

个人信息

年龄：24

电话：^_^

邮箱：randomTT@gmail.com

Github: <https://github.com/chaseSpace?tab=repositories>

Csdn 博客: http://blog.csdn.net/sc_lilei

工作期望：首先考虑重庆，其次成都

自我评价：经常关注走在前沿的各种技术和框架，闲暇时愿花时间折腾。有自己的技术博客，记录填过的坑和总结。

教育经历

专业：计算机网络技术

时间：2013.9-2016.6

院校：南宁职业技术学院（广西） 大专

技能掌握

Python:

掌握 python 基础语法，熟悉面向对象编程的概念，熟悉 python 语言特性。对 python 多线程、多进程有实践经验。了解 python 的内存管理机制以及如何写出优质、高效代码；熟悉 standard library 使用

数据抓取:

熟悉网页爬取原理，熟悉深度抓取、动态网页抓取、请求伪装、模拟登陆、代理反爬等技术及相应的经验总结；对优化爬取效率和稳定性有个人理解；熟悉常见反爬机制

爬虫框架:

熟悉 Scrapy 框架原理及其特性，熟练使用 scrapy-redis 搭建分布式爬虫；了解 scrapy 去重原理及其优缺点熟悉 BloomFilter 去重原理，并结合个人理解将其嵌入 Scrapy 框架以适应海量数据（上亿条）去重场景，极大地降低了大规模爬取场景时 Redis 服务器的内存消耗。有分布式爬虫实践经验，并将 scrapy 与 python 的 multiprocessing 进行结合使用，在真实场景下充分利用单台物理主机硬件资源（操作系统稳定下，内存占用可达 80%，爬取效率提高几十倍，依赖主机内存资源）。

数据库:

MySQL: 熟悉 SQL 语句，熟悉 MySQL 特性及优缺点，能够优化表结构，熟悉 MySQL 性能的检索优化。

Redis: 熟悉 Redis 常用操作语句；熟悉其优势及存储特性；熟悉 Redis 基于 Sentinel 的高可用集群原理及搭建。

MongoDB: 熟悉 MongoDB 特点，熟悉常用操作语句 insert/find/update/save/remove;熟悉 Replicat set 及 Sharding 的原理和搭建。

Continue ↓

工作经历

广西开云科技有限公司

2016.2-2018.2

职位：网络工程师

工作内容：与客户沟通项目实施方案，现场实施项目，售后技术支持

主要涉及的技术：Linux，路由交换、操作系统与存储

自学编程

2017.2-至今

个人实践

12306 抢票

- 通过 cookie 模拟登陆
- 通过 Fiddler 分析从查票到订票完成整个过程中的 http 请求
- 模拟请求自动完成刷票、订票

云起书院全站小说简介信息爬取

- 使用两台电脑+多进程分布式部署 Scrapy 进行爬取，20 个小时爬完 30 万条小说信息
- 支持断点续爬，使用 BloomFilter 进行去重，redis 服务内存消耗极少