



QCon 全球软件开发大会
INTERNATIONAL SOFTWARE
DEVELOPMENT CONFERENCE

BEIJING 2017

蘑菇街广告的排序：从历史数据学习到个性化 强化学习

蘑菇街邓钦华（花名问天）



促进软件开发领域知识与创新的传播



关注InfoQ官方信息
及时获取QCon软件开发者
大会演讲视频信息



扫码，获取限时优惠

ArchSummit
全球架构师峰会 2017 [深圳站]

2017年7月7-8日 深圳·华侨城洲际酒店

咨询热线：010-89880682

QCon

全球软件开发大会 [上海站]

2017年10月19-21日

咨询热线：010-64738142

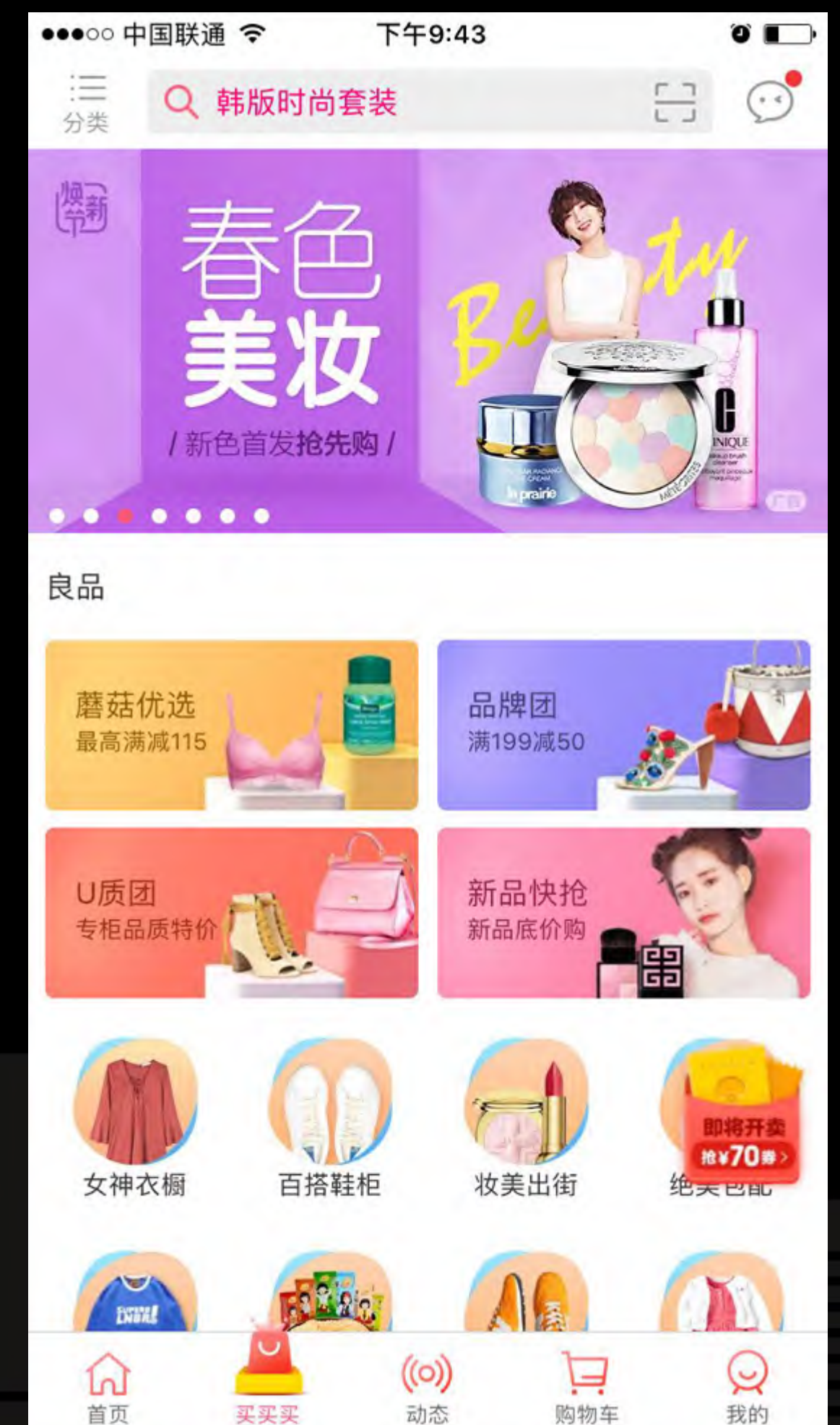
个人介绍

- 邓钦华，花名问天，在蘑菇街负责搜索系统、广告系统、流量系统、图像算法和机器学习团队
- 一直从事搜索推荐、机器学习和大数据系统的研发实践，参与开发过百度统计、百度关键词推荐、百度搜索广告系统、360 搜索广告系统、360 展示广告系统、360 推荐系统、迅雷大数据平台、迅雷数据统计分析平台等产品，从零搭建了蘑菇街广告体系、流量体系和搜索体系，并将图像技术用于搜索的排序。

