



第九届中国数据库技术大会  
DATABASE TECHNOLOGY CONFERENCE CHINA 2018

# 考拉数据业务实践

考拉数据技术与产品部 陈文翔

DTCC  
2018

2018.05.10 – 12 北京国际会议中心



IT168.com

ChinaUnix

ITPUB

# 大纲

- 考拉数据发展简介
- 数据仓库构建
- 数据营销业务实践

DTCC2018



# 数据业务发展历史

## 考拉数据发展

2014

- Q4海淘项目立项
- 数据分析师
- Excel

2015

- 1月首页上线
- Oracle->Hadoop
- Azkaban
- 优化ETL

2016

- 数据仓库重构
- 接入猛犸大数据
- 数据权限
- JStorm-实时大屏

2017

- Impala
- Kylin
- 标签挖掘
- 用户画像
- 个性化营销
- 接入网易有数
- 元数据管理

2018

- 实时数据仓库
  - Sqoop->Binlog
  - KUDU
- OpenApi
- 自动化营销
- 数据质量
- 数据地图

# 数据业务0->1挑战

---

- 业务复杂，变化快
- 数据时效性：离线->准实时->实时
- 海量数据产生的大量计算，元数据管理问题
- 如何保证数据质量

DTCC2018



# 数据仓库建设

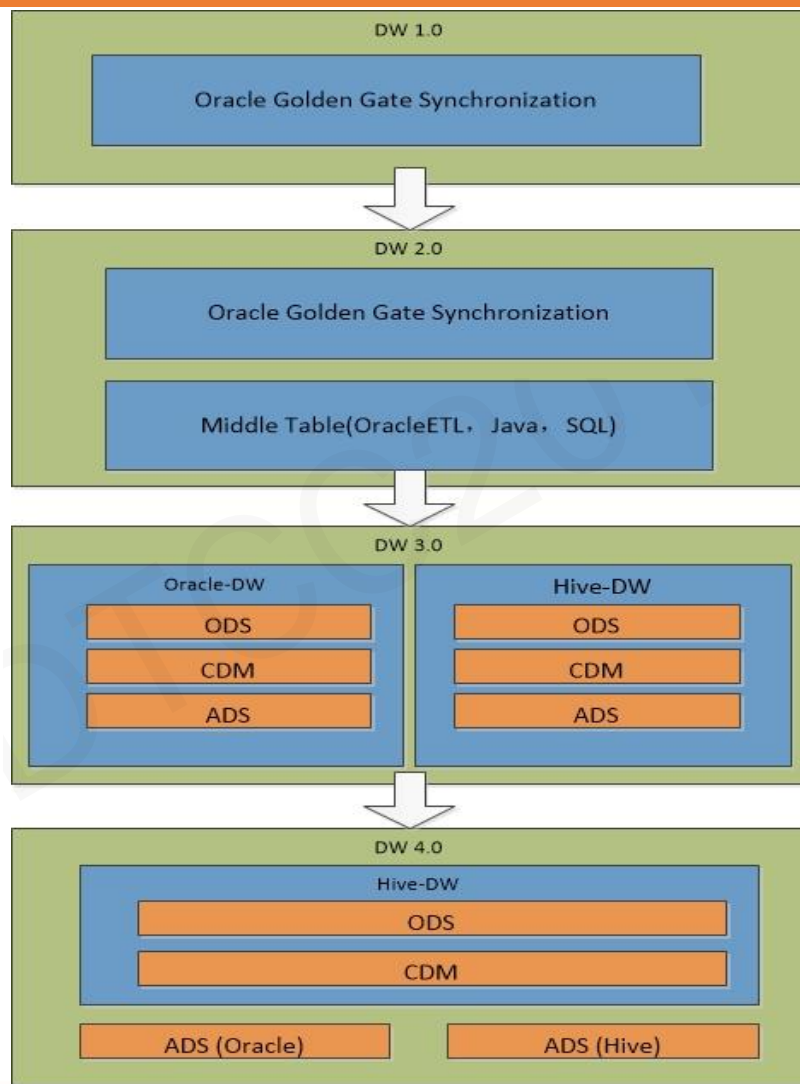
---

- 数据仓库层级重构
- 数据仓库实时化
- 数据元数据管理
- 数据监控质量管理

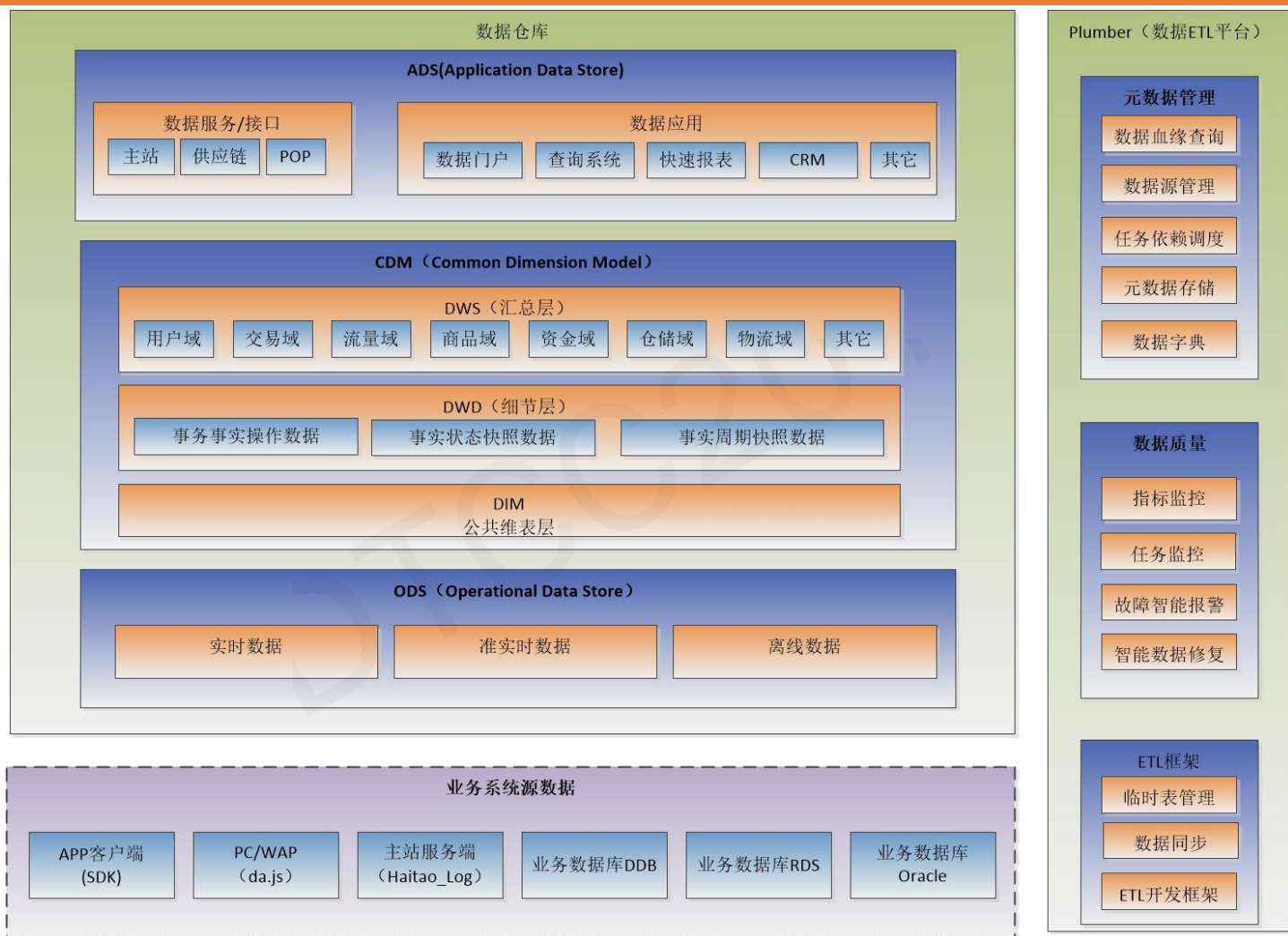
DTCC2018



# 数据仓库层级重构变迁

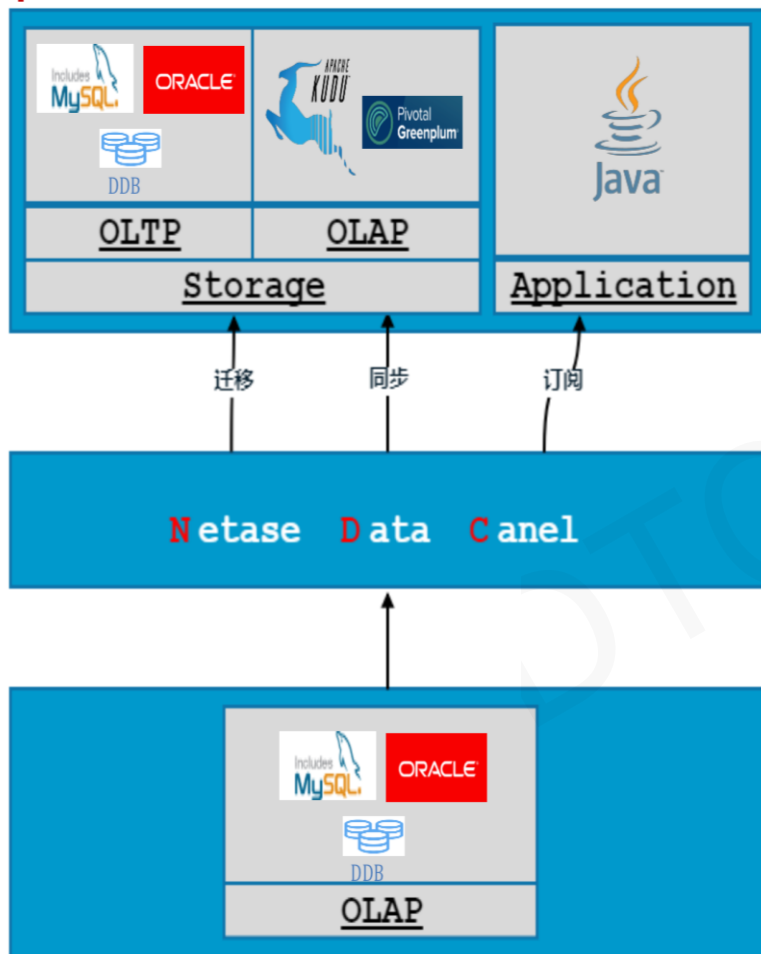


# 数据仓库层级重构





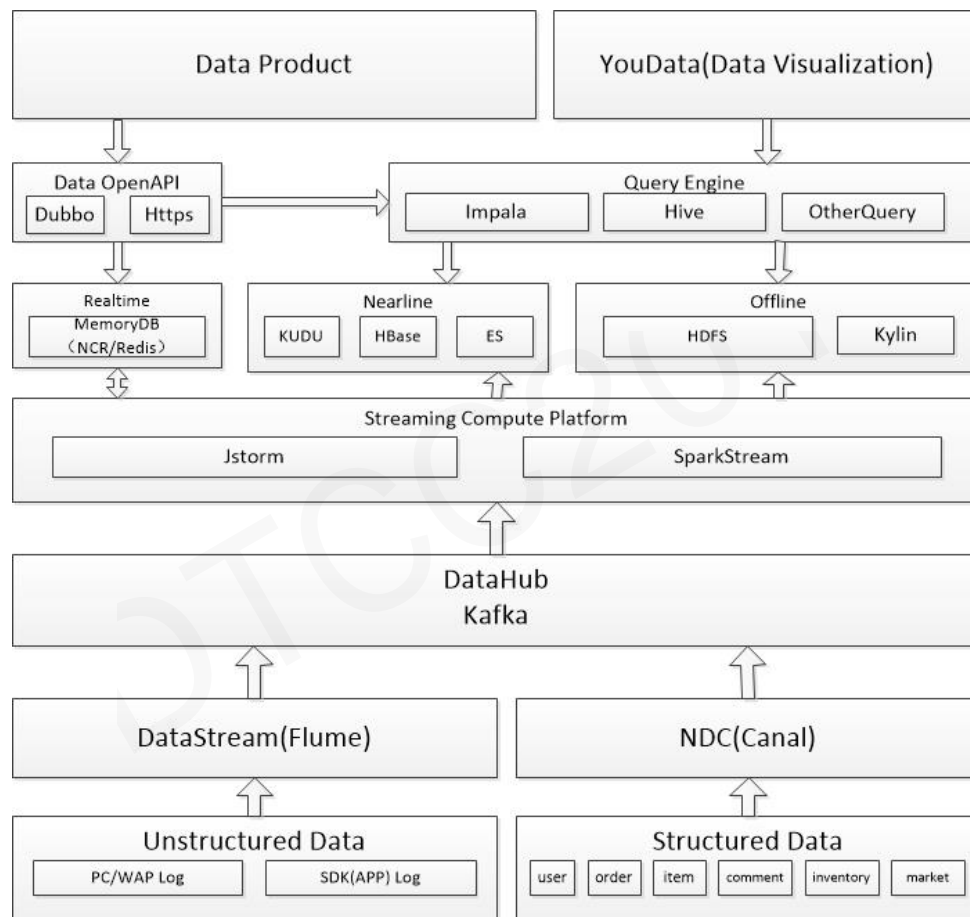
# NDC+KUDU 仓库实时化改造



- Binlog 增量订阅&消费模式
- 利用KUDU提供实时场景的表更新和事务查询
- 利用Impala作为实时和离线的混合查询引擎



# 仓库实时化改造

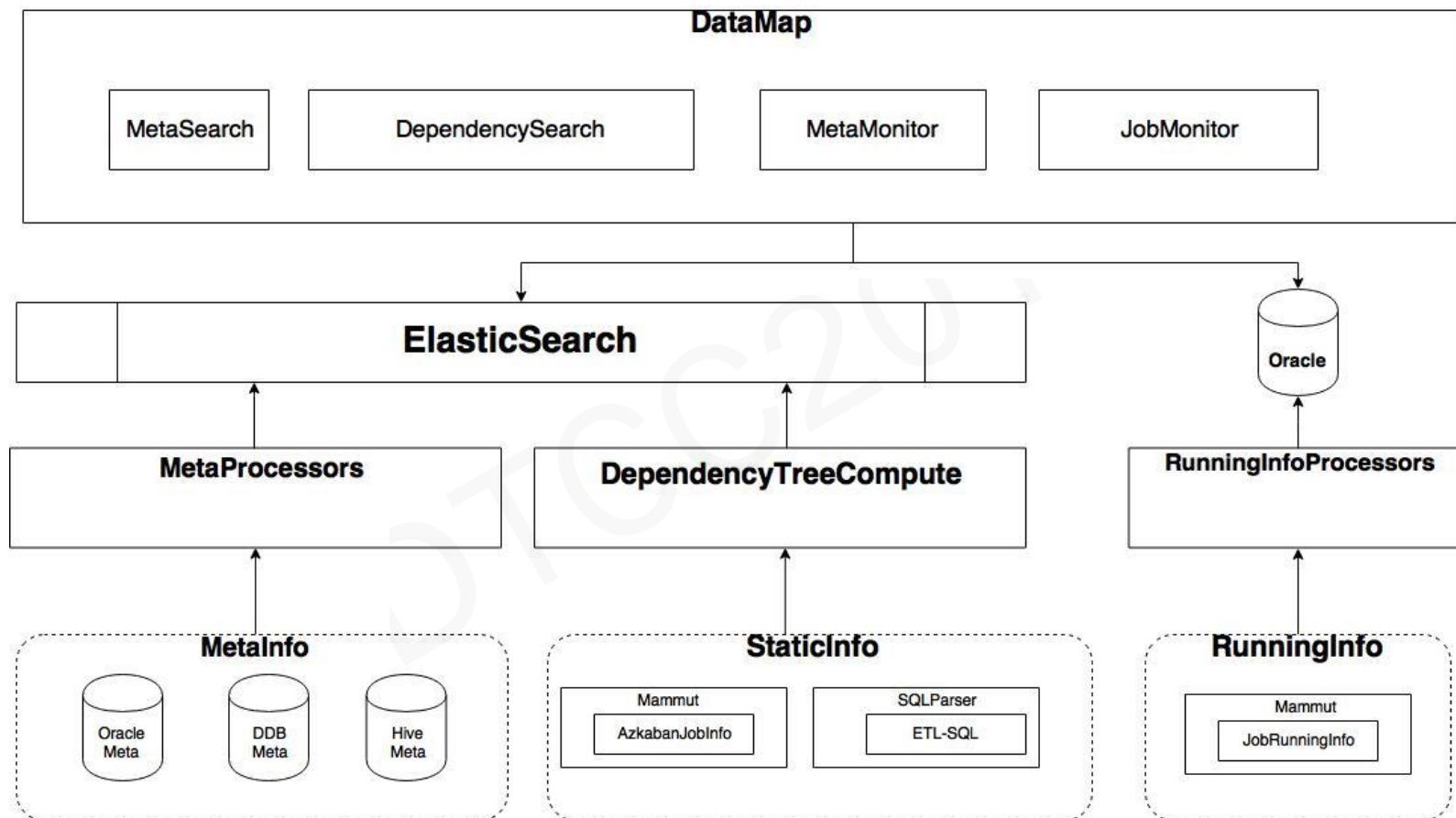


# 元数据管理平台

- 数据地图
  - 收集处理数据仓库，业务库元数据
  - 收集并处理任务运行状态
  - 利用ES搭建元数据搜索引擎
  - 提供元数据，任务依赖查询服务
- 监控系统
  - 业务-系统-模块-指标 监控（业务指标监控）
  - 任务流运行监控（猛犸大数据平台监控）
