

企业-DT应用方案

主讲人：张辉 阿里云资深专家



提纲

- DT 时代大数据方案思考
- DT 时代企业数据应用架构
- OLT - 新一代数据模型
- 场景化的计算模型
- 实践案例

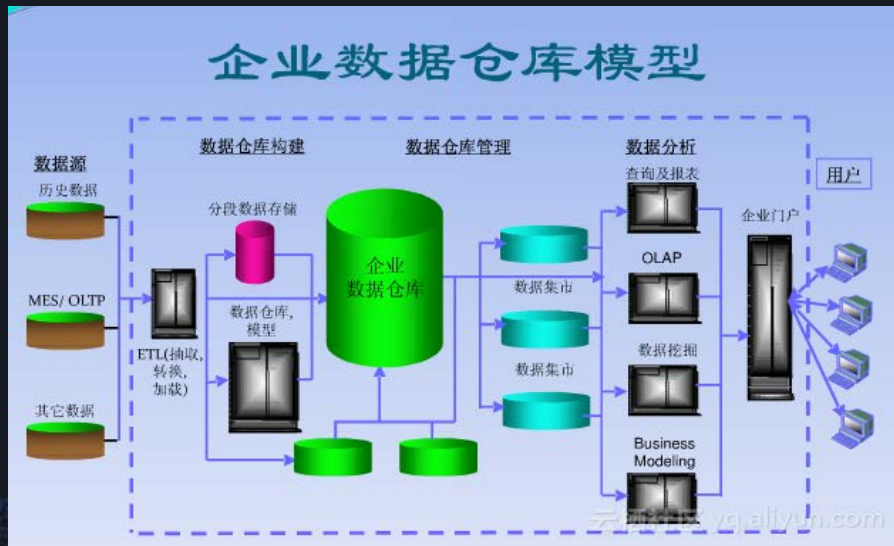


思考-三个问题

- 当下是否还需要一个复杂的EDW？
- 数据系统的目标用户是谁？
- 让数据适应计算能力还是计算适应场景？



思考-传统的架构



- 层次结构清楚，但不简洁
- 上层应用直接操作物理层
- 业务人员理解不了，技术人员维护成本高
- 大量数据冗余
- 单一层次应用模式受限
- 模型不稳定
-

思考-原因

- 计算能力不够
 - 为了适应计算能力在切分数据
 - 重心从场景化计算模型偏移
 - 垃圾数据越来越多
 - 业务人员离数据越来越远
 - 业务数据化就是报表



DT时代的数据应用架构



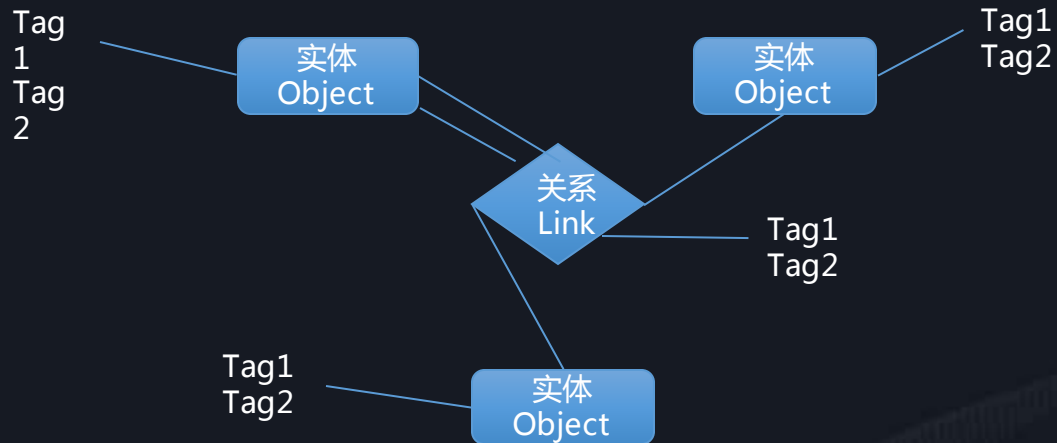
数据模型

主题模型



OLT模型

数据模型-OLT



OLT-Object

- **O : Object** , 即**实体**。表示分析的主体, 是外部物理世界真实事物的抽象, 比如电商场景的买家、卖家、商品等。

<Object>

<name>确定实体唯一性的命名</name>

<nick>中文别名</nick>

<benchmark>实体标准化定义</benchmark>

<datatype>实体数据类型</datatype>

<keyList>实体对应的物理表键信息</keyList>

<meta>其它元信息</meta>

</Object>

OLT-Link

- **L : Link , 即关系。**关系建立在两个及两个以上的实体上。比如电商场景的“购买关系” , “收藏关系” , “浏览关系” , “评价关系” 等。

