



2017第八届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2017

# O2O搜索优化实践之道

马尧

yao.ma02@ele.me

# 概要

01

饿了么o2o搜索推荐业务介绍

02

饿了么o2o搜索推荐的特殊挑战

03

饿了么在工程和算法上的实践

04

一些感悟



# 饿了么o2o搜索推荐介绍——业务形态

用户触达90%+

## 多入口来源

- 搜索入口
- 品类入口
- 首页列表
- 活动、会场入口...

## 多种产品形态

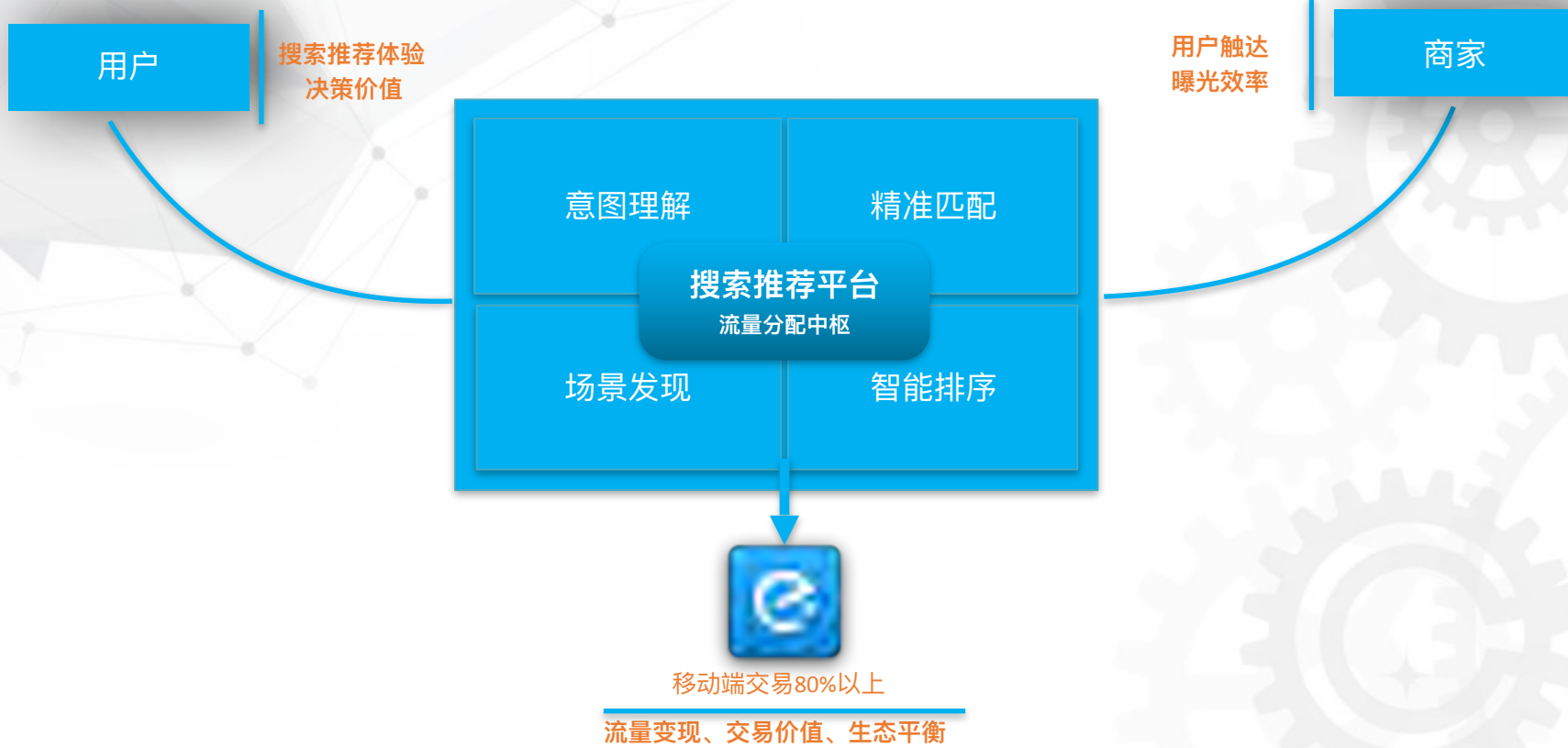
- 搜索、推荐、筛选
- 优惠、品质、效率

## 多品类

- 美食
- 果蔬生鲜
- 商超、鲜花、药店...



# 饿了么o2o搜索推荐介绍——定位



# 饿了么o2o搜索推荐的特殊挑战—— vs 传统搜索

- 移动化——流量订单基本全部来源于APP
- 场景化——时段、地点、季节、天气、用户相关
- 本地化——用户、商户信息皆是本地
- 短文本（无文本）
- 交易相关
- 高度闭环——信息反馈完整



