1. 我们知道，从一个数组里找一段(连续的)子数组求最大和，是一道经典的面试题，  
   方法很简单，只要O(n)的时间。把这个问题变一下，假设是一个循环数组呢？找一个  
   size<=n的子数组with最大和。  
     
   分析，很容易想到第一步，找个地方把循环数组切断，回到了原来的问题，然后在考虑  
   一下额外的情况。额外的情况就是：有可能最大和的子数组是跨越了切断点的？这种情  
   况的最大和怎么求呢？一个naive的方法能做到O(n)，但是需要O(n)的空间。巧妙的解  
   法就是，注意到所有数的和是固定的，考虑切断后的非循环数组，找一段从首开始+一  
   段从尾开始的两个子数组with最大和，等价于找一段子数组with min sum.  
     
   总结，要擅长利用等价性转换问题，从而将新的问题转变为一个已知有好solution的旧  
   问题。利用已知的经典问题来解决新问题，可以说是面试题目中相当重要的一个技巧  
     
   2. largest rectangular problem：问题是这样的，一个N×M的棋盘，上面的数字要么  
   是1，要么是0，那么要：a）最大的一个正方形全是1填充，b）最大的全是1的矩形。  
     
   a）是用动态规划做，虽然方法也很好，但是这里就不提了。b）问题感觉上要比a难很  
   多，为什么呢，因为rectangular比square有更大的自由度。不好用DP来做，分冶也不  
   合适。  
     
   这题的奥妙就在于，利用经典问题。什么经典问题呢？其实是另外一道面试题，其本身  
   也是有一定难度的题，题目是：给你一个统计直方图，假设每根柱子都是单位宽度，从  
   图的最左边一个紧挨一个排到图的最右边，求在这个图里找到一个最大矩形，它不跟任  
   何直方柱相交(边缘接触是允许的)。为什么提起这个题呢，故事是这样的，我之前没有  
   做出O(N\*M)解法的largest rect题，后来有一天遇到了这个直方图的题目，找到了很漂  
   亮的O(N)解法，猛然回顾起那道largest rect的题，这次就很轻松的搞定了。  
     
   3（鸣谢mittbbs jobhunting版上的一位面试官贡献自己出的题）有n个房间，小偷每天  
   偷一间，偷的规律简单说就是随机行走，如果今天偷了第i间屋子，明天有一半的几率  
   偷i-1，一半的几率偷i+1，注意如果刚好偷到了边界上，那么第二天只有唯一的选择。  
   如果你是警察，你只能每天选择一个房间蹲守，并且贼的手段相当高明，偷了一个房间  
   后，没有任何人能发觉该房间是否曾经被偷过。  
     
   提示：奇偶性。总结：注意观察题目中隐含的性质。  
     
   4. wild card匹配+搜索：假设你有一个dictionary(原题中是URL集合)，你要搜到到所  
   有与 \*a\*bc\*d 这样的输入所匹配的words。这里，\*是通配符，可以当成是任意个任意  
   字符(包括空)，怎么 预处理+搜索？如果输入是 ???a???b??cde 这类呢？ ‘?’代表  
   单个任意字符。如果输入是? \*的混合呢？

有道题目最近频繁出现，特地总结一下常规解法以及它的变体。有个经典变体我还没看  
到一个很经典的答案，希望有人补上，呵呵。  
  
1. The largest rectangle under a histogram  
  
http://www.informatik.uni-ulm.de/acm/Locals/2003/html/judge.html  
  
Given: An integer array represents a histogram  
  
Goal: Find the largest rectangle under the histogram.  
  
Key observation: 输入为一个整数数组。如果某元素比它两边的邻居都小（比如Ai)，  
那么高度大于Ai的矩阵要么在该元素左边，要么在该元素右边，不可能穿过Ai。利用这  
个性质，想办法遍历所有的矩形。  
  
Complexity O(N) where N is the size of the given array.  
  
  
2. Maximum subarray with all 1’s. (Generalization of problem 1)  
  
http://www.drdobbs.com/184410529  
  
Given A two-dimensional array b (M rows, N columns) of Boolean values ("0" a  
nd "1")  
  
Goal: Find the largest (most elements) rectangular subarray containing all o  
nes.   
  
Key observation: 从一边（假设右边）开始逐列扫描，构造直方图。每次构造出直方图  
来，用上面的解法求最大矩阵。每次构造直方图只需要O（M），解需要O（M），做N次。  
  
  
Complexity O(MN).  
  
  
3. Given a binary matrix, find out the maximum size square sub-matrix with a  
ll 1s. (Specialization of problem 2)  
  
http://geeksforgeeks.org/?p=6257  
  
Key observation:   
  
假设输入二位矩阵为M,构造辅助矩阵S  
  
If M[i][j] is 1 then  
  
        S[i][j] = min(S[i][j-1], S[i-1][j], S[i-1][j-1]) + 1  
  
Else /\*If M[i][j] is 0\*/  
  
        S[i][j] = 0  
  
Complexity: O(MN)  
  
4. Imagine you have a square matrix, where each cell is filled with either b  
lack or white. Design an algorithm to find the maximum subsquare such that a  
ll four borders are filled with black pixels. (variation of problem 3)

1. write a iterative function to calculate fibonacci sequence  
2. what is the difference of Delete table and truncat table.(about sql &   
database)  
3. what is ACID.(about sql or database)  
4. write a find() for BST, How to decide the element could not be found.((  
less than && No left child) || (large than && no right child))  
5. how to reverse a link list using one point  
(should I use recursive calling)  
6. find the sizeof(int) without using sizeof(). （大家可以讨论一下，我用(~  
0x0)右移位做的，但是不对方希望的答案）  
7. main()  
{  
    folk();  
    folk();  
    folk();  
    printf("abc\n");  
}  
How many abc will be printed?  
8. what is the difference between.  
kill proces\_id  
kill -9 process\_id  
  
另外一家类似职位的电面，全部是stl的问题，很简单。  
1. new a vector, what is the size(),  
2. how to prevent relocation when push elements into a vector. (reserve())  
3. why vector use relocation instead of chunks of memory.（keep compatible   
with C array）  
4. how is deque implemented?(uses chunks of memory)  
5. the different of set & map? what is the return value of find()? (an   
iterator of element )  
6. what is the member function needed to use set.find() or map.find().(less(  
))  
7. what if the element class doesn't have less() operator? (pass a functor   
to constructor as comparing function)  
8. for loop to find a element in a vector, what is the method the class need  
to support? (equal operator)  
9. For smart pointer, How it implement the reference counter(a pointer to a   
int or unsigned int)

1. count the number of 1s in binary representation 我用的4位4位hash的方法  
   但是没有考虑负数情况，在面试官提示下，意识到对负数向右shift,发生的是sign   
   extension, 所以这题要先对输入进行检查。  
   2. remove duplicates from linked list, suppose you can use stl list. 我用的  
   hash\_set, 这题很顺利。  
   3. 很长很长的string/file, 要我找出中间重复次数最多的character, character set  
   可能是任何char set, unicode... 我说mapreduce, 后来面试官说 如果是一台机器，8  
   个core, 1个process,怎么办。我说 似乎mapreduce需要在distributed system环境下(  
   但其实我不肯定，如果一台机器多进程/多线程)是不是就也可以mapreduce）.面试官又  
   给了提示，我才说multi-process,后改成multi-threading, 后来他问我那几个thread  
   怎么写协调，我说可以公用data segment in the process space, 但是要注意锁的问  
   题。最后他问我说mapreduce和multi-thread的tradeoff， 我说communication   
   overhead, 似乎他对我的答案满意。  
     
