第二轮 外星人字典原题，一个ABC小姐姐面的，说实话要不是这是面经题我真的不知道她到底要干啥。给我一个array Apple banana orange, 问blueberry放哪里 我擦嘞这是什么意思。。然后我猜猜是面经题，就说放 banana后面 她说that's it..。那是要输出啥还是要给一个array和一个单词输出插在哪里啊亲也没说。。接着我说那我们输出字母顺序吧》她说好。。然后我说这题是拓扑排序。。她问我 那你为什么不用别的排序呢。。？？还能这么问？？？。。，题目虽然做出来全程聊天尴尬。。

第三轮 行为树 地里有原题。 HashMap＋ tree做不难。 当时时间匆忙这个面经没怎么仔细看。不过应该现做都能做出来。

轮是论坛里常见面经 behavior tree。 后来打印我先写了level order traversal, 然后interviewer问这样怎么知道parent child 关系。好吧，我说那不能，那就写dfs吧。

通过用户访问网站的顺序输出行为树，输入格式 时间 用户 页面编号，输出为树形结构，记录每个页面的路径和访问次数，比如用户1 访问的路径为A->B->C,用户2 访问的路径为B->C->A，用户3访问的路径为A->B->D,则输出

A(2)---

| |---B(2)

| |----C(1)

| |----D(1)

|

B(1)---

|---C(1)

|----A(1)

TrieTree

第四轮 中国小哥面的 找出reverted sorted array找出最大值。然后写完了还有好久。。

感觉他现想了一道题。。比如给1 1 3 4 然后告诉你 有一个密码是由这几个数字组成的。。。现在有个API可以用来猜密码。。返回true／false.

然后API的input是个string 比如 1123132 那么API会测试 1123 1231 2313 3132看看这里面有没有正确密码。。。 好了其实就是说能不能有个string能cover所有的 combinations of given input 并不用那个猜密码的API..求最短String。。。然后我想到了用brute force然而不知道那是不是最短。。

最后小哥说有个什么算法（对不起名字不记得因为没听说过）可以做不过没指望我面试15分钟能写出那个。。然后就结束了。 他好像是临时来顶替面试的。。。

De Bruijin Sequence

一 二叉树横向遍历，复制带随机指针的链表

二 设计只读（？！）键值存储

三 实现行程长度编码

https://en.wikipedia.org/wiki/Run-length\_encoding

四 设计广告计数系统

就是统计每个广告点了多少次，纯粹计数，没有任何其他要求也没有feature层面的foillow up。设计的时候各个环节都追问（存储，消息队列追问的比较多）

给这些隐射

A - > B

B - > A,C

C - > B

D - > E

E -> D

求输出 [(A,B,C), (D,E)]

第二轮：TOP k elements

老白

第三轮: 老题，在其他面经里面见过

中国人

trie + DFS

具体题目：

给一个wordlist，然后N， 求问怎么用这里的word排成一个N\*N的矩阵，这样每一行，每一列都是一个wordlist里面的词

wordlist = [cat, oah, aee, coa, aae, the, abh, csw]

cat

oah

aee

不用每个词都用上。

这一轮老中不看算法，只关心上机跑不跑得过。。。

1. 数据结构和算法

给一堆公司门禁的安全日志，内容有时间，门编号，员工名称，是否有权限。根据是否不同的人在相同时间地点的出入记录推测哪些人是朋友，输出top10的朋友

followup：

如果公司很大，跨国公司，不同的工作地点（LA,NY,Beijing,blablabla)的日志，都存在一起，怎么海量日志找出top100

4. Problem Solvin

面经题：输入是一堆pin和一个col数，将这些pin放入col个list中返回，填充页面，规则是每次填一个pin的时候找当前像素数最少的那一列填

followup：忘记了。。

round 1: coding,给所有人的schedule，find all time intervals that all people are free

LeetCode 759

round 2: design homefeed

常规题，不过问的很

round 3: coding，design a hit counter for a website

基本上实现2个interface:

1. set(timestamp) - visit website

2. get() - return how many people visited website in the past 5 mins

LeetCode 362

1：Coding: 面试的小伙子在dropbox的allocateId版本上做了些修改， 给个值max, 范围0 到 max - 1, 第一个api allocate() 返回没用过的第一个值， 第一次肯定返回0， 然后1，

2. 还有另外一个api deallocate(int id) 把这个id回收，然后重新利用。 这时候就不是线性的增长了， 会把整个范围切成几个段。 从最开始map优化到treeMap 存 interval.