  - 数据接口，数据产品监控(哨兵监控)



# 数据地图



# 数据地图搜索结果

## 数据地图列表

表名	字段	信息
haitaodw.GLOBAL_CDM.DWS_KL_TRD_MTD		全站交易月累计表
haitaodw.GLOBAL_CDM.DWS_KL_TRD_TD		全站交易历史累计表
haitaodw.GLOBAL_CDM.DWS_KL_TRD_CAT_1M		类目交易月汇总表
haitaodw.GLOBAL_CDM.DWS_KL_TRD_CAT_1D		类目交易日汇总表
haitaodw.GLOBAL_CDM.DWS_KL_TRD_BUYER_1Y		买家交易年汇总表
haitaodw.GLOBAL_CDM.DWS_KL_TRD_IMPTYPE_CANCEL_1D		贸易方式交易关单日汇总表
haitaodw.GLOBAL_CDM.DWD_KL_TRD_PHONE_FIRST_DI		收件手机全站首单表
haitaodw.GLOBAL_CDM.DWS_KL_TRD_CANCEL_1M		全站交易关单月汇总表
haitaodw.GLOBAL_CDM.DWS_KL_TRD_SKU_CANCEL_1D		前台SKU交易关单日汇总表
haitaodw.GLOBAL_CDM.DWS_KL_TRD_GOODS_CANCEL_1D		前台商品交易关单日汇总表