VS

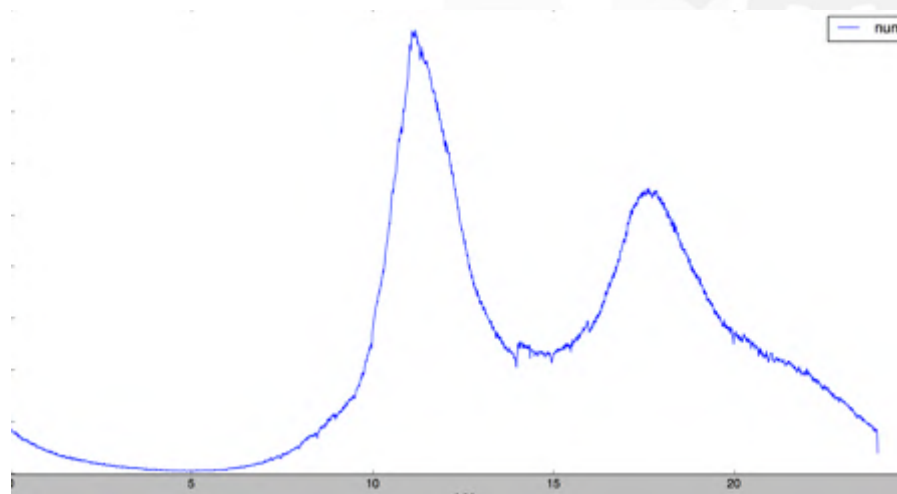


# 饿了么o2o搜索推荐的特殊挑战—— vs 电商平台

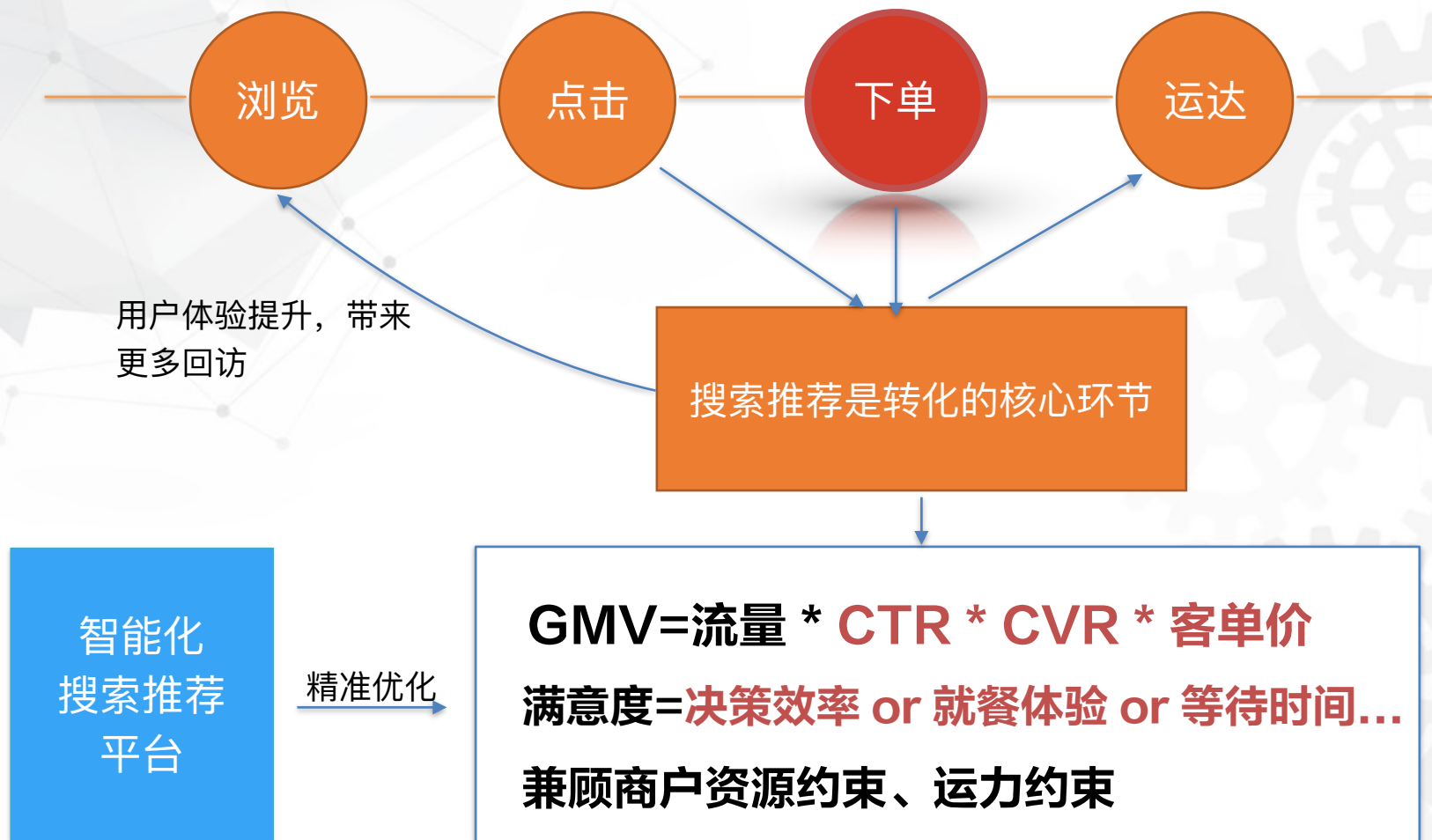
- 优化链路长
  - 浏览->点击->下单->运达
- 峰值效应明显
  - 高QPS；用户行为分布变化快；
- 即时资源约束
  - 餐厅资源、人力；配送运力；
- 时间敏感
  - 决策效率——快速帮助用户决策
  - 等待时间——端到端时间尽可能短



