

推荐系统 de 工程挑战

RecSys Engineering

陈天健

推荐与个性化部

2012.12

内容

RecSys Engineering

- 背景介绍
- 争分夺秒
- 重新思考
- 未来之路

背景介绍

Background

以搜索
满足用户的**主动表达**的需求



以推荐
挖掘并满足用户的**潜在需求**



国内趋势

Background

互联网产品

- 蜂拥升级个性化推荐功能
- 产品设计者思路变化

推荐技术

- 突破协同过滤，多采用复合算法
- 系统规模飞速成长

内容

RecSys Engineering

- 背景介绍
- 争分夺秒
- 重新思考
- 未来之路

工程挑战

RecSys Engineering



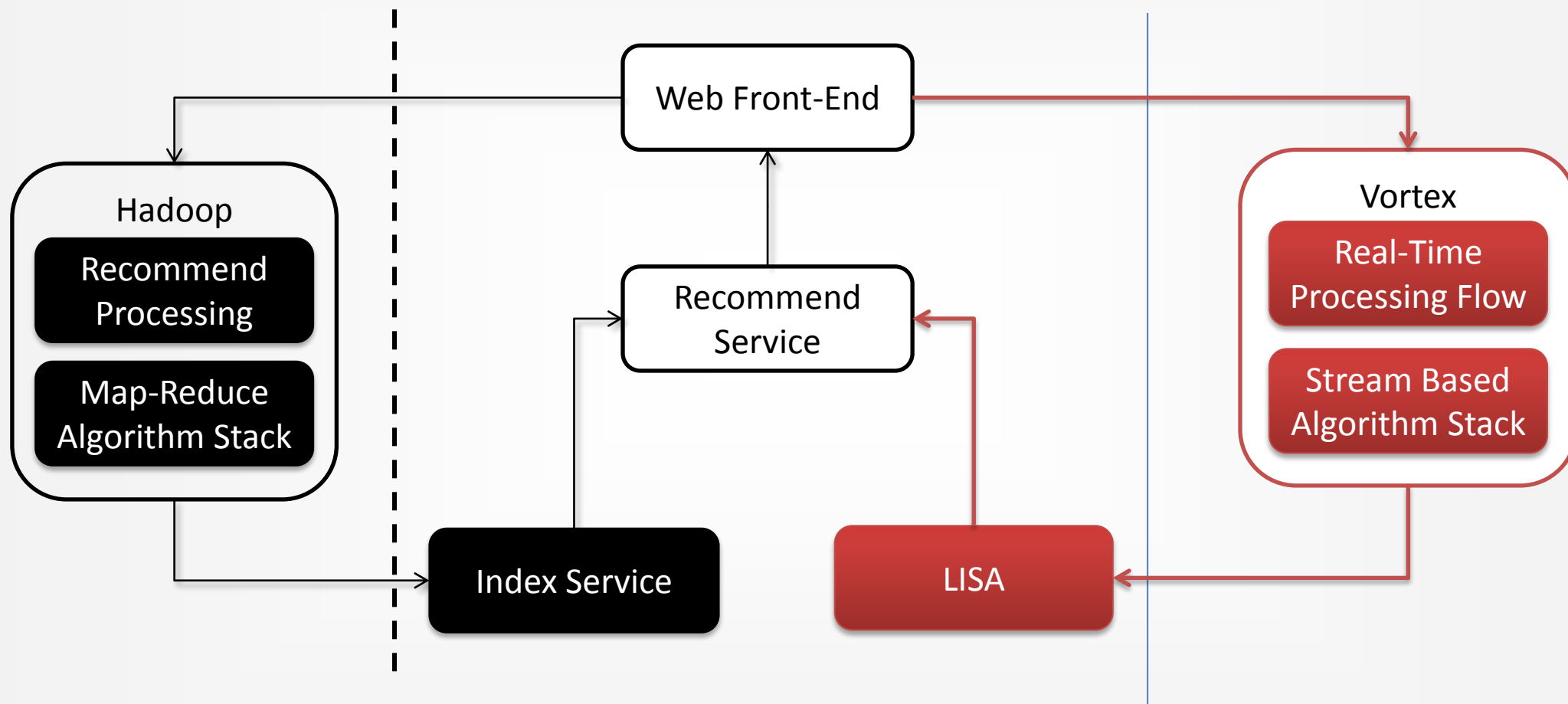
工程挑战

RecSys Engineering

当 大数据 遭遇 时效性？

