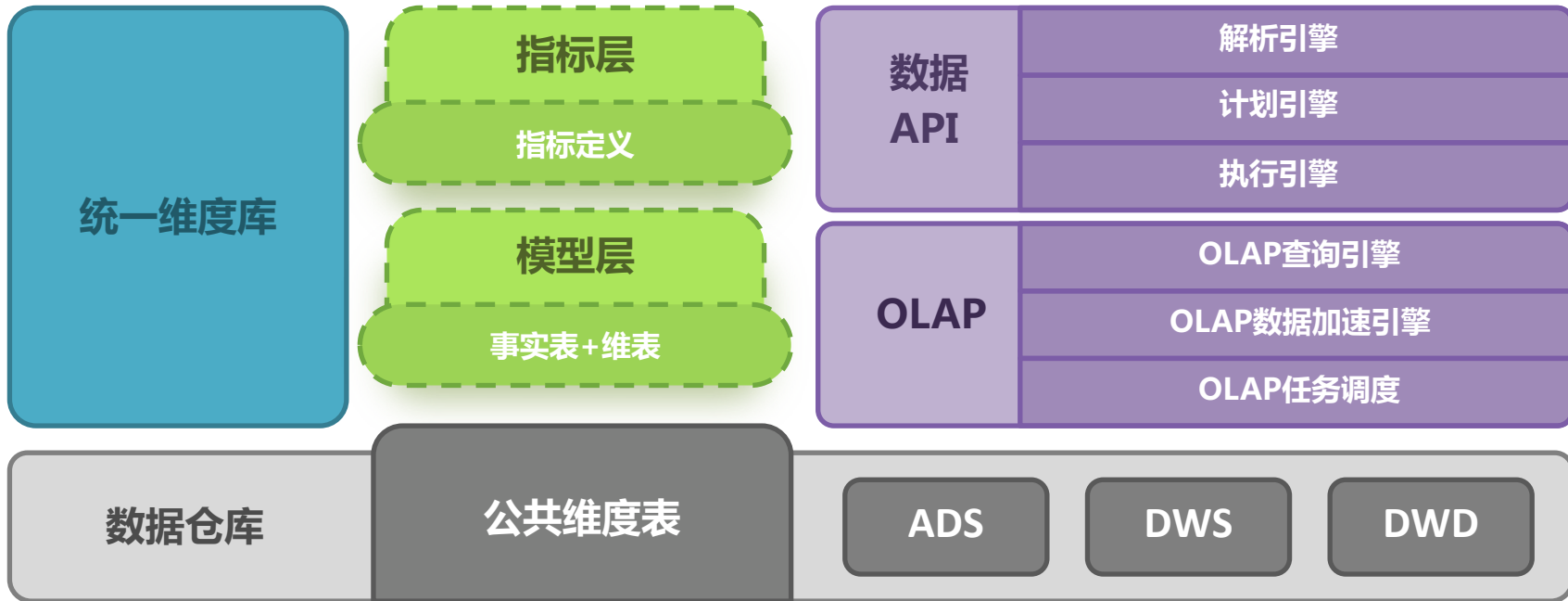


模型-指标-报表体系系统架构

天工数据层：可视化引擎



天工数据层：可视化引擎



01

总览

02

数据建模与指标化

03

维度管理

04

指标查询服务与OLAP引擎

05

总结

为什么是星型模型

避免数据冗余，
减少数据库的
空间

3NF

VS

反范式

OLTP

VS

OLAP

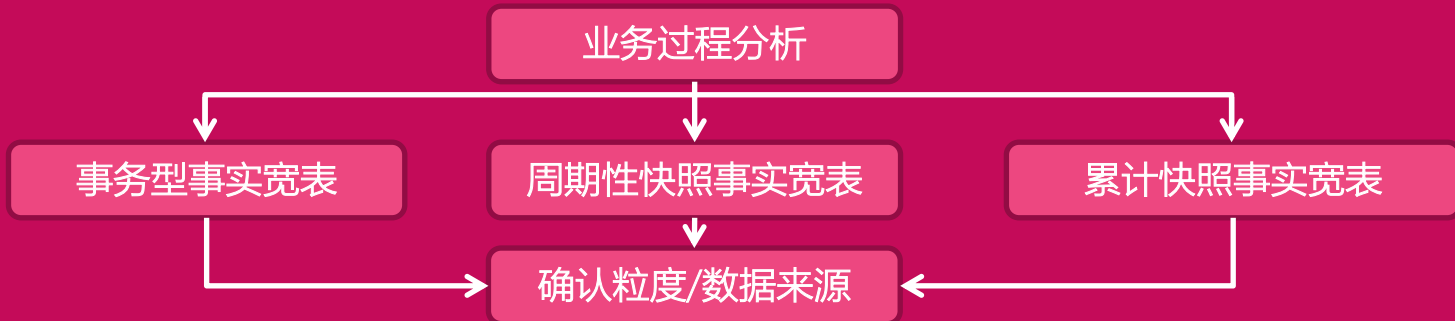
雪花

VS

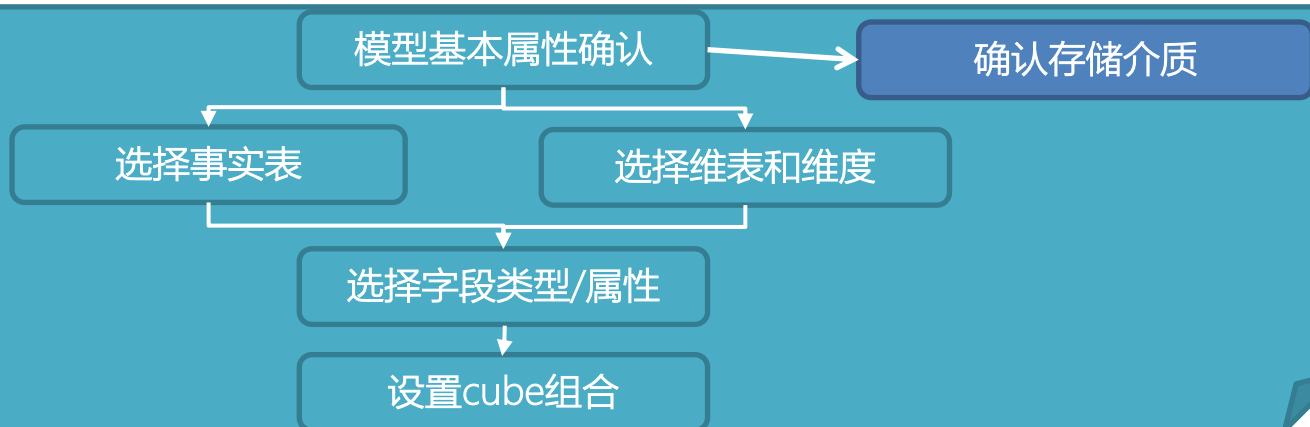
