

● 2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 猎头 - Onsite | Pass | 在职跳槽

共五轮：

1. behavior, 米国大哥, 问了超多
2. coding, (lc:898) 中国小哥, 人很好, 树序列化非序列化, 以及面经题目, 给一个矩阵里面是1和0求第一个1在的列, 特征是在同一行里面, 只要遇到一个1, 它右边都是1
3. 系统, type ahead
4. coding, 树的最长路径, 还有一道让我想想
5. coding, 实现把一个数字翻译成英语, 比如111, 输出one hundred and eleven

● 2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 网上海投 - HR筛选 | Other | 在职跳槽

国人大哥。Implemen an iterator class for pre-order of a binary tree. 实现 next (), hasNext()
Follow up: 实现 remove, remove 要求删除next () last visited node and its subtree.

其他都比较顺利, 就是对要求把node and its subtree 从original tree中删除的要求比较疑惑。因为没有reset的API, 所以没必要保留root。为了把node 从original tree中删除, 还要额外保留root, 觉得多此一举。弄得国人大哥有点不快, 觉得没我注意听他的要求。其实知道他的要求, 但觉得没必要。后来他重复了要求, 按他得做了, 保留了root, 但时间耽搁不少。

● 018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 内推 - 技术电面 Onsite | Pass | 在职跳槽
<http://www.1point3acres.com/bbs/forum.php?mod=viewthread&tid=426961&ctid=456>

面脸才真正感觉刷题是有用的。。。基本都是leetcode原题

电面两道 lc:438, 88

onsite

1. design fb messenger

2. behavior question加一道coding, 给一堆string, 然后给另外一个target string, 里面可能含有特殊符号*, 代表可以match任何的char, 问target是否可以match那堆string里的任何一个, 感觉是利口原题, 找不到题号了 (211 Add and Search Word - Data structure design)

3. 跟一个中国人吃饭, 大力宣传了fb的好, 感觉公司文化很好

3. 利口一零

4. 判断一个string是不是一个valid数字, 不用考虑E, 还有一道是给了一个 async copy function, 这个函数可以传一个callback函数, 就是说每次async copy一个file结束就可以自动call那个call back function, 题目要求是最开始有一个文件在一个机器上, 要求把这个文件以最快速度copy到50个机器上, 还有一个限制条件就是一个机器最多只能同时copy文件到另外三个机器

Q :design messenger, 就是按照一般的chat app 来设计吗? 有啥具体要求? async 这个 是 类似于 gossip那种吗? 望LZ提供思路啊

A:就是chat app吧, as

ync这个不是design是算法题

Q: 第二轮那题是不是:211 Add and Search Word - Data structure design

●018(4-6月) 码农类General 硕士 全职@Facebook - 猎头 - Onsite | Fail | 在职跳槽

<http://www.1point3acres.com/bbs/thread-425205-2-1.html>

1. 遲到的烙印, 人還不錯, 問了一題 Kth smallest element in 2D array, 秒完就開始聊簡歷

2. 加拿大華人, BQ, 問簡歷, 然後上了二題 validate BST, next permutation

3. 加拿大韓國人, lunch

4. 不知道是哪裡人, 一直對我笑, 笑的我心裡發寒, 估計就跪在這一輪

第一題是BST iterator, 第二題是array找peak(lc:162)) 第三題給城市人口數, 求一个根据人口数找城市的随机概率

題目基本上都做過, 但剛吃飽午飯, 腦袋不靈光, 沒有應對很好

5. 白人, system design, 問設計一個裝string的cache : 設計一個cache用來裝string的key-value pairs, 給memory大小, 給number of threads, 問你怎麼設計

6. 白人, BQ, 問和以前同事相處的經驗

Q: 请问楼主第四轮最后一题是类似这个面经吗: 写一个random 输出城市名字的函数。给了一个城市的map, 格式: 城市名, 人口数。根据城市人口数的比例, 来输出城市名字。比如 map: DC 90, SF 50, LA 60, 城市输出的概率分别是: 90/200, 50/200, 60/200. 不停的掉用这个函数,按照概率来输出城市的名字 对!

Q: 不知道楼主怎么答的这个system design的问题, 我简单说说我的想法, 楼主看看行不行。首先, 看这个问题应该是一个单机的应用程序, 不涉及到分布式相关的东西吧? 然后, cache的策略是什么? LRU, LFU, FIFO? 或者其他?

因为memory大小是给定的, 所以我觉得需要知道一下key, value的平均长度, 这样我们可以知道cache的容量大概是多少。

线程的数量用来做什么, 我没啥想法, 是要问CPU需要多少CORE么? 这个number怎么用呢? 还是只是考线程并发相关的东西?

A: 给memory大小是要你计算总共可以装多少字串, 算一个大概就好

线程数量是为了要考你data sharding, 其实就是把web的那一套搬到单机上来而已, 因为楼主没有web的经验, 他们就设计了一个单机版的system design

Q: 可以讲一下根据人口数找城市的随机概率的思路吗? 我想到的是generate an array that contains each city many times based on population. 这样可以O(1)找到答案, 但是memory上不efficient。

A: 这题楼主也是差不多这样回答的, 后来被问有没有办法做到 O(1) space, 楼主就愣在那了。

move fast, be bold, hack hack hack, teamwork :)

●2018(4-6月) MachineLearningEng 博士 全职@Facebook - 猎头 - Onsite | Fail | 在职跳槽

1. BQ + coding 三姐面的 coding是李口 九耀: 聊的时候感觉谈笑风生 题也基本一次过了 但后来feedback的时候好像就是这轮被阴了

2. coding 国人妹子 稀疏向量相乘: 没做过的题 现场做的 就是用什么数据结构来存稀疏向量 然后怎么做乘法 写了4个version with 不同的O(), 后来还算满意 但只做了一道题