争分夺秒

RecSys Engineering



关键系统技术

RecSys Engineering

Vortex 流式计算系统

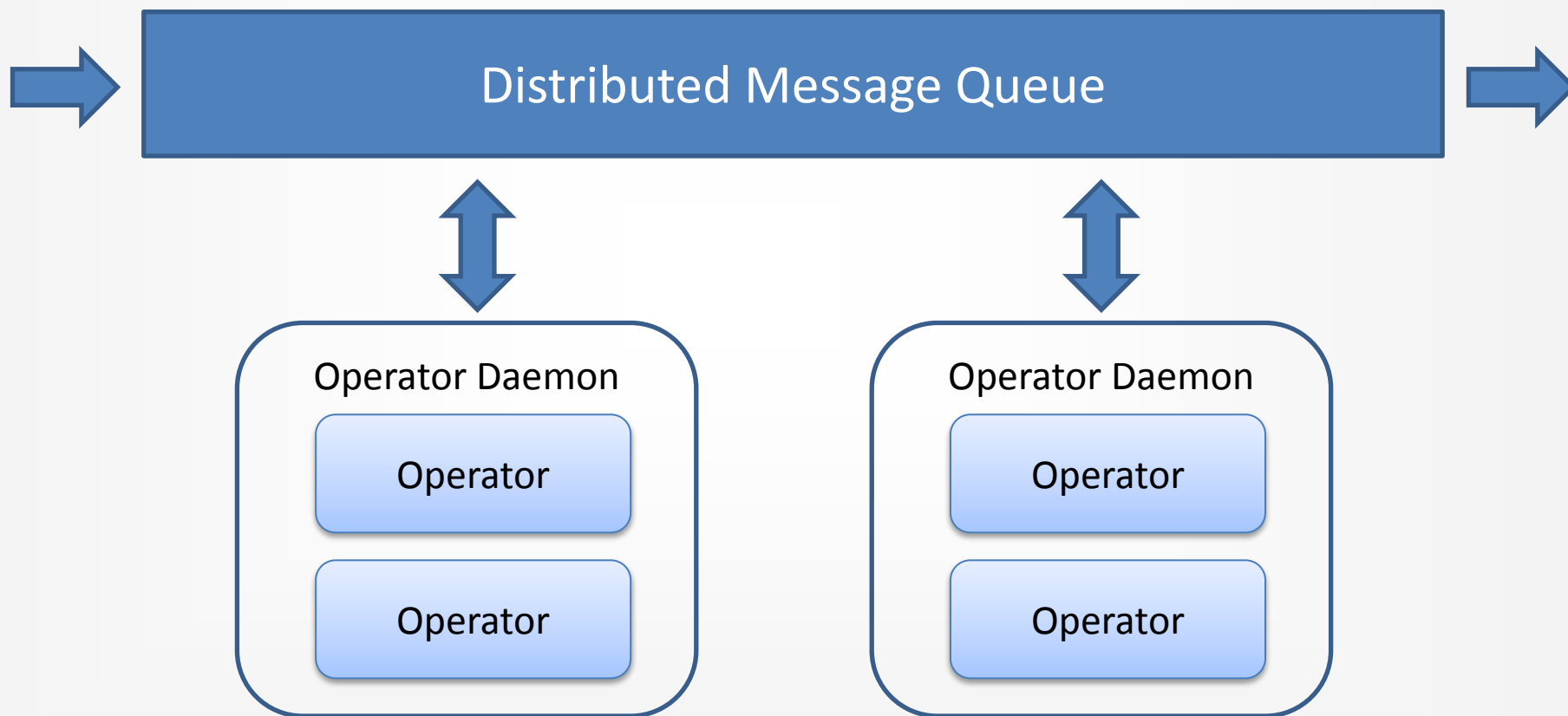
- 数据处理延迟从数小时级下降数秒级别
- 流量波动影响钝化

LISA 实时索引架构

- 实时索引|实时展现
- 索引|查询和更新操作融合

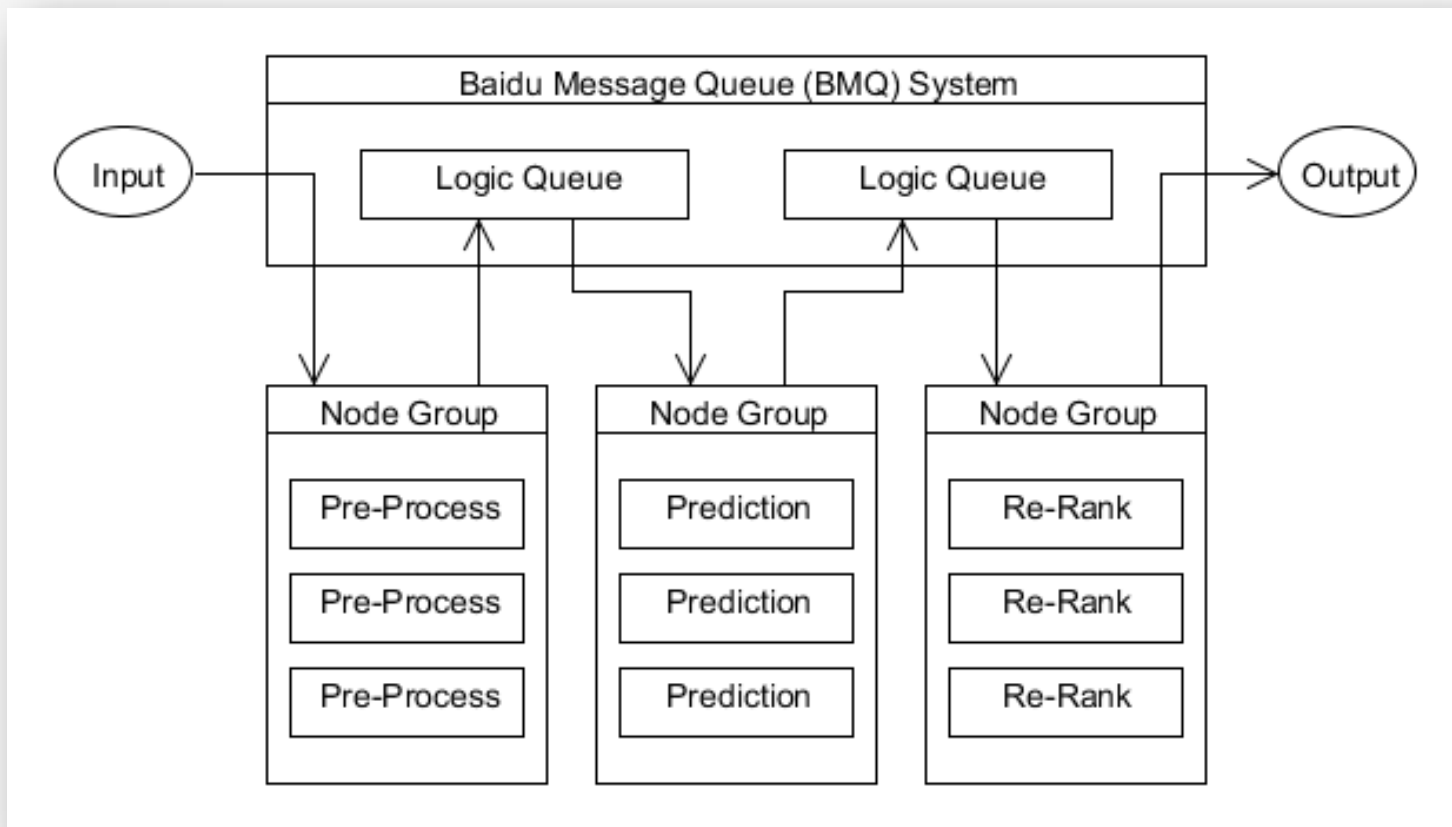
Vortex简图：Queue-Worker

RecSys Engineering



Vortex应用例子

RecSys Engineering



Vortex 技术特点

RecSys Engineering

简单可依赖

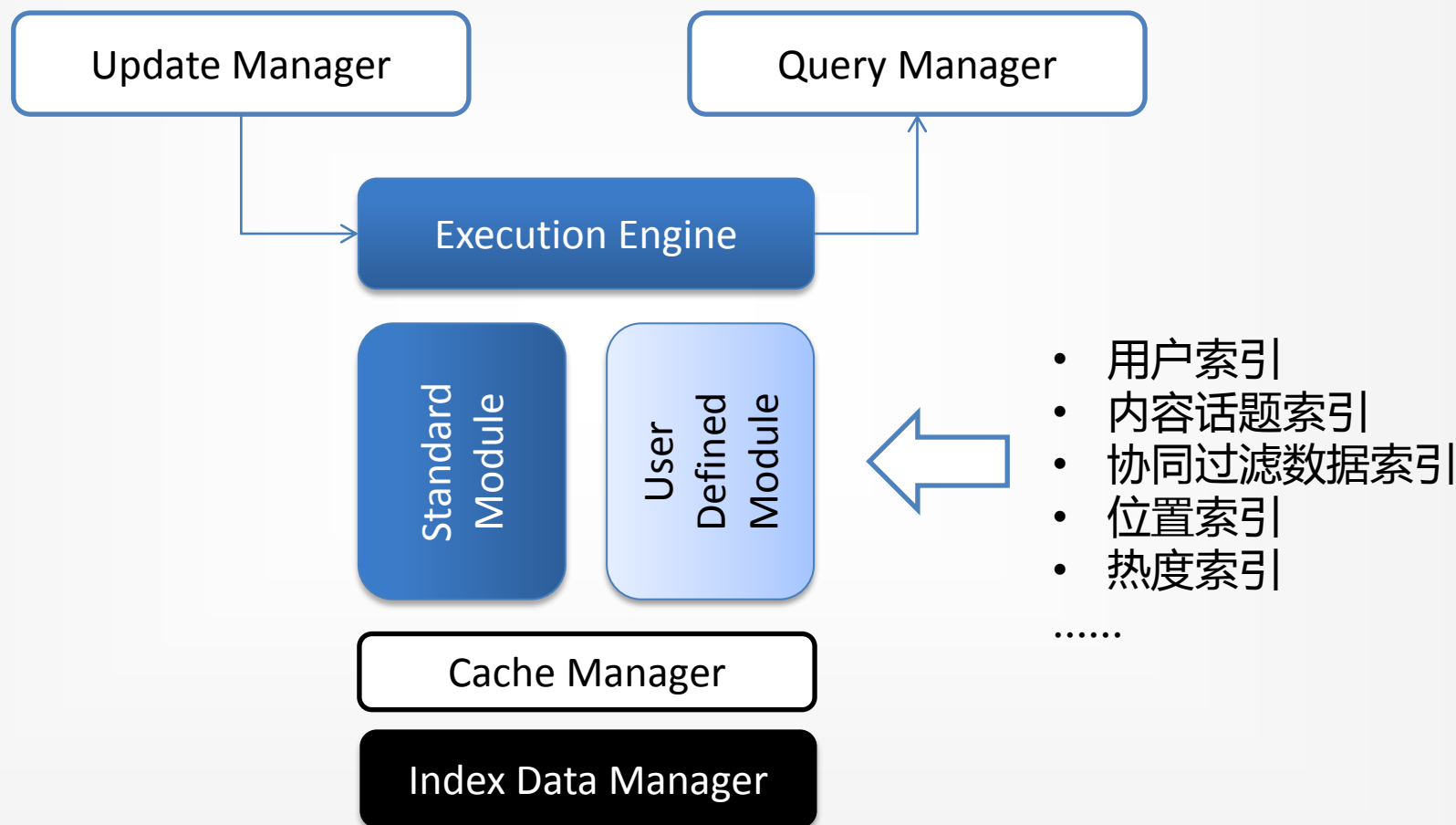
- 系统可用性好
- 功能扩展简单

系统灵活性

- 拓扑不局限于DAG