<Link>

<name>确定关系唯一性的命名</name>

<nick>中文别名</nick>

<objectList>实体 , 实体 , 实体 ,</objectList>

<meta>其它元信息</meta>

</Link>

OLT-Tag

- **T : Tag , 即标签。**标签挂接在某个实体或者某个关系上。比如买家的姓名、年龄、籍贯等基础标签属于买家这个实体；比如购买时间、金额、件数属于购买关系上的标签；比如搜索的关键词、搜索频次属于搜索关系，其他可以以此类推。

<tag>

<name>确定关系唯一性的命名</name>

<nick>中文别名</nick>

<category>标签所属类目</category>

标签归属实体或者关系

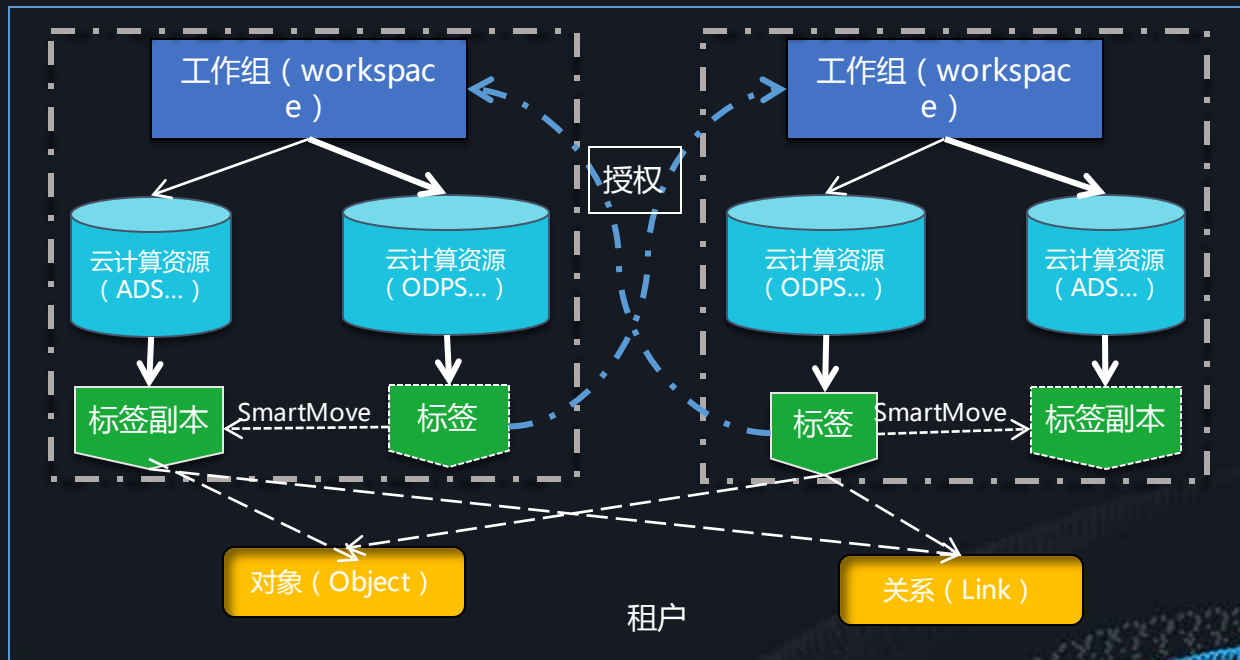
<table>标签归属物理表</table>