星型

允许数据适当
冗余，缩短操
作数据的时间，
用空间换取时
间

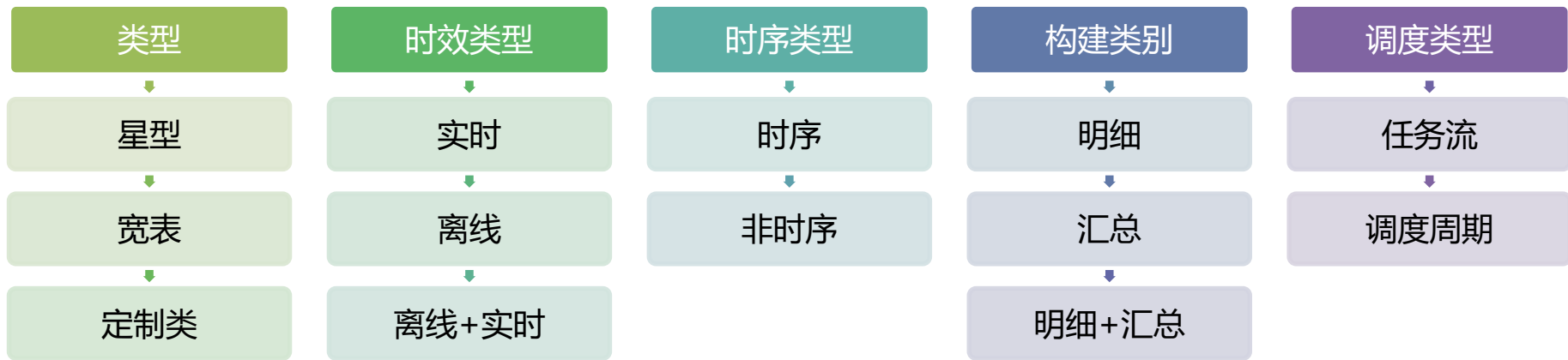
建模分析



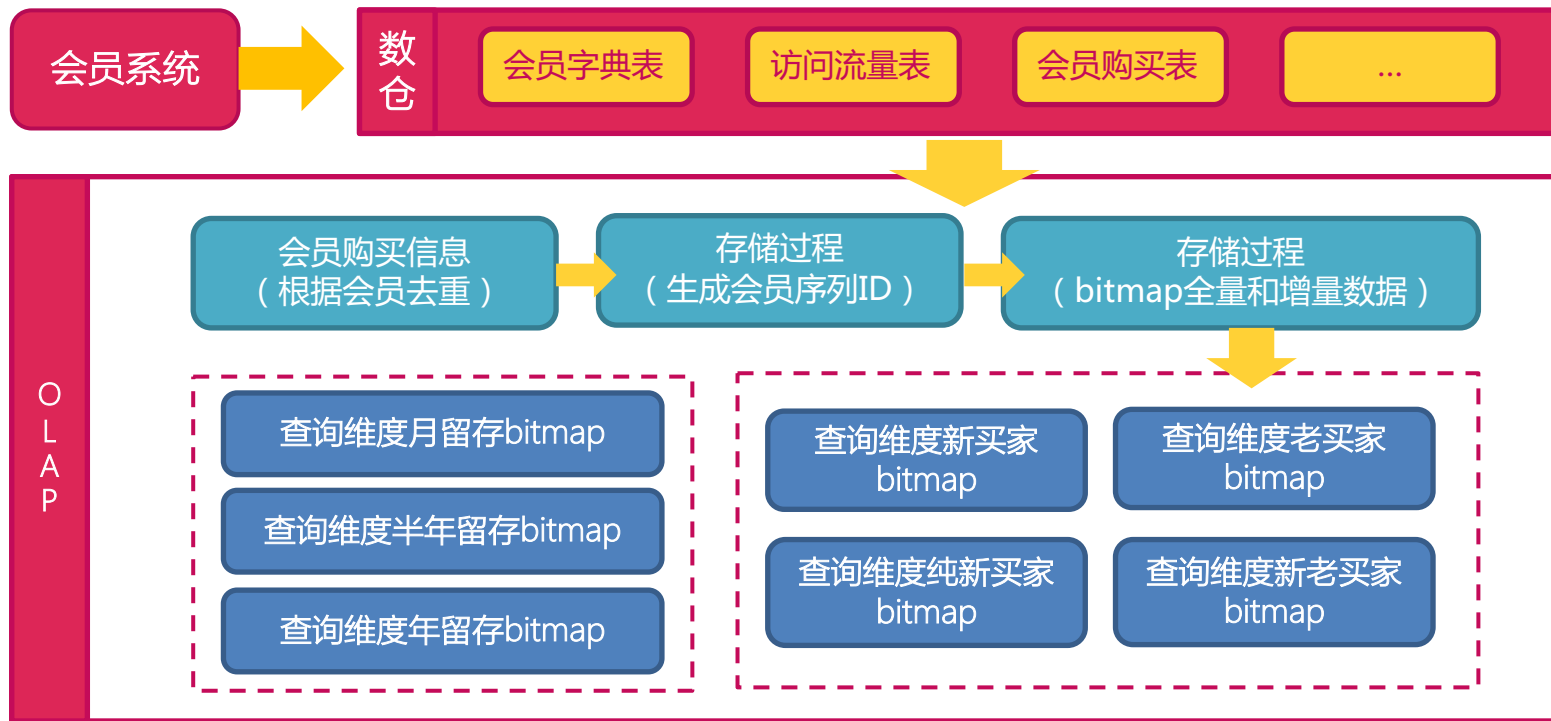
建模过程



模型基本属性







指标基础信息

- 多种时间粒度
- 多种时间周期
- 多单位换算

度量

- 计算函数：
max/min/count/count
distinct/sum/avg/abs
- 累计函数:lastday

衍生计算表达式

- 支持逻辑流
- 支持运算符
