



# 去哪儿酒店实时报价搜索技术分享

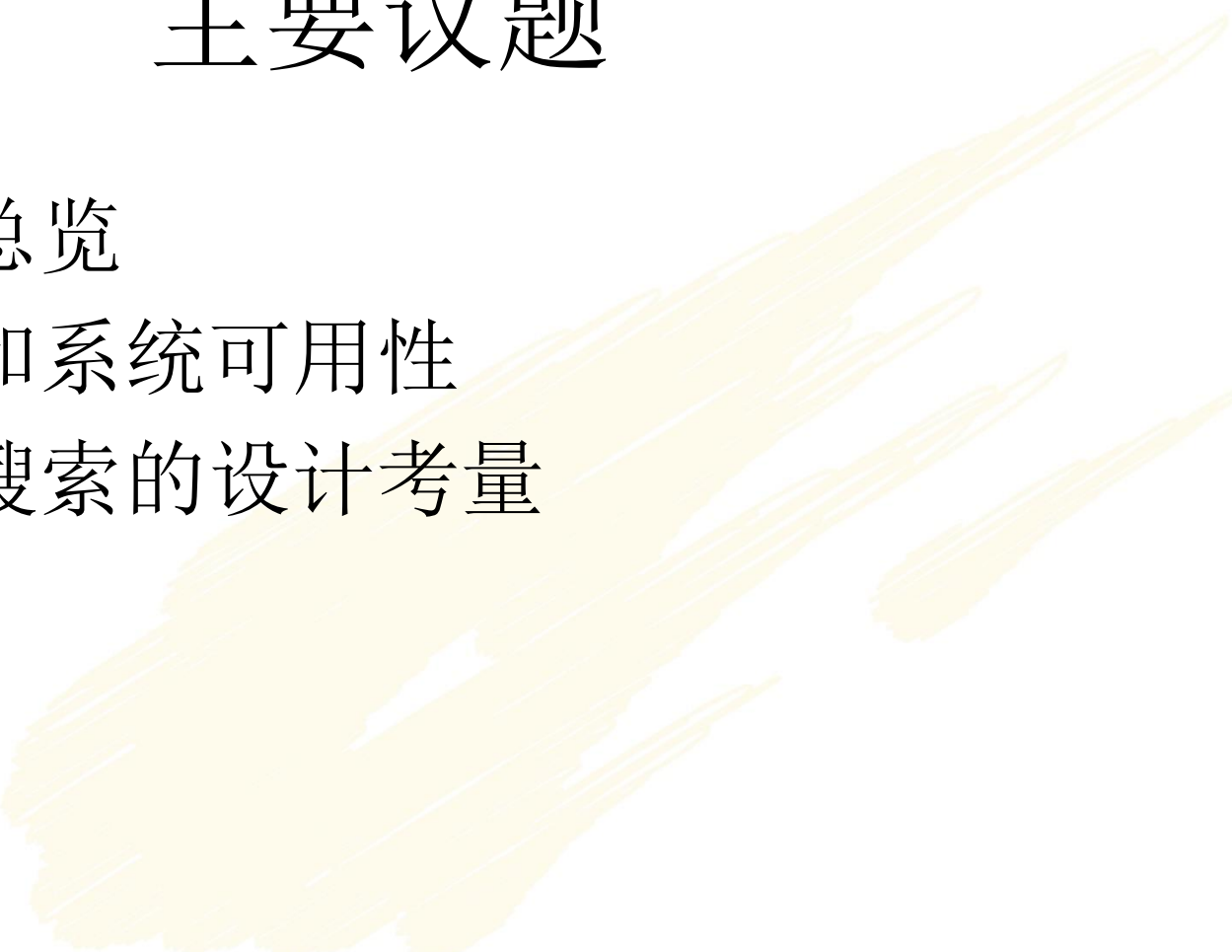
刘 玥

# 关于去哪儿酒店搜索

- 定位
  - 垂直搜索平台
- 目标
  - Smart Your Hotel Reservation
- 目前规模
  - 搜索210家酒店预订站点
  - 支持全球22699个城市
  - 覆盖368892家酒店



# 主要议题

- 系统结构总览
  - 服务拆分和系统可用性
  - 实时报价搜索的设计考量
  - 监控系统
- 

# Overview



登录 | 注册 | 我的订单 | 联系客服 | 收藏



有**3,020,071**次  
尝试从这里改变旅行

首页

机票

酒店

团购<sup>热</sup>

度假

火车票

攻略

旅图

门票<sup>新</sup>

酒店搜索

客栈民宿

酒店一口价<sup>折</sup>

高端酒店<sup>新</sup>

酒店夜销

酒店点评

精品酒店特选

我想找

北京

酒店名、地标、商圈，可组合搜索

日期

入住 2013-07-14

后天

离店 2013-07-15

周一

搜索

按价格: [¥200及以下](#) [¥200 - ¥300](#) [¥300 - ¥500](#) [¥500以上](#)

按档次: [经济型](#) [二星及其他](#) [三星舒适](#) [四星高档](#) [五星级豪华](#)

按热点: [王府井](#) [中关村](#) [首都机场](#) [西单](#) [天安门](#)

按品牌: [7天连锁\(132家\)](#) [如家快捷\(128家\)](#) [汉庭酒店\(91家\)](#) [速8\(82家\)](#) [格林豪泰\(52家\)](#) [99连锁\(50家\)](#)

## 低价之选

**210**家网站价格比较，团购、夜销、一口价多种优惠  
从客栈到精品酒店，从经济到豪华酒店，我们帮您发现最低价

## 丰富之选

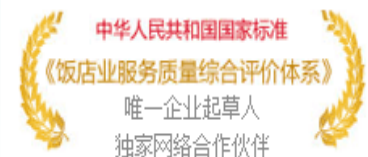
全球**22699**个城市、地区，  
**368892**家酒店

## 安心之选

**213**项标准严格筛选合作伙伴，担保通**100%**安心保障



国泰航空  
荣膺“全球最佳机舱服务员”大奖



中华人民共和国国家标准

《饭店业服务质量综合评价体系》

唯一企业起草人  
独家网络合作伙伴



# Overview

目的地	杭州	要找	武林广场	搜索
日期	入住 2013-07-14	后天	离店 2013-07-15	周一

16 家酒店满足条件 | [清空所有筛选项](#)

