

# 指标中台核心技术 - 自动化建模与统一服务

---

陈硕 快手 指标平台技术负责人



# 关于我



## 陈硕

- ▶ 参与并负责快手统一指标中台技术建设
- ▶ 专注于指标中台，Headless BI，分析引擎等数据管治与数据分析方向

# 目录 CONTENT

**01** 快手指标中台总体介绍

**03** 统一指标服务核心技术

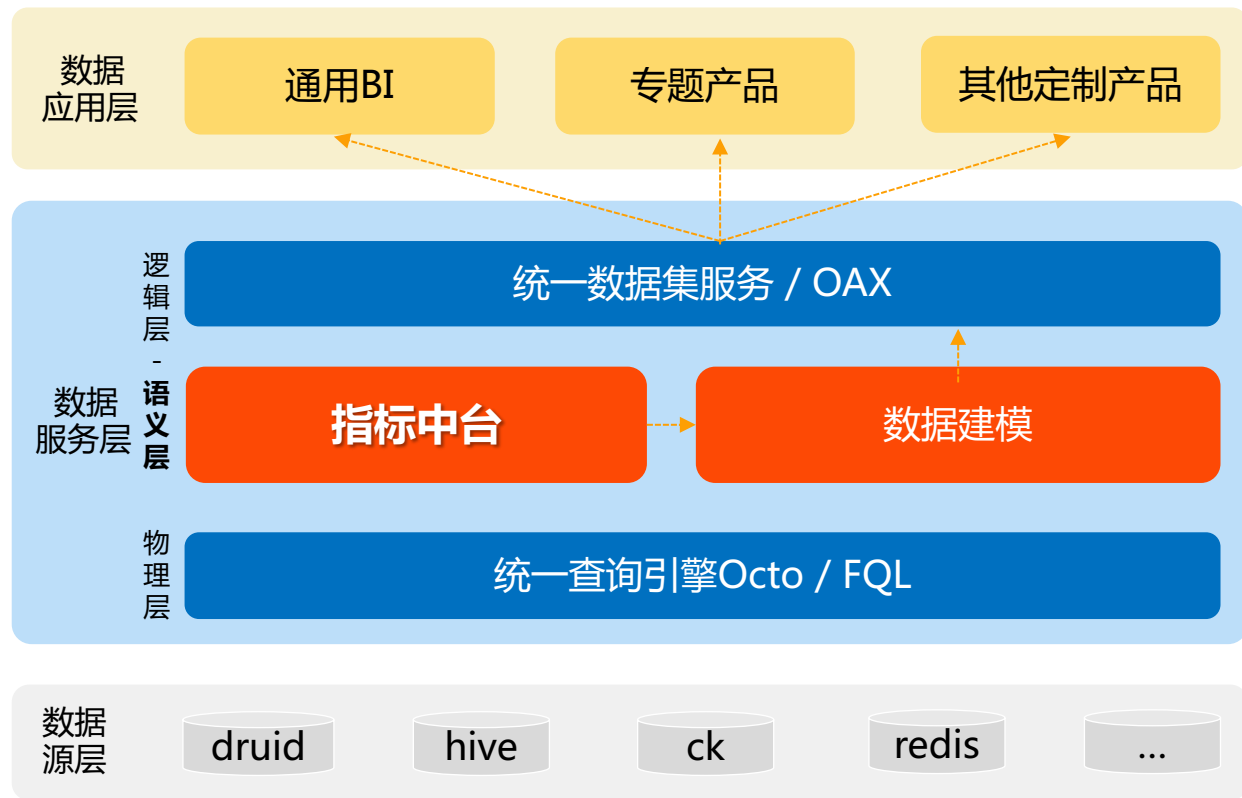
**02** 指标标准化管理

**04** 未来展望

# 01 快手指标中台总体介绍

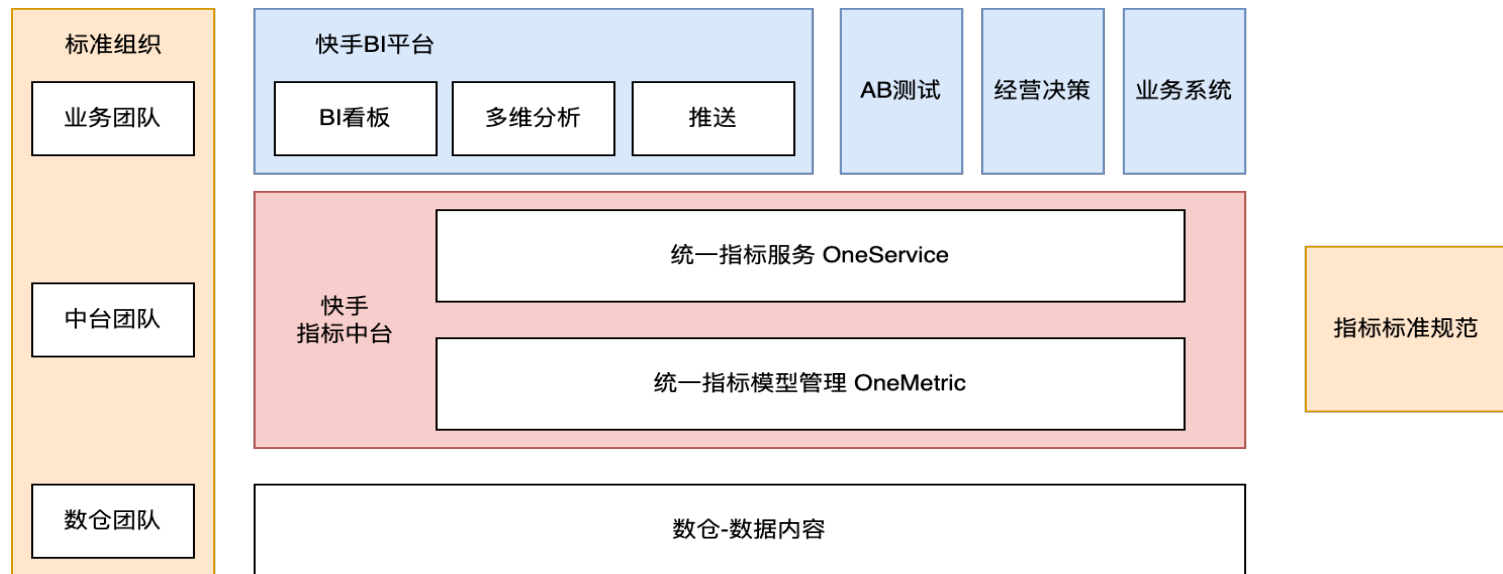


