

CH 8月2日 腾讯 第一轮电面

时长: 45min

hash 和 tree 的区别

进程间通信方式、共享内存具体怎么实现

epoll (细节)

网络通信方式 (windows、linux)

文件系统 (知道哪些, fat 和 ntfs 的区别, 磁盘 io 优化)

hdfs 的存储机制, 以及如何保持一致性

teamwork (问了很多, 分工协作, 遇到问题怎么解决, 是否了解 teammate 的工作, 从他们身上学到什么)

总结: 没有算法题, 大部分是操作系统底层的东西, 而且很多相关的实现细节。

腾讯

一面:

操作系统: 虚拟内存、多进程多线程区别、进程间通信、IO (同步阻塞、异步阻塞、同步非阻塞、异步非阻塞), epoll

网络: TCP 三次握手、四次挥手

C++: 虚函数表、指针和引用的区别;

做题: 求数组的最大连续子数组; 判断一棵二叉树是否是完全二叉树;

二面:

项目: 百度大搜索建索引的具体流程, 存储迁移系统问的很细;

用过哪些高级 C++ 语言特性? 答函数指针, 继续问函数指针和虚函数的区别;

做题: 写一个函数, 判断操作系统中栈的递增方向和内存逻辑地址递增方向是否一致; (非常蛋疼, 需要对内存理解非常透彻, 知道程序编译以后数据、指令在栈和堆上到底怎么放的)

腾讯

一面

说项目。。。

1. 有两个数组, 求交集, 不能用哈希。

$A[m], B[n]$,

排序后两个指针顺次遍历, $n \log n + m \log m + \max(m, n)$

follow up: 如果 N 远大于 M , 如何优化 (这种情况下, 上述算法还不如暴力 $M * N$)

只将 $A[M]$ 排序, 然后逐个二分查找, 复杂度 $m \log m + n \log m < n * m < n \log n + m \log m + \max(m, n)$

2. 内存对齐, pack 用法

3. const 修饰的量为什么不能修改? 实现原理?

4. C++内存布局
5. TCP 三次握手四次挥手，SYN 攻击如何防范（百度考过）
6. 类似 LC 上那个 IP 划分的一道代码题。
7. 设计题（围绕这个题问了好多，值得 mark 一记）

如何实现根据 IP，返回详细地址的功能。

开头出了个 BUG，扯到内外网 IP，网段划分等，这个完全不是很清楚，被教育了。

就是 A, B, C, D 段网段的划分之类。

然后问了内外网 IP 的映射问题，比如在外面看，中科院计算所可能都是一个 IP 地址，但是内网 IP 配成不同的。那么，内网一台机器发包到外网，经过网关时包的源地址变成了单位的对外 IP，从服务器返回到网关后，网关如何知道是内网哪一台机器需要的包？掩码什么的很不熟，答得磕磕绊绊。

然后在网关啊，内外网 IP 映射，路由器，路由表，TCP 包，UDP 包等网络知识上扯了好久后。。终于回到正题。

给你的海量数据字典是长这样的。比如：

10.10.1.0 - 10.10.10.255 北京

10.10.7.21 - 10.10.8.29 海淀

.....

.....

1. 0. 0. 0 - 1.0.XX. XXX 加拿大

写完后，又说，其实最原始的数据，是人工标注的一个个具体 IP 是哪儿。

然后先预处理抽取成上面这样，但数据粒度是不一样的，比如可能有的国家，假设拿加拿大做例子，因为原始数据的原因，只精确到国家这一级别为止了。而有的，比如可以精确到北京海淀，或者更进一步北京海淀中关村。

现在来一个 IP，如何高效尽可能精确定位所在位子。

讨论了因为粒度不均匀，而且是个范围，加上精确的级别不一致，并且每个级别由多少位划分，都无法确定，所以不好哈希，考虑类似前缀字典树进行这个查找，因为 IP 前缀一致，能搜到什么精度就是什么精度。大致这样，用树不错，不过这个具体应该是个 B 树，讨论了一下 B 树之类的，时间不够了没说完。。。回头再好好想想这个题~~(>_<)~~

二面

1. 给一个字符串，比如 12345<%abcEFGyhu%>abcwesd

把<%... %>中间的，凡是在字典中出现过的子串按规则替换，并直接原地修改成替换后的。比如字典是{["abc", "X"], ["yhu", "Y"]}

上面的串被原地变为 12345XEFGYabcwesd

觉得出第一题是为了有时间在我写的时候看简历。。。然后写差不多就简单问了下，水题没啥说的，过了后就聊了会项目。

2. 写一组字符串库函数，strlen, strcpy, strncpy, strcmp

3. 虚函数，纯虚函数，虚基类

4. 虚析构函数什么时候用，父类子类构造函数和析构函数的调用顺序

5. 内存对齐，什么时候用

6. 设计题

设计 QQ 书签。

明确了下题意，意思就是 QQ 浏览器有那种用户登录功能，然后可以在一个页面管理收藏夹。

先讨论数据库设计，ID 就是 QQ 号没得说。因为是海量数据，按 ID 哈希到 N 个数据库服务器存储，然后为了查询快，记得加缓存，缓存策略就是活跃度。用户可以自己创建各个文件夹分类，做一个索引。又讨论了下 NoSql，了解哪些关系型非关系型数据库。

然后讨论 WEB 部分，一开始 DNS 按照 URL 看是移动网还是联通网啊之类的，分发到最快的服务器，思路类似键入 URL 后发生了什么一顺溜画下来。每个层次画清楚了就好，然后分析性能瓶颈。讨论了一下连接池等。

细节记不清，面试官引导比较好，所以交流还挺愉快的。相当于把 WEB 的一些东西串了一遍。

上面都是在武汉 office 面的，Kevin 似乎比较满意，二面完了闲扯了一会儿说名额少，让我下午去酒店直接找老大说是他们组推荐来的，表达强烈想来武汉坚决不去深圳意愿。有可能悲剧因为名额，要是非常想来 TX 还是考虑一下深圳。

三面

这老大就没名额不想面，或者因为我暑假前投过安卓啥的印象不好，觉得方向不明确，杂而不精。

扯了半天你为啥非要回武汉呢，万一武汉的部门撤了还是要去深圳你来么，你这么纠结地点就为回家吃个饭么 - -|||

想回家图稳定为什么不去考公务员呢囧。。。这种气氛下开始面试可想而知。。

好吧，那说说项目吧。。

好吧，那问你几个问题吧。。

Linux shell 熟么，查找一行开头是 abc 的怎么搞，结果纠结与“abc*”匹配不是从头开始，比如 deabcf 也符合，必须 abc 打头咋整。。忘了，被鄙视。

考 sed awk 具体忘了。被鄙视。

python JAVA C++多线程比较。

Linux 进程通信哪几种，举例子应用。

阻塞和非阻塞时，receive 函数的区别。估计扯错了。每次都被 socket 虐 T_T，socket 不会还敢投后台简直找死。

解释一下 TCP 滑动窗口。

TCP 拥塞控制。

说说数据库的左连接是什么意思。一紧张没想起来，没答好。

画一个节点不少于 15 个的平衡二叉树。。(感觉很诡异，又被调戏了的感觉 T_T)

怀疑的乖乖画好。问，多少个节点？刚好 15 个，四层满二叉树。嗯，给他们标上序号。

囧。。乖乖一个个标了。。问，树有几种遍历方式？囧。。一般常规就三种。

嗯，写一下三种遍历吧。。懒得问递归非递归了直接写。。

被打断。。等等没让你写代码，按你自己的编号，写出三种序列(= =! 这么鄙视我么。。)

怀疑的一个个写了编写编解释。。因为节点多，写完前序中序估计他也受不了了
说行吧就这样。

知道红黑树么？知道，和他扯了下 STL map。

红黑树和平衡二叉树什么区别？着色规则之前记了一下但不是很清晰，以及旋转
复杂度。好在没让写代码。。

你写 JAVA 比 C++ 多么？ = =! 显然不是...