- 时间计算函数
- 异常数据

个性化筛选条件

- 统一维度支持
- 自定义维度支持
- 自定义参数支持

指标属性

- 可比
- 占比
- 同环比
- 均值

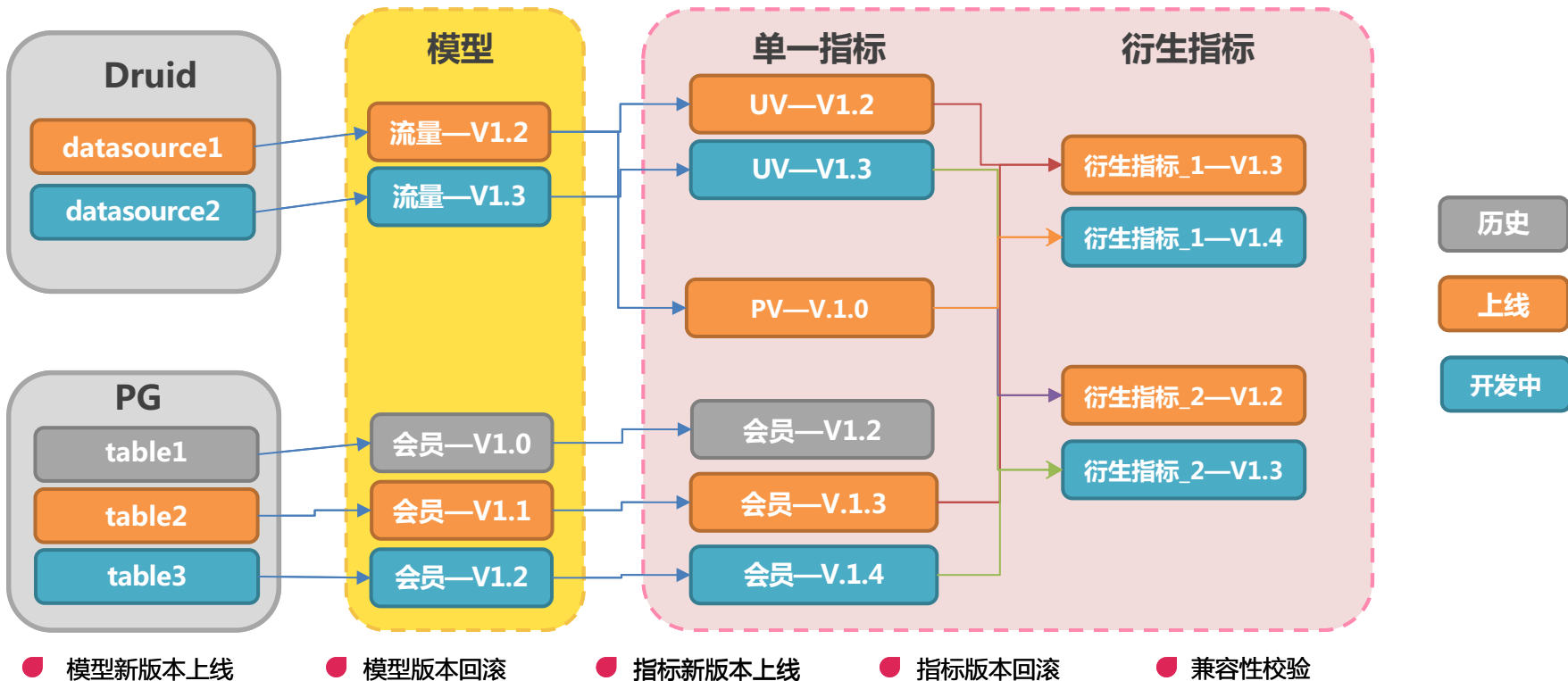
可视化测试

- 小天工
- 多维度数据验证
- 性能验证

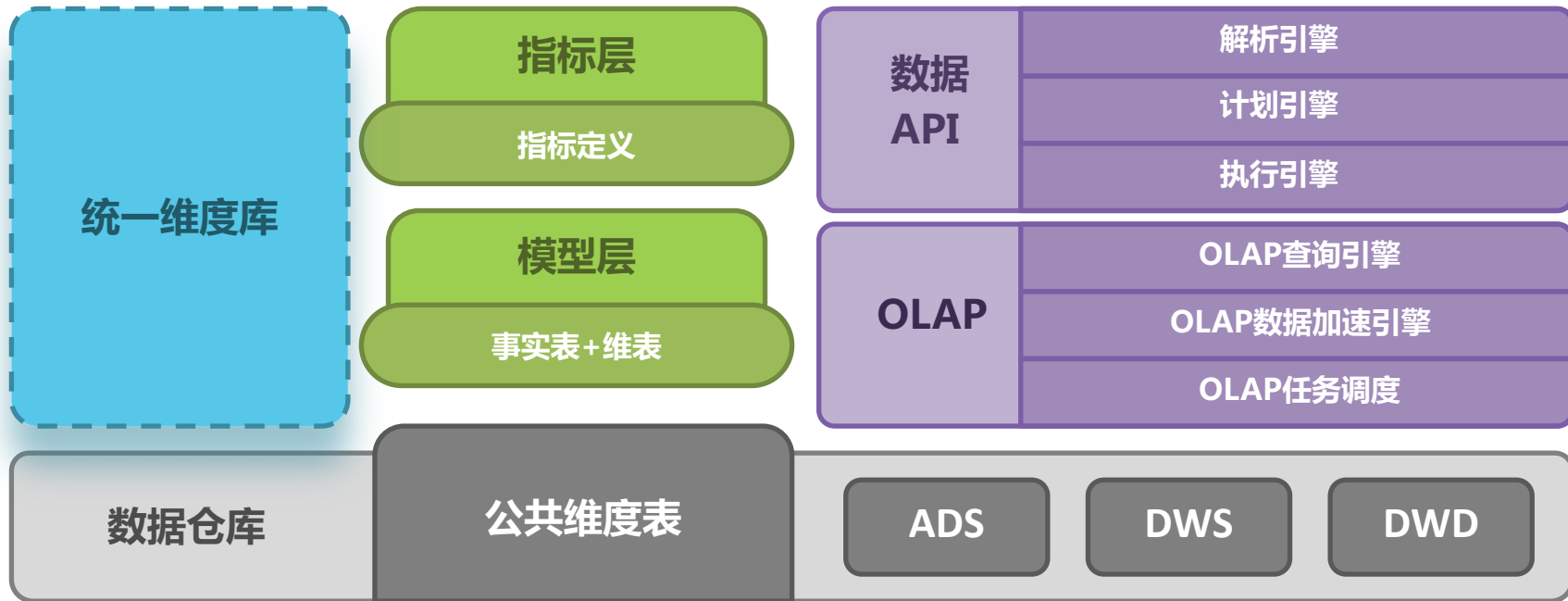
原则

- 状态分成上线/历史/开发中
- 上线单一指标来源于上线的模型
- 上线衍生指标中的单一指标必定是上线版本
- 历史版本模型有冻结期，冻结期结束此版本模型删除，关联历史指标下线