# 快手指标中台的定位



# 快手指标中台总体架构

**快手指标中台**，其核心设计思想是以指标管理驱动数据服务，实现了从技术语言到管理语言的抽象，对外提供统一的指标管理以及统一的指标服务能力，进而达到“**一处定义、多处使用**”。

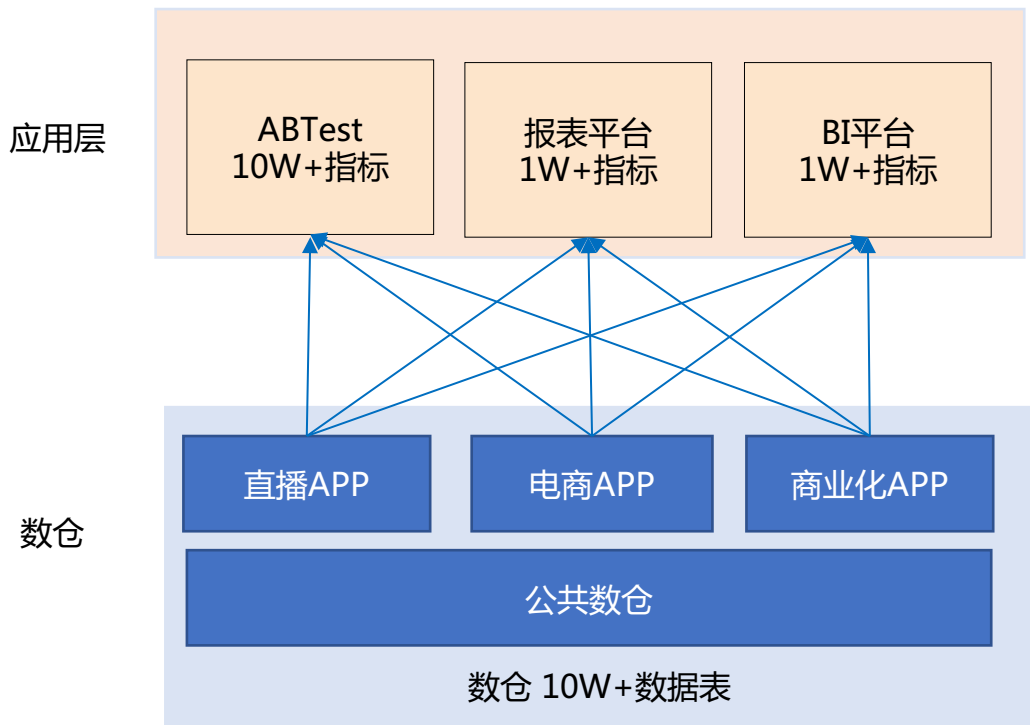


## 02 指标标准化管理



# 快手指标管理面临的问题

## 快手指标标准化管理之前...



- **指标管理不统一**

- 多处管理：成本较高
- 重复管理：指标泛滥

- **指标口径不统一**

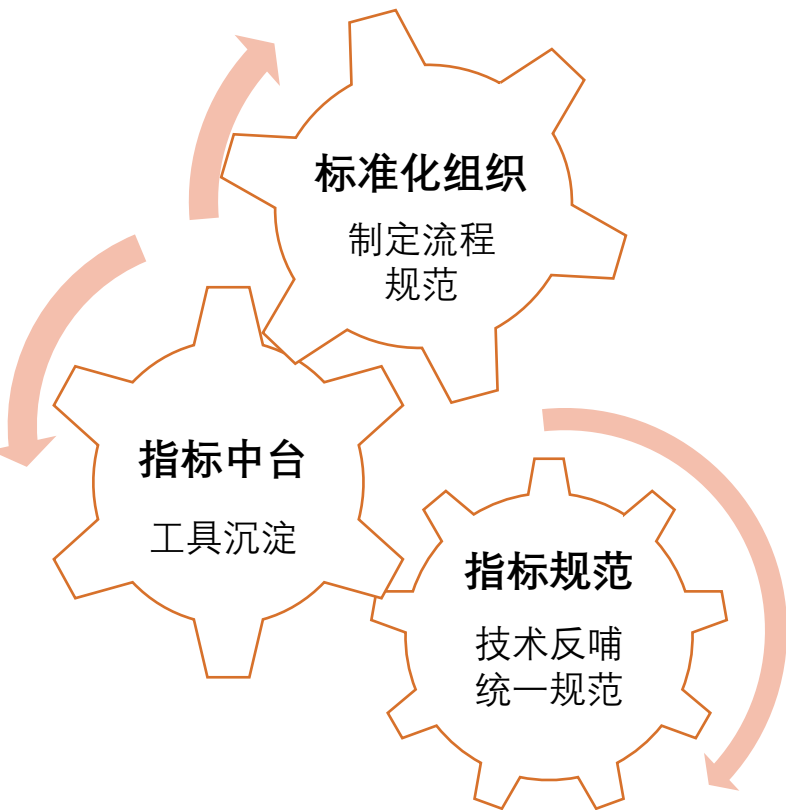