目录

- 蘑菇街和广告业务介绍
- 传统广告排序：从历史数据学习
- 传统广告排序在蘑菇街场景遇到的问题
- 我们的一些经验：个性化强化学习
- QA

蘑菇街和广告业务介绍



蘑菇街pc和app的页面

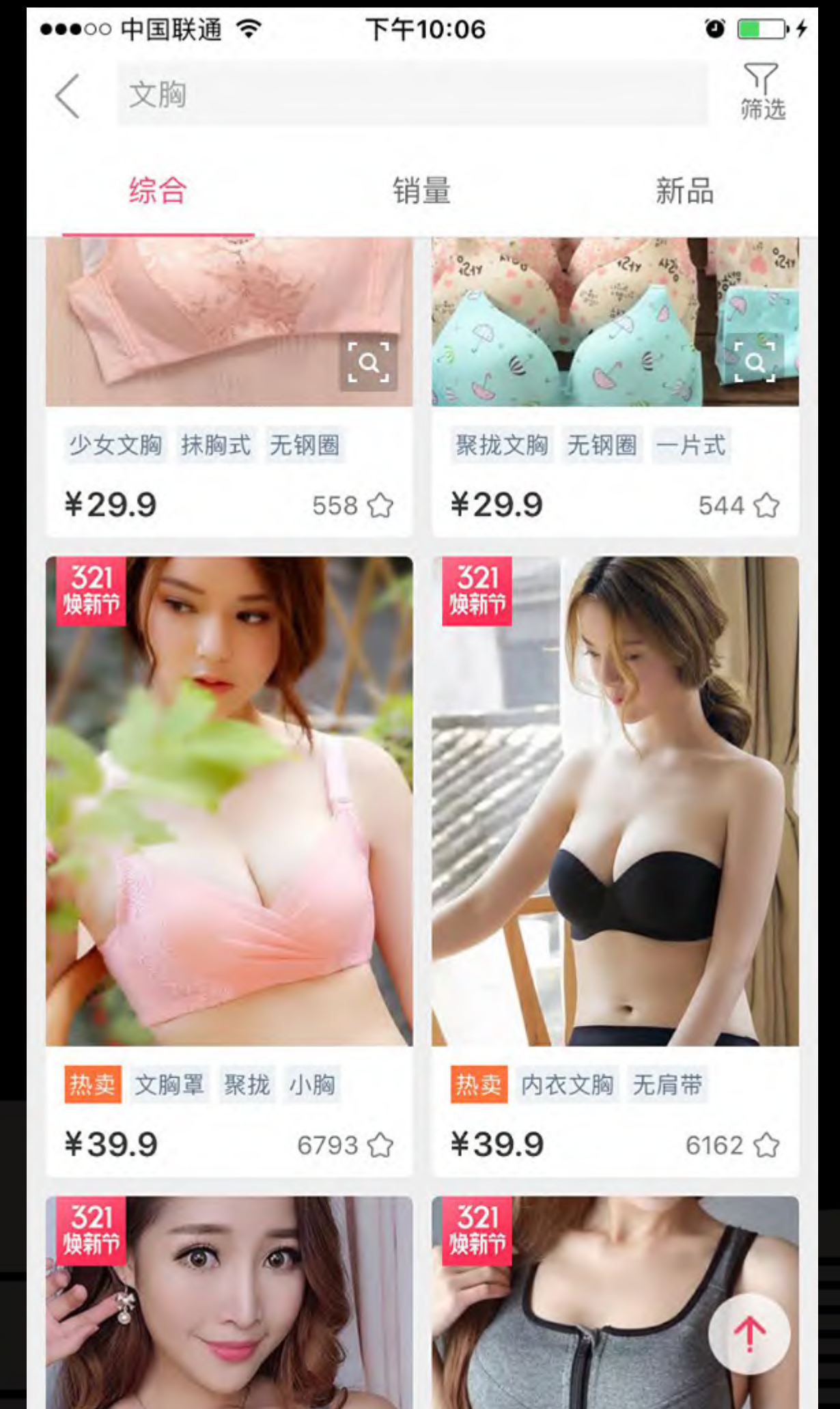
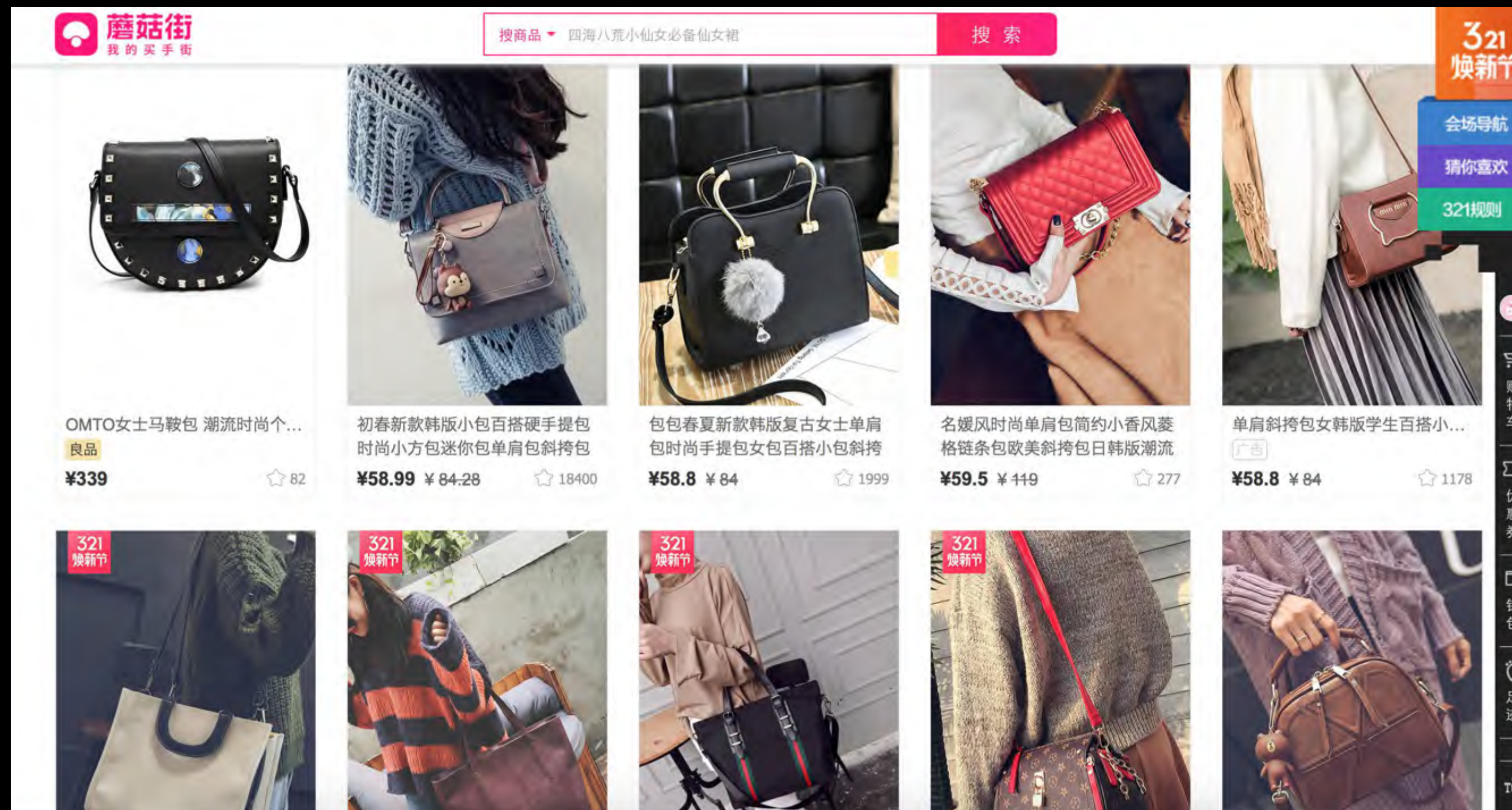
蘑菇街和广告业务介绍