3. 系统设计 国人大哥 typeahead: 心中窃喜 然后刚说到qps 大哥问你是不是看过这道题 一下子怔住了不知道怎么回答 后来实诚的说 确实以前见过 然后就换了道题 是关于如何search fb的post (比如搜apple,出来苹果公司消息的post) 全程阴着脸 不苟言笑 最后这轮好像也没过

4. ML 设计 国人大姐 问一些以前的项目经历 然后问的是 怎么设计fb的page模块的 search,比如搜san jose 出来的是地名 搜lady gaga 出来的是人的page.

5. coding 国人妹子 波峰波谷 一个每次只升或降1 的array (0,1,2,1,2,3) 求所有局部最大或最小 (example里应该是2 和 1). follow up 是局部最大最小很少的情况下怎么做到log(n) (二分法) 后来也算满意过了

●面试前在论坛里看了很多帖子, 受益良多, 特来此分享面经和复习资料。

背景: 我是CV方向的PhD new grad, 职位是Research Scientist (ML) 。

(首先是非典型timeline。。。)

Timeline: 5月左右被FB HR联系, 9/12电面, 11/13 onsite, 11/17 HR约下周一打电话告诉结果, 11/20 得知结果。

.本文原创自1point3acres论坛

面经：

因为签了NDA，不方便说太细，请见谅！

电面：高频题string

onsite: . Waral 博客有更多文章,

第一轮：简历，BQ 编程题：高频题 Trie

第二轮：ML [system](#) design. Newsfeed ranking. 这一轮是论坛里讨论的比较少的一个类型。我复习的方式是先复习一遍一些ML的经典算法，然后学习Facebook的ML视频（来自FB blog和YouTube搜索）。看得多了，会发现总结起来答题有章可循的。

我自己准备的时候在白板上对空气讲过几遍，但是面试的时候其实问题会特别发散，跳来跳去的，比如我提到“训练效果。。。”，对方就问“怎么知道好不好”。但是大体上都是很常规的问题。

第三 / 四轮：Coding. 都是面经题。 . from: 1point3acres

一点总结：

Coding: 面经题特别重要。然后我的感觉是一一定要和面试官沟通，多问问题，有交流，一定会加分的。

BQ: 往FB核心价值观上拉。

ML design: 把基础打好，多了解一下工业界的操作，总结一个ML系统的组成部分和每个部分的design要点。

In general: 和面试官保持好的互动，当作一个conversation，不要太stressed out.

最后是我总结的资料。一份是coding，包括leetcode FB的题，和一些论坛上看到的面经题；另一份是ML，包括我总结的经典ML算法cheatsheet，和一些design题。其中一部分来自于论坛里的一个大神的分享，感谢大神！

[FB coding 资料 Dropbox链接](#)

[ML design资料 Dropbox链接](#)

2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 内推 - **Onsite** | **Fail** | 在职跳槽

分享个面经，3轮coding + 1轮design。估计挂在bq和设计上了。

1. 李扣 衣无酒 + 而无伞
2. 溜而衣 + 溜而衣的followup
3. BQ + 而气四
4. 设计数据库

Q：设计数据库是什么意思？就是让你设计如何parse sql query以及实现、储存？high level谈话，提到面试官感兴趣的地方会打断，让你深入谈谈？design database with high consistency or high availability？

●2018(4-6月) [码农类General](#) 博士 [全职@Facebook](#) - 猎头 - [Onsite](#) | [Pass](#) | 在职跳槽

1. 波谷游戏，用字典树，跟里口而要而一样
2. BQ，你最自豪的是什么项目，跟同事，跟上司的conflict，你觉得你最大的失败是什么. 一亩-三分地，独家发布
3. 系统设计，fb的聊天，这轮感觉很一般，lz没有什么工作经验，尤其是很久没用TCP了，被问的很纠结 (facebook chat, support 100k user, don't need to keep history, one to one chat, 对于一个server到底能support多少个websocket追问了很久)
4. 双链表转二叉平衡树，写了 $O(n\log n)$ 和 $O(n)$ 时间复杂度的两种方法
5. 面经里面的括号那道题，跟里口不同的是，只需要输出一种结果就可以了。第二题是a的b次方 (里口三菱要的简化版本 lc:301)

然后。。。recruiter反应很迅速，说反馈都是positive的，第三天就送committee了，然后崩溃的是。。。committee要求加面一轮系统设计。晴天霹雳啊，带着忐忑悲壮的心情去了，运气很不错，面到短网址那道题，以前准备过，重点问的是如何把长网址换算到短网址，这个追着问了20多分钟，lz给了随机产生，MD5 hash和UUID三种方法，然后就过了。

Q: 最大的失败：其实这个我没准备过。我刚开始说以前做project有错误，他说这个不算failure。后来我觉得他想要big picture那种失败，就说当年做某个选择的时候，太过于保守，选择了稳定的方向，其实应该taking risk, blablabla

Q: recruiter和我说要面distributed system design 请问这和普通system design 一样吗 感觉你的两轮design和distributed system没什么关系啊

A: 最后都会发展到distributed system design上面的，多个server，多个DB，一般都是从单server开始，给一个可行solution，然后再多个server。会要求你给出high level design，就是画个图，根据那个图问你加个DB你怎么保持consistency这种，如果server crash了，你怎么处理

●2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 内推 - [Onsite](#) | [Other](#) | 在职跳槽 <http://www.1point3acres.com/bbs/forum.php?mod=viewthread&tid=421125&ctid=456>

hr安排了店面：number of islands, follow up 输入数据量很大该怎么办~

. Waral 博客有更多文章，

然后约了一周之后的昂塞：

1. lc:143，楼主[刷题](#)不精没做过，当场现做，还被好几次友情提示。

.留学论坛-一亩-三分地

2. 设计题，内容有些长，差不多就是给你了一段代码，你先指出哪里不太好。接着就问我该从什么地方修改 (比如设计那些api, 数据库, server怎么搭建)跪

午餐时间

3. 这一轮是BQ，然后出了一道比较简单的题，给你一个无序数组，一个target value, 如果有连续的任意个数加起来等于target，就返回true. (560?523?)