- 同名不同义：口径不一致
- 同义不同名：命名不规范

- **指标流程不统一**

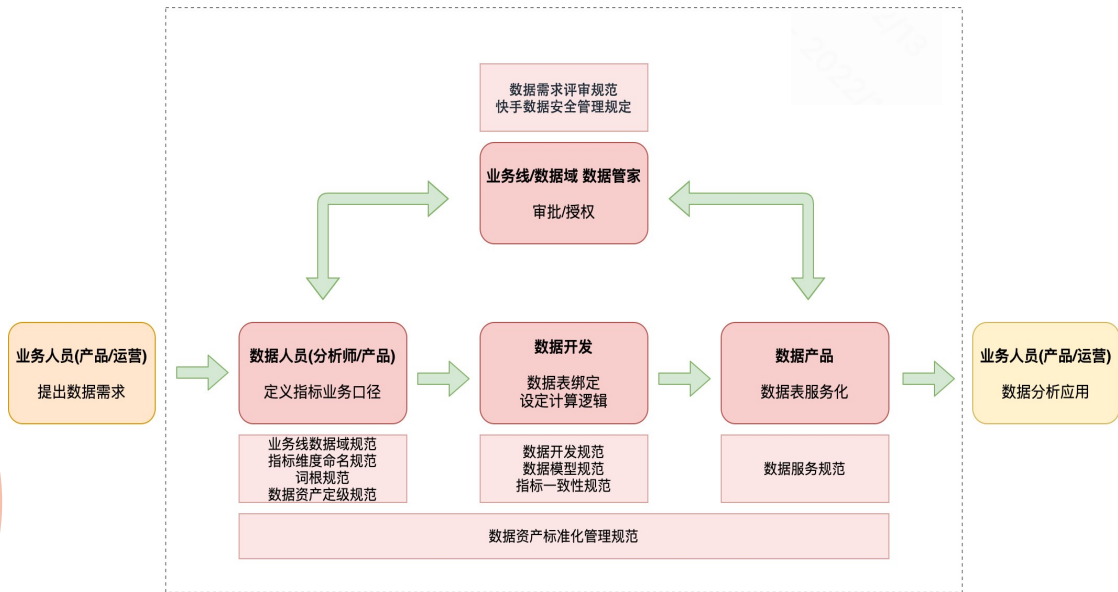
- 系统独立：管理不统一、流程不统一



# 方法论 - 指标标准化



## 指标标准化业务角色职责



# 指标体系标准化管理规范

## OSM指标拆解

### 业务目标 (O)

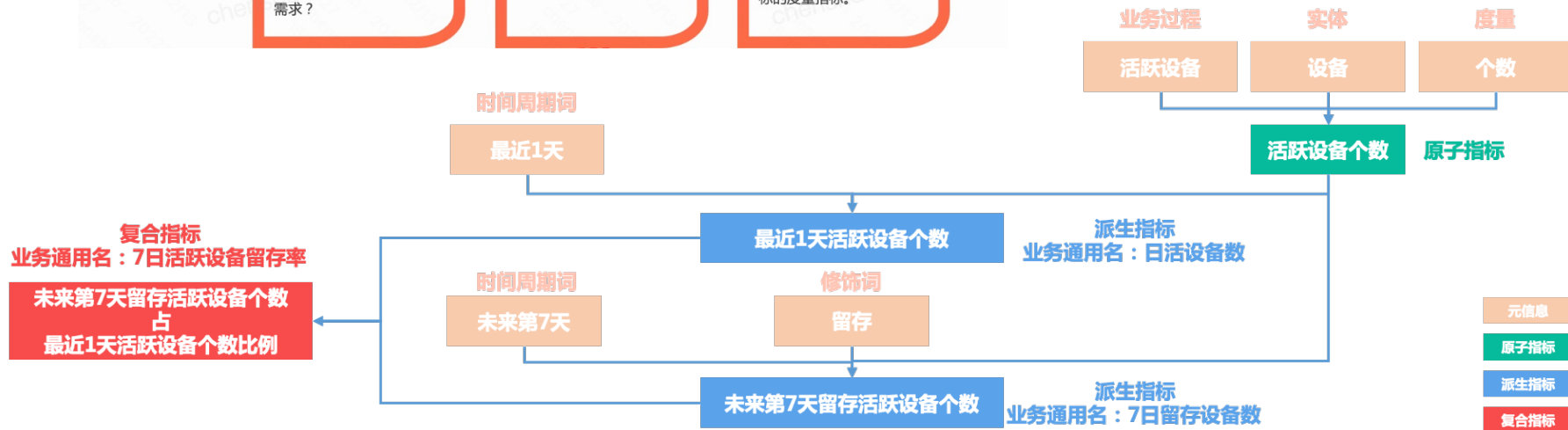
我们的业务、产品，存在的目的是什么、能够解决用户什么问题、满足用户的什么需求？

### 业务策略 (S)

为了达成上述目标，相应采取的业务策略有哪些？

