[HR Screen]

Actual questions: August 27 2019

- 1 + NULL =? NULL
- How to filter aggregations: HAVING
- UNION vs UNOION ALL: which is more performant
- Default sorting: Ascending
- Stack vs Queue
- recursion vs iteration
- keys in dictionary distinct? YES
- Why DE at FB

[Phone coding screen]

I was referred by one of the FB employee. One of the HR connected to me in a week and asked me about my experience. He was very happy with me and then came the online coding round using coderpad. I was asked 4 SQL and 3 Python questions. I think they have 5 SQL and 5 python question but depending on time bandwidth they ask you as much as possible. I did very well in SQL. I struggled with Python. But I cleared the interview. The reason I was sent to the next round was I shared my thought process at each level. For example. When they showed me the ERD I mentioned what are the facts and dimensions. I also mentioned that there should one more dimension for promotions. Basically gave my views on what could be better. Then with SQL I always started giving answers logically. by always giving the business context. Example-- One question was to find percentage increase in revenue due to promotions. I didnt even look at ERD and directly said promo_id should be in fact - 'sales'. Price of the product should be in dimensions. Ask the interviewer what exactly does she mean by revenue. Like are there any additional earning you get for a product.

- 60min Coding Interview SQL&Python using Coderpad, finish as many questions as you can.
- Explain your logic while you are coding.
- Cross questioning on performance improvements or alternative solutions.

[SQL]

[TABLES:]

四个tables，columns蛮多：

products：product_id, brand_name, class, price…

sales: product_id, store_id, transaction_date, customer_id, …

stores: store_id, store_type, country,product_id, customer_id, store_sales, promotion_id, transaction_date

customers: customer_id, gender, register_date, education, …

eg. **sales** volume by day by product by **store**

第三题the number of customers who have made at least one purchase in each group（定义了两个groups, 分别用两个维度来定义）

第四题要求对male, female分别计算三个值：1. 各有多少人；2. the number of days at least one sale；3. the number of days has at least one sale at one given store type（这一题差不多是这样吧，题目比较长，2.3需要理解一会儿，要求用一个query同时output, 写3的时候就很懵了，最后也没完全写完）

第一题是求没有消费过的customer个数

第二题应该是求product_class_id的交易额之类的

第三题是求gender和education_level等于两组给定情况的customer个数

第四题求分性别的消费者人数，购物天数，在某类型店铺的购物天数

第五题求注册日期距离第一次购物日期最近和最远的消费者，注册日期，首次购物日期和注册日期距离第一次购物日期的天数

1.select 买了a和b两种产品的顾客
2.每个product_class的 一个什么值，很简单的group by product_class_id
3.不记得了
4.统计 在所有加州(state=CA，不是California)的数据中，买过0种商品的顾客有几位，买过1种商品的有几位，and so on, group by 种类的数量。这道题要连接多张表。

1. percentage of male customers
2. percentage of revenue increase due to promotion (CASE WHEN promotion_id != 0 THEN revenue…) blablabla
3. customers that purchase both A and B -> 不要用subquery！！！！
4. top 3 sales product classes
   前四个做得太顺了 最后一题给了个有点难度的 记不太清了 经提示还是秒了 提问环

节我问了他一些**关于A/B test设计和product metrics设计方面的问题 包括click through rate啥的**

1. Coderpad SQL questions with buildin data schema.
2. The names of salespeople that have 2 or more orders.
3. When was the earliest and latest order made
4. cumulative sum
5. top 3 Products by sale
6. % using Case
7. Basic Having clause and one Set operator (Intersect) type question
8. 1. given table products(product_id, brand_name,product_name, product_price), find brand_names with avg price > 3 and having more then 5 products
9. 2.given table sales (sale_id, promotion_id, sale_amount, other columns), find % of how sales with promotions (promotion_id is null) are doing in comparison to all sales.

One SQL question was tricky because you have to look at the data to come up with the question, the question is to find the percent of valid promotion, condition to find valid promotion is not specified in the question, so you have to figure it from looking at the data.