价格范围: ☒ 不限 ☐ ¥200以下 ☐ ¥200-¥300 ☐ ¥300-¥500 ☐ ¥500以上 自定义  到

酒店级别: ☒ 不限 ☐ 经济型 ☐ 二星级/其他 ☐ 三星级/舒适 ☒ 四星级/高档 ☐ 五星级/豪华

连锁品牌: ☒ 不限 ☐ 万怡 ☐ 君亭 ☐ 书香酒店 ☐ 最佳西方

[更多](#) ▼

默认排序 ▼ 星级 ▲ 评分(5分) ▼ 价格 ▼ ☐ 当前有房酒店 ☐ 只看团购中酒店

## 折 杭州市武林广场商圈区域商务型酒店 高档型

该酒店位于**武林广场商圈**。位于文三路，地处杭州高新数码中心，东邻热闹繁盛的武林商业圈，西接闻名天下的西湖，距火车站约5公里，位置优越。

一口价6.2折  
为什么隐藏酒店名？

仅¥423起

[查看详情](#)

## 1 杭州武林万怡酒店 高档型

该酒店坐落于**武林广场商圈**内。位于杭州市政府旁边，毗邻历史悠久的武林广场，与西湖美景仅十分钟车程，地处位置无与伦

4.1分/23条点评

杭州好评第190

¥727起

[查看详情](#)

[< 看大地图](#)

☒ 随屏幕滚动 ☐ 拖动地图找酒店 ?



# Overview

杭州武林万怡酒店\_杭州武林万怡酒店预订及特惠价格查询-去哪儿Qunar.com - Windows Internet Explorer

http://hotel.qunar.com/city/hangzhou/dt-5863/?tag=hangzhou#fromDate=2013-07-1

杭州武林广场附近酒店\_杭... 杭州武林万怡酒店\_杭州...

1 豪华大床房 ¥727起 报价列表

2张房型图片 面积:34m² 楼层:4-13层 床型:大床1.8m 设施:上网(免费)

预订网站	综合评级	预订详情	价格	
同程网	5	豪华大床房 (不含早)	¥791	预订
		豪华大床房 (单份早餐) (含单早)	¥895	预订
		豪华大床房 (双份早餐) (含双早)	¥998	预订
来订吧	4.6	豪华大床房 (豪华大床房 (不含早)) (宽带收费)	¥791 返¥64	预订
		豪华大床房 (豪华大床房 (含单早)) (含单早) (宽带收费)	¥895 返¥73	预订
		豪华大床房 (豪华大床房 (含双早)) (含双早) (宽带收费)	¥998 返¥81	预订
自行千里	4.6	豪华大床房 需预付	¥615	订完
		豪华大床房 (含双早) 需预付	¥775	订完

2 豪华双床房 ¥733起 展开报价

3 套房 ¥855起 展开报价

交通信息

- 杭州萧山国际机场 驾车 | 公交 驾车距离: 大约30.3公里(约32分钟)
- 杭州火车站 驾车 | 公交 驾车距离: 大约6.1公里(约12分钟)
- 杭州东站 驾车 | 公交 驾车距离: 大约6.8公里(约11分钟)
- 杭州北站 驾车 | 公交 驾车距离: 大约11.5公里(约16分钟)
- 杭州南站火车站 驾车 | 公交 驾车距离: 大约23.3公里(约36分钟)
- 德清站 驾车 | 公交 驾车距离: 大约45.9公里(约43分钟)

您最近有出游计划吗? 将该酒店 收集到攻略

# 系统结构总览

- 系统的核心考量
  - 信息搜索准确和全面
  - 报价和房态的实时准确
  - 高可用性
  - 性能

# 系统结构总览

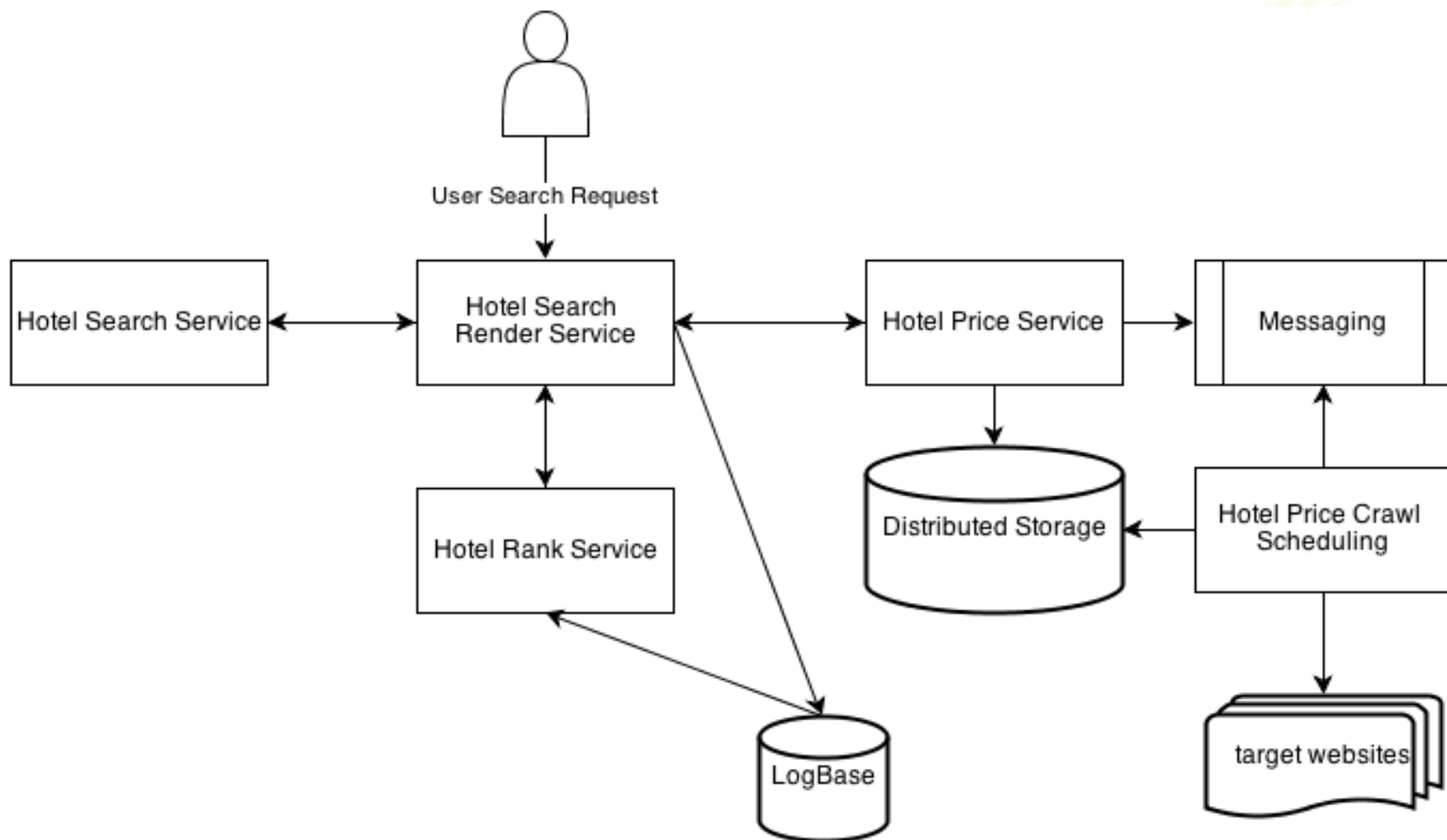
- 如何达到
  - 报价和房态的实时准确
    - 实时报价获取
  - 高可用性
    - 服务拆分
    - 监控和运维
  - 性能
    - 缓存设计
    - 监控和数字