模型、指标多版本体系



天工数据层：可视化引擎



01

总览

02

数据建模与指标化

03

维度管理

04

指标查询服务与OLAP引擎

05

总结



统一维度建设背景

标准

维度业务口径不统一，缺乏有效的管理流程来对其进行管理和约束，维度建设存在重复和歧义



成本

对于维度服务的开发，各个产品中心需要各自实现，造成开发成本重复投入。例如公司的维度。



工具

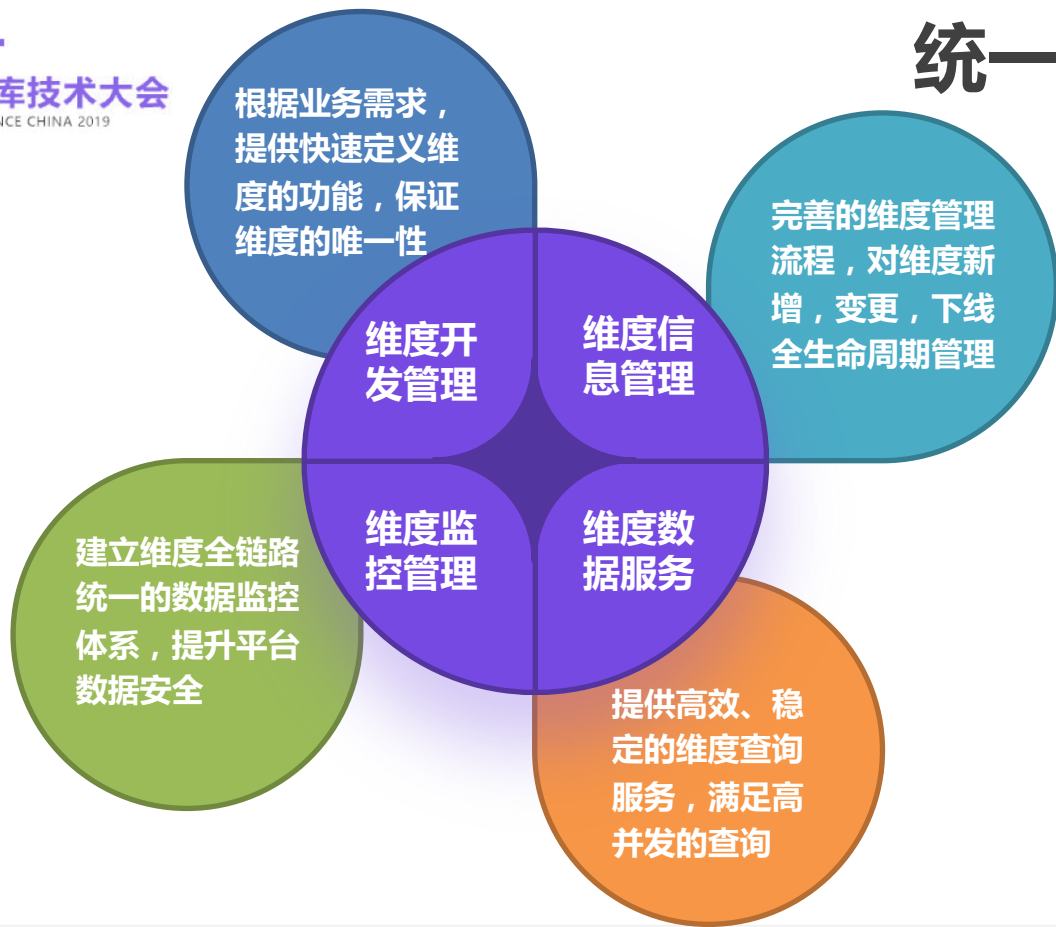
