
15期开学典礼

常见业务

OutLine

搜索引擎概述

广告系统概述

知名推荐系统架构

搜索引擎



北京天气 北京天气



百度一下

- 桥梁——引导用户找到所求
- 满足用户需求的过程
- 连接人与内容、人与服务

——八斗大数据——



北京时间



百度一下

网页

新闻

贴吧

知道

音乐

图片

视频

地图

文库

更多»

百度为您找到相关结果约100,000,000个

搜索工具

[北京时间 - 国家授时中心标准时间](#)

19:51 40

星期五

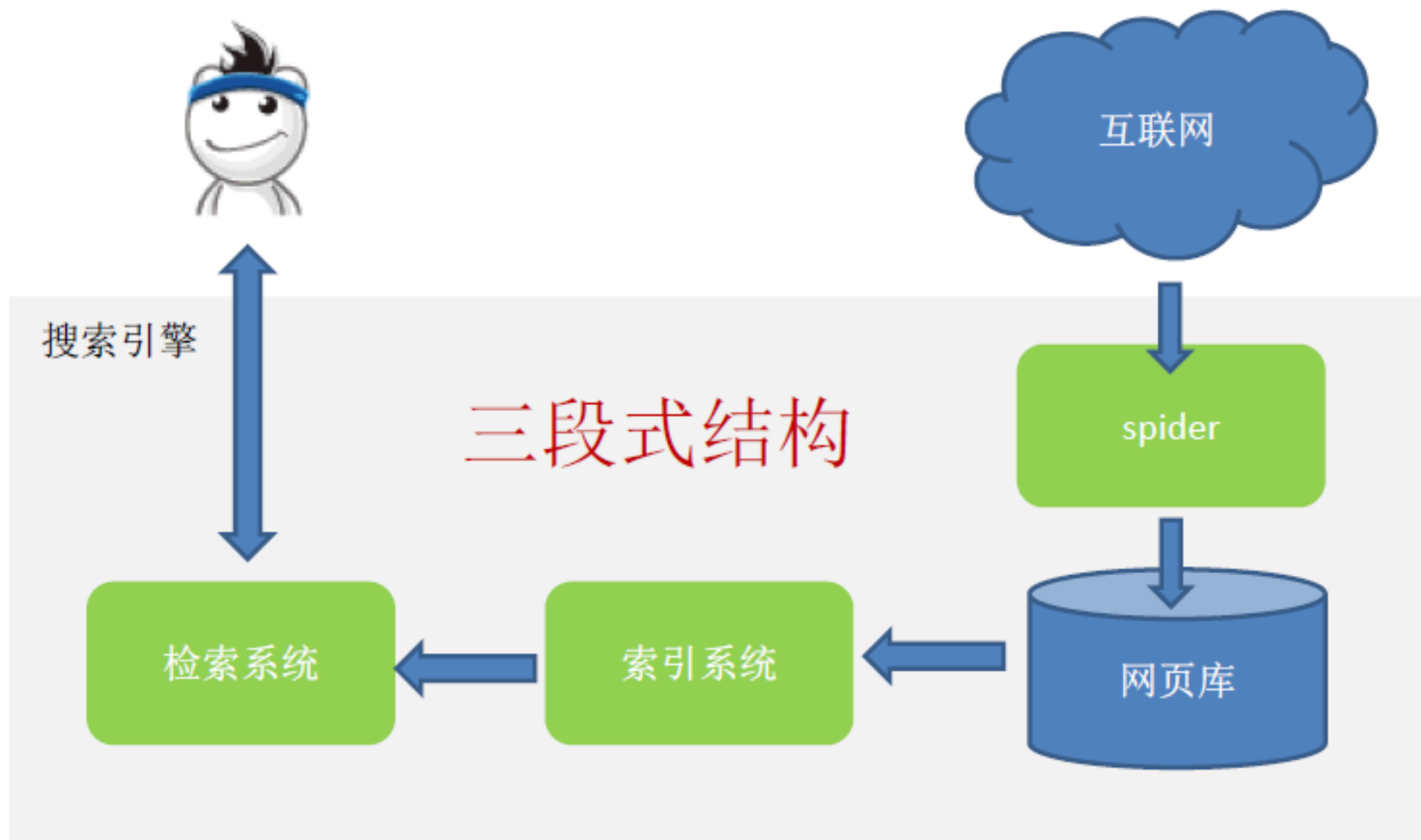
2017年3月17日

[open.baidu.com/](#)[北京时间,北京时间报时,北京时间校准器,现在北京时间几点校对...](#)

本站是国家授时中心同步提供的标准北京时间,可将现在北京时间校准精确到毫秒,您的网速也会影响时间精度,一般误差在半秒之内。本站标准北京时间的实时查询,与国家授...