VS



# 搜索推荐业务目标拆解



# 搜索推荐业务架构



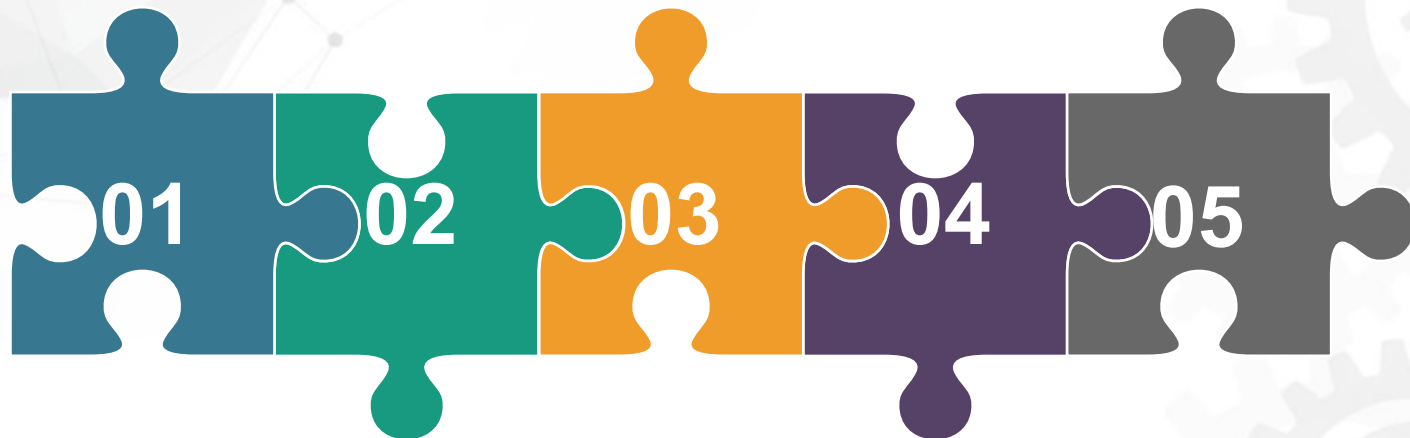


# 饿了么在工程和算法上的实践

搜索、Rank技术  
框架鸟瞰

体验升级  
Rank算法演进路  
线

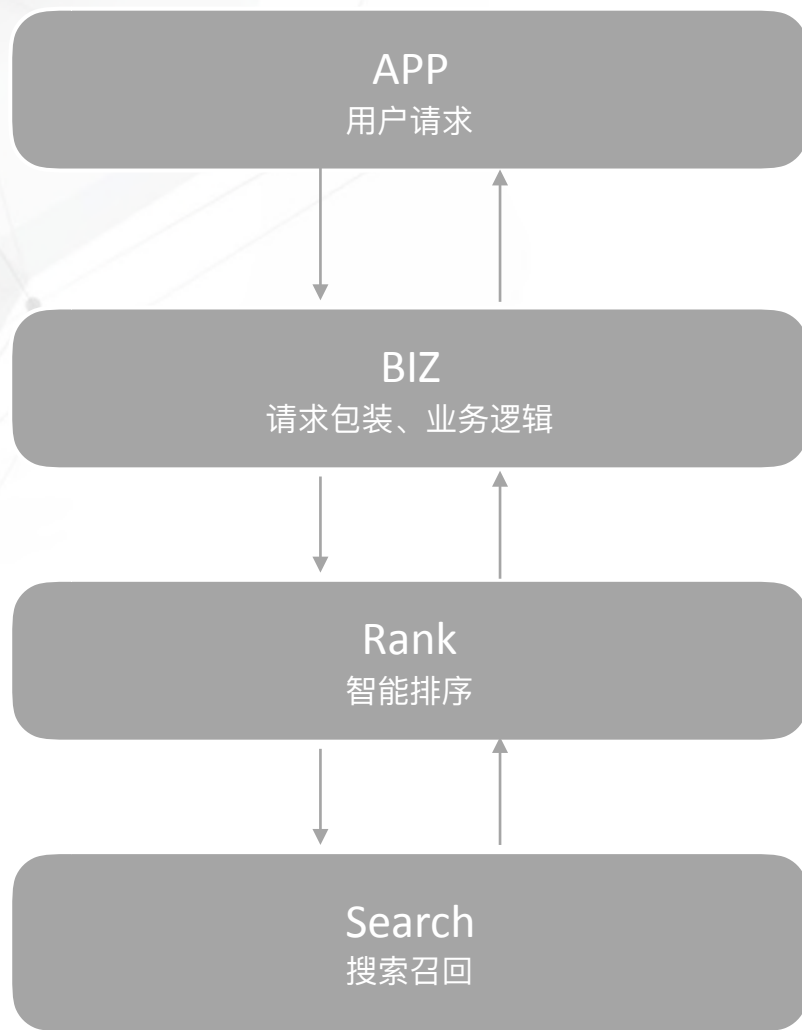
工具化提升效率



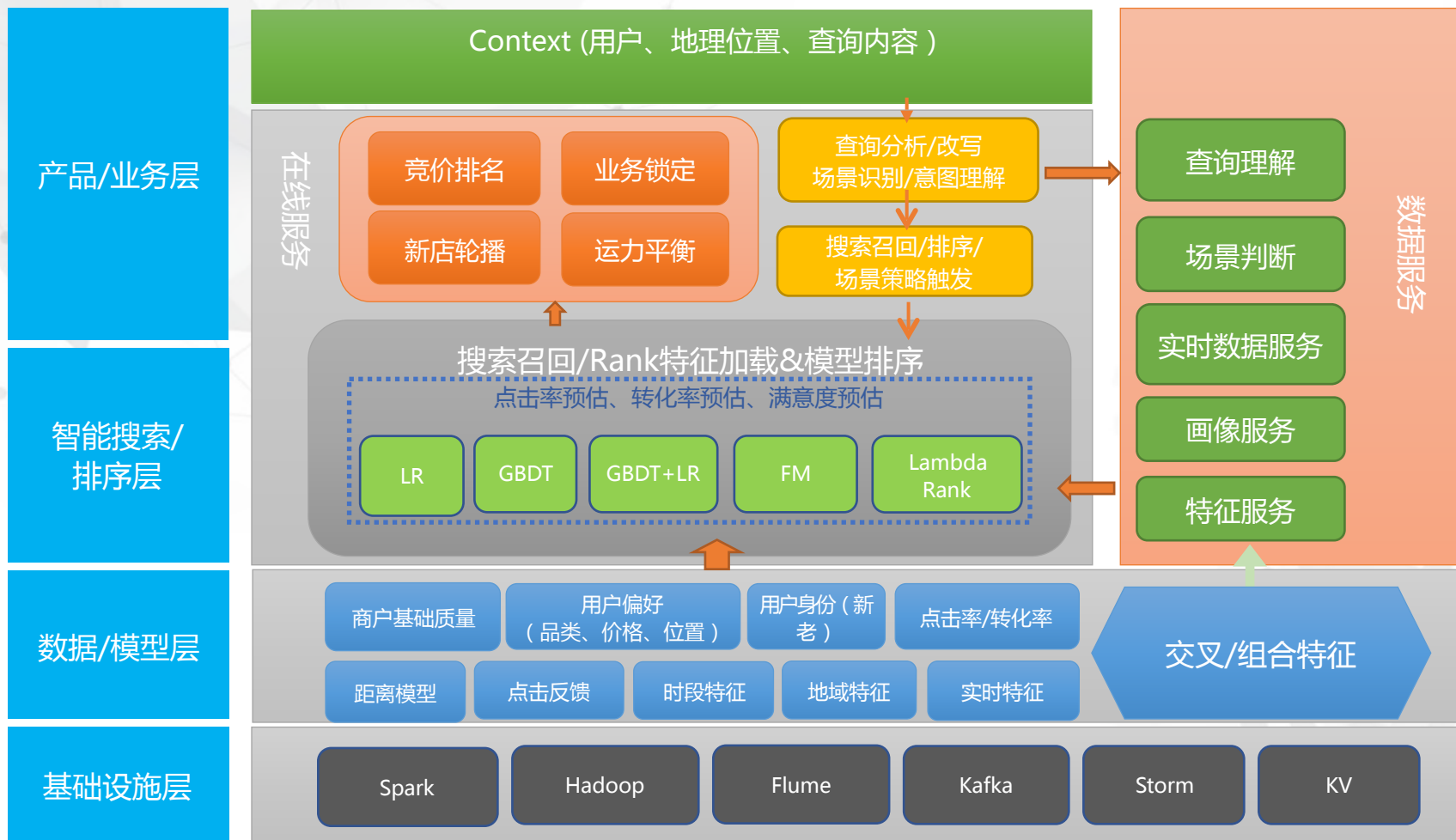
工程升级  
o2o分片搜索  
Rank工程架构改进

实时数据体系与在  
线学习

# 一次搜索请求的流程



# 搜索推荐技术架构





# 工程挑战及升级

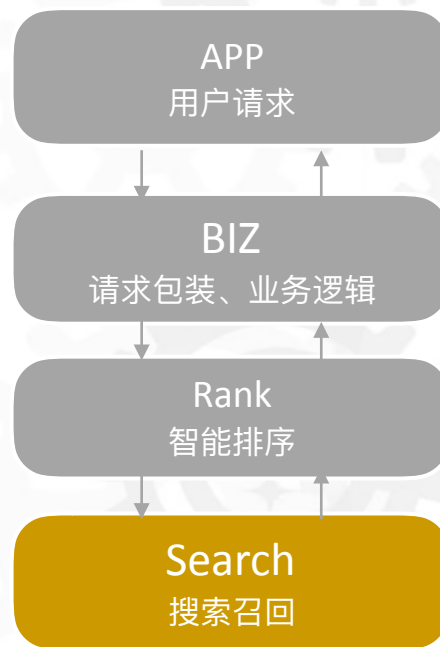
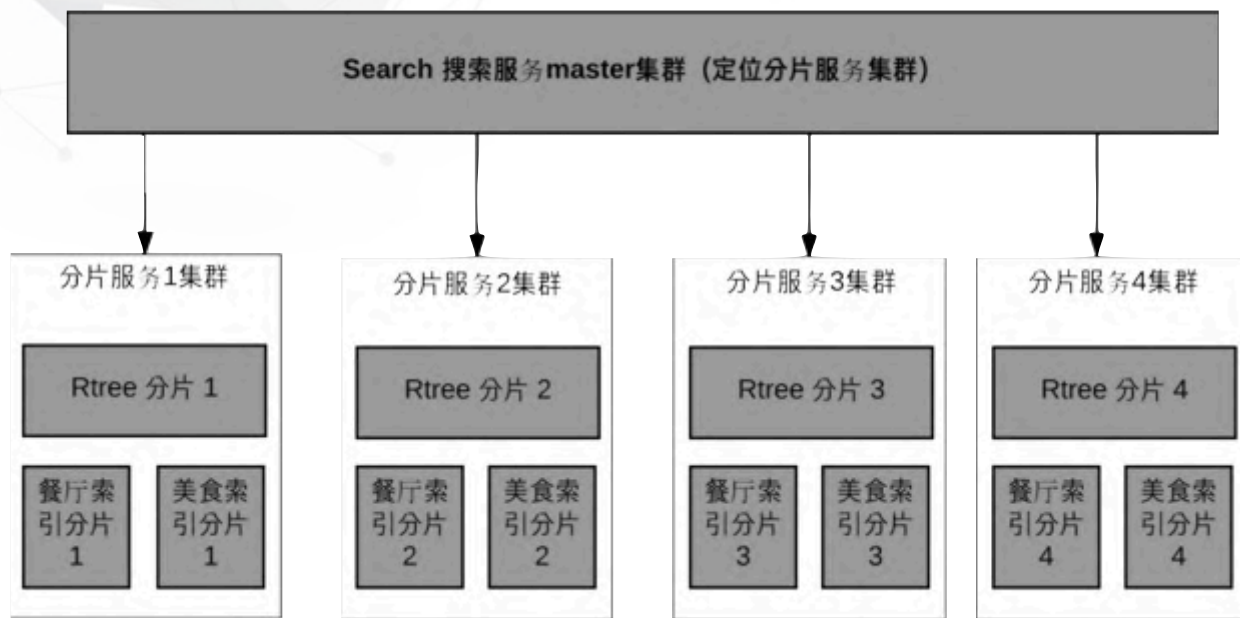
# 工程升级——o2o分片搜索

## 挑战

- 百万商户、上亿商品、上百个属性