Allocate Id

2: Coding: 给一个没有排序过的array, 返回能用array里能用binary搜到的个数。 比如[3, 2, 1] return 1, [3, 2, -1, 4] return 2。

[3, 2, -1, 4] 这里面有两个数可以用binary search找到， index 为1的“2”， 和index 为 2 的 “4”。 所以return 2 (可以找到两个)

2. System design, 华裔小哥，刚从推荐组转到vision组，设计yelp的一个功能，给定餐馆ID list R以及经纬坐标C，如何找出给定餐馆Q的 near 餐馆集合，如何scale。

3. Coding, 华裔小哥，蠡口 散把灵 + 聊project （他是P家一篇很有名文章的作者）

4. 原本schedule上该是ML design，但是搞错了进来的国人大叔说他分到的是coding，于是就面了coding，蠡口 留酒屋 和尔凌凌合体， 同时聊了一些behavior question， 遇到的最难的项目啊，如何处理conflict啊balabala。

岛屿问题

岛屿个数，遍历统计1的个数，有一个count++，同时在岛屿合并的时候count--，最后count就是个数

每个岛屿维护一个大小计数器

6. Coding，有点忘记了，但是是用trie 做，大概类似于蠡口 尔要要，裹了了Pins的外衣。

LeetCode 211

1.国人小姐姐。设计实现battle ship游戏。主要是代码实现，跟OOD没关系。.

3.印度小哥。LC277，LC759

Find the celebrity

Employee free time

4.白人小哥，类似json parser.输入是 string path 和 JSONObject. 比如path ="$.store.book.author", 要把这个JSONObject里面的author的value 返回。json是nested,并且有array

5.日本小哥，sorte integer array, 找occurence大于n/4的elements,然后扩展到n/k,然后扩展到array特别大，多机怎么处理。

二分查找寻找每个元素的frequency

找majority number的方法主要有：

排序+取中间点

未排序，摩尔voting算法

未排序，对整个数组遍历32次，对每一个比特上1出现超过半数的计为1，可以得到最后的值

Onsite:

1. coding，美国小哥，设计个接口，支持2个函数，ingest(String) 和 generat(int)。ingest 的 input 是一个String；generateString 的 input 是个整数，output 是个String，要求根据ingest 的input 的转移概率矩阵输出一个随机的字符串；字符串的长度最大是给定的整数，或者当遇到终止符 \n 的时候停止输出。

比如输入是

a b c \n

a c \n

那输出字符串里，第一个字符为a 的概率为100%， P(b|a) = 0.5, P(c|a) = 0.5, P(c|b) = 1, P(\n|c) = 1

3. system design，不知道哪里的小哥，design homefeed

4. domain specific，国人小哥，本来是问research，结果他不懂我的，我不懂他的，就变成了问coding。蠡口 物流 和 武器。

Merge Interval

1. System Design/ ML. 本来应该要考SD的，结果一上来看了我的背景说，我们直接考machine learning design吧，然后就是问设计pinterest的推荐系统

5. LC 698的变形，输出的结果是分组的情况。没有做到bug free，中间进行了debug，最后run了两个test case，时间到。

LeetCode 698

6. 总共有三小问，第一问：给定一个很长的数组N，每个数字都是0-255的数值，和各种pair的index的数组M （可能是[[1,100],[5,1000],etc]）。M的长度最大是N\*(N-1)/2. 问求M中所有pair的mean。用个presum，效率是O(M+N).

第二问：相同的数组，求M中所有pair的median，在预处理的时候给每一位计算当前总共有多少个0，多少个1，。。多少个255. 每一位用256的数组存储。效率O(M+N)。

第三问：当N特别大的时候，我们无法存进内存中，给定一个压缩指标K, K<<N， 问怎么在预处理的时候进行压缩，同时效率还是O(M+N). 答案是每次隔K个进行存储，这样在查找的时候每次需要K次，但是K是常数，所以效率还是O(M+N).