<var>物理字段名称</var>

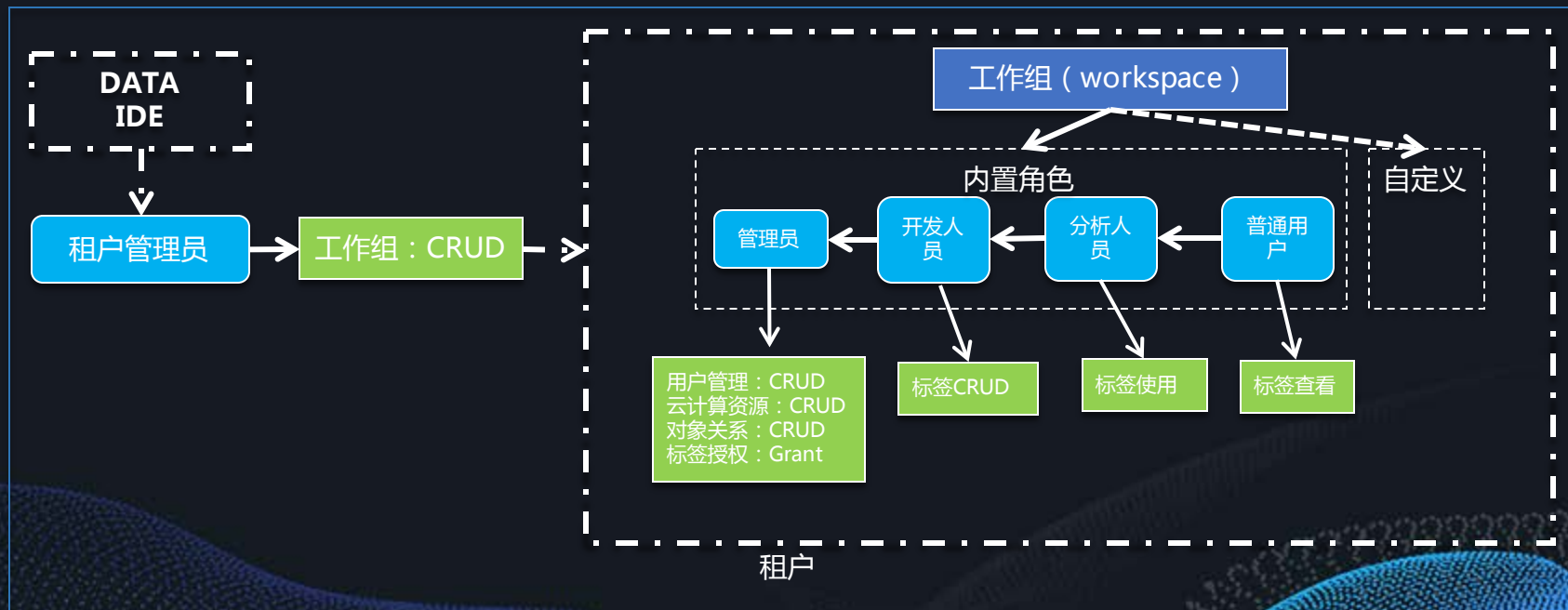
<meta>标签其它元信息</meta>

</tag>

OLT-共建共享



OLT-权限





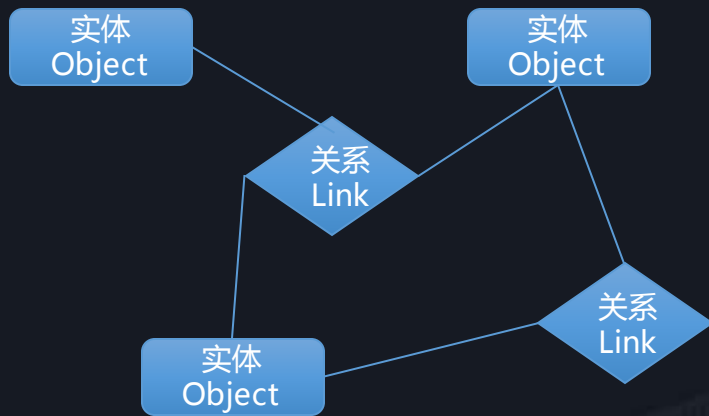
OLT-质量评估

- 实体、关系、标签统计量分布
- 链路数据分布分析
- 市场机制，应用驱动

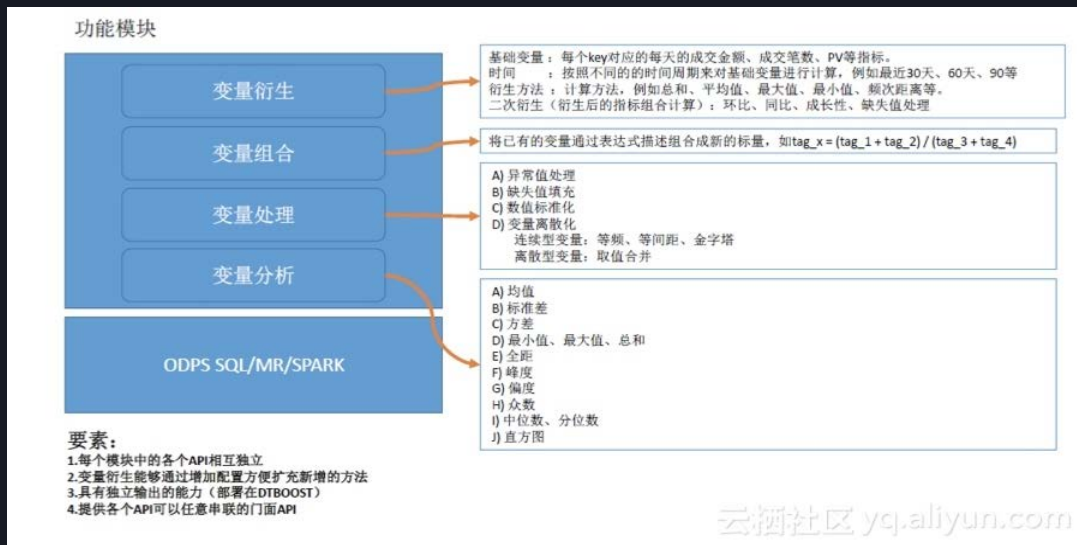


OLT-智能构建

- 自动发现实体
- 自动发现关系



OLT-标签工厂



OLT-标签工厂

时间序列上的衍生：

- 方法名称 方法描述
- cnt 变量在一定周期内的发生次数
- cntd 变量在一定周期内出现的不同值次数
- totv 变量在一定周期内的总和
- ttav 变量在一定周期内的均值
-

