



全球软件案例研究  
峰会

# 100

TOP 100 CASE STUDIES OF THE YEAR

全球软件案例研究峰会



TOP 100 CASE STUDIES  
OF THE YEAR

[www.top100summit.com](http://www.top100summit.com)



全球软件案例研究  
峰会

## 58同城推荐系统设计与实现

# 摘要

- 业务场景简介
- 推荐系统简介
- 常见推荐算法
- 推荐系统难点
- 推荐系统实现细节



TOP 100 CASE STUDIES  
OF THE YEAR

[www.top100summit.com](http://www.top100summit.com)

## 58同城推荐系统设计与实现

# 案例

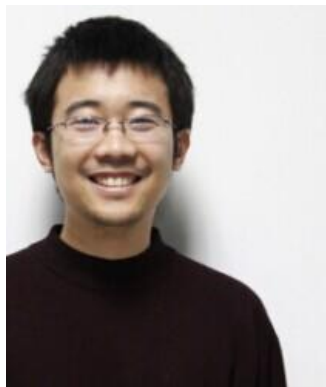
- 架构上：58同城推荐系统设计与实现细节
- 业务上：如何在招聘业务线实现简历与职位的推荐





全球软件案例研究  
峰会

# 沈剑



沈剑，分布式领域技术专家。

曾任百度高级工程师，参与过多个百度HI重大项目的研发。

现任**58同城**高级系统架构师，技术委员会主席。负责过58同城即时通讯，支付系统重构，摊销系统重构，数据库中间件，推荐系统等多个系统与项目的设计与实现。

**58同城**优秀讲师，曾多次代表58同城作为嘉宾参与velocity，系统架构师大会，lamper人社区，top100summit等技术会议，分享58同城的架构技术。



TOP 100 CASE STUDIES  
OF THE YEAR

[www.top100summit.com](http://www.top100summit.com)



全球软件案例研究  
峰会

推荐在58招聘

# 业务场景

- 简历与职位的推荐

您可能感兴趣的职位					<a href="#">查看更多</a>
<a href="#">月薪过万诚聘电话销售</a>	北京富立旺达商贸有限公司	5000-8000元/月	北京	2014-08-22	
推荐理由: <span>五险一金</span> <span>年底双薪</span> <span>加班补助</span> <span>全勤奖</span> <span>带薪年假</span>					
<a href="#">百度诚聘电话销售精英</a>	百度时代网络技术(北京)有限公..	5000-8000元/月	北京	2014-08-22	
推荐理由: <span>五险一金</span> <span>周末双休</span> <span>饭补</span> <span>商业保险</span>					
<a href="#">诚聘电话销售</a>	北京华夏盛典文化艺术有限公司	8000-12000元/月	北京	2014-08-22	
推荐理由: <span>五险一金</span> <span>包吃</span> <span>包住</span> <span>周末双休</span> <span>年底双薪</span>					



TOP 100 CASE STUDIES  
OF THE YEAR

www.top100summit.com

## 推荐在豆瓣

# 业务场景

- 豆瓣电影的推荐

驯龙高手2的预告片和图片 ····· (预告片25 | 图片416 | 添加图片)



喜欢这部电影的人也喜欢 ·····



驯龙高手



里约大冒险2



守护者联盟



精灵旅社



驯龙记：伯克岛的龙骑手 第一季



全球软件案例研究  
峰会

## 推荐在百度

# 业务场景

- 百度知道的推荐

等待您来回答

[更多提问 >](#)

我关注的关键词	我关注的分类	为我推荐的问题
dota 忍者村大战地图 忍者证明		0回答
dota传奇死亡脉冲区是什么版本		0回答
dota之最强血脉txt下载奇书网		0回答
有一首特劲爆的英文歌 里面有一句 "下拉下拉下拉里瑞拉...		2回答
谁有暗黑3 pc端的修改器? 别的不用, 只要增加资源就可...		1回答
谁有在纳米比亚玩过dota?求个帐号, 英文版, 进入冰封...		0回答
30 暗黑战神霸体效果有什么作用		0回答



TOP 100 CASE STUDIES  
OF THE YEAR

[www.top100summit.com](http://www.top100summit.com)

## 什么是推荐？

# 推荐简介

- 用户在在某个场景下对某个商品或信息产生了某种行为，系统会对另一些商品或信息进行推荐
- 要素：
  - (1) 用户 - user
  - (2) 场景 - scene
  - (3) 商品或信息 - item
  - (4) 行为 - action
  - (5) 系统 - recommendation-system
  - (6) 推荐结果集合 - recommendation-result / item-set