4. 给了一个map, key是联系人, value对应的是邮箱list。如果任意两个人有相同的邮箱, 那么他们就是一组的(如果A和B有某个相同邮箱, B和C也有某个相同邮箱, 但是A和C没有相同邮箱, 那么ABC也是一组的)。返回所有的组。同样跪, 又被友情提示了好多。(lc:721)

Q: 第四题是 期儿一 吧? 第三题有负数吗?

A:第三题应该有负数, 没问~汗. 第四题给的key是联系人, 所以不会重复, value是这个联系人的一些邮箱。看上去和其儿一有些像, 但是差别还是很大的。

+:有负数好像就只能 $O(n^2)$ 了, 不然全是正数的话就是一路加, 加到sum超过target就减掉第一个数, 是这个意思吧?

第四题我感觉key重复不重复好像也没啥关系。union find还是根据email来建的。利口那道题也不关名字啥事, 就是最后输出的时候用一下

+: 我觉得第三题应该还有限制条件, 如所有的数为非负整数什么的。否则如果输入有正有负, 那就没法提前终止而必须看到数组最后一个数。要是有这个限定条件, 算法就会简单很多

第四题 按联系人做union find 联系关系用hashmap+联系人ID来取得

比如 1: a@b.com, b@c.com. more info on 1point3acres

2 :b@c.com, d@e.com

hashmap 存成{ b@c.com :{1,2}, a@b.com:{1},d@e.com{2}}

之后每个key - value pair 在hashmap中找union关系 就可以了

Q: 电面follow up。数据量大到什么程度呢? fit 不进 memory 还是说 只是比较大 但是 dfs 可能会 stack over flow?

A: 具体我没有问有多大, 我觉得他只是想考考我有没有多个server同时处理的概念吧。我只是给他随便扯了一下, 因为我也不是很了解

●2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 猎头 - 技术电面 Onsite | [Fail](#) | 在职跳槽 西雅图!

店面: 国人面试官

无序数组找第K大的数字, 我用的PQ。

写一个BT的iterator, 要求顺序是后序遍历。最优解纠结了一下最后没写出来, 看着没时间了赶紧写了个先递归然后把数字存进list的方法。以为挂了, 没想到让我过了。感恩。

onsite: 两轮coding, 各45分钟。两轮system design, 也是各45分钟。中间有一轮behavior + coding。然后还有一顿中饭。

第一轮, 一个类似merge sort的题。做完看着还有点时间, 加一题, 也是BT的iterator, 这次是中序, 必须一分钟解决了。

第二轮, LC三药的变种, 不是求next, 而是求previous, 一个套路, 仔细解就好了。

第三轮, behavior + coding。45分钟问各种BQ, 没什么特别有印象的, 比较常规。最后15分钟给我来了道LC题号石榴的平方根的那变态题, 但是简化了, 两个数组长度一样, 写了个大概最后recruiter跟我说这轮是过了的。

第四轮, design一个系统, 客户上传文件, 假设我们有一个service X可以take 文件, 返回这个文件是不是有malware。这题主要关注几个点, 一是service X可能会需要比较久的时间来判断, 二

是尽量减少调用service X的次数，还有就是怎样通知用户检查的结果，因为是async的call。这轮过了。

第五轮，design类似yelp的系统，这题我准备过，数据结构用的quadtree。三哥面试官一上来就没打算让我过，无交流，玩手机。

Q：能不能讲一讲详细讲讲第四题怎么design？

A：并没有什么标准答案。async call那块可以让用户隔断时间去check结果是不是已经出来了，因为要消耗很长时间，一般的http connection不能维持很长时间。或者用long pulling。感觉面试官对于第一种方法也满意。

Q：LC题号石榴的平方根的那变态题 //也用PQ，行不行？

A：??? PQ怎么解？复杂度如何？面试官给完题目直接就让我用最优的二分解法了

●2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 内推 - [Onsite](#) | [Fail](#) | 在职跳槽

1. 第一轮behavior，问了最有挑战的项目，和同事发生conflict怎么办，同事有没有给你提过特别constructive的建议，最后一个没说太好，没准备过这个例子。。。。

2. 美国小姐姐，第二轮system design，楼主之前做的 iOS，所以直接问的 mobile design，感觉这轮面的最好。。。。。。问了个如何设计机票app

我来说说我怎么设计的：

(1) 先确认需求

(2) concept design，work flow 如何设计，比如user 先把 schedule (to,from, date) 等信息发给服务器，服务器返回一个flight lists，然后user 再从里面选一个flight，以及seat，还有一些extra information 返回给服务器。。。。等等

每一步要考虑一些请求，比如user 想back 怎么办，session 过期怎么办，等等问题

(3) high level design，网络层怎么设计，怎么handle 各种错误，比如http code 的错误，和业务逻辑有关的错误等等，有些请求发送前是不是要在本地数据库存一下，这样请求发送之后，万一失败了，我们可以不通知user 的情况下retry 几次

(4) 到这一步应该会问你 dive deep 到一些模块，比如用什么类啊，core data 怎么用啊，NSURLSession 怎么用啊之类的。。。。

. from: 1point3acres

3. 美国大叔，问了个clone graph 的问题。。。。楼主头一次写白班，真的不适应，写的很凌乱很凌乱。。。。最后没时间了，问了个followup 就随便问了个问题，然后没时间写了。。。。

4. 国人小哥，问了个如何clone tree，这一轮答的不太好，不是因为题难，是因为我想用java 写，写到一半发现小哥很多follow up 和 iOS 只是有关。。。。希望小哥你看到的话，高抬贵手。。。。真的是怨我最开始没说清。。。。

这轮问了下在算法的基础上问了很多和block 有关的知识。。。。

5. 最后 shadow 轮，这轮面得惨不忍睹，脑袋不转了，问了个 两个NSRange 相交的问题，BST 转成一个circular linked list

● 2018(1-3月) [码农类General](#) 本科 [全职@Facebook](#) - 内推 - **Onsite** | **Fail** | 在职跳槽

