百度历年笔试面试150题

**1**、用C语言实现一个revert函数，它的功能是将输入的字符串在原串上倒序后返回。  
  
**2**、用C语言实现函数void \* memmove(void \*dest, const void \*src, size\_t n)。memmove  
函数的功能是拷贝src所指的内存内容前n个字节到dest所指的地址上。  
分析：由于可以把任何类型的指针赋给void类型的指针，这个函数主要是实现各种数据类型的拷贝。  
  
**3**、有一根27厘米的细木杆，在第3厘米、7厘米、11厘米、17厘米、23厘米这五个位置上各有一只蚂蚁。  
木杆很细，不能同时通过一只蚂蚁。开始时，蚂蚁的头朝左还是朝右是任意的，它们只会朝前走或调头，但不会后退。  
当任意两只蚂蚁碰头时，两只蚂蚁会同时调头朝反方向走。假设蚂蚁们每秒钟可以走一厘米的距离。  
编写程序，求所有蚂蚁都离开木杆的最小时间和最大时间。  
  
**4**、给定一个存放整数的数组，重新排列数组使得数组左边为奇数，右边为偶数。  
要求：空间复杂度O(1)，时间复杂度为O（n）。  
  
**5**、在一维坐标轴上有n个区间段，求重合区间最长的两个区间段。  
  
**6**、系统有很多任务，任务之间有依赖，比如B依赖于A，则A执行完后B才能执行  
  （1）不考虑系统并行性，设计一个函数（Task \*Ptask,int Task\_num）不考虑并行度，最快的方法完成所有任务。  
  （2）考虑并行度，怎么设计

  typedef struct{

      int ID;

     int \* child;

      int child\_num;

  }Task;

  提供的函数:  
    bool doTask(int taskID);无阻塞的运行一个任务；  
    int waitTask(int timeout);返回运行完成的任务id，如果没有则返回-1；  
    bool killTask(int taskID);杀死进程  
  
**7**、解释下面ptr含义和不同  
double\* ptr = &value;  
    //ptr是一个指向double类型的指针，ptr的值可以改变，ptr所指向的value的值也可以改变   
const double\* ptr = &value  
    //ptr是一个指向const double类型的指针，ptr的值可以改变，ptr所指向的value的值不可以改变  
double\* const ptr=&value  
    //ptr是一个指向double类型的指针，ptr的值不可以改变，ptr所指向的value的值可以改变  
const double\* const ptr=&value  
    //ptr是一个指向const double类型的指针，ptr的值不可以改变，ptr所指向的value的值也不可以改变  
  
**8**、去掉const属性，例：  const double value = 0.0f;  double\* ptr = NULL;怎么才能让ptr指向value？  
    强制类型转换，去掉const属性，如ptr = <const\_cast double \*>(&value);  
[http://topic.csdn.net/u/201109 ... 09169](http://topic.csdn.net/u/20110925/16/e6248e53-1145-4815-8d24-9c9019d24bd8.html?seed=1665205011&r=75709169" \l "r_75709169" \t "_blank)  
  
**9**、一个数组保存了N个结构，每个结构保存了一个坐标，结构间的坐标都不相同，请问如何找到指定坐标的结构（除了遍历整个数组，是否有更好的办法）？（要么预先排序，二分查找。要么哈希。hash的话，坐标(x,y)你可以当做一个2位数，写一个哈希函数，把（x,y）直接转成“(x,y)”作为key，默认用string比较。或如Edward Lee所说，将坐标(x, y)作为 Hash 中的 key。例如(m, n)，通过 (m,n) 和 (n, m) 两次查找看是否在 HashMap 中。也可以在保存时就规定 (x, y) , x < y ，在插入之前做个判断。）  
  
**10**、百度最新面试题：现在有1千万个随机数，随机数的范围在1到1亿之间。现在要求写出一种算法，将1到1亿之间没有在随机数中的数求出来。  
（编程珠玑上有此类似的一题，如果有足够的内存的话可以用位图法，即开一个1亿位的bitset，内存为100m/8== 12.5m, 然后如果一个数有出现，对应的bitset上标记为1，最后统计bitset上为0的即可。）  
  
11、linux/unix远程登陆都用到了ssh服务，当网络出现错误时服务会中断，linux/unix端的程序会停止。为什么会这样？说下ssh的原理，解释中断的原理。  
  
12、利用互斥量和条件变量设计一个消息队列，具有以下功能：  
   1). 创建消息队列（消息中所含的元素）  
   2). 消息队列中插入消息  
   3). 取出一个消息（阻塞方式）  
   4). 取出第一消息（非阻塞方式）  
  
