**写在最前**

**请在收到HR发给您的面试题后，自行创建一个以自己姓名命名文件夹，将以下题目的py文件或txt文件放入其中标明题号，做完试题后将文件夹压缩之后发送给HR。**

**1.基础爬虫题:**

**目标网址：https://www.3839.com/top/hot.html**

**项目需求：爬取人气榜单所有游戏信息，游戏评价**

**题目要求：使用多线程、生产者消费者模式爬取,不得使用任何爬虫框架**

**游戏信息字段：**

**游戏id game\_id 83294**

**游戏名 game\_name 我的世界**

**游戏logo logo https://fs.img4399.com/sykb~sykb/20200116/09145170713**

**游戏介绍 introduce 原生游戏全面更新，冒险主题全面开启！**

**评分 score 8.3**

**评价数 comment\_count 13567**

**评论信息字段：**

**评论id omment\_id 244841561**

**用户id user\_id 19328128**

**用户名 username Ieo**

**用户头像 portrait https://imga.3839.com/19328128**

**评论时间 creat\_time 2020-03-01 09:32**

**评论内容 content 这四星都是给mc的，并不是给网易.......... 点赞数 like\_count 61**

**回复数 reply\_count 13**

**2.APP爬虫：**

**首先，详细的阐述一下对app爬虫的整个流程**

**然后，如果抓不到数据，该怎么处理？**

**3.算法题：**

**使用python实现冒泡排序和选择排序，并简单叙述**

**4.简答题:**

**1. 简述scrapy的运行机制**

**2. scrapy中如何自定义一个 Downloader Middleware**

**3. scrapy去重原理和scrapy-redis去重原理**

**4. scrapy和scrapy-redis有什么区别？**

**5. 简述你遇到过的反爬机制，你遇到过最棘手的一个是什么，最后如何解决的？**