- BQ + lc301 只需要一个结果
- lc125 + merge sorted iterator
- find anagram 就是sliding window解的那题 + find celebrity：题目是说有2个string, s1, s2, 找到s1中包含s2所有字母的最长substring
- POI：POI是什麽意思呢? point of interest? yelp?
- typeahead

上个月onsite的，面试的时候表现完美基本是bug free，今天recruiter通知说挂了，原因是coding太好，所以认为是见过题了。。。看来还是欠缺经验啊，同学们引以为戒，面试的时候尽量表演下

2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 内推 - **Onsite** | **Other** | 在职跳槽

过了一周了，还没出结果，上地里发个面经攒攒人品。

三轮：

1. bq轮，就聊聊最近做什么，处理人际关系，最讨厌和什么类型的人工作之类的。后面设计了一个简单的ood，一个总的receipt，里面有n个子note：很简单的，就是想象一本食谱书，里面有n个食谱，每个食谱都有不同分量的材料需求，最后返回一张购物清单去买材料，能够做所有的食谱。也不算严格的ood，就是自己想存储结构，然后写一个method去返回需要的结果。

楼主是怎么做的呢？我的想法是用一个hashmap，key存食谱，arraylist存相对应的材料。

2. receipt，然后每个receipt有n种材料，最后返回一个整理好的需要各种材料各多少的list

2. 2道题，一道题是680，另外一道忘了。。。

3. 在一个无序的数列里面，找到所有的subsets的min和max的和小于等于target，返回subsets的个数。可以优化到 $O(n)$ 。

补充内容 (2018-4-25 02:22):

第三题要先排序，所以复杂度应该是 $O(n\log n)$

给的无序数列可以排序，应该是 $O(n\log n)$ ，然后指针就可以做，程序主体部分 $O(n)$

第三题应该是LC的，排序后双指针从两头数，代码看着简单，想出来很难

[color=rgba(0, 0, 0, 0.87)]这是之前的一篇帖子关于FBI的首轮店面<http://www.1point3acres.com/bbs/forum.php?mod=viewthread&tid=345880&page=1#pid3695429>

[color=rgba(0, 0, 0, 0.87)]总的来说，我自己准备得一般，但运气比较好，给的题都做过，没做过的也不难。

[color=rgba(0, 0, 0, 0.87)]behavior和design比较自信，面完复盘觉得还是有一些答得不算特别好，但还好面试官手下留情。

[color=rgba(0, 0, 0, 0.87)]coding准备：主要就是LeetCode， onsite之前，花了几天时间把LeetCode上FBI tag的题全部重做了一次，并确认掌握了最优解。

[color=rgba(0, 0, 0, 0.87)]design准备：我准备主要靠得这些资源，

[color=rgba(0, 0, 0, 0.87)]1. youtube上的[system](#) design视频

[color=rgba(0, 0, 0, 0.87)]2. <https://www.educative.io/collection/5668639101419520/5649050225344512>

[color=rgba(0, 0, 0, 0.87)]3. <https://www.interviewbit.com/courses/system-design/>

[color=rgba(0, 0, 0, 0.87)]4. <https://github.com/FreemanZhang/system-design?files=1>

Onsite:

[color=rgba(0, 0, 0, 0.87)]round 1 : behavior question + coding (里口 柳儿依 简单版)

[color=rgba(0, 0, 0, 0.87)]round 2: 里口 二分图， interval简单变形(非里口FBI题)

[color=rgba(0, 0, 0, 0.87)]round 3: POI， 问得都不算了难，但我主动扩展了一些问题。

[color=rgba(0, 0, 0, 0.87)]round 4: 里口 齐八武， LRU cache

[color=rgba(0, 0, 0, 0.87)]round 5: Instagram， 问得都不算了难，但我主动扩展了一些问题。

2018(4-6月) MachineLearningEng 博士 全职@Facebook - 猎头 - HR筛选 | Other | 其他

一共四面 5月 24 lint819

1. BQ问题, 和上司的冲突, 过往的projects。里抠要散救 139
2. ML Design, spam ad detection。感觉这一轮出的题会跟面试你的人的背景有关。
3. 里抠吴遛, 里抠酒吧 56, 98
4. 里抠伞要要, 里抠旗遛 311, 76

2018(4-6月) 码农类General 博士 全职@Facebook - 猎头 - Onsite | Other | 在职跳槽 面试时间: 6.14 @西雅图

除1, 5轮外, 都是国人(包括午饭)。都非常 nice, 交流也都很好, 挂了也只怪自己学艺不精, 只是还是想攒点人品, 只因心存幻想。

1. 给一数组, 和最长容许的subarray长度, 依次找出所有subarray的最大值存入返回数组中。
2. BQ + rotated sorted array, 找最大值位置
3. 设计题, FB 朋友在线状态功能设计
- 4.1 判断输入 string 中括号是否匹配
- 4.2 0, 1 矩阵中, 找最左边的1的位置
- 5.1 (shadow round) 给一个新的字母表顺序, 判断给定的字符串数组是否是排好序的(按照新的顺序)
- 5.2 Follow up: 如果给一组排好序的字符串, 如何找出未知的新字母表顺序。只讲思路, 没写代码。

2018(4-6月) 码农类General 硕士 全职@Facebook - 猎头 - Onsite | Pass | 在职跳槽 面试时间: 五月初 <http://www.1point3acres.com/bbs/thread-429215-2-1.html>

第一轮聊职场聊人生。

第二轮: 1. Union of intervals: 给定两个list, 每个list都包含若干已经sort好的intervals, 计算所有interval的union, 用一个新的list装下并返回。

[(1,3), (6,7), (9, 12)].

