



Save Copy to Evernote

Last updated: Jan 18, 2019



**>> Average Package:**

Package: 150k + 250k/4 + 50k + 10%

First year total salary: 277K

Green Card: < half year

>> Positions:

1. Infra : 3 system design, hard onsite interview.
2. App: 1 system design.
3. tools developer experience. easy coding problem. Some system design.

>> System Design:

1. Store email using what database. noSql vs SQL?
2. topk shared URL in 24 hours
3. 设计linkedin的news feed, 基本要求是对所有的news, 每次用户登陆都显示根据某个排序条件（可能是很复杂的多个条件综合算出一个分数后的, 这个计算过程不一定可以被SQL表达）后的top K的新闻, 然后会问怎么做过去1分钟 / 5分钟 / 一天的top K, 跨机房啊, 一致性啊这些问题。
4. design calendar. 用户调动自己日历, 调取别人日历, 增加会议。Outlook
5. log process
6. 设计graph题, 给定一个 int[] getFridend(int user), O(1) complexity, 求两个users 是不是一级联系, 二级联系和三级联系。先在local解, 后来问图很大, 怎么scale到多个machine上
7. 实现线程池/ExecutorService, 跑需要延时的任务
8. publish and query. 存event id和内容, 内容大小不定。然后query API是 getEventSince(id, int bufferSize), 求怎么存储数据可以更利于high concurrency。这题答得不太好, 要了好多hint。最后面试官还说到想问lock, 但是没时间了。
9. tiny url. 最终挂了, 反馈是系统设计短网址不会架设多个服务器来处理请求,
10. 设计一个文本数据库, 对于query 高效的返回某个key在数据库中出现的的所有位置, 出现次数是否大于k
11. 设计一个传感器网络, 有大量的传感器实时发送数据, access point 向这些传感器发送query,如何快速得到结果
12. <https://www.1point3acres.com/bbs/thread-294114-1-1.html>
13. <https://www.1point3acres.com/bbs/forum.php?mod=viewthread&tid=175538&extra=&highlight=linkedin%2B%CF%B5%CD%B3&page=1>
14. <https://www.1point3acres.com/bbs/forum.php?mod=viewthread&tid=446923&extra=page%3D3%26filter%3Dsortid%26sortid%3D311%26searchoption%5B3088%5D%5Bvalue%5D%3D1%26searchoption%5B3088%5D%5Btype%5D%3Dradio%26searchoption%5B3046%5D%5Bvalue%5D%3D6%26searchoption%5B3046%5D%5Btype%5D%3Dradio%26searchoption%5B3109%5D%5Bvalue%5D%3D2%26searchoption%5B3109%5D%5Btype%5D%3Dradio%26sortid%3D311%26orderby%3Ddateline>
15. 闲扯单机的 再cross node.
16. append only datastore like HDFS
17. 设计, 三哥, 设计metrics collection and monitor system, 收集host的1second, 1minute, 1hour metrics。如何scale, 如果是10000 hosts怎么收集。提供各种方案, kafka, HDFS, pub-sub model。各种分析优缺点。然后an over workflow。这轮还算比较顺利。



ole

reaming system, 每个message 可以有 4Bytes ~ 16 KB, 需要考虑的有 retention, work load,

memory overflow, race condition。只需要考虑一个machine。不说 high level idea, 说具体的components, data structure