# 系统结构总览

- 现有主要服务
  - HotelSearch Render Service
    - 负责主要页面数据获取和展现
  - HotelSearch Service
    - 负责关键字检索
  - HotelSearch Rank Service
    - rank计算服务
  - HotelSearch Price Service
    - 负责实时房价房态搜索和计算
  - HotelSearch Price Crawl Service
    - 实时房价房态抓取服务

# 系统结构总览



# 服务拆分

- 拆分原则
  - 功能内聚的独立业务模块
- 拆分目的
  - 降低维护成本
  - 提高系统整体可用性
    - 故障隔离
    - 服务降级

# 服务拆分

- 服务化模块间通讯
  - http
    - nginx + QunarClient
  - rpc
    - dubbo
- 可用性和负载均衡
  - rpc
    - zookeeper
  - http
    - ngx-healthcheck
    - 负载均衡
      - roundrobin
      - 按ip或cookie哈希
      - 按搜索条件哈希



# 实时报价搜索

- Why

- 酒店的价格和房态，尤其是房态变化快
- 价格和房态的准确性是保证用户的搜索体验的前提条件

- Challenges

- 保证用户请求的响应效率
- 保证用户看到最新的价格和房态
- 降低对目标网站的抓取量

# 实时报价搜索

- 核心组件
  - 报价服务
  - 抓取服务
  - 消息中间件
  - 分布式缓存

# 实时报价搜索

- 当前规模
  - 报价服务每秒请求量峰值近1600qps，全天超过1亿次动态请求
  - 报价服务平均请求响应时间4ms
  - 报价抓取服务平均每秒处理完成4000个抓取请求，全天近2.5亿次报价抓取



# 实时报价搜索

- 如何应对
  - 服务拆分
  - 异步化
  - 缓存设计
  - 降低抓取量



# 实时报价搜索

- 服务拆分
  - 报价计算服务和抓取服务分拆
    - HotelSearch Price Service
    - HotelSearch Price Crawl Service
- 拆分意义
  - 功能侧重点不同
    - 报价服务
      - 缓存+计算系统
      - 对外接口响应时间是关键
    - 抓取服务
      - 定向抓取
      - 整体吞吐量和获取率是关键
- 服务间通信
  - 生产者消费者
  - 异步通讯