JS 掌握到什么程度？ 几乎不会

安卓会么？ 几乎不会

你本科不是计算机？那为什么读研要读计算机呢？。。 = =! 喜欢写代码不行么
囧

有什么爱好？。。我说写代码你信么囧

你还有什么要问的？其实刚毕业头几年在哪儿无所谓。。 嗯，不过 TX 的话我只
考虑武汉。

好吧，滚粗。。

百度互联网知识挖掘工程师

一面

1、strstr，【kmp 可破】

2、求极大值，先升后降，【夹逼可破】

3、两个 50 亿的文件，每个 url 64 字节，内存是 4G 的，（要求用 Python，不会）
【hash，分桶桶间无关系可破】

4、相交链表，【考虑环，可破】

5 无穷尽流中，找 1000 个等概率的 word 【蓄水池问题可破】

二面

1、字符串中，有数字有字母，找出最大的数字，int，不考虑溢出

2、Android 手机若干问题

3、项目细节

三面

1、 必胜策略，圆桌上按照坐标系四个象限，放四枚硬币

A，背对桌子，可以要求裁判任意翻转 0-4 枚硬币，告诉裁判，相对自己的
位置的哪枚硬币，比如说可以说相对自己位置的第几象限

B，看着桌子，可以任意将桌子转 90 度的倍数

硬币全正或者全翻，A 赢，有必胜策略么？那么必胜策略在最坏情况下是几
步呢？

2、100 亿个 query，统计 top100

3、然后问了一些 mapreduce 的原理巴拉巴拉的，大数据处理架构，有没有用过
哪些

时长: 1h

百度 2 面

商务搜索部, C++研发工程师

一面:

1. 问了问微信那边的情况, 发现面试官认识我组长和总监(狂汗。。。), 然后就开始聊微的项目, 大概二十分钟。
 2. 先要求介绍常见排序, 随后问了一个海量数据处理的题, 1 亿个 double 找出其中最大的 1w 个, 问内存消耗和时间复杂度。用小顶堆解决以后, 面试官要求降低时间消耗。我说了个堆的优化, 随后他说可以考虑并行计算。于是说多节点分布式计算解决, 面试官要求具体讲讲用 hadoop 如何实现。
 3. “这个题目你应该见过”, 面试官如是说到, 然后就说找两个链表的交叉节点。明显无法炫演技, 老实作答见过这个题, 并且给出了有环和无环时的解决方案。
 4. 对移动搜索的认识, 主要讲讲移动搜索和传统搜索有什么不同点或者着力点。
 5. 比较偏产品的题, 腾讯入股大众点评, 点评明显看中的是微信入口, 如果你作为点评决策人应该把什么形式的入口作为最重要的?
- 随后就是聊天, 闲聊了会儿等二面面试官过来。。。

二面:

1. 先介绍下自己, 介绍下项目, 然后开始讲项目内容, 和一面差不多。。。
2. 是否熟悉 cache? 介绍下 cache 的原理, 细说 cache 替换策略。我提到了 LRU, 面试官随即让我实现一个基于 LRU 的 cache, leet code 原题啊, 演技派节奏走起, 解决了。
3. 系统设计题, 关于主从节点的数据同步, 设计一个方案要求效率高。
4. map 和 unordered_map 的区别, 讲下红黑树的特性, 哈希表的映射策略等等。
5. 还是系统题, 1 个名人收到千万粉丝的关注, 如果短时间内有千万级别的访问数, 如何保证有效响应请求?

回答:

首先, 负载均衡, 面试官说你可以认为即使负载均衡还是会超过单机负荷, 别改变硬件。

其次, 去重策略, 因为名人发过的微博数毕竟有限, 可以在后端节点之上使用一个逻辑上层用于做请求去重, 这样就大大减少了真正对机器数据的访问量。

面试官: 名人两年前发的文也被存在这些节点中, 但是实际上要拉去两年前微博的用户非常少, 微博是具有时间热度的。

再次, 和前面的问题结合, 引入 cache, 缓存最近最热的微博, 这些微博被拉取的频率最高。

面试总结,

1. 世界太小了，一面面试官认识北京微信团队不少人，二面面试官从腾讯去的百度。。。
2. 和我以前面百度的经历不同，这次面试算法问的并不多。
3. 二面结束以后问了面试官后续安排，说是会有三面，只好静候佳音。。。

Qiango

时长：3h

百度

一面：

Coding1：翻转链表，给定链表头，返回倒转后的链

Coding2：非递归求树的 dfs

开放性题目：谈谈对 hash 的理解。我说了几个 hash 算法，如 time33，MD5，sha1 之类的，然后说了 hash_table 的实现，然后提了 bloom filter。之后版主提到了一致性 hash 也是一个亮点。

大数据处理：两个文件各有 50 亿个 URL，找出交集。参考 July 博客。

二面：

coding：给定一个数组，求逆序数。

开放性题目：给定两个数据集合，第一个数据集合描述了一系列对象，比如（刘德华，男），（海贼王，动漫）。第二个数据集合是用户的 Query。问怎么利用第二个数据集合来丰富第一个数据集合，比如（刘德华，男）可能扩展成（刘德华，男，歌手）。

三面：

翻硬币：A，B 两个人玩游戏，C 是裁判。AB 中间有一个桌子，A 背对着桌子，B 正对着桌子。桌子上放了四枚硬币，摆成了一个正方形，硬币初始正反的情况是随机的。现在由 A 任意说几个位置，C 会把相应位置的硬币翻转，然后 B 会把桌子旋转 90 度的整倍数。C 是绝对公正的裁判，当 A 的操作使得硬币都朝向一面了，A 赢。问 A 的必胜策略是什么。

百度

一面：

1、首先是问 C、C++

```
struct a{
```

```
int b;
```

```
char c;
```

```
} 占多大内存空间
```

2、堆和栈的区别

3、Singleton 的实现

4、常见的锁有哪些？

- 5、对于死锁的认识，以及解决方法。
- 6、hashMap 底层数据结构是什么
- 7、ConcurrentHashMap，多线程的情况下为何较优？
- 8、Linux 操作系统调度的基本单位是什么？线程还是进程？
- 9、一个程序，只希望同时只有一个进程在执行该程序应该怎么实现？
- 10、EPoll 和 Select
- 11、socket 的 time_wait 状态，以及 TCP 协议 3 次握手的详细分析，包括包头
- 12、两个排序数组中位数（写代码，二分破之）
- 13、树中两个节点的公共父节点（三种方式破之）
- 14、问项目

二面：

一直扯项目、吹牛逼

百度

- 1、tcp 和 udp 区别
- 2、tcp 怎么保证到达顺序有序
- 3、说说拥塞控制
- 4、说说虚地址
- 5、讲一下你对 hadoop 生态系统的理解
- 6、求树中，两个节点之间的最长距离，距离定义为经过节点数。
- 7、翻转链表

百度研发

1. 设计算法，ab 两个大文件中各有一亿大小 url，每个 url 占用 64 字节，内存只有 2G，怎么找到重复 url？
2. 设计一个类似于 qq 音乐中的猜你喜欢产品，可以从算法或者工程的角度回答。
3. 有一个很大的日志文件记录用户登陆情况，有三列记录，分别是用户 id，时间，登陆或登出。设计一个算法，计算某个时间在线用户的数量。

阿里

部门：阿里云

面试官提前一天打电话给我，约了第二天面试，让我准备好纸和笔，当时觉得是要写题的节奏，果断开始突击剑指 offer。

面试经过：

1、自我介绍，简单介绍就行了，感觉自我介绍就是走个流程，很快就被面试官打断，开始正式环节了。

2、开始问论文，首先要用简短的语言讲清楚文章的主要工作，然后接受细节上的拷问。遇到没有思考过的问题也不要紧，现场发挥，大胆跟面试官交流想法。而且有的问题，面试官也并没有期望你能答上来，他只是试探一下。

3、开始问项目，问项目的时候面试官对我的毕业设计内容比较感兴趣，就聊了起来。所以建议大家可以把毕业设计的工作写在简历上，虽然可能还没有做，但是也都是稍有深度的问题，而且之前开题肯定也调研了好多内容，比较有谈资。

4、写了一道算法题，我在电话这头写，写完念给面试官听，他再写。

题目：给定一个排序数组，对排序数组向右移位，根据移位后的结果，求得移动了多少位？数组元素无重复。

输入：移动后的排序数组

输出：移动的位数

这道题实际上就是求数组中的最小值。是剑指 offer 上的“旋转数组”一题。而且假定无重复元素，更简单。

于是迅速给出了一个二分查找最小值的解决方案。但是因为自己写完后没有主动跑测试用例，留下了 bug，在面试官的提醒下才找出来的。所以当听到“你能不能自己跑个例子”这句话的时候，多半是程序有 bug，骚年们一定要认真找啊。

时长：35min

1、项目。介绍项目，还是 cloudshare，而且还是问了我毕设相关的问题，感觉提前准备好问题讲给他听是真的很爽。还问了，出于什么目的做的这个项目，你在其中做了哪些工作。

2、算法。没有写直接讲思路。

题目：2sum。给定整数数组，求三个数 a、b、c，使得 $a+b=c$ 。要求不用额外空间， $O(n^2)$ 。

3、数学题。字符 a a b c d 的全排列数目。答案 60。

4、语言。

a) Java 的字节码是怎么执行的。

b) Java 里线程通信的方式用过哪些。

5、系统。

- a) 学过哪些分布式课程，对哪些分布式系统比较熟悉。
- b) HDFS 是怎么保证 namenode 的可用性的。

6、其他。

想做什么工作。底层、产品层 or 控制台。。。。

阿里

第一部分：在线做题，一小时。

1. 有 `char a[n], char b[m]` 两个数组， $n > m > 1000$ ，`b` 数组中的元素 `a` 中都有，现在需要生成数组 `c`，将 `a` 中有 `b` 中没有的元素都放到 `c` 里面，要求高效，用 `C/C++` 语言实现。