21. 求平均值 + follow up
22. LinkedIn Infra 3 system design: 就是KV Store (\approx 自制db), 找朋友, build invert index(/solr), 设计类kafka系统。。
23. <http://massivetechninterview.blogspot.com/2015/06/algorithm-how-to-count-number-of.html>
24. 倒排索引很多怎么存可以快速查找. Hot to make it distributed
25. design instagram
26. 设计一个系统可以统计出过去24小时里面top k exception。具体要聊到数据怎么从数据中心发到这个server端, 然后怎么存数据, 怎么生成top k exception
27. . 设计一个系统, client会发1B-16M的message到server, server要把message存在memory里并返回一个id给client, client可以用id fetch message。很细致的讨论了存哪些metadata, 每个metadata到底占多少空间, 为什么每个message都存metadata会非常不efficient, 怎么样优化存metadata的方式等等。
28. 假设已有如下两个function Output Compute(Input in) Output Merge(Output o1, Output o2) 要求在最多可以有K个thread的情况下实现 Output MergeAll(List<Input> inputs)。实际上是个multi-thread programming的问题, 需要写出比较detail的代码。
29. 然后看时间还挺多, 就出了道system design的题, 就是如何autocomplete, 先用trie树, 然后再聊到用cache来记录大部分结果。这轮问题不大。
30. <https://www.1point3acres.com/bbs/forum.php?mod=viewthread&tid=427788&extra=page%3D5%26filter%3Dsortid%26sortid%3D311%26searchoption%5B3088%5D%5Bvalue%5D%3D1%26searchoption%5B3088%5D%5Btype%5D%3Dradio%26searchoption%5B3046%5D%5Bvalue%5D%3D6%26searchoption%5B3046%5D%5Btype%5D%3Dradio%26searchoption%5B3109%5D%5Bvalue%5D%3D2%26searchoption%5B3109%5D%5Btype%5D%3Dradio%26sortid%3D311%26orderby%3Ddateline>
31. , 输入是文件, 要求query文件里面的内容, 这轮死了。
system design, 要求就是特别简单, client 存文件, 然后user query 内容, 文件是1MB。这个我也不知道怎么设计比较好。

当时有几个问题就是如果client 同时往一个database里面写的话需要lock, 怎么解决, 我就想说多分几个database 存, 但是呢就需要merge, 多久merge 一次。

面试官说的很清楚就是存文件, 大小确定, query里面的content, 返回文件。这些文件得存啊, 怎么存呢, 怎么存才能更快地query, 我一个数据库存文件。

32. Coding design: m blocks, each block has m bytes, support files read/write, design metadata and data, need to allocate memory, free, update and get
33. , 问一个design LinkedIn homepage, 需要有发一条状态share出去, 和显示feeds的功能, 主要问了很多database的问题, 感觉是个要求特别严格的manager, 有点不确定答的是否满足他的要求
34. ip blacklist service
35. 一个精心包装的 KV 缓存设计, 要能保存 500M 个网页, 查询时间不得多余 1s, QPS 等需要自己问。
36. design ip blacklist. malicious IP blacklist.



9. mute semaphore vs binary semaphore
10. what is memory map file
11. what is db transaction? How to ensure durability? how to restore if crash before commit.
12. TCP vs UDP. Which layer.
13. mmap是什么，读和写shared memory object操作系统内部会发生什么? -ba
14. Database Transaction
15. **Write-back vs Write-Through.**
16. hadoop vs spark
17. java里的final, finally, finalize
18. How does processes communicate with each other? IPC: different ways, RPC, shared memory, message passing, pipes, socket
19. 多个application 同时运行，memory 不够用了，OS 会怎么办。OS 如何 free, allocate memory的，namely, memory management. 比如page share, compress, 或者swap
20. 给一个文件，问你用程序读这个文件的时候，OS都发生了什么操作。怎么读大于内存的文件，怎样优化等等。。这个我答得不是很好，稍微讲了下思路，但是说我平时不怎么操作文件，就过了
21. paging, page demand, page fault, **Thrashing**,
22. Java exception。问的非常细，详细描述什么是exception，有哪些分类，为什么要用exception，而不是直接打印出错误 etc.

>> BQ:

1. talk about your weakness, how you overcome it.
2. behavior question: 你在项目中遇到过最难的问题是什么
3. 跟经理意见不一致怎么办
4. how to trouble shooting, how to improve your current project
5. How to improve linkedin website
6. how to code review
7. 簡歷解說，叫我把他當一個新人，把現在的部門在做的事情介紹給他聽
8. why cassandra
9. how much linked open source you know about?
10. Project you proud of
11. largest challenge.
12. why come linked?
13. 聊工作，怎么提高developer 的 productivity
14. your strongest and your weakness
15. 选一个你觉得最自豪的项目，然后标准是把面试官当作new hire on board, 给他讲这个project，关注大框架，如何design，同时也会问detail，基本45分钟吧，因为前面各自自我介绍5分钟，后面留10分钟问问题。我的面试官没有问很多，但之前看到别人也被问类似如果让你重新设计，你会improve哪里之类的问题。我的面试官是说这轮就是看你有没有technical communication的能力，就是跟组里人交流你做的东西，所以既有technical detail，又关注communication
- 16.