   我和面试官说honestly第一题 我见过了 （我不知道这是不是画蛇添足了！！！！！）  
   哎 总之 虽然觉得自己答得还可以 但是很担心！！  
     
   第二个人先问了一些暑期实习的项目问题  
   1. 设计一个counte class, 写一个函数算某机器最后一分钟处理任务的个数。我说用  
   circular array, 他说right way to go, 但是在细节上澄清了一会。不知道是不是我  
   的表达不够好。  
   2. vector of vector of ints, 请implement bool hasnext()和int next() 我说要用  
   一个变量track在nested vector中的位置，他提示用2个，然后我开始code, hasnext   
   code完了 然后也解释给他，似乎没有问题。next() 由于时间不够，但我把逻辑解释完  
   了，把关键的那4，5行code出来。

因为赶着年底毕业，九月底才开始投简历。这个offer来的太快，小startup就是动作快  
，从十月初联系我，到 offer, 就两周。那几个大公司的 on-campus interview还都没  
开始。也算是 hot startup,但这里肯定没人知道的，移民版知道这个公司的更多些。  
就不透露公司的名字和考题了，见谅。  
HR 联系之后，先是组里的头直接电面，问了一个他们实践中的问题，我没想出答案，  
但还是扯了扯。后来就在谈公司做什么，我有准备，问了很多问题。刚放下电话,HR问  
我什么时候作 coding test, 可以马上把题发给我, 就是fill Java class, 实现某些  
功能，一般给 2-3个小时。我想这么大工作量，不能拖，否则牵扯时间和精力，就说马  
上作，决定不准备了，冒一定风险。结果一个小时做完发给他们，小impress了一下。  
然后另一个 team lead 马上打电话二面，顺便考察一下是不是真是我自己写的程序。  
这就是一天三面，两电一编程，然后就给 onsite了。面了组里的4个lead,HR 和   
Founder。前两个基本都在问我的 research和 big picture, 没有考我，第三个问了两  
个经典算法，属于偏简单的，版上都有，出现过很多次的。有一个我当时不知道，提示  
着才做出来。感觉面试的时候脑子不是很转的开，没有准备的话，很多题都不可能马上  
想出来。反过来，有准备的，就当不知到结果，从头把思路说一遍。吃午饭的时候，拍  
了拍马屁，开了开玩笑，到下午果然放水。第一个人就问我 PhD 为什么不考虑   
academic job,这个之前 HR也问过，基本是 PhD必问的 behavioral question, 一定要  
准备。我就说整天synthetic data, abstract problem 不如实际的问题 motivating.  
然后就是我在问他做什么东西，追的很 detail, 就没他问我的机会了。第五个还是那  
个联系我的 HR, 人很精干，办事效率高，问了很多behavior的。 你最想作又没机会做  
的，我就说 leadership。 还有什么你总做不好，例子是 time management, work   
life balance, 我说了一些 research的艰辛。还有你跟人关系处的不好的例子，怎么  
解决的等等。最后是大老板，Founder, 向我夸赞他们的产品，策略和公司影响，还有  
公司的光明前景，我们组的重要性，还有股票来者有分，听得我直流口水。Onsite 后  
第二天晚上就给 口头 offer了。要了 reference，background check, 第四天正式   
offer,但只给我一周答复实践。觉得工资低了点，就大胆照着Google的 negotiate了一  
下，涨了一些，达到了 Google PhD的下限，觉得知足了。我这样的，去面 Google, 估  
计不到一成的胜率。  
再说其他面试经历，年初有一个猎头从 linkedIn上找到我要给我介绍什么 DE Shaw,   
Two Sigma. 我说年底要毕业，thesis 还八字没一撇，到时候再联系吧。到了九月底，  
联系那猎头，很热情的要了我的简历，还帮我改了改，说很好，我马上给你投 DE Shaw  
, Two Sigma. 没两天就受到裸拒，深受打击。后来一看，也不冤。这是两个名hedge   
fund 要不牛校(MIT, Stanford, Berkley, Ivy 级别的)，要不拿过竞赛奖的，否则基  
本不会考虑你。恍然大悟，原来这个该死的猎头本着中六合彩的心态拿我撞运气。  
从此发现了可以走quant dev/strategist这条路。从九月底，先投了两个 wall street  
的。 第一个是作 CDO相关的 quant strategist, 过了coding test, 和两轮 phone   
interview,然后问我什么时候来纽约，好安排 onsite. 我说你得出钱呀，被告知他们  
一般只招local,早说呀。费这么半天劲，就当练手了。另一个是 hedge fund 的 data   
specialist, 通过另一个猎头找的。面试我的是中国人 team lead。由于被 DE Shaw,   
Two Sigma拒的神经错乱，整天想着牛校，在结束之前竟然问人家是什么学校出身，被  
不愉快的谢绝回答。以为就完蛋了，猎头打来电话说不错，继续。 二面又是一个中国  
同胞，感觉比一面还好些，最后又问人家是什么学校出身，不长记性，又被不愉快的谢  
绝回答。至此就没消息了。肯定是觉得我人品有大问题。如果二位在版上混，请不要见  
怪。  
这两个公司，除了经典算法题如longest increasing subsequence, linear time   
selection, verify if S1 is substring of S2, 还都问了很多 C++， Java 概念题，  
如虚函数，如何用 destructor 实现 exception safe code,顺便提及为什么Java 不需  
要 destructor？因为 Java 有 garbage collection, 但是 Java如何 safely release  
external resource in case of exception? 答案是用 try/finally.用finalize 也  
可以，但不是同步的（asynchronous）。 还问了 smart pointer, reference   
counting, Java garbage collection, 这些都是相关的。也投了一些其他的金融类，  
都默拒了。看来还是找猎头好，至少不会默拒。  
上周和这周on campus 面了 Microsoft, Epic, Amazon, Facebook  
Microsoft: 从 numeric array 里找两个数之和最大。刚开始还以为是经典的找和为固  
定值的两个数，其实这个更简单。找最大的两个数就行。Simple O(n) algorithm. 都  
没有让优化到 N+lgN-2 个 comparison. Instead, 让考虑意外情况，比如数字太大，  
加起来overflow怎么办.  
EPIC: 竟然没有technical question ！（干什么来的？）, 跟我套了套近乎，说他N年  
以前选过一门database的课，跟我写在简历上TA的是同一门课，同一个professor.我说  
现在教的很不一样了，顺便吹了吹。继续说到他们在用一个很老很奇怪的database叫什  
么 CHRONICLE 之类的，70年代他们 CEO自己写的，现在还在用。还属于post-  
relational database,我后来在Wiki上也没找到这个什么 Chronicle.又说他们还在用  
VB,小惊讶了一下。后来在Glassdoor上看了一下EPIC的员工review，很多人在抱怨  
working on legacy code, old technology … 尽管EPIC在 healthcare 里占有很大份  
额，但得不到什么技术上的提高，作码工还是那种最累的，从 H1B visa 纪录看，pay  
的也不是很好的样子。劝大家能不去就不去。一家之言。  
Amazon: 这回是经典的从 array 里找两个数，和为一个特定输入的值。一种是用hash,  
expected O(N) time, 一种是sort之后从两头往中间scan, O(NlgN)。第二题是verify   
that a binary tree is a binary search tree. 就是 recursive in-order   
traversal, 然后看是不是都是单调递增。如果不用多余空间，只要再用到一个全局变  
量记录上一个访问节点的数值就行了。当时脑子短路，提示了一会儿才想出来。 这两  
个都要再纸上写code,他们存档讨论  
Facebook: 把一个binary tree变成double linked list。 也是写一个recursive in-  
order traversal, update pointer 的时候稍微有点tricky. 第二题是经典的计算 x/y  
不用 division operator. 我说用 log 和 exp, 他说可以。又问这些也不能用怎么办  
。 提示了一下 x, y 都是正整数的话，从 x减去y 一直减到0,看要减多少次，很   
obvious 是 bisection 了。面完之后说我肯定第二轮 on campus了,如果过了就是   
onsite。后来recruiter打来电话说根据我问的问题，他们决定让某个比较match的组约  
电话 interview时间,马上会联系我， 而不是继续下一轮on campus。等了一天还没消  
息，不知到是不是传说中婉拒的最高境界。  
这些大公司都只过了一轮，题也比较简单，见笑了。手头 offer催的紧，觉得不算太差  
，base pay 的和google也差不了太多，potential bonus 比例还高些，但要看公司业  
绩，所以很可能是望梅止渴。Stock给的也没google值钱, 不过YY起来还够用。还要赶  
着写论文，就决定从了。  
另外说一下Facebook现在给的stock确实很诱人，我委婉问interviewer他有多少股 (2   
年员工)，是不是很有钱了。 感觉他有上10万股，但很frustrated的样子说，不IPO就  
跟废纸一样。也许是怕我嫉妒他。听说二手市场买卖facebook的share有一定的成交量,  
但不知到是不是有限制，所以伦不上小喽啰门。  
最遗憾的是没有试试Google。找好了递简历的同学，但是一直觉得没准备好。想放到最  
后。估计现在去了也白给。手头只有一个 offer 就从了，但从各方面分析，也只能这  
样了。现在只剩一个月写thesis了，八字才刚有一撇:(