ETL开发人员/产品人员，存在大量手工配置表需要规则维护，缺乏快速开发工具。



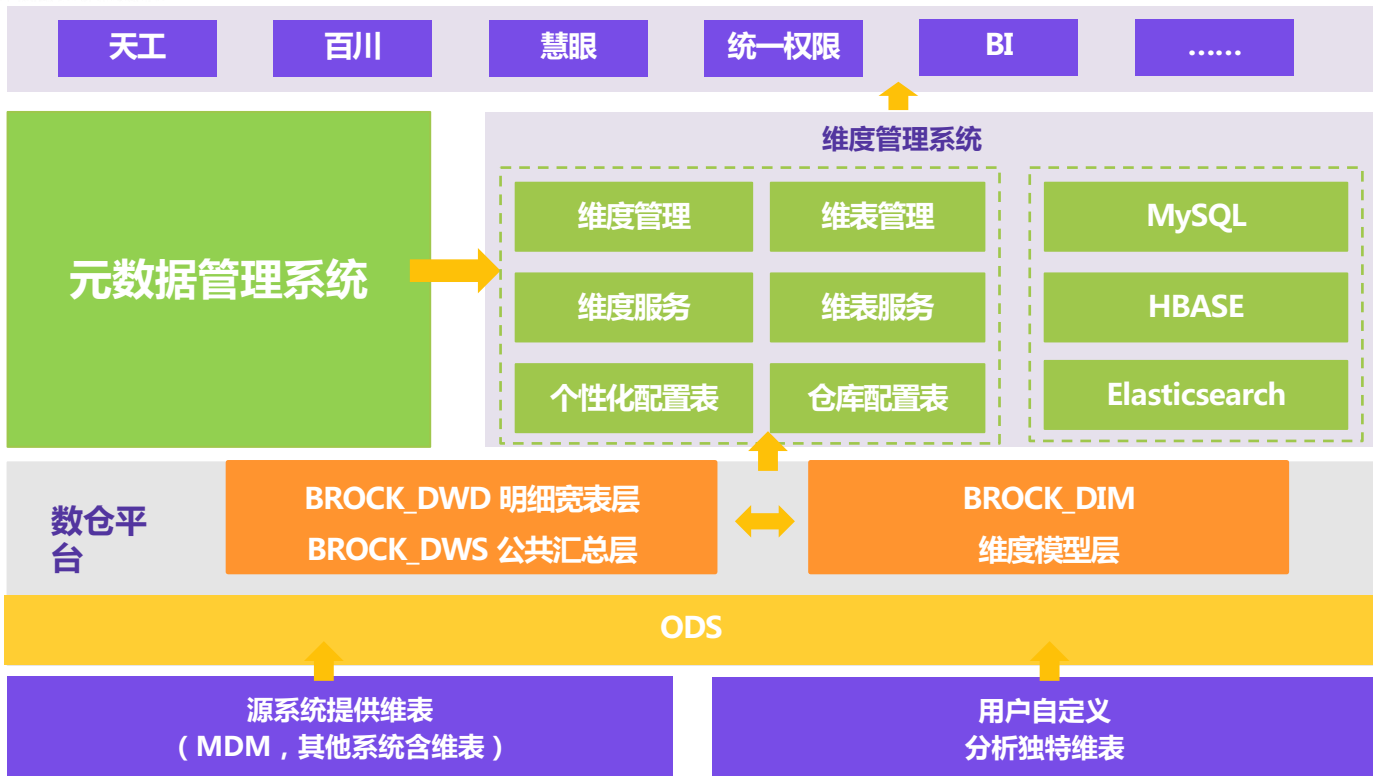
平台

业务人员想查询维度信息，缺乏可靠的公共平台去快速方便的查询。

统一维度管理目标



统一维度系统架构



维度类型

- 普通维度
- 父子维度
- 角色扮演维度
- 杂项维度
- 日历时间维度

维护分类

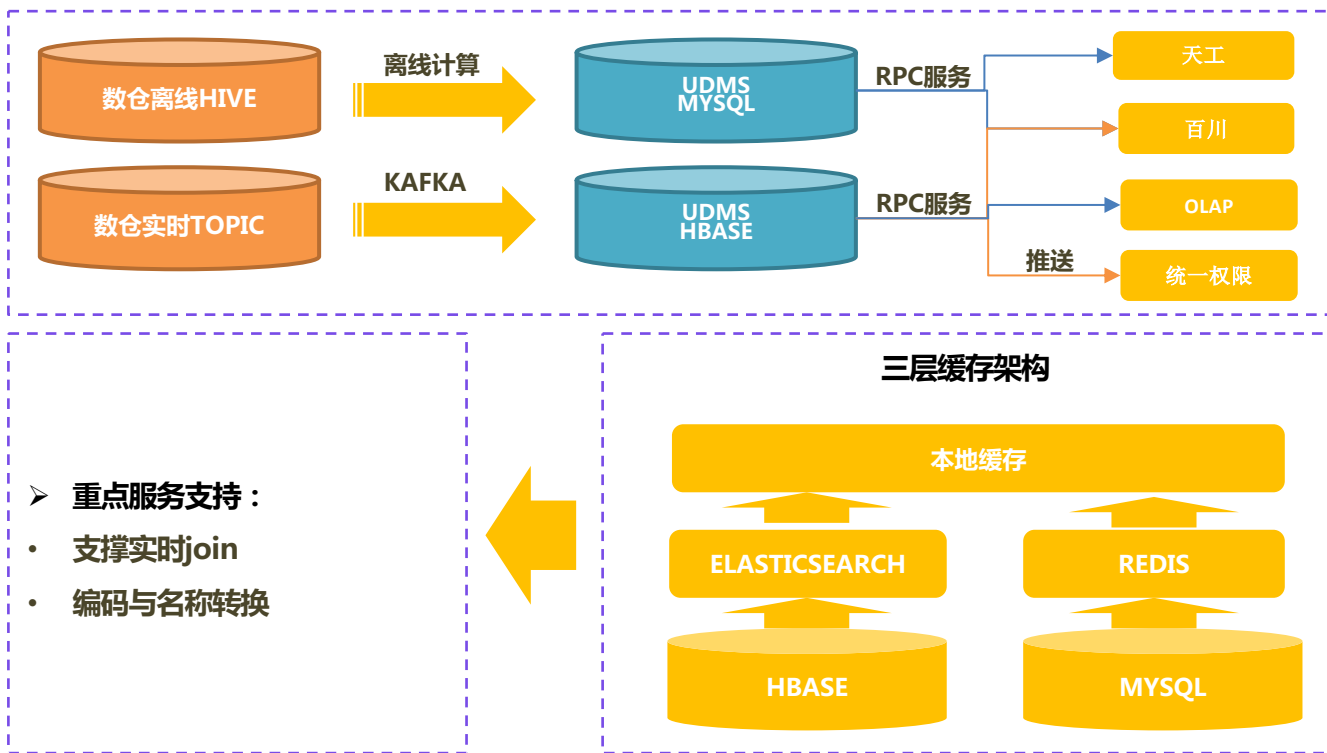
- 主数据维度
- 手工维度

时效分类