>> LeetCode



0, 46, 50, 53, 56, 57 (支持addInterval()和getWeightedSum) 61, 65 (don't think

Save Copy to Evernote

2, 153, 156, 160, s170 (优化find, getSum o(1)), 1

200, 200, 200 (max word), 212, 215, 235, 236, 240, 243, 244, 245, 254, 265, 270, 272, 282, 297

314, 339, 352, 364, 366, 373 (smallest product, without sort)

432, 464, 474,

516, 535,

605, 611, 612 (KNN), 671 (find 3rd minimum or kth minimum value), 698,

716, 730 (一个字符串可以去除任何数量的字符 输出所有的子串), 722, 726

>> Other Coding:

flip tree upside and down.

BST top K closest neighbour

实现hashmap, 全部都实现,

given a filter and a iterator to implement a validIterator 也出了解法

第二题是自己实现一个stack, 可以O(1)实现push, pull, getMiddle, 讨论了几分钟写完又跑了一下test.

coding, rotation string

two sum to k sum...

topology sort

N-Queue I, II

Fizz Buzz

word search,

word ladder,

paint house,

list intersection and union,

serialize and deserialize binary tree. follow up: if tree is deep.

implement max heap.

tree level by level

罗马数字转普通数字,

LRU,

find triplet,

是否一个图的所有店可以标记为2个不同的颜色。

Nested List Weight Sum 1&2

interval problem

已知[0, 1]的随机uniform分布函数01Uniform, 求用它来写[0, 6]的随机uniform分布函数。

打印所有的factor可以是负数, 心态崩了估计就挂这了。。

写了一道算法题, 给定一个社交网络, 打印出两个人之间的最短路径, 由于没有权重不需要考虑Dijkstra这种, BFS可以写, 然后就开始问ML的问题, 让我自己随便说一个熟悉的模型然后他问问题, 没什么特别的地方~

bi-bfs会更好点, bfs要 $\text{avg_edges}^{\text{distance}}$ 的时间, bi-bfs只要 $2 * \text{avg_edges}^{(\text{distance}/2)}$

, 一个白人小哥, 人也挺好, 首先互相介绍, 然后coding, 平面上找点最多的线那题, 不是很难, 主要是考虑一些边角情况, 斜率是无穷的时候怎么把等, 写的过程中不断考虑和加新的细节。因为小哥来的晚了些, 所以最后就问了这一题, 我写的还算快, 最后有时间问了一些简历上的细节, 感觉小哥也是非常满意的, 应该是一个不低的分。

挂人题: <https://www.1point3acres.com/bbs/thread-395874-1-1.html>

Coding (Tree traverse with Stack, with Queue, 写一个Queue, 有poll, offer, add, remove要求多线程, 不能用其他Generic 数据结构, 要求有Lock, 最后她说要用Doubly Linked List写)

3. Coding (Paint House III, 不能用最优解, 要每行解释, 最基本的DP, 然后再优化, 扫雷.529, minesweeper. 教训就是每个题不要只去注重一个最优解, 然后先把整个题的基本解给出来, 然后如果面试官要求需要优化再优化。我当时有的也没弄清楚, 有的就是死背了一个答案。看到这个题的时候还暗喜了下, 结果人家不让你这么



Nodes。找到他们的最低祖先。

5.2 find celebrity变种。不一样的地方就是没有给knows这个API。给的是一个boolean二维数组A。A[j]如果是true, 那么说明i follow j。根据数组A, 返回一个int所有人都follow他, 但是他不follow任何人。