Yahoo 电面 印度人  
1.电话键盘上1-》abc 2->cde... 现在来一堆数，未知长度，比如123456..... 请输出  
序列可能对应的所有字符串  
比如 123 输出acf, acg, ach, bcf...   
2.检测链表是否有环  
3.sql语句  
employee(id(primary key),name)  
employee\_bonus(id(primary key), bonus) （现在觉得这题 似乎有点问题，因为他和  
我说 id可以对应多个bonus, 那这还算是primary key吗。。。）  
请写sql 输出name和这个人bonus总和。  
  
MS on campus interview - first round  
1. 简历问题  
2. 给一个字符串检测是否是valid ip address, 这题他似乎是想看看我的思路，我说  
regular express, 他说要code, 我就写了一些，解释了一下。总之，这题真要追究起  
来，细节颇多，但因为我们每个人都只有30分钟，所以他没有要求完美的答案。  
  
我知道的其他ms校园面试题  
1. bst删除一个节点  
2. 矩阵，每行从左到右递增，每列从上到下递增，找一个数。  
  
Amazon所有轮的题目  
1. 数一个字符串中的单词数。  
2. 经典的 一堆数只有一个出现2次 其他数出现一次 怎么找出那个重复的数。请给出4  
种以上解法。。。包括brute force...  
3. longest palindrome.  
4. design restaurant reservation system.  
5. atoi, 只能用一个loop. 要考虑-213， 2-13， a2-13....各种情况。。 这题看似  
简单，但是要小心，容易出错。  
6. design a phonebook, 要求输入人名的时候电话随之更新。比如a给出一个电话，  
abc给一个，abcd再更新给出一个电话。trie的implementation, 假设每个节点存一个  
数据和26个指针。  
7. c++中static和virtual keyword, 请解释用法。  
amazon给我的感觉就是算法不一定很难，但喜欢考oo design.

1. 一个rotated的排序整数数组，比如A=[6，8，1，2，4，5]，写code找一个给定元素  
   ，并分析复杂度。其实就是binary search的变体，但是需要考虑两种A[m]中值的情况  
   加以判断。  
   2. 他谈到facebook的log，如果每个log文件有10 billion行，每行包括timestamp,   
   user\_id, visited page三个field。如果高效的统计一个月内用户访问量最多的十个网  
   页。假设文件已经按照timestamp排好序了。不用写code，谈想法就成。我的解法是，  
   如果内存足够，建立一个<网页,访问数>的hash表，每读入一条记录，该网页key对应的  
   项加一。但是加入内存不够大，可能需要map reduce方法。面试官马上说，内存根本不  
   是问题，用个4G或者8G内存来算这个月访问量很值得，不要省内存。然后我说，那hash  
   表就好了。面试官表示赞许，接着提出引申问题：  
   3.假设我们已经算了一个月的，如果这个滑动窗口要往前移动一个delta量，就是说，  
   已经算好timestamp=[t\_1..t\_x]区间的访问量，要计算timestamp=[t\_2..t\_x+1]的访问  
   量，提出高效算法。当然，最brute force的方法就是针对[t\_2..t\_x+1]再建一个hash  
   表，显然这不是面试官想要的。我提出来，可以建个multi-level hash表，外层为网页  
   ，内层为timestamp，但是仔细想想也不算好。面试官也没说他们的解法是啥。

1 external sort  
2 一道正态随机的题目。我到现在还不太明白。。。。  
3 print BST in level order  
4 实现linkend里查找两个人之间connection的功能。（如果每人有100个熟人，假设任  
何两个人之间只隔6个人，需要space 100^6，内存放不下。所以改用同时从两边bfs,   
需要space 2\*100^3）  
5 合并两个直方图，新图是原来两图的和。（直方图用点的array储存，比如一个直方  
图有两个矩形:x 2-3, y 4; x 3-5, y 3.表示成{(2,4), (3,3), (5,0)}.计算的新图点  
对就类似成combine two sorted array了）  
  
apple网投，一个月后phone。  
1 c++的多态有哪些实现方法。（继承，template specification，好像还有一个，忘  
了。。。。）  
2 为啥用mutex(to avoid unstable states)  
  
ebay网上投简历，1轮phone interview+onsite. onsite从9点面到6点，一个小时一个  
人，累死了。。。。签了保密协议，具体题目就不说了（能记住的只有下面这些了。。  
。）。  
1 design patten. factory method, factory, visitor, bridge, class design.  
2 反转link list  
3 external sort  
4 实现查找两点间最短路径  
5 1堆数，找出所有和为定值的三个数

#include <stdio.h>

void permutation(char \* p\_str, char \* p\_begin)  
{  
    if(!p\_str || !p\_begin)  
    {  
        return;  
    }

    /\*  
     \* If p\_begin points to the end of string,  
     \* this round of permutation is finished,  
     \* print the permuted string.  
     \*/  
    if('\0' == \*p\_begin)  
    {  
        printf("%s\n", p\_str);  
    }  
    /\* Otherwise, permute string. \*/  
    else  
    {  
        char \* p\_ch;

        for(p\_ch = p\_begin; \*p\_ch != '\0'; ++p\_ch)  
        {  
            char temp;

            /\* Swap p\_ch and p\_begin. \*/  
            temp = \*p\_ch;  
            \*p\_ch = \*p\_begin;  
            \*p\_begin = temp;

            permutation(p\_str, p\_begin + 1);

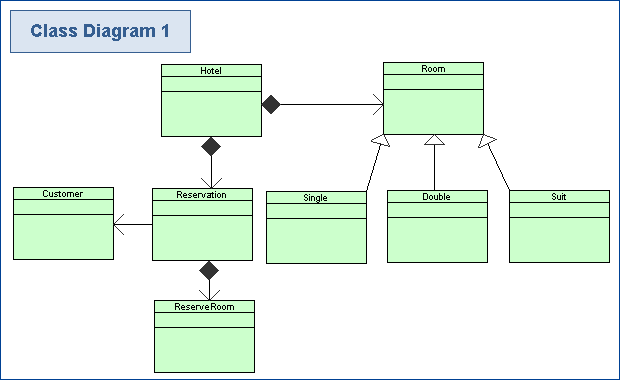
            /\* Restore p\_ch and p\_begin. \*/  
            temp = \*p\_ch;  
            \*p\_ch = \*p\_begin;  
            \*p\_begin = temp;  
        }  
    }  
}

int main(int argc, char \* argv[])  
{  
      
      
    char strr[4]="123";  
    char strd[4];  
permutation(strr, strr);  
    return 0;  
}

UTF8解码器  
线段overlap  
死锁机制  
fair streaming sample  
最不好的是,忘记可以用bitmap记录sparse data,选择了hashtable.