# 猛犸中项目依赖关系



# 数据业务实践

---

- 数据仓库->报表->生产资料

DTCC 2018

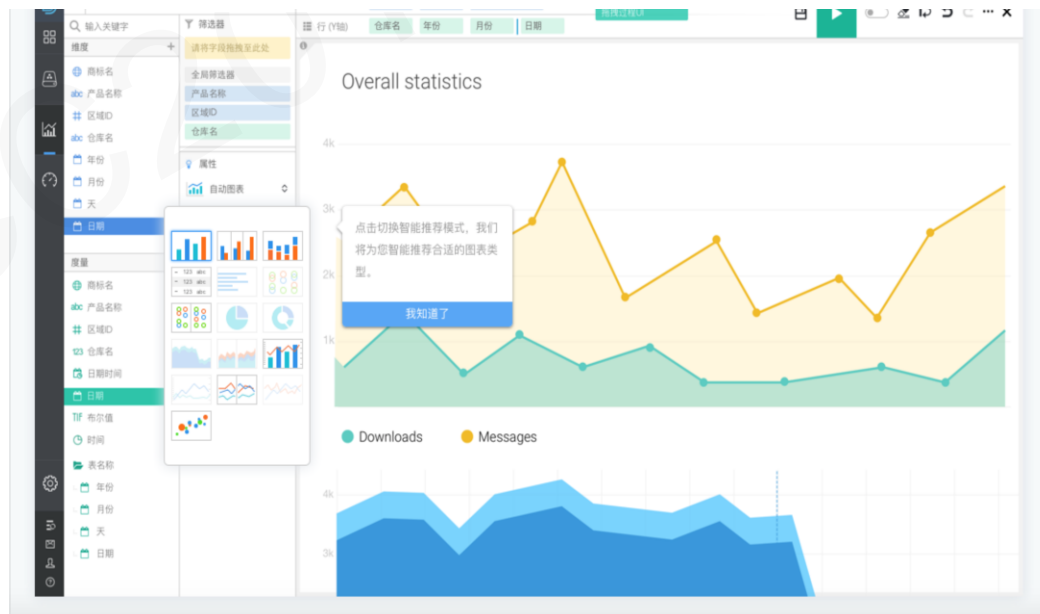
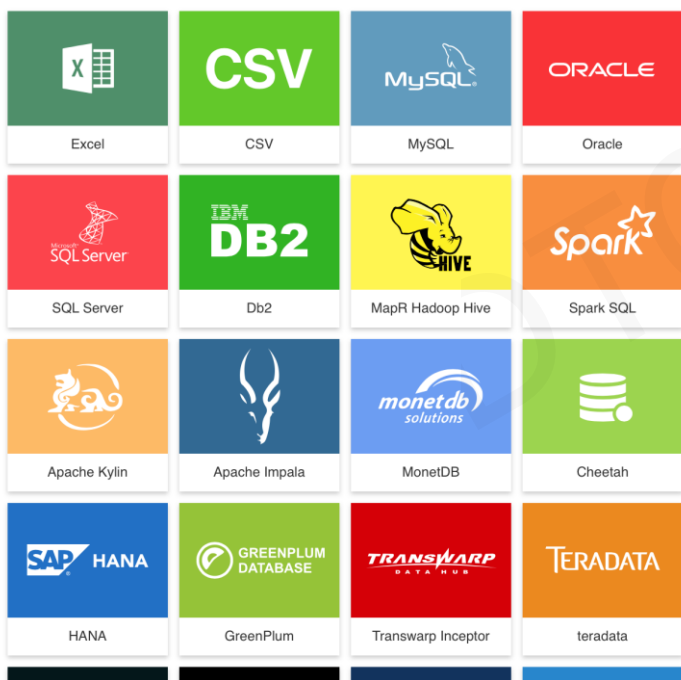
# 网易有数提升报表开发效率