13、对已排好序的数组A，一般来说可用二分查找可以很快找到。现有一特殊数组A[]，它是循环递增的，如A[]={ 17 19 20 25 1 4 7 9}，试在这样的数组中找一元素x，看看是否存在。  
请写出你的算法，必要时可写伪代码，并分析其空间、时间复杂度。  
  
14、动态链接库与静态链接库的区别  
静态链接库是.lib格式的文件，一般在工程的设置界面加入工程中，程序编译时会把lib文件的代码加入你的程序中因此会增加代码大小，你的程序一运行lib代码强制被装入你程序的运行空间，不能手动移除lib代码。  
  动态链接库是程序运行时动态装入内存的模块，格式\*.dll，在程序运行时可以随意加载和移除，节省内存空间。  
  在大型的软件项目中一般要实现很多功能，如果把所有单独的功能写成一个个lib文件的话，程序运行的时候要占用很大的内存空间，导致运行缓慢；但是如果将功能写成dll文件，就可以在用到该功能的时候调用功能对应的dll文件，不用这个功能时将dll文件移除内存，这样可以节省内存空间。）  
   
15、指针与引用的区别  
相同点：1. 都是地址的概念；  
指针指向一块内存，它的内容是所指内存的地址；引用是某块内存的别名。  
区别：  
1). 指针是一个实体，而引用仅是个别名；  
2). 引用使用时无需解引用(\*)，指针需要解引用；  
3). 引用只能在定义时被初始化一次，之后不可变；指针可变；  
4). 引用没有 const，指针有 const；  
5). 引用不能为空，指针可以为空；  
6). “sizeof 引用”得到的是所指向的变量(对象)的大小，而“sizeof 指针”得到的是指针本身(所指向的变量或对象的地址)的大小；  
7). 指针和引用的自增(++)运算意义不一样；  
8).从内存分配上看：程序为指针变量分配内存区域，而引用不需要分配内存区域。）  
   
16、进程与线程的区别  
①从概念上：  
进程：一个程序对一个数据集的动态执行过程，是分配资源的基本单位。  
线程：一个进程内的基本调度单位。  
线程的划分尺度小于进程，一个进程包含一个或者更多的线程。  
②从执行过程中来看：  
进程：拥有独立的内存单元，而多个线程共享内存，从而提高了应用程序的运行效率。  
线程：每一个独立的线程，都有一个程序运行的入口、顺序执行序列、和程序的出口。但是线程不能够独立的执行，必须依存在应用程序中，由应用程序提供多个线程执行控制。  
③从逻辑角度来看：（重要区别）  
多线程的意义在于一个应用程序中，有多个执行部分可以同时执行。但是，操作系统并没有将多个线程看做多个独立的应用，来实现进程的调度和管理及资源分配。）  
   
17、函数调用入栈出栈的过程。  
   
18、c++对象模型与虚表。  
   
19、海量数据处理，以及如何解决Hash冲突等问题。  
  
20、系统设计，概率算法。  
  
21、判断一个数的所有因数的个数是偶数还是奇数（只需要你判断因数的个数是偶数个还是奇数个，那么可以这么做@滨湖&&土豆：那只在计算质因数的过程中统计一下当前质因数出现的次数，如果出现奇数次则结果为偶，然后可以立即返回；如果每个质因数的次数都是偶数，那么结果为奇。如果该数是平方数 结果就为奇  否则就为偶了）。  
  
22、写一个C的函数，输入整数N，输出整数M，M满足：M是2的n次方，且是不大于N中最大的2的n次方。例如，输入4,5,6,7，都是输出4 。  
   
23、C++中虚拟函数的实现机制。  
   
24、写出选择排序的代码及快速排序的算法。  
  
 25、你认为什么排序算法最好？  
   
26、tcp/ip的那几层协议，IP是否是可靠的？为什么？  
   
27、进程和线程的区别和联系，什么情况下用多线程，什么时候用多进程?  
   