组合标签支持的表达式以及函数：

- 计算运算： $+$, $-$, $*$, $/$, $\%$
- 数学函数:abs,acos,asin,atan,ceil,conv,cos,.....



计算模型-业务场景化的计算模型

- 什么是计算模型？
- 计算模型不是单一的算法
- 计算模型是一个工作流

计算模型-分析引擎

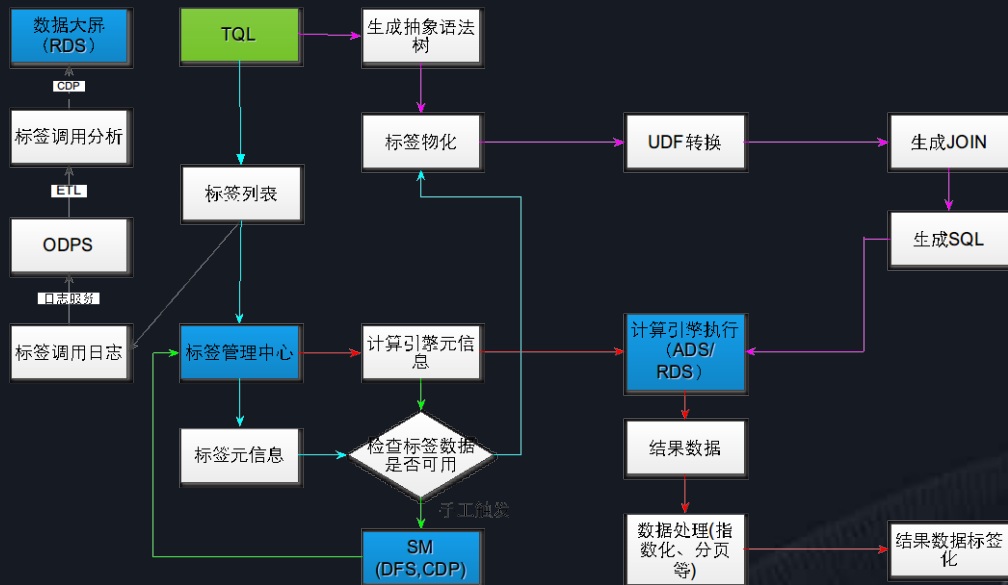
TQL :

```
SELECT fm_customer.lname AS 名字, fm_customer.address1 AS 地址1, fm_customer.city AS 城市, fm_customer.phone1 AS 电话1, fm_customer.yearly_income AS 年收入 FROM fm_customer WHERE fm_customer.gender = "M"
```

