

罗纲

电话：18377336165

邮箱：1148728004@qq.com



教育背景

(2013-2017) 桂林理工大学 本科毕业生 应用物理学(统计计算方向)

- **主修课程：**java，Python，C 语言，通信原理，计算机原理，统计学原理，概率论与数理统计，线性代数，计算物理，数理方程，高等数学。
- **工作技能及经历：**有淘宝，腾讯，Baidu，jd 等网站的数据抓取经历，熟悉分布式爬虫。熟练使用 Scrapy 和 Urllib 以及 selenium+phantomjs，IP 代理池，IP 代理接口，用户代理池，模拟验证码登录等防反爬技术。熟练使用正则表达式，xpath，fiddler 抓包等，熟悉 redis，MySQL 数据库。熟悉 Docker，Hadoop 集群的部署及 elasticsearch 的使用。

工作经历

2016.06-2016.12 Python 爬虫工程师 北京联云天下科技有限公司 北京市

使用语言：Python, Linux shell

开发环境：centos6.5

使用工具：MySQL, fiddler, SVN, pycharm

使用的相关模块和框架: Scrapy, Urllib。

工作内容: 使用 IP 代理，模仿浏览器登录，cookie 登录等技术。通过正则表达式，xpath，fiddler 抓包进行分析网页获取淘宝、当当网、IT 桔子等相应公司网站的数据，编写 python 脚本进行数据的清洗，统计分析并编写数据分析文档，分析数据字段的信息设计相应的 mysql 数据库表，将获得的有效数据导入 MySQL 数据库。进一步检查数据表的合理性并优化，编写数据挖掘文档。独自在 centos 下建立可远程登录的 redis 服务器，使用 python 处理 redis 数据库的数据并导入 mysql 数据库。在 Python 端处理好数据格式，满足写入数据库使用数据类型的需求。编写 redis 安装文档并使用安装文档来安装生产环境的 redis 数据库，封装 python 调用 mysql 数据库的函数。

2017.01-2017.05 数据开发工程师 360 企业安全集团 北京市

工作环境：Centos7.3

使用的技术：hadoop，docker，elasticsearch

使用语言：JAVA，C 语言，Linux shell

工作内容：在 Centos7.3 下部署 hadoop2.6.0 完全分布式集群，并使用 hadoop 提供的与 C 语言连接的 API 导入数据，调研 hdfs 的写入和读取性能，编写 mapreduce 程序分析相关数据。

Docker 容器中部署 hadoop2.6.0+hbase+spark+ookeeper 集群，应公司需求把 docker 下的集群封装入系统镜像，应用 docker 对生产环境的隔离优势，干净的开发设计，满足公司对数据并发，聚合，搜索和分析所需环境的需求，并方便后续的软件生态部署。在 Centos7 下安装 elasticsearch 搜索服务器，通过 JAVA 的 API 在 Elasticsearch 服务器中批量导入数据和建立索引，优化建立的索引和 elasticsearch 搜索服务器副本设计和存储的策略。

技能爱好

• 专业技能：

Python 爬虫：熟练使用 Scrapy 和 Urllib 以及 selenium+phantomjs，IP 代理池，IP 代理接口，用户代理池，模拟验证码登录等防反爬技术。熟练使用正则表达式，xpath，fiddler 抓包等，熟悉分布式爬虫的部署。

Java：熟悉 java 编程，熟悉 jar 包的使用和工作原理。

Linux shell：精通 Linux 基本命令，能编写常用的 Linux shell 脚本。

Redis：熟悉 redis 服务器的配置和使用，曾在工作中管理公司 Oracle 数据和 redis 数据库之间的数据操作，应工作需求设计 redis 的各项参数满足功能需求。

Mysql：熟悉 mysql 数据库相关操作，以及 mysql 的 java，python 等接口的使用。

Hadoop：熟悉 hadoop2.x 实现的技术原理，理解相关 hadoop 生态的组件实现，对 mapreduce 实现的原理有一定的理解，有 mapreduce 使用的经历。

Docker：在 centos7 下熟练编写 dockerfile 制作满足需求的镜像，理解 docker 的镜像使用的优势和相关的应用场景，能部署相关的 docker 应用环境。

• 业余爱好：

篮球，游泳，网球，骑行

个人评价

IT 技术男，有责任感，能担当复杂的开发任务，团队意识强，善于沟通，擅长数据处理方面的技术，乐于学习新技术，喜欢从事数据采集和数据挖掘相关的工作。喜欢和同事一起交流讨论研发技术及问题，一起喝可乐一起加班。