2. 在二叉排序树上面找出第 3 大的节点。注意：不能把二叉树全量存储到另外的存储空间，比如存储到数组中，然后取出数组的第三个元素。

3. 求出 100000 以内符合下面条件的数字：

它是可以分解的整数，且所有数位上的数字和等于其全部质数因子的数字总和。

例如： $9975=3*5*5*7*19$ ， $9+9+7+5=30$ ， $3+5+5+7+1+9=30$

4. 给定一个无序的整数数组，求出数组中的最长递增子序列的长度；

5. 编写一个扑克牌的随机发牌程序，要求将 52 张牌随机均等不重复的分给 4 个选手。要求用 `C/C++` 编写。用 `1,2,...,13` 代表 13 张牌，结果存放在 4 个数组即可。

第二部分：电话聊天，45min

- 1. 讲解解题思路及代码
- 2. 聊项目
- 3. 工作意向

最后，感谢强哥和铁男的讨论、支持、帮助及精神鼓励，感激不尽~~~~

时长：45min

阿里

我这边阿里主要面了两面三轮，大概情况如下：

一面：

电面，主要是问项目中的算法细节，没问其他的。

二面：

我这边主要是面的阿里妈妈的算法工程师，搜索广告事业线的部门，所以面试题主要是跟算法相关的，题其实也没几道，但问的比较深；然后就是依托项目的算法，问的比较深。面了两个人，每个人大概一个半小时左右。

一面面试官主要问了我两个项目，一个是我现在手上做的矩阵分解的项目，一个是以前大作业做的微博抑郁预测，关于项目个人觉得应该注意以下几点：

1. 这个项目的算法要给他讲清楚，因为面试官可能并不熟悉你的算法，最好有一些简单的推导公式给他说明。
2. 结合你的算法说明你的应用场景，这个东西你不说他也会问，你的应用场景是怎么套到你的算法中去的，其中包括一些特征处理的东西。
3. 你的算法和类似算法的比较，比如我的 SVDFeature 算法中有通过一个对数函数来映射模型，他会让你比较 Logistic 和你这个算法。
4. 还有就是模型的目标函数，就是损失函数，这个两面都问了。
5. 然后他可能会问你的模型能不能用到他们的应用场景中去。

然后就是一些海量数据处理的问题，我这边两个面试官问了三道：

1. 给你一个文件，每一个记录都是以 `userid:action` (`userid` 可能在这个文件中出现多次，`action` 有多种 `action`)，另一个文件每一个记录是以 `userid:userinfo` 的形式存在（这里 `userid` 是 Unique 的），你怎样把这两个文件 join 起来表示成 `userid:userinfo:(actions)`。
2. 给你一个有几十亿记录的大文件，每一个记录是一个单词，然后再给你几个单词，你怎样很快判断这几个单词是否在里面。

以上两个是一面面试官问的，说思路就可以，下面蛋疼的来了

二面面试官问了个简单的海量数据处理的问题，但每一步都问你为什么，我感觉他是在看你的思考过程，然后让你说这样做合不合理。

3. 给你一个 1 亿记录的文档，每一个记录是一个 query，怎样统计出出现次数最多的前 top10。（这里面涉及很多为什么，比如你认为每一个 query 的长度大概是多长，为什么这么长，我这里每一个 query 是汉字，那么汉字你认为每一个时多大能够表示，为什么，诸如类似的这些为什么，结合他们的应用场景，也就是淘宝的应用场景估计每一个 query 大概几个词，然后你根据你的推断来计算这每一 query 大概多长，然后看能够装进单机的内存，然后根据是单机版的还是分布式的写代码）

除此之外，分布式计算框架，hadoop 需要了解一下，我这块儿不是特别了解，但通过别的方法绕过去了，比如 `map/reduce` java 版的我没写过，我说我写过 `pig` 统计数据，然后用 `pig` 给他说的思路；然后 `hadoop` 大概框架我知道，但不是很深的理解，但我用过 `spark`，就跟他扯 `spark`。

大概就是这些东西，没问类似 `leetcode` 的算法。

后续更新百度面试的情况。

阿里 2 面

面试的组是做淘宝中间件的，其实就是淘宝的服务器了，双 11 支撑整个淘宝流量的主要组件都是这个组做的。

面试过程就是很随意的聊天，他一边介绍他们的工作，我一边问着。中间穿插着让我讲讲项目，并在项目的基础上扩展，说几个还记得的具体问题吧：

1. 多台 MySQL 组成一个分布式数据库，怎么处理死锁？

没答上来，最后面试官告诉我用两阶段锁。

2. 遇到系统较慢时，用户会不断点击刷新之类的按钮，增加系统负载，怎么处理？

用户点击一次刷新按钮后自动锁定，直到请求返回或者超时之后接触锁定。

3. 现在需要做一个简单的 Java 应用，对用户的每个请求都返回服务器当前的时间戳，服务器 32 核，内存等都足够。但是任意一个核获取系统时间的系统调用 `getTime` 都需要花费 1ms，在这 1ms 中会锁定总线，故整台机器 1s 内最多可以执行 1k 次 `getTime` 操作。现在想用这台机器支持 10w 的访问量，应该怎么办？给用户返回的时间戳精确到秒，且允许 $\pm 1s$ 的误差。

显然是要想办法复用 `getTime` 操作，减少 `getTime` 的使用次数。问题就是在于如何控制自动 `getTime` 的间隔。我的整体思路是用生产者-消费者模型，一个单独的线程通过 `getTime` 将时间戳放到特定内存中，其他的线程去读。在控制 `getTime` 线程的时间间隔上卡住了。开始是想利用 CPU 的调度策略，保证每秒里 `getTime` 线程都会被运行一次，但面试官说不能对调度做任何假设。然后想到了利用硬件中断来实现间隔，面试官肯定了这个思路，问具体实现。我说可以用 C++/ASM 写好再用 JNI 调用，面试官让再想，最后还是没想出来。。。面试官问我知不知道 `Sleep` 函数，其实我知道，但是这种情况下只好假装无知地说以为 `Sleep` 函数是采用轮询的方法而不是中断的方法。面试官没在追究了。

感想：阿里这轮面试特别扣项目，尤其是和面试官所在组相关的项目，重点是考察是不是在做项目的过程中对涉及的技术有足够的了解（不能说深入，但应保证不会由于对第三方技术了解不够而影响自己的项目，比如第一个问题中我的项目涉及到事务的正确性，但是我没答上来，算是个反例了）。比较重视思考的过程，只要中间思路对了，结果有点偏差也 OK。这样纯聊天的面试中积极和面试官交流思路，不要觉得思路挫就不说，然后对面试官的思路积极响应，如果觉得面试官思路不对也可以提出自己的看法。

阿里

时长：1h

部门：数字娱乐（交叉面）

1、数据结构

a) 问平时用哪些数据结构，列举了一些算法题里经常用的数据结构，然后问我平时有在哪里用到栈，我说工程里没用过，只在做题时用过，然后给他扯了一个用两个栈实现队列的算法。

b) 数组和链表的区别。如何在链表中插入一个节点。

2、排序算法

a) 问我平时用哪些排序算法，列举了一些，然后写了一个插入排序（基本什么没说写什么，服了）。然后问了一下复杂度。

3、多线程

a) 怎么实现线程同步。

b) 场景：创建一批线程去执行任务，如何获知所有的线程都执行完了。

c) 线程调度算法，讲一下 LRU 的实现。

4、其他

a) 如果后台加密文件的加密算法被破解了，怎么在保证服务可用的前提下，防止文件被盗取。（开放题目）

b) 项目里的工作、项目的改进

c) 自身的优缺点

阿里终面

技术面：

继续扣项目。貌似这轮面试会决定最终的评级，所以面试官的态度就是让我看看你最牛是啥样子。

技术上问了三个题：

第一个是构造函数可不可以为虚函数。他的期望答案就是构造函数为虚会编译错误。但我没有验证过，所以就说了为什么不能构造函数不能为虚函数，最后他算是认可了吧。

第二个是构造函数可不可以调用虚函数。这个不知道可不可以，我就瞎扯说跟编译器实现有关，取决于编译器什么时候把虚函数表放置到对应的对象中。事后查了下发现 C++ 规定了这种行为的具体实现

<http://stackoverflow.com/questions/962132/calling-virtual-functions-inside-constructors>。

第三个是写段代码判断一台机器是大端机还是小端机。利用 c++ 中的 union 即可。

也问了很多项目的事情。这个面试官很推崇阅读源代码，感觉阿里的技术文化是比较务实的，看一个牛逼不牛逼就是看有多少深入的技术实践，这里的技术实践主要是指深入钻研一些系统，比如 hadoop, nginx, MySQL 啥的