连接数据

建立数据模型

制作报告并分享

添加数据连接





# 普通报表->诊断细节报表

- 一般报表
  - 表格型
  - 概览分析
  - 业务分析



- 进化报表
  - 诊断
  - 对比
  - 下钻



# 数据业务实践

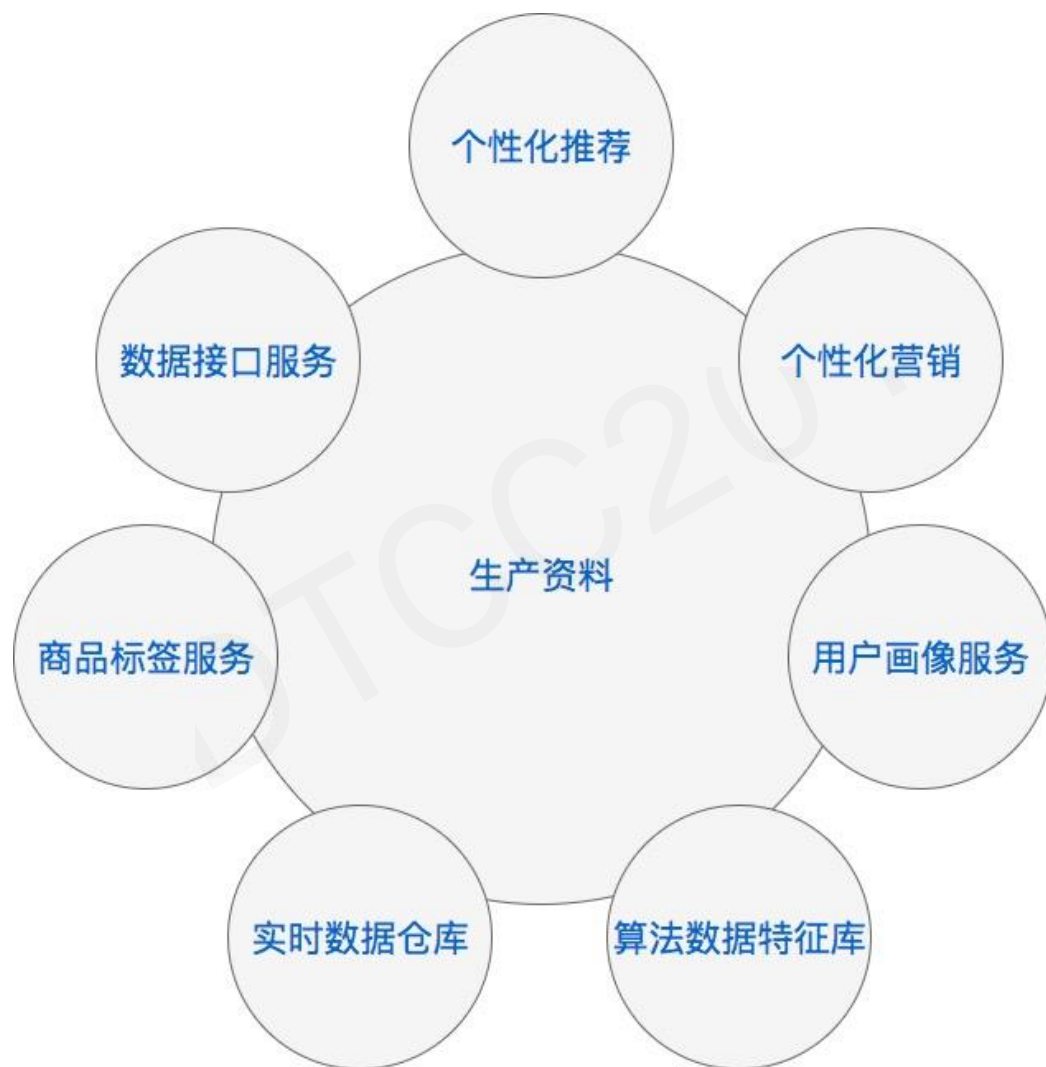
---

- 数据仓库->报表->生产资料

DTCC 2018



# 数据生产资料



# 为什么要利用数据做用户营销?

- 我知道我的营销费有一半被浪费了，但是我不知道是那一半

-----约翰 沃纳梅克

- 数据精准营销的三个难点：
  - 如何找到目标人群
  - 如何适配最优的资源营销策略
  - 如何节约成本

# 技术分解

- 从计算广告角度看，营销广告的核心问题就是为多组用户与环境的组合找到最合适的广告营销投放策略以优化整体广告活动的利润。
- 最优化
- 广告
  - r函数内影响参数定义：
  - a表示本次投放的广告（例如：以商品，品牌，类目，活动生成的营销主题）
  - u表示广告投放的用户群体，
  - c表示广告的环境，场景（包括上下文场景，资源位，弹窗，邮件EDM等触达方式）