28、指针数组和数组指针的区别。  
   
29、查找单链表的中间结点。  
   
30、最近在实验室课题研究或工作中遇到的技术难点，怎么解决的？  
   
31、sizeof和strlen的区别。  
   
32、malloc-free和new-delete的区别  
   
33、大数据量中找中位数。  
   
34、堆和栈的区别。  
   
35、描述函数调用的整个过程。  
   
36、在一个两维平面上有三个不在一条直线上的点。请问能够作出几条与这些点距离相同的线？  
  
37、假设有一台迷你计算机，1KB的内存，1MHZ的cpu，已知该计算机执行的程序可出现确定性终止(非死循环)，问如何求得这台计算机上程序运行的最长时间，可以做出任何大胆的假设。  
  
38、C++ STL里面的vector的实现机制，  
1).当调用push\_back成员函数时，怎么实现？  
（粗略的说@owen，内存足则直接 placement new构造对象，否则扩充内存，转移对象，新对象placement new上去。具体的参见此文：[http://blog.csdn.net/v\_july\_v/ ... 81522](http://blog.csdn.net/v_july_v/article/details/6681522" \t "_blank)）  
2).当调用clear成员函数时，做什么操作，如果要释放内存该怎么做。  
（调用析构函数，内存不释放。 clear没有释放内存，只是将数组中的元素置为空了，释放内存需要delete。）  
  
39、函数foo找错，该函数的作用是将一个字符串中的a-z的字母的频数找出来

void foo(char a[100], int cnt[256])

{

    memset(cnt, 0, sizeof(cnt));

    while (\*a != '\0')

    {

        ++cnt[\*a];

        ++a;

    }

    for (char c = 'a'; c <= 'z'; ++c)

    {

        printf("%c:%d\n", c, cnt[c]);

    }

}

int main()

{

    char a[100] = "百度abc";

    int cnt[256];

    foo(a, cnt);

    return 0;

}

40、设子数组A[0:k]和A[k+1:N-1]已排好序(0≤K≤N-1)。试设计一个合并这2个子数组为排好序的数组A[0:N-1]的算法。要求算法在最坏情况下所用的计算时间为O(N)，只用到O(1)的辅助空间。  
  
41、一个单词如果交换其所含字母顺序，得到的单词称为兄弟单词，例如mary和army是兄弟单词，即所含字母是一样的，只是字母顺序不同，用户输入一个单词，要求在一个字典中找出该单词的所有兄弟单词，并输出。给出相应的数据结构及算法。要求时间和空间复杂度尽可能低  
目前思想：

struct {

   char  data；

   int  n

}；

根据数学定理：任何一个大于1的自然数N，都可以唯一分解成有限个质数的乘积 N=(P\_1^a1)\*(P\_2^a2)......(P\_n^an) , 这里P\_1<P\_2<...<P\_n是质数，且唯一。  
例如  
a=2 b=3 c=5 d=7 e=11...  
f(abcd)=2\*3\*5\*7=210   
然后字典里找乘积210的位数相同的一定是这5个字母组合的单词就是兄弟单词 。  
  
42、设rand（s，t）返回[s,t]之间的随机小数，利用该函数在一个半径为R的圆内找随机n个点，并给出时间复杂度分析。  
  
43、为分析用户行为，系统常需存储用户的一些query，但因query非常多，故系统不能全存，设系统每天只存m个query，现设计一个算法，对用户请求的query进行随机选择m个，请给一个方案，使得每个query被抽中的概率相等，并分析之，注意：不到最后一刻，并不知用户的总请求量。  
  
44、C++ STL中vector的相关问题：  
    （1）、调用push\_back时，其内部的内存分配是如何进行的？  
    （2）、调用clear时，内部是如何具体实现的？若想将其内存释放，该如何操作？  
  
45、正常用户端每分钟最多发一个请求至服务端，服务端需做一个异常客户端行为的过滤系统，设服务器在某一刻收到客户端A的一个请求，则1分钟内的客户端任何其它请求都需要被过滤，现知每一客户端都有一个IPv6地址可作为其ID，客户端个数太多，以至于无法全部放到单台服务器的内存hash表中，现需简单设计一个系统，使用支持高效的过滤，可使用多台机器，但要求使用的机器越少越好，请将关键的设计和思想用图表和代码表现出来。  
  
46、如p([1,2,3])输出：  
[123]、[132]、[213]、[231]、[321]、[323]  
 求一个组合函数  
如p([1,2,3])输出：  
[1]、[2]、[3]、[1,2]、[2,3]、[1,3]、[1,2,3]  
这两问可以用伪代码。  
  
47、如何快速访问ipv6地址呢？ipv6地址如何存放？  
  
48、正常用户端每分钟最多发一个请求至服务端，服务端需做一个异常客户端行为的过滤系统，设服务器在某一刻收到客户端A的一个请求，则1分钟内的客户端任何其它请求都需要被过滤，现知每一客户端都有一个IPv6地址可作为其ID，客户端个数太多，以至于无法全部放到单台服务器的内存hash表中，现需简单设计一个系统，使用支持高效的过滤，可使用多台机器，但要求使用的机器越少越好，请将关键的设计和思想用图表和代码表现出来。   
  