## 什么是推荐-举例

# 推荐简介

- 用户在58同城发布了一份简历

- 要素：

user-info: uid

scene-info: entry、local、cateid

item-info: jid

action: post

**system: rec-system**

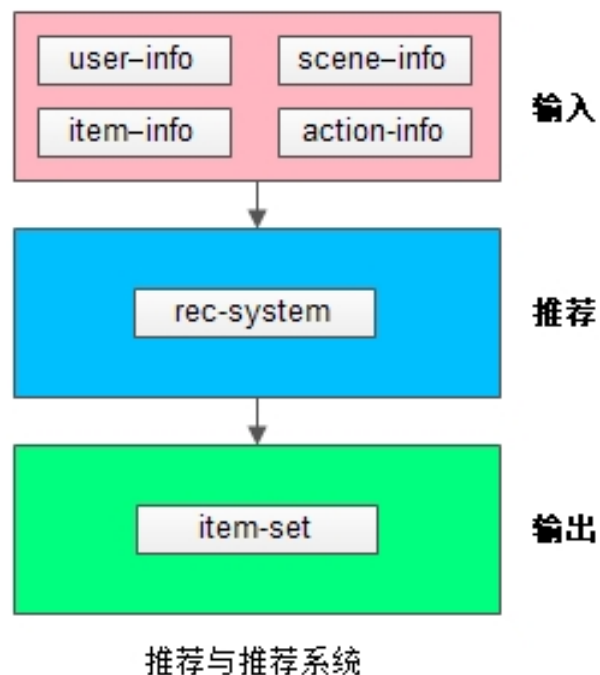
item-set: set<zid>



## 推荐系统简介

# 推荐简介

- 用户在58同城发布了一份简历



## 常见算法

- 协同过滤：collaborative filtering Recommendation
- 原理：用户的相似喜好进行推荐
- 举例：商家下载简历的推荐

	jid1	jid2	jid3	jid4	jid5	jid6	...	jid1000w
uid1	yes	yes	yes				...	
uid2	yes	yes	yes	yes			...	
uid3	yes	yes	yes		yes	yes	...	
uid4				yes		yes	...	
...								
uid100w								

## 常见算法

### 历史行为

jid1, download  
jid2, download  
zid1, post

### 详情

jid1 (司机, 北京, 8000月薪, 5年经验, NULL)  
jid2 (司机, 北京, NULL, 2年经验, 硕士研究生)  
zid1 (司机, 北京, NULL, 3年经验, NULL)

### 共性

(司机, 北京, NULL, 2年经验, NULL)

- 内容推荐: content-based Recommendation
- 原理: 抽取共有属性
- 举例: 商家下载简历的推荐
- 步骤:
  - (1) 历史行为收集
  - (2) id详情查询
  - (3) 共性内容挖掘
  - (4) 推荐



## 综合排序-CTR预估

### 常见算法

用户产品：ctr

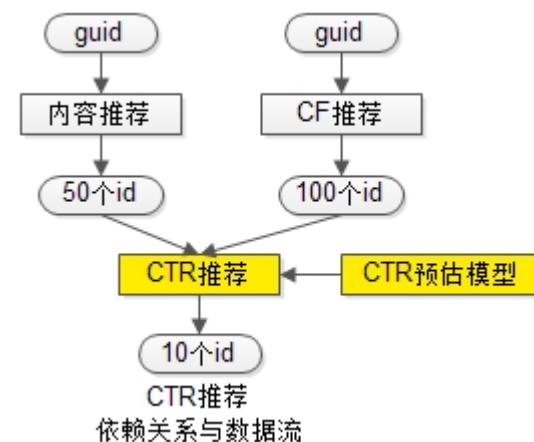
CPC商业产品：ctr \* price

CPA商业产品：ctr

地区+职位+薪酬范围+工作经验+学历  
发帖时间+是否下载过+是否浏览过+...

ctr-score = a\*f(地区) + b\*f(职位) + c\*f(薪酬) + d\*f(工作经验) + e\*f(学历)  
+ f\*f(发帖时间) + g\*f(是否下载过) + h\*f(是否浏览过) + ...