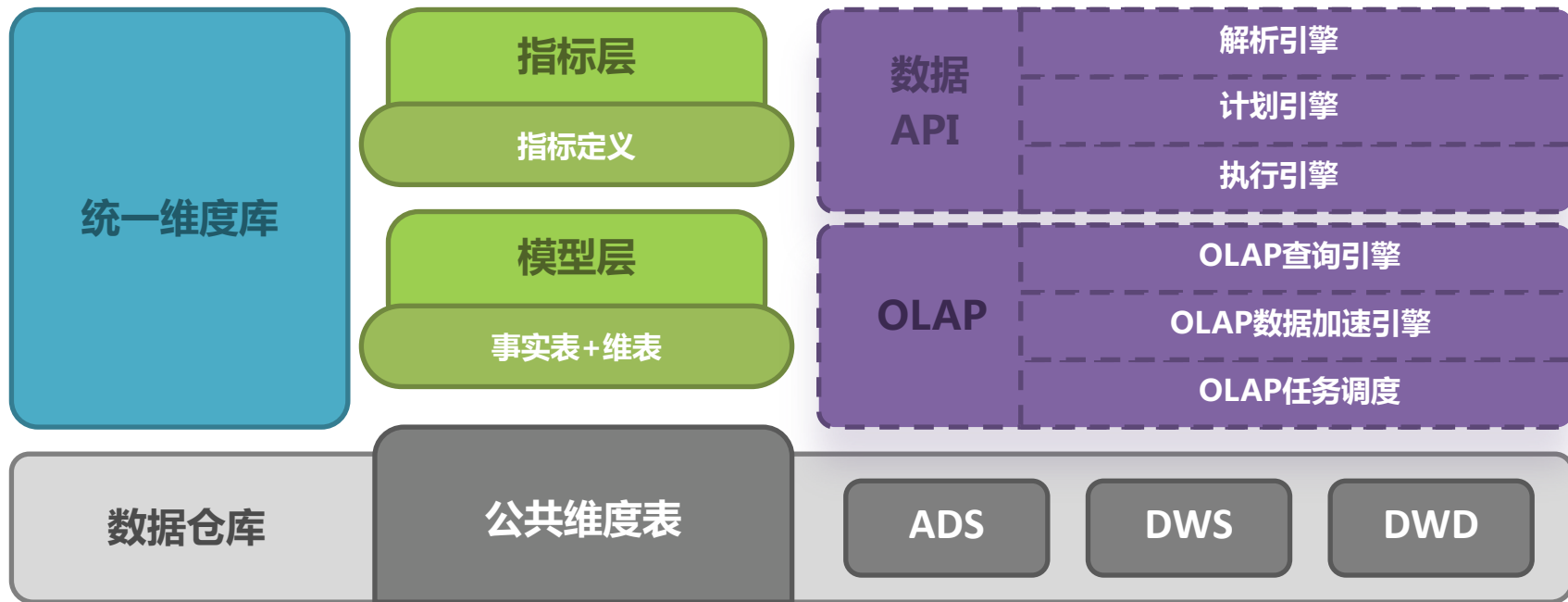
- 实时
- 离线

层次类型

- 层级维度
- 非层级维度



天工数据层：可视化引擎



01

总览

02

数据建模与指标化

03

维度管理

04

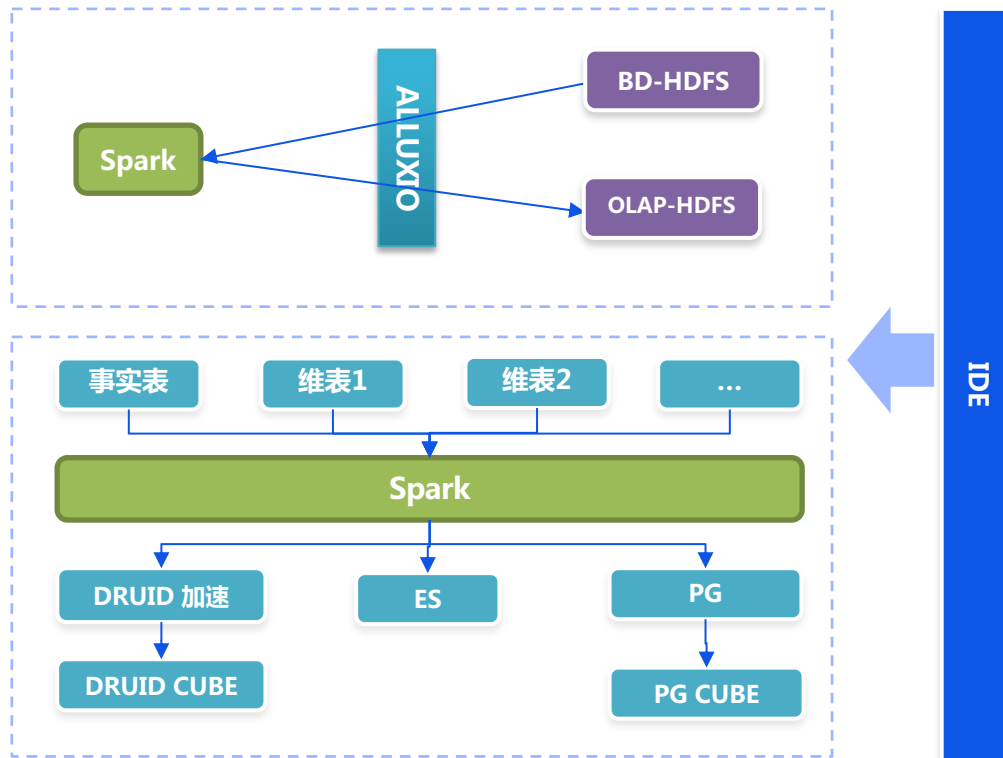
指标查询服务与OLAP引擎

05

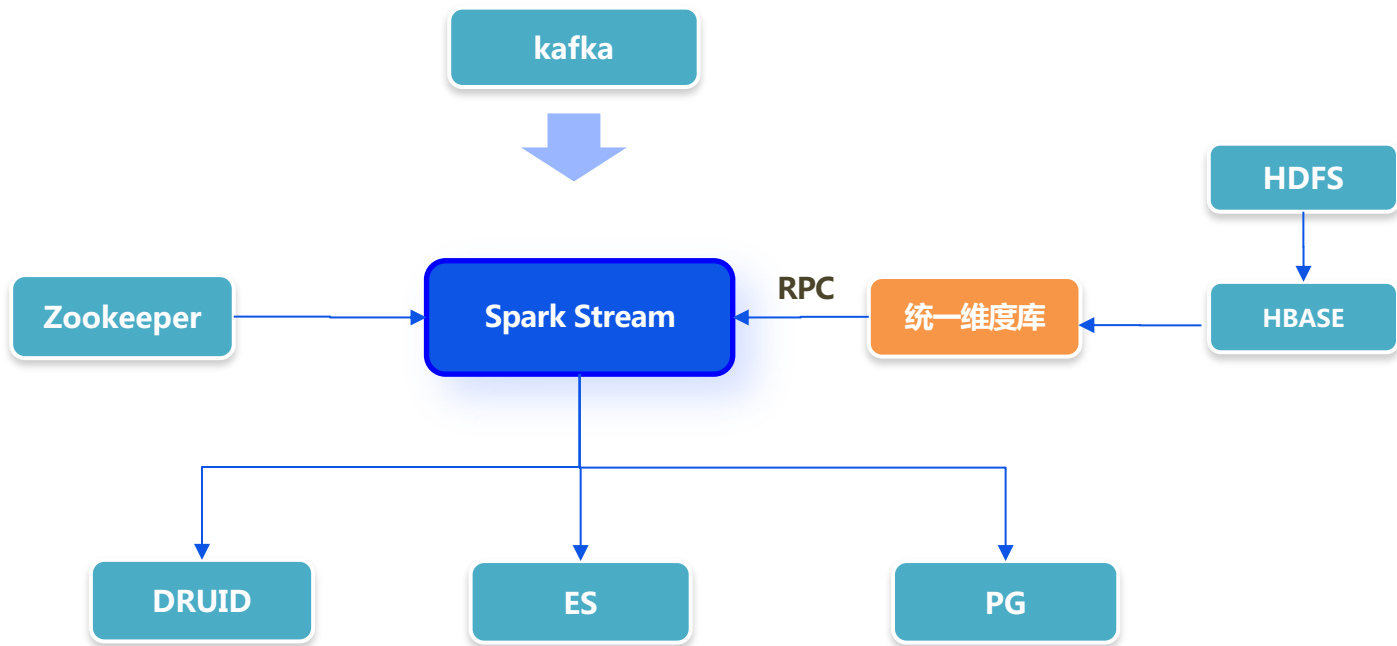
总结



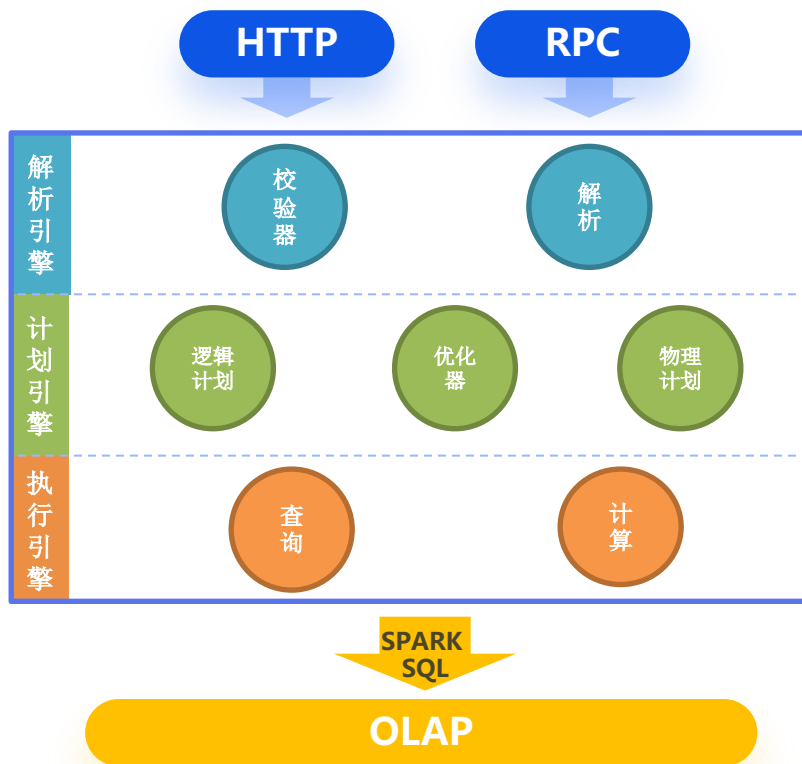
