# 搭建个性化推荐引擎系统

## 推荐系统概念和应用场景

信息过载问题：

信息时代，接受的信息太多，用户使用信息效率下降。

长尾问题：

长尾理论认为，由于成本和效率的因素，当商品储存流通展示的场地和渠道足够宽广，商品生产成本急剧下降并且销售成本急剧降低时，几乎任何以前看似需求极低的产品，只要有卖，都会有人买。这些需求和销量不高的的产品所占据共同的市场份额，可以和主流产品的市场份额相比，甚至更大。所以需要建立小众商品和关注他们的人群的匹配。

推荐系统时解决上方两个问题的好方法。

推荐系统：

利用电子商务网站向客户提供商品信息和建议，帮助用户决定购买什么产品，默尼销售人员帮助客户完成购物过程。

用户建模模块，推荐对象建模模块，推荐算法模块。

个性化推荐系统：

个性化推荐是根据用户的兴趣特点和购买行为，向用户推荐感兴趣的信息和商品。个性化推荐系统是建立在海量数据挖掘基础上的一种高级商务智能平台，以帮助电子商务网站为其顾客购物提供完全个性化的决策支持和信息服务。

采样才算一个好推荐系统：

满意度：多用下面几个衡量

准确度：

覆盖率：长尾信息的曝光度

多样性：（高度回归问题），权重

惊喜度：之前没有推出 但是客户很满意的指标

健壮度：

实时性：

信任度：

广泛应用领域：

电子商务，影视网站，网络电台，社交网络

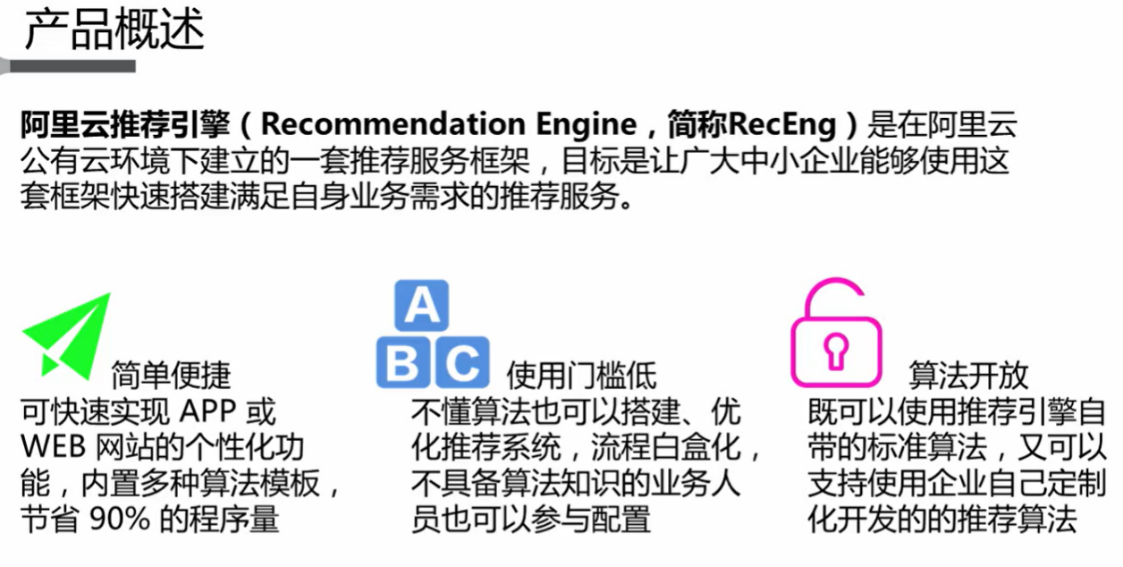
个性化阅读，位置服务，个性化邮件，个性化广告

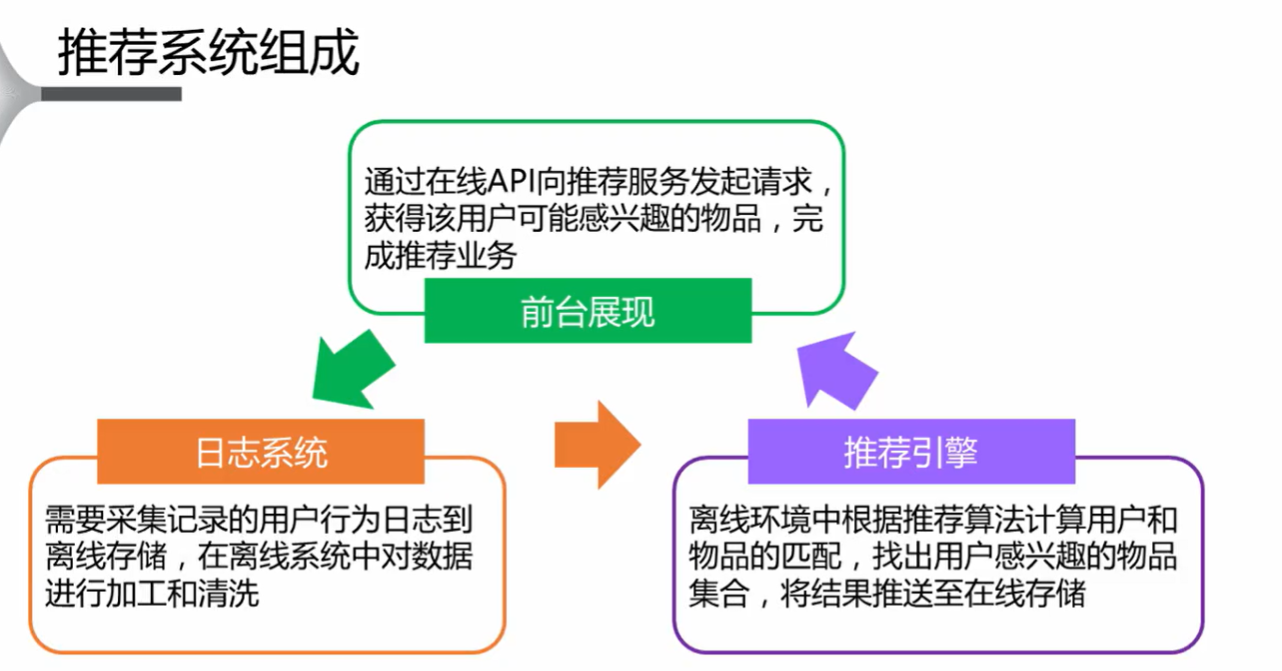
## 推荐引擎的算法原理

引擎的主要推荐算法有如下四类：

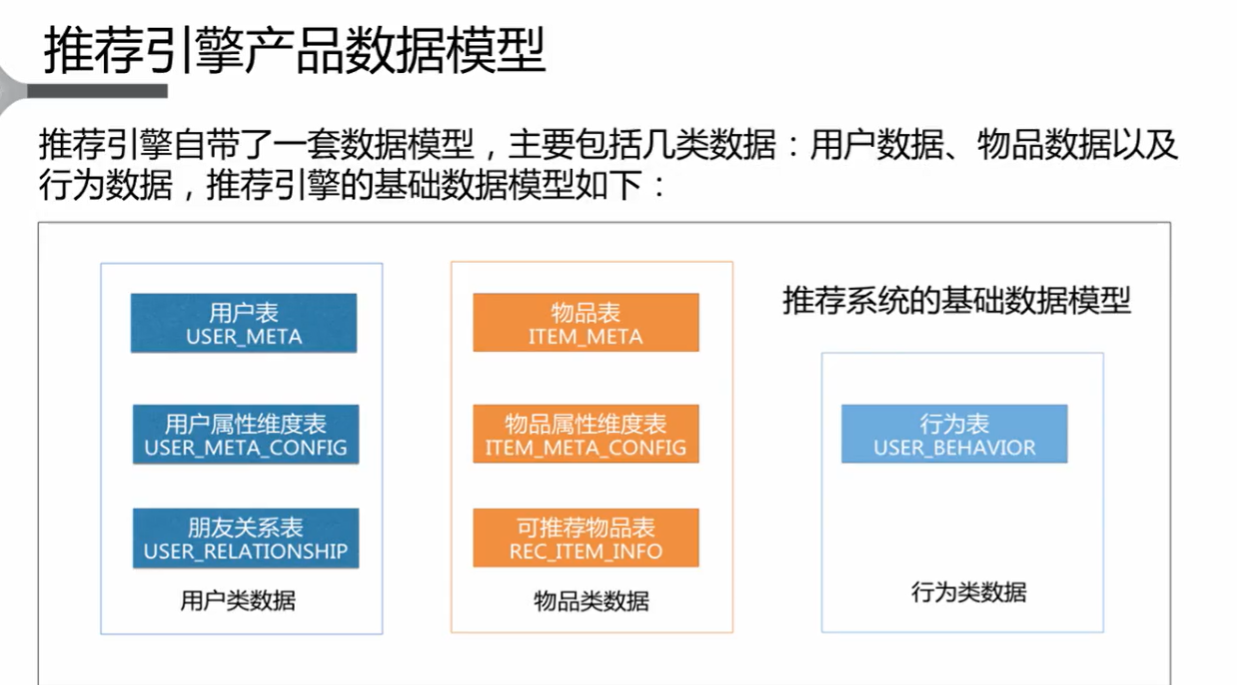
1. 基于人口统计学的推荐
   1. 用户信息建模
   2. 计算用户相似度
   3. 根据相似用户推荐
   4. 优点：不需要当前用户对物品喜好的历史问题；对新用户没有冷启动问题；不依赖物品本身
   5. 缺点：需要收集用户的信息，敏感信息；算法比较粗糙，无法应用于电影，音乐等；用户的信息可能变更。
2. 基于内容的推荐
   1. 物品信息建模
   2. 计算物品相似度
   3. 根据相似物品推荐
   4. 优点：不需要用户的信息，只需要用户本身的物品纪律即可；没有启动问题，能很好地分析用户的口味。
   5. 缺点：需要对物品进行建模，推荐的质量依赖于对物品完整的建模程度。建模仅仅依赖于物品的本身特征，没有考虑人对物品的依赖程度
3. 基于用户的协同过滤
   1. 用户偏好模型
   2. 找到相近的用户群
   3. 根据相似用户推荐
   4. 优点：推荐物品内容上完全不相关，所以可以发现用户的潜在兴趣，能够对用户进行准确的个性化推荐，对用户进行准确的分类
   5. 缺点：新用户有冷启动问题。用户对物品喜好会发生变化，需要及时更新
4. 基于物品的协同过滤
   1. 搜集购物篮数据
   2. 物品相似度模型
   3. 根据物品相似度推荐
   4. 优点：基于用户的基础上改良的一种策略。物品的个数远远小于用户，并且物品间的相似度更加稳定。实时性更好。
   5. 缺点：需要以历史数据为基础，抓取和建模用户的偏好后，很难进行更改，并且根据用户的喜好改变而进行改变，从而导致不够灵活。不是所有场景都合适。