- CF算法推荐了50个jid，内容推荐算法推荐了100个jid，最终页面只需要返回5个jid，如何返回？哪个排前面？
- 综合排序
- 什么决定综合排序？
- CTR由什么决定？
- CTR预估打分公式？



## 线下+线上的系统

# 难点 设计 实现

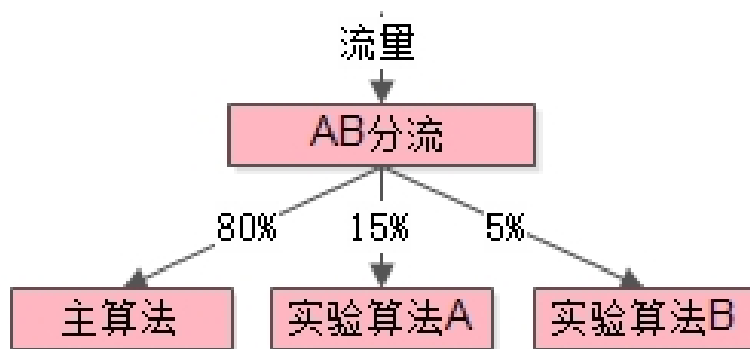
- 线下系统
- 线上系统
- 几个问题
  - (1) 线下数据如何存储?
  - (2) 线上数据如何存储?
  - (3) 如何进行数据转化?



## 支持实验的系统

# 难点 设计 实现

- 如何做算法测试？
- 如何快速支持一个算法平台？
- 如何实现分流AB测？



支持实验的系统

# 支持实验的系统-分流平台

## 难点 设计 实现

- 作用？需求？
  - (1) 支持random分流 (2) 支持取模分流
  - (3) 支持与或非表达式 (4) 支持集合操作 (5) 配置热加载
- 支持哪些属性的与或非，集合操作？

```
# 招聘数据路径↵
zhaopin.datapath: /opt/data/zhaopin/↵
# 招聘组合筛选条件↵
# 以|分隔，A表示实验平台，1表示优先级，第三部分表示组合筛选条件，多个条件以逗号分隔↵
zhaopin.default: A|70↵
zhaopin.condition: A|1|entry.entryDeviceType=IOS,visitor.visitorUserId^(uid_whitelist)↵
zhaopin.condition: A|2|visitor.visitorCookieid=58tongcheng↵
zhaopin.condition: B|3|visitor.visitorUserIP=127.0.0.1↵
zhaopin.condition: A|4|business.businessEntityId%5%0↵
zhaopin.condition: B|5|default↵
```