[(2,4), (8,10)]

返回 [(1,4), (6,7), (8,12)]

2. 在grid中找最左侧的1

类似如下grid：每个横行只要出现1，右边的数字都会是1。例子中的最左侧的1是倒数第二行第二列的那个。

```
0 0 0 0 1 1
0 0 1 1 1 1
0 0 0 0 0 1
0 0 1 1 1 1
0 1 1 1 1 1
0 0 0 1 1 1
```

第三轮：

手机按键上跑国际象棋的马

```
1 2 3
4 5 6
7 8 9
0
```

马只能跳日字，马从1开始跳，跳n次，问总共能有多少条路径。

我先给出了dp解，然后让我用bfs/dfs再写，然后优化，分析复杂度

note：跳马方向没有限制，只要你按照日字能跳到的地方都算是能走的路径。

请问一下楼主第三轮可以跳到之前有跳过的点吗？返回值就是一个int就可以了是吗？

是的，来来回回跳都无所谓。就返回个数值，不需要记录路径。

第四轮：

系统设计points of interests

1百万个interest points记录，实现像yelp一样的那种根据地点返回最近的五个interest point。

着重分析如何对points进行建模，参考geohash，网上有相关文章

第五轮：

完全开放，自由发挥，不领着你，让你自己去分析。中途时不时我会问要不要继续分析某个东西，他会说你继续。。。。

我侧重的是数据存储，读取，数据库设计，HA，load balance相关的东西，feature在一开始小分析了下后，主要用来帮助进行数据分析了。

Q&A

Q:HA 是指什么？instagram 的数据储存 指的是图片的方面还是 text 的content？如果是图片感觉数据库没有太大发挥的空间？

A:HA是high availability。你说的数据库存储应该是指如何建表或者如何存图片信息？我讨论的数据库是如何sharding。

Q: 我指的是 存储 因该会需要图片 想请问图片在sharding 跟index strucutre 上 有跟一班的 text document 有不一样吗? 因为一班图片较大 不知道interviewer 有无要求 or 楼主怎么答的?

A:面试重点不在你关注的图片或文档存储的方式, 而是存到哪里。你要就instagram量级的图片数据量进行分析, 并理解instagram用户的读写需求, 因为这些影响图片存在哪个server中能最优化performance。举个例子, 不同用户上传图片, 如果系统对用户主页读取响应要求很高, 是不是某个用户的图片最好能存储在同一个db server内? 如果系统对新鲜数据最感兴趣, 老数据没必要低latency展示, 是不是最好新鲜的数据能存储在统一的区域内? 这些问题才是系统设计关注的东西。

Q: 着重分析如何对points进行建模

A:是的, 划分区域的算法

Q: 您能说说m+n的思路么?

A:要求只需要找到全局最左的1, 所以可以记录第i行的最左1所在位置j, 然后看i+1行j的位置是否为1, 如果为1, 则在i+1行继续向左找新的最左, 如果为0, 则i+1行不可能提供最左1, 可以直接去i+2行。

你观察一下, 这个算法处理的点其实是从右上角到左下角的折线, 是 $O(m+n)$ 的。如果还要优化, 则利用你说的Binary Search的方法, 向左找新的1的时候, 使用binary search。

Q: 楼主跳马能给说说思路吗!

A:先用hashmap记录下来从某个位置能跳到哪些其他位置, 可以二维dp (简化为1维) 推导, 记录第k次移动move到手机i位置处总共有多少种路径。或者bfs/dfs运算中使用memorialization记录跟dp相似的信息。

Q: 想问下楼主POI这个题目, quadtree和geohash如何选择的? 感觉应该两个都可以, 但是一轮面试是不够说两个的。如果面试官有偏好的话应该就是展开说他想要你说的那个, 但是GeoHash解释起来复杂一些。

A:我的理解: 从算法层面去谈, geohash本质是跟quadtree一样的, 或者说geohash是quadtree的一种应用, 都是利用空间填充曲线把二维信息转成一维。从实际使用角度去谈, 你也看到了geohash带着hash这个字眼, 这个是影响你数据库使用的, key-value的构建, indexing和B-Tree查询, 进行分布式存储等等, 加上很多语言提供了原生支持, 所以业界geohash大热, 是因为方便使用。如果面试官比较侧重算法方向, 你估计要答的就是网格索引到quadtree甚至到r-tree, 但是如果涉及存储和index相关的事情, 我感觉geohash还更好说点。

2018(4-6月) [码农类General](#) 博士 [全职@Facebook](#) - 内推 - [Onsite](#) | [Other](#) | 在职跳槽 面试时间: 六月初

题都不算了, 有2轮都是国人, 挺照顾我的。

1. 纯behavior,聊足了一个小时

2. 利口 气流. 76

3. System design Facebook messenger, 这轮是国人大哥, 非常nice。开始先用中文和我讲design 的基本要求, 让我放松。感谢!

4. 图的连通问题, 一个dfs 就解决, 不过我答得不算好, 最后还有一个小bug, 虽然我已经发现并指出问题了, 但时间来不及改了。

5. 利口 以儿流 126

2018(4-6月) 码农类General 本科 全职@Facebook - 内推 - **Onsite** | **Other** | 在职跳槽

基本全是高频题。。可惜有道非lc题我写的并不对, 面的时候被抓住bug一直纠结不出来, 导致没时间问第2题。教训就是还是要code review一下看看有没有更好的解法。并且整个过程都会要求人肉test run, 非常花时间

1) BQ conflicts, 项目, 最后5分钟单调递增 是判断一个数组是否是单调的

2) Coding 1: 高频任务安排。就是挂在这题上。不要求排序的版本, 要确保那个queue的使用能提高空间效率, 光跑过自己写的test case还不够 - - 最后时间都花在跑test case上, 做完这题只剩10分钟, 连可改顺序的扩展都没问就让我问题了。跪。。

3) 午饭瞎扯

4) system design 1: 都在瞎扯, 提了几种方案, 感觉都行, 聊了最后什么结果都没。。也不知道他们怎么写feedback

5) system design 2: 这轮彻底挂了, 花了好多时间在client和API上, 而本意是要考察 distributed systems. 还是要自己掌握时间控制节奏。

4) Coding 2: 面了三道忘记内容了, 都是lc高频

5) Coding 3: 面了三道忘记内容了, 唯一记得是有移动零, 但是要求是minimum write, 和lc不一样。

note: 他要求可以是不保持顺序的, 并且其实只需要把非0的部分移动到左边, 右边可以不管, 返回最后非零的长度就可以了。比如 {1, 0, 2, 0, 3} 操作以后应当变成 {1, 3, 2, 0, 3} return 3 只需要1次write.

