# Two Sigma面经2022

1. **两个马23SDE实习技术面经**

RT，3轮视频技术面，第一面和二三面之间隔了一段时间。属实没想到自己能进final round

第一面：设计一个database class，实现创建、插入、查找功能，自己写test case。非常感谢大哥带飞，如果他能看到这篇的话

第二面：给两个数组表示好友关系，要求邀请一些人，这些人之间互相不为好友，求最多的邀请人数。应该是挂在这里了

第三面：地里最近比较常见的image matching： https://www.1point3acres.com/bbs/thread-909099-1-‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍1.html

三轮面试的面试官人都很好，还帮着给提示debug hhhh。HR也很给力回复很快。interview guide里说穿你觉得舒服的衣服就可以了，这点真的很爱。总之面试体验很棒，无奈本人实在是太菜了，最后没过

Cr:https://www.1point3acres.com/bbs/thread-944374-1-1.html

1. **TwoSigma QR VO技术店面**

之前GHC投的简历，今天约了4轮面试。第一轮代码，第二轮open case，第三轮统计，第四轮BQ。

全程没有太多时间聊天，基本都在做题。

第一轮VO：Coding，Leetcode 200. Number of island的变种，返回一串数字，代表相同大小的岛屿的数量值。用Hashmap做，修bug时间有点久，跑通过用的时间刚刚好，但是没来得及问太多问题了。应该是需要很快做完能和面试官交流，lz代码熟练度还是不太够需要再练习。

第二轮，open question，预测未来一小时内共享单车的数量，也是经典老题目了，面试的是个哈佛的统计小哥。之前复习了kaggle上经典的答案，从获得数据到处理数据，小哥问了很多followup问题，除了天气温度湿度，还有如何获取一些spatial和time相关的数据，比如如何预测人流量（游客的数量，可以通过酒店预订，周围博物馆门票，公交地铁售票量），共享单车和公共交通的距离，如何将公共交通的时间数据分类处理（用什么关系做weighting），感觉考察的是考虑问题的全面性。当然feature engineering问题很多，lz之前对时间序列，没有空间数据太多经验，花了比较久。如何可视化数据，考察multilinearity，选择相关性强的feature，这部分答得比较快。然后考察对随机森‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍林的预测，树fitting的是什么，然后时间就不够了。总体来看lz对数据处理和ml模型还不够熟练，导致前面回答的不够快，问题没有考察完。

为了准备这个面试大概一个月左右，前面花了大量时间补统计和线性回归，如果时间更充足会多花点时间复习ml的知识。不过lz通过这次面试也整理了很多quant资料（考察的基本都在题库里），如果感兴趣的小伙伴欢迎加我分享，一起再战～

Cr：<https://www.1point3acres.com/bbs/thread-943605-1-1.html>

1. **Two Sigma Intern 三个行为面**

一小时，详细问了项目和各种带情景条件的BQ

一小时，继续问了更多BQ，看起来不太想招人的样子

半小时，简略问BQ和项目，是等级很高的manager

Cr:https://www.1point3acres.com/bbs/thread-940913-1-1.html

**4.Two Sigma Intern 三个技术面**

Connect 6

不相交的区间个数

nearest cities

全都是原题，可惜lz最后一面太累了，就没写出来，真的好可‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍惜

Cr：<https://www.1point3acres.com/bbs/thread-940637-1-1.html>

**5.TwoSigma OA 附思路以浏览**

原题见图

第一题 股票拍卖

有n份股票，有k个bid，bid里包括userId, price, quantity, timestamp。bid分发要求从价格高的开始，同价格的按照timestamp从小的开始每人发一份，如果股票够多就再发一轮，直到股票发完，或者所有bid都满足。要求返回一份股票都没有领到的userId，按照从小到大排序。