HR 面：

会问有什么项目印象比较深，什么比赛印象比较深，为什么学习计算机，理想是啥，bat 三家怎么选之类的。

回答问题的目的就是刻画自己，比如我是借数模比赛夸自己有跨领域的团队合作经验，然后讲讲自己从小学电脑，对这个很热爱啊，平时很刻苦很喜欢学习啥的。也分析了一下微信这个产品，顺便谈谈人生谈谈理想。最后问了 bat 都有 offer 话选哪家，对这种问题我的策略是不直接回答，谈谈自己对 offer 选择的策略，剩下的让 hr 自己去想吧。

总结一下整个阿里面试的感受：

阿里的技术文化就是两个字：务实。

面试中体现的很明显了，一切从你做的东西出发，看你怎么做怎么想。也会从语言，OS，网络这些工作中会用到的基础知识出发提问。阿里对于大牛的定义是对一个可以直接产生业务价值的系统有着深入的了解。比如 nginx+刀片机可以取代 cisco 的专门的负载均衡设备，但是成本大大降低，所以深入了解 nginx 就是有直接业务价值的。

我猜他们认为算法对工作的帮助作用不够直接，所以不怎么提问。

阿里 1 面

阿里校招一面 不到一个小时

研发面试会分为 java 和 c++_ 分别来面，我面的 java。没问数据结构算法这类的题目，据说其他的同学有问。

1. 自我介绍。我介绍项目是打断了我，说先介绍大概情况，再谈项目。
2. 使用 java 多长时间。主要用到里面的什么技术？
3. java 线程安全的机制？

静态的 synchronized 和非静态的区别？

4. jdk 版本问题

最新的版本是什么？每个版本之间的区别了解吗？做了哪些改进？

5. 总结一下 java 的特点。

提供垃圾回收。怎么实现的？

怎么判断一个对象要回收？回答：不可达。怎么不可达？Gcroot 怎么形成的？。

6. jvm 内存划分？怎么监测查看其内存具体情况，就是堆使用的情况等？
具？然后他问怎么实现的？

7. 对 jvm 和 java 熟悉到什么程度

8. spring 的特点？

HTTP 协议的重点是什么？

Tcp 协议三次握手，滑动窗口

Java 序列化：XML，json

9. 项目：问项目的一些实现，你做了哪些工作。从用户角度你觉得你们项目还要做想哪些改进（觉得像面产品的）。。。还有就是点领域知识。

10. 你想做那个方向的？除了对性能测试领域还对什么其他领域有关注？还做过其他项目吗？想去哪个城市。。。。

谢谢大家！

阿里

一面

1. 自我介绍

2. 简单问了问实习和项目

3. 算法题

把一个 int 数转换成它的读音

二面

1. 自我介绍

2. 两个非递减单链表，都可能含有重复元素，要求将他们合并，合并后的为递增，即重复元素只保留一个

我写了一个，有两个问题：

1) 没有 delete 不要的节点

2) 他问如果两个链表有交点怎么办, 这样由于 delete 的一个点在公共部分的话另一个链表也断了

三面

1. 自我介绍
2. 项目介绍
3. 写一个单例模式
4. 一个数组, 分成 m 段, 把这 m 段打乱顺序, 即段内有序, 段间无序, 让排序。并且问复杂度
5. 给一些矩形, 下面的边都在 x 轴上, 相互覆盖, 求轮廓

阿里

商业基础平台部, 研发工程师

一面:

- 1、快排和归并排序的时间复杂度, 快排和归并排序比较、应用场景 (我就从额外空间和稳定性两方面答的)
- 2、大文件, 里面存的一个个字符串, 对字符串进行排序
- 3、对 int 流实时求中位数, 内存够用
- 4、手写算法: 求数组逆序对的算法
- 5、其他时间在聊项目、聊工作 (部门主要做什么 blablabla)

二面:

- 1、问了 C++ 的问题, 问语言特性 (我不会!)
- 2、线程和进程的区别、线程之间通信方式、进程之间通信方式、线程之间共享数据的同步
- 3、给定应用场景: 很多线程, 共同操作一块数据, 其中读的线程很多、写的线程很少。
设计一种数据结构 (其实就是设计一种方法), 满足: a) 写不能影响读的性能 b) 写可以牺牲一点实时性
- 4、手写算法: 单项链表, 奇偶位节点互换

5、两个大文件，里面存的都是整数，找出两个文件中共同的整数

三面：

三面一直都在聊项目，讲技术，其中涉及到分布式的比较多，比如：设计分布式系统应该注意哪些关键问题啊。。。

心得：

- 1、有些题目，不能因为见过就马上开口，一定要想清楚、自己把过程走一遍再说
- 2、回答问题的时候可以发散性强一点，用自己熟悉的知识掩盖在问题上的知识漏

豌豆荚

时间：1.5h

没有聊项目，主要问了 Java 语言相关的，还有 web 服务相关的知识，然后做了一道题

- 1、JVM 内存是怎么划分的
- 2、什么时候会发生 minor gc
- 3、用过 spring 哪些注解
- 4、@Autowired 和@Resource 的区别
- 5、AOP 的实现原理
- 6、抽象类和接口的区别
- 7、泛型的好处，List<T> 和 List<Object>一样么
- 8、有没有了解过 java 6、java 7、java 8 的区别
- 9、IOException 是 RuntimeException 么

10、session 和 cookie 的区别

11、像淘宝的保持在线是怎么实现的

算法：

请编码实现把一串文本里面的任意的不超过五位的中文数字转换成阿拉伯数字，如二十三转换成 23，一百二十转换成 120，输出转换以后的数字。

豌豆荚电面

1. 设计算法，ab 两个大文件中各有一亿大小 url，每个 url 占用 64 字节，内存只有 2G，怎么找到重复 url？
2. 设计一个类似于 qq 音乐中的猜你喜欢产品，可以从算法或者工程的角度回答。
3. 有一个很大的日志文件记录用户登陆情况，有三列记录，分别是用户 id，时间，登陆或登出。设计一个算法，计算某个时间在线用户的数量。

豌豆荚

1. 设计算法，ab 两个大文件中各有一亿大小 url，每个 url 占用 64 字节，内存只有 2G，怎么找到重复 url？
2. 设计一个类似于 qq 音乐中的猜你喜欢产品，可以从算法或者工程的角度回答。
3. 有一个很大的日志文件记录用户登陆情况，有三列记录，分别是用户 id，时间，登陆或登出。设计一个算法，计算某个时间在线用户的数量。

豌豆荚四面

时长：2h

1、为 n 个人分配 id。

输入：数组 $a[n]$ ， $a[i]$ 代表第 i 个人要求 id 的范围为 $1 \sim a[i]$

输出：有多少种分配方案

2、x 为正整数，初始值为 1。x 一次能乘以一个质数或者乘以 x 的一个因子。

问：x 经过多少次运算能得到 A 的 B 次方（A 和 B 为输入，正整数）

3、其余时间全在聊项目

豌豆荚

三面

一、浅聊项目；

二、讨论 vector，讨论 string

1、vector，多线程 push_back();

2、string b = a;发生了什么，争论了半天

三、100 台机器，磁盘各有 1G 的 query，

1、统计所有 query 的出现次数

2、排序所有 query

要求最快

四面

一、浅聊项目；

二、循环数组，找某数字的 index；

豌豆荚终面：

CEO 面，问希望豌豆荚有 Google，MS，Hulu，有道这些公司的哪些特征？
然后就是让我问问题，我没啥问题就让我走了。一共也就 10 分钟多的样子。

豌豆荚

一面：java 多线程相关的语言特性，实现单例（注意多线程安全）

第一题是给个数组，求 $\max(a[i] - a[j]) \quad i > j$

第二题是画了个树，然后给了这个树的遍历方式，然后和他 check 你明白了他想要的遍历方式（其实就是中序，如果有右孩子的话就再输出一根），然后写了两段代码，一个是给树输出遍历结果，一个是给遍历结果然后还原树。

二面：leetcode 上那个矩阵每行递增每列递增，让你找个 target 的。。但是他不同意那个右上角开始的 $O(n)$ ，硬是让用二分，然后写复杂度。之后又是语言特性，还让比较 java 和 c++ 的不同之类的。。。

三面：基本全部是 map reduce，找出现次数 top 1000 的 query，给出现次数排序，还有做两个表的 join，join 这个题一直在各种改条件，一个表很大会怎么样，很稀疏会怎样，比较小会怎样之类的。最后让写找 top 1000 的代码，恩，代码。

然后问了 hashmap 怎么做线程安全

四面： 问问项目，扯了扯推荐系统相关的。然后正则项的作用，11 和 12 的场景。

前三个面试官全程高冷>< 被虐的，唉。。

豌豆荚

2014/09/03 豌豆荚 后端研发工程师面试

1 面：

长度为 n 的数组求 topk，要求用 partition 实现并分析平均复杂度。

2 面：

LRU Cache 以及 unordered_hash 的实现，问了个对于哈希桶数量的统计学估算
两个叶子节点寻找最近的公共祖先，据说可以用递归一次遍历解，我写了个非递归。