拿到offer了回馈一下 system design:

一定要按照

1) clarify requirement

2) understand scope (qps, datasize, estimate resources)

3) high level design (画画图, 分几个步骤)

4) API and Data schema

5) deep dive / bottleneck

这个顺序来，自己控制时间。楼主在讨论需求和scope的时候就开始讨论这个scale下可能碰到的问题，一发散就把大框架忘了，最后剩下7分钟的时候被提醒，赶紧把大框架火速说完，非常狼狈

两道题是 1) 给spotify设计most listened songs的功能。2) facebook chess game

Coding 真的不记得了，一共就100-200道高频，全都刷烂就对了。这次找工一圈面下来，感觉FB的面试是刷题收益最高的。其他家的都没什么用。

2018(4-6月) [码农类General](#) 本科 [全职@Facebook](#) - Other - **Onsite** | **Fail** | 在职跳槽

1. 经理，BQ + 而其把 - 我tm居然把mid写while外面去了，然后他撇了嘴说你这有个小bug自己找找，然后我看了两分钟改回去了，讲真，我没想到这是挂的理由，这经理和我都是前一天晚上飞了六小时来的湾区，明显很疲惫的感觉

2. 国人，而异唔 两种方法说了说挑heap做

wood cut 二分法的题 (就是给你一堆木头的长度woods，和一个常数K，求把这堆木头平均切成K份的最大长度，二分[1, max(woods)]就好)

问了十几分钟问题

3. 国人，易而思 没啥可说的

易齐 做完问我复杂度，我说 3^n ，然后他问我怎么分析的，然后分析了半天把自己绕进去了，他告诉我应该是 n^3 ，然后问了十几分钟问题，问完我"恍然大悟"说我明白为啥是 n^3 了，hmm其实我不明白 :(

coding轮的feedback是我coding good，要理解为啥这样实现...我估计是评价栏有好的fb和坏的fb，recruiter就把需要提高的部分说了说

4. 国人，设计：脸熟用户上传心情status - 高兴、快乐、悲伤之类的，要求1。是能上传 2。是用户输入语句能联想心情关键词 感觉是结合了typeahead、存储和mem。这轮挺纠结的，feedback只说要加强对存储和mem的理解，overall应该不是neg

5. 白女+影子，设计：爬虫去爬Wikipedia所有页面，能做到一个链接只读一次，感觉是结合短链接的设计理念来谈

note: 第五题设计考察的是分布式处理，跟tinyurl无关。涉及hashing和distributed file system, distributed cache, 集群管理, high availability,

failover一系列问题。可以搜索distributed web crawling, 网上有一些文章在讲.

楼主回复: 你说的都问到了

2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 猎头 - [Onsite](#) | [Other](#) | 在职跳槽

一天下来5轮:

- 1) career/behaviour
- 2) coding
- lunch--
- 3) [system](#) design
- 4) coding
- 5) career/behaviour #2

干货: 蠡口 山药, 无私伞, 还有一个变形没有的

SD: 经典的高频 instagram feeds

整个下来觉得普遍比较nice, 无论后面怎么写. 不会当面斥责你. 我实际上5年前面过, 但是相比这次, 之前那个就是没准备.

基本的感想是:

后来recruiter有给我feedback,基本上每一轮都没谈好. 认栽, 是我自己不好.

基本上就是, behavior就是提前要准备好具体例子. 我太随意, 没准备好.

leetcode嘛, 常刷.

design嘛, 常聊, 常看youtube. 论坛

- coding 部分, leetcode还是好好刷, 未必是原题, 有原题, 有原题变形加条件. 但都不是用来刁难你的. 我做不好是我的问题, 没准备够. 地里的牛们是稳稳的. 把leetcode刷够归纳总结一下类别应该问题不大.

- system design部分. 就是网上的超高频题. 感觉和朋友同事练习一下还是很有帮助的. 但是基本还是得自己多练几遍. 主要是练一下被面试官打断思路怎么有序地重新组织. 因为虽然是同样的题. 但是每个面试官的兴趣都不一样. 你准备的未必就能用上, 但是细节都是类似的, 准备的不会浪费. 至少不慌.

system design 油管上有几个tutorial讲得是很有道理的, 明确告诉他你会什么不会什么. 自己也不会慌, 你听到他说他这块也不熟的时候, 你也知道用什么语言和他沟通, 不用瞎猜, 还是反过来不慌.

- 面试前一晚在酒店喝了个小酒和bartender聊一下, 他提醒了一个, 我觉得对我有帮助. 就如果面试官面无表情, 那是他们训练出来的, 来stress test 你的. 无论事实如何, 保持这个想法, 心里还是会沉稳.