最近开始准备OO design questions, 这两天学习了一下design pattern, 用Amazon了  
经典的Hotel Reservation OO   
Design的问题练习了一下。希望能和版上各位高手交流一下。  
  
(1) 问题1: 这个system应该就那些类  
      在version 1设计，我定义了如下类Room, SingleRoom, DoubleRoom, SuitRoom,  
Hotel, Customer,  Reservation, Date。 因为感觉这个design,主要就是如何设计check different types  
of rooms' availability和make reservation。  
  
(2) 问题2: 这些类的关系应该如何  
      其实就是"has-a" and "is-a" 的设计，也就是用inheritance或是composition  
。根据直觉定了了如下的UML图  
  
以上两步估计大部分人都能得道，不过如何才能打动interviewer呢？怎么才能提高这  
个设计呢？  
我想可以从两各方面入手：（1）program with interface (不过这个好像是java   
programmer的rule) （2）多用design pattern, 因为design pattern能decouple code,这样就能提高code的  
resuability. 在什么情况下，如何使用design pattern就是门学问了，需要经验阿！  
  
不过对于这个问题，我想了些改进方法。  
(3) 问题3：如何create object   
用factory method pattern来建造不同的Room object, 虽然目前的design只有三个  
Room-based classes,不过随着以后系统复杂度的增加，可能回有n个Room-based classes。我猜想design的时候考  
虑日后扩展的flexibility，应该也是interviewer喜欢看到的把。  
  
(4) 问题4： 如何hold多个Room-based objects  
是用一个list来hold所有的SingleRoom, DoubleRoom, SuitRoom的object,还是用三个  
list来hold各自的object? 这个没有想好，目前就用了一个list, 因为估计总room的数目不会太大，对单个list进行  
iterate的效率不会太差。  
  
(5) 问题5: 如何有效的check availability.  
假设有一个customer要查询每种room从某天到某天的availability, 那么需要对所有的  
Room subclasses都定义一个查询函数checkAvailabity()。这样的话不是很好的design。因为日后要加其他的查询：  
如Room的朝向（如，有没有view之类的:-))，那又再改interface, 然后对所有的classes添加函数。麻烦！  
  
所以， 我考虑用strategy pattern可以日后extend各种查询。  
  
(6) 问题6： 找到availability, 如何设计reservation  
回想一下，make a reservation其实需要多个步骤的，需要填用户信息，需要填信用卡  
（如果是pre-pay的话），需要计算tax, 还可以用discount code的，最后得到total amount。这些步骤都是  
sequential的，有一定的次序。所以我们可以用builder pattern来generate所有这些步骤。  
  
  
写个code（看下面link），欢迎各位排砖。  
http://www.ideone.com/KsDCV  
  
  
  
  
  
  
--  
  
※ 修改:·langqinren 於 Nov 13 10:29:29 2010 修改本文·[FROM: 71.62.]  
※ 来源:·WWW 未名空间站 海外: mitbbs.com 中国: mitbbs.cn·[FROM: 198.82.]  
  
此主题相关图片如下：

HotelReservation.gif

(5967 字节) [删除]  


Phone1：  
Behavioral: Your biggest challenge, do you know our product?  
Tech: 经典的html里找email的题 (using regex)  
      找anagram  
      Deck shuffle algorithm  
      Two stacks for a queue  
  
Phone2：  
      N-way merge和时间复杂度 (n-way 和 2-way的比较)  
      手机输入提示功能 (trie)  
  
两个phone都不难 很快拿到onsite 同时面两个组 onsite发现 A组全是白人 B组全是阿  
三 结果被阿三给放倒了...  
Onsite1 - 老美 A组manager： 问了问profile, 给了一个oop design的问题 不是常见的  
电梯或家具题, 完全是他们所做项目的设计. 这题回答的一般, 最初给的答案不是他想  
要的。。。在提示后 完成设计   
  
Onsite2 - 阿三 可能是Bar raiser, 因为不是A组或B组的人  
       1. 超大文件 查找unique数字 (hash, bitmap, 外排)  
       2. Load balancer问题: 一个balancer, 两个servers, 如果其中一个server   
down了, 问如何设计balancing algorithm使得用户感觉不到延迟(e.g. 如果使用round   
robin的算法 需要从server A返回不available的信号, 再转向server B, 就会有延迟  
）要求只有两个servers available. (这题给了两个答案：read/write through cache,   
和recent response time 都被否决 不知道想要什么)  
       3. 一个balancer， 多个servers 问如何设计 回答说master-slaves +   
hashing (没有说consistent hashing 因为想弄明白他的问题). 结果果然没有问  
consistent hashing, 问如何优化slaves的storage，回答sharding。 这题不置可否，讨  
论了一会时间到  
  
Onsite3 - 老美 水过  
      1. linked list 找random k元素 (reservoir sampling)  
      2. 又是anagram  
      这个老美感觉是来应付交差的。。。 后面问了一堆behavioral question...  
  
Onsite4 - 阿三panel B组manager和B组成员  
      B组全部都是阿三。。。 问machine learning的问题 都是和Amazon实际问题相关 没  
有准备。。。回答的差  
  
Onsite5 -  A组 东欧人  
       1. C++多继承转Java code (using combination)  
       2. user activities log, 找出最长的session (一小时之内的activities被认  
为在一个session里). 分别问了单机 多机的情况, 然后coding

deep copy linked list: 倆pointer, next, random  
Linked list merge sort  
write iterator for a list of arraylist  
最自豪的事儿  
找到一张图片里面所有的word 的bounding box  
比较两个directory是不是完全一样  
OO design for chess game,design render loop  
number of 1 in binary representation for integer  
find kth to last item in linked list  
好多behavior 问题  
怎样实现traceroute  
tcp vs udp  
caching, indexing in db, load balancing, sticky session etc.  
怎样优化web service,找bottle neck

Round 1:  
  
1. C++ Copy Constructor, 包括接口和deep vs. shallow的区别，这题答得一般，有些  
细节没搞得特别透彻  
  
2. Hash Table的实现和技术要点 (仔细看过一遍wiki就没问题了）  
  
3. Large file, multiple lines, how to get any line in equal probablity. 这题  
可以问得很深入，比如文件太大内存无法装入如何办。  
  
我回答的思路：  
  
内存够文件内容就都装入内存，然后randomGenerator选一行,one pass  
  
内存不够可以记录每行的偏移值在内存，这样之后可以fseek到那一行去读取  
  
如果偏移值都放不下，可以divide into ranges, 当然这个range的stepsize不好选，可  
以预估，也可以动态改变（到这一层其实大致给些思路就ok了）  
  
  
Round 2:  
  
1. Research problem  
  
2. How to build a service to generate an unique number for each client reque  
st, 这个问题也可以问得很深入。  
  
我回答的思路：  
  
systemTime + IP --> IP不唯一 --> systemTime + ConnectionID   
  
multiple requests --> build a distributed service so we need add hardware si  
gniture to make sure the unqiueness   
  
Another solutions is building a hashTable to store all generated numbers for  
a check before sending number to client  
  
拿到Onsite ...  
  
  
Microsoft:   
  
Some behavior questions    
  
Singleton pattern concept, write a Singleton class (in multithreading enviro  
nment)  
  
拿到Onsite ... (特别感谢一位谈不上很熟的朋友给了很多帮助，所以这个onsite拿得  
最顺利)  
  
  
a startup:  
  
String2int, 这题不难，但要考虑很多细节，比如+/-号，溢出等  
  
Hash Table 实现与技术要点，与stl:map的区别和优劣  
  
finding missing integer (enough memory solution, not enough memory solution)  
, 基本上就是用bit-vector的思路，如果内存不够就1bit to a range  
  
拿到Onsite ...  
  