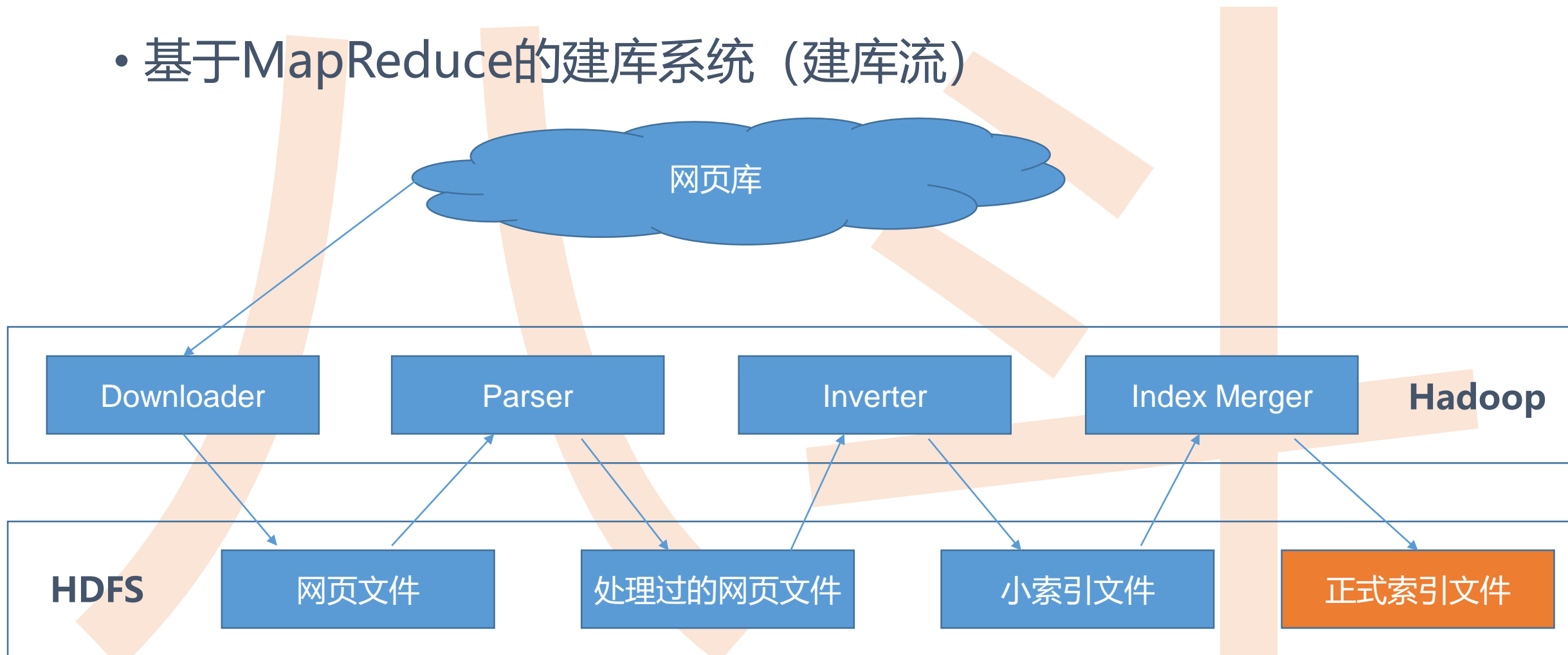
[www.114time.com/](#) - 百度快照 - 评价

搜索引擎



搜索引擎

- 基于MapReduce的建库系统（建库流）



搜索引擎

- 基于MapReduce的建库系统（建库流）

- 目的：建立供检索使用的索引和摘要
- 输入：网页
- 输出：索引和摘要
- 处理：多轮map-reduce
- 页面分析和处理（parser-extractor）
- 页面属性小库输出（splitter）
- 小库正排转倒排（invert-index）
- 小库合并大库（index merge）