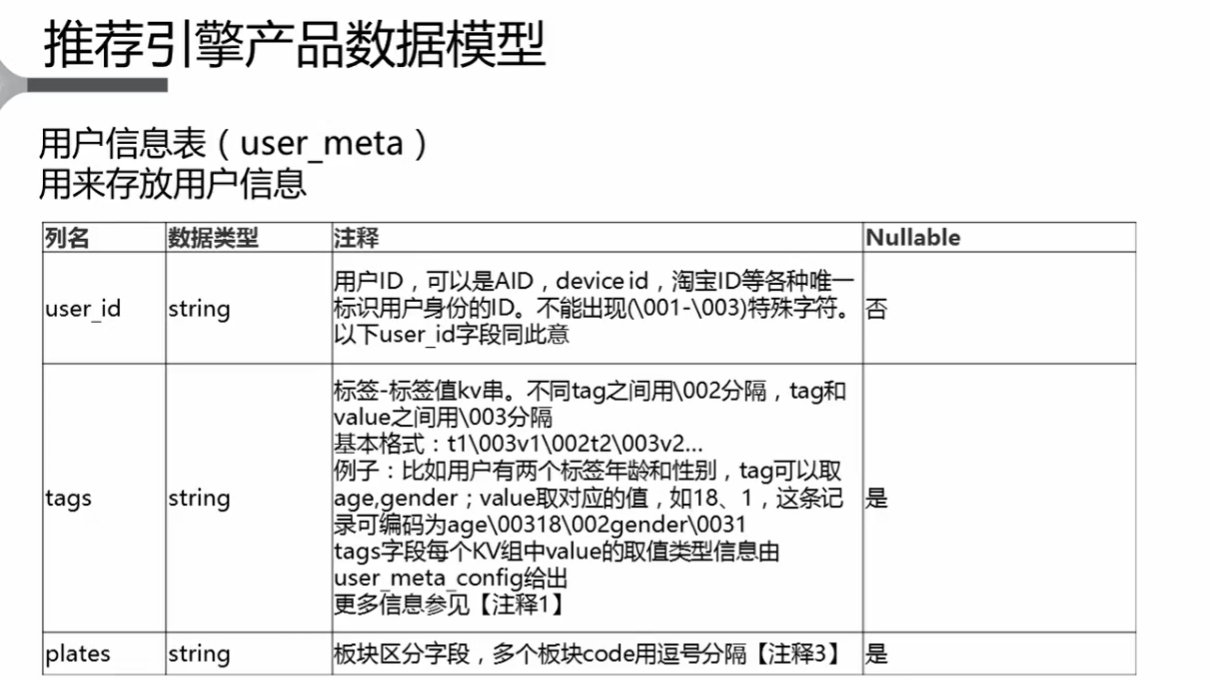
## 推荐引擎产品RecEng介绍

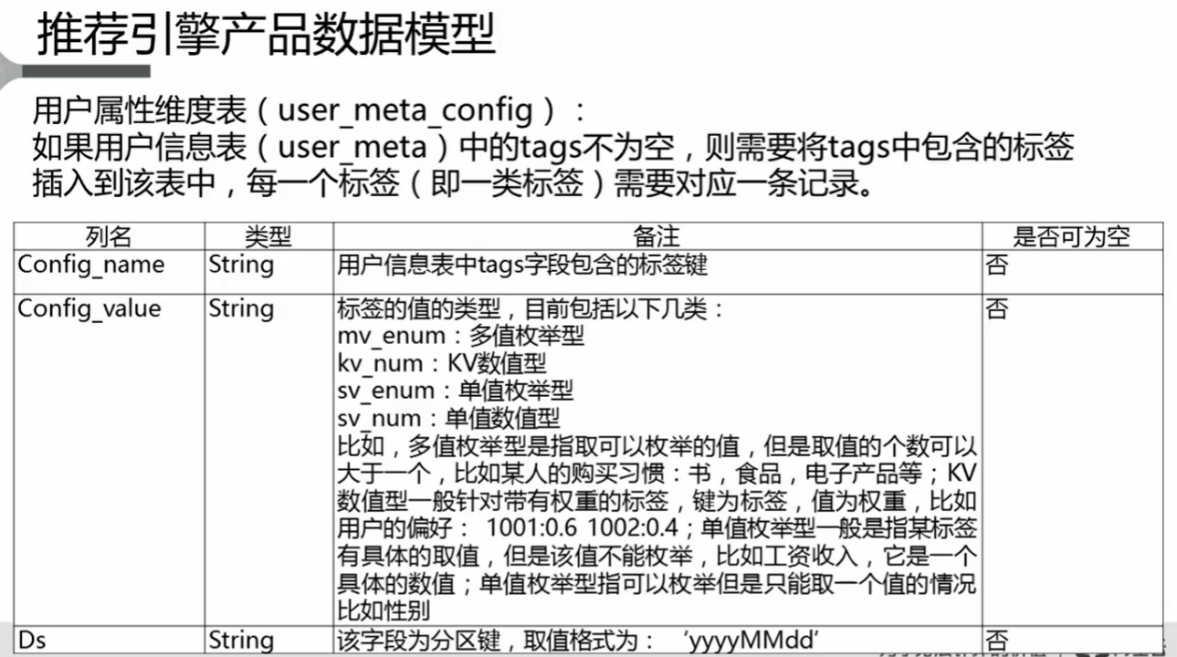




