Trie

平衡二叉树 insert interval

Union find # of island

[**https://www.evernote.com/shard/s424/client/snv?noteGuid=2804d7e0-c93c-455d-9b79-027caa1750c5&noteKey=48c247d32264b0e5739c5fc27c7eaff1&sn=https%3A%2F%2Fwww.evernote.com%2Fshard%2Fs424%2Fsh%2F2804d7e0-c93c-455d-9b79-027caa1750c5%2F48c247d32264b0e5739c5fc27c7eaff1&title=dropbox-algorithm**](https://www.evernote.com/shard/s424/client/snv?noteGuid=2804d7e0-c93c-455d-9b79-027caa1750c5&noteKey=48c247d32264b0e5739c5fc27c7eaff1&sn=https%3A%2F%2Fwww.evernote.com%2Fshard%2Fs424%2Fsh%2F2804d7e0-c93c-455d-9b79-027caa1750c5%2F48c247d32264b0e5739c5fc27c7eaff1&title=dropbox-algorithm)

[**https://www.evernote.com/shard/s424/client/snv?noteGuid=abe1558e-b512-4ff0-929b-f305221166e0&noteKey=3feb3686c7b488d341ddfec1629a836f&sn=https%3A%2F%2Fwww.evernote.com%2Fshard%2Fs424%2Fsh%2Fabe1558e-b512-4ff0-929b-f305221166e0%2F3feb3686c7b488d341ddfec1629a836f&title=dropbox-design**](https://www.evernote.com/shard/s424/client/snv?noteGuid=abe1558e-b512-4ff0-929b-f305221166e0&noteKey=3feb3686c7b488d341ddfec1629a836f&sn=https%3A%2F%2Fwww.evernote.com%2Fshard%2Fs424%2Fsh%2Fabe1558e-b512-4ff0-929b-f305221166e0%2F3feb3686c7b488d341ddfec1629a836f&title=dropbox-design)

[**https://www.1point3acres.com/bbs/thread-311458-1-1.html**](https://www.1point3acres.com/bbs/thread-311458-1-1.html)

**2017.12**

1.     Hr 十一点之前带我tour了一下 人非常好

2.     If byte[] is contained with a file

a.     Naïve way of reading into the sized buffer by a sliding window of the size and then compare byte to byte

b.     然后面试官说 你这样重复读了很多次有没有什么办法可以避免这些重复呀

我说那就用linked list吧

c.     Linked list structure to avoid repetitive reading byte

d.     然后面试官说但是你还是mn的time complexity呀 有没有什么办法改进呀

e.     我说那就用rolling hash吧 用自己的语言解释的，假装我从来没听说过这个

f.      说了一会儿以后 他说不错，但是玩意byte array 非常大，你那个方法算出来的hash value overflow了咋办，

g.     然后我就卡了非常久， 实在没啥思路我就尝试着和面试官说一下我的difficulty在什么地方 我这么想是不是on the right track？

h.     最后好说歹说 才要到hint 他说可以decrease the number by mod呀

i.      我想了一会儿以后又说 我的difficulty在怎么mod了以后如何recover呀

j.      然后他就说了一些mod 是closed under addition and substraction, and multiplication blah blah blah

k.     那我就懂了 相当于最后的solution就是每次calculate完hash value, mod by some large prime number 得到a, 然后算头上那个数的value，那个value再mod 得到b，相当于每次slide window的时候就是a-b+last byte value 这个value来和byte array的hash value mod by that prime number比较 然后如果一样的话就byte by byte得比较，如果不一样的话就move on

l.      最后总算是想出来了 面试官说了一个good job

3.     一个中国妹子？带我吃食堂 原来那也是一场lunch interview呀 要提供feedback的 反正就是聊聊天 还蛮轻松的

4.     get top m most viewd photos – 这个面试官 感觉人非常好 最后和我聊了十几分钟天反正就是感觉人很好的小哥

a.     priority queue+hashmap

b.     然后就被问如果get top m被叫很多次 也不能重新iter一边 那咋办勒

c.     我想了一会儿就说用treeset来做

d.     反正他又说了还可以稍微得improve一点 说了一个再加一个hashmap 存node 这样子可以减少update tree的次数

e.     后面就是聊天啦

5.     id allocator

a.     这道题让我算精确的size in 多少mega bytes

b.     按顺序讲了四个solution、

c.     queue+hashmap(bitset)