2018(4-6月) **码农类General** 本科 **全职@Facebook** - Other - **Onsite** | **Fail** | fresh grad应届毕业生

第一轮, shifted sorted array search. 给了一个最优解, 国人小哥 (国人小姐病了, 临时换成了国人小哥) 说了句good, 接下来, 一个从未见过的题, 应该用dfs做, 我不知道咋描述具体内容, (实在听不太懂国人小哥的英文) 对比并找到了最好的做法, 但是没时间写了, 可惜 (没听说过那个概念)。

第二轮: 吃饭, 一小时, 国人小哥, 问我怎么还有力气问问题。

第三轮: 一道move zero, 必须用最少的swap, 解决之后有个xiao bug (out of bound error) 在提醒下, 纠正。get random city, 数学概率题, 但是我在想有没有更好方法的时候, 国人姐好像觉得我不太会, 一直想提示我, 我的习惯就是在作的过程中筛选方法。后来问能不能优化, 由于一部分哦 $O(n)$ 第二部分 $O(n)$ 或者 $O(\log n)$ 怎么样都是 $O(n)$ 为啥还要binary search 每明白, 最后一秒想到了方法描述了没时间写第二个binary search, 但是我脑残, 把自己linear search的代码擦了, 真是醉了。

第四轮: 后来abc 大哥, phd, 大聊, 没时间code 了, 说我总共做了6到题了, 也不需要code了, 结束了。

4/19店面

第一个题是 蠡口 耳扒伞, 用了同向双指针, 然后让最小write次数, 我用了相向双指针。问时间复杂度, n

相向双指针的时候, 三个肉眼手跑例子, 他自己把while看成if然后说我错了, 后来我有点懵逼, 就改, 墨迹了几分钟, 然后我坚持用原来的代码,

4/17 onsite.

三轮：

1. bq轮，就聊聊最近做什么，处理人际关系，最讨厌和什么类型的人工作之类的。后面设计了一个简单的ood，一个总的receipt，里面有n个子receipt，然后每个receipt有n种材料，最后返回一个整理好的需要各种材料各多少的list
2. 2道题，一道题是680， 另外一道忘了。。。
3. 在一个无序的数列里面，找到所有的subsets的min和max的和小于等于target，返回subsets的个数。可以优化到 $O(n)$ 。

补充内容 (2018-4-25 02:22):

第三题要先排序，所以复杂度应该是 $O(n\log n)$ 二个题是给一个数组，问有多少个集合，满足这个集合中的最大最小值之和小于k，k是输入的一个参数。

2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 内推 - [技术电面](#) | [Other](#) | fresh grad应届毕业生

给我电面的是个萌萌的伊朗小哥，虽然口音有点捉急，但不影响总体交流。

第一题：哈希

Input:

A seq of int <3, 4, 1, 2>

A target int 5

Output:

whether a continuous subseq sums up to the target

target 5 -> True

target 4 -> True

target 6 -> False

follow up: 分析时间复杂度，空间复杂度，改成输出最短长度的subsequence的长度，以及有什么test case

第二题：拓扑排序+BFS

n jobs 0 .. n-1 single processor.Executing each job taked one unit of time i, j -> before starting i, we should have finished j Input:n

<(i,j)> list of job dependencies

Output:

min req time for finishing job index 0

Input:

6

1 2

1 3

4 1

0 1

4 5

Output: 4 (3, 2, 1, 0)

2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 猎头 - **Onsite** | **Pass** | 在职跳槽

1. behavior, 米国大哥, 问了超多
2. coding, 中国小哥, 人很好, 树序列化非序列化, 以及面经题目, 给一个矩阵里面是1和0求第一个1在的列, 特征是在同一行里面, 只要遇到一个1, 它右边都是1 (lc: 898)
3. 系统, 泰普额害的
4. coding, 树的最长路径, 还有一道让我想想
5. coding, 实现把一个数字翻译成英语, 比如111, 输出one hundred and eleven

2018(4-6月) [码农类General](#) 硕士 [全职@Facebook](#) - 内推 - **Onsite** | **Fail** | fresh grad
应届毕业生 SWE 岗位

电面: Task schedule, 带两个 follow up, 电面时候还没刷到, 自己弄出来了, bug free 跑通。

Onsite:

1. 国人大叔, Sliding Window Maximum, 没有 FB 的 tag 所以我刷题时候就看了一眼没手动写。
我开口说这个特别像 CNN 的 maxpool, 换来冷漠脸。。。我記得有个 $O(n)$ 算法, 但是我觉得当场写不出来。给出了两个 $O(n \log k)$ 的思路, 一个基于 BST, 一个基于 Heap。面试官说都不好, 你有没有想过 PriorityQueue。。。我说我用 Heap 就是把它当 PriorityQueue 的。这时面试官说 $O(n \log k)$ good enough。写代码吧。
写了一半面试官对我的 PriorityQueue 的实现很有疑问, 一度问出 "PriorityQueue 一般用什么数据结构" (我说 Heap), "Heap 一般用什么实现" (我说 Array)。
然后就让我解释如何在 MaxHeap 里移除 Max 元素, 添加元素等等, 这种基础的 Heap 操作问题。说我操作不对, 我又给了一个具体例子, 他最后说 okay 这个可行, 浪费了超过 15 分钟。

我知道 Heap 有多种实现思维，但是我发誓我用的就是 MIT 教材的实现思维。

这时候我已经知道这一轮挂了，不过淡定写完了主代码，PriorityQueue 有几个函数没时间实现。面试官说不要担心（鬼才信）。

2. 亚裔姐姐，culture fit。

此外问了一道 BST 相关的题，不要求写代码，细节我忘了，我给了一个递归思路。

3. 印度小哥。(lintcode 的捌亿玖 819

merge ranges

还有一个字符串排序题：排序 "ABC" "CAC" "AB" 但已知 $C > A > B$ 这种。

写了一个 $O(nk)$ 实现

都一口气写成，剩下很多时间就闲聊。

(第一轮应该是挂了。第二轮亚裔姐姐竟然看着我的简历说出了：Oh you have REAL work experiences。这样的话，估计之前也没看我简历。第三轮应该没问题。虽然有 4 年工作经验，但是按照 new grad 标准的面试，只有三轮，没有设计轮。朋友强推换来的面试，没有去 argue。已通知挂掉，recruiter 拒绝透露 feedback。改年再试吧。

教训就是还是要刷题，FB 根本不看你背景)