49、一个单词单词字母交换，可得另一个单词，如army->mary，成为兄弟单词。提供一个单词，在字典中找到它的兄弟。描述数据结构和查询过程。  
评点：同去年9月份的一道题，见此文第3题：[http://blog.csdn.net/v\_july\_v/ ... 03368](http://blog.csdn.net/v_july_v/article/details/6803368" \t "_blank)。  
  
50、线程和进程区别和联系。什么是“线程安全”  
  
51、C和C++怎样分配和释放内存，区别是什么  
  
52、一个url指向的页面里面有另一个url,最终有一个url指向之前出现过的url或空，这两种情形都定义为null。这样构成一个单链表。给两条这样单链表，判断里面是否存在同样的url。url以亿级计，资源不足以hash。  
  
53、数组al[0,mid-1] 和 al[mid,num-1]，都分别有序。将其merge成有序数组al[0,num-1]，要求空间复杂度O(1)。  
  
54、百度搜索框的suggestion，比如输入“北京”，搜索框下面会以北京为前缀，展示“北京爱情故事”、“北京公交”、“北京医院”等等搜索词，输入“结构之”，会提示“结构之法”，“结构之法 算法之道”等搜索词。  
请问，如何设计此系统，使得空间和时间复杂度尽量低。  
评点：①直接上Trie树「Trie树的介绍见：从Trie树（字典树）谈到后缀树」 +  TOP K「hashmap+堆，hashmap+堆 统计出如10个近似的热词，也就是说，只存与关键词近似的比如10个热词，详细方法可参看此文第2个题的讲解：[http://blog.csdn.net/v\_july\_v/ ... 82693](http://blog.csdn.net/v_july_v/article/details/7382693" \t "_blank)」？   
②or Double-array trie tree？同时，StackOverflow上也有两个讨论帖子：[http://stackoverflow.com/quest ... plete](http://stackoverflow.com/questions/2901831/algorithm-for-autocomplete" \t "_blank)，[http://stackoverflow.com/quest ... e-c-c](http://stackoverflow.com/questions/1783652/what-is-the-best-autocomplete-suggest-algorithm-datastructure-c-c" \t "_blank)。  
③此外，这里有一篇关于“拼写错误检查”问题的介绍，或许对你有所启示：[http://blog.afterthedeadline.c ... ions/](http://blog.afterthedeadline.com/2010/01/29/how-i-trie-to-make-spelling-suggestions/" \t "_blank)。。  
  
55、不使用随机数的洗牌算法，详情：[http://topic.csdn.net/u/201208 ... .html](http://topic.csdn.net/u/20120831/10/C837A419-DFD4-4326-897C-669909BD2086.html" \t "_blank)。  
  
56、公司组织一次羽毛球比赛，采用淘汰制，假设公司共有1001人，如果要评出“公司羽毛球第一高手”的称号，至少需要进行多少场比赛？请简述设计过程，并编写代码模拟比赛过程（语言不限，可以使用伪代码）。  
  
57、一百个灯泡排成一排，第一轮将所有灯泡打开，第二轮每隔一个灯泡关掉一个，即排在偶数的灯泡都被关掉，第三轮每隔两个灯泡，将开着的灯泡关掉，关掉的灯泡打开，以此类推，第100轮结束的时候，还有几盏灯泡亮着？  
58、假定有20个有序数组，每个数组中有500个数字，数字类型32位uint数值，降序排列，现在需要取出这10000个数字中最大的500个，怎么做？  
  
59、手机上通常采用九键键盘输入，即：1—9个数字键分别对应一定的英文字母（如：2对应ABC，3对应DEF，……，9对应WXYZ）。因此，用户可以方便的输入中文内容，比如，用户输入“926”，可以对应“WXYZ”，“ABC”和“MNO”的一系列组合“WAN”，“YAN”，“ZAO”等，这些对应“万”，“严”，“早”等汉字的中文拼音。  
要求：  
现在我们把这样的输入方式应用在我们的手机联系人查找功能上，有一个联系人列表UserList，记录了（姓名，手机号）这样的组合，通过输入的数组字符串NumStr，按照下面的规则把对应的联系人查找出来，返回一个ResultList。  
规则：  
1).手机号能连续部分匹配输入的数字字符串NumStr，如：输入NumStr=926，则手机号为13926811111会被查找出来。  
2).联系人姓名中的汉字转化成拼音后能够连续匹配输入数字字符串NumStr对应的英文字母组合，如：输入NumStr=926，则联系人“王二”，“万事通”，“李艳”会被查找出来，因为“王二”的“王”的拼音“WANG”中含有“WAN”，和“926”能匹配。  
输入：联系人列表UserList<UserName,PhoneNo>;汉字拼音映射表Dict;数字字符串NumStr。  
输出：符合规则的联系人列表ResultList<UserName,PhoneNo>。  
  