d.     bitset

e.     based on second solution, divide the array into different sections

f.      segment tree

g.     最后写了第一个和第三个的implementation

6.     总体experience感觉还是蛮好的

还有个中国妹子后来我们一起出来的，她说她被问到game of life, bit torrent,还有个啥给忘了

**2018.2**  
第一轮：highest minimum value. 给一个2d array, 从第一列出发到最后一列，每次只能往右上，右和右下移动， 返回所有可能路径的最小值中最大的一个。dp解决，分析时间空间复杂度，follow up是如果给的数组特别大怎么办，然后说到disk的读写head, 要尽量减少读写指针的交换。  
  
第二轮：phone number combination. 多一个dictionary， 只输出dictionary里面有的单词。follow up是dictionary 里面存了长度至少为三的单词，输出的结果可以是两个词的组合。例如 8474833 可以是visited也可以是 tip + huff（忘记具体例子..） 之类的，只要这些词在dictionary中。然后再优化，用到trie tree， （用trie tree存词然后检查前缀来剪枝）讨论了一下字典树的原理，还要写insert a word的function。  
  
第三轮：allocate id. 要求多种方法解决，当时讨论了queue + set, bit set, tree

**2018.3**  
1 Sharpness Value  
2 Max island size  
Follow up: what if we can flip some cell, return the current max number of island size after the flip...   
Union find ba... check leetcode, number of islands series  
3. Web Crawler

[https://paper.dropbox.com/doc/In ... 7npLjRBjrAKA5AOBsaM](https://paper.dropbox.com/doc/Interview-Notes-Dropbox-wl7npLjRBjrAKA5AOBsaM)

**2018.3.10**

三轮technical分别是：1. substring search: rolling hash; 2. phone number找words; 3. allocate, realease id

**2018.3.26**

11：00 - 12：00 第一轮技术面， 题目就是 highest sharpness 地里超级多面经的，就不详细讲啦。Follow up 也是一样的。  
12：00 - 13：00 午饭时间，会有一个小哥哥或者小姐姐陪你吃个午饭，过程中随意聊天~  
13：00 - 14：00 Demo, 熟悉一下 Dropbox是怎么用的，会有一两个小问题，答不答应该不计入总成绩。  
14：00 - 15：00 这一轮有点惨，是个白人小哥哥，他的图一画出来我以为是 Game of Life, 但是实际上是没什么资料的# of islands，而且是变形版的。题目是给你一个2d array，找出图中最大 island 的面积。每一个岛只有四个 neighbor，上下左右，不走对角线。我把他当成新题来做的，靠临场发挥，跟面试官确认过眼神之后开始码，过程比较顺利，没想到最后竟然给了 test case 去 run。唔...就出现 bug 了，然而那个时候我脑子已经不怎么转了，小哥哥帮忙改了一行码之后就成功了。这个可能成为挂点吧。。god bless。  
15：00 - 16：00 Allocate Id, 这个也是原题啦。碰到原题大家不要表现太明显，演技！演技要上线，展现你的思考过程！

**2018.11**

上周五 onsite 发一下面经 攒个人品第一轮 国人小哥问的 isdownloadcomplete的问题 和地里的基本一摸一样 （interval tree）  
第二轮 一道给了一个网址 让返回所有这个网址可以访问到的地址 给了一个 getlink的API. 写的时候忘了考虑 重复的link 会有cycle的问题 然后稳bottleneck. 最后问了怎么解决 （多线程只要写 high level）  
第三轮 老题duplicate fil 只给了一个根目录 然后让你用提供的几个API来做。其他思路完全一样 follow up也和地里面经一样