Other SQL questions are based on a join, group by, top(3)

1. Percentage increase in revenue compared to promoted and non-promoted products.

2. Products classes that has the highest transactions

3. Count of Customers who bought 2 items type (A,B)

[PYTHON]

2、给一个list里面都是数字，输出一个map里面是数字以及对应的计数

3、给一个list里面都是数字并且含有重复数字，输出去重后的结果，当相同数字出现第二次及以上就不输出它，仅输出第一次出现的

4、接下去是个坑，其实是个原题就是flat integer list，然而给定的形式是List<Object>，compile的时候当Object应该是一个List的时候，我尝试了各种策略都不能convert成List<Object>。。面试官也写不了Java，她做了一个奇特的操作，她把正确答案粘贴给我。。。我就懵逼了。于是这个就不知道算过不算过，就下一题了

5、这题是给一个数组，还有一个pct，求不小于这个pct的数值。面试官只让我说思路，说完嘴跑了几个test case就没时间了，她说extra题让我别写，但很怕她在坑我，都感觉自己在挂的边缘徘徊了。需要注意的是pct是按照递增来看的，不过数组不一定是有序的需要先把数组sort成递增的，再做操作。
对这份工作的expectation是什么

3.Neighbor那道题，就是根据input list return出每个人有多少小伙伴

1. **Valid IP address**
    a. edge case to remember is if there are alphanumeric characters
2. graph node count
3. **average word length in a sentence.**
    a. some of the edge cases are having spaces in the beginning and end of the words, returning a float instead of int, returning None for blank input.
4. Flatten a nested dictionary
5. Binary Search
6. closest distance
7. **valid parentheses ***
8. **valid brackets**
9. Facebook Friends Tree (?)
10. dictionary
11. list of lists
12. **flatten list of lists.**
13. **Friends problem [[A,B], [B,D],[E]...] (List of lists) (count friends)**

    1. Average length of letters to words.
    2. [[A],[A,B],[A,C],[B,D],[C,A]] -- Find the alphabet with highest neighbors? -- (Wasnt able to solve because of time limit but the interviewer was like I get what I want to convey.. I gave her an algo of what I would I have done)

You have a 2-D array of friends like [[A,B],[A,C],[B,D],[B,C],[R,M], [S],[P], [A]] Write a function that creates a dictionary of how many friends each person has. People can

have 0 to many friends. However, there won't be repeat relationships like [A,B] and [B,A] and neither will there be more than 2 people in a relationship

Object Oriented Design for a data set, Data Structures : Questions on Trees.
3sums
python: avg length of sentence , valid IP address

Dot Product with Sparse Vectors, Dot product of sparse vectors stored as hashmaps

how to find duplicates in an array

1. Flatten array: [1,2,3[4,5,[6]]] -> [1,2,3,4,5,6]
2. Number of words in a sentence ‘Hello World’ -> 2
3. Valid Parenthesis -> 注意这个是带星号的那个题 (*)
4. Valid anagram
5. Count number of substrings in a string

[Onsite]

working session with colleague. design a table to get desired result followed by SQL and the n python.

Given a multi-step product feature, write SQL to see how well this feature is doing (loading times, step completion %). Then use Python to constantly update average step time as new values stream in, given that there are too many to store in memory.

All on site questions are regarding DW design, table design, Schema design and etc. The interviewers ask sql questions based on the DW that I designed for them. The onsite questions are no match with coder pad interview. You need to have a good understanding of system architecture and system design. I personally could not finish a python question . The question was about updating metrics (I defined the metrics to answer earlier question) for a fb module as stream of new data is coming in. As interviewer point out I was on right track having a session id and user id to de-dup same user activity, but it was not enough i guess.

Given this data calculate this KPI How would you design the data model for this scenario

he in person interview will consist of 3 problem solving interviews, in my experience you will be given a scenario of a data engineering problem (i.e how to be identify the metrics for performance for this specific feature) and you will be expected to write SQL and actual code for the context of the problem itself. There will also be a more standard behavioural interview portion, asking you about your work experience, how you deal with interpersonal problems that sort of thing. Finally there will be a casual lunch conversation where you can ask about the work culture and other day to day questions.