# 如何搭建用户营销平台

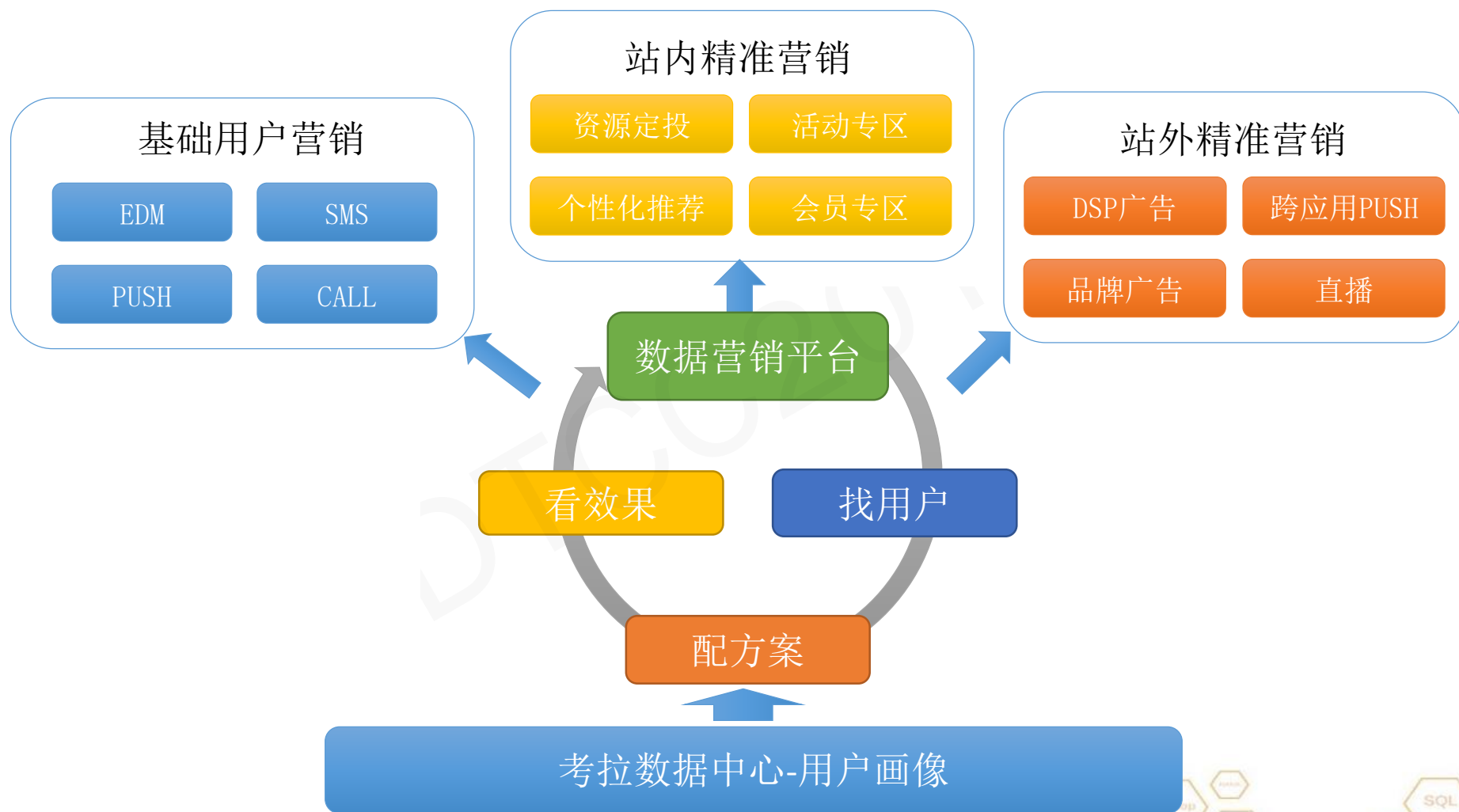
1 定制化用户分群（受众定向）

2 标签人群查询（查询引擎）

3 资源广告投放服务（投放引擎）

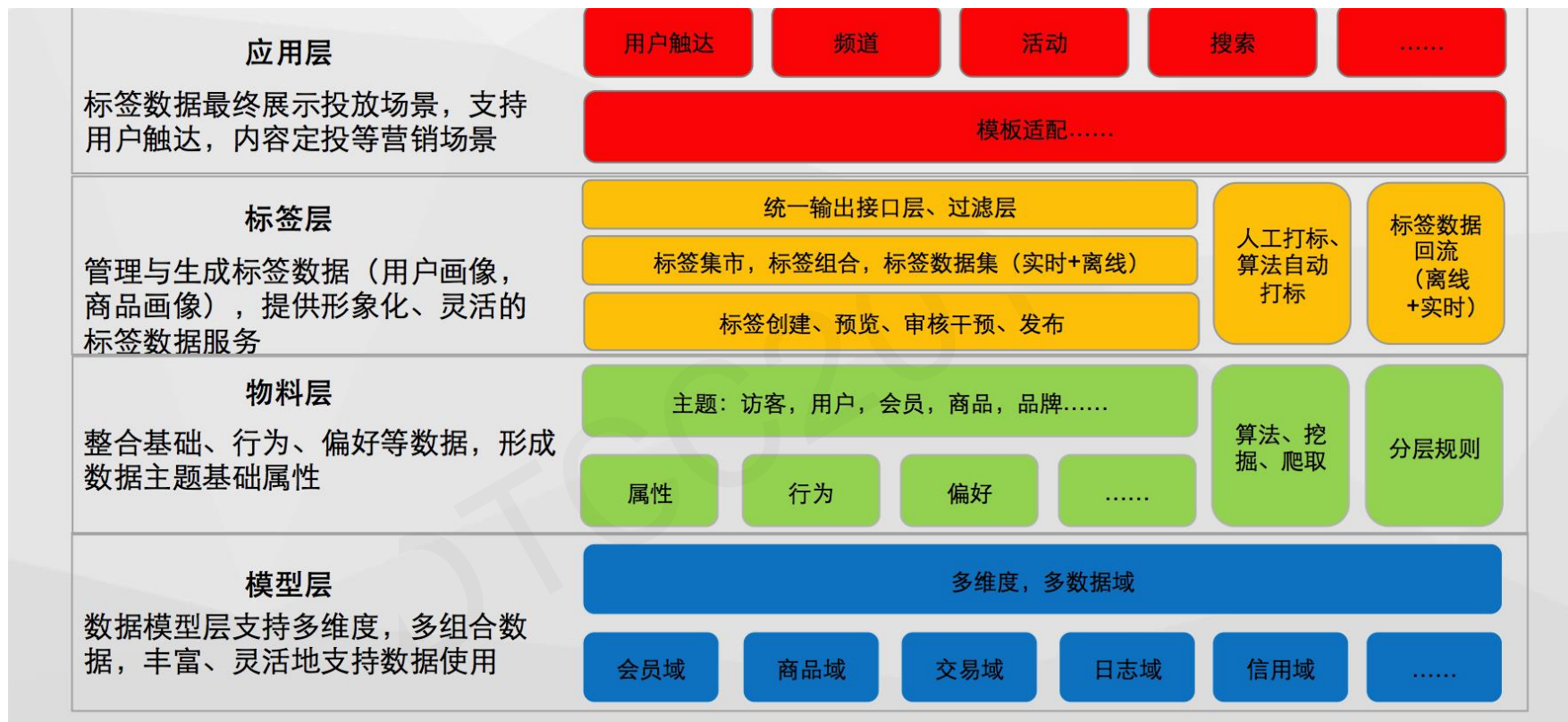
4 营销效果反馈平台（A/B效果评估）

# 业务流程图





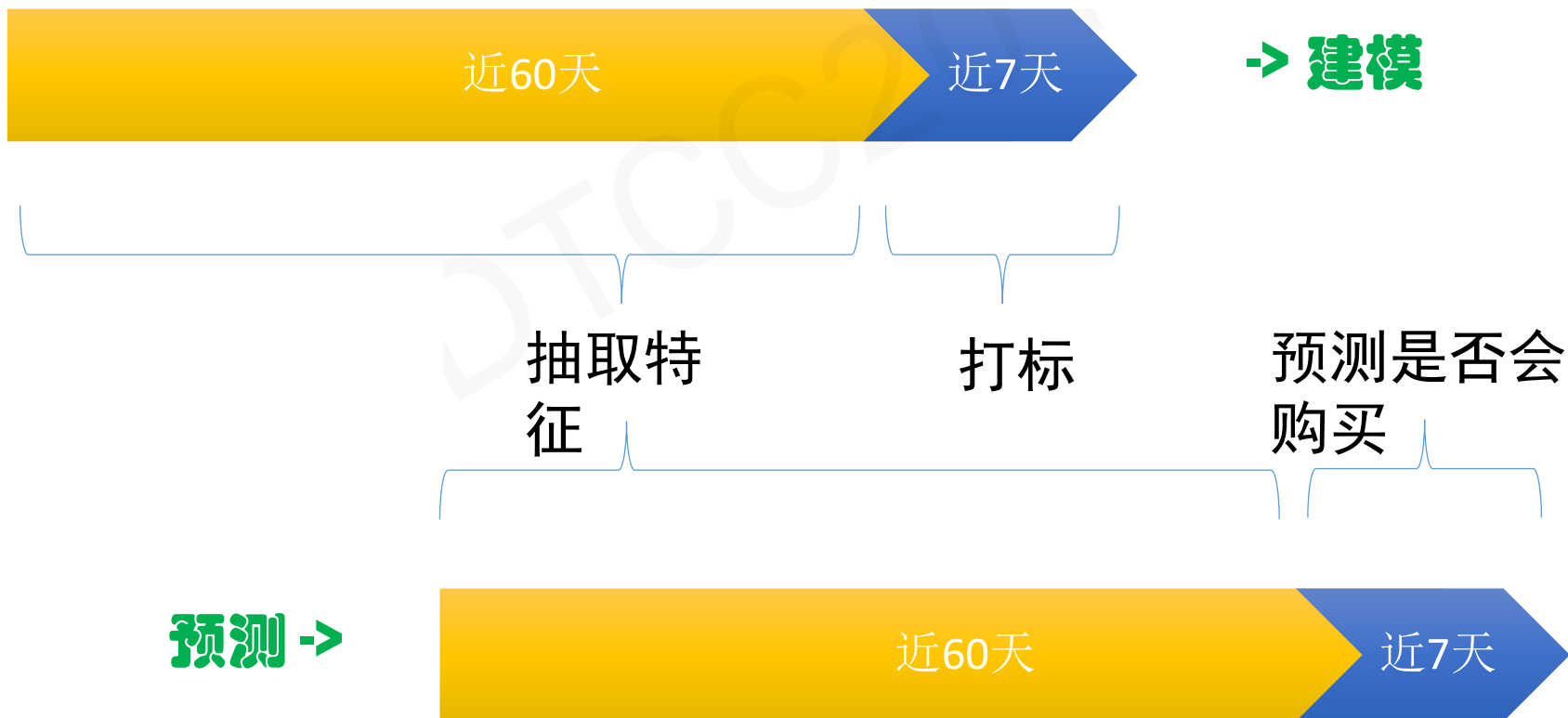
# 受众定向-标签挖掘分层



# 用户购买预测中的行为特征

用户属性特征和近期的行为特征 → 用户未来是否会购买

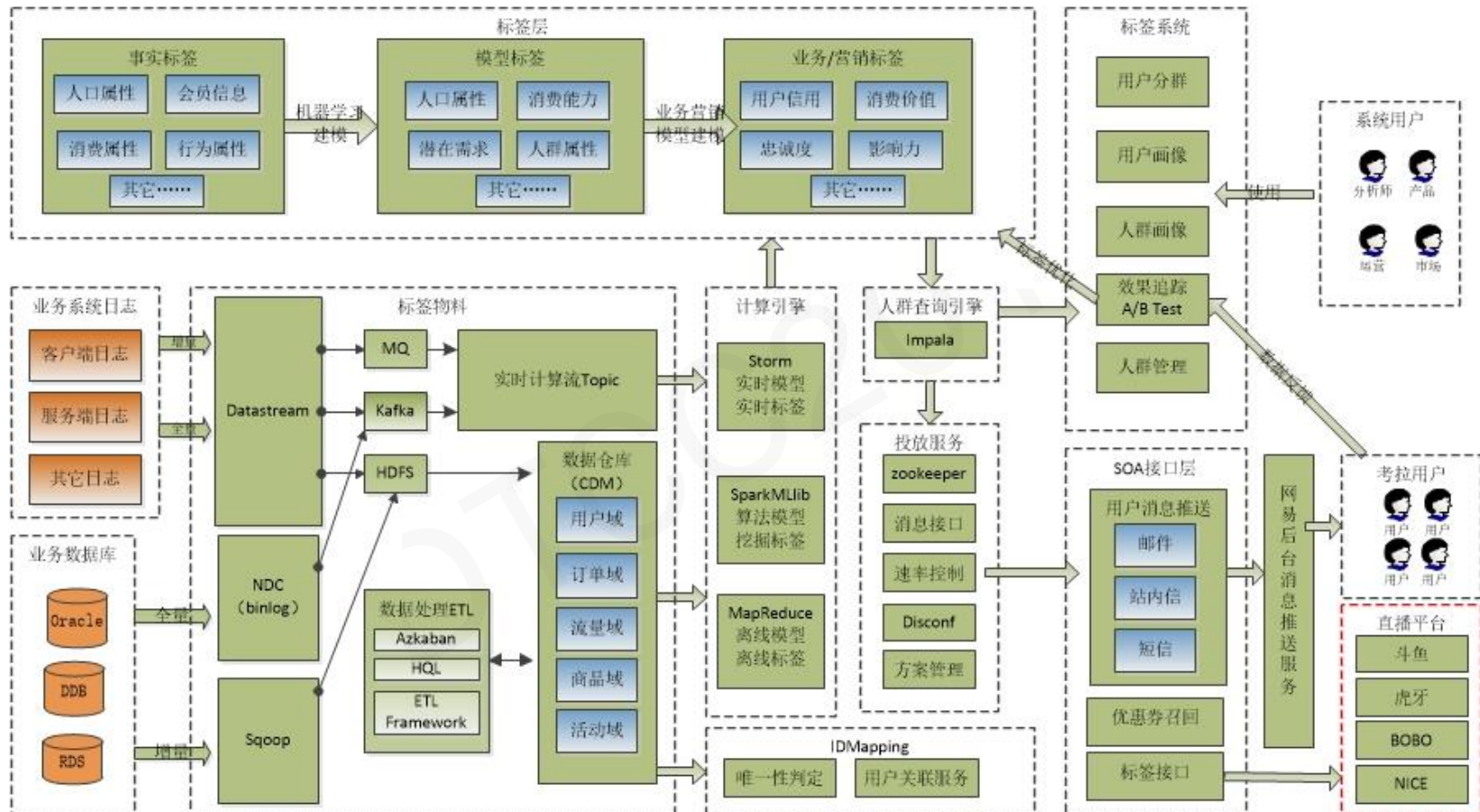
时间窗口：



# 特征类型

- 涉及用户行为
  - 用户到商品的订单
  - 加购商品（数量、时间）
  - 浏览商品详情页
    - 来源入口、活动、购物车、收藏、我的足迹）
  - 搜索点击
  - 分享
  - 评价
- 时序切片（7天天然周期） 7天\*8段
- 涉及的统计口径（来源于经验）
  - 均值、方差、求和、差分的均值、最大值、最小值
  - 最早时间、最晚时间、次数、发生天数、平均时间间隔

# 关键技术实施



# 方案配置

## 方案配置

标签主题 用户/人群



### 方案配置-静态人群

[进入>](#)



### 方案配置-动态人群

[进入>](#)



### 方案配置-AB测试

[进入>](#)



### 方案配置-全站推送

[进入>](#)



### 方案配置-自动化营销

[进入>](#)



### 方案配置-手机号发短信

[进入>](#)

