

# 电商搜索引擎的架构设计 和性能优化

吴英昊

WeChat: [wyh267](#)

Weibo: [weibo.com/wuyinghao](#)



# 提纲

- ◆ 什么是电商的搜索引擎
- ◆ 电商搜索引擎的架构和优化方案
- ◆ 实际遇到的问题 and 解决方案



**什么是电商搜索引擎**



# 两种不同的搜索引擎

**dangdang.com** 尾品汇 1折起

百年孤独

全部分类 搜索

热搜: 当当优品 尖货1折抢 茵曼 李光耀 蚊帐 高级搜索

全部商品详细分类 尾品汇 新品闪购 图书 电子书 服装 运动户外 孕婴童 家居 当当优品 电器城 当当

你是不是想找: 马尔克斯 | 霍乱时期的爱情 | 生命不能 | 施咸荣 | 百年孤独英文版 | 百年孤独 马尔克斯 | 追风筝的人 | 致加西亚的信

全部 > 百年孤独

世界名著 美洲 欧洲 更多选项

默认排序 销量 好评 出版时间 价格 收货地

**五星** 百年孤独 诺贝尔文学奖获得者加西亚·马尔克斯巅峰杰作

**¥29.30** ¥39.50 (7.4折)

(哥伦比亚) 马尔克斯 著, 范晔 译 / 2011-06-01 / 南海出版社

★★★★★ 383820条评论

当当自营

如何阅读一本小说 加西亚·马尔克斯 相关推荐 ↓ 《迷途》 永年经典译文! 真实的玻利瓦尔只存在于历史深处, 这本

加入购物车 收藏

是什么??!

买什么??!

**Baidu 百度** 百年孤独

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1,550,000个

为您推荐: [百年孤独读后感](#) [百年孤独经典语录](#) [百年孤独txt下载](#) [百年孤独 电影](#)

**百年孤独\_百度百科**



《百年孤独》，是哥伦比亚作家加西亚·马尔克斯的代表作，也是拉丁美洲魔幻现实主义文学的代表作，被誉为“再现拉丁美洲历史社会图景的鸿篇巨著”。作品描写了布恩迪亚家族七代人的传奇故事，以及加勒比海沿岸小镇马孔多的百年兴衰，反映了...

[内容梗概](#) [人物介绍](#) [创作背景](#) [点评鉴赏](#) [成就影响](#) [更多>>](#)

[baike.baidu.com/ 2015-01-24](#)

**百年孤独在线阅读**



百年孤独·内容提要·第一章 第二章 第三章 第四章 第五章 第六章 第七章 第八章 第九章 第十章 第十一章 第十二章 第十三章 十四章 第十五章 第...

[www.uus8.org/c/60/0...](#) - 百度快照 - 94%好评

**百年孤独 (豆瓣)**



# 电商搜索引擎的特点

- ◆ 没有爬虫系统
- ◆ 数据非常结构化
- ◆ 过滤功能的需求甚至大于搜索本身
- ◆ 支持各种维度的排序
- ◆ 数据实时性要求非常高
- ◆ 绝对不能丢品
- ◆ 与推荐系统和广告系统的融合
- ◆ 需要绝对的高可用，不宕机



