##### 数据分析师 Business Analyst

岗位描述Job Description

在这里，你可以与业务线同学一起搭建业务监控指标体系并产品化；   
在这里，你可以基于业务监控指标体系，及时发现业务问题，进行异动指标分析或深度的专题分析定位问题并寻求解决方案；   
在这里，你可以为业务团队提供专题分析、数据分析与挖掘、模型及算法等相关服务；   
在这里，你需要主动的寻找机会获得资源并落地；提炼数据产品需求，提供数据产品解决方案，并最终推动数据产品落地；   
在这里，你需要组织数据技术与产品相关的理念、技能、工具的培训。

岗位要求Qualifications

如果你，熟练掌握excel、sql、ppt；   
如果你，拥有良好的沟通表达能力；   
如果你，所学专业是数学、统计、运筹学或其他专业；   
如果你对行为分析感兴趣，喜欢从数据中发现规律，想踏入大数据的门槛，那就加入数据分析师的行列吧！

工作地点Location

北京市(Beijing),杭州市(Hangzhou)

参加面试的城市或地区Interview City or Region

远程(Remote Interviews)

电面第一轮，（四轮面试，不得不说考核的真的很细致）

面试地点：阿里巴巴-杭州

我只收到了一轮电面。应聘数据分析师。  
1.自我介绍  
我噼里啪啦讲了一堆自己的专业，项目和性格，感觉没有很好的把自己的特征和岗位结合起来介绍，有点失败啦。  
2.专业技能例如会不会SQL之类，有没有做过分析的项目  
回答的是学习过SQL，但是项目中没有很多的数据，不太需要用的。项目里面用的都是SPSS之类统计分析软件、  
3.项目里面有没有用到SQL  
回答的是没有用过。

**面试官的问题：**

问面试的姐姐问了几个问题： 1.数据分析师和数据分析工程师这两个工作岗位在你的理解看来有什么不同？（问的第一个问题，我回答的不好呢，其实往往第一个问题很重要，大家以后要注意） 2.映象比较深刻的有哪些项目，为什么？ 3.如果让你去评判忠诚度，你会怎么建立评价的指标？

答1.我说的是数据分析师是对数据分析，需要对业务产品有很深的了解。数据分析工程师应该是数据分析过程中的一些模型的改进吧，还有数据维护，数据管理之类的。（好像回答的是错的。。） 2.自己的某项目 3.大家自己想吧。

简历中你参加过某某竞赛，你能具体讲一讲吗？

聊聊你之前做的项目？

你写过什么数据分析的报告？

问了一些基础的数据库知识。

你觉得你凭什么能够胜任这个职位？

SQL你知道怎么实现多表之间的链接？R怎么实现这一应用，你做过哪些数据分析的项目？时间序列项目有做过吗？

简历中你参加过某某竞赛，你能具体讲一讲吗？

面试阿里数据研发岗，无缘三面

导语：先随便说说自己的情况，一个研一学生，之前有过两年的工作经验，研究生方向是模式识别，目前做图像处理、识别，基础知识嘛，动手编程能力尚可，毕竟做过几年码农，之前做java的，理论基础相对来说薄弱的一塌糊涂！好，废话不多说，下面谈谈这次面试的经历，仅供大家参考：

三月一号，阿里的内推正式开启，刚好研一的我其实本来是没打算投的，因为学的并不多，但朋友太过热心，帮我内推数据研发岗，于是就抱着试试的态度，接着就是在等待中等待，在等待中温习。。

三月十一号，晚上十点左右，接到阿里技术中心那边的电话，第一面是技术面，问题很重理论，面试官很有素养，问的问题涉及面也挺广，问题主要如下，答案我就不一一给出了，给出的也是仅供参考，网上大概都能搜到：