### 业务度量 (M)

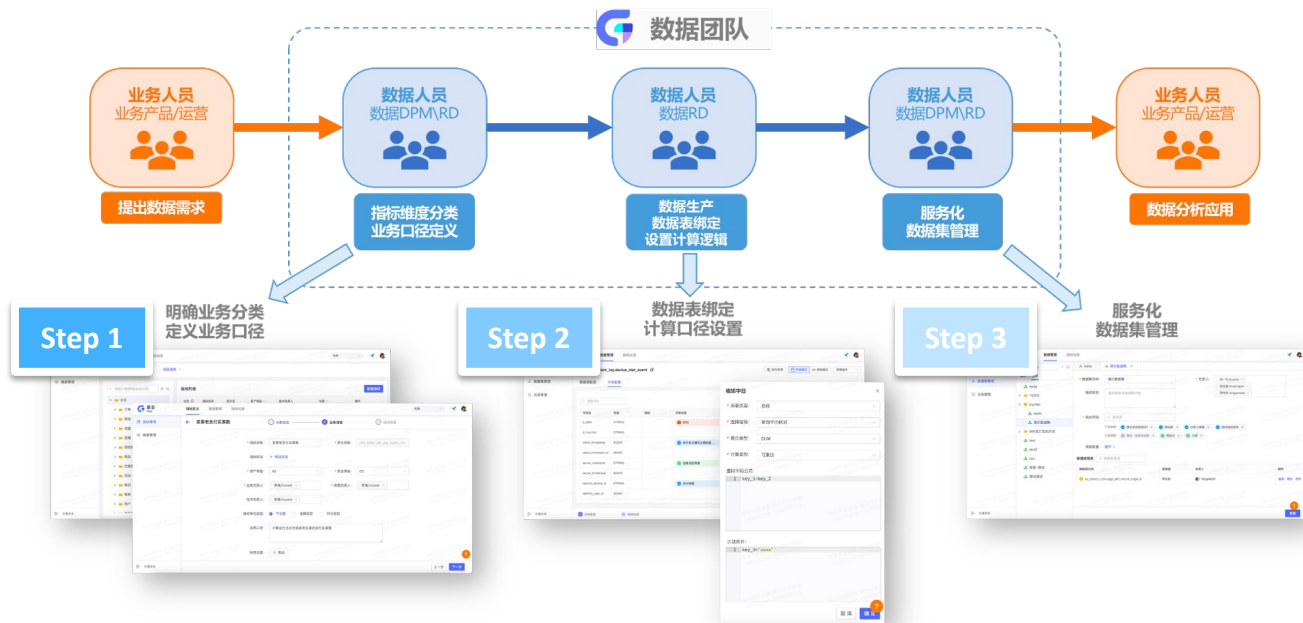
反映业务策略有效性、反映策略执行是否能达成业务目标的度量指标。



# 指标标准化流程

## 指标标准化流程

1. 定义指标业务口径
2. 定义指标计算口径
3. 定义数据集



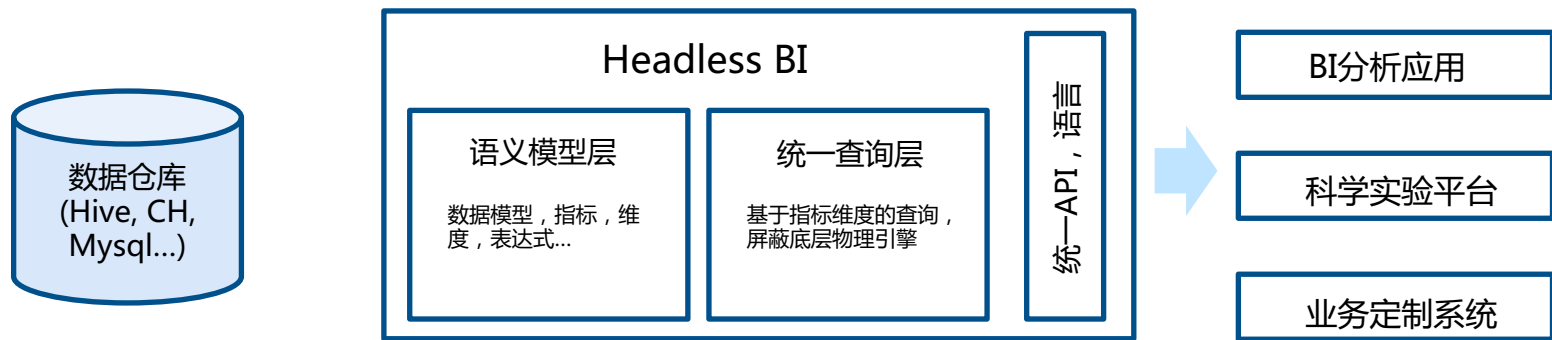
## 03 统一指标服务核心技术



# 统一指标服务核心技术

- 设计理念：**Headless BI**
- 核心技术1：**自动化建模**
- 核心技术2：**统一指标服务**

