面试：微众银行（软通外包 hr：黄霞）

时间：20211112 15:00 （**本来14:30，面试官迟到半小时！！！**）

地点：深圳湾科技生态园汉堡王

对于这个地点其实我一开始也是疑惑的

**1.自我介绍**

A：略

**2.简单说一说做过的项目**

A：（按照宝典上的，根据自己的理解和做过的项目，扯）

**3.在项目中主要负责哪一部分**

A：（当然是ETL啊……大概不容易露馅？倒也不是我怕他问我存储过程……毕竟一直都复制粘贴要是让我现场写岂不是完蛋）

**4.你对贷款的业务有了解吗**

A：（当然……不了解啊！！！但是也不能直说嘛，）就说

平常工作中更关注取数逻辑啊这方面的，对于贷款其实本身了解得并不多，但对于贵公司的贷款其实我也有看过，比如有微粒贷、微业贷等等

【这件事告诉我们，面试之前一定要查一查对方银行的业务，必要时转移话题】

**5.（指着简历上瞎掰的银行）这个银行主要有哪些贷款产品**

A：（自己瞎掰的银行，跪着也要查完资料！！官网上都有，实在不行整那种村镇银行啊，哪些贷款产品都简单粗暴好理解）

**6.有用过spark吗**

A：（废话，当然没用过）啊这，这个我好像只是略有耳闻，但是没用过呢

**7.（追着项目细节问了很多，但太多专有名词我一下子没记住是啥）**

不良贷款率这个词挺烫嘴的，我一直想说这个词，一直说成逾期啥啥啥一直没说出来╮(╯▽╰)╭感觉他已经发现了我是假冒伪劣产品的样子

于是没得办法，只好和他说拨备贷款率之类的，得亏之前把泰隆银行的年末总结看了一遍，随意扯了一些数据，转移了他的视线

大概是发现我是假的？于是他开始问数据库类的了（谢天谢地终于结束项目问题了再扯我也编不下去了啊）

**8.这个项目你都用到了哪些维度？（好像还问了有哪些指标）**

A：（天地良心，我是真不记得……）嗯……好像，就拿这个逾期用户画像来说吧，比如用户的性别啊，年龄段啊，地区啊，职业啊，还有各种认证这些，因为是想了解一般逾期的用户都有哪些特征，这样银行在日后贷款审核的时候才能加强把控，降低这个，逾期……不是，不良贷款率（依旧烫嘴）

（别学我，真的好好看看那些个表，最好闭着眼背一遍表名表头啥的，我这都是现场编的，磕磕绊绊，简直不要太明显）

**【大概项目追问不下去了，开始问技术类】**

**9.你对数据仓库有了解吗，说一下自己的看法**

A：传统数仓三层（宝典有）

大数据的分层（宝典有）

**10.有写过Shell脚本吗**

A：有，但都是最初级的，在做项目改造升级的时候有接触过，就是把原来的脚本复制粘贴来，加个echo让他在日志里输出时间，然后因为sql语句不能直接执行，前面还要加个hive -e，大概就这些了，再往深的我就不太清楚了，但如果未来项目有需要，我也是会去学习的（说太多容易被追问常用命令，想想算了）

**11.现在有一张表，一亿的数据，然后是每天全量更新，总共50个字段，但现在我需要再加一个字段，也就是51个字段，要怎么做？**

A：（？？？？一亿数据你全量更新？增量他不香吗？？）啊……为什么要全量呢？

面试官：我就是要全量

我：（你高兴就好）额……那，干嘛要加个字段

面试官：我需要啊

我：（行，你赢）哦……我之前还没碰到过这种，嗯……（实际上一边问一边脑袋里疯狂想）

我：那这个，嗯，你更新的字段，这里面的数据，是都有吗？还是空值？还是赋值0？（其实我也知道是有数据，就是想拖时间思考）

面试官：当然有数据

我：哦哦，也就是说你会给我数据对吧，然后这给我的数据，或者说这新的表，里面也有能和旧表关联的字段，是这样吗？（这时候已经想到了关联，反正就是，要横向增加，要么关联，要么case when啊）