3 面：

旋转数组的查找（345612）
数组循环右移 k 位的实现

4 面：

开放题，给个场景，要求设计策略。（尼玛这货是个搞算法的！！！）

n 个 app，每个 app 包含以下信息：

1.title 2.description（描述） 3.comments（用户评论）

需要把这些 app 分成 15 种类别，这些类别是已知的。

显然是个分类问题，于是各种扯，扯特征提取、分类算法选择、计算指纹，监督还是半监督，怎么解决过拟合。。。

总结：前三轮感觉面的还算 happy，第四轮被逼的死去活来，这是要给跪的节奏。。。

对，还想吐槽一句，每轮都问项目，都在那抠细节是什么尿性，难道每一轮面试官就不能交流下前一轮问的东西吗？？？

豌豆荚

一面：

1. strcpy，不考虑覆盖和给定内存的长度的有效性
2. 设计实现 malloc 和 free。假设可以使用的系统内存是 char buf[MAX_SIZE]，所有动态申请的内存都需要在这个 buf 上申请，比如想实现 vector 话，就要自己在 buf 中划一个区域维护。

函数原型如下：

```
void* malloc(int size);  
void free(void* mem);
```

3. 随机生成语法树：假设有 a AND b, NOT a, a OR b 三种语法树，给定两个函数 random_operator(), random_char()用来随机生成 AND/NOT/OR 和各种字符。实现一个函数，Node* generate()，用于返回一个随机生成的语法树。

generate 是需要有参数的，面试官故意不告诉你，看你能不能发现这一点，等你发现后，会告诉你参数是树的高度，所以函数原型是 Node* generate(int height);

二面：

1. leetcode 上的 LRU cache
2. 给定一棵树和两个节点，找最近公共祖先。没有父指针。

三面：

1. 实现 RWLock，可以使用 Mutex，以及++/--操作都是原子的。

豌豆荚

1 面：聊项目，主要聊爬虫和异构检索

题目：一个数组，前半部分升序，后半部分降序，在这个数组中查找某个数

2 面：聊项目，还是聊爬虫和异构检索

题目：大数的加法，考虑负数

3 面：聊项目，还 TM 是聊爬虫和异构检索及推送

题目：将链表 $a \rightarrow b \rightarrow c \rightarrow d \rightarrow e \rightarrow f \rightarrow g$ 转化为 $g \rightarrow e \rightarrow f \rightarrow c \rightarrow d \rightarrow a \rightarrow b$

4 面：聊项目，终于聊了我的 Bench4Q 啊，一顿神侃，聊的很开心（他询问一些在项目中可能出现的问题，看你怎么解决）

聊人生，聊理想

豌豆荚

一面

1.问项目

2.杨辉三角，输入 row，返回 $\text{List}<\text{List}<\text{Integer}>>$

3.二叉树层次遍历，输入 root，返回 $\text{List}<\text{String}>$

二三面

见其他人的面经文档

2014-09-15 豌豆荚 研发工程师 CEO 面

CEO 还是老套路，先根据个人项目或者经历问几个问题，然后问我还有没有问题要问的。本着基本是个跪的思想，我问了七八个问题算是恶心一下他，有的问题他答的很是模糊估计是涉及商业机密的。大概就这尿性，坐等被跪的节奏。。。

人人 onsite

一面：

1、二叉树变形为双向链表。

2、大数乘法。

二面：

1、字符串去除重复地和多余地空格和换行。

- 2、大数据单词去重。
- 3、一个最弱地排列组合概率题，不好意思写出来，就当大家都会。

三面：

- 1、项目
- 2、设计模式用到哪些，策略模式怎么用地，写 double check 地 singleton
- 3、聊人生，谈理想。

人人 onsite

一面：

问项目

pow (a, n)：递归+迭代

二面：

bst to double linked list

人人

一面

- 1.两个有序数组求交集
- 2.大数加法与大数减法

二面

- 1.一个 $O(1)$ 时间内找到最小值的栈（min 栈）

三面

经理面，聊人生聊理想

人人

人人三面：

问了点项目相关的东西，然后就是聊聊人人网的后台发展历史，以及现在的架构。

人人

【一面】

- 1. 求数组第 k 大的数

2. 写个快排

3. 一个字符串，最少添加几个字符，变成回文

【二面】

扯项目，最后扯到你参加过什么社团？做过什么义工？爱好是什么……

【三面】

二面没做题啊？来

1. 一个 int 数用汉语读出来，比如 1234，读作“一千二百三十四”

2. 一个有序数组长 n ，分成 m 块，然后打乱顺序，要求重新排序，怎么排，复杂度多少

问什么没有和我聊人生 聊理想？……

其他人的人人面试题

1. 顺时针打印矩阵

2. 输出最小值的栈

3. 大数相乘

4. 写一个堆排序（他说自己用库 `priority_queue`，这题没写）

5. 生成旋转矩阵（leetcode 原题）

文本中去掉多余空行 连续空格

人人

1、找出一个数组里，两段连续字段和的最大值

2、原地去掉一个有序数组中的重复数字

3、不断的扣项目细节，直到发现你应该做但是没有做的东西为止。例如：你在进行性能测试的时候，都需要调哪些参数啊？（blabla）如果你想达到单机 6w 以上并发应该怎么调啊？（没跳过）-》告诉你怎么调都不行

搜狐

围绕项目，问的很杂。

1. redis 是什么？
2. 对 nginx 的了解。
3. 对 tomcat 的理解。
4. 最感兴趣的方向。
5. http 请求有哪些类型？
6. hbase 和传统的数据库有什么区别？

搜狐

2014-09-12 搜狐大数据中心 数据挖掘 一面

搜狐面试感觉比较看重项目，也可能是因为项目和他们在做的有点相关，在项目方面抠的很细，问了很多项目逻辑的问题，不过关于底层实现涉及的不是很多。在问项目的过程中问了一个问题：

1. 关于倒排拉链如何求交

由于在写代码的时候我用了个 stl 的 `set_intersection`，面试官好像觉得我对 stl 的掌握程度还不错，关于语言特性上的问题基本没问。问了问编译系统的问题，比如对 `MakeFile` 的掌握程度，我说我现在都是用 blade 于是就在讨论 blade 的工作模式。还有就是动态库与静态库在链接的时候有什么区别。

大概就这么多，就写了一个代码，其他的都是讨论，感觉面试官人很 NICE。

搜狐

2014/9/18 搜狐大数据中心 二面

先是和中心的两个 boss 聊了聊，由于我想搞研发，给我找了个研发的工程师进行了一轮面试。这哥们口音很重，工程能力应该很强，问的问题我觉得很有难度。

1. 聊项目，抠细节，抠的非常细，在有的细小的点上很深入。
2. 工程问题，如何实现一个线程池，保证高效。
3. 如何用 C 语言设计一个 C++（把 C++之父那本书和深入理解 C++对象模型的内容说了一遍）
4. 实现一个 `hash_map`，完成 c 风格字符串到字符串的映射，要求不使用 stl 容器，map 中保存完整字符串而不是指针。要求是给一个定长内存作为这个 map 的可用空间，需要把内存空间用到极致，与此同时尽量提升查找速度，保证多线程安全，支持 LRU 等等。

总结：对工程要求很高，顿感自己太年轻了，面试中各种被教育。这把虐的我心服口服。。。

搜狗

一面

- 1、问了项目，特别细节
- 2、合并 n 个数组，lc 上题
- 3、找出 n 个数组相同的数
- 4、返回 `vector` 引用
- 5、红黑树咋实现的。。。 (居然问这个) 不懂
- 6、C++写一下单例 (`static` 的，不满意)

二面

问了一个，如何找同义词库 (给你一坨网页，给你一坨 `query`)
如何建立一个搜索推荐词库 (类似百度搜索下面有一条的)

三面

自我介绍

然后给一个数组如何压缩 (没想到好方法)

1 亿多个字符串文件，找出中位数

然后简述一下你对 `cache` 失效时间的做法。

搜狗 1 面

【无线搜索部】

1. 项目

2. 做题

- * 去掉一个字符串中前后的空格 (中间不用管)
- * 搜索框中输入一个查询字符串，会有提示。提示是从最近的用户 `log` 中提取出来的，假如是 100 万条，每一条是关键词+词频。怎么实现这个提示功能。

搜狗

一面

1. 太复杂，记不清了……
2. 写一个 `string` 类，实现赋值函数
3. 一个矩阵，求 2×2 的子矩阵中和最大的那个
4. 一个字符串，求出现 n 个不同字符的最短距离

5. 10 亿个 doc 文件，每个约 10W 词，认为两个文本之间词汇出现次数差别小于 1/5 就是一类，求分类

二面

1. string to integer

2. GPS 如何定位

3. 聊天

搜狗

2014/9/19 搜狗搜索事业部 数据挖掘研究员

先做了一套卷子，是 2013 年搜狗校招的卷子，尼玛两个小时的体量，坑爹啊。做了一个半小时实在是不想做了就交卷了，但是前一个同学貌似还没面试结束。墨迹了好一会儿，后来终于给我临时找了个面试官。小哥是北大毕业的，感觉工作也没几年，人很 NICE。先聊了聊项目，主要是聊天没有深问，后来问了我一个题：