- 空间索引、餐厅索引、美食索引等多维度索引
- 单节点性能受限于索引大小及复杂度

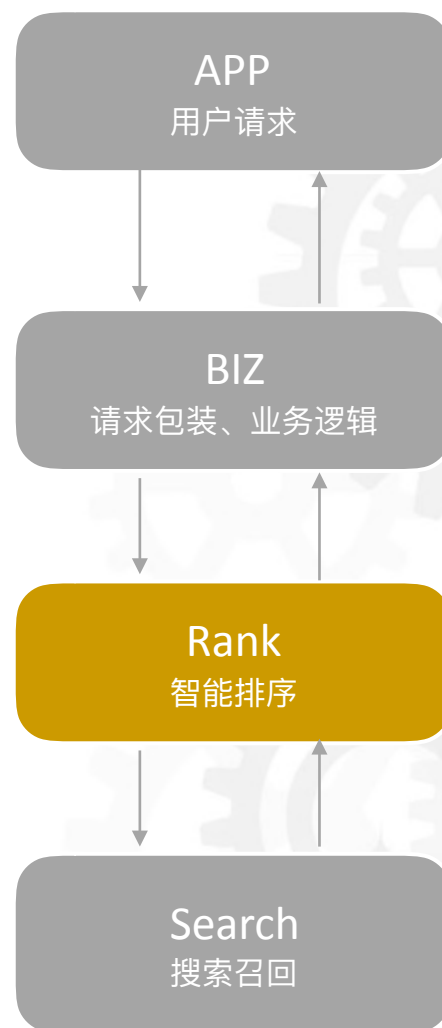
## 方案

- 基于商户配送范围进行分片，根据用户定位地址进行路由，横向拓展方便
- 单节点索引大小降为1/N，性能提升
- 集群的并发能力提升



# Rank工程挑战

- 高调用量、高QPS：
  - 核心入口，每日数亿次调用
  - 峰值效应明显，高峰期qps是正常10倍以上
- 重排数多：
  - 单次Rank重排数5000+
- 多属性、多来源
  - 一次Rank需用到300+属性及上百用户、Context属性
  - 涉及十多个业务数据源，来自搜索索引/MongoDB/Mysql/Redis/RocksDB/RPC
- 高时效性
  - 部分属性必须秒级更新
- 低时延
  - 30-40ms的平均响应时间