60、10亿个int型整数，如何找出重复出现的数字。  
  
61、有2G的一个文本文档，文件每行存储的是一个句子，每个单词是用空格隔开的。问：输入一个句子，如何找到和它最相似的前10个句子。（提示：可用倒排文档）。  
  
62、一个处理器最多能处理m个任务。现在有n个任务需要完成，每个任务都有自己完成所需的时间。此外每个任务之间有依赖性，比如任务Ａ开始执行的前提是任务Ｂ必须完成。设计一个调度算法，使得这n这任务的完成时间最小。  
  
63、有一个排序二叉树，数据类型是int型，如何找出中间大的元素。  
  
64、一个N个元素的整形数组，如何找出前K个最大的元素。  
  
65、给定一个凸四边形，如何判断一个点在这个平面上。  
点评：本题的讨论及参考答案请见这：[http://www.51nod.com/question/ ... 3D669](http://www.51nod.com/question/index.html" \l "!questionId=669" \t "_blank)。  
  
66、堆和栈的区别。  
  
67、问如何数出自己头上的头发。  
  
68、给定一数组，输出满足2a=b（a，b代表数组中的数）的数对，要求时间复杂度尽量低。  
  
69、搜索引擎多线程中每个线程占用多少内存？如果搜索引擎存储网页内存占用太大怎么解决？  
  
70、有很多url，例如\*.baidu.com，\*.sina.com ......  
现在给你一个sports.sina.com 快速匹配出是\*.sina.com。点评：老题，此前blog内曾整理过。  
  
71、找出字符串的编辑距离，即把一个字符串s1最少经过多少步操作变成编程字符串s2，操作有三种，添加一个字符，删除一个字符，修改一个字符（只要听过编辑距离，知道往动态规划上想，很快就可以找到解法）。  
  
点评：请看链接：[http://blog.csdn.net/Lost\_Pain ... 57334](http://blog.csdn.net/Lost_Painting/article/details/6457334" \t "_blank)。  
  
72、编程实现memcopy，注意考虑目标内存空间和源空间重叠的时候。  
  
73、实现简单的一个查找二叉树的深度的函数。  
  
74、进程和线程的区别。  
  
75、一个有序数组（从小到大排列），数组中的数据有正有负，求这个数组中的最小绝对值。  
  
76、链表倒数第n个元素。  
  
77、有一个函数fun能返回0和1两个值，返回0和1的概率都是1/2，问怎么利用这个函数得到另一个函数fun2，使fun2也只能返回0和1，且返回0的概率为1/4,返回1的概率为3/4。（如果返回0的概率为0.3而返回1的概率为0.7呢）。  
  
78、有8个球，其中有7个球的质量相同，另一个与其他球的质量不同（且不知道是比其他球重还是轻），请问在最坏的情况下，最少需要多少次就能找出这个不同质量的球。  
  
79、有一个数组a，设有一个值n。在数组中找到两个元素a[i]和a[j]，使得a[i]+a[j]等于n，求出所有满足以上条件的i和j。  
  
80、1万个元素的数组，90%的元素都是1到100的数，10%的元素是101--10000的数，如何高效排序。  
  
81、用简单语句描述数据库操作的步骤 。  
  
82、写出TCP/IP的四层结构。   
  
83、什么是MVC结构，并描述各层结构的作用 。  
  
84、字母a-z，数字0-9，现需要其中任意3个作为密码，请输出所有可能组合。（伪码\C\C++\JAVA）  
点评：如本文评论下第198楼所述，即从26+10=36个不同字符中选取3个字符的组合，用递归及非递归两种方法，可以参照以下链接：  
[http://blog.csdn.net/wumuzi520 ... 87501](http://blog.csdn.net/wumuzi520/article/details/8087501" \t "_blank)（从n个数中选取m个数的组合数）。  
  