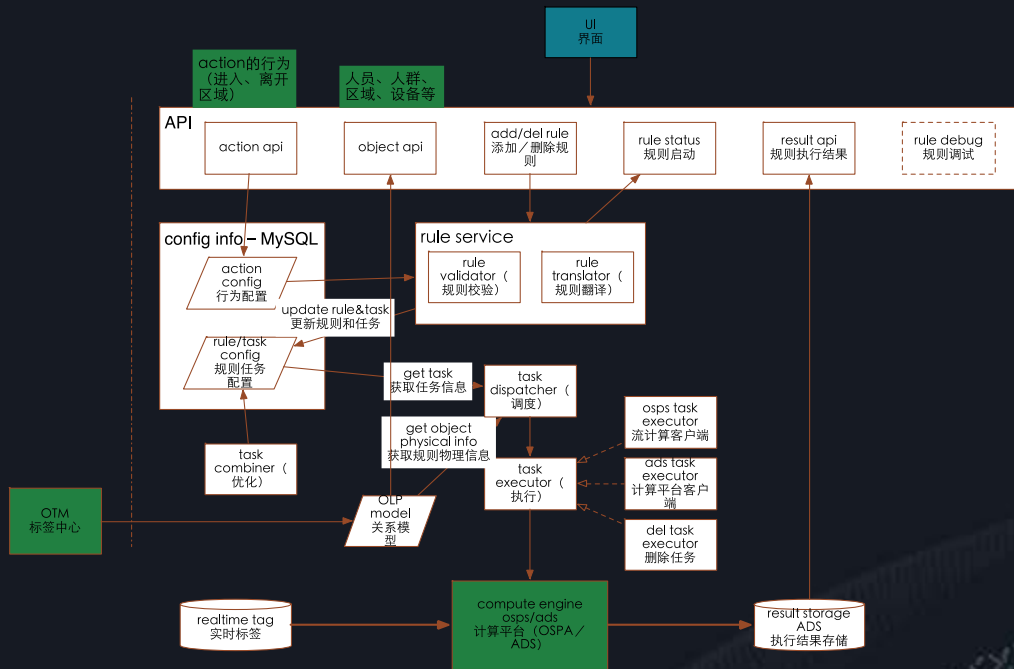
API :

```
https://dtplus-cn-shanghai.data.aliyuncs.com/dataplus_10044/porana/analysis/api/template/bida_foodmart_tql1?workspaceCode=
```