- 与BVC计算网络融合

L.I.S.A

RecSys Engineering



L.I.S.A技术特点

RecSys Engineering

简单可依赖

- 索引功能模块组件化

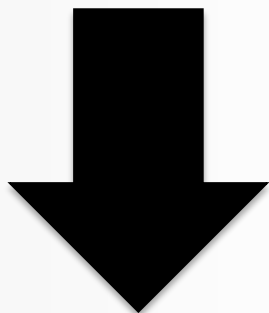
所见即所得

- 从周期更新到实时更新

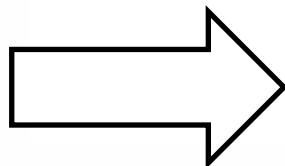
时效性的优势

RecSys Engineering

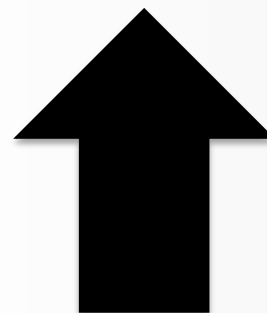
更新延迟



2小时 => 5秒



点击率



3x

内容

RecSys Engineering

- 背景介绍
- 争分夺秒
- 重新思考
- 未来之路

工程挑战

RecSys Engineering

时效性 并不是唯一挑战

算法升级周期

RecSys Engineering

公司	更新周期