**标签静态数据：**即方案设置后，生成固定标签人群，不随用户数据变化而变化。

**标签动态数据：**为每天按标签规则重新匹配用户，人群会随用户每天状态变化而变化。

**用同一人群，划分成不同组，测试方案效果。**如：选10%的人进行投放方案A，选另外10%的人投放方案B。

**针对所有用户投放，**包括历史以来账号人群，及当日新增（准）实时人群。

**配置自动化营销规则与优惠券模板后，**每天定时触发，给用户生成专属优惠券、内容，并通知用户。如：CDE加购商品营销。

**支持手动填写手机号，发送短信，**不对手机号做转成考拉账号验证等。



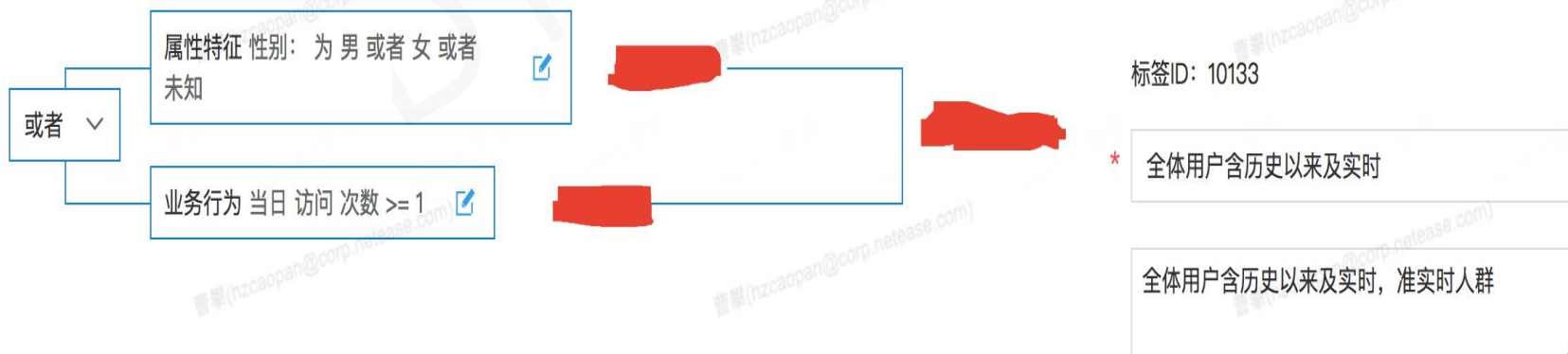
## 案例一：最基本的一个条件人群

已选规则条件 ?



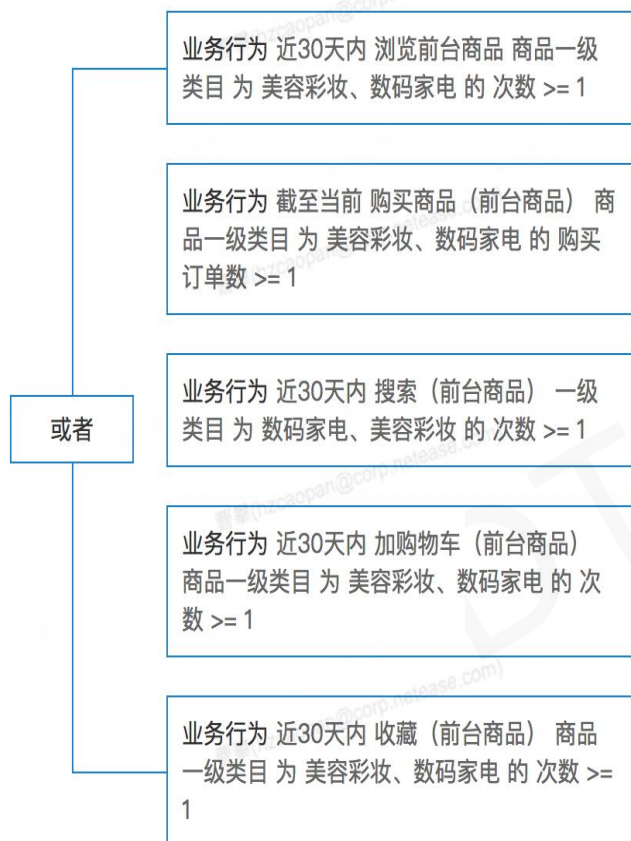
## 案例二：多个条件，“或者”关系，包含实时条件，离线条件

已选规则条件 ?



## 案例三：多个条件人群，“或者”关系

已选规则条件 ?



标签ID: 271

\* 美妆&数码偏好用户

请输入标签描述

画像分析

下载明细

另存为新标签



# 未来要做的事情

---

- 实时化推进
- 算法特征数据仓库建设
- 营销业务程序化投放
- 算法实验效果平台

DTCC2018

# THANKS





讲师申请

联系电话（微信号）：18612470168

关注“ITPUB”更多  
技术干货等你来拿~

与百度外卖、京东、魅族等先后合作系列分享活动



## 让学习更简单

微学堂是以ChinaUnix、ITPUB所组建的微信群为载体，定期邀请嘉宾对热点话题、技术难题、新产品发布等进行移动端的在线直播活动。

截至目前，累计举办活动期数60+，参与人次40000+。

## ITPUB学院

ITPUB学院是盛拓传媒IT168企业事业部（ITPUB）旗下  
企业级在线学习咨询平台  
历经18年技术社区平台发展  
汇聚5000万技术用户  
紧随企业一线IT技术需求  
打造全方式技术培训与技术咨询服务  
提供包括企业应用方案培训咨询（包括企业内训）  
个人实战技能培训（包括认证培训）  
在内的全方位IT技术培训咨询服务

ITPUB学院讲师均来自于企业  
一些工程师、架构师、技术经理和CTO  
大会演讲专家1800+  
社区版主和博客专家500+

## 培训特色

无限次免费播放  
随时随地在线观看  
碎片化时间集中学习  
聚焦知识点详细解读  
讲师在线答疑  
强大的技术人脉圈

## 八大课程体系

基础架构设计与建设  
大数据平台  
应用架构设计与开发  
系统运维与数据库  
传统企业数字化转型  
人工智能  
区块链  
移动开发与SEO



## 联系我们

联系人：黄老师  
电话：010-59127187  
邮箱：edu@itpub.net  
网址：edu.itpub.net  
培训微信号：18500940168