  
Google  
  
Round 1:  
  
1. 矩阵乘法 （紧张导致脑袋不清晰，虽然最后代码正确但中间出了挺多纰漏）  
  
2. 一些基础的数据结构，arrary, list, hash, tree等等  
  
虽然自我感觉糟糕还是被容许第二次电话面试  
  
Round 2 （一个语气挺tough的同胞）：  
  
1. 股票的买入卖出，max the profit, 如果有多个pair产生max profit如何处理（这题因为自己编程做过所以答得应该挺好）

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | [[回复](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Fmitbbs_postdoc.php%3Fboard%3DJobHunting%26reid%3D31742735%26gid%3D31742735&sa=D&sntz=1&usg=AFQjCNFG4ma0tURqiFtfUSQGy--s9ZR-Ng)] [[回信给作者](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Fmitbbs_mailbox.php%3Fboard%3DJobHunting%26file%3DM.1290634966_2.I0%26userid%3Dlinkaixin123%26title%3DRe%3A%2520Amazon%25B5%25E7%25C3%25E6%2B%2528%25C3%25E6%25BE%25AD%2529&sa=D&sntz=1&usg=AFQjCNFfKS8N2VxLu6ybWtmvALNuQTgspw)] [[本篇全文](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Farticle%2FJobHunting%2F31742735_0.html&sa=D&sntz=1&usg=AFQjCNHYsUhIDuM3eUcnIVQw5eUEwJLYNA)] [[本讨论区](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Fbbsdoc%2FJobHunting.html&sa=D&sntz=1&usg=AFQjCNHLQ8eH7AgliZznHoAI-cOc6_cZIA)] [[修改](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Fmitbbs_bbsedit.php%3Fbrdname%3DJobHunting%26title%3DAmazon%25B5%25E7%25C3%25E6%2B%2528%25C3%25E6%25BE%25AD%2529%26author%3Dlinkaixin123%26file%3DM.1290634966_2.I0%26id%3D31742735%26gid%3D31742735&sa=D&sntz=1&usg=AFQjCNEyruY9bJ3jYabRJSVIymXANaiLog)] [[删除](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Farticle_t%2FJobHunting%2F31742735.html&sa=D&sntz=1&usg=AFQjCNGJW0zn8ai2x0gYP9ihjVzsVyFq_A)] [[转寄](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Fmitbbs_forward.php%3Fboard%3DJobHunting%26id%3D31742735&sa=D&sntz=1&usg=AFQjCNGeKXbyJlevgmXqk-Ne7uUusAVP_g)] [[转贴](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Fmitbbs_bbsccc.php%3Fboard%3DJobHunting%26file%3DM.1290634966_2.I0%26title%3DAmazon%25B5%25E7%25C3%25E6%2B%2528%25C3%25E6%25BE%25AD%2529%26author%3Dlinkaixin123%26id%3D31742735&sa=D&sntz=1&usg=AFQjCNHXBypDAmCpfHoAepGpDaNJooWzTA)] [[收藏](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Fshare%2Ffav_input.php%3Ftype%3D1%26board_id%3D106%26article_id%3D31742735%26boardId%3D106%26area%3D4%26articleId%3D31742735%26club_flag%3D0&sa=D&sntz=1&usg=AFQjCNF72Rcr8Ka4PVjzUR8mmJdcSiAmVg)] [[举报](http://www.google.com/url?q=http%3A%2F%2Fwww.mitbbs.com%2Fmitbbs_report.php%3Ftype%3D1%26board_id%3D106%26article_id%3D31742735&sa=D&sntz=1&usg=AFQjCNFqQboB2QUGXLbZG1RbddRuR723xg)] |  | |  |  |  |  | | --- | --- | --- | --- | |  | 0 |  | 0 | | [ 1 ] | |  |  | | --- | | 发信人: linkaixin123 (林开心), 信区: JobHunting 标  题: Amazon电面 (面经) 发信站: BBS 未名空间站 (Wed Nov 24 16:42:46 2010, 美东)  报个面经，攒一下RP。  几个问题，都是经典题，另外考了不少Java的基础知识，还有一道设计题。  == Java == 1. interface vs abstract class 2. the usages of 'final' keyword 3. overload vs override 4. concept of 'equals' and how to override (need to override hashCode() also) 5. checked vs unchecked exception  == Misc == 1. 50,000 HTML pages, extract the ones with phone numbers (regex) 2. Find the odd-occurence number in an array (I told him that I have seen  the problem before and he asked the next one) 3. Find the duplicate number in an array of N ints (1<=x<=N-1) 4. Design a transportation system (warehouses, routes, trucks, fuel, MPG,  etc)   == Homework == Write a Card Game program and send it to him before evening. | |
|  |

1. the difference of new/delete malloc/free. what if new/free, and malloc/  
   delete?  
   2. sizeof an empty class  
   3. sizeof an empty class with virtual function.  
   4. can we call virtual function in constructor/destructor?  
   5. what is pure virtual function?  
   6. How to disable class initialized on heap,  
   7. How disable class initialized on stack,  
   8. where static varible saved, heap or stack?(bss, data)  
   9. what does static mean in C++  
   10. what is 4 default functions for empty class  
   11. when throw exceptions, why use reference to catch exception(这题  
   12. what is the syntax to catch all exception?  
   13. can constructor/destructor throw exception.  
   14. what will happen if 2 or more exceptions happened?  
   15. why we need exception in C++?(what is the benefit we can have by using exception)  
   16. What the design pattern do you use?  
   17. Tell me all STL containers you know.  
   18. Can we use auto\_ptr as element in  set or map?

== Data Structures & Complexity ==  
2. Lookup in linked list and array (sorted, unsorted)  
3. Sorting strategies (comparison-based & non-comparison-based)  
4. Lookup, insert, delete in hash table.  
5. How to resolve collision (chaining, open addressing)  
6. How to support delete with using open addressing  
7. What affects collision (hash table size & hash function)  
8. What the complexity when using dynamic array for hash table?   
(insert 1M elements takes how many steps when we always copy over => 2^n (  
geometric serie))  
9. What is BST, balanced BST? How to maintain balance with inputs like 1, 2,  
3, 4, 5 ... (red-black tree)  
  
== OS concepts ==  
1. What is process & thread? Difference between kernel thread & application   
thread.   
2. What is virtual memory? What is it good for?   
3. How to implement thread? (save/restore registers during context switch)  
4. Write a program to check if stack grows downward or upward. (A calls B   
and prints the addresses of two local variables)  
  
== Design ==  
1. Design an web-site for playing card games. (what classes, what are the   
responsibilities, users vs players)  
== OO ==  
1. What is inheritence and polymorphism?   
2. How to overload a function? Why return type is not enough?  
3. Is-a vs Has-a  
4. aggregation vs composition  
5. Strategy design pattern  
6. Issues with multiple inheritance (diamond problem)  
7. Class member vs instance member  
8. Implementation inheritance vs interface inheritance  
9. How to reuse code? (inheritence, template, library, external program, web  
-service, FPGA, etc)

1st:  
1. 讨论我的博士研究项目  
2. 如果SNMP agent不能获取数据，或者获取的数据不符合预期，如何诊断该问题？  
3。我做过的最有挑战的项目是什么？  
4。用邮件写代码,然后讨论我写的代码：  
unsigned char \* get(int sizeOfArray, int sizeOfRecord);  
void release(unsigned char\* ptr);  
  
该函数可以实现：  
unsigned char \*\* array = get(5, 10);  
snprintf( array[0], 10, “hello world\n”);  
snprintf( array[1], 10, “hello again\n”);  
  
