

计算机科学基础

- 计算机理论
- 计算机系统
- 计算机网络
- 软件工程
- 线性代数
- 离散数学
- 统计学
- 微积分

编程基础

- C/C++
- Java
- Python

算法基础

- 排序算法
- 图搜索
- 字符串匹配
- 数据结构
- 算法分析

上层应用

- 大型软件开发
- 框架开发
- 高性能计算
- 机器学习
- 计算机图形学
- 网络安全
- 等等

求职流程

投递简历

网上笔试

(基础编程问题和算法题)

网上/线下面试

(岗位相关的理论问题)

二次面试

(可能会手撕代码或者案例分析不同公司不一样)

(三次面试)

入职

Skill1. Sql 数据库取数，hive/spark/mysql类似，会一个就行，通用代码逻辑

SQL : Structured Query Language

MySQL: 开源RDBMS

Hive (SQL) 作用于Hadoop

Spark (SQL) 作用于Spark

简单来说，单学一个数据库取数建库增删查改个人认为先从基础sql开始学起，后两者作用于分布式框架用于处理更复杂的场景，优先学习使用环境必然是MySQL

关系型数据库中阶

设计数据库

高效增删查改

事务 回复 (recover) 取消 (abort)

并发

关系型数据库高阶

从抽象的程度进一步改进设计以提高效率

考虑底层硬盘对数据库query的效率影响等等

2. Python/R 数据分析，建模，可视化

Python

R

这俩定位差不多，个人建议二选一（推荐py），先从一个开始，主要的学习流程估计会先从包用法开始学起，然后学完用法后做一些数据分析练习题

Python 主要熟悉 numpy pandas matplotlib scikit-learn等等

3. tableau/powerBI可视化工具

这个我没有用过，但我查了下教程

简单来说，就是用来打印数据 让数据更客观

这个部分优先级中等偏下（因为基本上都是图形化操作界面（为的就是好操作易上手），丢进去就可视，事实上你用matplotlib 也能做可视化 就是画的没他好看）

4. Excel(函数，透视表，图表，VBA)

函数 透视表 图表 同3

VBA 个人认为是一项陈旧的技术，局限性极大。VBA最早是为Microsoft Office开发的宏编程语言。它允许用户编写自定义宏和脚本。其次，VBA只能用于excel上，excel在数据处理有个致命问题（而且不能跨平台）即当数据量大了后卡的不行，现在python基本上替代了VBA的工作。学习优先级不高

5. 统计学/机器学习算法，如随机森林，KNN，逻辑回归，决策树，聚类，朴素贝叶斯算法，SVM等

统计学

经典机器学习算法：随机森林，KNN，逻辑回归，决策树，聚类，朴素贝叶斯算法，SVM

这里学习路线上没有捷径，只能一个一个来，没什么好说的。

理想情况下，学习数学和算法设计流程以及应用场景->简单动手写一写->简单分析->学习调包->根据面经继续学习

现代机器学习算法：简单讨论一下神经网络等