1. 求三国杀里神吕蒙技能的期望（从牌堆上方拿 5 张牌，可以拿走与花色数相等的卡）
2. 无限牌堆和有限牌堆的期望相等吗？为什么

问完之后开始聊天，问我平时喜欢干啥。我说喜欢玩 DOTA，结果这哥们眼睛一亮，开始追问天梯多少分。。。balabala

总结：本来抱着基本是个跪的信念去的，去了以后又是那么多的笔试题，当时就想往回走了。后来想了想既然来了，这么随意就放弃也太坑爹了。果然，人生就是有太多的不可预测，不过面试结果咋样至少我觉得这次面试挺 happy！

猿题库

1 面：

- 1) 旋转数组查找目标值；
- 2) 有序数组中，只有两个值发生了交换，查找目标值；
- 3) 链表的分段旋转

2 面：

- 1) 两个树节点的公共父节点
- 2) 忘了（但是超简单）
- 3) M 个长度不同的链表，按从小到大输出（代码量较大）

猿题库

1、旋转图片，图片由矩阵表示

2、输入一个字符串，处理之后，输出的字符串中不含 5 以及 '10'

3、（比较有意思）有一段长度为 10 公里的公路，在公路上分布了若干个路牌，每个路牌有两面：左面标识的是当前路牌到终点的距离，

右面标识的是当前路牌到起点的距离（以公里记）。但是路牌的质量并不好，可能会被风吹之后反转了。

问：给定一个数组，代表这段公路上每个路牌左面的标识，问该数组是否是一个合法的路牌序列。

例如：数组 8, 4, 2 合法，数组 8, 8（可反转成 2），1 也合法，但是 5, 6, 6 一定不合法。

猿题库 1-3

一面 60min:

1、用数组实现两个栈；

2、Rotate Image;

3、Path Sum II;

4、判断 n 个矩形是否有重叠（原题坐标范围很小，可以暴力；我打算 zb 写一个没有坐标限制的线段树，不过没写完）。

二面 60min:

1、聊项目；

2、Reverse Words in a String;

3、删除链表中 value 等于 target 的节点；

4、判断一棵二叉树是否是另一颗的子树。

5、一个 5000 长度的字符串，输入 10W 次查询，每次查询[s,e]区间内有多少个回文

三面 35min: \\

1、大浮点数加法；

2、聊项目；

3、扯淡。

4、用过哪些 C++类库，说 vector，栈，队列，然后让自己想功能，实现一下 vector

美团

1. 寻找链表的倒数第 k 个值
2. 两种树之间的相互转换：多叉树和斐波那契树（只包含左孩子和兄弟节点的指针）
3. 十进制数转 N 进制数
4. 封闭一个机房内最近 5 分钟内访问流量最大的 IP，如果是多个机房怎么办
5. TCP 和 UDP 的区别，虚函数及内存布局，进程和线程的区别

美团

1. 自我介绍
2. 讲一个你认为有挑战的项目
3. 不会 java 非要问我 java
4. shell 了解到什么程度
5. 网络编程会吗
6. 说说虚函数、模板
7. 说说进程和线程的区别和联系
8. n 个数里随机选出 m 个
9. 求链表倒数第 k 个节点
10. 十进制转 n 进制

美团

1 面 70 分钟

1、项目

2、基础：Integer 缓存，HashMap 的与负载因子相关的问题，java 堆的分区，mysql 的联合索引，java 线程通信机制，多线程的应用场景和不适宜场景，LRU cache 实现原理

3、题： N 个数随机取 M 个数。。链表倒数第 k 个元素。。。10 进制转。。。10 进制转汉字（类似于 leetcode 转罗马数字，考察特殊用例，表示不同不懂壹贰叁。。。）感觉是现从自己的 mac 上找的题而且不写 java 的样子，所以沟通有点不在一个节奏上。

2 面 60 分钟

1、项目+聊平常的 code 的乱七八糟的事。

2、做题：1、平面上 n 个点坐标为整数，以点为圆心做单位圆，求所有的圆覆盖的面积。

2、同样是 n 个点，这些点都是地雷，地雷的爆炸覆盖半径都一样是 r ，覆盖范围内的地雷会被引爆，指定一颗引爆，求能够全部爆炸的最小半径 r 。

3 面 30 分钟

一副很鄙视我的样子的面试官，就带着个小米手机就进来了，先问了项目，又鄙视了一会儿，问了我上过什么课程，我说了些高大上的，又质疑我是不是没学过计算机基础课，我说本科时间长了有点忘了，他说他毕业 10 年了都没忘。。。感觉当时时间凝固了 10 秒。。。问 socket、tcp 拥塞，然后问虚拟内存，然后问 windows 怎么调，调过没有。。我说如实说没调过。。他又质疑我是不是学计算机的。问我的 windows 慢吗？什么时候慢？原因。。又聊起 mac，问我知道不知道 mac 怎么锁屏。。。我下意识以为有什么 shell 命令，就说不知道。。然后他呵呵了，问我为什么要用 mac。。。我就说我都把盖子扣了就走人了，因为是固态硬盘不用担心弄坏，然后在讨论固态硬盘为什么快。。他表示不同意我说的。。。看我用过 google map 的 api，然后打开了他的小米手机给我看他的百度地图，提问：怎么找两点最短路线，百度是什么策略找路线。。又呵呵我。。然后就没有然后了，走之前冒昧的想问问他平常用什么技术，他凶狠地说 php, python, ruby 什么都用。。NM。。

美团

一面 60min

- 1、链表倒数第 k 个节点
- 2、堆排序
- 3、螺旋打印 $1-n^2$
- 4、java 反射
- 5、HashMap
- 6、项目

二面 60min

- 1、聊项目
- 2、聊感兴趣的技术（我说了 java、scala、函数式、平时看的技术博客 blah blah）
- 3、设计一个电影票预订系统（接口、数据结构、类、存储、并发，重点是接口设计）

三面 40min

- 1、瞎扯人生理想职业规划心路历程
- 2、穿插着问了一个虚拟内存和虚拟地址空间

有道笔试

请大家把记得的题目写在这：

1. 贡献一个自己做错了的题：

n 个杯子杯口向上，每次翻转 $n-1$ 个杯子，多少次翻转后杯子可以全部向下。

n 为偶数， $f(n) = n$

n 为奇数，不可能完成

//edit by C

简答题

1. n 个杯子，最初都朝上，每次可翻转 $n-1$ 个杯子，问经过多少次反转可以使所有杯子都朝下

2. 安卓 App 如何使用 C 代码

3. 程序改错

编程题

1. $1 \rightarrow 2 \rightarrow 3 \rightarrow \dots \rightarrow n-1 \rightarrow n$ 变为 $1 \rightarrow n \rightarrow 2 \rightarrow n-1 \rightarrow 3 \dots$

2. 忘了

算法题

给一个序列，每个元素是一个人，每个人有两个域：weight 和 height，现在要求上面的人必须比下面的人 weight 和 height 都要小，问最多可以叠加多少人(动态规划)

微信 2 面

面试官超高冷，一共面了接近两个小时，问了四个算法题，然后挑项目问了一下。

1. 问区间 ab 有多少个数的不同素因子个数有 2 个，例如 $10=2*5$, $20=2*2*5$

2. 给出已知 n 天的股价，现在允许在这 n 天里面进行两次买卖，问如何进行可以使得得到的收益最多

3. 给出一个 $n*m$ 矩阵

请按照下面例子的规律来输出这个矩阵

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

1->2->5->9->6->3->4->7->0->1->8->2

4. 给出一个字符串 s, 判断 s 中是否含有一个长度为 k 且其中字符各不相同的子串

问项目时感觉被鄙视了，意思就是做的不够深入。

知乎

搜索研发
一到那里就拿出 mac 让在 vim 中写代码，编程题只有一个，其余都在聊天。2 个工程师，2 个 hr+联合创业者。但是要实习才能转正。

知乎的话题是分层次的：
输入数据如下：

A	C
A	D
A	E
B	G
F	A
F	B
H	J

格式化输出：

F	A
	C
	D
	E
	B
	G
H	J

其实就是类似输出一颗树，打印出来，python 实现。

小米

一面：

- 1、字符串转浮点数。
- 2、给一个字典，和一个字符串 s，找出字典中所有满足以下条件的字符串 t：
交换 t 中的字符顺序可以得到 s。

二面：

- 1、双向链表中的节点两两 swap。
- 2、给出一个双向链表，其中有一个指针指歪了（可能是某个节点的 next，也可能是 pre，但是指向的还是链表中的一个节点），找出这个指针并修正。

据说应届生只有两面。

小米

2014-09-17 小米 服务器开发 1 面

妈了个蛋，一定要吐槽下这次面试，是我面试以来体验最糟糕的！！！！

上来自我介绍，我开始说自己的经历，说实习经历、大致介绍实验室项目情况，这哥们首先不关心项目，皱了皱眉问：你们中科院不管学生？我大概介绍了科院的情况。他追问：你研一是不是不去上课？我TM的想抽他，告诉他我上课只是上的少。感觉他对这个很反感，莫名其妙。