# Headless BI - 理念与核心技术



**Headless BI**  
**三要素**

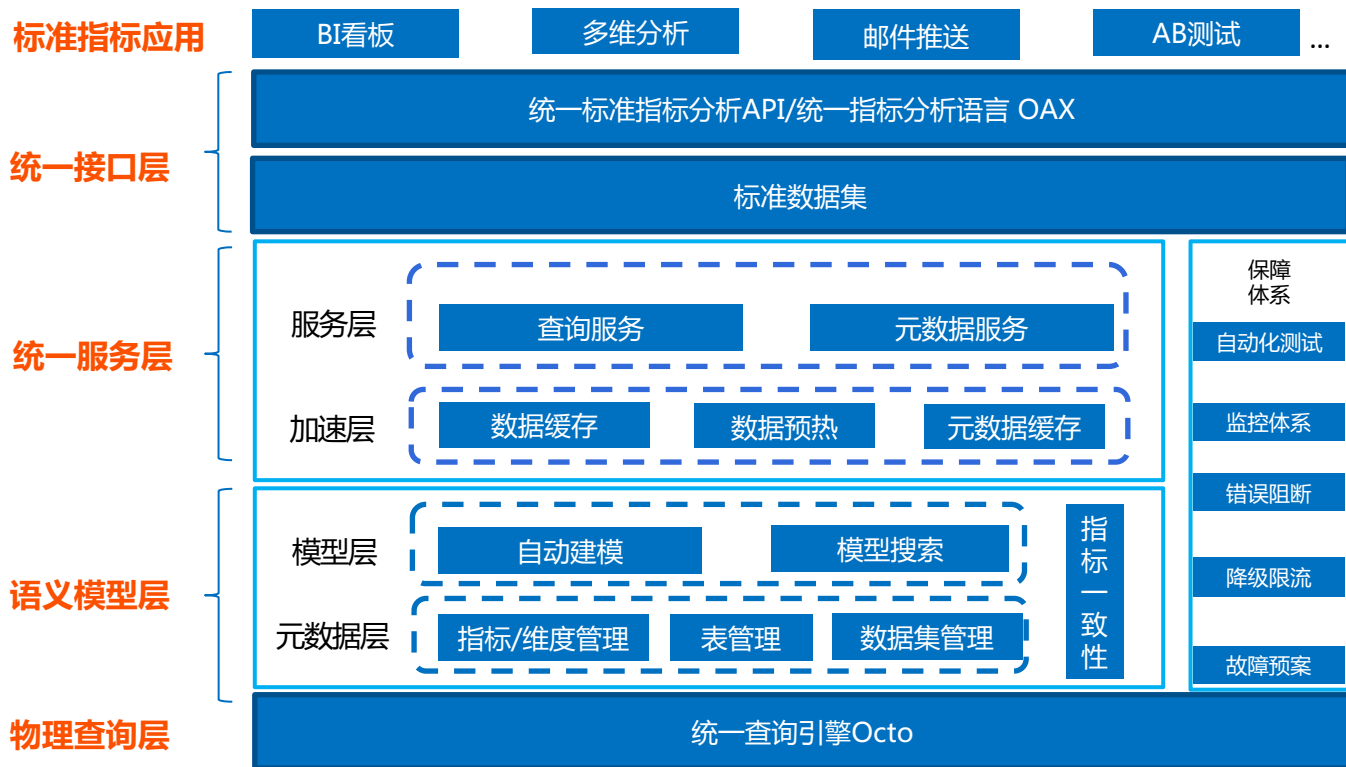
**统一指标**

**统一查询**

**统一API**

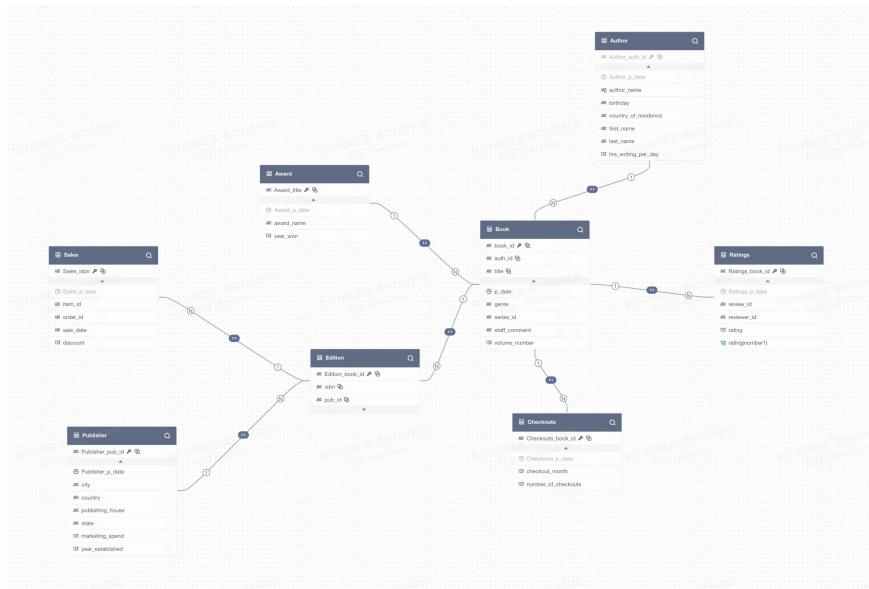
 **快手** |  **DataFun**

# Headless BI - 快手指标中台BI体系技术架构



# 核心技术1：自动化建模 – 初衷

## 传统BI建模 – 手动设置关联关系



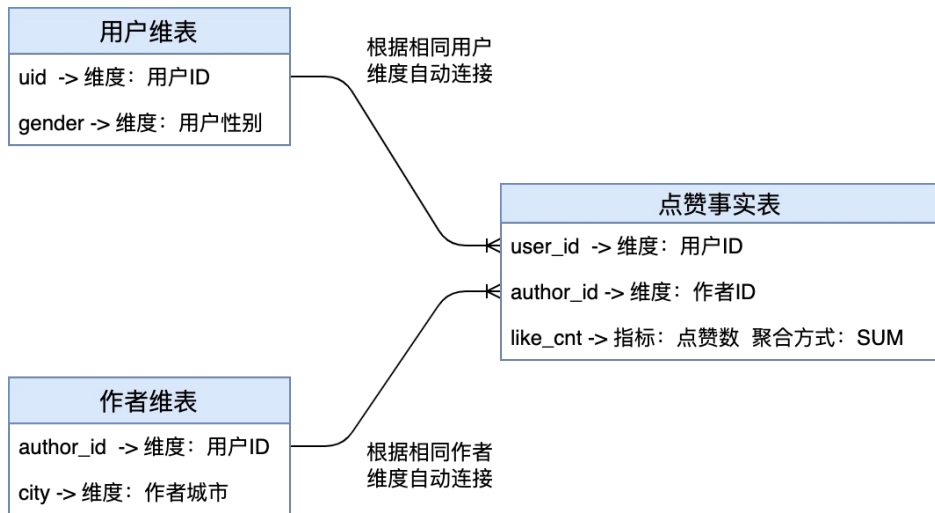