Netflix	3 个月
eBay	3~6个月
Google+	约1个月
LinkedIn	1~2个月
Facebook	2~4个月

注：非官方渠道来源信息

算法寿命缩短的原因

RecSys Engineering



工程挑战

RecSys Engineering



重新思考

RecSys Engineering

每个推荐系统都要重起炉灶？

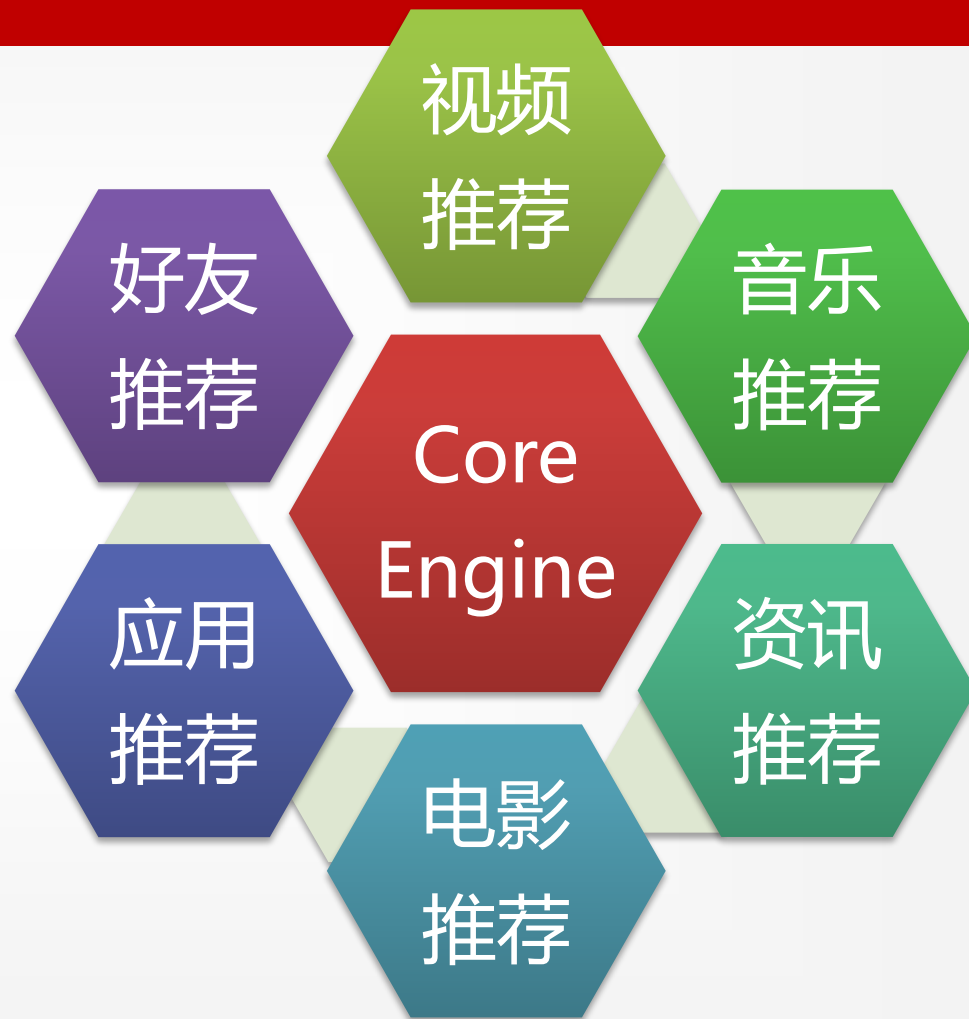
从推荐系统到推荐引擎

RecSys Engineering

以推荐引擎为核心

构建各种推荐系统和应用

最大限度复用数据、算法、系统资源



推荐引擎的规模化效应

RecSys Engineering



- 成本：推荐系统工程成本摊低
- 算法：问题的维度变化，从“艺术”转向“科学”
- 数据：不同应用共享数据，克服数据稀疏问题
- 效果：推荐产品功能横向打通，多领域满足用户需求

