Python

熟悉python面向对象，元编程

熟练掌握python装饰器

熟悉python多线程，多进程，协程开发，熟悉select，epoll

熟悉掌握python网络编程

熟悉python界面编程

熟悉python office办公自动化，邮件操作

Linux

熟悉linux 平台软件开发

熟悉Linux 指令，awk指令

熟悉shell脚本，python脚本

熟悉Linux网络编程

掌握部署维护 nginx，samba ，ftp等服务器

数据库

熟悉mysql、oracle数据库，熟练使用sql语句

熟悉sql联合查询，分组，having语句

熟悉索引的维护调优，慢查询分析

熟悉redis，mongdb等非关系型数据库

爬虫

熟练掌握网页抓取原理及解析技术，对SQL优化有一定的经验

熟悉request，urllib库

熟悉基于正则表达式、XPath、CSS等网页信息抽取技术

熟悉MySQL、Redis、MongoDB等数据库

熟悉http协议，各种web认证机制

熟悉常见的爬虫反爬虫策略

熟悉scrapy/pyspider等爬虫框架

数据分析

熟练使用numpy数学运算

熟悉掌握pandas数据分析

熟练使用matplotlib数据可视化

熟悉scipy数据统计包

熟练使用scikit-learn进行数据挖掘

熟练掌握逻辑斯蒂回归，决策树，支持向量机等机器学习算法

熟悉朴素贝叶斯，聚类

了解神经网络，深度学习原理

熟悉tensorflow等开源深度学习框架

熟悉Hadoop、Spark大数据框架

熟练掌握mapreduce编程

熟悉HDFS分布式存储

使用Hadoop 制作电影推荐引擎

熟悉python Spark大数据分析

熟练使用Spark推荐系统

熟练使用Spark机器学习