全部 > 母婴用品 > 尿裤 > 纸尿裤/片 > 尿不湿

品牌

雀氏  
ChinUS

Anerle  
安儿乐

好奇  
HUGGIES

MamyPolo  
妈咪宝贝

Merries  
帮宝适

Pampers  
帮宝适

可爱宝贝

GOO.N

Suitsy  
舒比奇

Carefor  
爱舒

moony

丽贝乐  
Libero

适用年龄

0-6个月

7-12个月

1-2岁

适用性别

男

女

通用

产地

中国

日本

韩国

美洲

欧洲

海外其他

尿裤型号

NB

S

M

L

XL

XXL

其它尺寸

包装

正常包装

箱装

组合/套装

礼盒装

收起 ^

默认排序

销量 ↓

好评 ↓

最新 ↓

价格 ↑↓

¥

- ¥

收货地

☐ 促销

☐ 当当自营

☐ 手机专享价

☐ 预售

☐ 尾品汇

过滤功能的需求大于搜索本身



# 电商搜索引擎的架构

# 实现方式

- ◆ Lucene + 自己封装
- ◆ Solr
- ◆ Elastic Search
- ◆ 自己来



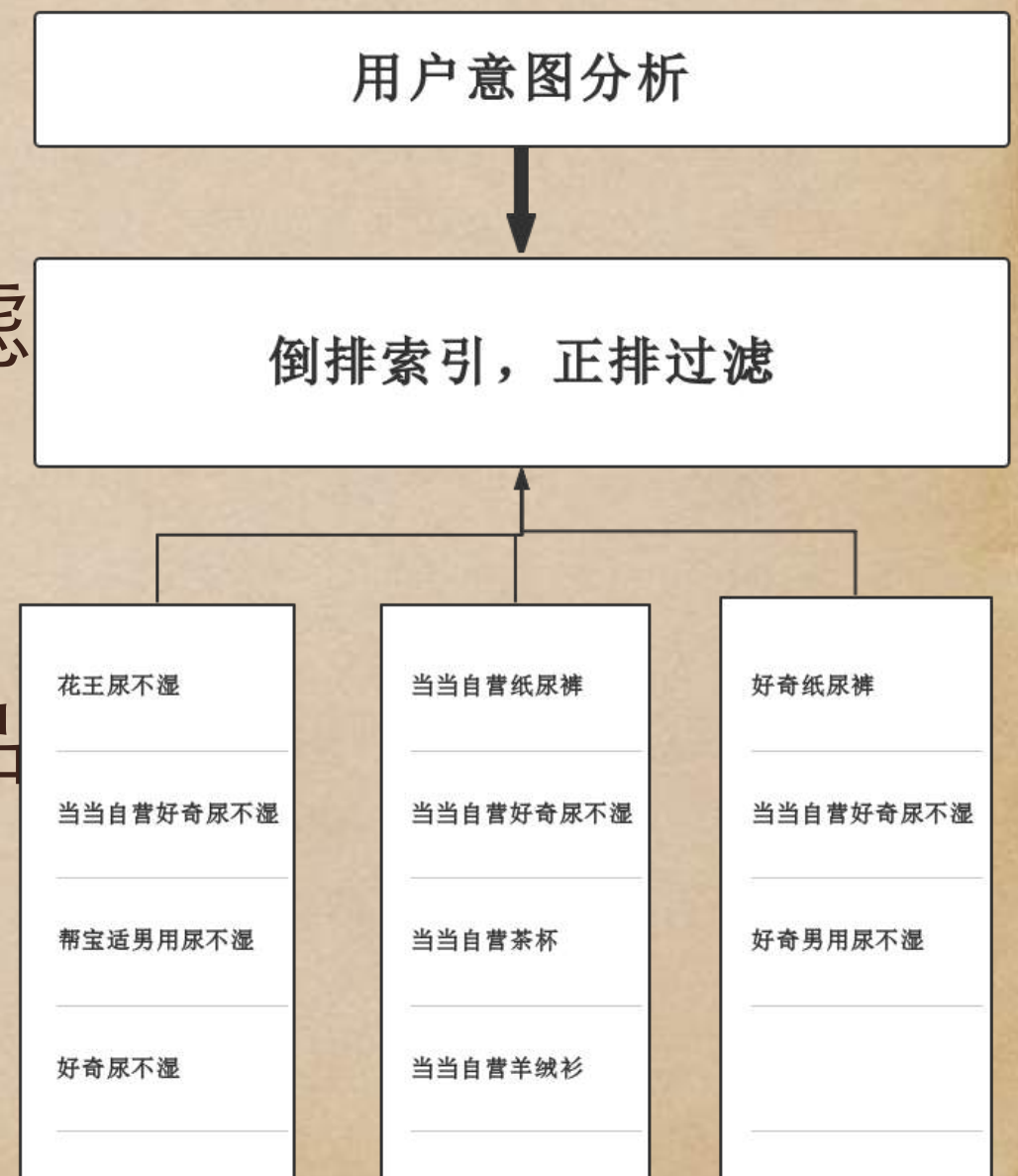
# 电商搜索引擎标配模块

- ◆ 检索模块
- ◆ 数据更新模块
- ◆ 排序模块
- ◆ 分词模块
- ◆ query分析模块



# 检索模块

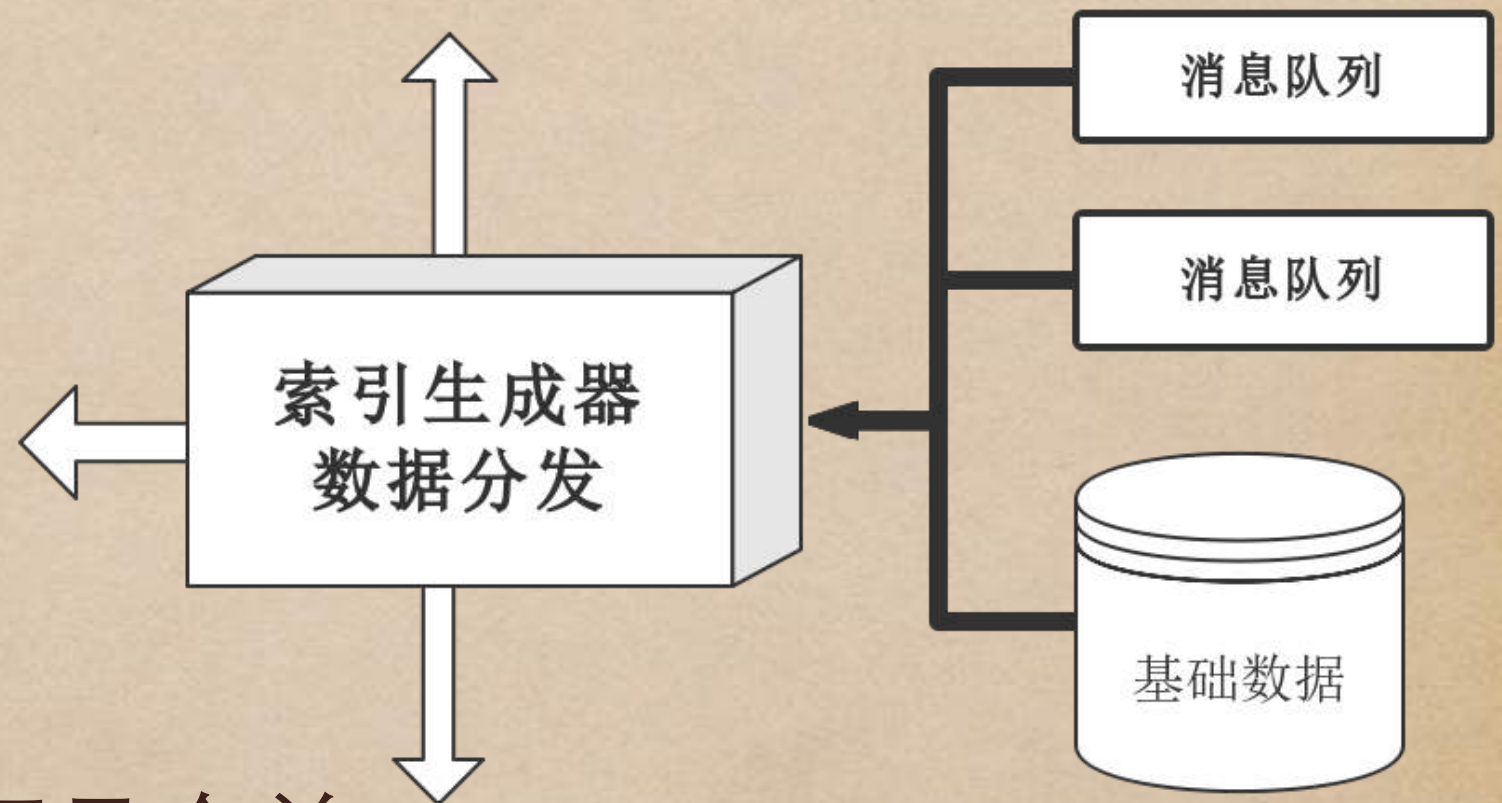
- ◆ 用户意图分析
- ◆ 分词系统
- ◆ 倒排索引召回，正排索引过滤
- ◆ Http对外服务框架
- ◆ 完全自己实现和使用开源产品
- ◆ C++ 还是 JAVA 还是其他



