# 刘少华

学历: 统招一本 毕业院校:武汉科技大学

工作经验:两年 目前状态:在职

个人博客: https://github.com/linghuazaii/blog

个人GitHub: https://github.com/linghuazaii?utf8=√&tab=repositories&q=&type=source&language=

联系方式: +86 15910693294

个人邮箱: charlesliu.cn.bj@gmail.com

求职意向:中级软件开发工程师

## 技能树

• Linux C/C++ (熟练)

- TCP / IP, 网络编程, 高性能Server (熟悉)
- Mutil-threading, concurrency (熟悉)
- Python (熟悉)
  - Django (了解)
- Http (了解)
- PHP (了解)
- 开源的项目(了解,看过文档,用过,可聊,但是无法深入到代码层面):
  - Memcached
  - Kafka
  - Zookeeper
  - o Redis
  - Hbase
  - Thrift
- Web后端架构设计的一些思想
  - 。 分布式 (了解)
  - RPC (了解)
  - 。 服务发现 (了解)

# 工作经历

# 安远国际 (Kloudpeak)

安远国际是一家做聚合新闻的海外今日头条,app名称Gündem,主要市场在土耳其,其他市场包括德国、印尼以及巴西。目前的的变现模式主要是广告变现的方式,另外为其他创业型想做内容的公司提供后端的技术支持,后端的整体架构不需要变动即可切换语言,更改新闻抓取源,也提供Client端的技术支持。同时还有直播,音乐等试水产品。

# 2016/5 - 至今 后端开发 (时间线回顾,倒叙)

- 负责个性化推荐结构化日志等收集,以及后期的GBDT预测
  - 。 初版用MongoDB, MongoDB混合索引占用内存过高,后转用Hbase
  - 。 Hbase结构多次调整,先阶段结构化日志先写到Kafka,Client端日志也是写到Kafka,便于统一处理。
  - 。 负责Kafka到Hbase中间件的编写
  - 。 负责GBDT预测部分 (xgboost)

#### • 后端推荐的性能优化

- o 内存占用过高:推荐服务用两块内存来存储数据库里的新闻Item,没两分钟更新一次做指针置换,总计60000多条数据。推荐服务占用内存6G,远远超过预估值,问题出在 safe\_strncpy 函数传递的Length不对,导致 strncpy 写整块内存,引发PAGE FAULT,进而PAGE SWAP,glibc所有的string函数都会 memset 整块内存。修复后,RES占用1.5G。
- 推荐服务耗时降低:
  - 历史过滤优化:原先历史过滤N个Item,会一直在History\_Map里查找是否存在历史,时间复杂度O(NlogM)。由于两个数据集合本身以续,采用增量比较的方式,时间复杂度O(N+M)。
    - + 日志系统更改:原先的日志系统,日志IO会阻塞业务,当日志量过大的时候,会影响QPS,新的日志系统采用异步设计,日志会先写到队列,一个工作在 SCHED\_IDLE 模式的后台线程会将队列里的日志写到文件里去,最大限度的减小对系统性能的影响。示例代码见:https://github.com/linghuazaii/Charles-Logging. + 日志系统更改:原先的日志系统,日志IO会阻塞业务,当日志量过大的时候,会影响QPS,新的日志系统采用异步设计,日志会先写到队列,一个工作在 SCHED\_IDLE 模式的后台线程会将队列里的日志写到文件里去,最大限度的减小对系统性能的影响。示例代码见:https://github.com/linghuazaii/Charles-Logging

### • 新闻搜索的搭建 (python)

- 。 以增量的方式将数据库的新闻Load到ElasticSearch, 建crontab每日更新
- 。 以MultiMatch的方式搜索ES, title和content匹配赋不同的权重
- 分布式Http Downloader
  - 预想是做一个高性能的下载功能,为爬虫提供服务,稍微重构即可做线上的分布式图片处理服务器
  - 整个项目分为两个模块: DownloadProxy和DownloadServer
  - 用libmicrohttpd做简易的HTTP Server,用libcurl做现在功能,根据官方文档,配置调到性能最优,禁用DNS,自己用Redis做全局的DNS cache
  - 用Zookeeper做服务发现和集群, Server动态刷新zk更新自己的任务负载, Proxy可以根据Server复杂选择最优的Server 分发请求
  - 。 各种对象池,提升性能
- Http Proxy
  - o 预想是取代AWS ELB和openresty的lua来节省服务器费用,后测试性能不如openresty lua暂时搁置
  - 。 用libmicrohttpd做建议的HTTP Server来处理请求和HTTP包的解析
  - Client和Http Proxy之间加上建议的TCP连接池, P/C模式
  - 用Zookeeper做Server集群和服务发现
  - 。 用BKDR Hash做一个简单的分发机制
  - 。 插件化的设计,后端业务接入可以动态调整
- Kafka Proxy
  - 。 Kafka Proxy的主要目的是为所有业务提供一个统一的HTTP接口,包括python的Django后端,Client端的Android和iOS以及C++版的推荐和相关新闻
  - 。 同样libmicrohttpd做简易的HTTP Server,解析HTTP包
  - 做了一个简易的工厂模式的架子,很容易扩充
- 其他的一些杂项
  - Bug以及一些小feature,太多太琐碎,无法——罗列

## Websense (韦伯森斯网络安全有限公司)

Websense是一家致力于Cyber Security的安全公司,主要致力于Web Security,Data Security,Mobile Security和Mail Security,后被Raytheon 19亿美元收购。

## 2014 / 7 - 2015/5 后端开发

- 负责解决产品Bug和稳定性问题,包括一些琐碎的bug,字符集,UI,内存泄漏等
- 负责参与组内测试框架等搭建,GTEST/GMOCK

## 杂项

• 英语: CET-6

• 性格:偏内向,但是也比较容易相处

• 爱好:骑行,拍拍东西,研究其他一些东西,但是都不深,时间有限