## 快手指标中台复杂性

2780 × 570 = **1584600**  
事实表 维表 待确认关系数量

手动建模工作量巨大



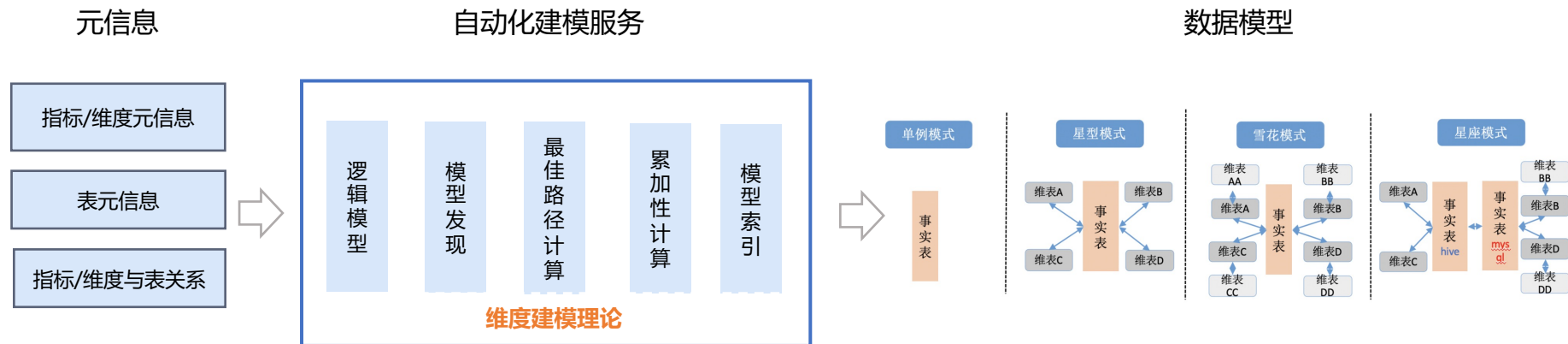
# 核心技术1：自动化建模 – 设计



## 自动化建模

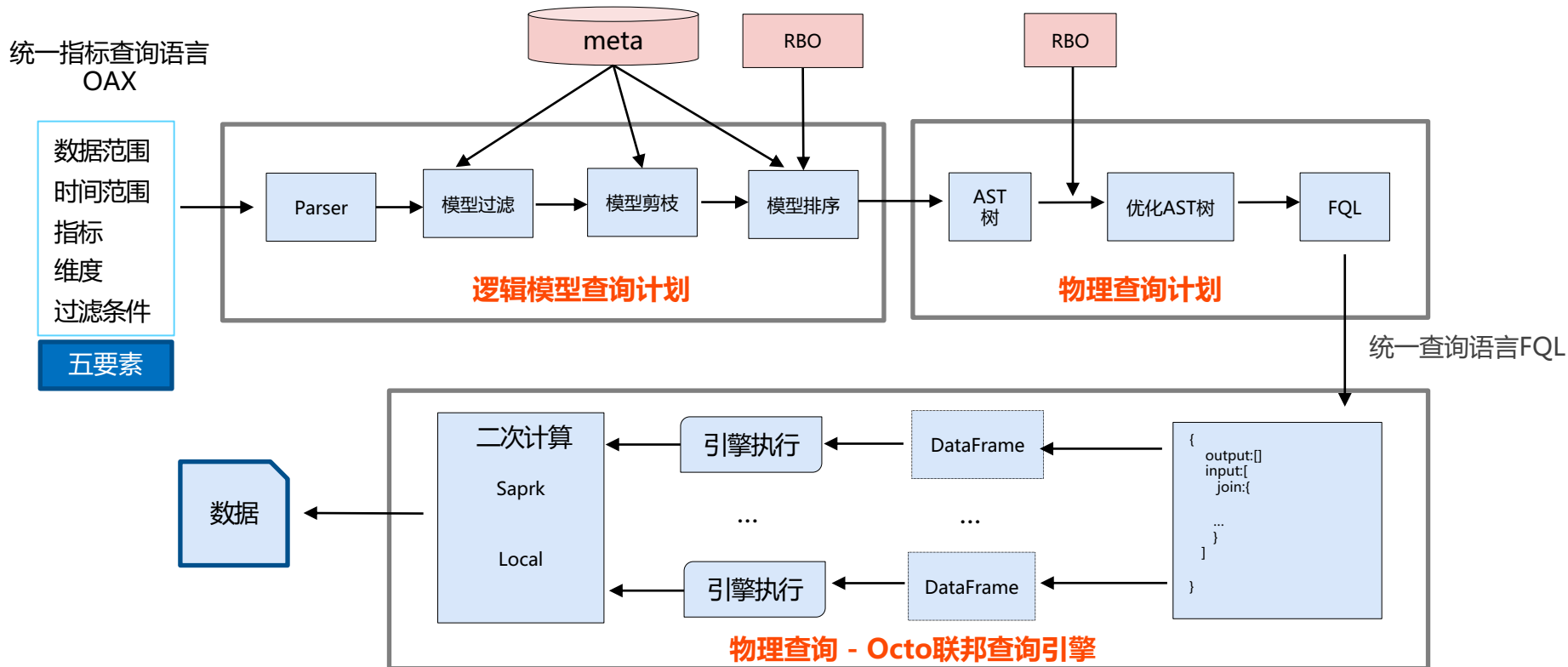
- 根据数据表的维度绑定关系，自动构建关联关系
  - 关联字段
  - 基数关系
  - 完整引用关系
  - ...
- 优势：**效率与质量**
  - 通过维度关联，自动化构建大量模型
  - 不同模型之间，数据表关系可以复用
  - 单个数据表变更，可以自动化重新构建关联模型