开始抠项目，在那纠结搜索引擎是不是必须得用 cache？我说数据规模没那么大的时候只要能够快速定位倒排索引的文件 offset 就能保持高效，然后他又在那纠结读写一次文件要多久时间等等，好吧这些也就算了。。。之后他问我你刚才提到过跳表，给我实现以下吧？我TM没实现过，我说先讲讲逻辑行不行，然后给他讲了讲逻辑。他又说，那你实现一下跳表吧？我说我不知道跳表节点里面具体怎么存！他说：就根据你的想法实现一个，先写跳表的插入和查询吧？

妈蛋，开始苦逼之旅，我在那写代码，他一会儿一个电话一会儿去面吴桑，来回折腾。实现完了代码，他告诉我一般不是这么实现的，让我用他的思路再完整实现一次。终于实现了，回来他在纸上画了几个图，说这里面还有挺大问题，让我再完整实现下，改改问题。于是我TM又开始改，发现是一个指针指错了地方，改完想找他反驳。。。我凑，这货直接不看代码了，给我安排了个二面完事！！！！

2 面

二面是师兄面的，感觉比较轻松愉快，出了两个题，师兄说如果题目见过就告诉他，做见过的题目就没啥意思了。然后出了两个确实有应用场景的题：

1. 配置文件的解析，配置文件类似 `a:123 b:1{a}2` 其中 `a:123` 表示参数 `a` 的值是 123，而 `b:1{a}2` 表示 `b` 的取值要依赖于 `a` 的取值，在这个例子中 `b` 的取值就是 `b:11232`。要求给一个配置文件，解析出其中所有配置项的值，配置项可能是循环依赖，这样就无法正确解析，要求让这些不能解析的参数取值为 NULL。编码实现。
2. 类似 `nginx` 的加权轮询策略，有 `N` 个机器，每个机器有个权重，让你根据这些权重做负载均衡，编码实现。

就这尿性了，感觉给跪是妥妥的，就这样吧。

小米

- 1、链表 mergesort
 - 2、系统设计题：小米每周二有抢购，那么要防止黄牛作弊，做这样一个 Antispam 的服务，在每次用户登录的时候会调用验证是否有作弊嫌疑；并且在用户登录之后操作的过程中，每个操作都会生成 log 并交由 Antispam 分析处理。
确定了数据存储的格式之后，开始讨论一层层的组成。包括需要满足 5wKPS 的要求。
 - 3、有依赖性的程序模块并行编译（拓扑排序）
 - 4、最长连续子序列
- 注：面试官没有准备，都是临时想的题

小米

一面

1. 太复杂，记不清了……
2. 写一个 string 类，实现赋值函数
3. 一个矩阵，求 2x2 的子矩阵中和最大的那个
4. 一个字符串，求出现 n 个不同字符的最短距离
5. 10 亿个 doc 文件，每个约 10W 词，认为两个文本之间词汇出现次数差别小于 1/5 就是一类，求分类

二面

1. string to integer
2. GPS 如何定位
3. 聊天

小米

讨论了下项目和机器学习的模型，然后就是做题，每一面都如此。

题目：

一面：

1. 判断一棵树是不是判断一棵树是不是二叉查找树
2. 判断单链表是不是回文

二面

1. 链表翻转
2. 给定一个包含字符串的字典，和一段文本，查找一种切分

三面：

1. 讨论了下 map reduce
2. 给定一个包含字符串的字典，和一段文本，用字典中的词匹配文本，最长匹配多长。（和二面 2 题类似）

小米

先问问了问基础知识和项目，基础知识就是 jvm 垃圾回收、多线程安全这些东西。

一面：

- 1: 查找数组中第 k 大得数。（我用的堆）
- 2: 根据 1, 0 的二维矩阵求色块，例如

1	1	0
1	0	1
0	0	1

上图色块个数是 4

先写了个暴力，后优化，只讲了思路

二面：

- 1: 把 BST 拍成链表
- 2: 有向无环图，邮递员送信，求最短路径
邮递员从一个节点出发，到达目的地后可以不回来。每条边权重就是两个村子的距离（动态规划，写了递推关系，不要求实现）

Yahoo

一面 50min

（这轮面试官比较 suck，总爱问 RTFM 的问题）

一些简单 java 基础（具体忘了）；
怎么远程调试 Java 程序（jpd）；
如果服务器上的 java 程序 hang 住了，怎么查看具体是哪个线程的问题（jstack, jconsole）；
几个 linux 命令（不会）；
写一个 Java Singleton（面试官也不知道 volatile）；
一些 web 相关的问题（web 容器、web 框架、netty、thrift）；
一些 TCP 细节（三次握手、ack 机制等）；
最后问了几题 LC 原题，我说都见过，就只说了思路，没写代码（Copy list with random pointer，环形链表找入口，其他忘了）；

二面 50min

- 1、聊项目（问的面很广，但是不深）；
- 2、给你一堆时间戳，和一些区间，过滤出在这些区间的时间戳；如果时间戳特别多怎么办？如果时间都是一天内的（单位为秒）怎么办？

三面 部门 Leader 40min

- 1、从数组中找到两个不相交的子序列，和最大；
- 2、简单聊项目；
- 3、设计微博时间线；
- 4、用 MapReduce 解：输入是一个单向的 follow 关系（比如 A follow 了 B、C，B follow 了 A、C、D），找出每个人的二度好友；

Yahoo

一共三面：

螺旋矩阵

不用乘除求 $a/b=c\dots d$

$A_1A_2A_3\dots A_nb_1b_2\dots b_n$ 转化成 $a_1b_1a_2b_2\dots a_nb_n$

C++ cast，模版，lambda 实现，c++和 java，进程线程，虚拟内存物理内存

Yahoo

1 面:

- 1) 在一个 $n*n$ 的矩阵中, 有一个起始点, 和一个终点, 而且包含障碍点, 寻找最短路径。
- 2) 一个二叉树, 每个节点除了 left, right 指针之外, 还有一个 other 指针, 复制整个树。

2 面:

- 1) K 个排序链表, 寻找所有链表都包含的最小值
- 2) 一个数组, 寻找 $\text{sum}(i, j)$ 绝对值最小的
- 3) Singleton

3 面:

- 1) 纯聊项目, 找共鸣
- 2) 写一个读写者问题

Yahoo

一面:

1. 链表的排序, 要求 ($n \log n$)
- 2: 根据 `random3()` 求 `random2`, `random5`, `random100`。(有提示)
- 3: 求全排列, 要求偶数位置的数比其前一个的奇数位置的数字大, 例如 1, 2, 3, 4 正确, 1, 3, 2, 4 正确。(只说思路)
- 4: 进程和线程的区别, 给你三个文件, 对文件的内容排序, 使用三个线程处理还是用三个进程处理好? 理由?

二面:

- 1: "abc 0124cdg789*" 类似这样的字符串, 把里面的数字加一, 变为 "abc 0125cdg790*", 没有负数的情况
- 2: 写出 `queue` 的接口和数据结构, 问了 java 里面的同步机制。
- 3: 字符串问题: "abcabcabc" 类似这样的字符串, 判断他是不是由一个子串不断重复形成的。"abcdabc" == false. 只说思路
- 4: 简单问了一下项目

三面:

- 1: 问项目, 主要看表达是否清晰。
- 2: 问 `hashMap` 的数据结构, 插入, 查找, 删除。hash 里面的 Table 扩容问题。如果数据太大, 内存中存不下放到磁盘上文件中的 hash 问题, 每个文件应该是多大? =====至少是一个内存页=====当时根本没想到, 面试官会给出答案。一致性 hash 问题。
- 3: tcp 和 upd 的区别。tcp 流量控制、接收的数据有序。
- 4: 说前面的面试反馈你操作系统基础知识不好, 他们问了你什么问题? 进程中占用的资源有哪些, 内存有哪几部分?
- 5: 512 bit 的 m 和 8 bit 的 n, 问 n 在 m 中出现的次数。

例如：（缩短了 m）m:1111 1110 n:1111. 出现的次数是 3.