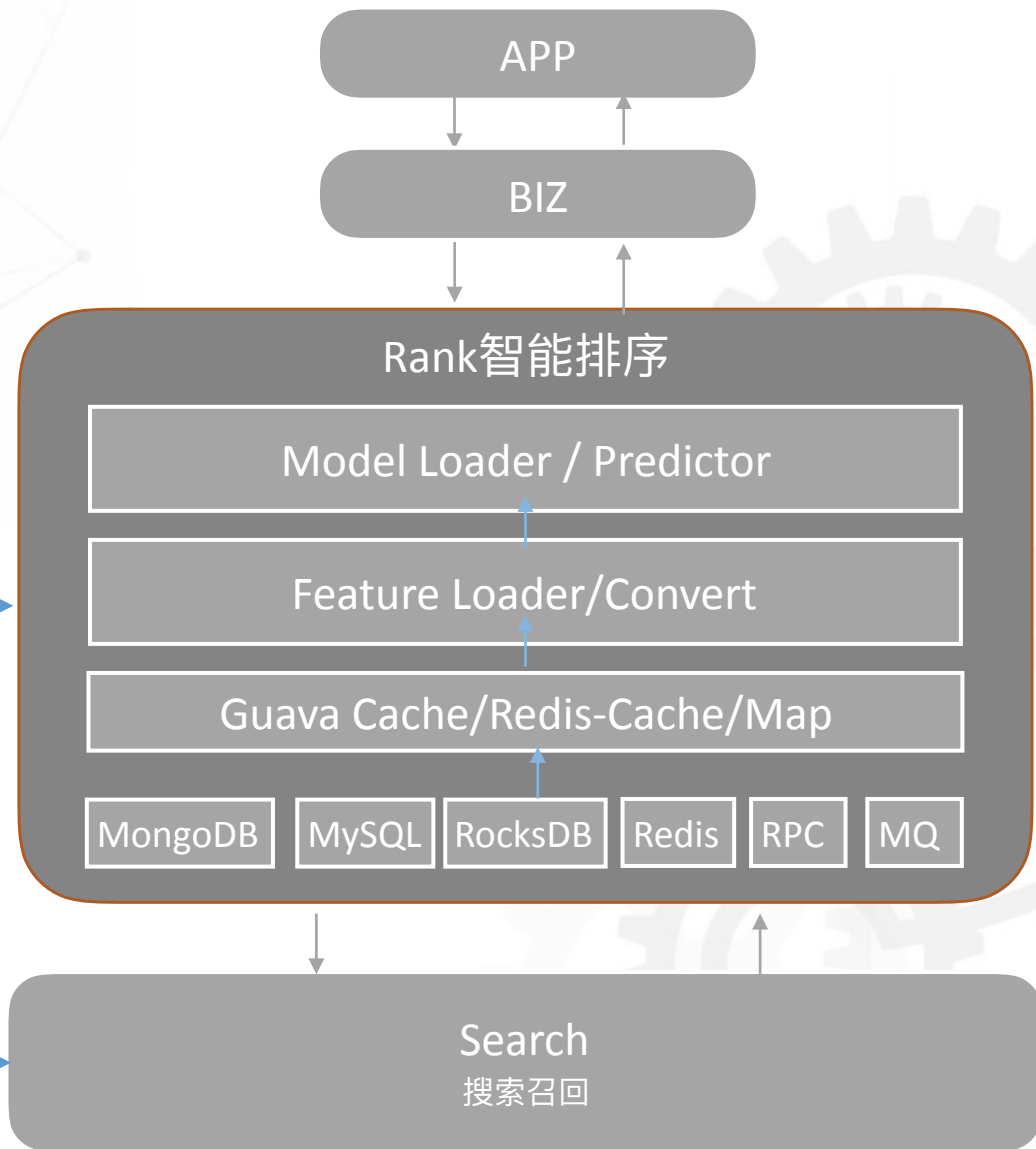


# Rank工程升级

## • 存储设计

- 从单一Mysql到引入RocksDB/Redis/Mongo等多种KV
- 不同KV针对不同需求，Redis和实时数据打通，RocksDB存储天级/小时级商户属性
- 低时延/高时效
  - 多种缓存，减少RPC调用和数据库访问
  - 接入MQ，主动更新状态缓存

- 通过MQ实时更新索引，过滤无效候选集
- 引入简单模型粗排，降低重排数



# Rank工程升级——商户特征存储方案

早期方案：MySQL

服务启动全量加载，发现新数据再异步全量加载

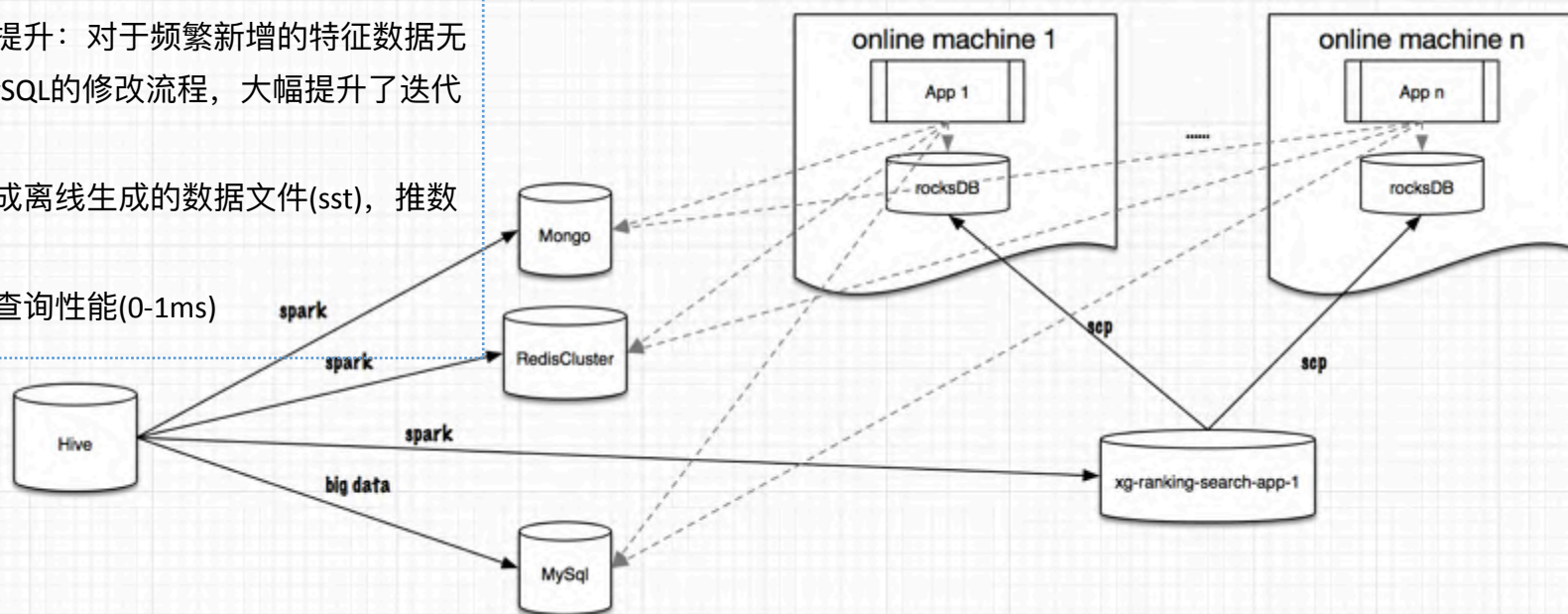
- 优点：服务启动成功即可接收请求, Mysql无压力
- 缺点：启动慢

过渡


- MySQL+Guava Cache做懒加载：启动迅速，但随着QPS升高和商户特征丰富，cache miss时MySQL的压力越来越大，重启抖动非常严重
- Mongo DB：导出数据成本高、开发不灵活
- 搜索索引：高IO成本