计算模型-分析引擎



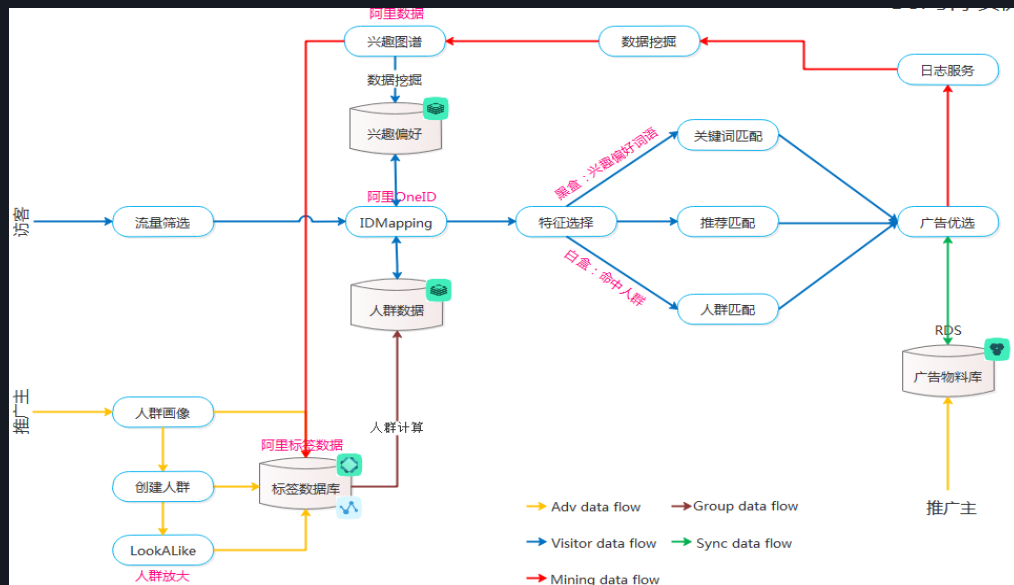
计算模型-规则引擎



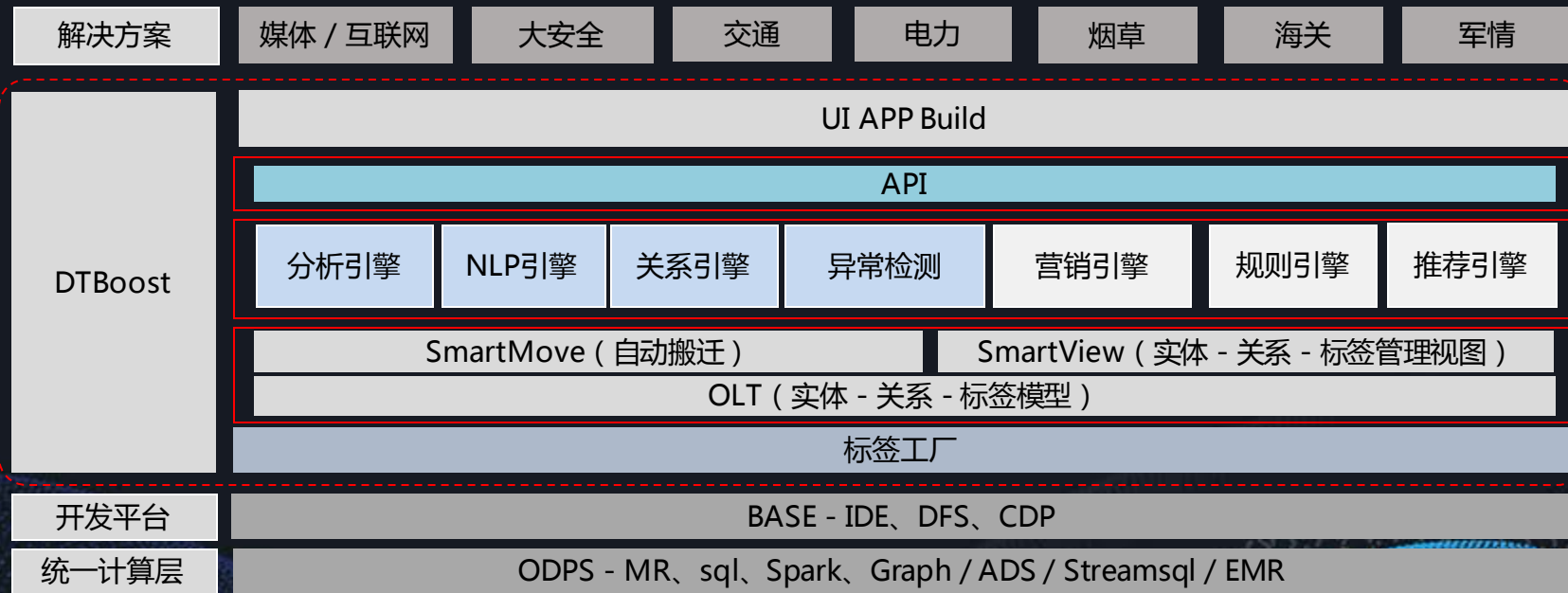
计算模型-关系引擎



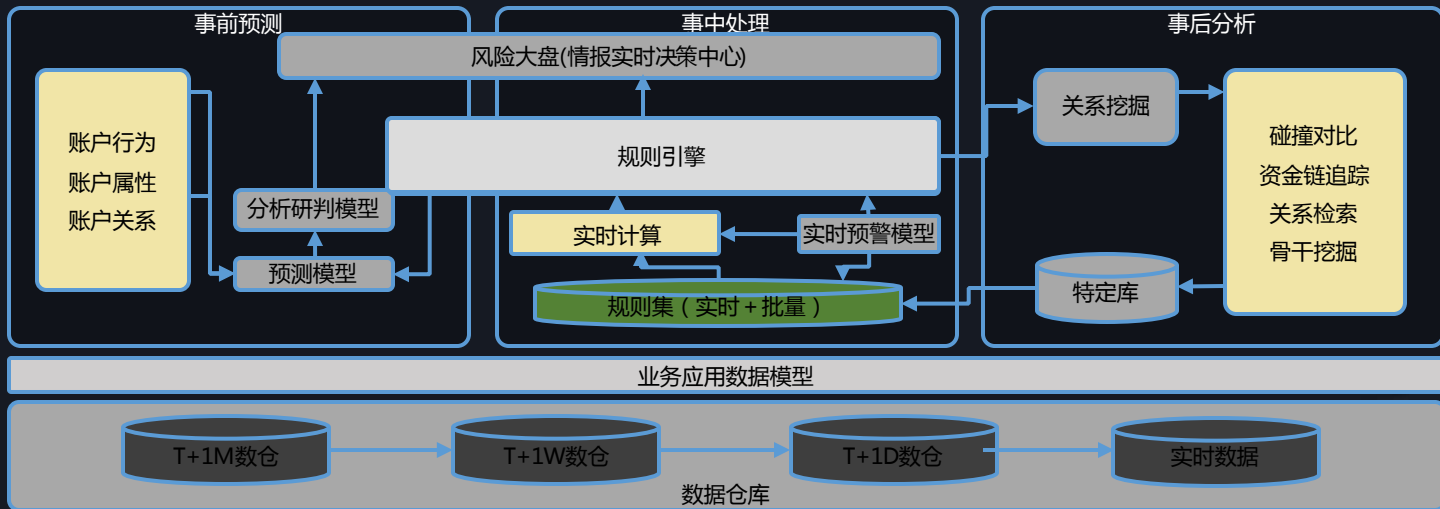
计算模型-营销引擎



企业级数据应用方案



案例-风控预警



案例-统一用户OLT模型

物理表

人口档案 酒店住宿 火车出行

数据源1

电子围栏 电察

数据源2

网页浏览 购物消费

数据源3

标签视图

1. 业务可理解
2. 跨库、跨计算
3. 基于明细数据建模
4. 逻辑模型方便拓展

标签挖掘

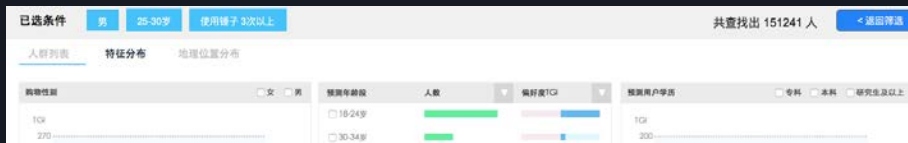
标签注册

统一标签市场

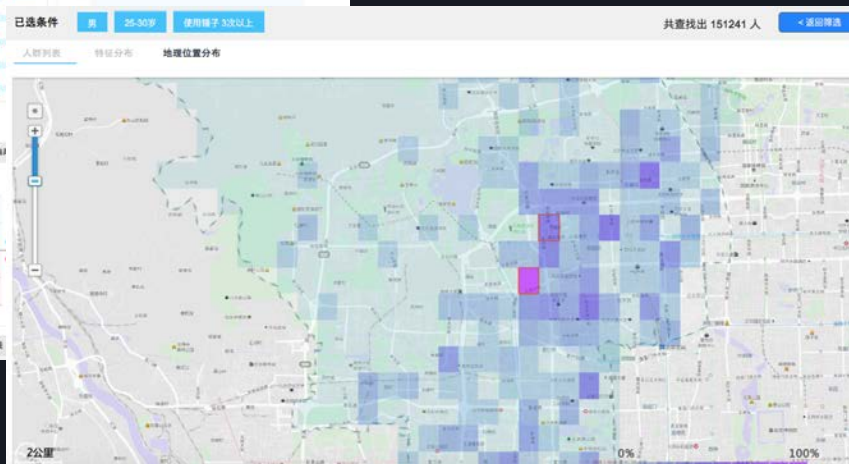
基础属性	偏好习惯	线上消费	地理位置	社交关系	用户财富	垂直行业
基本特征(27)	性别	根据用户上传人的证件信息得出	TCIF	195,709,745	968,428	
教育情况(1)	预测性别	参考用户淘宝网购数据计算得出	TCIF	289,115,872	2,841,876	
能力价值(8)	预测年龄	根据用户上传人的证件信息得出	TCIF	291,516,745	865,967	
人生阶段(10)	星座	根据用户上传人的证件信息得出	TCIF	195,715,696	116,599	
设备信息(35)	年代	根据用户上传人的证件信息得出	TCIF	195,255,955	0	
身体特征(3)	生肖	根据用户上传人的证件信息得出	TCIF	195,715,696	0	
职业信息(3)	生肖	根据用户上传人的证件信息得出	TCIF	196,787,237	116,563	
注册认证(2)	女性身高[eval评价]	女性身高, 数据来源于天猫评价	idf	9,998,906	0	
	女鞋尺码[eval评价]	女鞋尺码, 数据来源于天猫评价	idf	2,386,891	0	
	女性体重[eval评价]	女性体重, 数据来源于天猫评价	idf	9,042,839	0	