思路

先sort，然后安拍序处理就行了。需要注意的是，这题只要求每个bid最少拿一份股票，所以很简单。

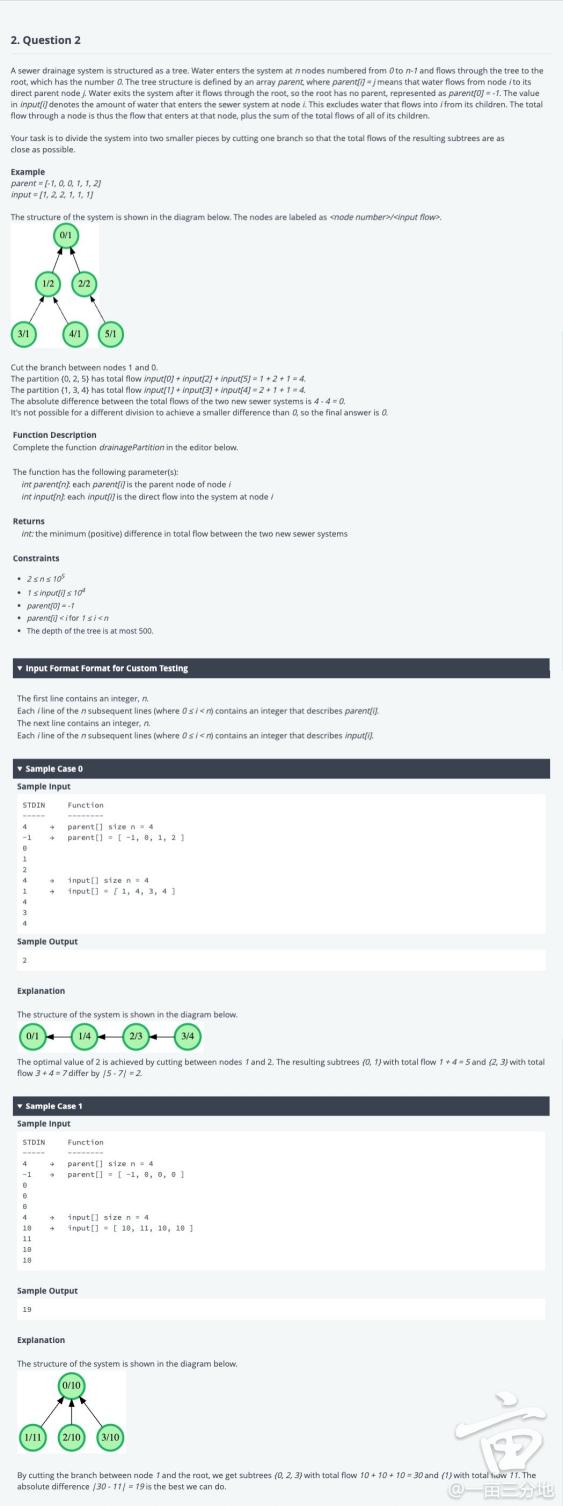
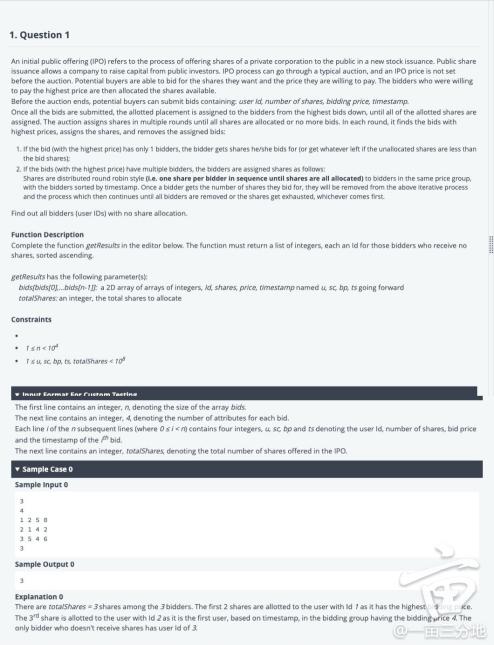
第二题 树状流量切分

给了两个等长的list，一个是parent，一个是size。对于第i个，parent<i>表示父节点的index，size<i>表示净流入的流量大小。一个节点的总流量是自己的净流入量和所有子节点的流量的综合。要求把这个树分成两个树，让两个树的总流量的差值最小。返回最小差值。

思路

数据结构上比较像线段树，所以父节点的index永远小于子节点。我们可以‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍从后往前扫描，1pass得到每个节点的总流量。

对于总流量S，两个树的总流量分别就等于N和S-N，那么最小差值就是求min(abs(S-2N)。2pass就能解决这个问题了



https://www.1point3acres.com/bbs/thread-939983-1-1.html

**6.二马VO**

猎头联系的，然后OA，之后onsite。recruiter没按我给的时间安排面试，最后安排成了四天面试，中间她也不断在改时间，反正也无所谓了，毕竟有时差(我在西部)，自己就按她的安排来的。问她target level，她说看面试。(其实到最后被拒，我都不知道我面的level，被拒电话还是猎头打来的。在levels.fyi上，我在自己公司本身的level是cover他们家4和5的)