- 蘑菇街业务介绍
 1. 导购+电商
 2. 买手优选+红人经济
 3. 社会化电商



蘑菇街和广告业务介绍

- 原生广告，广告结果和自然结果混合在一起
- cpc计费，广义二阶竞价
- 独立的广告位置+广告投放业务系统



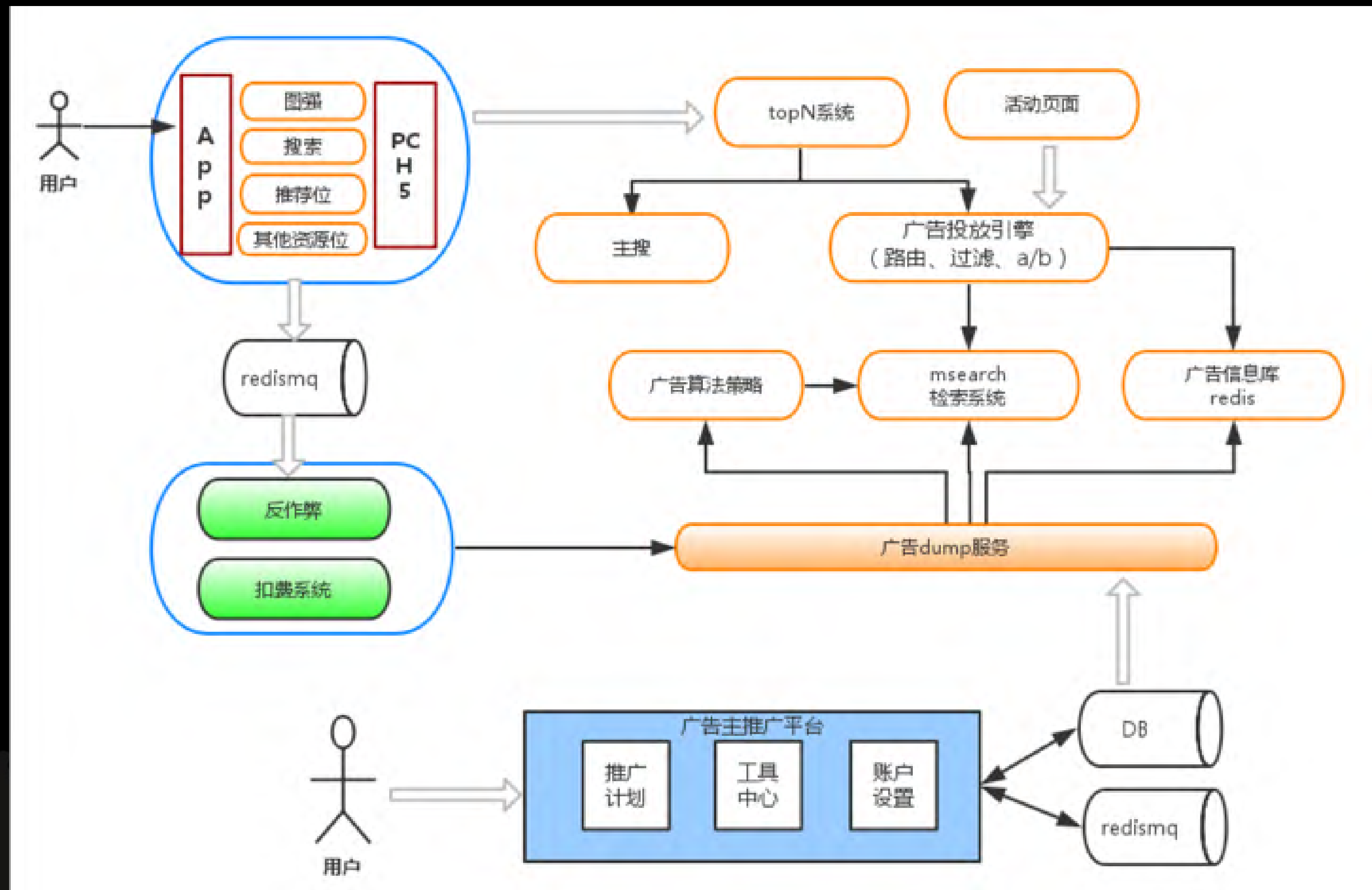
蘑菇街和广告业务介绍

- 蘑菇街业务特点

1. 用户逛街式浏览，图片的美观对用户体验影响很大
2. 移动端访问占比超过80%，用户浏览量大
3. 商品存在明显的时效性和季节周期性
4. 商家在库存压力下，对流量获取的稳定性和可控性存在天然的诉求