# 实时报价搜索

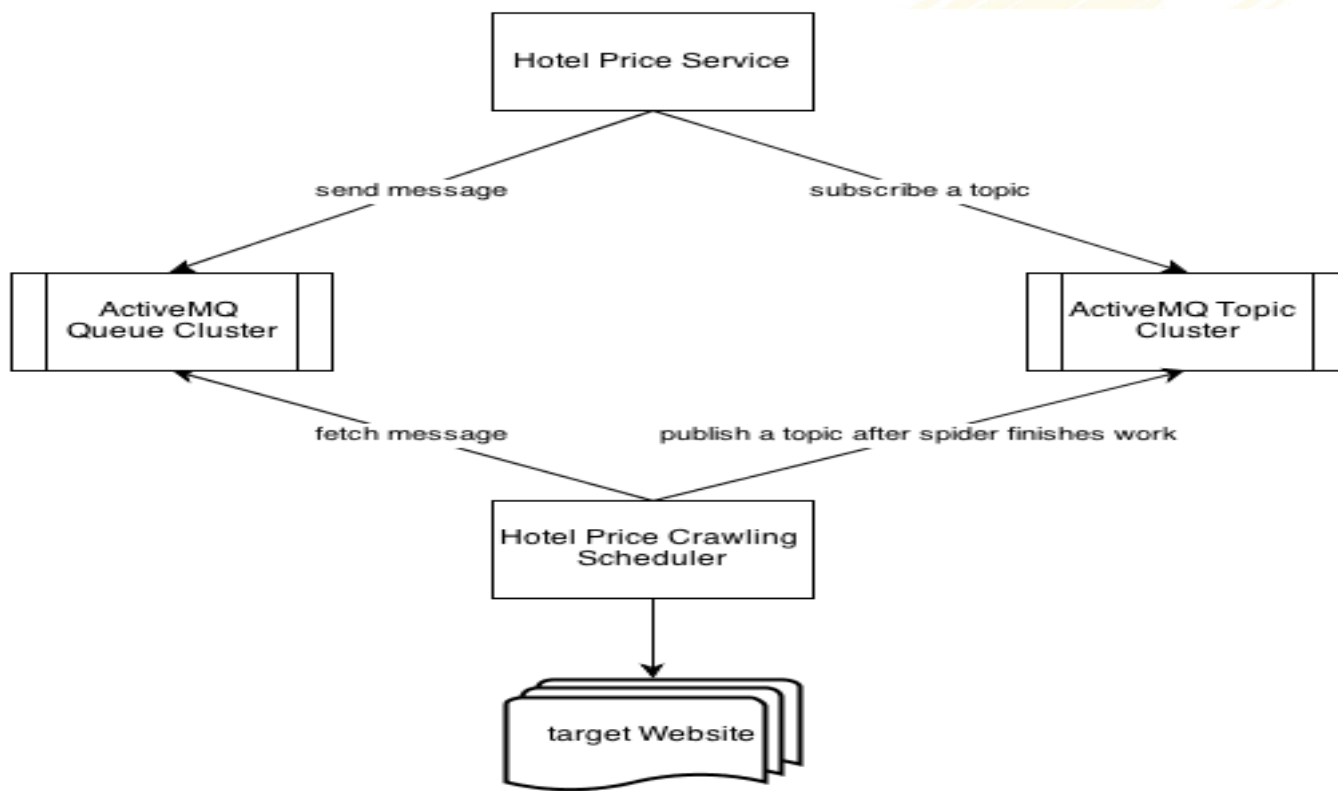
- 异步化
  - 消息中间件
  - 线程池隔离
  - 前端动态请求

# 实时报价搜索

- 消息中间件
  - 用于报价计算服务和抓取服务之间的通信
  - 方案选择
    - activemq
  - 基于业务模式进行配置
    - 不需要持久化
    - 过期消息可以丢弃
      - 直接丢弃
        - » `<sharedDeadLetterStrategy processExpired="false" />`
    - 提高吞吐量
      - async send & async ack
      - 接收到消息后异步处理
    - 可用性考量
      - client端连接串配置成failover协议

# 实时报价搜索

- 消息中间件





# 实时报价搜索

- 线程池隔离

- 目的

- 提高性能
      - 独立任务串行执行转并行执行
    - 提高整体可用性
      - 最大限度降低外部资源依赖失效的影响
      - 使得整体吞吐量和响应时间可控
    - 避免资源争用

- 关注点

- 线程池容量合理设置
    - 任务超时设置
    - 利用监控，持续运维

# 实时报价搜索

## • 应用场景

Utour 众信旅游

与最爱的宝贝相伴  
私享全家度假时光! 立即预订

62 家酒店满足条件

价格范围: 不限 ¥200以下 ¥200-¥300 ¥300-¥500 ¥500以上 自定义 到

酒店级别: 不限 经济型 二星级/其他 三星级/舒适 四星级/高档 五星级/豪华

连锁品牌: 不限 布丁 汉庭酒店 怡莱快捷 宜必思 速8 全季酒店 君亭

默认排序 星级 评分(5分) 价格 当前有房酒店 只看团购中酒店

酒店名称	经济型	评分	点评数	杭州好评第	价格	查看详情
1 汉庭酒店杭州学院路店	该酒店坐落于黄龙体育中心商圈内, 坐落于高新电子信息企业云集的西湖区学苑路上, 文三路和文二路中段。网友评价说“交通方便, 房间很小, 性价比高, 服务很好, 地理位置很好, 性价比不错”。	3.1分	221条点评	609	¥154起	查看详情
2 宜必思杭州文三路酒店	该酒店坐落于黄龙体育中心商圈内, 位于西湖区文三路和黄姑山路交汇处, 颐高旗舰广场旁边, 紧邻九莲新村。	4.5分	13条点评	18	¥150起	查看详情
3 杭州嘉海大酒店	该酒店坐落于黄龙体育中心商圈内, 位于杭城中心地段, 毗邻西湖风景区。网友评价说“环境不错”。	19条点评	暂无评分		¥388起	查看详情
4 杭州雅逸假日酒店	该酒店坐落于黄龙体育中心商圈内, 位于杭州市西湖区外东山路, 属于西湖景区核心区域, 紧邻西湖景区, 距浙江大学玉泉校区需步行3分钟。	2.9分	17条点评	626	¥109起	查看详情
5 布丁酒店杭州西湖店	该酒店坐落于黄龙体育中心商圈内, 坐落于杭州西湖区, 邻近植物园, 由植物园, 近西湖景区。	4分	550条点评	267	¥159起	查看详情

该商圈共399家酒店

西湖位于杭州城西, 包含了众多名胜古迹。它分为北山、南山和湖中三个游览区。北山有灵隐、岳庙等景点; 南山有虎跑、六和塔等景点。湖的一山、二堤、三岛闻名世界。

查看本区域全部酒店>>

酒店推广

大鹏网报价

我的订单 我的收藏 我的点评 我是行家 我的图片 我的回应 我的徽章 更多

# 实时报价搜索

- 应用场景

- 背景

- 用户提交一个酒店搜索，搜索列表页在得到符合条件的酒店列表后，需要远程调用报价服务的接口获取酒店报价
    - 报价服务的任何问题都不能影响到搜索结果列表页的展示
    - 如果结果列表酒店数量太多，报价服务串行获取报价效率较低，调用方等待时间较长

# 实时报价搜索

- 应用场景

- 解决方案

- 单独创建一个线程池来处理对报价服务的调用
    - 对于超过一定阈值的酒店数的报价查询请求，拆成多个任务执行

- 设计考量

- 线程池配置
    - 超时设置
      - 经验值，我们一般是选择被调用方平均响应时间\*10
    - 超时而处理
      - 任务需要能够响应中断
      - 超时要回收资源
      - 超时需要记录监控
      - 超时需要将调用条件记录到日志中，供线下分析



# 实时报价搜索

- 应用场景

- 线程池参数选择

- corePoolSize

- 该请求qps历史峰值\*增长系数（1.5，依据业务预期而定）\*调用外部接口任务超时时间（s）

- queue

- LinkedBlockingQueue，无界队列，资源不可控，不考虑
      - ArrayBlockingQueue，会产生任务等待，适合线下任务，只考虑吞吐量而不关注响应时间时使用
      - SynchronousQueue，Direct handoff，有任务立即尝试开线程执行，在关注响应时间时选择