OutLine

搜索引擎概述

广告系统概述

知名推荐系统架构

广告系统

• 清单:

	淘宝	百度	Google
自主搜索流量(P4P)	直通车	搜索推广	AdWords
联盟(外部)流量	阿里妈妈	网盟推广	AdSense
品牌广告	钻石展位	百度精准广告	
按成交计费	淘宝客		

广告系统

- 搜索广告点击计费
(CPC)



广告系统

- 展示广告展现计费
(CPM)



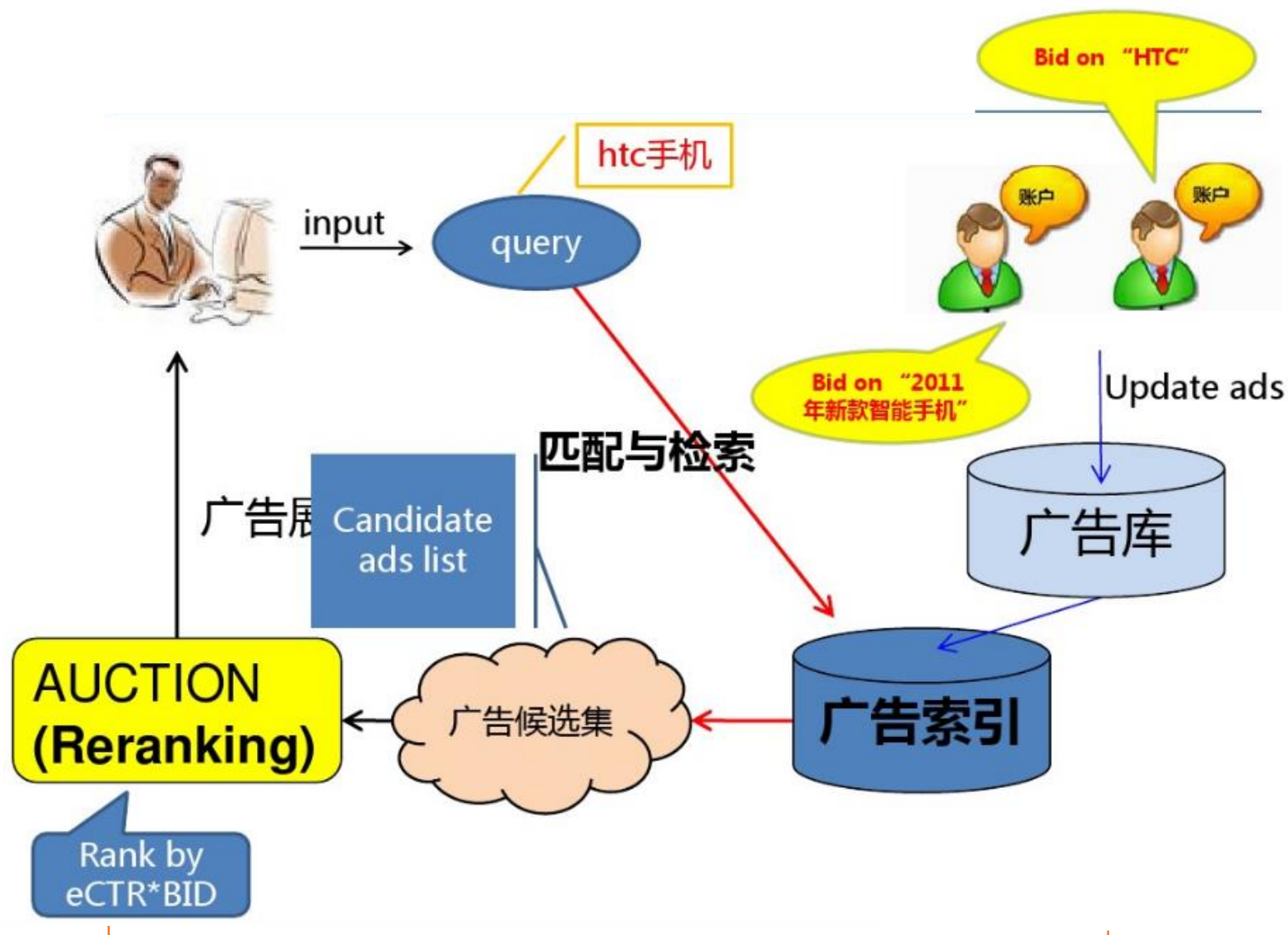
• 淘宝客 (CPS)