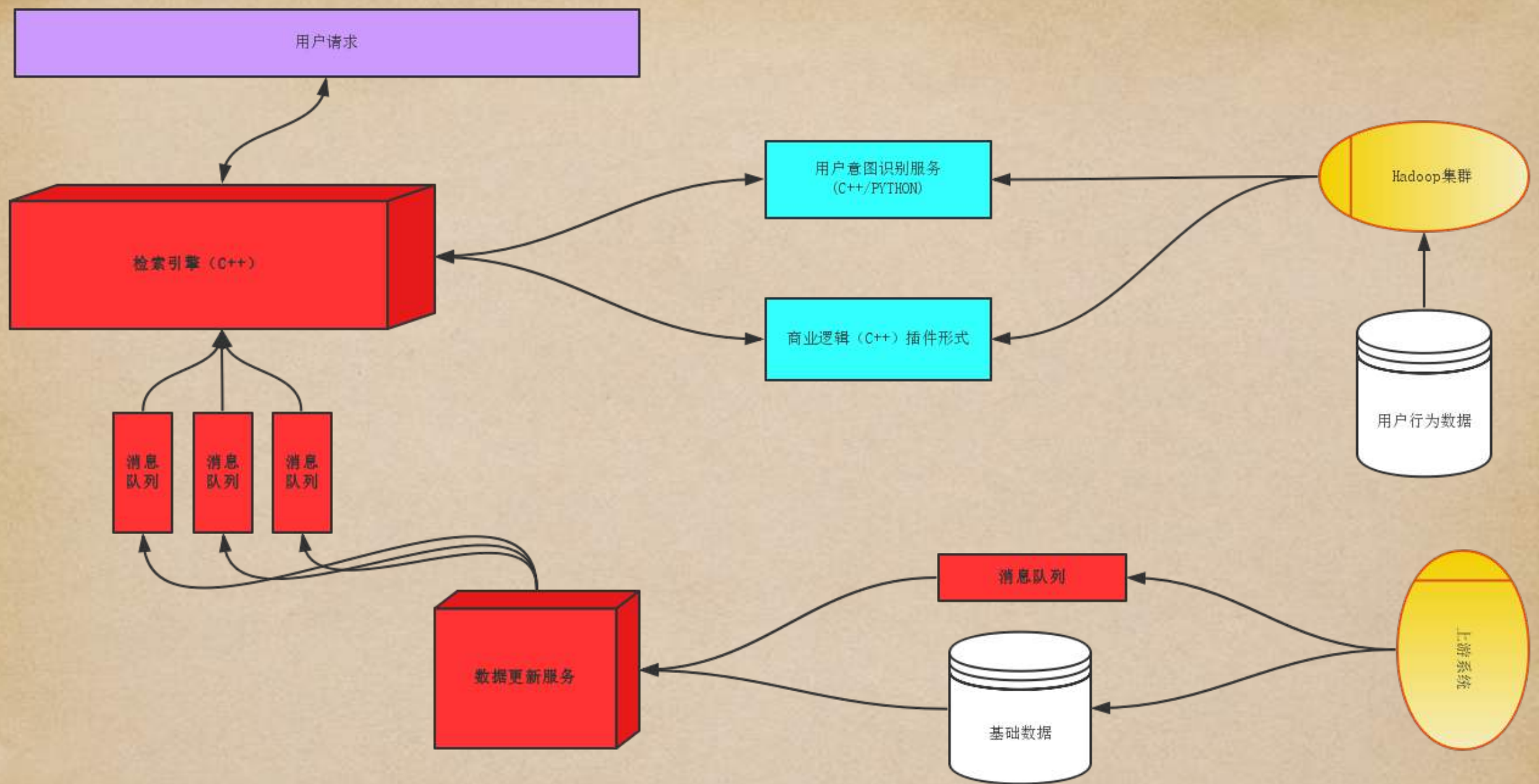


# 数据更新模块

- ◆ 索引生成方式
- ◆ 索引更新方式
- ◆ 索引优化
- ◆ 更新优化
- ◆ 和检索系统分开还是合并
- ◆ 消息队列??