共七轮，3轮technical，4轮mgr，共六个半小时。

第一轮，白男，原题，merge多个stream，stream的元素包括值和时间，但是值是delta

第二轮，三姐，原题，top K，三姐迟到了，匆匆忙忙找题和贴题，说完思路，她同意，然后我准备写。她说她把sample和function帮我贴出来，结果一看sample，发现我们都理解错了。我们之前理解成了如果出现(A, 1)和(A, 2)，就merge成(A, 3)，结果一看sample，其实就是1变成了2，不需要相加，然后又重新说了下思路，写完了。感觉三姐就丝毫没准备。

第三轮，亚裔男，原题，connect 6，他改成了connect 5。把代码写了，他说不需要test，说我思路是对的。然后问了好多关于处理数据时partition和ordering的问题。

第四轮，白男，senior mgr，这轮预先说了是半小时。他迟到了几分钟，也没说道歉，然后说了好几次wait for a second，回他自己的信息以及找我的简历。打哈欠，问了些平常问题，想做啥。最无趣的一轮。

第五轮，白男，mgr。主要就是针对他组里的ser‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍vices在聊，他说他们面临哪些问题和challenges，问我有什么想法和建议。这轮的面试形式我觉得还挺好的，就是聊实际项目，真的能帮助双方看是否match。

第六轮，黑女，mgr。迟到了12分钟，然后也是BQ，没啥follow ups，聊的不深。都是答完了就进入下一个问题。

第七轮，印裔男，无口音，mgr，和第五轮差不多。

周五面完，四天后的周二猎头打电话过来说unfortunately...。问他有无feedback，他说TS说不match。

Cr:https://www.1point3acres.com/bbs/thread-914608-1-1.html

**7.Two Sigma DS 挂经**

基本都是之前地里面出现过的面经第一轮oa 就是地里面千年不变的几道题，线性插值，纽约温度预测。。。

第二轮和hm的店面，主要问了一些背景的问题，case study是问怎么用google search data 去预测公司（Starbucks）的revenue。

vo三轮：

第一轮：怎么预测纽约apartment的价格，用哪些features，features和房价的关系是什么。。。

第二轮：如果给你10几年的url的数据，有什么business use case？怎么预测电商的revenue？如果这些url可以track到user ids, 但是很多user 中途消失了，怎么correct这个bias？

第三轮：利口122, 虽然写出来了，但是写得太慢，debug了很久，应该是挂在这了

Cr:https://www.1point3acres.com/bbs/thread-939873-1-1.html

**8.two sigma vo**

1.bloom filter

2. 设计kv store

3. 2d array treasure hunt

Cr:https://www.1point3acres.com/bbs/thread-933563-1-1.html

**9.Two Sigma OA老题目**

两马通过GHC拿的Code Test。

地里老题目

第一题IPO求没有分到shares的bidder，输出的时候忘记转成list导致有3个test没过

第二题下水道树结构，构建一个TreeNode就很好做了。

大家按照地里的好好准备一定都能过的

我看之前大家都是分享图，我分享一下文字的原题吧

原题复制在下面：

第一题

一家公司 sellshares.com 网站上注册IPO。本网站上的所有股票均可在称为竞价窗口的特定时间范围内竞标。在竞价窗口结束时，拍卖逻辑用于确定有多少可用股份归哪个竞标者所有，直到所有可用股份都被分配完毕，或者所有竞标者都收到了他们竞标的股份，以较早者为准。

出价以<用户 ID、份额数、出价价格、时间戳的形式从用户那里到达>直到出价窗口关闭。

拍卖逻辑按如下方式将股份分配给投标人：

价格最高的竞标者获得他们竞价的股份数量

如果多个竞标者以相同的价格出价，则按如下方式为竞标者分配股份： 同一价格组中的每个投标人将连续分配一股，每个竞标者根据其时间戳排列在组内。一旦竞标者获得他们竞标的股票数量，他们将被从上述迭代过程中移除，然后继续该过程，直到所有竞标者被移除或股份用尽，以先到者为准。

