

腾讯个性化推荐系统 设计及实现

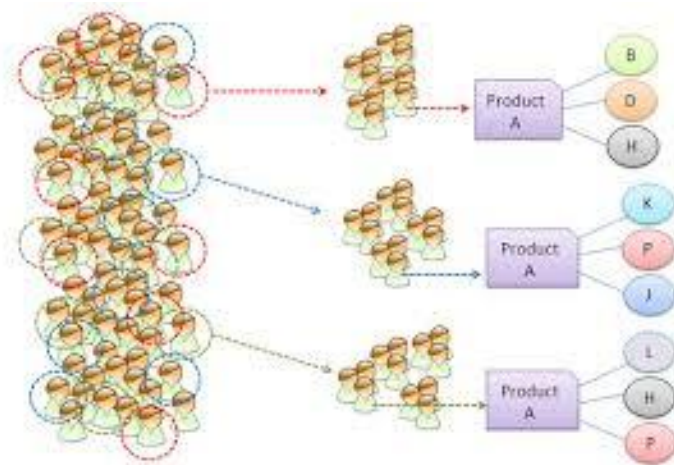
目录

一. 推荐 简介

二. 问题 分析

三. 解决 之道

四. 两大 平台



推荐系统一直在发展

2006-06-01 至 2013-07-24

● 推荐系统



KDDCup 2012 by Tencent

<http://www.kddcup2012.org/>

Login

KDD Cup 2012
from Tencent

[Sign Up](#) | [Home](#) | [Track 1](#) | [Track 2](#) | [Workshop](#) | [FAQ](#)



Tencent
**2012
KDD Cup**
Sponsored by Tencent Inc.

We are pleased to announce the official winners of this year's KDD Cup.

For Track 1:

1st Place, Team ACMClass@SJTU
2nd Place, Steffen Rendle
3rd Place, Team FICO Model Builder

For track 2:



腾讯精准推荐 - 视频推荐

<http://v.qq.com/>

电影 > 院线 > 人再囧途之泰囧

人再囧途之泰囧

互联网高校APP大赛



收藏

分享

下载

用手机看

5720

965

10658.7 万次播放

详情 | 正片列表 | 精华片段 | 明星动态



人再囧途之泰囧

8.9

主演：徐峥 王宝强 黄渤

导演：徐峥 类型：院线 喜剧 地区：内地 年份：2012

看神马

徐峥、王宝强、黄渤“囧神”三人组泰国行，末日狂欢搞笑升级，“与其等死，不如笑死！”

简介：“难兄难弟”此次演绎的《人再囧途之泰囧》是全新人物，全新戏码，影片主要讲述了在泰国发生的闹事……

更多


腾讯精准推荐 - 电商推荐

<http://www.yixun.com/>

潮流服饰 >

运动鞋靴 >

票务旅游, 充值中心 >

公寓! 健康又环保!
库存: 
立即抢购


是好空调! 美的明星
库存: 
立即抢购

评, 今日特价限量
库存: 
立即抢购

¥1599.00 **去看看**

易迅价 ~~¥1999.00~~ 节省 ¥400.00 已售出 113件


旗舰店 FLAGSHIP STORE

猜你喜欢  逛更多  换一批



BUFFALO 巴法络旗舰店	acer 宏碁旗舰店
Microsoft 微软旗舰店	PHILIPS 飞利浦旗舰店
Hisense 海信旗舰店	Kingston 金士顿旗舰店

¥29.00
WD 西部数据 鱼保护包 (S)

浏览过此商品的用户最终购买了

 Samsung 三星 S7572 3G (GSM/WCDMA) 双卡双 ¥898.00	 Samsung 三星 S7568 3G (GSM/TD-SCDMA) 白色 ¥758.00	 三星 S7562 3G手机 (纯白) WCDMA/GSM 双卡双待 (加 ¥1038.00	 Samsung 三星 I8258 3G (GSM/TD-SCDMA) 手机 白 ¥928.00	 全网底价 三星 S7562i 3G (GSM/WCDMA) 双卡双 ¥788.00
-----------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------