18年底面经详解：

<https://www.1point3acres.com/bbs/thread-467473-1-1.html>

**2019.2**

1 allocid  
2 爬虫  
3 设计log系统, 问的很细节很底层，没有按照系统网上系统设计的套路来什么先给high level啊啥的，直接给我问懵逼

**2018.3.26**

多线程爬虫 max photo 还有NASA题

**2019.3**

1 access file

他给的folder结构如下，意思是A是总文件夹，A底下有BC两个文件夹，C底下有D文件夹，等等，(child,parent)。  
[(A, None)  
(B, A)  
(C, A)  
(D, C)  
...]  
我用BFS写的所以这个结构不会用，他说当然可以重构一个，但是要用他的结构去initialize。  
我重新构造的是parent到children的map:  
{A: [B,C], C:[D]}

2 phone number

3 allocate id

1. # Improve time complexity: least time complexity
2. # allocate O(1), deallocate O(1), space O(2n)
3. # Bit array: least space complexity, slow in allocate. From 1point 3acres bbs
4. # allocate O(n), deallocate O(1), space: n/8 Bytes
5. # Segment tree: efficient when allocation far more than deallocation
6. # allocate O(logn), deallocate O(logn), space: 2n-1 bits, (2n-1)/8 Bytes

实现一个分配id的类，id可以是0~MAX\_ID-1中的一个值，分配出去的id必须唯一，类里面还有个回收id的函数，总而言之就是实现这个类的两个函数allocate和dellocate，数组结构自己设计。我说了两个方案，一个用queue，一个用bool数组。他让我先讨论用queue的情况，写了代码，用queue很直接，然后他让改进时间和空间，方案就是用一个自增的整数，表示还没被分配的id，初始化为0，再有一个队列，每当dellocate时，就将这个回收的id放到一个队列里。每当allocate时，先判断这个队列里有没有id，有的话拿出来返回，没有就拿那个整数，并另其+=1，这样的话构造函数的复杂度就变成O(1)了，同时空间也省了很多。接着让讨论bool数组的情况，这个方法比较糟糕，初始化O(n),allocateO(n),唯一好处就是空间比queue省一点。然后他让改进，想了一会无果，他提示有没有办法可以让allocate更快呢，快32倍，快64倍。然后我就说用bit，每次我可以扫32位（int），然后如果这个int大于0的话，就说明这里面有一个id可用，然后就可以用常数级的方法找出这个id。他说还有没有更快的，比如存储一些临时信息？我就基于前面的方法和这个hint想到了用线段树，这样的话，初始化O(n),allocateO(logn),dellocateO(logn)，然后他让我写了代码，假设线段树已经建好了，要我实现allocate函数，pls note that 在实现的过程中，我假设一个数节点存着区间开始和结束点，可用id数量，左孩子和右孩子指针。写完后他让我优化空间，我说不需要开始和结束点，因为可以通过计算算出来。让我再优化，我说可用id数量可以换成一个bool值。让我再优化，让我省去两个指针，我说或许我可以用堆的数组实现方式来做，但是这不一定是一颗满二叉树，他说那假设maxid是2^n吧，我说那可以了，我只用一个数组就好了，a*的左右孩子分别为a[2i+1],a[2i+2]。然后最终方案就变成了只需要2\*MAX\_ID-1个bit的空间来实现。总结一下就是用queue耗空间省时间，数组耗时间省空间。*

**Questions to Ask**

1. Could you share some exciting amazing tech success in Dropbox? For example, how did dropbox built its own infrastructure (move away from AWS) with a small team?
2. People share that they are always surprised working at Dropbox. Could you share some of the fun or exciting moments?
3. Not a place to work but a place to grow, what are some of the training opportunities out there or how do people grow their technical skills?
4. The company is adding more excitement to it e.g. Dropbox Paper, what other features will be coming? what’s the strategic direction that the company is heading to?
5. Collaborate & help your teammates, as a team, or how do cross functional collaborates happen, any company-wide initiatives?
6. What are some of the important skills that you learnt from working at Dropbox? Or, what are some of the skills to succeed in working environment in Dropbox?
7. What are the tech stacks for the application dropbox offered? For example, the desktop application.
8. How’s dropbox’s team structure? I mean what are some of the teams and what are they responsible for? group by products etc?
9. What are the development lifecycle / methodology for the produce team e.g. how quick do they prototype or iterate quick to incorporate market changes/ customers needs.
10. How does your team train up new recruits?