Facebook：再次回到西雅图，考虑到living cost，特别选的西雅图office。结果没有持续刷题，和朋友在弯曲吃喝一个周末，状态差了很多。上去第一轮碰了个原题（我明确的记得leetcode contest时候我做出来了。。），没做出来。第二轮面behavior加一个简单coding，但是面试官最后持续问你觉得有没有错误了。。然后就彻底崩了。中午吃饭时候，中国小哥一直安慰我，还好下午两论都很友好。。。但是以为肯定挂了的。结果过了3天给了加面，特地推了一个星期，好好刷了4天的题，然后基本算是秒了,45分钟做了3个题，但是还是非常没有底（加面前一天，在地里看到说加面通过率极低），不过实际上和我一起的另外一个，我们都加面了，也都过了。

2018(1-3月) [码农类General](#) 博士 [全职](#)@[Facebook](#) - 内推 - [技术电面](#) | [Other](#) | fresh grad应届毕业生

一轮：看不出国籍拜仁但口音不好理解，里口股票买卖原题一二（还好没问三四），二要简单数学证明为啥贪心策略是对的

二轮：冷漠国人大哥，花了巨多时间相互介绍导致没时间写代码。。。深刻教训。。。第一题是常见题，每行都是先0后1（比如00011111），输出0出现最早的那一行；第二题给两个BST，要merge输出，要求只能写空间复杂度最优的iteration的算法，就是用两个stack分别对两个BST同时做inorder，问这题时就10min了，疯狂敲完code结果是bug一堆，于是就挂了

4. 可爱和蔼的印度小哥。。input是一个string，输出所有的palindrome substring的count，重复的也需要计算；一个tree多了一条边变成了graph，找到挂的那个边，和利口上面那题很像，但是input是root node

5. 判断两个input tree是不是similar, similar 的定义是两棵树任一结点可以通过swap左右子树后相等。

●感觉地里加面经特别少。LZ 四月五号经历了一场堪称是人生创伤后的面试。感觉自己挂定了，一周后拿到了加试email，加试coding

1.Common Lowest Ancestor

不一定两个node都在tree中。LZ没注意这点，最后还补写了个find，看看这个node在不在其中。

Time Complexity $O(N)$.

Space 如果算栈那就是worst $O(N)$ best $O(\log N)$.不算 $O(1)$

follow up:如果你可以修改tree。缩减time complexity你要怎么做？

LZ解法是每一层追加一个List<TreeNode> child. 然后存全部的child和自身。

此时因为每次我们只用找半边，time 缩减到 $O(\log N)$ 。

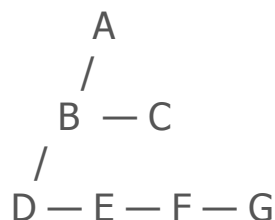
但是space 最差就是 N^2 ，最好就是 $\log N * N$ 。

2.

给定一个tree

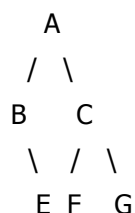


将树变成：



树的right point to sibling.

如果出现：



那么对B来说， $B.left = E$ $B.right = C$

LZ给的解释BFS+hashMap存每层，然后逐层修改。同时注意获取本层时要获取下面一层的leftmost，并且将本层的leftmost.left 修改成下一层的leftMost。

最后问问题的时候，LZ除了问面试官 general question，还问了你觉得我哪里还要再改进。面试官直接说第二题你应该找更简单的方法去做。

希望看到第二题的同学们认真读一下题，**tree的数据结构里没有next指针**。并不是利口原题。

2017(10-12月) **码农类General** 博士 **全职@Facebook** - Other - **技术电面** | **Other** | fresh grad应届毕业生

店面前刷了tag里面所有的题目，然后两个都不在tag里面。。。

第一题是给一个排好序的数组，然后算出来这个数组排好序后的平方数，给的例子全是整数，我一想这也太简单了，几秒钟就写完了，然后还说我假设这些数字都可以用32个bit表示。。。转念一想有可能有负数。。。然后问了一下，然后interviewer说对的，各种integer都有可能！但是很快也就写完了，我自己过了一个例子，感觉没问题。

然后就开始**第二题，判断一个数组是不是经过最多一次swap就变成完全排好序**。开始的时候函数的signature是interviewer写的，然后里面写着const, 然后纠结了很久的const，然后就问他，我可以改这个数组吗？对方说可以，然后我就把那个const删了，然后开心可以改数组的时候，把原来的思路打断了，然后总共两种case（交换的数字是挨着的，和不是挨着的），然后忽略了一个case，我自己跑test case的时候还没发现！因为还有很多时间，然后后来改好了，最后其实还有把名字写错，唉。。interviewer指出这个case的时候，有安慰说还有很多时间。。。感觉第二题我应该先把brute force的方法说一下来得，

- ML是做System，工作内容是设计高效的Data pipeline以及稳定的productionize models。
- DS是做modeling，工作重点是将business question转化成ML model，并持续改进模型。

Recommendation system: Netflix 如何推荐, 亚马逊产品怎么推荐, Fb的Newsfeed怎么排序等等。这些问题面试ML SDE的时候也会问到。

●machine learning design的问题一般都是问设计一个推荐系统, 广告排序, 还有一般的监督学习的系统。所以准备起来要了解各类的推荐系统的优缺点, 以及最新的embedding的方法, 可以看一下这个博客, <https://medium.com/the-graph/applying-deep-learning-to-related-pins-a6fee3c92f5e>。

现在为止, 面的两家pinterest和facebook都是围绕着推荐系统展开的, 中间会问到各种小问题, 比如feature提取, model不够好怎么办。

另外ML的问题, 建议最好按照sd一样, 按照逻辑点进行回答。

1. 厘清问题, 该问题属于什么类型的machine learning问题, 比如监督性学习, 比如推荐系统。
2. 明确输出目标, 明白该问题的目标是什么
3. 数据收集, 了解一下都有哪些数据可以给我们利用
4. 数据特征提取, 把你想到的特征进行分类, 都有哪些domain的特征可以进行提取。
5. 模型, 有哪些模型可以使用, 然后点出各个模型的优缺点是哪些
6. 评估, 你怎么进行模型的评估好坏, 评价指标是什么, 还有怎么改进。
7. 画出框图, 然后进行优化。