第一个问题问的是MapReduce二次排序，主要讲怎么实现二次排序，由于还没看过Mapreduce相关东西，于是第一问就被阉割了大哭

\*\*答：\*\*先了解下二次排序吧，在MapReduce操作时，我们知道传递的<key,value>会按照key的大小进行排序，最后输出的结果是按照key排过序的。有的时候我们在key排序的基础上，对value也进行排序。这种需求就是二次排序。数据处理分为四个阶段：（1）Mapper任务会接收输入分片，然后不断的调用map函数，对记录进行处理。处理完毕后，转换为新的<key,value>输出。（2）对map函数输出的<key, value>调用分区函数，对数据进行分区。不同分区的数据会被送到不同的Reducer任务中。（3）对于不同分区的数据，会按照key进行排序，这里的key必须实现WritableComparable接口。该接口实现了Comparable接口，因此可以进行比较排序。（4）对于排序后的<key,value>，会按照key进行分组。如果key相同，那么相同key的<key,value>就被分到一个组中。最终，每个分组会调用一次reduce函数。（5）排序、分组后的数据会被送到Reducer节点。在MapReduce的体系结构中，我们没有看到对value的排序操作。怎么实现对value的排序哪？这就需要我们变通的去实现这个需求。

\*\*变通手段：\*\*我们可以把key和value联合起来作为新的key，记作newkey。这时，newkey含有两个字段，假设分别是k,v。这里的k和v是原来的key和value。原来的value还是不变。这样，value就同时在newkey和value的位置。我们再实现newkey的比较规则，先按照key排序，在key相同的基础上再按照value排序。在分组时，再按照原来的key进行分组，就不会影响原有的分组逻辑了。最后在输出的时候，只把原有的key、value输出，就可以变通的实现了二次排序的需求。

第二个问题相对简单，就是让我说说数据挖掘十大算法，再就是就决策树的C4.5算法以及其核心ID3算法做具体的解释，也就是看自己了解多少了，然后是数据挖掘十大算法的svm,k-means算法，要求能够分析算法原理，熟悉基本应用，实现并阐述！

接下来第三个问题问到了hbase，只要大致了解就行！

除此之外，第四个问题问了事务处理，要求说什么时候用到事务，平时我们一般怎么控制事务，一般借助什么控制，优点！这个问题对于我相对来说比上面几个问题都容易，当时我主要阐述了sql事物以及java事物处理、常见的事务处理手段，这里有篇文章总结的不错，大家可以看看：http://lavasoft.blog.51cto.com/62575/53815/

接着便是第五问，问的sql调优，这个楼主当时没听明白估计应该是优化吧？！我就大概讲了下sql性能调优，面试官也没怎么说，就嗯了几下。

第六个问题很简单，大致问了java内存块以及关于多线程的理解！多线程就不多说了，各种面经上面都能看到详细答案，java内存块的话主要说说堆、栈、静态区以及代码区用来存放什么，说的详细点的话还可以扯扯垃圾回收器的机制，当然如果不懂就不要说。

最后又问了一些统计学相关的问题，就大致问问，问自己的统计学是不是科班出身，我说自学的，面试官就说还要好好的学学打好基础，最后反正还给了我很多建议，这次一面时间大概是30多分钟，自己不懂的可以向面试官求惑。

一面大致上面那些，二面还是电话面，在前天晚上，我是被惨虐了，是说项目经验，由于才研一。。哎也不多找借口吧，本来想讲我工作时候的java项目的，但是面官说我面试的是数据岗，要我挑个机器学习算法做的项目出来，我当时就蒙菜了，于是我就只好讲了个比较简单的小项目，甚至不能称之为项目吧，就是导师让我练手用的，果真说完，最后面试官说项目太简单了，不过也没多大事，接着就就我的项目提问，问我什么是特征选取，怎么选块，svm的核函数如何推出来，如何自定义核函数，二分类和多分类中参数选择等等问题，因为要说算法推导过程，所以最好是准备纸笔，我当时因为刚从外面和朋友吃饭回来在路上，没法只能找个安静的地方，算法推导的就都随便糊弄了下，最后说推不出来！还问了如果矩阵太大如何解决等等很细的问题，反正就针对项目给你提很多问题，问的很细，大概持续五十分钟左右。被虐的巨惨，但自己不知道的问题面试官都给了解答，最后建议我打好基础，那些推导也耐心的推了一遍虽然我当时没听太懂，但这素养确实很好！昨晚简历状态果真改变，已回绝！看来果真基础不够！

三面的话，自己没经历，就说说一些道听途说的消息吧，有小伙伴说三面是hr面，具体是就意向问题，城市选择问题，部门问题谈论，hr面之后，有些还会有一次技术加面。也或者是两个boss交叉面，不过交叉面不是每个人都会遇到，看手气了，大致就这些吧，博主写完躲角落里静静去！！难过委屈大哭