RocksDB

- 避免了大批量的请求落入数据库
- 减少IO成本：避免了数据传输的消耗
- 开发效率提升：对于频繁新增的特征数据无需再走MySQL的修改流程，大幅提升了迭代效率。
- 方便的集成离线生成的数据文件(sst)，推数方便
- 非常高的查询性能(0-1ms)







# 体验升级 Rank算法演进路线

# Rank算法演进路线



# Rank算法演进路线——数据升级

## 初版数据

### 数据规模

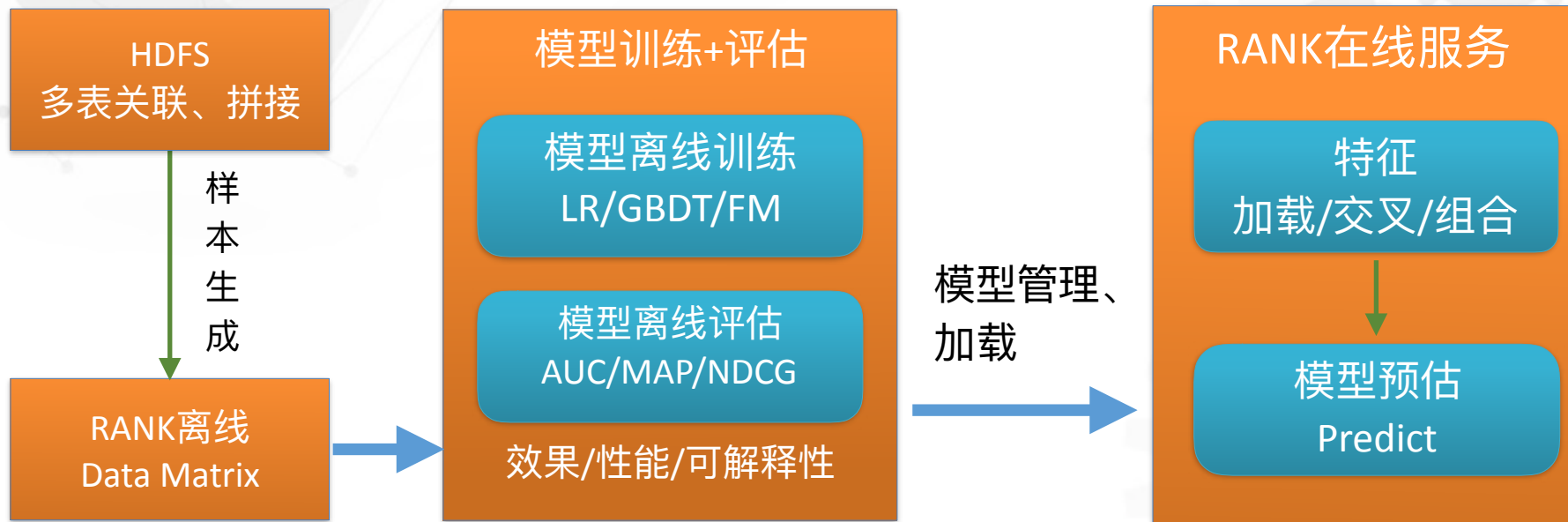
- 千万以内样本，特征几十维
- 单机训练

### 数据精准性

- 样本生成依赖于离线多表关联，线上  
线下数据经常不一致
- app埋点不精准

### 数据时效性

- 离线T+1样本
- 离线T+1特征
- 离线T+1报表



# Rank算法演进路线——数据升级

## 数据规模

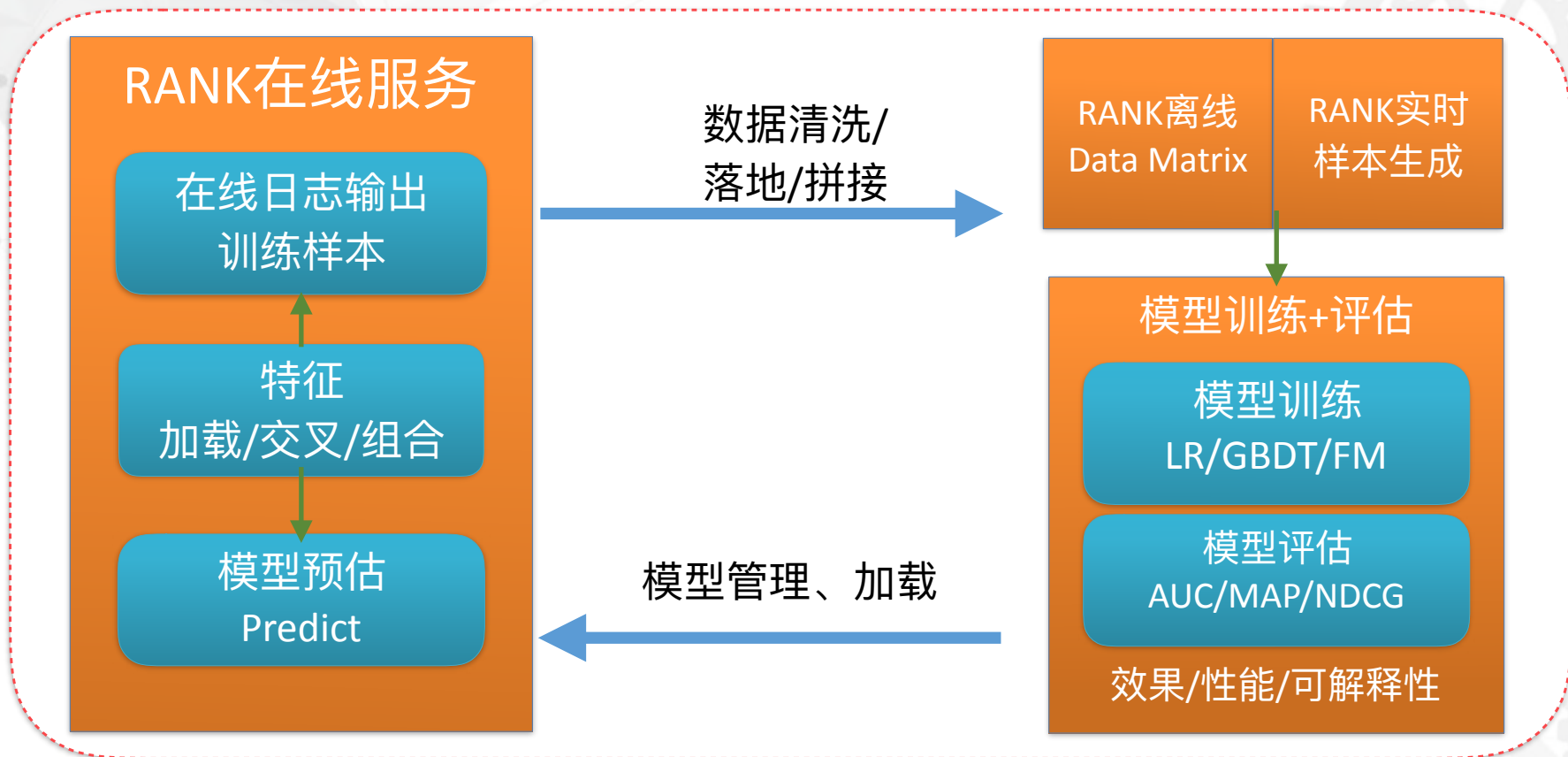
- 模型训练样本量高速增长
- 单机—>分布式

## 数据精准性

- 线上日志直接输出训练样本，避免离线数据关联的不一致
- 正负样本生成依赖于精准埋点

## 数据时效性

- 离线样本—>实时样本
- 离线特征—>实时特征
- 离线报表—>实时监控



# Rank算法演进路线——模型升级

## 复杂度

- 规则—>线性—>非线性—>融合模型

## 时效性

- 天级—>时段级—>小时级—>流式更新

## 优化目标

- 点击率(1)—>点击率+转化率+GMV(2) —>(2)+用户整体体验(3)