列出分配共享后甚至没有获得一个共享的所有用户的用户 ID。

输入

出价：由整数、Id、股票、价格、时间戳组成的二维数组，名为 u、sc、bp、ts 前进

total\_shares：一个整数，要分配

的总份额 输出

一个整数列表，每个整数都是那些没有收到股份的投标人的 ID，升序

排序

示例

1：输入：

出价 = [[1， 5， 5， 0]， [2， 7， 8， 1]， [3， 7， 5， 1]， [4， 10， 3， 3]]

total\_shares = 18

输出： 4

说明：

最高出价是用户 ID 2 以 8 的价格为 7 股，因此用户获得 7 股，剩下 11 股分配给较低的价格。ID 为 1 和 3 的用户分别出价 5 5 和 7 股，出价方 1 的时间戳较早。经过 5 次迭代，分配了 10 股，其中 5 股分别分配给这两个投标人。投标人 1 已全部分配，投标人 3 还有 2 股要购买，还剩 1 股要分配。它进入投标人 3，所有股份都已分配。投标人 4 是唯一没有获得股份的投标人。

约束

1<=n<=10^4

1<=u， sc， bp， ts， total\_shares<=10^9

第二题

下水道排水系统的结构为一棵树。水从编号从0到n-1的n个节点进入系统，并通过树流到编号为0的根部。

树结构由数组父节点定义，其中 parent[i] = j 表示水从节点 i 流向其直接父节点 j。水在流经根部后离开系统，因此根没有父根，表示为 parent[0] = -1。input[i] 中的值表示在节点 i 处进入下水道系统的水量。蒂S 排除了从其子项流入 i 的水。

因此，通过节点的总流是进入该节点的流加上其所有子节点的总流之和。

您的任务是通过切割一个分支将系统分成两个较小的部分，以便生成的子树的总流量尽可能接近。

示例：

父级 = [-1， 0， 0， 1， 1， 2]输入 = [1， 2， 2， 1， 1， 1]

输入

剪切节点 1 和 0 之间的分支。

分区 {0， 2， 5） 具有总流量输入 [0] + 输入 [2] + 输入 [5] = 1+2+1 = 4。

分区 {1， 3， 4） 具有总流输入 [1] + 输入 [3] + 输入 [4] = 2 + 1 + 1 = 4。

两个新下水道系统的总流量之间的绝对差值为4 -4 = 0。不同的除法不可能实现小于 0 的差异，因此最终答案为 0。

Cr：<https://www.1point3acres.com/bbs/thread-931403-1-1.html>

**10.2s vo挂经**

2s的vo真的顶 连面三小时三轮。。。

第一轮friends invite，几乎是小哥手把手带着我写的，需要dp来做，地里好像没看到完整的题目描述，放图片里了

第二轮第三轮是两个design题，一个是设计hashtable，需要考虑hash collision，hash函数是会给的。需要自己在unit test里面写test case。另一个是Huffman编码题，地里出现过。

第二第三轮楼主感觉挺好的，都做完了也测完了，但是由于第一轮表现过于拉垮，第二天收到拒信说没有进bq轮。不过早有耳闻2s很难，本来也‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍没有觉得能过hhh不知道图片为啥上传不了。。。第一题大意给一些数组，数组里的两个人是朋友，需要找到最大的set使得这个set里面的人两两不是直接朋友，返回这个set里面的人数

Cr:https://www.1point3acres.com/bbs/thread-929359-1-1.html

**11.TwoSigma 社招 OA**

刚做了TwoSigma的code test, 把面经分享一下，顺便求个米。

第一题是求ipo股票assign的问题，列出一系列得bid price跟数量，以及可分配得share，按照bid price从高到底assign。如果有很多人价格一样，那么这部分人就按round robin，根据timestamp顺序从最小的开始，依次每人assign 1 share，直到有人assign满了，就离开这个队列，继续assign其他的人。问最后哪些人没有assign股票。这题也不难，基本就按照题目的规则，用一个HashMap记录bid price和哪些人出的这个价格。然后对bid price排个序，根据Map找出这些人以及他们的要的assign的share, 依次减去shar‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍e，直到可assign的share为0，然后记录剩下没assign的人的ID。逻辑不难，就是要写一些代码。

第二题是一道高频题翼散散酒，稍微变动了一下。不是binary tree，而是多child的树，求得最小差，不过方法基本跟lc上的一样

Cr：<https://www.1point3acres.com/bbs/thread-926284-1-1.html>

**12.2sigma QR**

OA: 全部都是老题目, 180min 2+1. 第一题是linear interpolator, 第二题是NYC temperature, 第三题是第二题的最后一问从两个town变成五个town.

VO: 五轮.

第一轮是一位KCG的前辈人非常的好, 不做alpha而是monetization. 面到一半我就告诉他了我对execution没有兴趣. 他问了一个technical的题目, uniform从0到1的随机数一直sample直到加起来超过1, 问期望要抽取多少次. 答案是e.