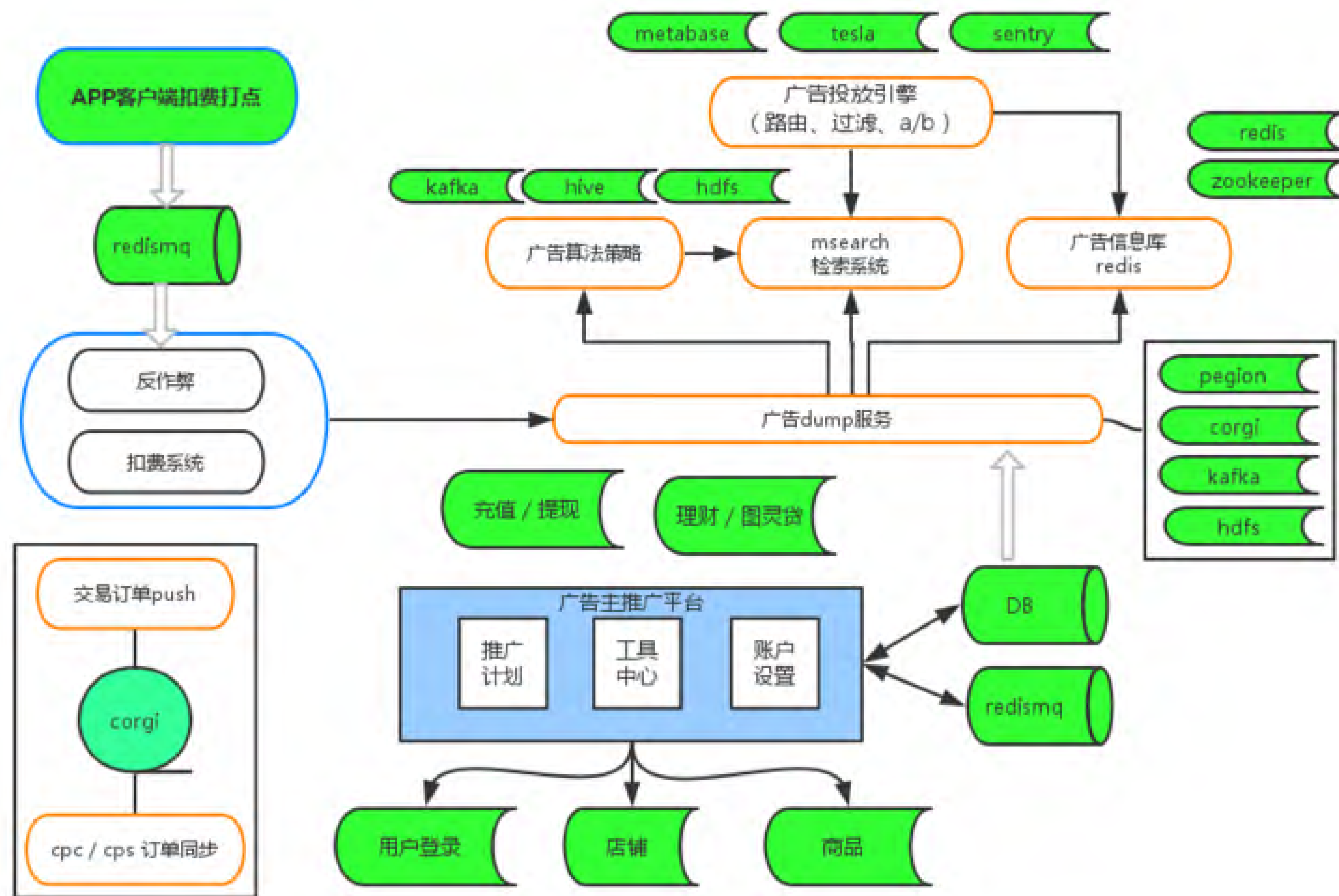
蘑菇街和广告业务介绍

- 蘑菇街广告业务



蘑菇街和广告业务介绍

• 蘑菇街广告业务



传统广告排序：从历史数据学习

- 广告排序历史

1. 竞价排名，发明者overture

优点？问题？

2. 进化：质量度*出价排名+广义二阶竞价，发明者google

优点？问题？

传统广告排序：从历史数据学习

- 如何预估质量度：点击率预估

1. 统计性模型

优点？问题？

2. 统计机器学习模型

LR：

$$P(Y = 0|X) = \frac{1}{1 + \exp(w_0 + \sum_{i=1}^n w_i X_i)}$$

传统广告排序：从历史数据学习

- Why LR ?
 1. 概率模型，预测值有物理含义，而不是简单的排序值
 2. 最简单的non-trivial 算法，预测值更平滑