## 设计

- 精细粒度分时段、分场景模型 (offline) +在线模型 (online)
- 点击率、转化率、用户满意度融合模型



# Rank算法演进路线——特征升级

## 特征设计

- 人工设计—>人工设计+自动构造
- 分时段、分入口、分地域自动组合
- GBDT+LR、FM等

## 特征选择

- 信息增益、卡方、互信息
- 正则化+指标量化(权重、Auc、MAP)+特征集合搜索



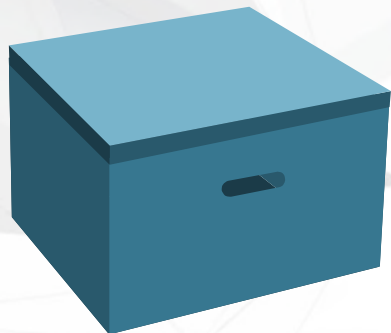
## 特征监控

- 覆盖率、范围、异常值
- 离线监控—>实时监控

## 特征时效

- 离线特征—>  
离线分时段—>  
离线+实时特征

# Rank算法演进路线——业务理解升级



## 单步优化—>多步优化 兼顾点击、转化与交易额

用户行为步骤:

浏览—>点击—>下单

点击率模型—>

点击率模型+转化率模型+客单价  
预估

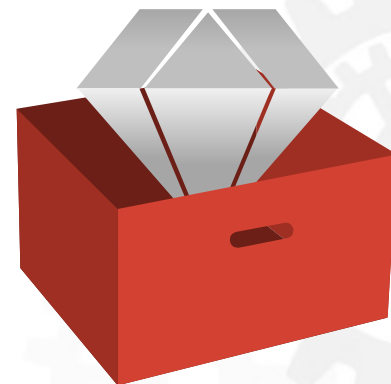


## 平台特殊目标 资源约束、时间敏感

交易转化—>

交易转化+用户决策效率+端到端  
时长

考虑餐厅负荷、运力负荷（策略  
调整）



## 兼顾生态平衡

政策倾斜/

新店加权/

老店拯救

防止过度个性化E&E  
各业务线流量分配均衡

# 实时数据体系与在线学习



# 实时数据体系设计

## ❖ 应用层

特征工程

策略控制

在线学习

实时报表

## ❖ 数据类型

曝光

点击

下单

加购

搜索

收藏

## ❖ 基础架构

Storm

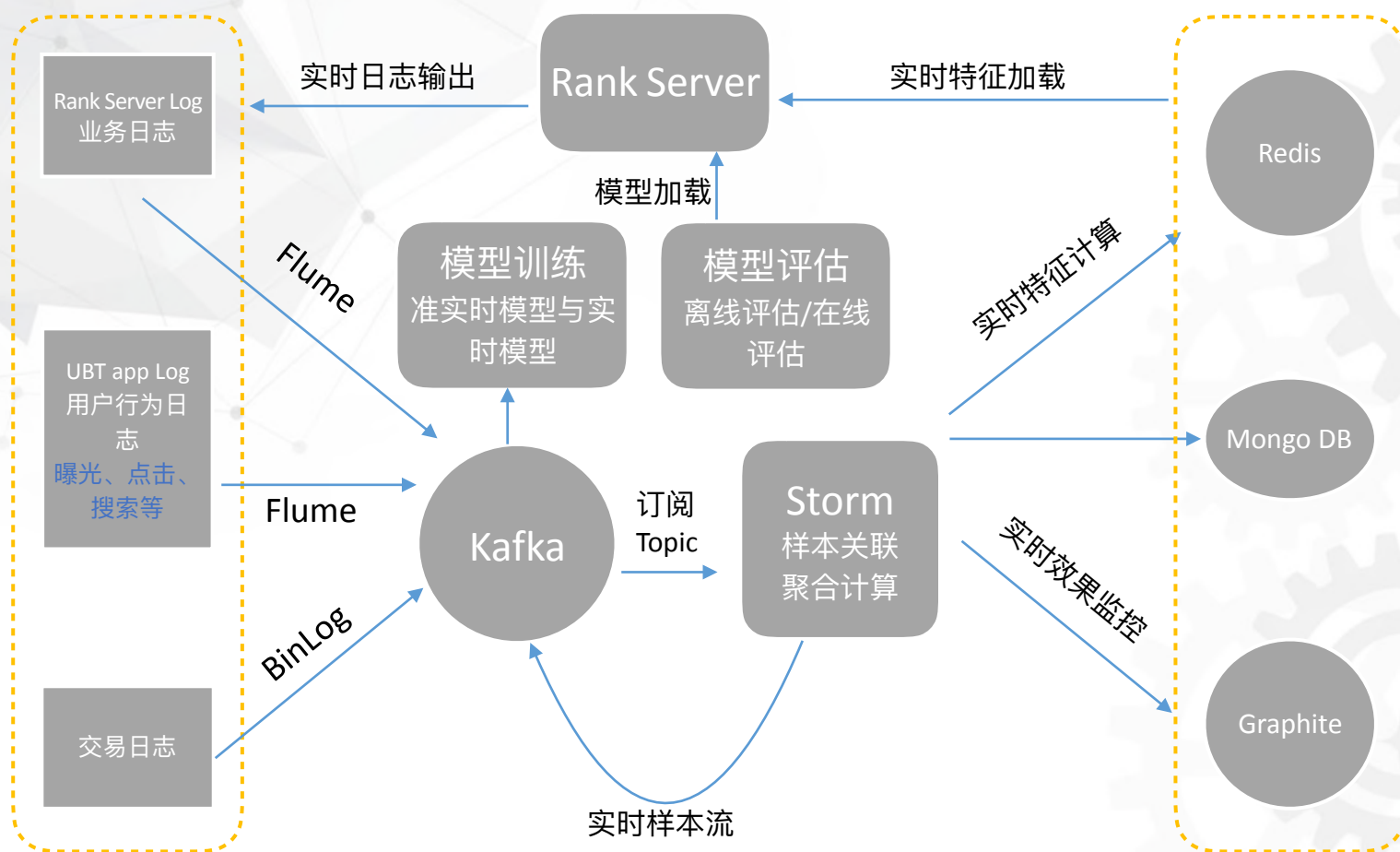
Kafka

Redis

DRC

# 实时数据体系与在线学习

实时样本生成、在线训练、特征工程、实时效果报表



# 实时数据体系——实时报表

策略效果即时感知&即时决策

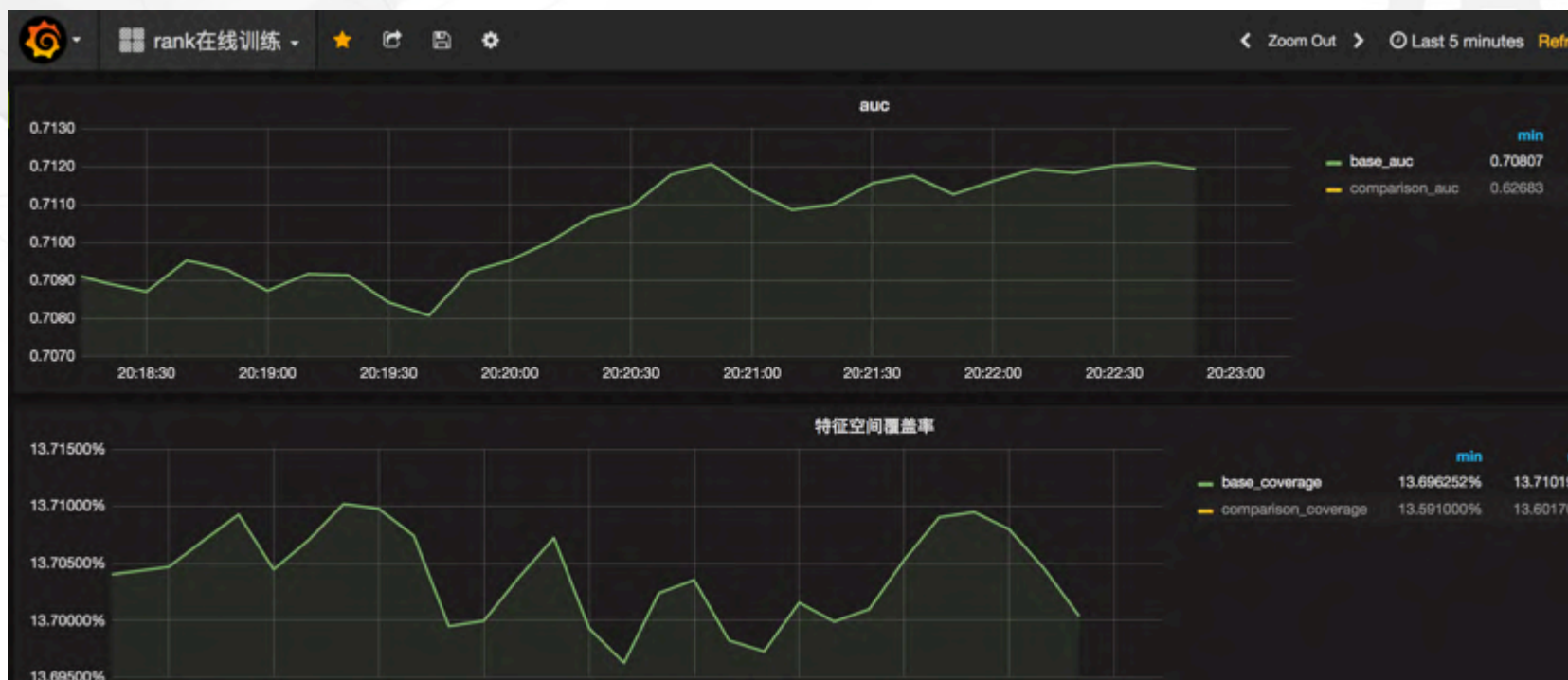
分入口、分算法版本、分指标



# 实时数据体系与在线学习

## 在线评估 vs 离线评估

- 在线auc与特征覆盖率监控
- 防止线上线下数据不一致带来的评估部准确问题
- 效果评估即时反馈



# 工具化提升效率

# 工具化提升效率——排名解释系统

## 作用

- 技术、产品、业务目标协同
- 快速定位排名原因
- 兼顾效果与可解释性

## 动态灵活

- 真实场景模拟—指定地址及入口
- 指定用户信息排查
- 指定不同算法版本对比

## Rank因素

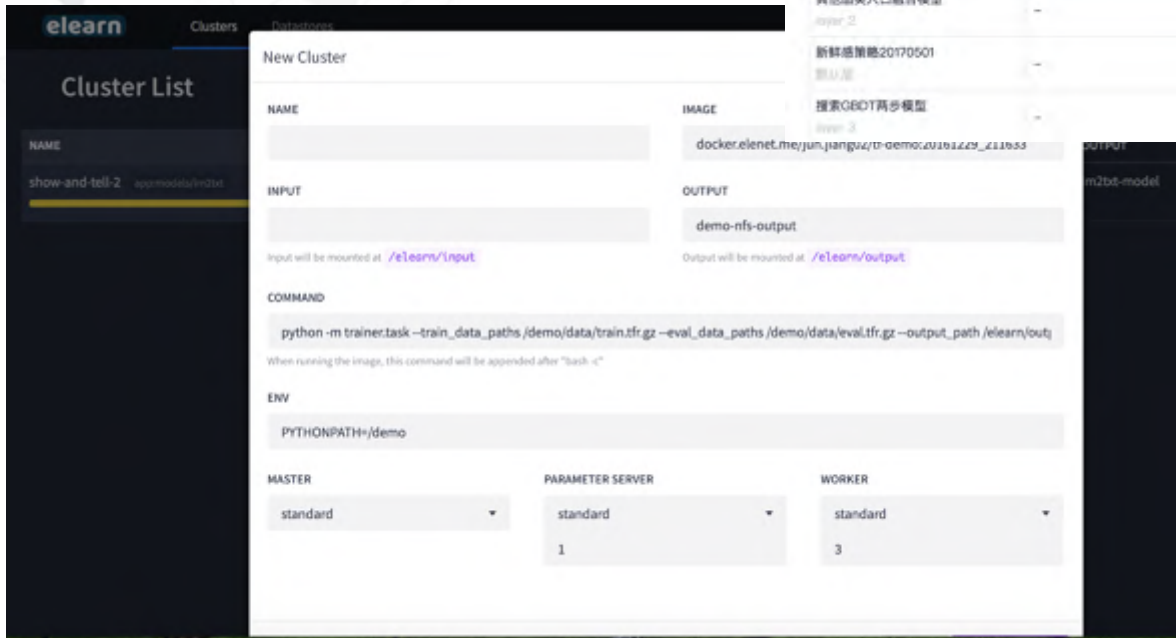
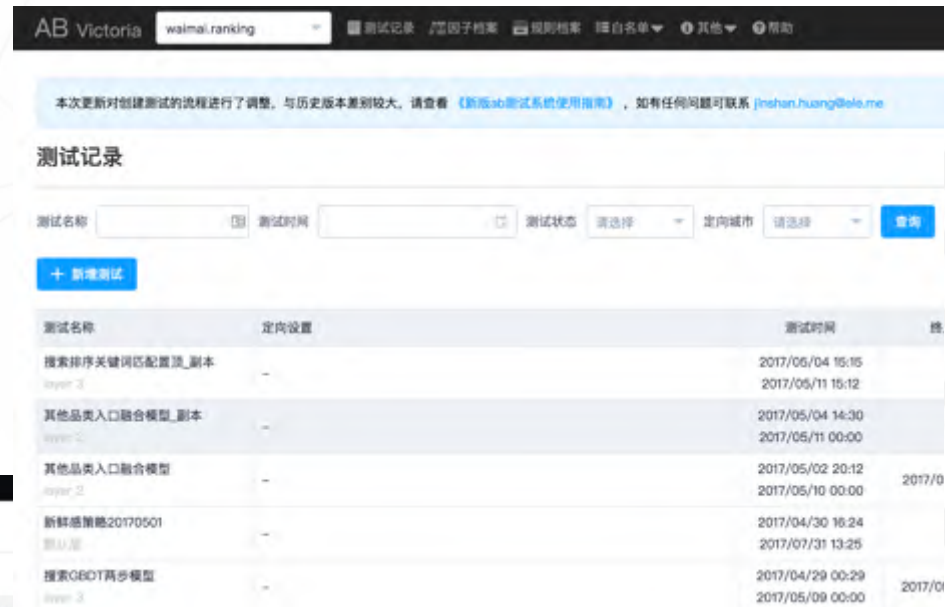
- 展示排名决定因素
- Top Feature 打分明细
- 政策、产品、业务因素