购买过此商品的用户还买了

 小米手机 2S 3G (GSM/WCDMA) 手机 ¥1848.00	 KALAIDENG 卡来登 唯系列 Trend Duoss S7562 多功能 ¥55.00	 全网底价 Samsung 三星 Micro SDHC(TF) 16G ¥78.90	 雷尊 ipad mini 智能保护套 黑 色 (前盖+磨砂后壳+高清耐 ¥78.00	 Apple 苹果 iPhone 5 (16G) 3G (GSM/WCDMA) 手机 黑 ¥4738.00
---------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------

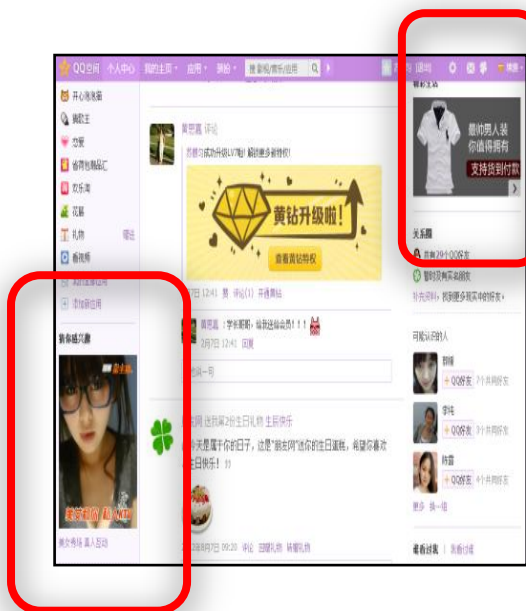
腾讯精准推荐 - 广告推荐

<http://e.qq.com/>

QQ秀



空间个人中心首页



空间应用中心

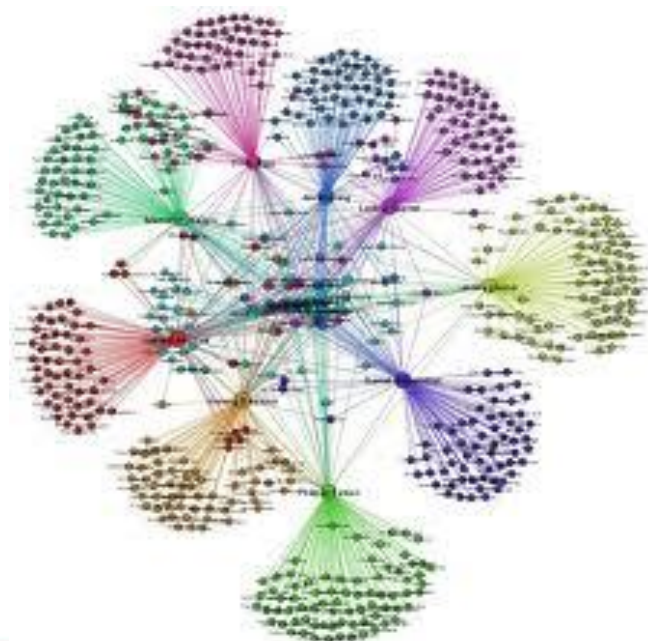


大数据

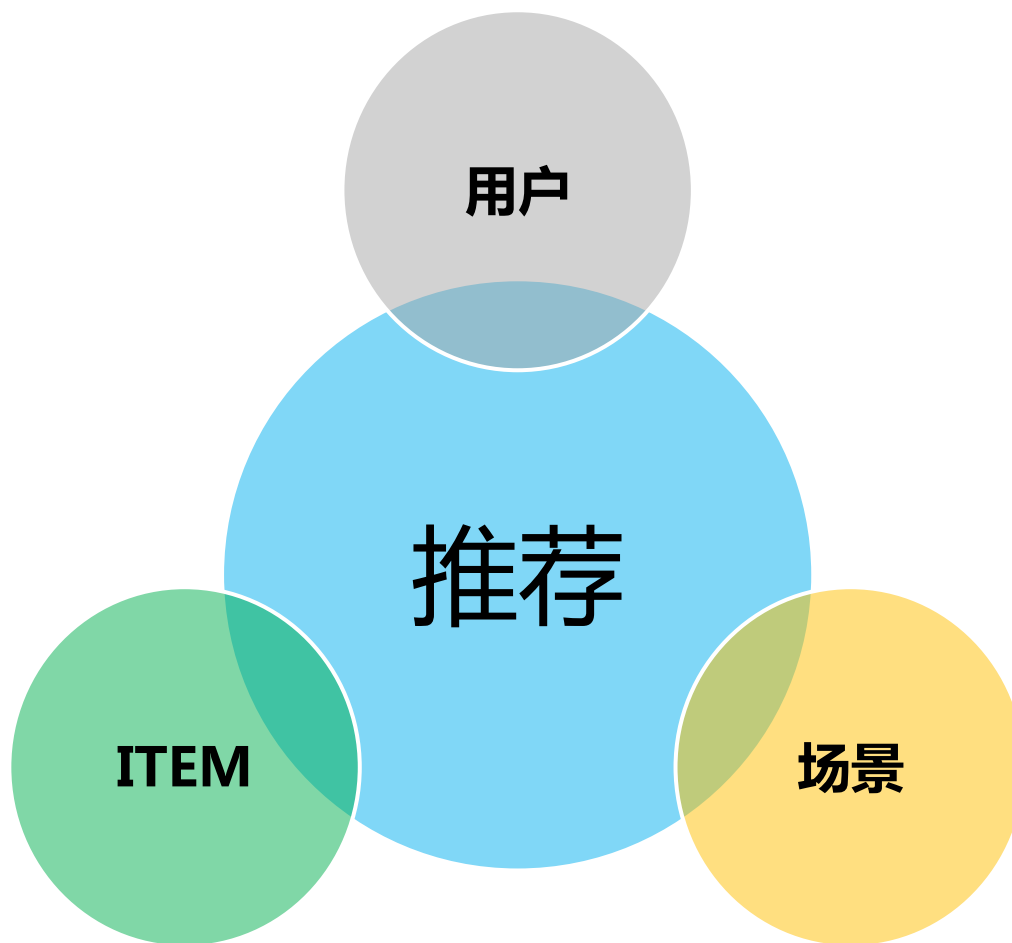
- 800 mil 8亿活跃用户
- 100 bn 1000亿用户关系链
- 100 PB 数据仓库存储100PB+
-

目录

- 一. 推荐 简介
- 二. 问题 分析
- 三. 解决 之道
- 四. 两大 平台



推荐中的 3P



用户

- 点击数据非常稀疏
 - 大部分用户，一个月内都没有点击
 - 平均1000次曝光，点击 $1 \sim 8$
- 曝光数据偏态(以某个广告位为例)
 - 40%的用户，一天内，曝光 = 1
 - 80%的用户，一天内，曝光 < 3

ITEM (广告)

- 素材内容对点击影响很大

0.20%

8X

1.70%



ITEM (广告)

- 新广告的冷启动
 - 每天新增几十万新广告
 - 没有用户反馈信息
 - 自身信息匮乏

场景（广告位）

- 上下文较少
 - 没有明显的意图带入
 - 没有固定的页面内容
- 广告位众多
 - 网页
 - 客户端
 - 手机 等



挑战

- 50 ms 用户体验，响应时延<50毫秒
- 10 bn 100亿推荐请求
- 3-9 广告生命周期仅有3-9天左右
- 99% 99%以上曝光无点击
- 1‰ FACEBOOK平均点击率小于1‰