引擎的核心功能

RecSys Engineering

Exploit

使用过程优化

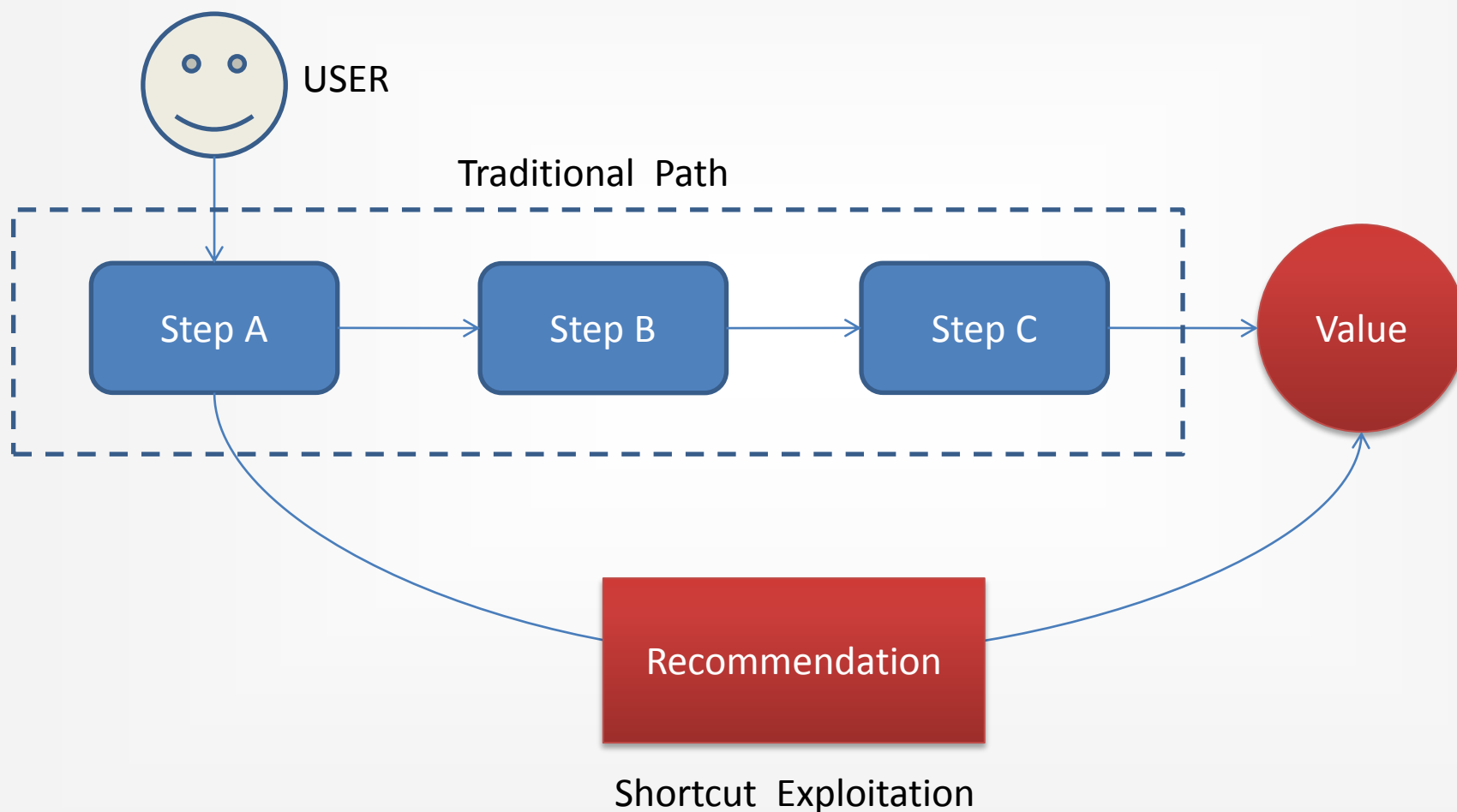


Explore

用户兴趣发现

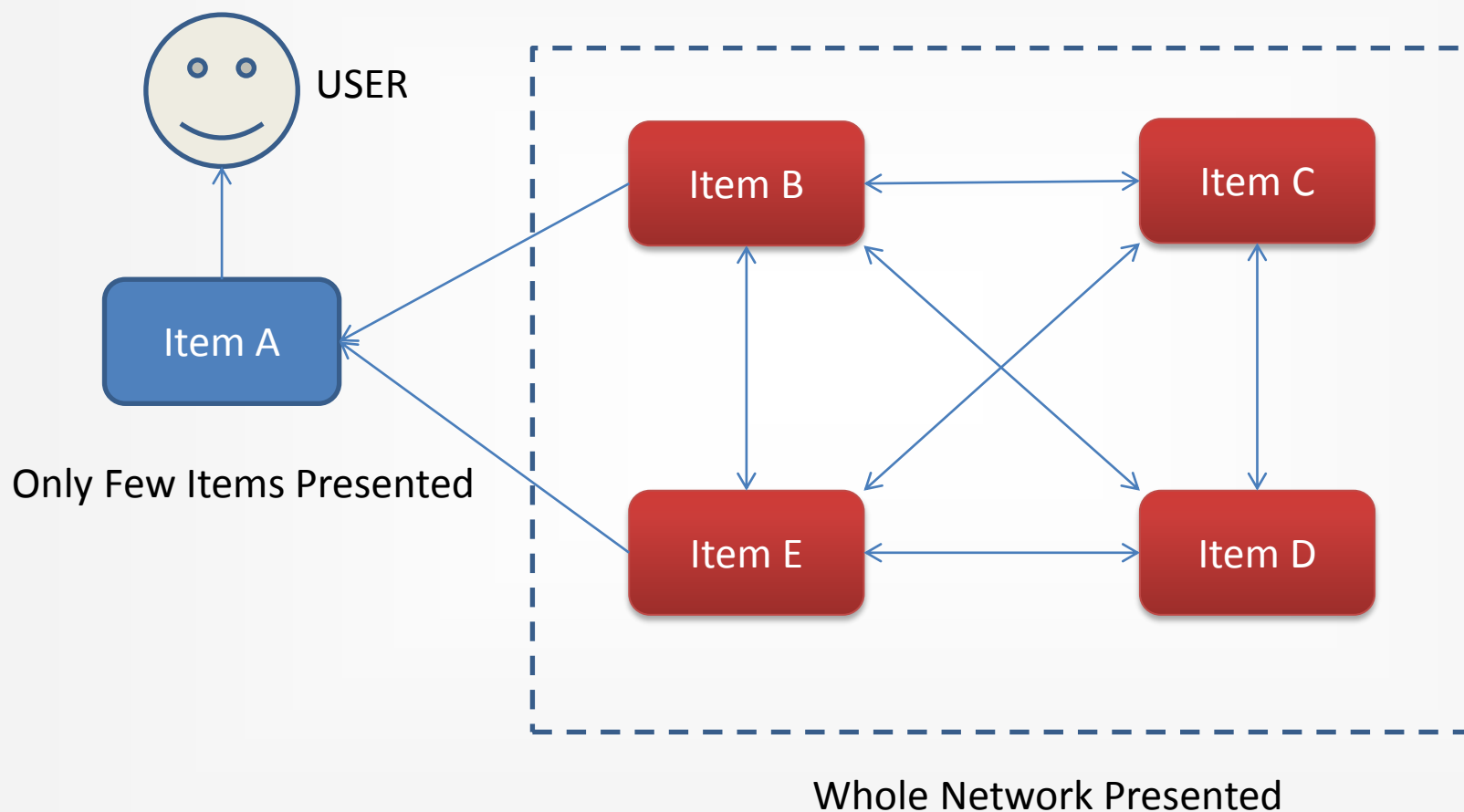
核心功能之一：过程优化

RecSys Engineering



核心功能之二：兴趣发现

RecSys Engineering



技术特点

百度推荐引擎 CoreEngine 1.0

推荐算法

- 召回算法 排序算法

系统能力

- 实时更新 超大规模计算 交互和实验

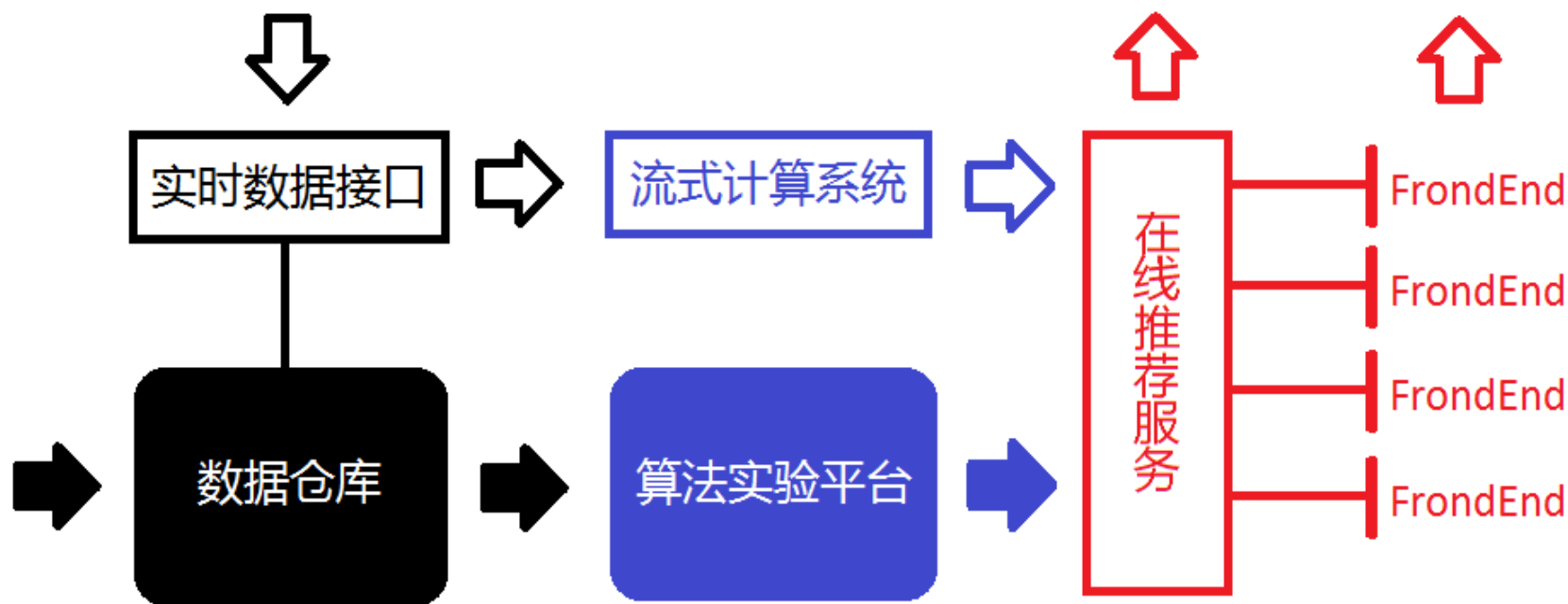
推荐算法模块

百度推荐引擎 CoreEngine 1.0



系统视图

百度推荐引擎 CoreEngine 1.0



系统能力