# 核心技术1：自动化建模 – 流程



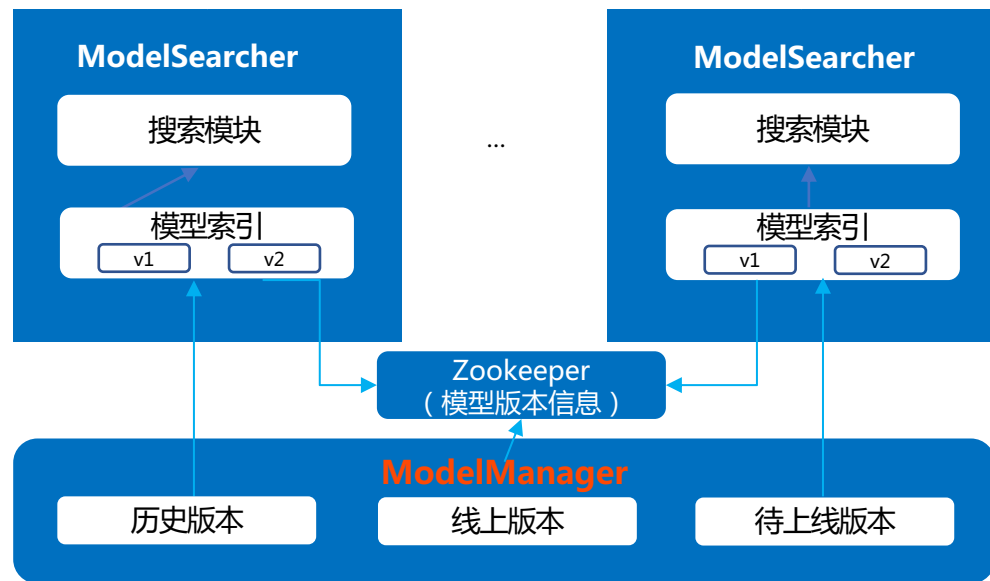
自动化构建2780个模型，全量模型构建时间小于3s

## 核心技术2：统一指标服务 – 统一指标查询流程



# 核心技术2：统一指标服务 – 模型搜索

## 模型搜索服务



**筛选阶段：** 指标维度条件-> **N**个满足条件的模型

指标筛选	淘汰指标不匹配的表
维度筛选	淘汰维度不匹配的表
范围筛选	淘汰数据范围（行）不匹配的表
日期筛选	淘汰数据日期范围不满足的表

**排序阶段：** **N**个满足条件的模型 -> **Top3**最优模型

生产排序	取早不取晚，优先取完成时间早的表
效率排序	取快不取慢，优先取效率高的表
维度排序	取粗不取细，优先取粗粒度表
手动排序	自定义优先级，取优先级高的表

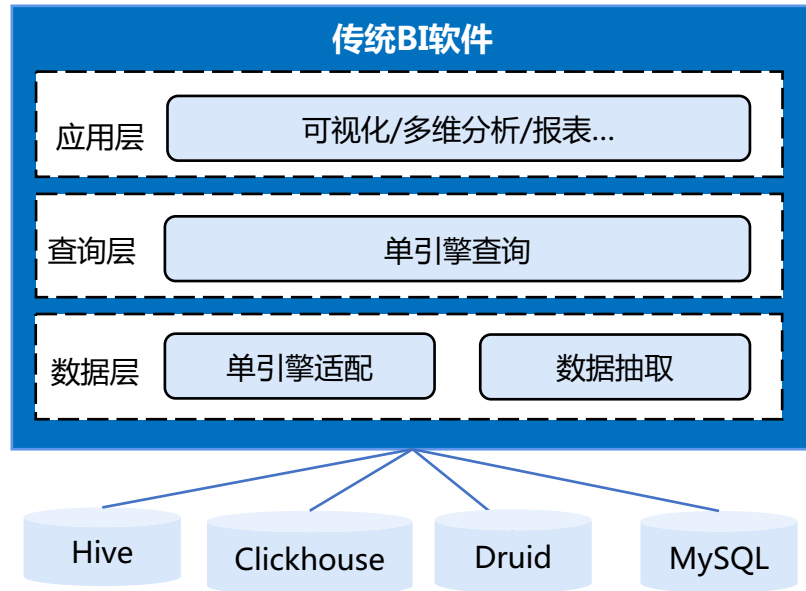
## 核心技术2：统一指标服务 – 传统BI物理查询层的问题

### 问题1：多引擎数据抽取到单引擎查询

- 成本问题：增加数据复制的存储与加工成本

### 问题2：BI查询计算能力内置

- 数据孤岛：BI体系内数据无法对外部应用服务



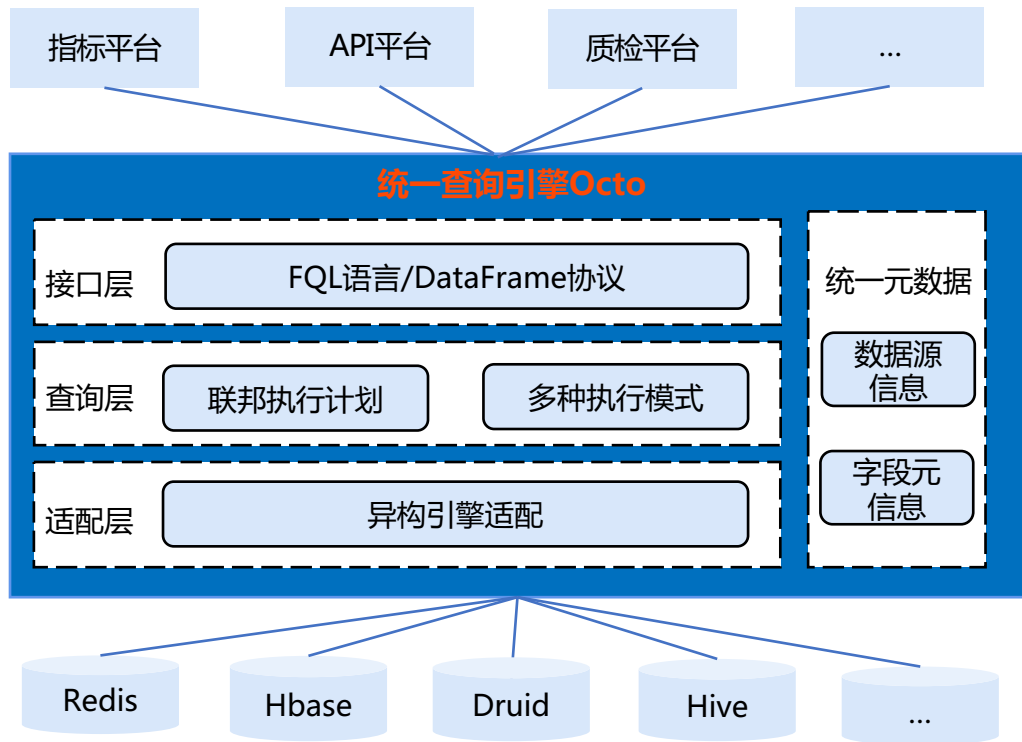
# 核心技术2：统一指标服务 – 统一查询引擎 Octo

## 解决问题

- 联邦查询：异构数据源联邦查询问题
- 开放能力：对接上层多种应用

## 核心技术

- **统一查询语言**：提供统一的基于Substrait协议扩展的查询语言FQL (Federation Query Language)
- **统一数据协议**：提供底层基于Apache Arrow数据格式封装的DataFrame协议