数据引自 <http://techcrunch.com/2012/06/19/facebook-mobile-ads/>

目录

一. 推荐 简介

二. 问题 分析

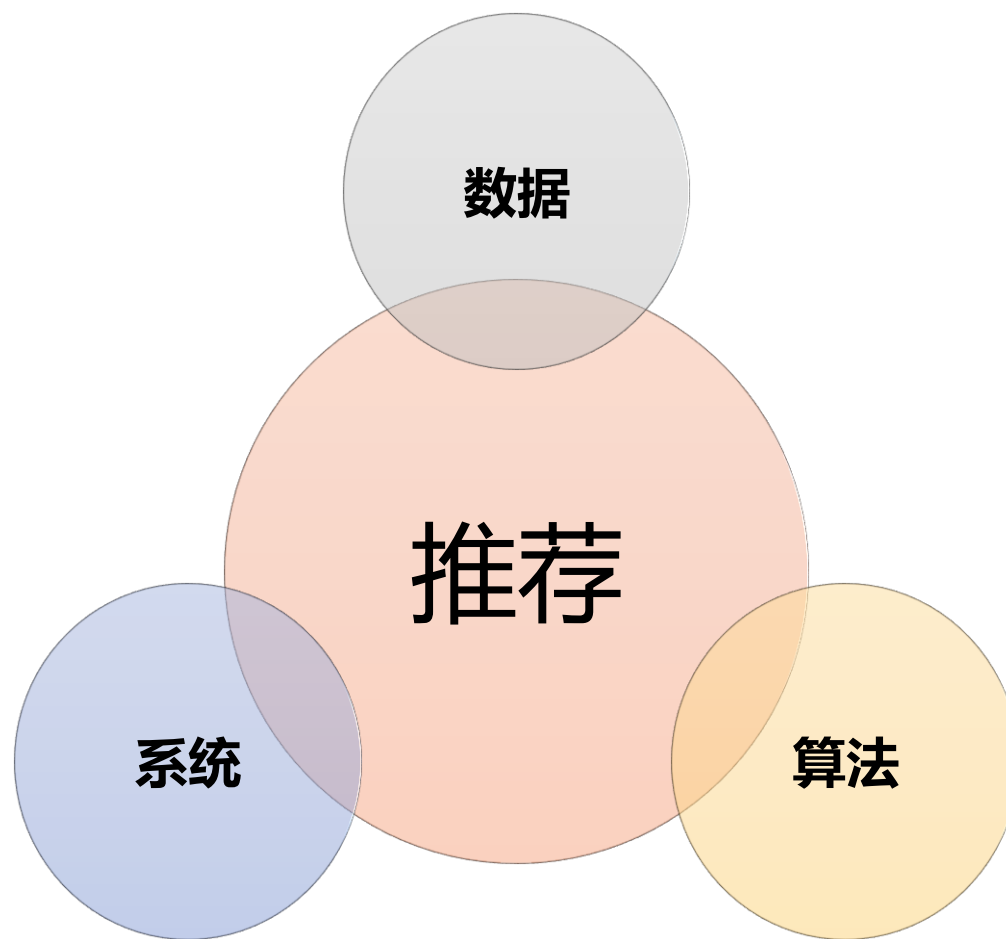
三. 解决 之道

四. 两大 平台

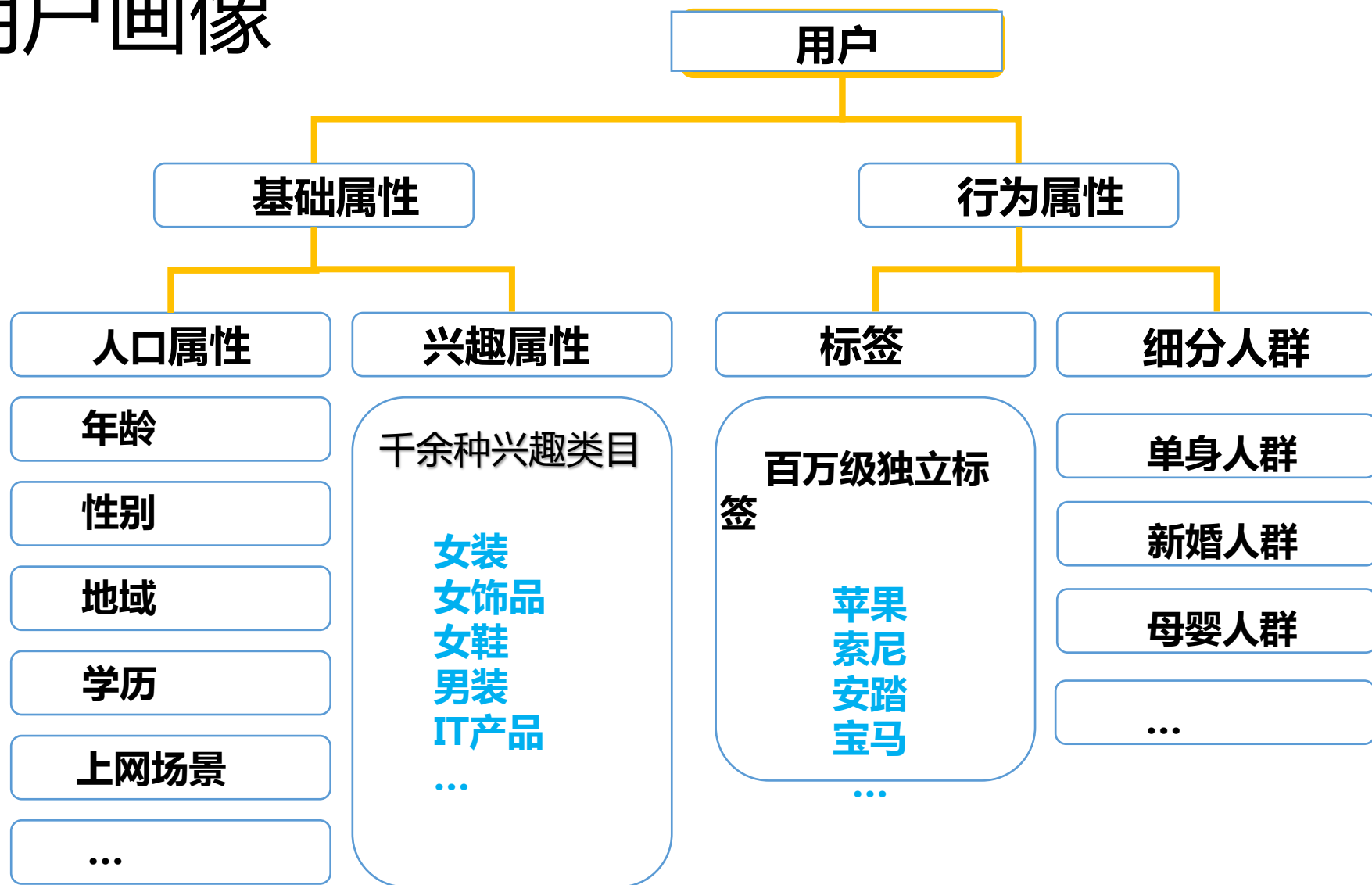
Recommender
Systems



解决之道 3S



用户画像



ITEM的图像特征

- 图像特征

- 亮度
- 饱和度
- 色彩度
- 对比度
- 尖锐度等

- 图像相似度

- Item_id → img_id

Example – 广告图片相似度



聚类id:

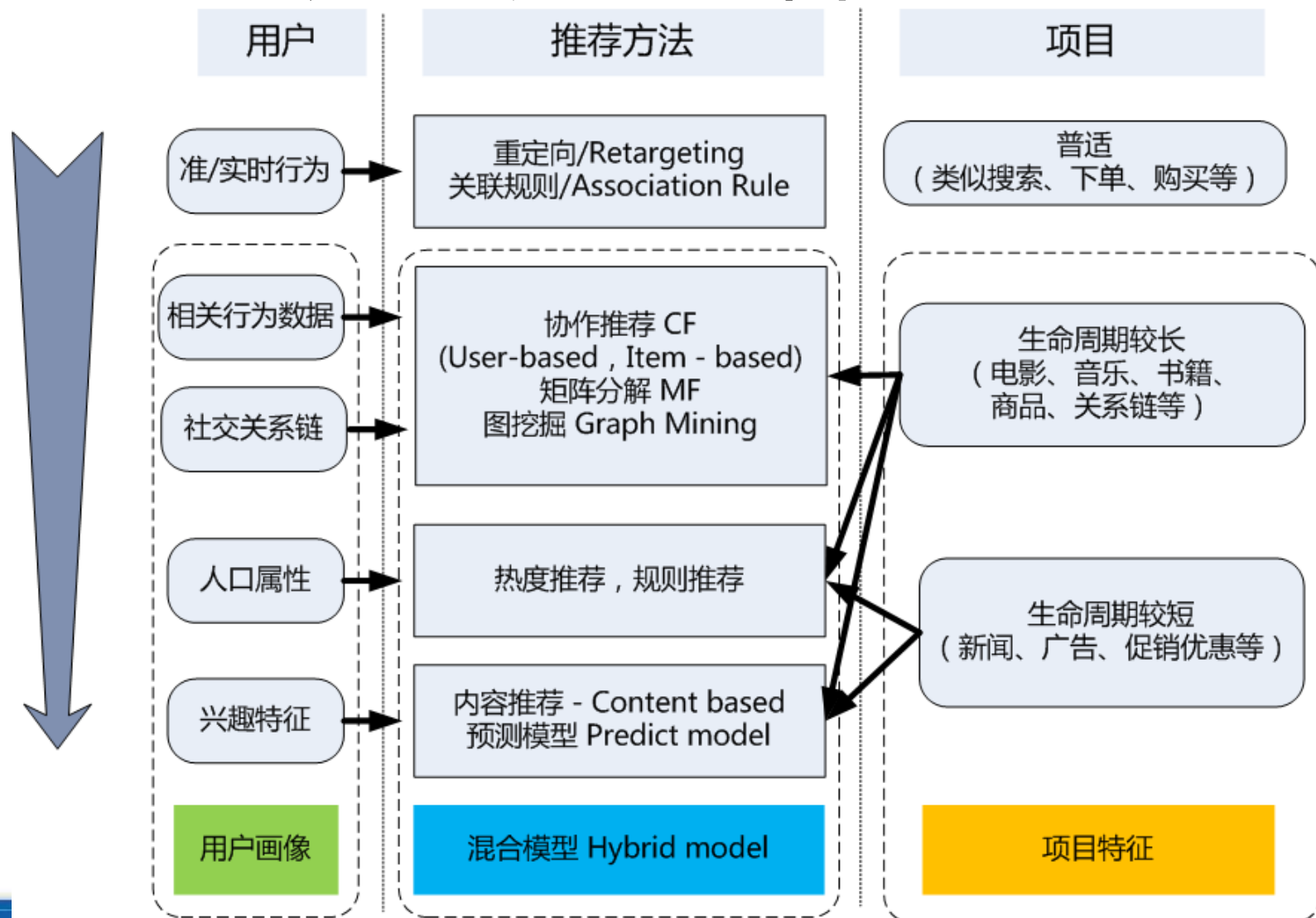
1711127890547892055

包含3106 个这样的类似广告素材

对于同一人群，相似图片CTR接近

广告ID	click	impression	CTR
8509989	88	55,560	0.16%
8509980	116	62,091	0.19%
8509912	166	93,099	0.18%