离线模型加速方案



实时模型加速方案

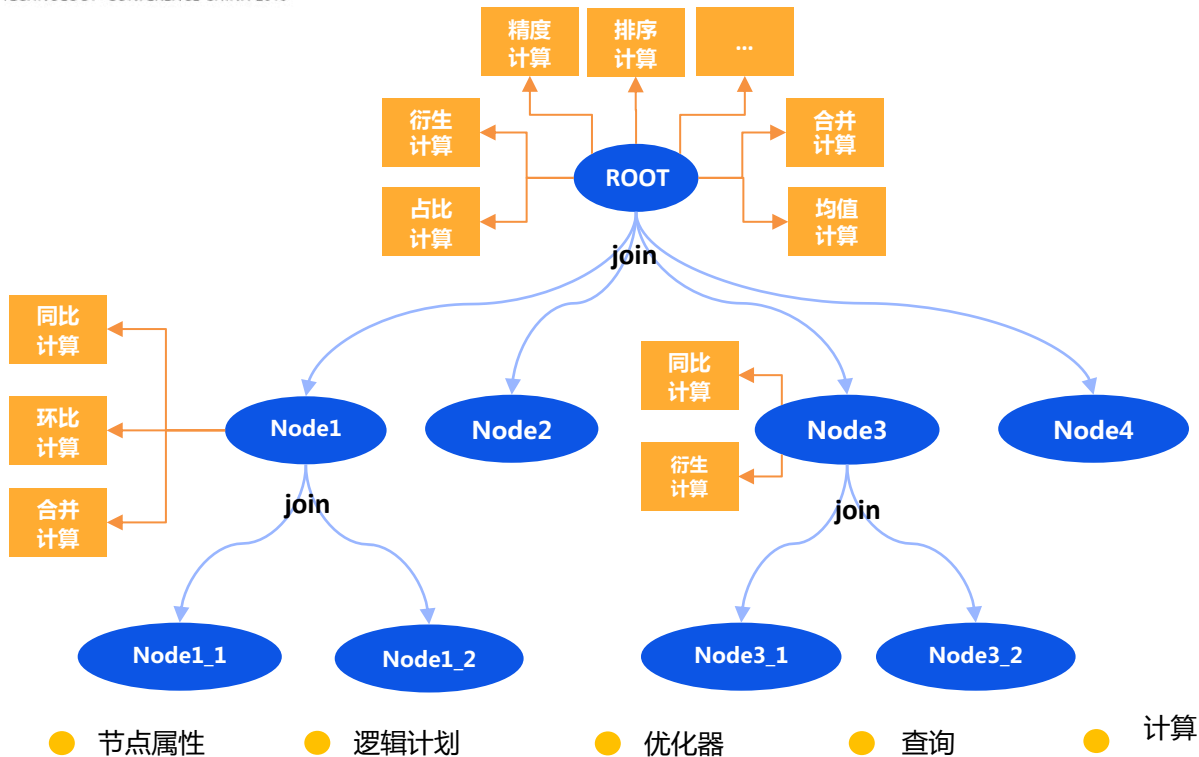


指标数据查询服务



- 基本规则校验后，将指标查询进行解析，获得解析器结果
- 按照模型，时间粒度，过滤维度等创建逻辑计划；
- 经过sort优化器，join优化器，cost优化器等优化器，优化执行计划
- 生成物理计划
- 多线程查询引擎
- 自动化计算器，包括：join计算器（衍生指标计算器，同环比计算器），占比计算器