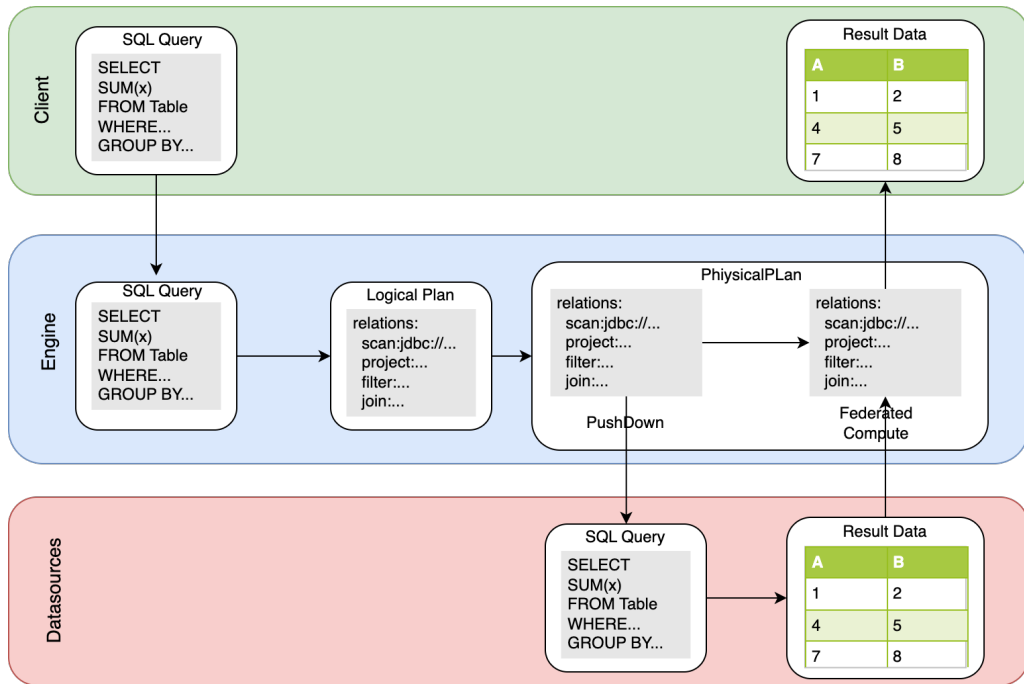
# 核心技术2：统一指标服务 – 联邦查询引擎的设计问题

## 问题1：查询表述不统一

- 不同层次之间，查询表述需要反复转化
- 对接的不同数据源对于查询缺少同一表述抽象

## 问题2：数据格式与协议不统一

- 冗余的格式转化：数据传输格式与数据处理格式不统一，查询的不同层次之间需要反复的序列化与反序列化与格式的转换



# 核心技术2：统一指标服务 – FQL和Dataframe

## 语言定义 FQL

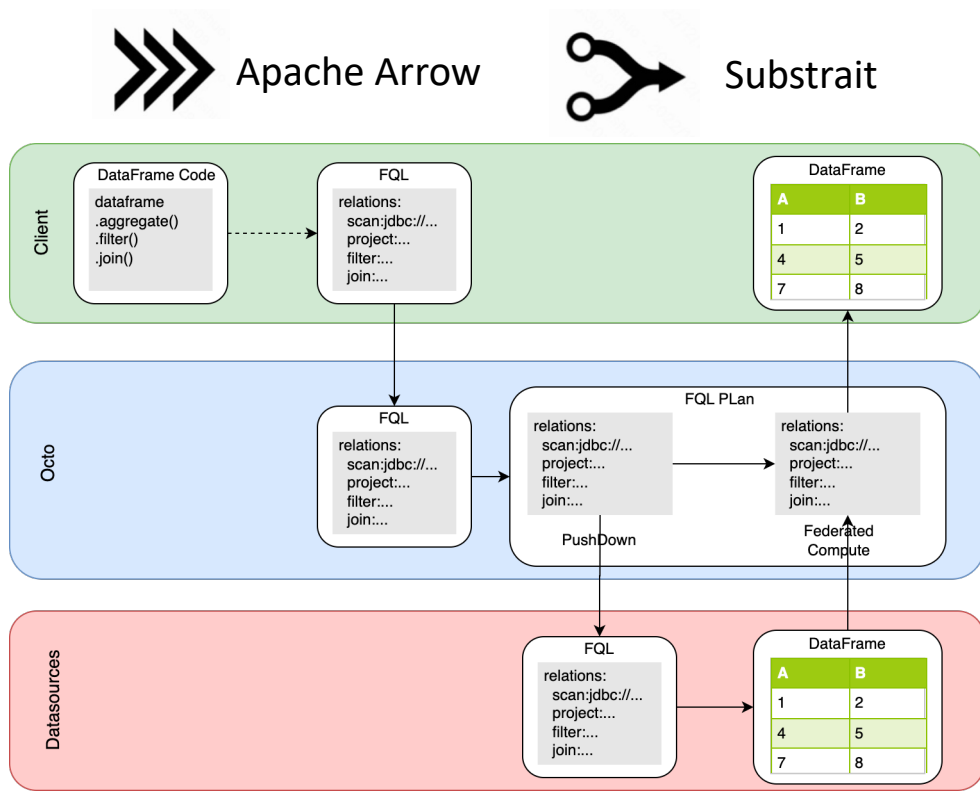
基于Substrait协议扩展的查询语言，旨在表示数据中台领域的跨引擎查询以及二次计算过程的。扩展点：

- 分析领域自定义函数：时间处理函数，方差/协方差函数...
- 计算节点：NativeQueryScanNode，利用底层引擎Native语言进行查询

## 数据抽象 DataFrame

基于Apache Arrow进行数据封装，利用DataFrame设计定义表格数据计算算子：

- Apache Arrow格式：高效传输，向量化执行
- Dataframe：定义表格数据抽象与计算API，屏蔽底层实现细节





## 04 未来展望



# 未来展望

## 智能化建模

- 根据用户历史查询推断模型关系
- 根据数据表元数据（字段/注释/生产血缘...）推断模型关系

## 智能化加速

- 利用用户历史查询进行自动化指标维度预生产，提升查询性能

欢迎技术交流



快手大数据公众号

# 非常感谢您的观看

---

 快手 |  DataFun.