算法 - 运用之妙 知行合一



Example - 泰囧的推荐结果

用户行为



人在囧途
王宝强爆笑喜剧



《教父2》
三巨头大闹泰国



十二生肖
成龙寻宝之旅



车在囧途
人在囧途兄弟篇



低俗喜剧
彭浩翔回归重口味



西游·降魔篇
新版《大话西游》

用户分群

男性 30-35



偷神家族
元华/胡慧中



北京遇上西雅图
跨国爱情奇遇



青春期2: 青春失乐园
问题少女疯狂爱



万有引力
全明星阵容诠释爱情



巨蟒惊魂
泰国版狂蟒之灾



毒战
人心比毒更毒

用户分群

女性 25-29



人在囧途
王宝强爆笑喜剧



《教父2》
三巨头大闹泰国



车在囧途
人在囧途兄弟篇



十二生肖
成龙寻宝之旅



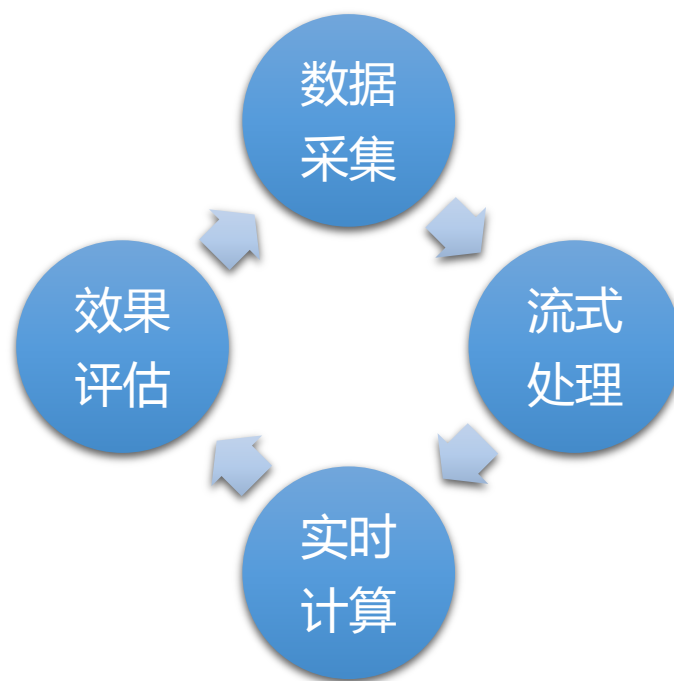
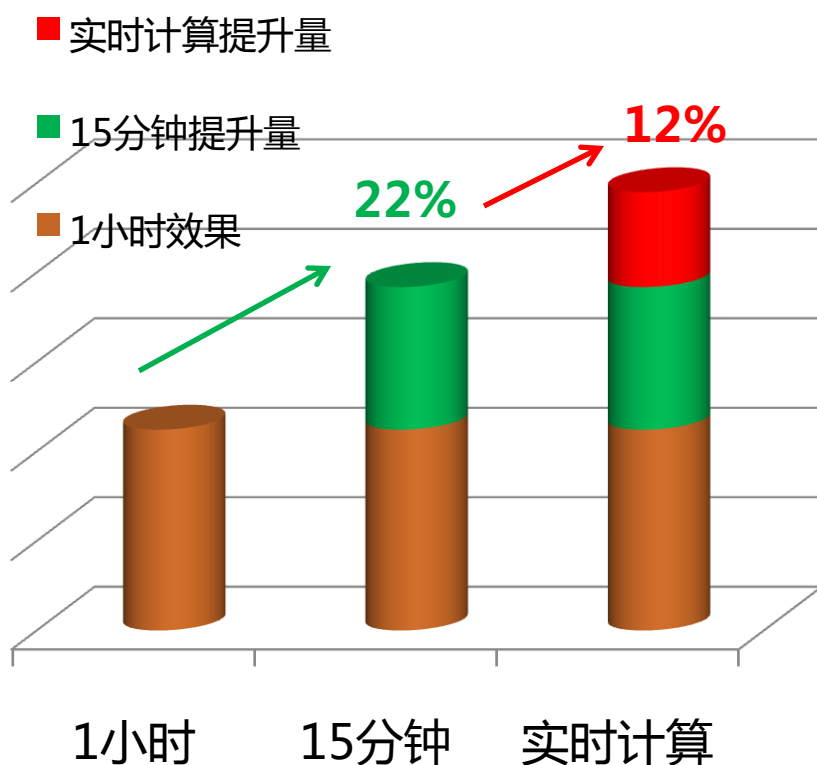
西游·降魔篇
新版《大话西游》