搜索广告

- 搜索广告
 - 参与的三方：网民、广告主、搜索平台
- 广告触发：keyword Targeting
 - 广告主和网民通过关键词表达需求
 - 网民输入的query和广告主购买的keyword进行匹配
- CTR预估 (click/show)：机器学习
 - 点击率用于广告的排序
 - 保护网民的利益和提升搜索引擎的收益
- 广告排序：关键词广告拍卖机制
 - 每一次的广告展现都是一次动态的拍卖
 - 排序函数： $ctr * bid$

搜索广告



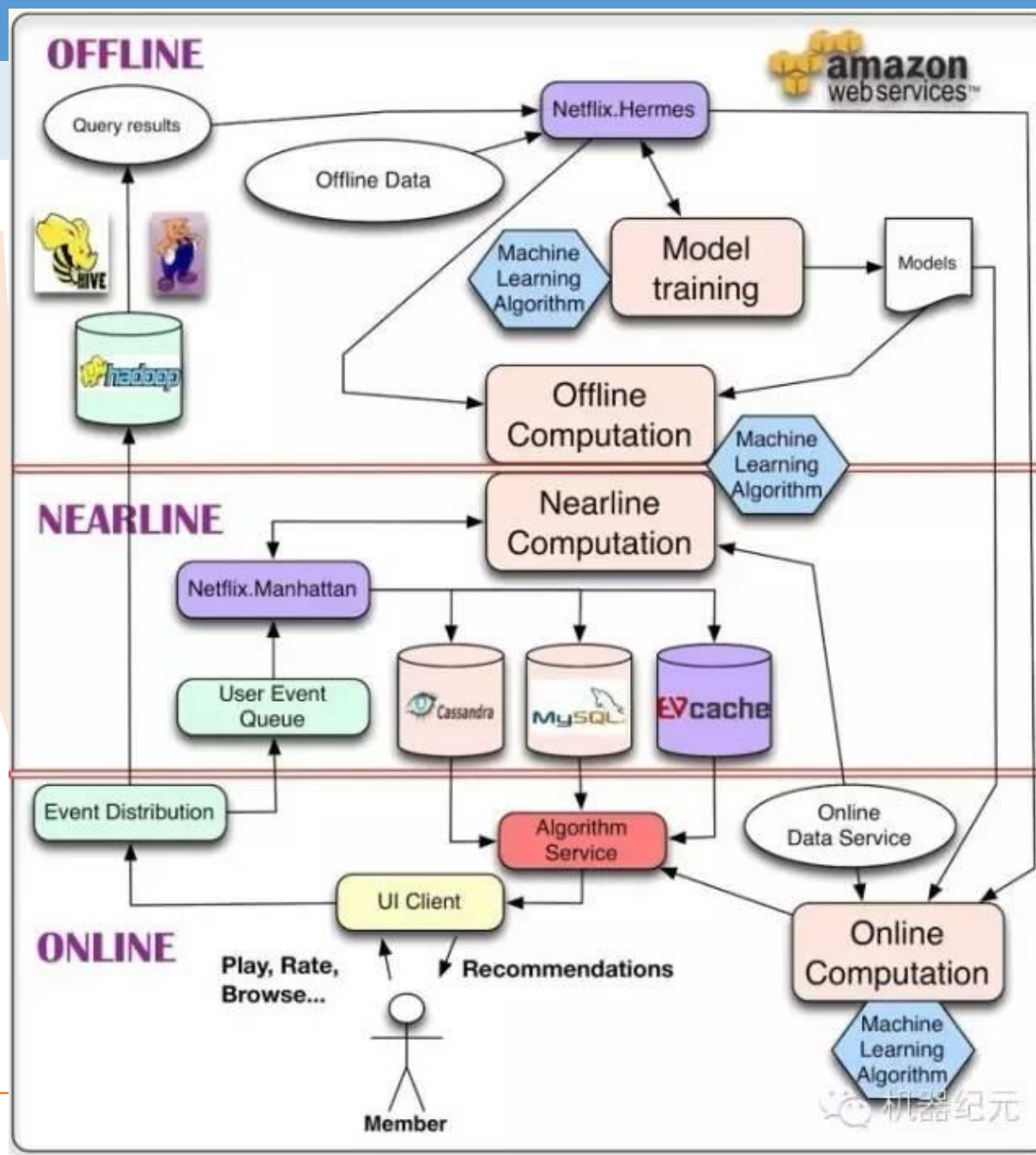
OutLine

搜索引擎概述

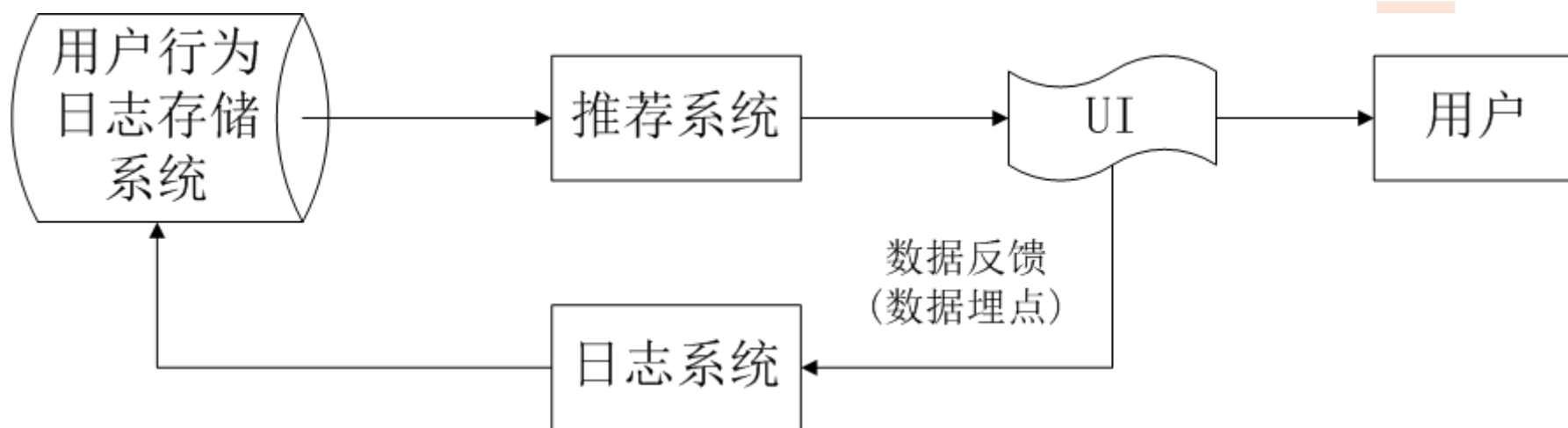
广告系统概述

知名推荐系统架构

NetFlix