 3. 简单够快，容易处理稀疏问题+易于并行实现

传统广告排序：从历史数据学习

数据日志

用户反馈

实时数据

在线服务_{服务}

离线数据传输

在线基础服务

在线数据挖掘

离线数据处理和数
据挖掘

离线数据模型

传统广告排序：从历史数据学习

- 进一步

1. GBRT+LR

变化？

2. Online learning : FTRL

变化？



传统广告排序在蘑菇街场景遇到的问题

商品的季节性和时效性

用户逛街式浏览的导购需求

个性化和时尚趋势的变化

我们的一些经验：个性化强化学习



收益和代价



ee广告展现形式



商家是否要参加



我们的一些经验：个性化强化学习

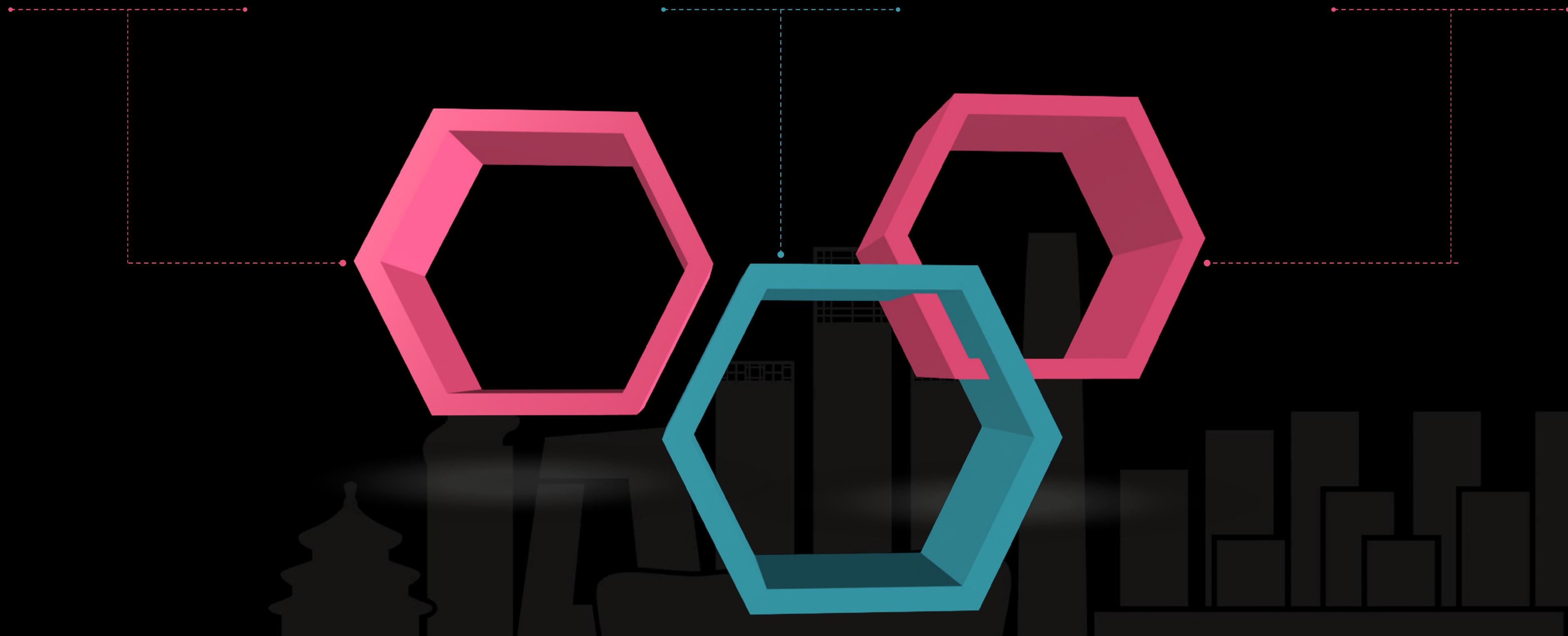
- 我们的一些思考：产品
 1. 新广告的比例和长期收益，评估指标是什么？
 2. 混排还是独立位置？
 3. 商家参与好处和缺点？