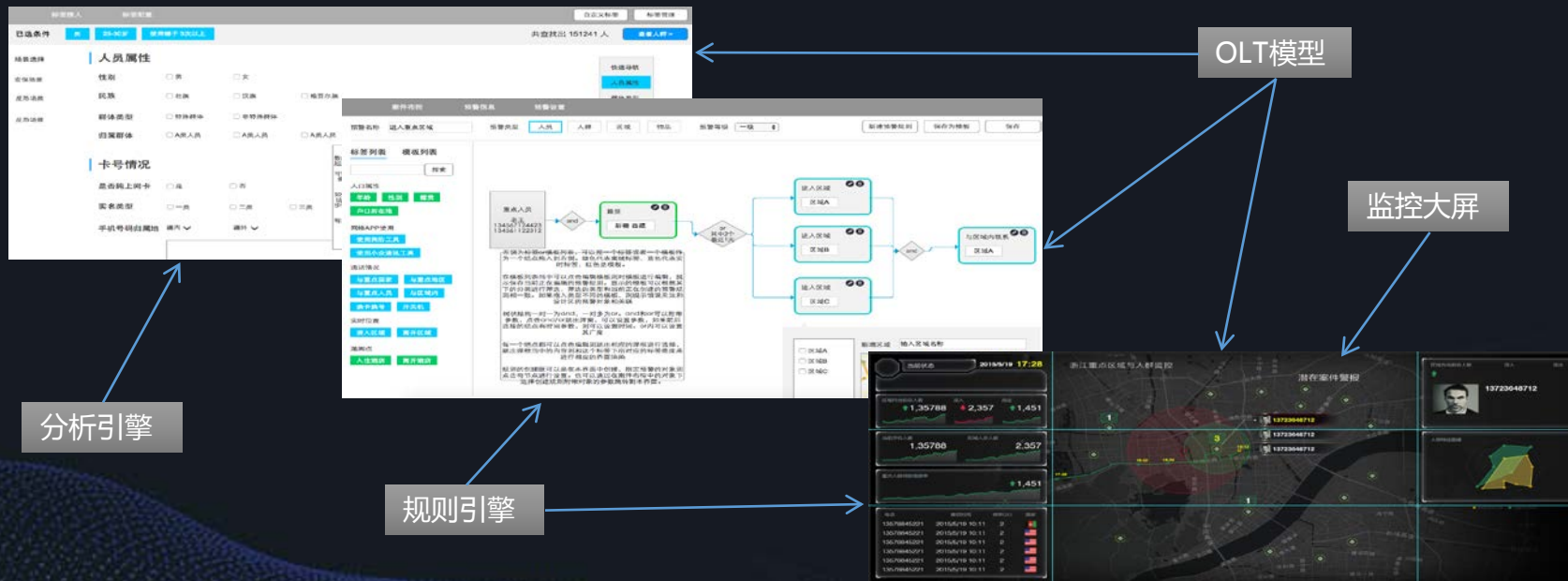
案例-事前分析



结合分析结构与交互式分析应用框架，能够快速搭建基于明细数据，让业务人员能够进行灵活筛选、钻取、联动式的分析研判产品



案例-事中监控预警



案例-事后追踪





2017云栖大会·成都峰会
THE COMPUTING CONFERENCE



产品-阿里云数加平台

- 画像分析
- 规则引擎
- 营销引擎
- 关系网络分析



2017云栖大会·成都峰会
THE COMPUTING CONFERENCE



飞天·智能

APSARA INTELLIGENCE

2017云栖大会·成都峰会

5月23日 成都世纪城天堂洲际大酒店