6. Coding -- coin change II 变种。给一个Integer Set和一个int target, 找到所有相加和为target的subset的数量。

经典的岛屿 (0, 1二维矩阵) 个数, 改成不返回岛屿个数, 返回每个岛屿的一个点。在白板上写的很认真, 走了几个例子, 然后follow up了一下, 如input矩阵太大了 只能一块一块load怎么办

101. Symmetric Tree

<http://www.1point3acres.com/bbs/thread-279704-1-1.html-baidu>

1point3acres

有一个优化的地方就是, 当无法在cache里面找到key时, 需要从data base 里面去取, 这个时候如果新取到的key的rank 要小于cache里面的最小rank 时, 不需要更新cache

1. 实现一个二叉搜索树, 问我都能实现啥功能, 我说insert, find, findMin/findMax, 然后让我实现了insert.
2. 按行打印出来这个二叉树 (利口妖灵儿)。
3. 这俩题没啥说的, 然后还剩不到半小时了, 问我给一堆点的集合, 找到其中距离指定Point最近的k个点, 这题比较常见
做法是用pq, 时间复杂度 $O(n \log k)$, 然后他问完复杂度和用的方法, 直接问我有没有更好的, 我说可以用快排的思路来做,
平均期望 $O(n)$, worst case $O(n^2)$, 他让我用这种方法写, 所以小伙伴们注意这题要会这两种方法, 虽然我个人感觉电面,

code部分就是让实现一个 hashtable, 给正常的Map interface, 让实现 add, remove, get等。做完以后要求写成thread safe的, 不能synchronize method或者synchronize this

implement a special stack, peek(), push(), pop(),

pop current max element in the stack, may not the top one popMax(),

retrieve current max element in the stack, same, may not the top one peekMax()

as fast as possible. popmax 和 pop 我刚开始想的都是 $O(N)$, 后来把其中一个优化成 $O(1)$

push(), peek(), pop(),

popMax(), peekMax()

。都是原题。max stack要写logn的优化版本。我用的double linkedlist + priority queue

```
public class RetainBestCache<K, T extends Rankable> {
    Waral 博客有更多文章,. From 1point
    3acres bbs
```

```
    private Map<K, T> cache;. visit 1point3acres for more.
    private Map<Long, Set<K>> rankingOfObject;
    private DataSource<K, T> dataSource;. 留学论坛-一亩-三分地
    private int maxSizeOfCache; 来源一亩.三分地论坛.
    . from: 1point3acres
    /* Constructor with a data source (assumed to be slow) and a cache size */
    /* Gets some data. If possible, retrieves it from cache to be fast. If the data is not
    cached,
```




on the data source. If the cache is full, attempt to cache the

with lowest rank among the ones that it has a
ie, the cache may choose any T with lowest rank to evict.

Save Copy to Evernote

*/

具体参考这个帖子第二道: <http://www.1point3acres.com/bbs/thread-279704-1-1.html>

- : BST找Top K Closet node, 用inorder和heap做的, 面试官貌似又不太了解Java (?), 他问我queue的offer是什么, 然后又出了道利口160
- 215. Kth Largest Element in an Array变种, find k closest point to (0,0) in 2d plane.
- 然后写代码。给你List<String> list1, List<Character> list2. 让你用list2里面的字母, 组成字符串, 但字符串必须在list1里面. 如何并行计算?
- find and replace in string. 有点像 利口霸散散, 但是不给starting inde, 只给s, source_str 和 target_str. 要注意如何handle corner case, 比如原词是"acc", 替换词是"accacc", 讨论应该输出什么, 如何处理。
- <https://www.geeksforgeeks.org/flip-binary-tree/>,
- <https://www.1point3acres.com/bbs/forum.php?mod=viewthread&tid=433813&extra=page%3D5%26filter%3Dsortid%26sortid%3D311%26searchoption%5B3088%5D%5Bvalue%5D%3D1%26searchoption%5B3088%5D%5Btype%5D%3Dradio%26searchoption%5B3046%5D%5Bvalue%5D%3D6%26searchoption%5B3046%5D%5Btype%5D%3Dradio%26searchoption%5B3109%5D%5Bvalue%5D%3D2%26searchoption%5B3109%5D%5Btype%5D%3Dradio%26sortid%3D311%26orderby%3Ddateline>