5。Java的基本概念  
  
2nd  
1。Apache的log file如何找访问量最大的网页 （用linux shell写个小script）  
2。如果某网站访问量突然增加，可能是什么情况发生，如何确定各种情况（1。暂时的  
Popularity激增 2. DDOS Attack 3. 网站添加新的内容）  
3。Java基本概念+设计扑克牌的类  
4。读reverse string的代码（基于stack和对换位置）  
  
Onsite Interview  
1. 很高级别的一个manager，介绍group, 各种behavior questions, 无任何技术问题  
。我早上8:00开始interview的，估计manager还没想好题，或者不像一大早就为难我把  
：D  
  
2. Bar raiser; 如何实现phone Book（我的答案是trie,）, 并要求写出insert函数；  
外加一推java的基本概念  
3。给一个maze：for example:4 0 5 6 1 0， 每个number代表最多可以走的步数，问  
是否能从a[0]走到a[n-1]; 问如何判断Web Service 运行正常，怎样解释response   
time的variance, 谈谈botnet  
4. 网络题：MTU Discovery, Switch&Router, IP header, VLAN怎样实现的，路由表怎  
样实现的，bitmap，hashtable.  还写了一个很简单的程序  
5。HashMap如何实现的；   
Userid  PageID  
A          1  
A          2  
A          3  
B          2  
B          3  
C          1  
B          4  
A          4  
找出最常用的3个访问序列：  
对于用户A：1-2-3， 2-3-4  
用户B：2-3-4  
2-3-4 是最常见的  
写atoi的程序  
设计rent movie的类  
6。lunch with hiring manager. 我对该职位的理解，为什么感兴趣，如果加入team会  
如何做啊，还有QA部分； 饭后问了到网络架构题

第一轮是一个印度大哥，PhD的研究方向（machine learning)跟我很类似。问了：  
1. 如何向不懂machine learning的人解释LASSO和l\_1 regularization。  
2. 解释什么是PCA；为什么二维数据中PCA找的主方向和做线性回归的fitting不一样。  
3. distributed median，（unsorted）数字分布在几台机器上，设计分布式算法找到  
它们的median，要考虑网络通讯的overhead。（先local sorting，然后一个数一个数  
去看是不是median就行，结果我设计了一个巨复杂的recursive algorithm...还好那个  
印度大哥比较nice）  
4. BST中任意节点的下一个数字。这个大家都会吧，呵呵  
  
第二轮是中国人，后来查了linkedin发现是当年本科学校的传奇人物。  
1. 在整数中找3个数，使它们和为0  
2. 怎么设计一个并行的linear regression solver  
3. 什么是decision tree, 怎么去建一个DT (用Gini墒判断用哪个attribute），什么是Gini墒。  
4. maximum area under histogram （忘记了O(n)算法，简单说了思路）

设计一个web cache server，假设存储网页数量是10个billion，打算怎么设计  
1. Is everything in-memory possible?   
2. If one machine is not enough, how to partition? What's the partition key?   
3. What's the index strategy? Hash based or tree based? If use hash based   
index, how to handle the hash-collision? (I remember Database textbook has   
discussion on this)  
4. If memory is not enough, how to handle?   
5. What's the complexity to check whether a URL is in cache or not? (is   
bloom filter useful?)  
6. Do we need a meta table on which table/machine is where for what?  
7. May also need to consider Replica for reliability.  
  
Big table (in memory version) may be a good reference here.

1、1~1000的数中有一个重复的数，把这个数找出来。  
2、找出Binary Search Tree树的Median number。  
这两道题的要求都是提出尽量多的算法，并给出复杂度。  
3、第2题我说了一个算法：  
如果事先知道每个节点的子节点个数，假设节点总数n是奇数，m=(n+1)/2，根的左子树的节点数是k，右子树(n-k-1)。这样的话，就把找中数转换成找第m小的数：  
if (m == k+1) return root;  
else if (m <= k) find\_m(root->lchild, m);  
else find\_m(root->rchild, m-k-1);  
我说复杂度可以达到O(log n)（不包括预处理的时间。现在想一想不准确，请各位帮忙看看这种算法对不对，如果对的话复杂度是多少？  
  
然后他就让我比较几个时间复杂度：O(n)、O(n^2)、O(log n)、O(n log n)，问哪个是最快的。是不是任何时候O(log n)都比O(n)的算法好？他想要的答案大概是："Big O notation discards multiplicative constants on the running time, and ignores efficiency for low input sizes, it does not always reveal the fastest algorithm in practice or for practically-sized data sets." http://en.wikipedia.org/wiki/Big\_O\_notation  
4、写出创建我说的bst的过程的代码，即二叉树中的每个节点都包含其子节点的个数。  
我当时我一个地方写错了，一个参数没有加引用，但是他当时也没看出来，对我的答案还比较满意。

第一个，聊了很久ERP方面的东西，最后只问了一道技术题。  
  
An array of continuous integers, only one number is duplicate, find out   
this number. What about there is a number is missing, how to find out   
both the duplicate and missing numbers.  
  
第二个，问了heap跟stack的区别，以及什么情况下会OutOfMenoryException。然后就  
是技术  
题了。  
  
Reverse a string. e.g. he is a man => man a si eh   
Reverse a string without reversing the words. e.g. he is a man => man a   
is he  
Find the square root of a float number.  
  
午饭后第三个，同样聊了很久我现在做的ERP方面的东西，问了下5年的career path和  
遇到的  
most challenging technical problem in previous projects, 然后就是技术题了：  
1) Give three database tables, Invoice, InvoiceLineItem, and PartMaster.   
Find the design problems. 主要就是要变成一个normalized design. 要问清楚具体的  
business requirements，然后针对这个requirement进行修改.  
2）Write sql to find the most recent invoice date for all customers. 我记  
得是这个  
  
SELECT customer\_name, max(invoice\_date)   
FROM Customer, Invoice  
WHERE Customer.customer\_id=Invoice.customer\_id  
GROUP BY customer\_name.  
  
对了，Customer table是normalized后得出的。  
  
3) Given a list of methods as follows.  
  
Station radio.getStation();   
void setToNext();  
void setToPrevious();  
  
String station.getFrequency(); // Station的Unique id.  
int station.getStrength();  
  
Find the top 6 strongest signal stations： List   
findSixStrongestStations(Radio radio){...}  
  
第四个，A list (LinkedList) of n numbers, p1->p2->...->pn. Find k random   
numbers.  
1) n is known  
2) n is unknown  
  
Given a BST, find the successor of a given node. The node class is as   
follows.  
  
class Node{  
Node lc;  
Node rc;  
Node parent;  
}  
  
最后一个是director, 聊了一下他们的ERP的架构，使用的语言。说了下这两组的区别  
和主要进行  
的项目。最后一个技术题。  
  
Given a file's full-path, e.g. c:\foo\bar\dir12345  
given a positive number n, find an abbreviated path from the full lpath   
whose length <= n with the following conditions:  
1) Drive must be included, i.e. "c:" should be in the result.  
2) The suffix part has higher priority to be included since it contains   
more important information.  
3) Each part can not be partially included in the result.  
4) The skipped part represented by "...".

给一个array of integers, 需要多少个comparison能找到min? how about min and   
max? what's the best worst-  
case? (3n/2)   
  
怎么merge两个sorted的array，n 个呢？  
  
给一个tree，把同层的nodes都连起来，assume每个node都有个指针，你要让它指向下  
一个同层的node。  
  
count 1's in an integer.  
  
mergeSort, mergeSort without recursion, why nlogn?  
  
Find whether one string is a subset of another string (not need to be   
contiguous, but order should match).   
  
Print the nodes on the exterior of a binary tree in a anti-clockwise order,   
i.e., nodes on left edge, then leaf   
nodes, then nodes on right edge.  
  