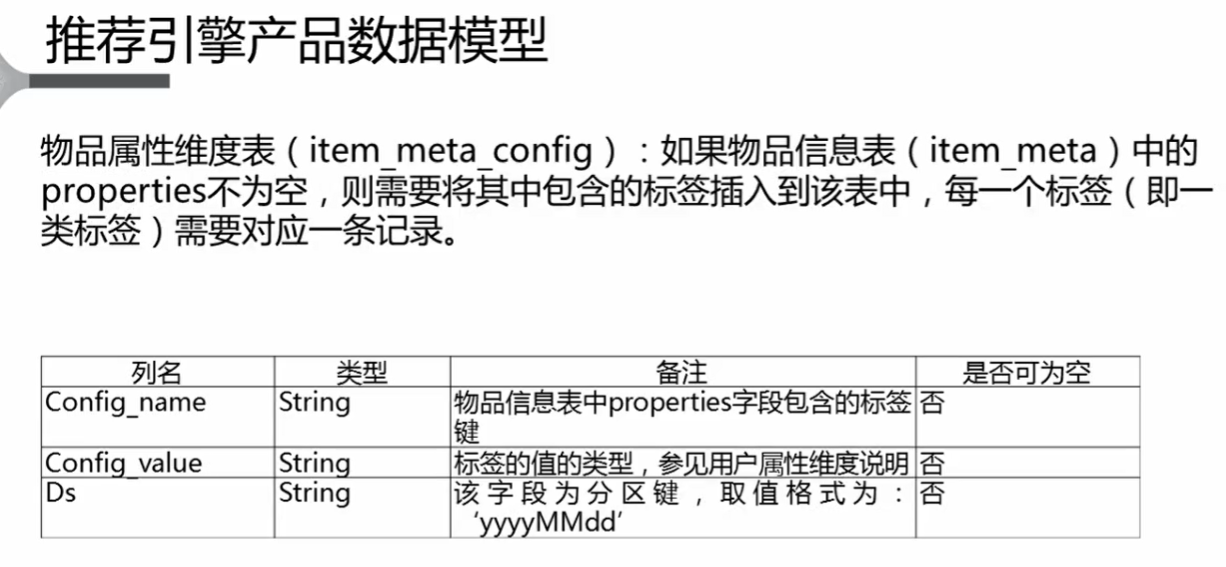




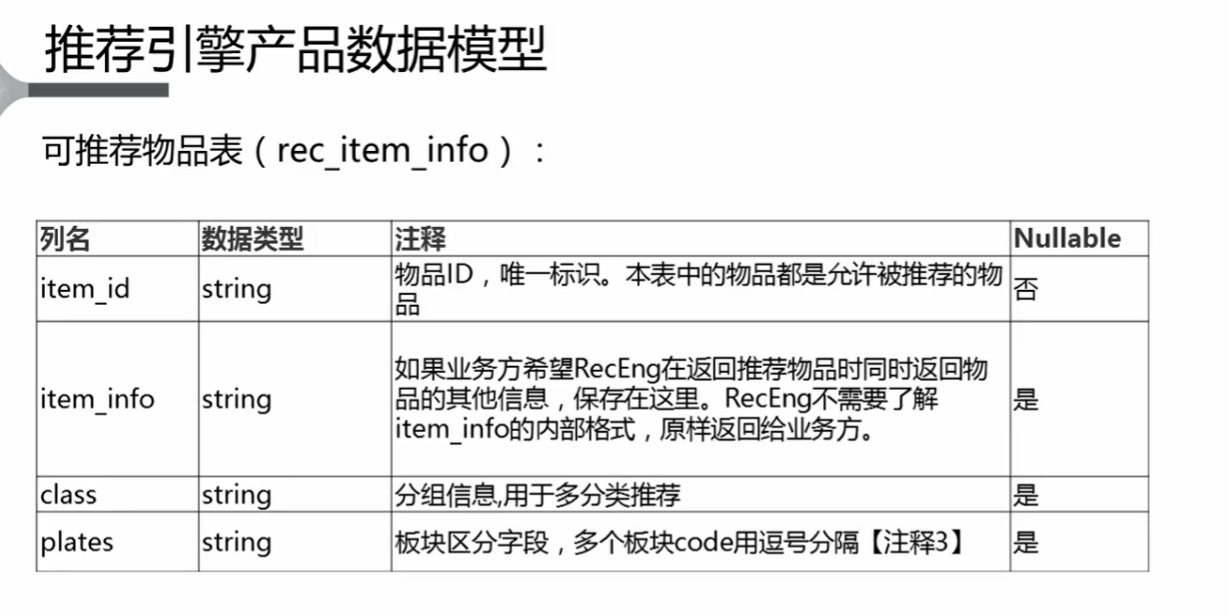












## 实践：搭建电商推荐系统