IncreaseFrequency(int) //让某个int的freq + 1, 如果是一个没出现过的数字, freq设为1

DecreaseFrequency(int) //让某个int的freq - 1, 最低为0, freq=0的数字不会再继续降, 尝试decrease没出现过的数字则报错

int GetFreq(int) // 返回一个freq=input的数字, 有多个数字有那个freq的话返回任意一个, 没有那个freq的数字存在则报错

这真是我自己作死, 三哥先问了个easy, 给两个sorted list, 然后merge成一个, 要求是merge后的list里不能有重复

之后的follow up是merge k个sorted list, 当时太累, 忘了表演推理过程了, 直接说用queue做, 三哥装傻让我给他走个case, 我走完之后三哥直接换题了.....

还是k个sorted list, 可能有重复, 现在让你按序找出所有list中都必须存在的元素, 比如

[1, 3, 5, 7]

[1, 1, 3, 5, 7] => return [1, 7]

[1, 4, 7, 9]

一堆人站在路上, 找一个点大家集合。求出这个点, 大家走动距离最小。其实就是 array 找 median。

写了 quick select, 讨论了一下大数据 bucketing, map reduce 的思路。

S

Not under linked tag

course schedule ii	https://leetcode.com/problems/course-schedule-ii/description/
rotate list	https://leetcode.com/problems/rotate-list/


<https://leetcode.c>
[Save Copy to Evernote](#)

Reverse String	https://leetcode.com/problems/reverse-words-in-a-string/description/
Longest Palindromic Substring	https://leetcode.com/problems/longest-palindromic-substring/
Minimum Window Substring	https://leetcode.com/problems/minimum-window-substring/description/
Edit Distance	https://leetcode.com/problems/edit-distance/description/
Lowest Common Ancestor of a Binary Tree	https://leetcode.com/problems/lowest-common-ancestor-of-a-binary-tree/description/
Regular Expression Matching	https://leetcode.com/problems/regular-expression-matching/description/
Validate Binary Search Tree	https://leetcode.com/problems/validate-binary-search-tree/description/
Find Minimum in Rotated Sorted Array	https://leetcode.com/problems/find-minimum-in-rotated-sorted-array/description/



```

    private List<Length> intervals;

    /** * Adds an interval [from, to] into internal structure. */
    void addInterval(int from, int to);

    /** * Returns a total length covered by intervals. * If several intervals
    intersect, intersection should be counted only once. * Example: * * addInterval(3,
    6) * addInterval(8, 9) * addInterval(1, 5) * * getTotalCoveredLength() -> 6 * i.e.
    [1,5] and [3,6] intersect and give a total covered interval [1,6] * [1,6] and [8,9]
    don't intersect so total covered length is a sum for both intervals, that is 6. * *
    ----- * ____ * ----- * * 0 1 2 3 4 5 6 7 8 9 10 * */
    int getTotalCoveredLength();
}

public class MyIntervals implements Intervals {
    List<Length> l = new LinkedList<Length>();
    @Override
    public void addInterval(int from, int to) {
        l.add(new Length(x,y));
    }

    @Override
    public int getTotalCoveredLength {
        Collections.sort(l);
        int retLength = 0;
        Length lastone = new Length(0,0);
        for(Length len : l) {
            if(len.x > lastLen.y) { //locate apart
                totalLen += len.y - len.x;
                lastLen = len;
            } else if(len.y > lastLen.y) { //overlapping
                totalLen += len.y - lastLen.y;
                lastLen = len;
            }
            //注意这里不需要考虑如果后一个在前一个里面会怎么样, 因为lastLen会维持一样, 写一次仍然
            跟前一个做比较
        }
        return totalLen;
    }
}

public class Length implements Comparable<Length> {
    public int x, y;
    public Length(int x, int y) {
        this.x = x;
        this.y = y;
    }
    @Override

```