# 工具化提升效率——AB引擎/训练平台

## AB引擎

- 集成分流与报表功能
- 可视化配置
- 方便高速算法迭代
- 白名单便于算法模拟及预演



## 训练平台

- 模型训练与评估
- 可视化配置
- 图形化指标展示
- 交互性强

# 工具化提升效率——Rank在线模拟&Debug系统

可视化集成

- App效果同步展示
- 精准展示曝光、点击、下单

算法信息校验

- 特征信息、算法版本
- 可视化Debug

UBT日志 交易日志 RANK日志 RANK效果 实时热图 1064851 查询 停止刷新

```
rank_id: 2915c0c8ea0546a6b7416db75398ad22, user_id: 1064851, module: CHANNEL, subModule: CHANNEL_FOOD, algoVersion: algo_default, appVersion: IOS_7.7.1, cityId: 1
```

```
0: {"exposure":1,"index":0,"shopId":2151030,"shopInfo":{"brandName":"御膳糖糖 百联中环店","score":829.7164,"rankType":"自然排名"}
1: {"exposure":1,"index":1,"shopId":2320841,"shopInfo":{"brandName":"妙厨(金沙江路店)","score":1079.9321,"rankType":"自然排名"}
2: {"exposure":1,"index":2,"shopId":146122,"shopInfo":{"brandName":"味千拉面 近铁店","score":828.79395,"click":1,"rankType":"自然排名"}
3: {"exposure":1,"index":3,"shopId":202767,"shopInfo":{"brandName":"汉堡王 118J 旗舰店20864","score":824.7323,"rankType":"自然排名"}
4: {"orderInfo":{"totalFee":37,"discountInfo":[{"id":102,"price":3,"name":"餐盒"}, {"id":12,"price":0,"name":"百事可乐/7喜1份"}], "id":300,"price":4,"name":"会员减配送费"}, "deliverFee":5, "orderId":"1206139356010419276"}, {"exposure":1,"index":4,"order":1}
5: {"exposure":1,"index":5,"shopId":920068,"shopInfo":{"brandName":"丰收日 近铁广场店","score":800.5636,"click":1,"rankType":"自然排名"}
6: {"exposure":1,"index":6,"shopId":2157036,"shopInfo":{"brandName":"小杨生煎 118店","score":786.24255,"click":1,"rankType":"自然排名"}
7: {"exposure":1,"index":7,"shopId":146086,"shopInfo":{"brandName":"曾田稻香(近铁广场)","score":773.3600,"rankType":"自然排名"}
8: {"exposure":1,"index":8,"shopId":155336699,"shopInfo":{"brandName":"福新鸡","score":0,"rankType":"新店轮播"}
9: {"exposure":1,"index":9,"shopId":159904,"shopInfo":{"brandName":"开心厨房","score":0,"rankType":"竞价排名"}
10: {"index":10,"shopId":402075,"shopInfo":{"brandName":"新白鹿 百联中环","score":772.032,"rankType":"自然排名"}
11: {"index":11,"shopId":360126,"shopInfo":{"brandName":"棒约翰比萨(金沙江路店)","score":756.55695,"rankType":"自然排名"}
12: {"index":12,"shopId":166657,"shopInfo":{"brandName":"四海游龙台湾风味坊 近铁店","score":706.0028,"rankType":"自然排名"}
13: {"index":13,"shopId":135158,"shopInfo":{"brandName":"食其家 118广场店","score":700.5586,"rankType":"自然排名"}
14: {"index":14,"shopId":26931,"shopInfo":{"brandName":"好宴一家","score":428.49133,"rankType":"竞价排名"}
15: {"index":15,"shopId":1396802,"shopInfo":{"brandName":"有馍有样腊汁肉夹馍","score":667.35913,"rankType":"自然排名"}
16: {"index":16,"shopId":301187,"shopInfo":{"brandName":"避风塘 中环店","score":663.96704,"rankType":"自然排名"}
17: {"index":17,"shopId":151696333,"shopInfo":{"brandName":"沙县小吃 定边路","score":0,"rankType":"新店轮播"}
18: {"index":18,"shopId":389037,"shopInfo":{"brandName":"达美乐比萨 中江路店","score":645.8556,"rankType":"自然排名"}
```

Features

```
[0:0.7906209, 1:1140.6, 2:14677.155, 3:14764.0, 4:0.2242778, 5:49.822697, 6:309.6, 7:48,26424, 8:42.738297, 9:28,332474, 10:82.71978, 11:47.40663, 12:0.07305776, 13:0.06628039, 14:0.013261057, 15:304.1, 16:0.62866473, 17:43.39598, 18:1455.9, 19:20902.2, 20:22,685371, 21:0.90068024, 22:244854.12, 23:5145.0, 24:0.0016662983, 25:4.448367, 26:128.0, 29:1.0, 30:1.0, 31:1.0, 32:1.0, 34:1.0, 35:1.0, 36:1.0, 37:1.0, 38:1.0, 40:20.0, 45:1.0, 46:1.0, 47:1.0, 49:1.0, 51:1.0]
```



# 工具化提升效率——搜索运营/Debug系统

## 运营系统

- 词库管理/标签管理/运营配置
- 抽检/标注
- 降低人工介入门槛
- 即时配置、即时生效

## Debug系统

- case排查
- 一键模拟
- 可视化数据校验

## 关键字搜索接口

关键字:

经纬度: 121.38098984958741, 31.23211920261383 [地图](#)

分类名:

筛选项 (“属性值”为必选):

1. 配送方式:  蜂鸟专送  空
2. 优惠活动:  新用户优惠  特价秒杀  下单立减  赠品优惠  进店领券  下单返券  空
3. 人均消费:  20元以下  20-40元  40元以上  空
4. 商家属性:  在线支付  开发票  新店  外卖保  品牌商家  准时达  属性值

服务器选项:  server1  server2  server3  server4  server5  server6  server7  server8  server9  ppe

## 搜索运营管理后台

搜索运营管理后台

- 基础词库Excel操作
- 基础词库管理
- 同义Excel操作
- 同义词管理
- 热词管理
- 品牌词同步
- 关键字白名单

关键词

只显示有效词  否  是

词类型  全选  品牌名  商户名  美食名  品类词  属性词  场景词

修改开始日期

修改结束日期

修改人

关键词	词类型	关键词所属品类
大理香云菌过桥米线	商户名,品牌名 419137	米粉丝面-100,香锅砂锅-100,麻辣烫-100
不二龙虾	商户名,品牌名 415185	海鲜:100,蛋炒饭:100,小龙虾:100
香芒记	商户名,品牌名 417212	包子粥店-100,地方小吃:100,米粉饭团:100
韩式炸鸡炸鸡店	商户名,品牌名 4148	炸鸡炸鸡-100,地方小吃:100,日韩料理:100
二人帮	品牌名 422123	

### Navigation

- 关键字管理
  - 关键字管理
  - 新建/更新/删除
  - 热词管理
  - 一键模拟
  - 一键标注
- 运营配置
  - 新美食上传
  - 优惠配置
  - 去无折扣
  - 满减活动配置
- 管理入口
  - 快捷入口
  - 人工标注
- 新建/删除
- 操作入口
  - 新建入口
  - 删除入口
  - 标注入口
  - 删除入口

```

{
  "brand": {
    "brand_id": 1, "brand_name": "肯德基", "brand_type": "品牌名", "brand_status": "正常", "brand_desc": "肯德基", "brand_logo": "http://img.kfc.com.cn/kfc/brand/logo/1.png", "brand_url": "http://www.kfc.com.cn", "brand_phone": "4009988888", "brand_address": "上海市浦东新区世纪大道100号", "brand_latitude": 31.23211920261383, "brand_longitude": 121.38098984958741, "brand_created_at": "2017-04-27 22:30:05", "brand_updated_at": "2017-04-27 22:30:05"}
  },
  "item": {
    "item_id": 1, "item_name": "香辣鸡腿堡", "item_type": "品牌名", "item_status": "正常", "item_desc": "香辣鸡腿堡", "item_logo": "http://img.kfc.com.cn/kfc/item/logo/1.png", "item_url": "http://www.kfc.com.cn/item/1", "item_phone": "4009988888", "item_address": "上海市浦东新区世纪大道100号", "item_latitude": 31.23211920261383, "item_longitude": 121.38098984958741, "item_created_at": "2017-04-27 22:30:05", "item_updated_at": "2017-04-27 22:30:05"}
  },
  "promotion": {
    "promotion_id": 1, "promotion_name": "满减活动", "promotion_type": "品牌名", "promotion_status": "正常", "promotion_desc": "满减活动", "promotion_logo": "http://img.kfc.com.cn/kfc/promotion/logo/1.png", "promotion_url": "http://www.kfc.com.cn/promotion/1", "promotion_phone": "4009988888", "promotion_address": "上海市浦东新区世纪大道100号", "promotion_latitude": 31.23211920261383, "promotion_longitude": 121.38098984958741, "promotion_created_at": "2017-04-27 22:30:05", "promotion_updated_at": "2017-04-27 22:30:05"}
  },
  "activity": {
    "activity_id": 1, "activity_name": "优惠券", "activity_type": "品牌名", "activity_status": "正常", "activity_desc": "优惠券", "activity_logo": "http://img.kfc.com.cn/kfc/activity/logo/1.png", "activity_url": "http://www.kfc.com.cn/activity/1", "activity_phone": "4009988888", "activity_address": "上海市浦东新区世纪大道100号", "activity_latitude": 31.23211920261383, "activity_longitude": 121.38098984958741, "activity_created_at": "2017-04-27 22:30:05", "activity_updated_at": "2017-04-27 22:30:05"}
  }
}

```

有效 2017-04-27 22:30:05

# 一些感悟

- 建立健康合理、符合公司战略目标的业务指标
- 强化业务理解，技术驱动业务
- 重视技术，但不唯技术
- 基于现有体系快速迭代

# 欢迎关注

饿科技



## We are hiring

- 算法工程师
- 搜索工程师
- Java工程师
- 深度学习专家
- 自然语言处理工程师

[yao.ma02@ele.me](mailto:yao.ma02@ele.me)



# THANKS

SequeMedia  
威拓传媒

IT168.com

ITPUB

ChinaUnix