我们的一些经验：个性化强化学习

强化学习模型和传统模型如何结合

如何挑选商品和用户：刻画能力和泛化能力

个性化和强化学习的结合



我们的一些经验：个性化强化学习

- 我们的思考：系统
 1. 分阶段模型？Bayes模型？
 2. explore不止一个商品，还要考虑商品的推广型
 3. 新品的个性化初始概率

我们的一些经验：个性化强化学习

- bayes模型
 1. 商品冷启动相当于先验，先验分布为beta分布，商品的先验概率预估
 2. 好处？缺点？
 3. 置信度 and 样本抽样

我们的一些经验：个性化强化学习

- 强化学习
 1. epsilon-Greedy
 2. **Naïve UCB1**
 3. **UCB Bernoulli fixed 95% confidence intervals**
 4. **UCB1-Tuned**
 5. **UCB-SEGMENT**
 6. **Contextual Bandit (linucb)**

我们的一些经验：个性化强化学习

- 强化学习：UCB Bernoulli fixed 95% confidence intervals

我们假设商品是否点击符合二项分布，那么在我们的bandit算法中，我们就使用Binomial confidence intervals来计算每个商品的置信上界，每次进行离线算法分更新时，我们可以可以知道：

Totals -- 商品总计被选择的次数；

Success -- 商品展现并被点击的次数

1. Estimate_Mean = Success / Totals = \hat{p}

2. Estimate_Variance = $\hat{p} (1 - \hat{p})$

3. standard deviation = $\sqrt{\hat{p} (1 - \hat{p})}$

- 将该符合二项式分布的随机变量归一化转化为N(0, 1)分布，通过95%的置信区间，可以得到需要的上界：

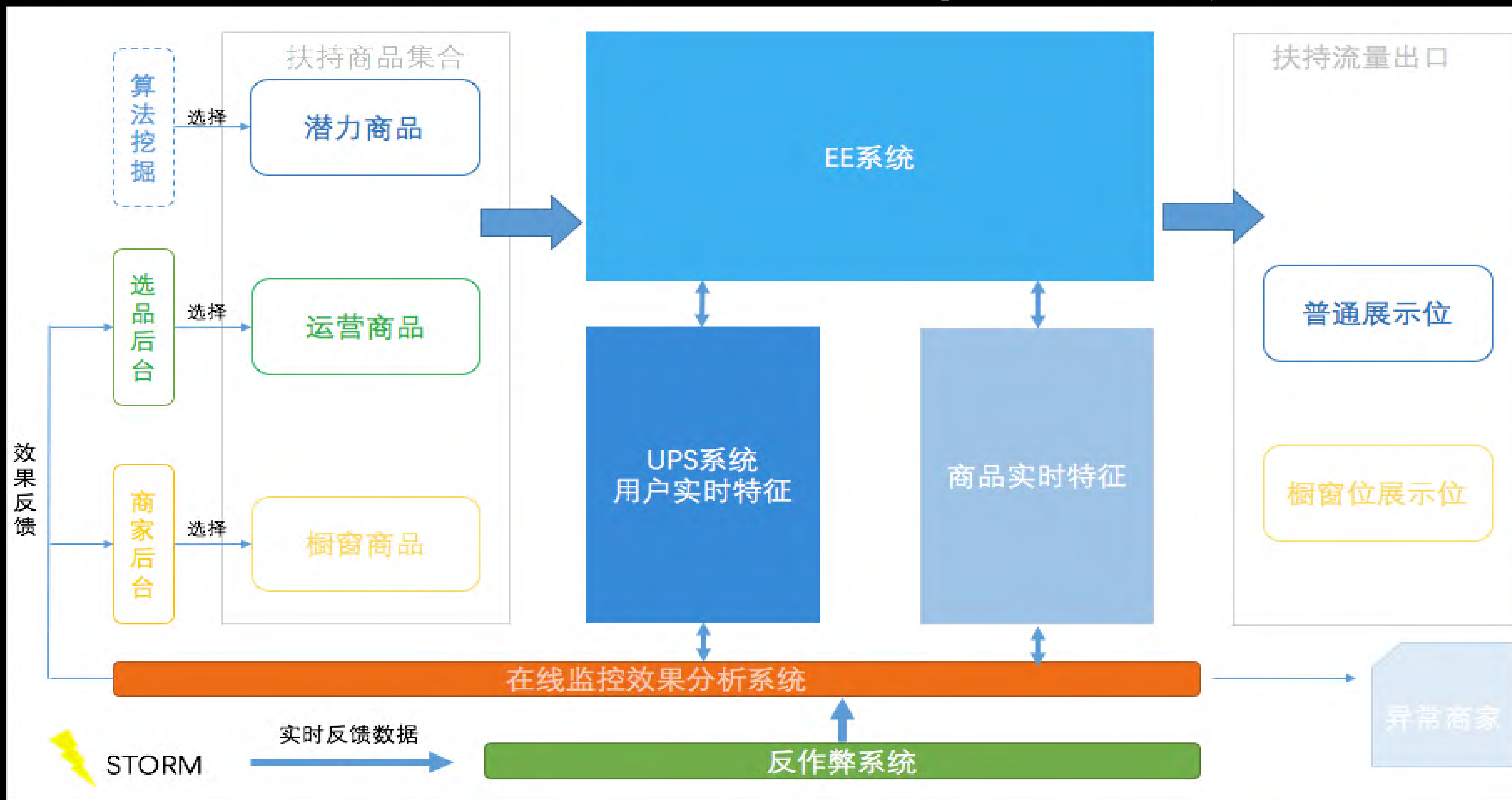
- $S_{bandit} = \hat{p} + 1.96\sqrt{\hat{p} (1 - \hat{p})/Totals}$