- maxPoolSize

- $\max(\text{该请求qps历史峰值} * \text{增长系数}, \text{corePoolSize})$

- rejection policy

- 直接丢弃
      - 记录监控

# 实时报价搜索

- 前端js动态请求

- 目的

- 报价服务不会等所有代理商都抓回报价
    - 将最新报价及时展现给用户

- 设计考量

- 前端按照特定间隔poll报价服务接口
    - poll的间隔根据以下因素动态改变
      - 发送的报价抓取请求回数比例
      - 随时间衰减
      - 报价服务接口连续返回相同结果集的次数

# 实时报价搜索

- 缓存设计

- 进程内缓存的意义

- 提高整体性能
    - 减少外部资源依赖，提高可用性

# 实时报价搜索

- 缓存设计
  - 缓存分类
    - 基础信息缓存
      - 酒店的报价抓取元数据
      - 全量缓存
        - » 系统启动时加载
      - 增量更新
    - 报价缓存
      - 进程内缓存
        - » LRU
        - » 命中率是关键
        - » 多级策略
        - » 控制读写锁粒度
      - 分布式缓存
        - » memcached

# 实时报价搜索

- 构建内存缓存的基本原则
  - 关注单个缓存对象的大小
  - 要正确的使用数据类型
  - 值域固定的字符串对象常量化
    - `String.intern()`
    - 自定义String Pool
  - 要监控缓存命中率
  - 要估算总的内存开销
  - 缓存要有明确的容量上限
  - 对于LRU缓存，不同业务数据尽量不要共用同一LRU队列



# 实时报价搜索

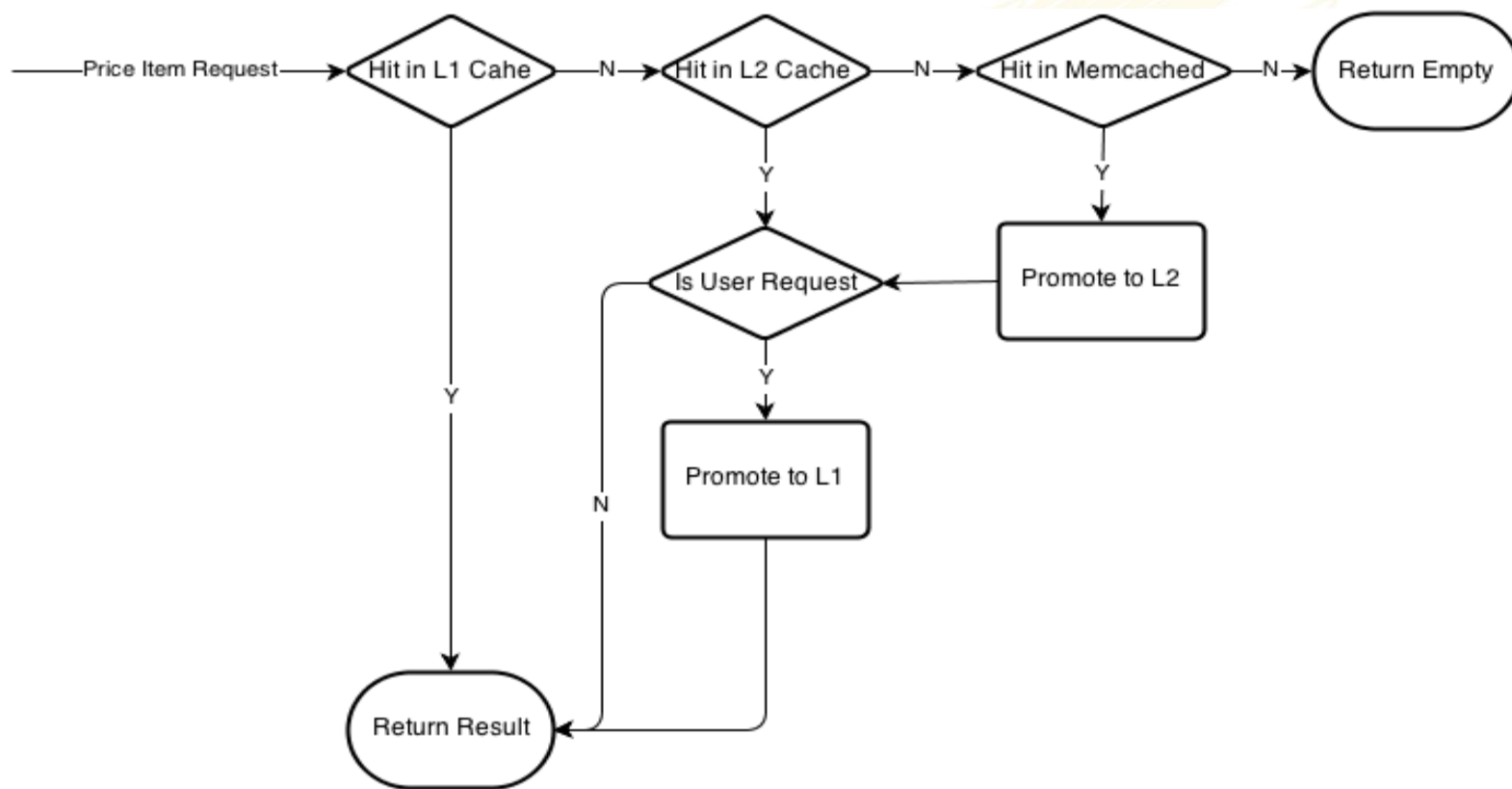
- 报价多级缓存

- 目的

- 确保用户报价请求，即一级缓存的命中率
    - 确保各代理商报价抓取结果合并时的缓存命中率
    - 减少和memcached的通信开销
    - 减少序列化反序列化的开销，控制负载