当这模块合在一起时



# 出问题是不可避免的

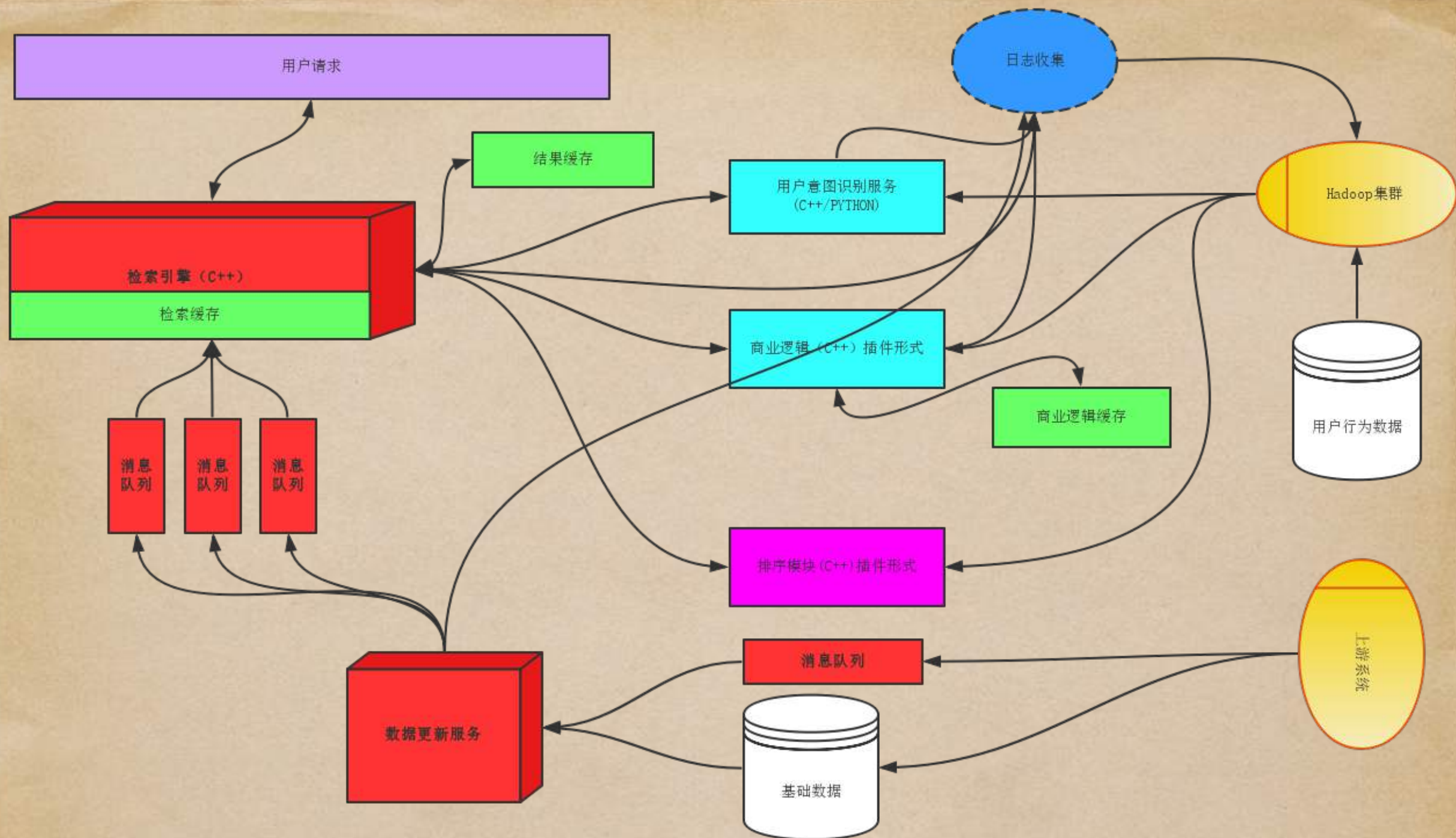
- ◆ 新上线，出BUG，怎么回滚
- ◆ 高并发，没法分库分表，没法静态化，没法排队，怎么办
- ◆ 问题追踪，日志系统和监控系统