我们的一些经验：个性化强化学习

- 一些优化

1. $S_{algo} = \gamma S_{bandit} + \delta S_{round-robin}$
2. 每个样本的权重不一样，用户权威度
3. 用户分群和商品属性分类的更新
4. 概率的个性化预估、初始概率预估
5. 参数的自动化调整

我们的一些经验：个性化强化学习



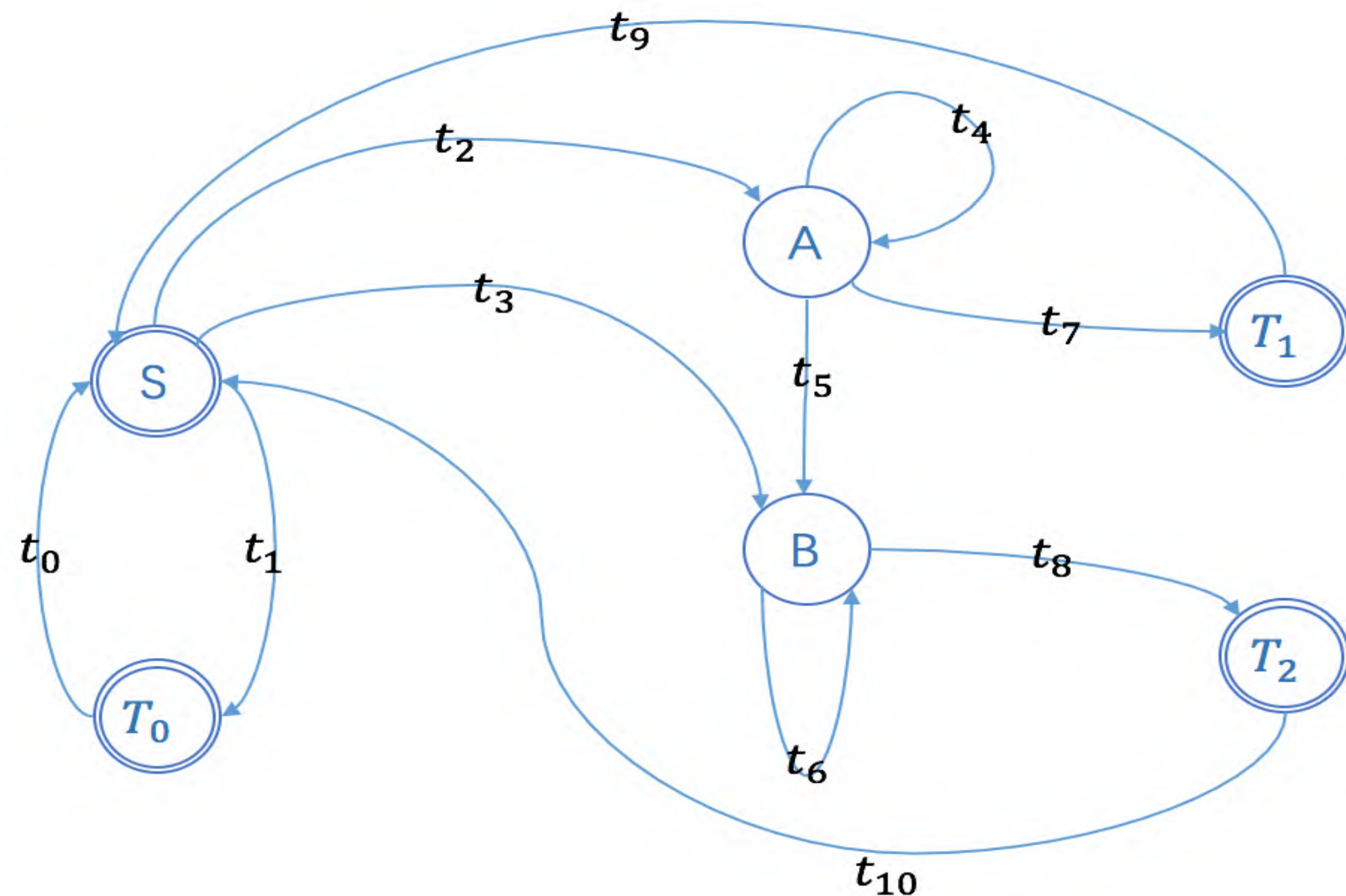
我们的一些经验：个性化强化学习



终止/起始状态



中间状态



状态转移说明:

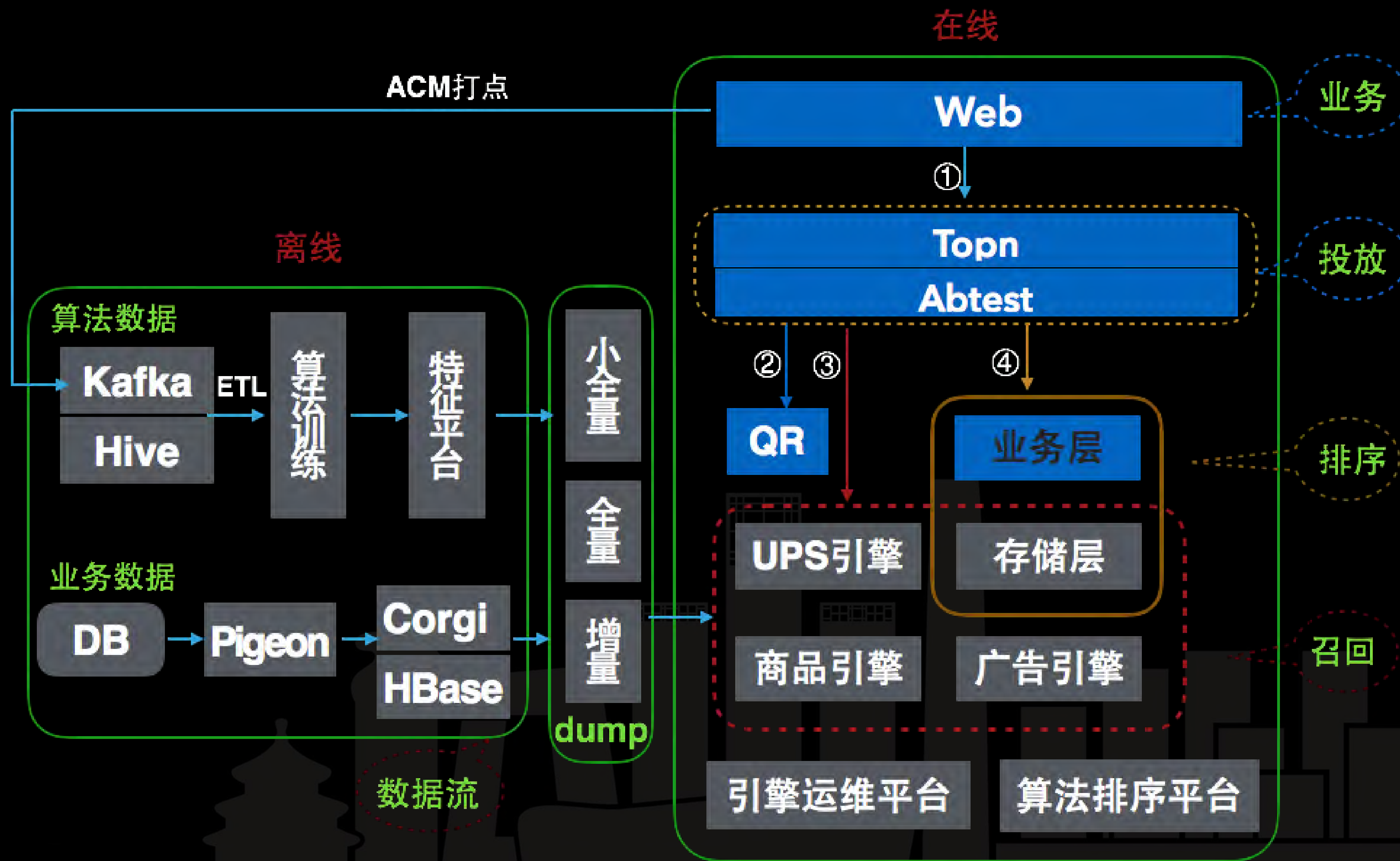
- t_0 商品被放入推广橱窗
- t_1 商品从橱窗被移除/失效
- t_2 橱窗商品被判定为低置信度商品（曝光不足，信息不够），需要先保守扶持并收集信息
- t_3 橱窗商品被判定为高置信度商品，进入放量扶持曝光位进行曝光的竞争
- t_4 商品表现数据持续积累
- t_5 商品数据收集足够且商品表现符合预期，进入放量扶持阶段进行曝光竞争
- t_6 放量扶持阶段，通过橱窗位不断获取流量收益
- t_7 收益表现足够置信，但商品潜力不足，未能进入放量扶持阶段
- t_8 商品表现发现未被低估，停止扶持流量
- t_9 商家尝试再次放入橱窗位，功能上支持，但算法要支持一定时间内的结论一致性
- t_{10} 同 t_9

状态说明:

- S 初始状态:商品被放入橱窗位
- A 新品扶持状态:商品被判断为新品，有保留地扶持并检验该商品是否可能被低估
- B 放量扶持状态: 商品为确认效果被低估同时业务目标需要放量扶持

- T_0 结束状态：移出橱窗位、商品下线等，停止扶持原因清晰。
- T_1 正常结束状态：新品验证阶段不符合预期，需要引导，给出结论报告和后续操作建议
- T_2 正常结束状态：放量阶段流量扶持过多衰减，需要引导并说明原因，并给出后续操作建议

我们的一些经验：个性化强化学习



我们的一些经验：个性化强化学习

- 没有讲到的
 1. 样本抽样和特征
 2. 反作弊的做法
 3. 复杂模型融合





关注QCon微信公众号，
获得更多干货！

Thanks!



主办方 **Geekbang** > **InfoQ**
极客邦科技