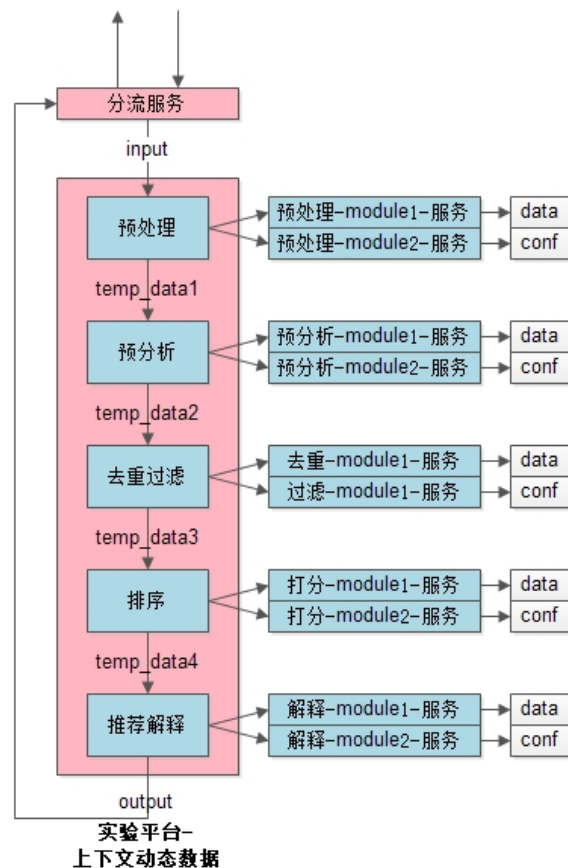




# 支持实验的系统-推荐内核

## 难点 设计 实现

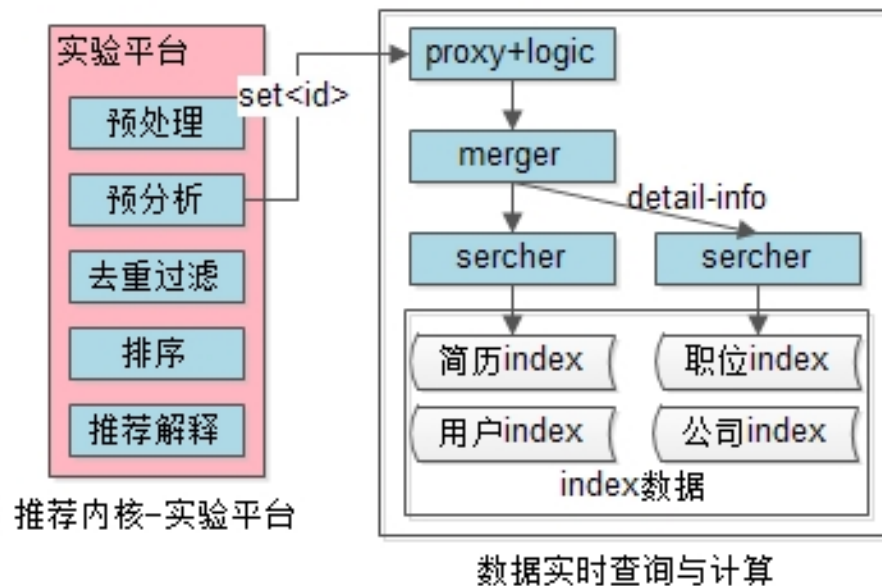
- 作用？需求？
- 设计与实现
  - (1) 算法平台的抽象
  - (2) 实验平台的扩展
  - (3) 上下文动态数据
  - (4) 异步框架+状态机



# 实时计算的检索系统

## 难点 设计 实现

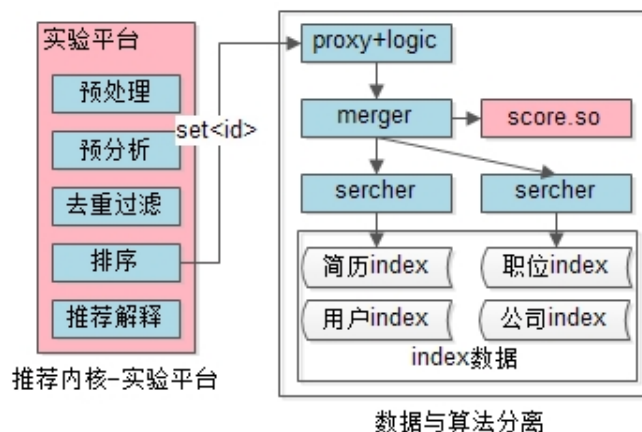
- 作用？
- 通用需求
  - (1) 正排数据
  - (2) 倒排索引
  - (3) 数据
  - (4) map-reduce
- 业务需求



## 工程+算法的系统

# 难点 设计 实现

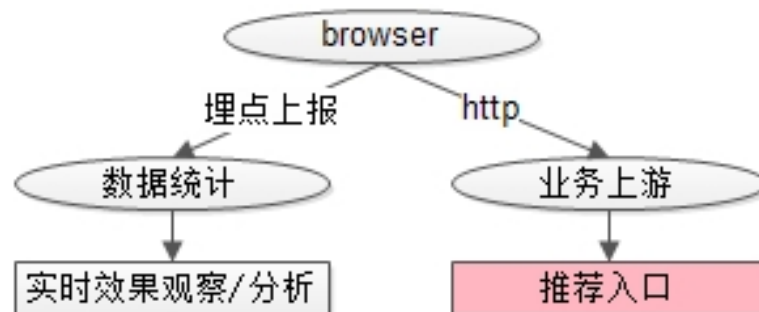
- 如何让工程和算法解耦？
- 线下算法如何分离？
- 实验平台算法如何分离？
- 排序打分算法如何分离？