# 如何解决问题优化架构

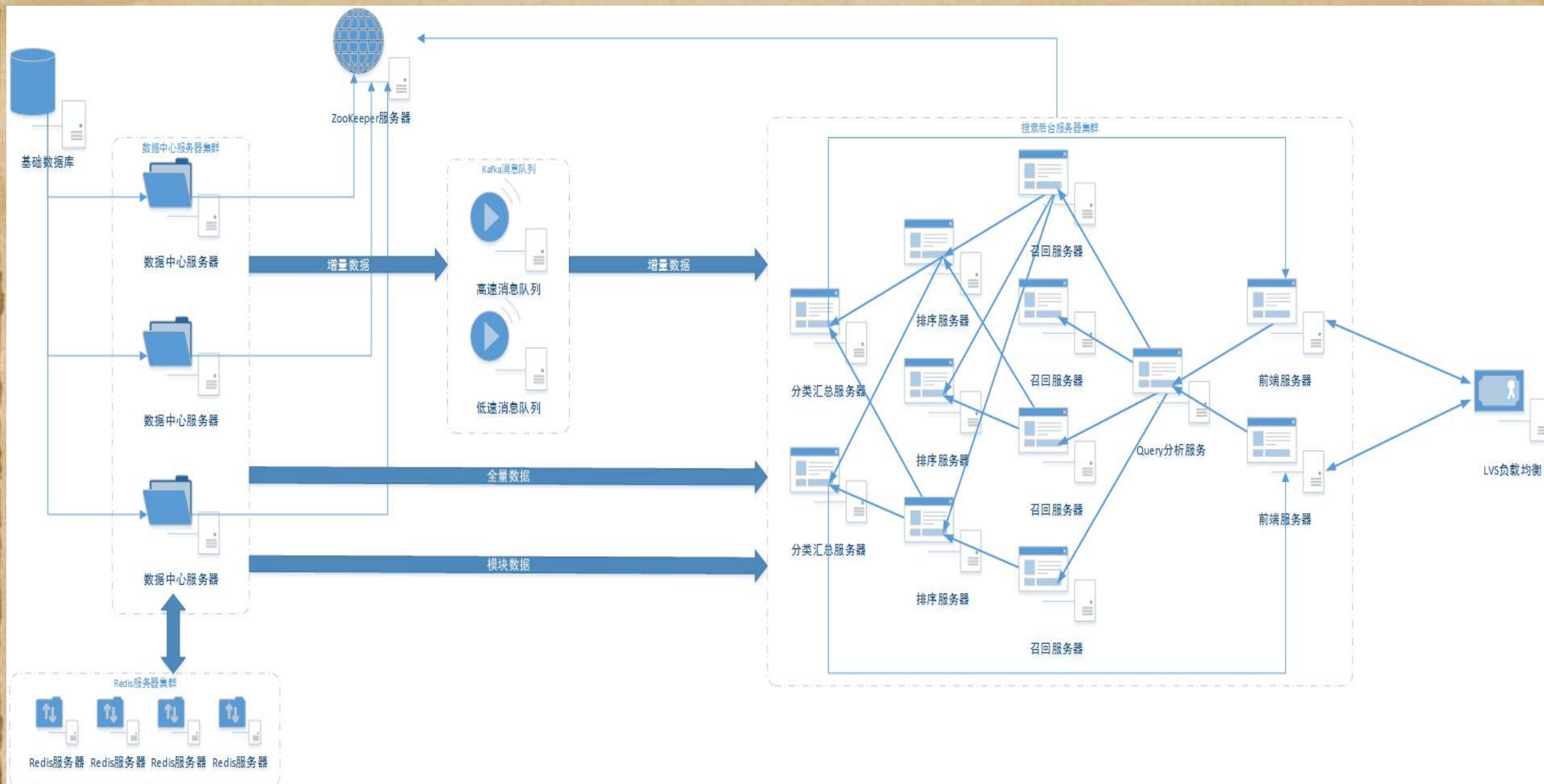
- ◆ 自动化上线，自动化回滚
- ◆ 缓存，横向扩展
- ◆ FLUME日志系统





带缓存，带日志监控的架构





部署架构图



# 闲扯几点

- ◆ 集群机器间通讯采用了什么方式? ZMQ
- ◆ 是否需要zookeeper集中管理?
- ◆ 完全自主实现整个搜索系统是否有必要?
- ◆ 整个系统的性能到底如何?
- ◆ 大促期间的压力如何?



# 实际遇到的问题 and 解决方案



# 系统冷启动

业务场景：  
系统刚刚启动的时候，查询  
时间非常长

- ◆ 为什么存在冷启动的问题
- ◆ 冷启动有什么影响
- ◆ 能避免冷启动吗？





# 如何避免冷启动

- ◆ MMAP 还是 Redis ???
- ◆ 一次性读入内存?
- ◆ 减少全量索引频率
- ◆ 在线索引优化，去除碎片化



# “秒”级更新

业务场景：

大促期间，大量商品在同一时间批量变价

- ◆ 何为“秒”级更新
- ◆ 哪些东西需要“秒”级更新的
- ◆ 真的能做到“秒”级吗？





# 怎么“秒”级更新

- ◆ 索引分片，更多的机器分摊更新量级
- ◆ 给更新的消息队列加上优先级，优先更新某些
- ◆ 提前更新





# 还有很多

- ◆ 搜索排序
- ◆ 分词
- ◆ Query分析
- ◆ 机器学习
- ◆ 推荐广告结合



谢谢大家