面试官：有的

我：那就把新表的数据和旧表关联，select的时候再加一个字段就好啦（其实我说得很明白了真的）

面试官：你写出来

我：（……行，你是面试官你最大，我写）

Select 前50个字段，51字段（新）

from 旧表 o

join 新表 n

on o.cust\_id=n.cust\_id

【这个答案不一定对，cust\_id也是我临时想的，所以**平时对表头一定要熟悉熟悉再熟悉，特别是这种常用的关联字段和主键**，一看就知道有没有用过】

**【话锋一转，他又看到了我电商的项目，又回到项目问题】**

**12.你说说这个电商项目**

A：（！！！糟了，光顾着看银行，电商这个其实准备得很少，再加上下意识觉得他简单好记好理解，但实际上说起来却不是这样……）

**13.这个平台主要卖什么**

A：……就一些服装啊，电子产品之类的，是一个比较小的平台（……感觉快瞒不住了嘤嘤嘤）

**14.你这个项目学到了什么**

A：嗯……这个项目吧，理解起来不复杂，不像银行有那么多的专业名词还有一些复杂难以理解的指标计算，但这个项目麻烦就麻烦在，他的小表特别多，比如性别一个表，地区一个表，颜色一个表，尺码一个表，有时候我需要同时管理很多很多张表才能得到需要的东西，这时候就需要好好去理解Mapping文档里的内容

**【气氛凝滞，就在我以为一切要结束之时（快半小时了），他，打开了手里的纸，给我了笔试题……】**

**15.你把这两题写一写**

A：（其实我后面还有一场面试如果10分钟以后再不走就真的来不及要迟到了！！！为了节省时间……我一边说一边简写）

**1）评分90分以上的总额**

思路：关联表，group by姓名，sum（amt），having score >90（其实应该写where，但我脑抽不知道为什么写了having）

**2）语句优化**

\*号不能使用，换成具体字段，然后先过滤再关联，不要关联成一张超大表以后再过滤会很慢

**【到这里，我以为一切就真的结束了，题也写完了对吧，问也问完了，真的求求你快放我走吧我真的要迟到了，所以我大胆地问了一句——】**

嗯，写完了，还有吗？

面试官：（瞅了我一眼，拿起笔，在纸上，手写，了，一，道，题）

我：？？？？？？？？？还来？还不够吗？

**3）把abcdf1234转化成4321fdcba**

A：这个要用函数吧，我工作没接触过，还不太了解（我不会，快让我走）

**4）（又拿起笔，又写）把一行的abcdf1234，变成一列abcdf1234**

A：（脑袋中飘过行转列但好像不太对……）额……这个我也没试过干这样的事儿啊（放！我！走！）

**5）（很无奈）那你把你用过的函数都写一写**

A：那成（用了平生最快的笔速，写得我自己都快不认得的字）

max min avg sum count ，后面加个over变成分析，嗯，这10个了

lag lead

rank，dense\_rank，row\_number

to\_char/number/date

round,trunc,mod……

（其实平时记得很熟，但当时很急，越急越想不出来）最后强行挽尊……嗯好像还用过floor，大概就差不多吧

**6）lag和lead有什么区别？**

A：一个往上一个往下

**7）你有什么想问我的吗**

A：（没有没有我们快结束吧！）嗯……就，如果我进去后还有什么不会的，比如大数据这一块，就不是很熟，你们会有人带我吗？

面试官：这个啊，我都还不会

我：？？？？啊……这……额

面试官：没事，有个专家在，他会教我们

我：哦哦……这样啊

**——————结束——————**

**总结：**

1. 要对简历里每个项目都很熟悉，起码表名表头，哪些维度，事实表有啥，最后需求是啥得整明白
2. 没听过的一些技术就直接说以前没接触过得了，别啥啥都会的，给自己挖坑
3. 对自己瞎扯的银行要熟悉，起码查一下，有哪些业务啥的
4. 对面试的银行也要熟悉，方便转移话题
5. 没事儿多读一读背一背那些个指标（特别是烫嘴的），函数一定要用手写一写，不然容易出错
6. 两场面试尽量别安排在一个上午或者一个下午，除非两场离得特别近。你是不会迟到，但是面试官可能会啊！所以我最后那场面试还是迟到了╮(╯▽╰)╭