第二轮是securities的一个quant, 都不是hiring manager. 交流下来他的alpha经验非常有限, 本身入行时间就不长在大厂里面板砖学习速度也不快. 我对高频无感, 直接就pass了. 他也没有问什么technical.

第三轮是一个做alternative data的. 交流了一下发现他们做的特别复杂, 想招个人来专门combine不同的alternative data. 我告诉他你怎么combine别的signal就怎么combine这个, 整那些花里胡哨的没什么用. 我再告诉他确实alternative data有特别的方法来combine, 他又说这个实现起来难度大. 然后我随便提了几个烂大街的alternative data alpha发现他的理解非常肤浅. 我估计他就是专门在datanomics这个组整花活的吧.

第四轮非常敷衍, 问了我一些经历没有任何细节, 到一半突然开始做题. N个有corr的identical RV, sample mean为什么是unbiased estimator, sample variance来做variance estimator可以么. 一个linear regression如果X append duplicate, y appen‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍d zero, beta和tscore怎么变. 毫无新意的问题. 我问他要招什么样的人, 他说要招researcher. 这就好像你去KCF下单, 然后说我就要一份肯德基.

第五轮更加敷衍, 问的问题都是philosophical的. 比如问我为什么要跳槽, 我说我现在做的alpha不够深入不够polished. 他问题我为什么要做的深入, 花50%的时间做到95%的效果不是很好么. 算了吧我们可能不是一个频道的.

Cr:https://www.1point3acres.com/bbs/thread-922840-1-1.html

**13.2sigma intern 挂经**

第一轮是coding

题目是把一个类似regex的expression给parse完，之后用parse出来的规则处理给的string

不是很难，具体有点忘了

第二轮是data science case study

给了一个纽约citi bike的case，问怎么model station用车的demand和supply

要想一些有用的variables‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍和数据处理办法

第三轮是纯stats

如果现有一个linear regression，现在把所有data points都复制粘贴了一遍，重新run regression，问what will happen to the weights, slope, correlation, std error 等？

Cr：<https://www.1point3acres.com/bbs/thread-919732-1-1.html>

1. **wo Sigma VO**

三轮长度一小时的技术面

1. Image matching - 先找出两个grid mismatch的点， 然后DFS；

2. Top K in N streams - 用Heap做，要理解题意自己写test case，题不难但要仔细看懂题干。这里面试官一直在打断我，我觉得他提的设计有一点问题，和他探讨了一会儿，最后没来得及cover所有的edge case；

3. Implement Hash table - 问hash table是怎么实现的，题目是蠡口 奇凌溜， 也要自己写test case。这里没想起来hash table的具体实现细节，面试官还给我发‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍了wiki让我自己看，看懂了开始写；

Cr：<https://www.1point3acres.com/bbs/thread-909099-1-1.html>

1. **两🐎 9轮onsite**

漫长而无聊的体验，面到后面都麻了，一共2轮coding，1轮distributed system design，6轮team match + BQ，目前还在等结果。

第一轮：原题- max product from largest group of friends，题目看面经：https://sammy-sheng.gitbook.io/two-sigma/。 Implementation我没看面经-（没看懂）；自己local提前写了一遍，并当场通过和面试官交流debug出了问题，反响不错。

第二轮：没见过的新题。类似于leetcode的serialize and deserialize binary tree，也是写两个method。但是serialize和deserialize的方法是题目制定的，理解algorithm花了好久，最后勉强写出了serialize的implementation，deserialize只说了下想法。整体感觉一般。

第三轮：distributed system design，具体题目忘了，方正不是面经的OOD，难度一般，面试官水平一般，没问出什么有价值的问题。

前4轮BQ：3个Sr manager，分别问了背景和BQ，主要问了confilct还有disagree with others（这两个是一个模版）。1个所谓的Bar raiser，不是match的组，也问了一些project的问题。

‍‍‍‌‍‌‌‍‌‍‍‌‌‌‍‌‍‍‌‍

后两个BQ：1个manager，貌似这个manager report给上3轮的其中之一，问了一些比较贴近日常工作的问题，比如你愿不愿意take一个legacy system，有哪些trade off等。1个partner quantitative researcher，问了一些big data相关的问题。

还没出结果，整体感觉就是又臭又长，说是team match，其实完全是单向的，然后感觉面试成本太高。

Cr:https://www.1point3acres.com/bbs/thread-908304-1-1.html