给出的接口是 `int repeat (int []m, char n)`

6: 提问环节

总结：面试过程中好多问题没答好，但是面试官说有些东西不知道没关系，主要看思路，沟通能力和态度。

HULU onsite

1. 设计算法，ab 两个大文件中各有一亿大小 url，每个 url 占用 64 字节，内存只有 2G，怎么找到重复 url？
2. 设计一个类似于 qq 音乐中的猜你喜欢产品，可以从算法或者工程的角度回答。
3. 有一个很大的日志文件记录用户登陆情况，有三列记录，分别是用户 id，时间，登陆或登出。设计一个算法，计算某个时间在线用户的数量。

HULU

1. 求区间[A, B) 之间有多少个数字满足 digit 上没有 2, 5, 8
2. 一个完全二叉树，要插入一个新节点，要求比 $O(n)$ 更好

比如

```
      1
     2 3
    4 ?
      1
     2 3
   4 5 ?
```

这两种情况，问号为要插入的点

HULU

Round 1

1. 文件是每一行是两个值 a, b，用 mapreduce 算法实现

`select a, count(distinct(b)) from table group by a`

要求一阶段的 mapreduce

2. word ladder

3. 将 1G 文件考到 n 台机的方案
4. 一个数组，要求分成 k 份，使得每一部分和的最大值最小
5. 海量网页，求超链接互相指向对方的 pair 的方案
6. 项目，被鄙视项目太杂了（我又想吐槽实验室这么榨学生干活）

Round 2

1. 简历上写的这几种语言的特性 C++, Scala, Java, Python; 用了 Scala 的什么特性
2. 两台机的同一份文件，第一台机的这份文件某一 byte 被修改过，用最小的传输 bytes 确认修改的位置；分析各种方法的代价；有 k bytes 修改时
3. 一个单调上升的数组，返回所有 pair，使得商为给定的数；正负情况

Round 3

先鄙视下没论文，本科的论文是给师兄做做实验

1. 输入流是字符，在屏幕打印输入的这些单词，要求每行长度不超过 M
2. 求数组出现次数超过 $n/2$ 的数； n/k 的情况；讨论下不同数据结构实现（用数组、map, hash）的 tradeoff
3. 抛硬币，求第一次出现正正、正负的期望

HULU

一面：
项目

1. 有序数组两个数之和为 k，两个数之差为 k，两个数之积, 两个数之商
2. 平面上有 n 个矩形，可能相交，求所有矩形覆盖的面积，各种数据结构实现的 tradeoff

二面：

1. 字符串表示的高精度加法，需要判断输入合法性。
2. 大小为 n 的数组，包含 $1-n$ ，要求不使用计数 (in-place 也算空间。。) 的方法，确认是否所有的数都出现
3. 判断一个树是否为完全二叉树。 完全二叉树各种问。。

三面：

项目

1. 两人取数必胜策略，讨论各种先取，后取。。
2. 实现一个 random 函数，返回数组中的元素，要求 x_i 出现的概率与 x_i 的大小成正比， x_i 为整数？double 类型？平均复杂度证明。。。

各种虐。。

HULU 电面

50min

- 1、简单聊项目；
- 2、给一个 double 数组: a_0, a_1, \dots, a_{n-1} ，你可以在这些数字之间添加加号、乘号、括号，问能得到最大的值是多少；
- 3、找到数组中出现超过 $1/2$ 、 $1/3$ 、 $1/k$ 的元素；
- 4、中序遍历+后序遍历还原二叉树
- 5、数组 int, 把所有负数移到 0 前，正数在 0 后，保持相对位置，常数空间，时间尽量少。。。我已经用到归并的思路 $O(N \log N)$ ，还是问有没有更快 (囧 rz...)

(两题算法题都没写代码，只说了思路)

Facebook

今天面试没看到脸。。。

先问了一个问题，问我为啥想去 facebook，事先没有准备，就 blabla 想到啥说啥。。。

然后做了一道算法题：dotProduct

对于数组 $A = [a_1, a_2, \dots, a_n]$, $B = [b_1, b_2, \dots, b_n]$, dotProduct 就是求：
 $A \cdot B = a_1 * b_1 + a_2 * b_2 + \dots + a_n * b_n$ 。

面试官给出了程序模板：

```
double dotProduct(int id1, int id2);
```

```
vector getVector(int id);
```

说明：dotProduct 就是给定 vector1 和 vector2 的 id，然后求 vector1.vector2

getVector 是根据 id 获取 vector，不需要实现，只要调用就可以了。

经过：

1、先用最笨的方法， $O(N)$ 解决。

2、面试官说假设两个 vector 中，0 的个数远大于非零个数，假设非零个数为 $M1$ 和 $M2$ ，能不能给出一个 $O(M)$ 级别的算法。

然后我就抱怨说，要忽略零也总是要遍历一遍数组的啊。然后他说不用考虑 getVector 的实现，而且假设 getVector 能够返回去掉零之后的 vector。

然后我就说那返回 vector<Pair>吧，pair 是<value, index>二元组，value 代表原来的值，index 代表值在原数组的位置，然后改写了程序（出了一个 bug。。。被鄙视了）。

3、这时 45 分钟已过，面试官说，我再问一个问题，你不用写了，说下思路就好。他说 2 中的方法是 $O(\max\{M1, M2\})$ 的，现在假设 $N \gg M1 \gg M2$ ，能不能设计一个小于 $O(M1)$ 的算法。想了一会没思路，最后面试官提示 $O(M2 * \lg M1) < O(M1)$ 。于是二分查找。。。搞定。

然后问我还有什么问题么。。。支支吾吾。。。I have no questions。。。

Facebook

面试官是个 Native Speaker

面试真不应该选在顶楼的，下雨，早上很闷，真是失算

面了 45 分钟，前面 10 分钟，聊了面试官自己做什么的，他是设计一些 API 给别人用，然后问了我项目，让我说一点能体现我水平，我的编程层次的项目（压力好大），我举了个有意思的实验室的项目，得到了他的肯定。考虑到他是整 API 的，我也给他留了个话匣子，说面完后可以讨论 REST, Thrift.

接着开始上题了：

给一系列点，然后判断能不能构成多边形，面试官 刷刷的写了个类出来（真想哭，别人都是把题目先在 colledit 上粘出来了，为毛我这种待遇哎），

看到这个题我第一反应是要求求凸包，于是我问面试官给我一些样例 来 clarify 一下这个问题，坑爹面试官拒绝给，而让我自己想，我先列举了两个点情况，然后 3 个点，共线情况，然后说 给的点至少要能构成一个三角形，然后他说为啥（囧 我纳闷，这不是很显然么），我又写了一个直线的一般情况来说明，还有重点的情况来说明，他说这确实不能构成多边形，难道还有啥情况我没考虑的么，我想了下没做声，面试官列了一个中间包了一个点的图形，问我能不能构成，我说可以啊，除了中间的点，能构成个凸包，面试官说不是要求一个凸包，囧，我郁闷了，那是要求啥，我停顿了一分钟，他又写了个样例，我一看是个四边形，说可以，他说不可以，我又看了一下，我说按这个连接方法可以，然后他说按照顺序连，他不认为用户会想要知道怎么练的细节。我想哭，你咋不早说啊，这时都讨论了 15 分钟了！！！！我说要一个个点判断新加的线段是不是与前面相交，我急着编码，还没完全考虑清楚全部共线和存在重合点的情况，先把主体的先写出来吧，两层循环暴力之，码了一会儿，留个了判断相交的函数 `isIntersect()`，写完准备开始处理共线情况，这时面试官说话了，存在 bug，我慌了找了下，貌似循环边界条件不对（大概思索了两分钟，现在想起来当时反应怎么那么慢），我改了下，解释了下，然后面试官说，这是个 bug，还有，我继续查循环，怎么看都没有了啊，面试官提示我 X 行，我又看了会（真是弱到极致），边界用例少=号。时间剩下 3 分钟，面试官叫停，开始聊天，我本来准备问很多的，就问了两个问题，扯了下他关于 Thrift, Rest 经验。FB 就这样跪了，没想到 FB 会出计算几何题，没想到面试官会是这种面试形式，总之跪得非常不爽，努力刷题，复习基础哎，实力还是太弱了

CH 8 月 12 日 google 第一轮电面

1 Given an english character string. It may contain both up/lower case characters or just up or lower chars. We want to transform the string into the following format: all the up chars are on the left, and all lower chars are on the right. Only one operation is supported: to replace the lower case with its up case in place or vice versa. The Q is: the minimal operations to transform the string into the expected format?

Example:

AbCDEFG -> ABCDEFG, the # of operations is: 1

AaBccccD -> AABccccd, the # is 2

2 There is one array of nuts (螺母) and one array of bolts, each with size 'n'. Each nut fits exactly one bolt and vice-versa. You do not have rule at hand so you can not know the exact size of nuts and bolts. When you compare a nut with a bolt, you can get one of the three results: tight, loose, fit. Try to design an algorithm to find the unique mapping between nuts and bolts?

（说思路，不要求写代码）

#3 Implement an algorithm to print all the valid combinations of n-pairs of parentheses. Valid here means properly opened and closed. Example:

input: 3 pairs of parentheses.

output: ()()(), ()(()), (()()), (((()))

Google 电面

1. 设计算法，ab 两个大文件中各有一亿大小 url，每个 url 占用 64 字节，内存只有 2G，怎么找到重复 url？
2. 设计一个类似于 qq 音乐中的猜你喜欢产品，可以从算法或者工程的角度回答。
3. 有一个很大的日志文件记录用户登陆情况，有三列记录，分别是用户 id，时间，登陆或登出。设计一个算法，计算某个时间在线用户的数量。