1. 给一个字典list of strings，和一个2D的matrix，把string填到matrix里，使得matrix的每行每列都是字典里的单词

字典中的字符串可以任意长度，但你填到matrix的时候只能填符合matrix长度的字符串，matrix的长宽可以不等，要求是用字典中的字符串填matrix之后，matrix的每行和每列都必须是字典中的单词。

比如字典中含有一下字符串abc, aaac, ace, ade, aaa, ace, ccc, bacc, caec, abcdde，填到一个4x3的矩阵，填好的matrix可以是这样的：

abc

aaa

ace

ccc

2. 利口尔酒，剩的时间询问了一些关于组里做的东西

LeetCode 29

3. OOD，实现一个游戏，游戏就是两个选手可以在一个2D board上造船，船就是一条线，可以是横着的也可以是竖着的，两个选手互相攻击对方，直到一个选手的船全部被攻击了，游戏结束

1. System Design auto-complete热门搜索。基本的读写，serivce的交互，如何优化，如何跟新。

2. Log process。找出一个log file里的某些特定内容的line。Follow up，定义一个session的开始，结束，如何整理出log file里的所有session。这题挺简单的，但可以讨论很多。

Interva相关

4. 场景忘记了，但是是可以用bfs或者union find做一些归并，整理。不难，但可以讨论很多。

follow up是内存不够了怎么做？如果用map reduce怎么merge?

1. 如何设计一个typeahead service

2. 给定一个整数数组A和另一个整数s，问有多少sub array 的和等于s

Sys design，设计pinterest的feed，pin的排序按照评分和新鲜程度综合排序，并且要求有paging功能，用户访问的时候要有accuracy（不同的page里pin不能重复出现），同时新的pin要及时feed给用户。

2. 算法，deep copy无向全链接的图，并且有存在自环

3. 算法，leetcode 20->301，followup是求最少删几个字符可以使原字符串合法。

包含呀！如果只是有(的话，不就太简单了。

我的思路是，原来代码里return false的地方，就直接删除该字符，字符串处理完最后栈里剩下都删除。这就是总共需要删除的字符数。现场跟面试官讨论了一下，基本是正确的，把各种情况都处理了

LeetCode 301

4. 职位相关知识，这轮纯聊天，问项目经历偏多，问到知识点有DB clustered idx，memory allocateor。

1.longest increasing sequence in the binary tree

2.系统设计 inverted index。。答的不好

3.distributed system

4.min stack ，要有popmin-》leetcode原题

3 two sum变种4个数凑target，o(n^2),

pow(m, n)

2. System design 他们组里一个实际业务，搭个etl pipline system

4. RMQ, 线段树

5 encode decode strategy using byte

1. 系统设计 设计home feeds

2. 平均得分top K的

如果可以限定得分精确度 比如只保留一位小数 可以开个数组表示得分(1 - 5分保留一位小数就是 0 - 50 长度51) 每个数组里面是得到相同平均评分的app 可以用个双向链表 方便增删 加个哈希方便快速定位 这样的话 增加新评分时间复杂度是O(1) 获取topK使O(k + L) L是数组长度 根据你的得分精度而定 其余思路 用个BST或者HashHeap之类的估计也可以 这都我自己想的 至于面试官期待的是哪个我也不知道

3. 有序序列 是否存在个数超过N/K个的数字

二分

4. 给个单词黑名单 和一段文本 判断这个文本是不是还有其中任意一个单词

Trie

1. HM 聊天，介绍组

2. System design a notification 类似 feed

3. Find celebrities 利口原题

4. Mapreduce 做 grouping pair

5. Alien dictionary 利口原题

系統設計：

設計一個餐廳推薦系統 (類似Yelp)，假設餐廳有textual features (name, description), image和location。使用者訊息有query, location, etc

coding 1：

next permutation

coding 2：

有一道單選題，有L個選項，拿去問N個人，把大家的答案存在一個array，這個array的size會等於N，且array element可能的值為 {1, 2, ..., L}

求大家對問題的答案有共識的概率，如果不用概率closed-form解，要怎麼寫code估計這個概率？

有共識的定義如下：

如果有大於等於N/2個人，給了一樣的答案，則定義為有共識