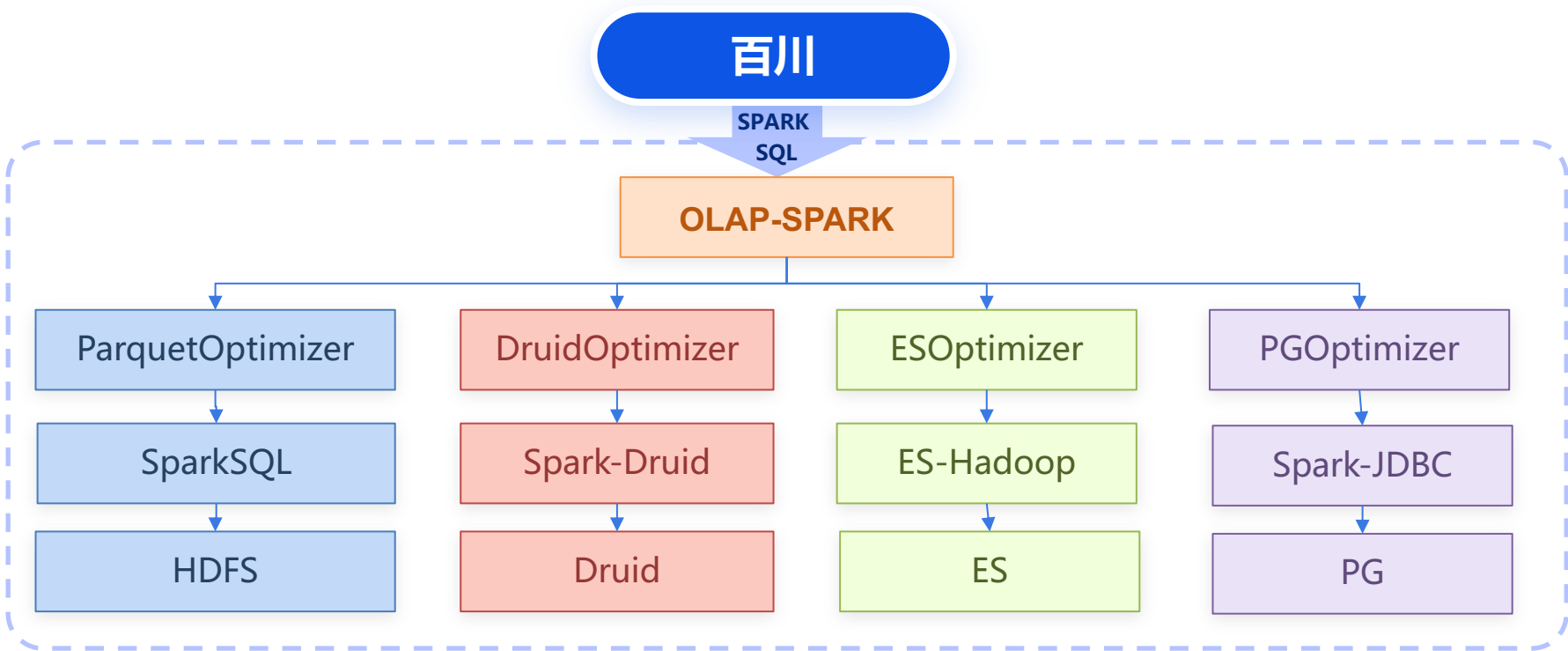
指标数据查询服务—执行计划



- 排序优化器
- Cube优化器
- Join优化器
- Cost优化器
- 路由优化器

- 计划节点
- 计算属性





01

总览

02

数据建模与指标化

03

维度管理

04

指标查询服务与OLAP引擎

05

总结



- 精确去重
- Cube自动优化
- DWS、ADS层优化

功能

- 高基数查询
- 深入COST模型
优化引导

性能

- 加密
- 解密

安全



THANKS