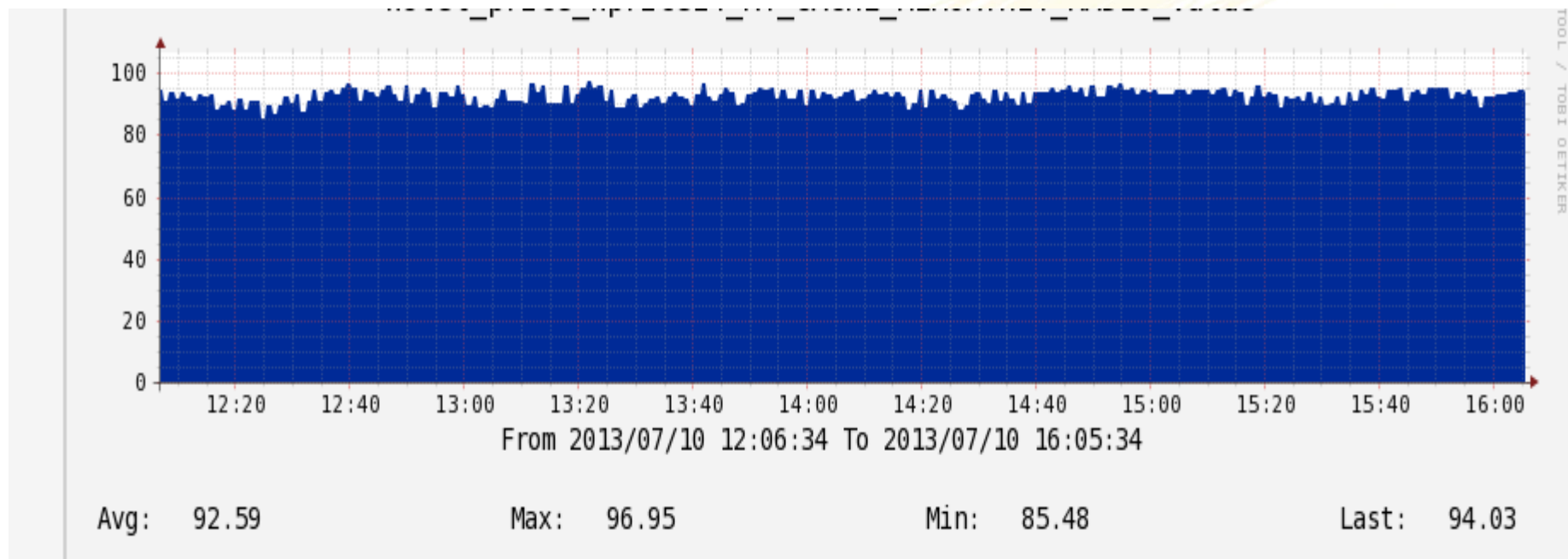
# 实时报价搜索

- 报价多级缓存



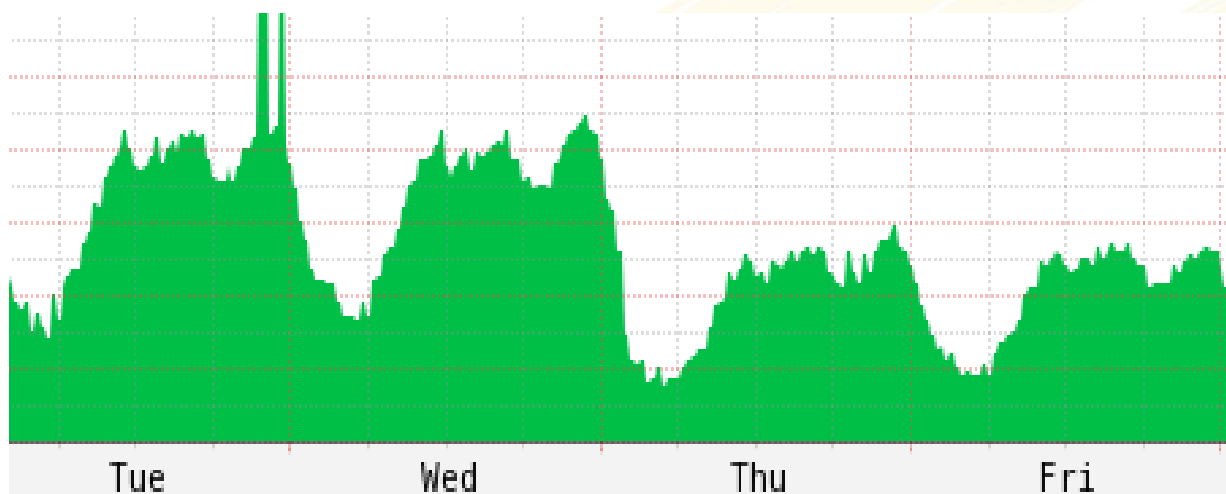
# 实时报价搜索

- 报价一级缓存命中率



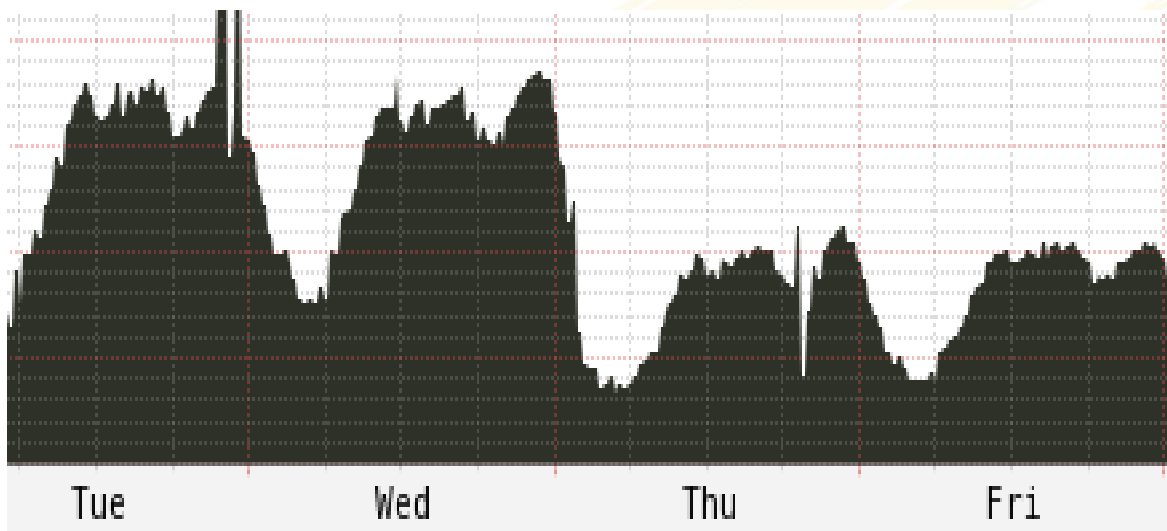
# 实时报价搜索

- 内存缓存换出到memcached的次数优化前后对比



# 实时报价搜索

- 从memcached换入到内存缓存的次数优化前后对比



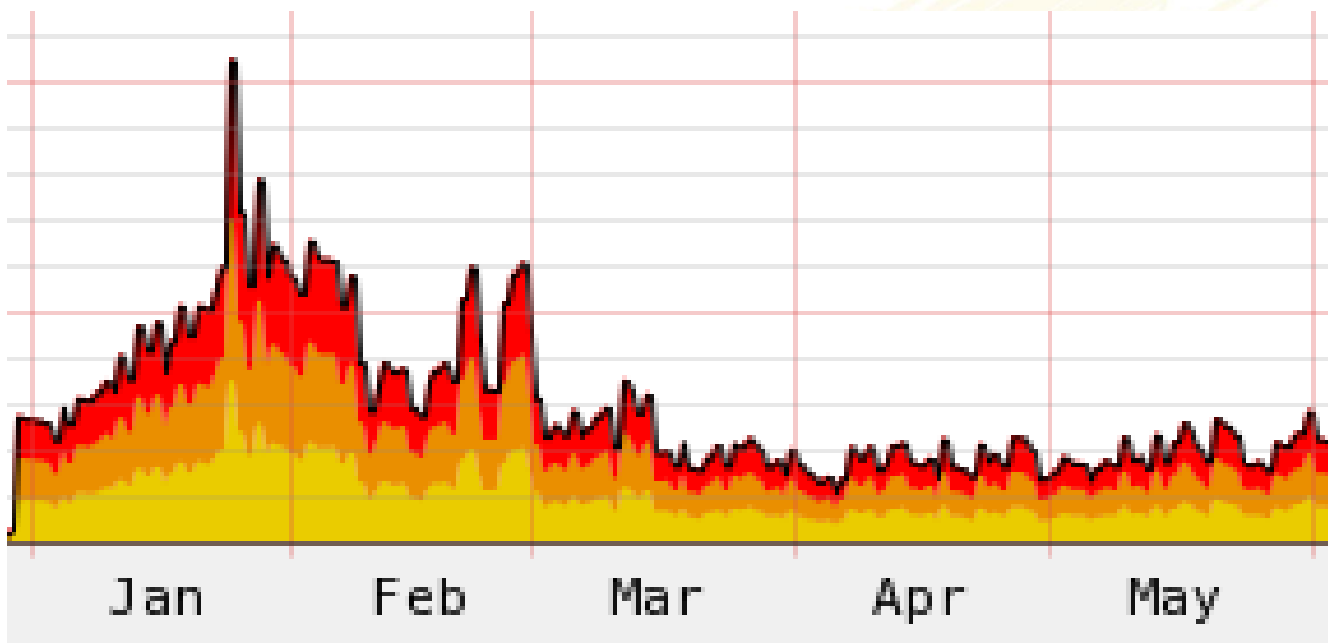


# 实时报价搜索

- Memcached的使用
  - Slab Allocation Mechanism
    - 防止类似大小的不同业务数据缓存争用同一slab，导致eviction
  - 尽可能提高缓存命中率
    - 保证memcached集群的总容量符合业务需要
    - 缓存失效会导致对代理商网站的大量抓取
    - 缓存一定要有合理的过期时间
    - 定期主动清除过期数据，减少eviction
      - qmemcached-patch
  - 使用高效的序列化反序列化工具
    - kryo
  - 确保可用性
    - 监控和报警
    - daemontools

# 实时报价搜索

- kryo替换java原生序列化/反序列化的效果



# 实时报价搜索

- 降低抓取量

- 目的

- 降低目标网站的负载
    - 降低双方的成本

- 手段

- 为每个目标网站设置cachetime
      - 依据当前抓取量监控
      - 依据目标网站房态房价变化频率监控
      - 人工配置+动态调整
    - 定时线下全量抓取配合线上用户触发抓取
    - 与合作客户配合进行变价推送

# 运维和监控

- 意义
  - 好的系统是运维出来的
  - 及时发现问题
  - 掌握系统运行状况
  - 找出系统优化方向
  - 为技术决策提供依据
- 分类
  - 实时监控和运维脚本
  - 监控系统

# 运维和监控

- 实时负载监控

```
14:52:18 up 155 days, 20:19, 1 user, load average: 1.02, 0.97, 0.92