Find two integers in an array that sum to a target integer.

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | **【谷歌（Google）】——谷歌笔试面试题集锦** | | 发布者 ：admin 时间 ：[2010-04-27] |  |  | | --- | | **一、笔试题**    1、假设在n进制下，下面的等式成立，n值是（） 567\*456=150216 a、 9 b、 10 c、 12 d、 18 2、文法G:S->uvSvu|w所识别的语言是：（） a、uvw\*vu b、（uvwvu）\* c、uv(uv)\*wvu(vu)\* d、(uv)\*w(vu)\* 3、如下程序段输出是：（） char str[][10]={"Hello","Google"}; char \*p=str[0]; count<<strlen(p 10); a、0 b、5 c、6 d、10 4、cnt=0 　　while(x!=1){ 　　　　cnt=cnt 1; 　　　　if(x&1==0) 　　　　　　x=x/2; 　　　　else 　　x=3\*x 1; } count<<cnt<<end1; 当n=11时，输出：（） a、12 b、13 c、14 d、15 5、下面哪项不是链表优于数组的特点？（） a、方便删除 b、方便插入 c、长度可变 d、存储空间小 6、如何减少换页错误？（） a、进程倾向于占用CPU   b、访问局部性（locality of reference）满足进程要求 ) c、进程倾向于占用I/O   d、使用基于最短剩余时间（shortest remaining time）的调度机制  e、减少页大小 7、80x86中，十进制数-3用16位二进制数表示为？（） 8、假定符号-、\*、$分别代表减法、乘法和指数运算，且 1)三个运算符优先级顺序是：-最高，\*其次，$最低； 2)运算符运算时为左结合。请计算3-2\*4$1\*2$3的值：（） a、4096 b、-61 c、64 d、-80 e、512 9、下列伪代码中，参数是引用传递，结果是？（） calc(double p, double q, double r) {q=q-1.0;r=r+p} main(){ double a = 2.5, b = 9.0; calc(b-a, a, a); print(a); } a、1.5 b、2.5 c、10.5 d、8 e、6.5 10、求输出结果：（） int foo(int x, int y){ if(x <=0 || y <= 0) return 1; return 3 \* foo(x - 1, y / 2); } printf("%d ", foo(3, 5)); a、81 b、27 c、9 d、3 e、1 11、下列哪个数据结构在优先队列中被最广泛使用？（） a、堆 b、数组 c、双向链表 d、图 e、向量 12、以下算法描述了一个在n国元素的双向链表中找到第k个元素的 方法（k >= 1且k <= n）： 如果k <= n - k，从链表开始往前进k-1个元素。 否则，从终点出发，往回走n - k个元素。 这个算法的时间代价是？（） a、θ(nlogn) b、θ(max{k, n - k}) c、θ(k + (n - k)) d、θ(max{k, k - n}) e、θ(min{k, n - k}) 13、有一个由10个顶点组成的图，每个顶点有6个度，那么这个图有几条边？（） a、60 b、30 c、20 d、80 e、90 14、正则表达式L = x\*(x|yx+)。下列哪个字符串不符号L（） a、x b、xyxyx c、xyx d、yxx e、yx8 15、为读取一块数据而准备磁盘驱动器的总时间包括（） a、等待时间 b、寻道时间 c、传输时间 d、等待时间加寻道时间 e、等待时间加寻道时间加传输时间 16、Fibonacci,求f(4）使用递归调用f(1)的次数（） f(n) = f(n-1)+f(n-2), f(0)=0, f(1)=1 a、5 b、4 c、3 d、4以上 17、有关哈希表正确的说法（不定项）（） a、哈希表的效率和哈希函数。。。。相关 b、哈希表的解决冲突方法慢，回影响哈希表效率 c、使用链表哈希可使内存紧凑 18、下列排序方法最差情况时间复杂度为O(n^2)的是：（） a、插入 b、归并 c、冒泡 d、快速  19、写一段程序判定一个有向图G中节点w是否从节点v可达。（假如G中存在一条从v至w的路径就说节点w是从v可达的）。以下算法是用C 写成的，在bool Reachable函数中，你可以写出自己的算法。 class Graph{ public: int NumberOfNodes();//返回节点的总数 bool HasEdge(int u,int v);//u,v是节点个数，从零开始依次递增，当有一条从u到v的边时，返回true }； bool Reachable(Graph&G, int v, int w){ //请写入你的算法 } 20、给定一棵所有边的长度均为整数的树，现要求延长其中某些边，使得从根到任意节点的路径长度相等。问满足要求的树的边长度之和最小是多少?请写出你的算法，并分析时间复杂度。 21、以下函数的结果？ int cal(int x) {  if(x==0) return 0; else return x+cal(x-1); } 22、以下程序的结果？  void foo(int\*a, int\* b)  {  \*a = \*a+\*b;  \*b = \*a-\*b;  \*a = \*a-\*b; }  void main() { int a=1, b=2, c=3; foo(&a,&b);  foo(&b,&c);  foo(&c,&a); printf("%d, %d, %d", a,b,c); } 23、T(n) = 25T(n/5)+n^2的时间复杂度？ 24、n个顶点，m条边的全连通图，至少去掉几条边才能构成一棵树？ 25、正则表达式(01|10|1001|0110)\*与下列哪个表达式一样？ 26、实现两个N\*N矩阵的乘法，矩阵由一维数组表示。 27、长度为n的整数数组，找出其中任意(n-1)个乘积最大的那一组，只能用乘法，不可以用除法。要求对算法的时间复杂度和空间复杂度作出分析，不要求写程序。 28、打印出一个二叉树的内容。 29、在一个字符串中找到第一个只出现一次的字符。如abaccdeff，输出b。 30、求一个二叉树的高度，如果只有root结点，高度为0。 31、将稀疏疏组中的非零元素提取出来，用链表表示。 32、两个n维数组，已排序，为升序。设计算法求2n的数中第n大的数。要求分析时间和空间复杂度。不用给出代码。 33、给定一个长度为N的整数数组（元素有正有负），求所有元素之和。最大的一个子数组。分析算法时空复杂度。不必写代码。 答：最大子序列 问题： 给定一整数序列A1， A2，... An （可能有负数），求A1~An的一个子序列Ai~Aj，使得Ai到Aj的和最大 例如： 整数序列-2, 11, -4, 13, -5, 2, -5, -3, 12, -9的最大子序列的和为20。 对于这个问题，最简单也是最容易想到的那就是穷举所有子序列的方法。利用三重循环，依次求出所有子序列的和然后取最大的那个。当然算法复杂度会达到O(n^3)。显然这种方法不是最优的，下面给出一个算法复杂度为O(n)的线性算法实现，算法的来源于Programming Pearls一书。 在给出线性算法之前，先来看一个对穷举算法进行优化的算法，它的算法复杂度为O(n^2)。其实这个算法只是对对穷举算法稍微做了一些修改：其实子序列的和我们并不需要每次都重新计算一遍。假设Sum(i, j)是A[i] ... A[j]的和，那么Sum(i, j+1) = Sum(i, j) + A[j+1]。利用这一个递推，我们就可以得到下面这个算法： int max\_sub(int a[],int size) { 　　int i,j,v,max=a[0]; 　　for(i=0;i<size;i++) 　　{ 　　　　v=0; 　　　　for(j=i;j<size;j++) 　　　　{ 　　　　　　v=v+a[j];//Sum(i, j+1) = Sum(i, j) + A[j+1] 　　　　　　　　if(v>max) 　　　　　　　　　max=v; 　　　　}  　　} 　　return max; }那怎样才能达到线性复杂度呢？这里运用动态规划的思想。先看一下源代码实现： int max\_sub2(int a[], int size) { 　　int i,max=0,temp\_sum=0; 　　for(i=0;i<size;i++) 　　{ 　　　　　　temp\_sum+=a[i]; 　　　　　　if(temp\_sum>max) 　　　　　　　　max=temp\_sum; 　　　　　　else if(temp\_sum<0) 　　　　　　　　temp\_sum=0; 　　} 　　return max; }  在这一遍扫描数组当中，从左到右记录当前子序列的和temp\_sum，若这个和不断增加，那么最大子序列的和max也不断增加(不断更新max)。如果往前扫描中遇到负数，那么当前子序列的和将会减小。此时temp\_sum 将会小于max，当然max也就不更新。如果temp\_sum降到0时，说明前面已经扫描的那一段就可以抛弃了，这时将temp\_sum置为0。然后，temp\_sum将从后面开始将这个子段进行分析，若有比当前max大的子段，继续更新max。这样一趟扫描结果也就出来了。    **二、面试题**    1、求直方图的最大内接矩形，假设每个细条的宽度为1，这个题很hot，两个人来问。我没想出什么好的算法。  2、NxN行列有序的矩阵查找一个数。以前有人遇到过。O(N)的时间复杂度。  3、给定一篇文章，求包含所有单词的最短摘要。O(N)的时间复杂度。  4、将MxN的矩阵转秩，要求O(1)的空间复杂度。参考群论中cyclic group,group generator  5、开放式问题，怎么避免重复抓取网页。  6、开放式问题，有些网站每天只允许有限次访问，怎么抓取网页使得索引尽量全面和新鲜。  7、写一个singleton pattern的例子。  8、vector vs. arraylist, growth strategy & complexity。  9、在C++文件中只declare class A, 但不以任何方式define class A, 是做什么用。  10、virtual function。  11、讨论html vs. xhtml vs. xml。  12、描述在浏览器中敲入一个网址后所发生的事情.dns,cache等。  13、给一个长度为n的整数数组，只允许用乘法不允许用除法，计算任意(n-1)个数的组合乘积中最大的一组。。。写出算法的时空复杂度。  1. Int to string (pay attention to the smallest negative number)  2. Given two sorted list, find the k smallest number  (binary search)  3. You can win three kinds of basketball points, 1 point, 2 points, and 3  points. Given a total score X, print out all the combination to compose X. ( recursion/ Dp)  4. Given a stream of integers, at a given time, there is a number appeared  more than half time, how to find this number.   (classic streaming algorithm)  //////////////////////////////////////////////面试书recruiter推荐 1) Mark Joshi:  "Quant Job Interviews:  Questions and Answers".   I have  heard very good things about this book. 2) Xinfeng Zhou,  "A Practical Guide to Quantitative Finance Interviews" 个人觉得非常有用， 大部分问题都在这两本上。 算法，C++, stochastic calculus 就看比较标准的几本。   - sqrt(i)=?  - You and me roll a dice,first one gets a six wins. You roll first. what  is the probability of you winning?  - A stair of n steps. Each time you step up 1 or 2 steps. How many  different ways are there to reach the top? what is the asymptotic limit?  - Moment generating function of standard model.  - Write a simple function to return Fibonacci number F(n). how to avoid  overflow?   - Do you know any variance reduction technique? What is control variate method? How and why it works?   - What is the future contract? Under what situation is the future  contract priced higher than forward price? What is the effect of the recent quantitative easing of fed?  - Explain one sorting algorithm. How does the merge sort work? What is  the running time for merge sort?  How does the quick sort works? What is the running time for the quick  sort? In practice, how do to choose among these two algorithm     - Given a Brownian motion starting from 0 and two barriers 'a' and '-b',  what is the probability of hitting barrier 'a' before hitting barrier '-b'? What if it is a process with drift?  - given a set of y\_i's and x\_i's, what is the linear regression beta?  how to interpret in terms of (co)variance? what is the correlation  between x\_i and the residuals?  - what is the c++ sort function? explain the heap sort. why running time  is O(nlgn)? what is quick sort? why it is O(nlgn)  - A european call option with strike 40 expires in 1 year. The current  spot  price is 50. what happens to the call price as the volatility goes  to infinity?  - Given two buckets of volumes 5L and 3L, how to obtain water of 4L?  - A stair with 10 levels. Each time you step up either 1 level or 2  levels. How many different ways are there to get to the top?  - what is AR, GARCH? what is GARCH used for in practice?  - what are the programming languages that you used? how confident are  you in programming?  - Describe one of the Monte Carlo simulations you have done.    - Given Brownian motion W, integrate WdW from 0 to T. how to interpret  (dW)^2 = dt? how to do the integral by definition? (Shreve's book)  - What is beta given vector observations: Y and X? what is the  requirement on X to make it work? what if two observables in X are  highly correlated? what technique can be used to fix the problem?   - what is semi-positive matrix? Prove covariance matrix is semi- positive.  - Fibonacci sequence. write a function that recursively calculate F(n).  what is the running time? what is the asymptotic limit of F(n)? How to  obtain simple lower and upper bound of F(n) when n is large.  - Give an situation when you could have a memory leak.   - Randomly pick three points, what is the probability that they will be  on the same hemisphere?    - How to price European option using Monte Carlo? (Assuming stock price  follows geometric brownian motion) Is Monte Carlo applicable in pricing  American option? How does the binomial tree pricing work for American option?   - Place cubes on a 10x10 grid. Place 1, 2 ,3 ..., 10 cubes in the first  line. 2,3,4,...,11 cubes in the second line. 3,4,5,...,12 cubes in the  third line. ... 10,11,12,...,19 cubes in the 10th line. what is the  total number of cubes(intuitively)?  - X1, X2 are standard normal, with correlation rho. What is the variance  of X=X1+X2.   - X1 and X2 are independent random variable with pdf f and g. what is  the pdf of X=X1+X2? What is the relation between Fourier coefficient of  convolution of two functions and   that of individual functions?  - Explain the copy constructor.  - Given uniform(0,1), how to generate a random variable with pdf f ?  - Given an example of variance reduction technique.  - What is the distribution of independent standard normals?   - Given two arrays, each with distinct integers, give an efficient  algorithm to find out the common elements.  - How to draw normal random given uniform variable generator?   - Show that geometric Brownian motion leads to log normal.   - A survey claims that 70% of people like coffee, 80% of people like  tea, what is the upper and lower bound on the percentage of people like  both coffee and tea?  - On a highway, the probability of observing at least one car in 20min  is 609/625, what is the probability of observing at least one car in  5min?  - Derive the Black-Scholes differential equation.  - Random walk  starts from 0. what is the probability of hitting -2  before hitting 3?  - Toss a dice. What is the expected number of tosses before getting two  consecutive 6's?  - Given a simple example of shell script that repeat certain task 10  times.  - What is a Brownian motion? What is the correlation of W(t1) and W(t2)?  - Algorithm to reverse the characters in a char array. Algorithm to  reverse the word order in a sentence with words seperated by white  space.   - 100 bulbs initially all off. First switch (on <-->off) the states of  1,2,...,100'th bulbs. The second time switch the states of  2,4,6,...,100'th bulbs. (the i'th time switch the states of i,2i,3i,...  bulbs). The last time switch the states of 100'th bulb. What bulbs are  on in the end?    - Asked about resume 100 times. Explained my research 100 times.  - Describe your thesis/research.   - Describe class hierarchy of any large computing system that I have  experience on.    - What is the "volatile" keyword?   - What is the "static" keyword?  - If one tosses a fair coin 5 times, what is the probability of not  getting 2 heads or 2 tails in a row?  - The ransom note problem. Design an algorithm that makes up the ransom  note from characters in a magazine.  - Introduce yourself  - What is a virtual function?  - Describe any sort algorithm  - How to derive Black-Scholes equation | |