百度推荐引擎 CoreEngine 1.0

时效性

- 用户模型更新延迟 < 5sec
- 内容索引更新延迟 < 50sec

计算规模

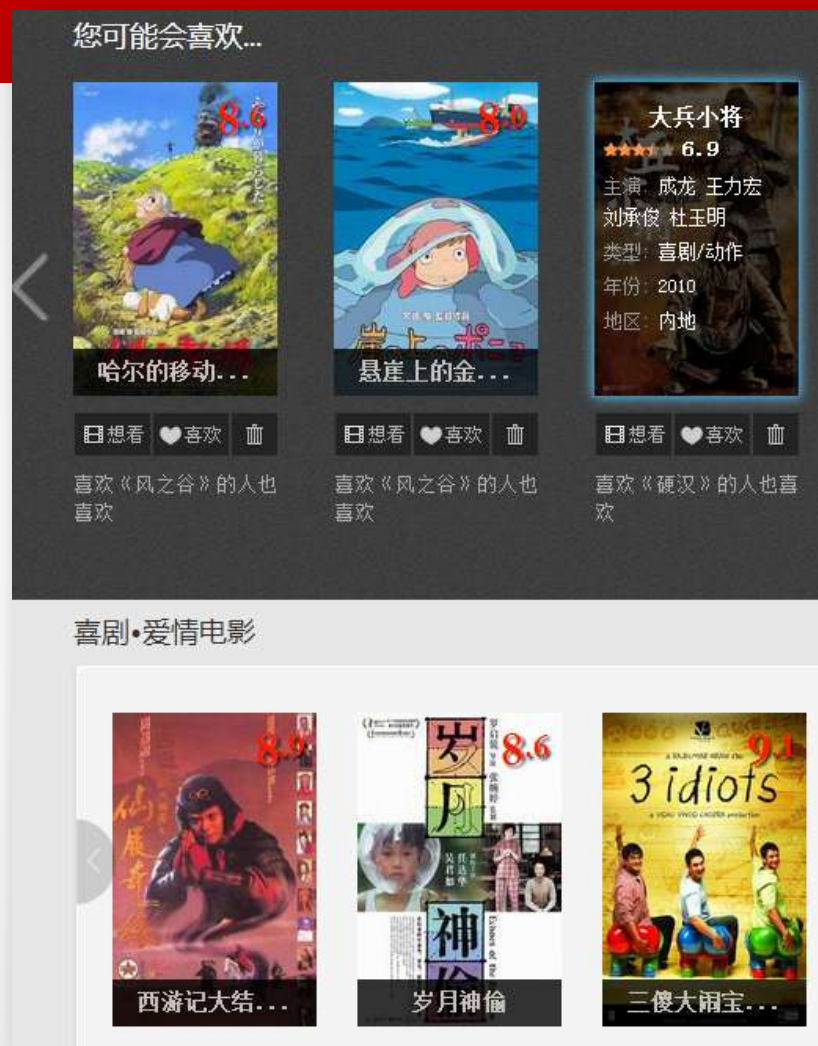
- 流式计算：
> 2M pairs/sec
- 在线计算：
> 1M pairs/sec

发现和优化并重：百度随心看

百度推荐引擎实践

- 个性化电影推荐
- Item关联和本体关联复合技术
- 点击率高于传统列表页数倍

<http://suixinkan.baidu.com/>



发现和优化并重：百度新闻2.0

百度推荐引擎实践

- 个性化资讯推荐
- 机器学习用户建模和本体关联复合技术
- 差异化竞争，后发先至



<http://app.news.baidu.com/>

偏重Exploit：百度“找”知道

百度推荐引擎实践

- 服务过亿知道用户
- 智能学习用户兴趣
- 及时回答率提高50%



<http://zhidao.baidu.com/>

偏重Explore：音乐随心听

百度推荐引擎实践

- 自适应用户喜好变化
- 音乐专家系统技术

<http://fm.baidu.com/>



偏重Explore：相关电影推荐

百度推荐引擎实践

- 相似电影推荐
- 协同过滤和本体关联复合技术

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

类似盗梦空间的电影_热门结果推荐



禁闭岛

★★★★★ 8.3

莱昂纳多·迪卡普里奥主演的心理/悬疑电影



蝴蝶效应1

★★★★★ 8.6

美国心理/悬疑电影



记忆碎片

★★★★★ 8.5

克里斯托弗·诺兰导演的悬疑/犯罪电影



致命魔术

★★★★★ 9.8

克里斯托弗·诺兰导演的悬疑科幻电影



蝙蝠侠前传2：...