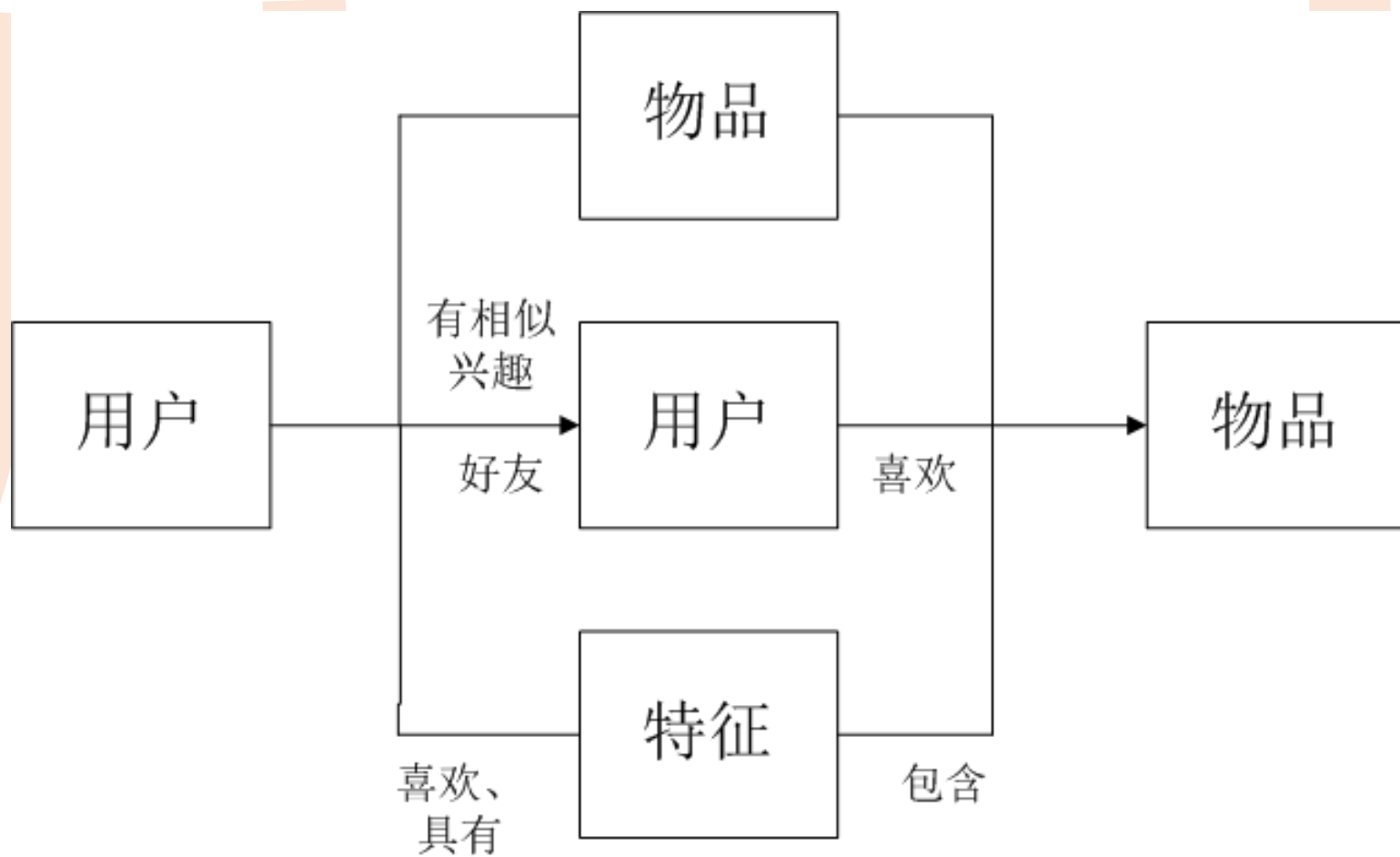


NetFlix

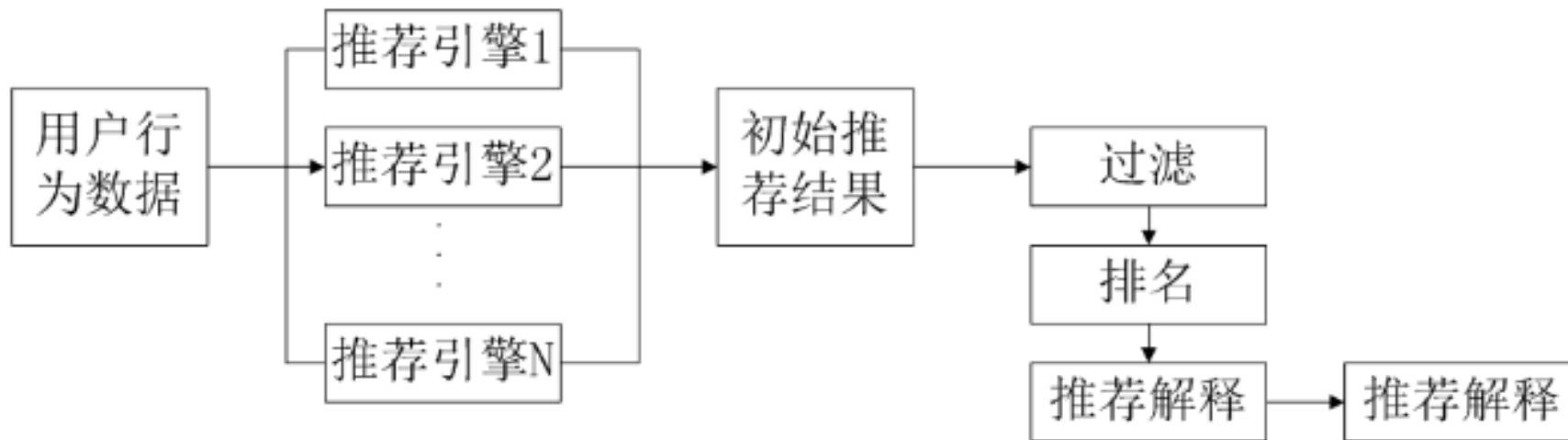


- 界面UI那一块包含3块东西：1) 通过一定方式展示推荐物品(物品标题、缩略图、简介等)；2) 给的推荐理由；3) 数据反馈改进个性化推荐；
- 关于用户数据的存放地方：1)数据库/缓存用来实时取数据；2) hdfs文件上面；