## 效果实时监测的系统

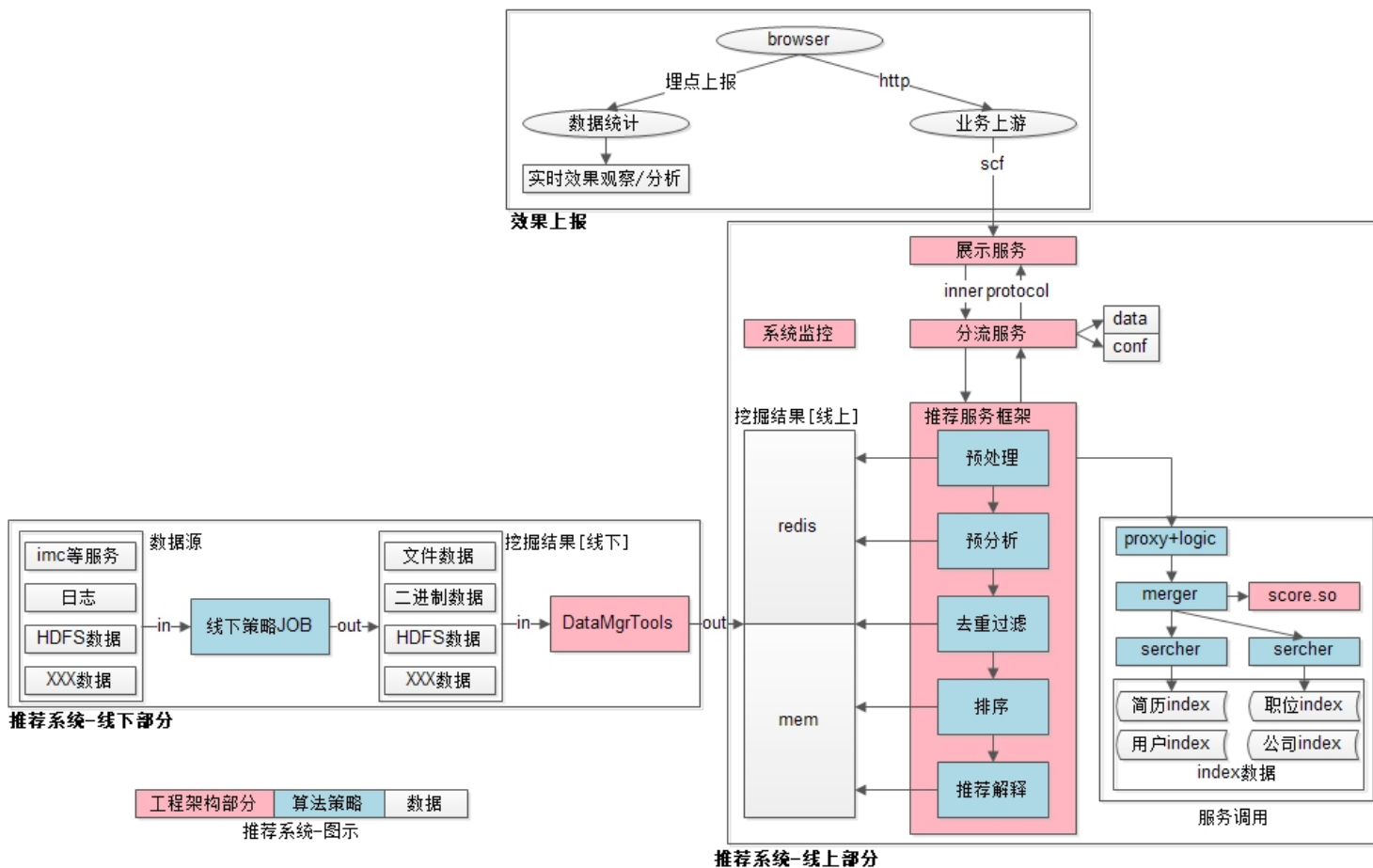
# 难点 设计 实现

- 作用？
- 步骤
  - (1) 结果展示
  - (2) 结果上报，点击上报
  - (3) 数据收集，数据统计
  - (4) 数据展示



## 难点 设计 实现

## 总揽



## 总结+启示



### 案例启示

- 总结
  - (1) 线下+线上的系统
  - (2) 支持实验的系统
  - (3) 实时计算的检索系统
  - (4) 工程+算法的系统
  - (5) 效果实时监测的系统
- 启示：一切不以支持业务为目的的技术选型及架构设计都是耍流氓





全球软件案例研究  
峰会

# 讨论 & 交流



TOP 100 CASE STUDIES  
OF THE YEAR

[www.top100summit.com](http://www.top100summit.com)



全 球 软 件 案 例 研 究  
峰 会

# Thank You!



TOP 100 CASE STUDIES  
OF THE YEAR

[www.top100summit.com](http://www.top100summit.com)