14:52:18 up 155 days, 20:16, 0 users, load average: 0.90, 1.32, 1.34
14:52:18 up 155 days, 20:16, 0 users, load average: 1.30, 1.49, 1.50
14:52:18 up 155 days, 20:21, 0 users, load average: 1.68, 1.46, 1.39
14:52:18 up 155 days, 20:21, 0 users, load average: 0.84, 0.94, 0.91
14:52:18 up 155 days, 20:20, 0 users, load average: 1.40, 1.14, 1.03
14:52:18 up 155 days, 20:19, 0 users, load average: 1.04, 1.13, 0.98
14:52:18 up 155 days, 20:20, 0 users, load average: 3.87, 2.82, 2.59
14:52:18 up 155 days, 20:17, 0 users, load average: 3.54, 2.92, 2.55
14:52:18 up 155 days, 19:50, 0 users, load average: 0.78, 1.14, 1.12
14:52:18 up 155 days, 20:18, 0 users, load average: 1.23, 1.06, 1.00
14:52:18 up 155 days, 20:17, 0 users, load average: 1.77, 1.40, 1.23
14:52:18 up 155 days, 19:48, 0 users, load average: 0.93, 1.13, 1.11
14:52:18 up 155 days, 20:19, 0 users, load average: 2.02, 1.72, 1.50
14:52:18 up 155 days, 20:17, 0 users, load average: 1.49, 1.79, 1.94
```