抽象出三种推荐方式

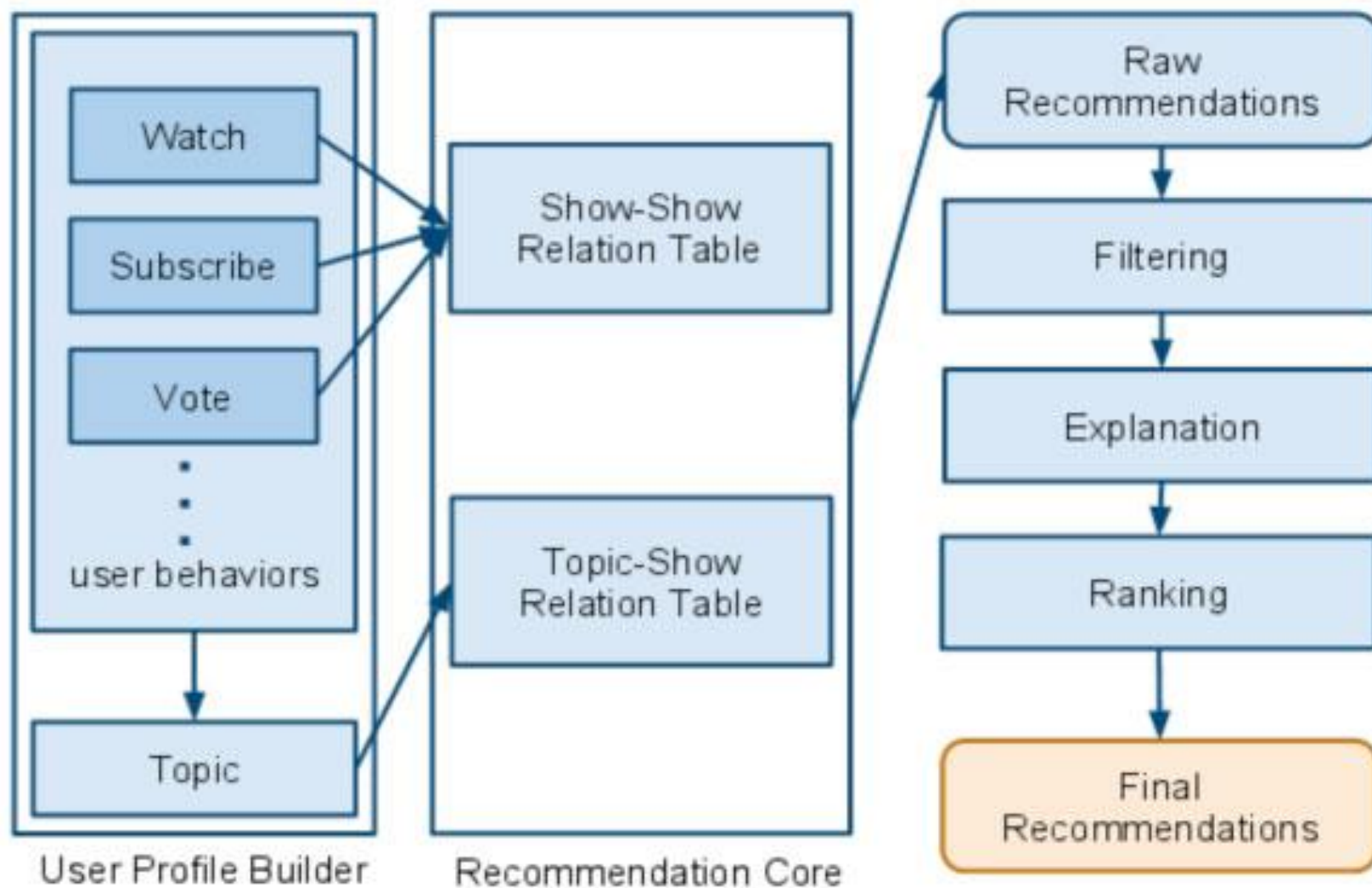


抽象出三种推荐方式

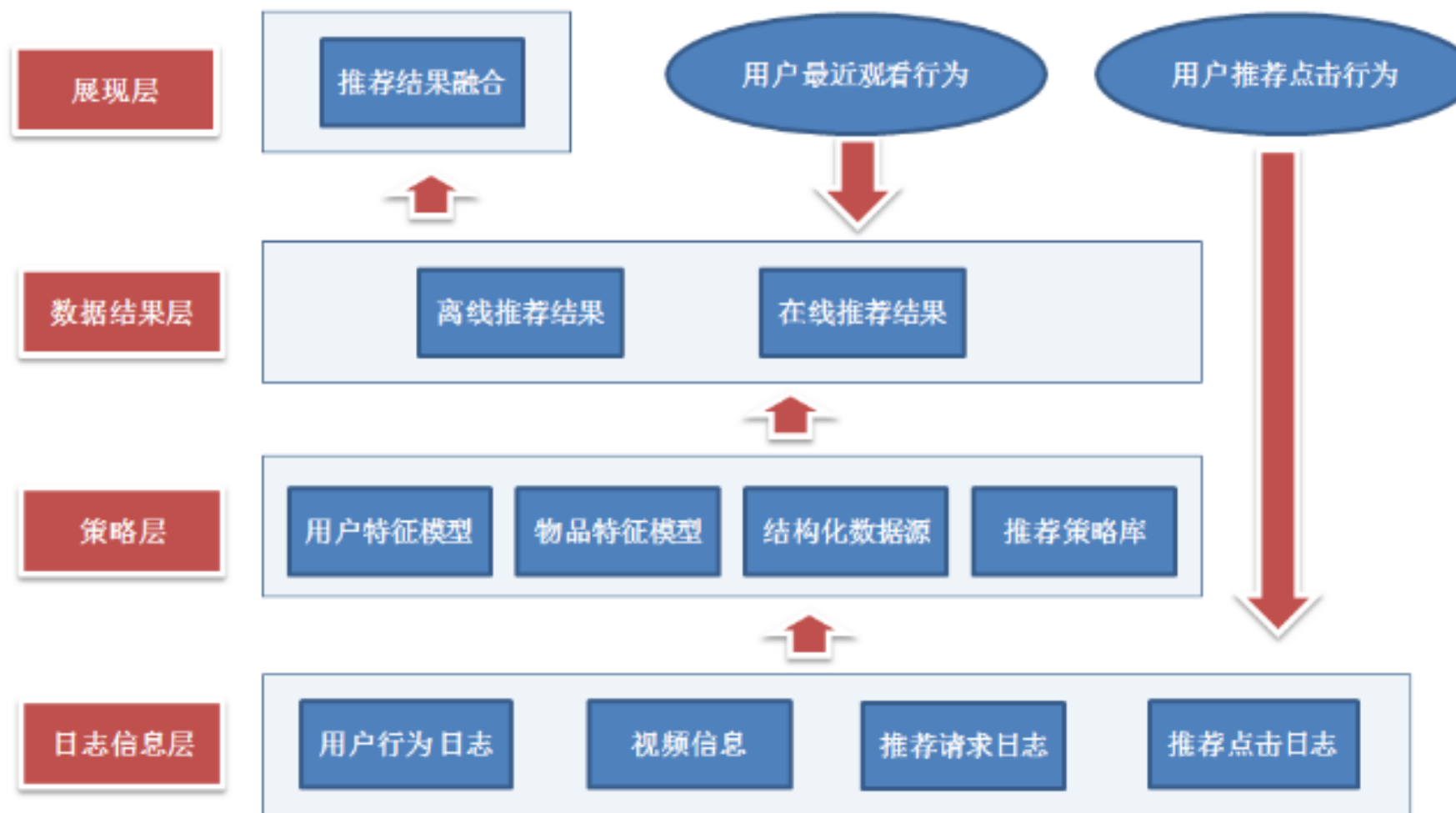


- 推荐引擎的构建来源于不同的数据源(也就是用户的特征有很多种类, 例如统计的、行为的、主题的)+不同的推荐模型算法, 推荐引擎的架构可以试多样化的(实时推荐的+离线推荐的), 然后融合推荐结果 (人工规则+模型结果), 融合方式多样的, 有线性加权的或者切换式的等

H U L U 推荐架构



优酷推荐架构



Q & A

@八斗学院