低俗喜剧
彭浩翔回归重口味

系统 - 流式计算，实时计算

CTR LIFT



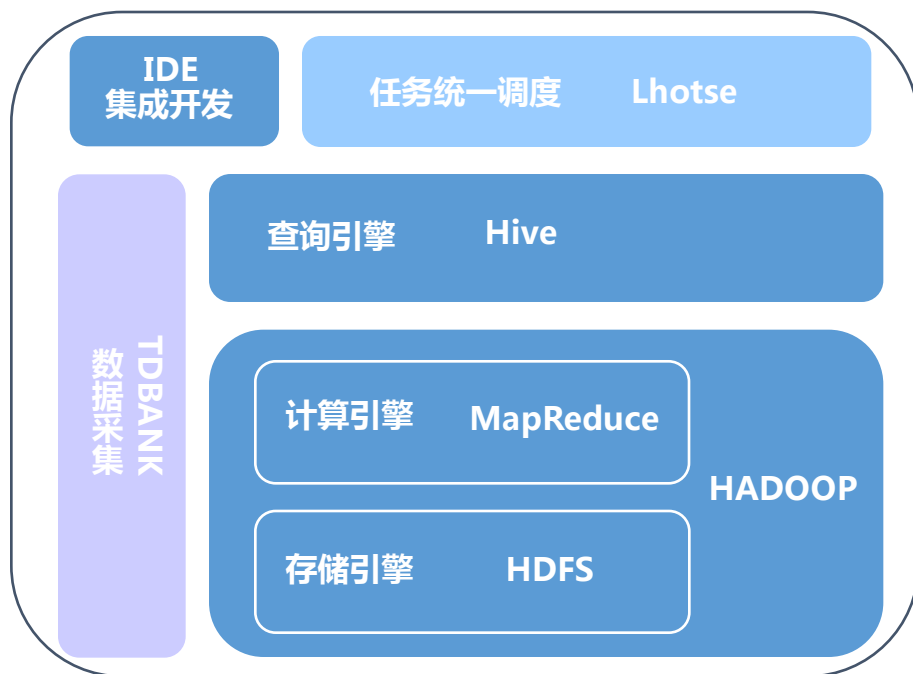
目录

- 一. 推荐 简介
- 二. 问题 分析
- 三. 解决 之道
- 四. 两大 平台



核心平台之一：腾讯分布式数据仓库（TDW）

- ✓ 基于开源hadoop和hive进行大量优化和改造
- ✓ 单集群4400台（业界顶级规模），存储容量100PB

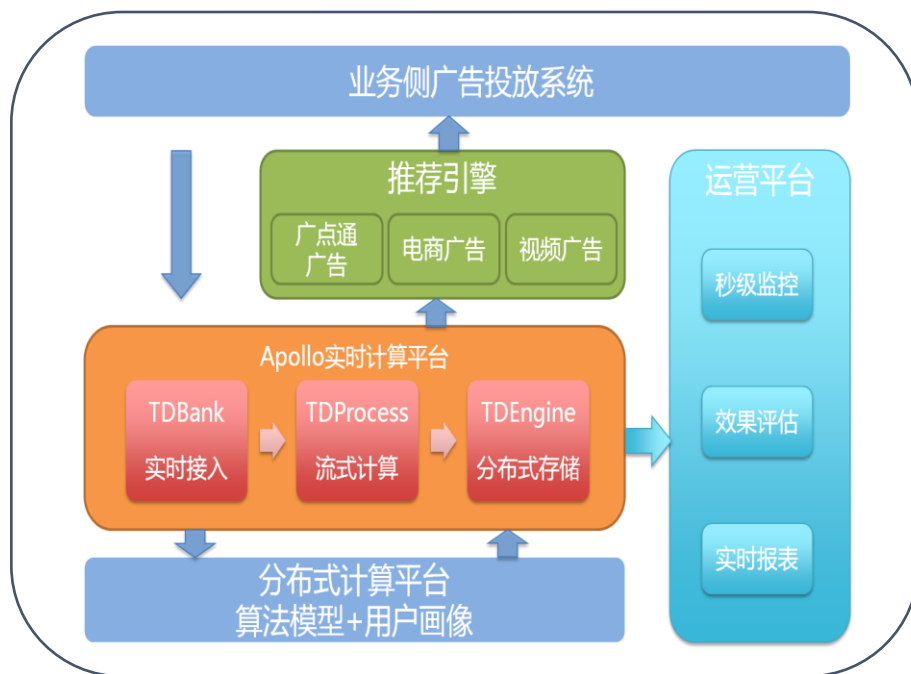


关键技术

- ✓ Hadoop Master (NN/JT) 节点实现并行扩展，支持灾难时自动热切
- ✓ Hive&Pig功能丰富，支持传统数据库的标准语法，提供可视化集成开发环境
- ✓ Lhotse一站式任务管理，每天支撑10万级任务调度，可平行扩展

核心平台之二：实时推荐平台（APOLLO）

- ✓泛平台支撑，满足腾讯各类个性化推荐需求
- ✓海量数据在线处理，日推荐请求300亿，实时计算30000亿
- ✓算法精准，平台高效，毫秒级响应



关键技术

- ✓多种算法模型灵活适配，LR、RDT、SVD等
- ✓简单高效的扩展能力，数据动态伸缩，上层无感知
- ✓多副本数据容灾，服务可用度99.995%以上
- ✓多级缓存技术，有效解决分布式计算的数据CoLocation问题

Reference

- <http://recsys.acm.org/>
 - [RecSys 2013](#) (Hong Kong)
 - [RecSys 2012](#) (Dublin)
 - [RecSys 2011](#) (Chicago)
 - [RecSys 2010](#) (Barcelona)
 - [RecSys 2009](#) (New York)
 - [RecSys 2008](#) (Lausanne)
 - [RecSys 2007](#) (Minnesota)
- http://en.wikipedia.org/wiki/Recommender_system

A stylized, light blue globe is positioned on the left side of the slide, showing the outlines of continents and a grid of latitude and longitude lines. The rest of the slide has a solid dark blue background.

谢谢！