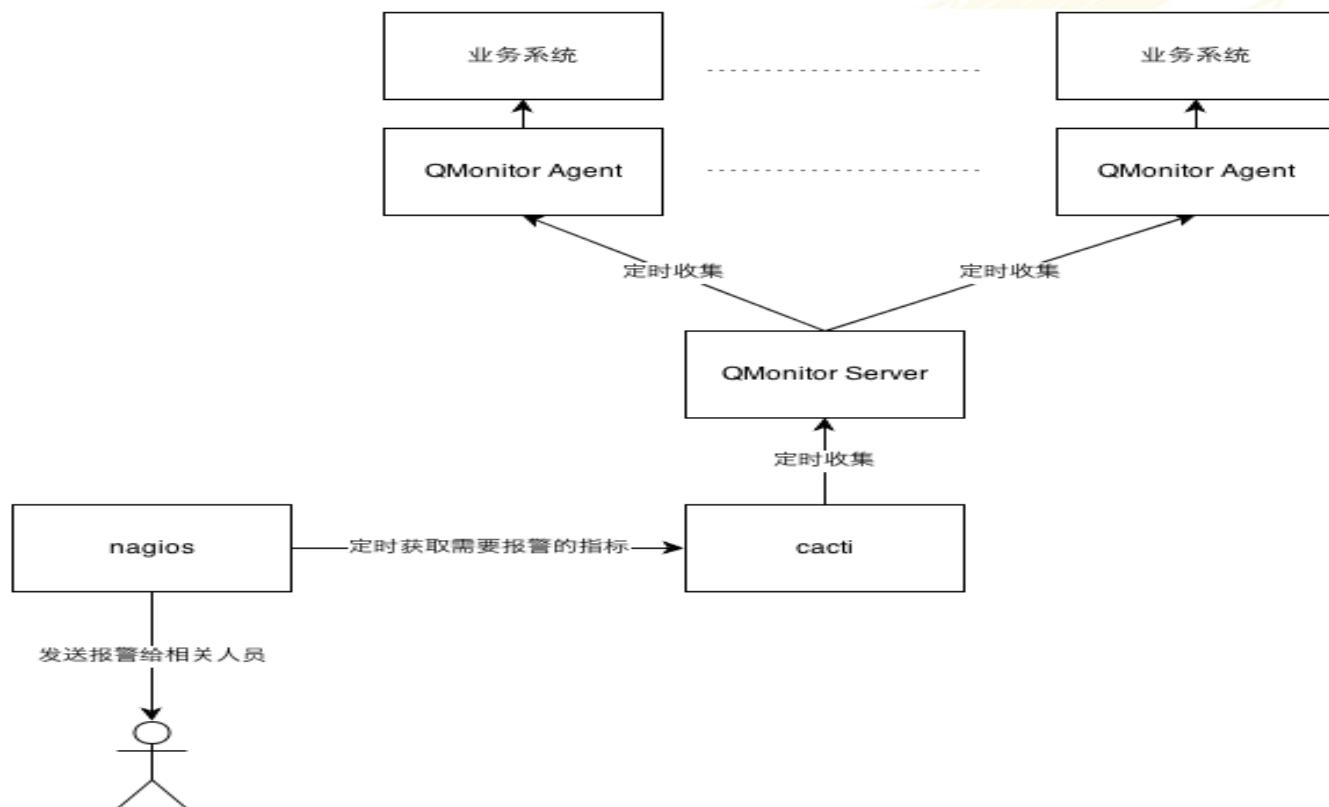
# 运维和监控

- 实时日志监控

2012-9-22 21318 com.qunar.hotel.roomType.SimpleRoomTypeRawTreeService.run(SimpleRoomTypeRawTreeService.java:137) INFO update tree seq:lijiang_3092 2012-9-22 21319 com.qunar.hotel.roomType.SimpleRoomTypeRawTreeService.run(SimpleRoomTypeRawTreeService.java:137) INFO update tree seq:dalian_4410	ntry WARNING: Timed out waiting for operation: htc/2012-09-24/wapstar0000/~s00jhkg854mONXhcgGTFBnm~ 2012-9-22 21318 com.qunar.hotel.util.RoomTypeCache.get(RoomTypeCache.java:78) INFO room cache hits with key:hangzhou_8138	2012-09-22 21:31:18 RoomTypeCache INFO Sep 22, 2012 2:13:18 AM ge INFO
awTreeService.run(SimpleRoomTypeRawTreeService.java:137) INFO update tree seq:guangzhou_9271 Sep 22, 2012 2:13:18 AM qunar.concurrent.ThreadRecycles clear WARNING: 资源回收器捕获到未回收的资源: Thread[http-8080-29;5,main] qunar.agile.TraceLogger@1a364245	2012-9-22 21317 com.qunar.hotel.util.RoomTypeCache.get(RoomTypeCache.java:78) INFO room cache hits with key:yuancheng_2027 2012-9-22 21317 com.qunar.hotel.util.RoomTypeCache.get(RoomTypeCache.java:78) INFO room cache hits with key:sanya_2129	) INFO 2012-09-22 21:31:17 Raw ) INFO
Sep 22, 2012 2:13:19 AM com.qunar.hotel.wrappers.WrapperCredibility getWrapperPriceWarning WARNING: 该房型下无可选价格均值: 三居室公寓房 Sep 22, 2012 2:13:19 AM com.qunar.hotel.wrappers.WrapperCredibility getWrapperPriceWarning WARNING: 该房型下无可选价格均值: 二居室公寓房	2012-9-22 21318 com.qunar.hotel.util.RoomTypeCache.get(RoomTypeCache.java:78) INFO room cache hits with key:chengdu_1894 2012-9-22 21318 com.qunar.hotel.util.RoomTypeCache.get(RoomTypeCache.java:73) INFO room cache missed, generate new key:guilin_9	2012-09-22 21:31:18 RoomTypeCache INFO 2012-09-22 21:31:18 RoomTypeCache INFO
etInstance(CacheTimeQueue.java:57)	tInstance(CacheTimeQueue.java:57)	get

# 运维和监控

- 监控系统





# Thank you for coming

mail: [yue.liu@qunar.com](mailto:yue.liu@qunar.com)  
gtalk: [the.6th.month@gmail.com](mailto:the.6th.month@gmail.com)