★★★★★ 8.8

克里斯托弗·诺兰导演的科幻/犯罪电影



零号嫌疑犯

★★★★★ 6.6

美国心理/悬疑电影



K星异客

★★★★★ 8.4

美国心理/悬疑电影



蝴蝶效应3

★★★★★ 6.4

美国心理/悬疑电影

1 2 3

video.baidu.com 2012-10-23 -

内容

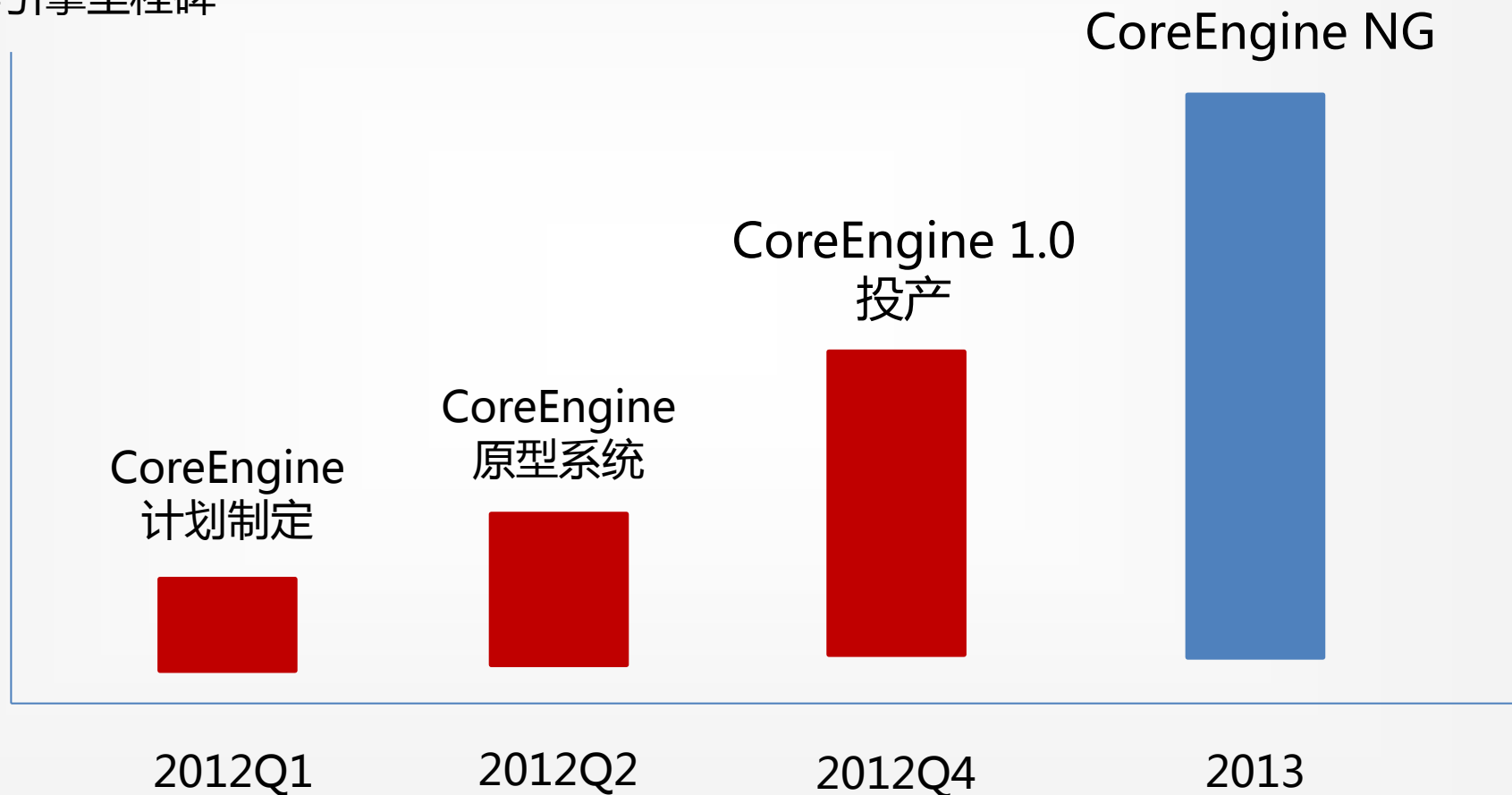
RecSys Engineering

- 背景介绍
- 争分夺秒
- 重新思考
- 未来之路

百度推荐引擎发展

RecSys Engineering

推荐引擎里程碑



百度推荐引擎的未来

RecSys Engineering

百度域内

- 对于百度产品的推荐功能进行完善支持
- 通过推荐引擎实践，保障新产品开发

百度域外

- 逐步向合作伙伴开放推荐引擎能力
- 支持国内推荐技术社区发展

再次谢谢各位莅临

RecSys Engineering

http://hr.baidu.com/jobPages/jobDetail_2835.html

百度技术沙龙

畅想

交流

争鸣

聚会

关注我们：t.baidu-tech.com

资料下载和详细介绍：infoq.com/cn/zones/baidu-salon

“畅想•交流•争鸣•聚会”是百度技术沙龙的宗旨。百度技术沙龙是由百度与InfoQ中文站定期组织的线下技术交流活动。目的是让中高端技术人员有一个相对自由的思想交流和交友沟通的平台。主要分讲师分享和OpenSpace两个关键环节，每期只关注一个焦点话题。

讲师分享和现场Q&A让大家了解百度和其他知名网站技术支持的先进实践经验，OpenSpace环节是百度技术沙龙主题的升华和展开，提供一个自由交流的平台。针对当期主题，参与者人人都可以发起话题，展开讨论。

InfoQ 策划·组织·实施

关注我们：weibo.com/infoqchina