85、实现字符串反转函数。  
  
86、给定字符函数a、插入 b、删除 c、替换   
例如字符串A=acegf，字符串B=adef，最少需要2步操作将A转换为B，  
即第一步将c替换为d，第二步将g删除；   
1).请问将字符串A=gumbo转换为字符串B=gambol，最少需要几步操作，列出如何操作  
2).任意字符串A和字符串B，如何计算最小操作次数，计算思路，并给出递归公式  
3).实现代码（注意代码风格与效率）  
  
87、RSA SecurID安全系统   
应用场景：这是一种用户登录验证手段，例如银行登录系统，这个设备显示6位数字，每60秒变一次，再经过服务器认证，通过则允许登录。问How to design this system？   
1).系统设计思路？服务器端为何能有效认证动态密码的正确性？   
2).如果是千万量级永固，给出系统设计图示或说明，要求子功能模块划分清晰，给出关键的数据结构或数据库表结构。   
考虑用户量级的影响和扩展性，用户密码的随机性等，如果设计系统以支持这几个因素.   
3).系统算法升级时，服务器端和设备端可能都要有所修改，如何设计系统，能够使得升级过程（包括可能的设备替换或重设）尽量平滑？  
  
88、什么是RISC。  
  
89、通过后序、中序求前序 。  
  
90、重写与重载的区别 。  
  
91、判断两个数组中是否有相同的数字 。  
  
92、1000瓶水中找 出有毒的那瓶，毒性一周后发作，一周内最少需要多少只老鼠 。  
  
93、系统设计 email客户端，支持多账户和pop3等协议   
1). 请写出可能的至少5个用例；   
 2). 使用sqlite存储帐户、已收信息、已发信息、附件、草稿，请设计合理的表结构   
 3). pop3等协议等接口已完成，请给出email客户端的模块设计图。  
  
94、百度地图里的路线查询：给定两个站点，如果没有直达的路线，如何找到换乘次数最少的路线？  
点评：蚂蚁算法？还是广搜，或A\*算法？   
   
95、有一箱苹果，3个一包还剩2个，5个一包还剩3个，7个一包还剩2个，求N个满足以上条件的苹果个数。  
  
96、用递归算法写一个函数，求字符串最长连续字符的长度，比如aaaabbcc的长度为4，aabb的长度为2，ab的长度为1。  
  
97、假设一个大小为100亿个数据的数组，该数组是从小到大排好序的，现在该数组分成若干段，每个段的数据长度小于20「也就是说：题目并没有说每段数据的size 相同，只是说每个段的 size < 20 而已」，然后将每段的数据进行乱序（即：段内数据乱序），形成一个新数组。请写一个算法，将所有数据从小到大进行排序，并说明时间复杂度。  
点评：  
思路一、如@四万万网友所说：维护一个20个元素大小的小根堆，然后排序，每次pop取出小根堆上最小的一个元素（log20），然后继续遍历原始数组后续的(N-20)个元素，总共pop (N-20)次20个元素小根堆的log20的调整操作。  
思路二@飘零虾、如果原数组是a[]，那么a[i+20]>=a[i]恒成立（因为每段乱序区间都是小于20的，那么向后取20，必然是更大的区间的元素）。  
第一个数组：取第0、20、40、60、80...  
第二个数组：取第1、21、41、61、81...  
...  
第20个数组：取第19、39、59、79...     （上述每个数组100亿/20 个元素）  
共计20个数组，每个数组100亿/20 个元素「注：这5亿个元素已经有序，不需要再排序」，且这20个数组都是有序的，然后对这20个数组进行归并，每次归并20个元素。时间复杂度跟上述思路一一样，也是N\*logK（N=100亿，K=20）。  
此外，读者@木叶漂舟直接按每组20个排序,将排好的20个与前20个调整拼接,调整两端接头处的元素,写了个简单地demo： [http://t.cn/zlELAzs](http://t.cn/zlELAzs" \t "_blank)。不过，复杂度有点高，目前来说中规中矩的思路还是如上文中@四万万网友 所说思路一「@张玮-marihees按照思路一：[http://weibo.com/1580904460/z1v5jxJ9P](http://weibo.com/1580904460/z1v5jxJ9P" \t "_blank)，写了一份代码：[http://codepad.org/T5jIUFPG](http://codepad.org/T5jIUFPG" \t "_blank)，欢迎查看」。  
  
98、一在线推送服务，同时为10万个用户提供服务，对于每个用户服务从10万首歌的曲库中为他们随机选择一首，同一用户不能推送重复的，设计方案，内存尽可能小，写出数据结构与算法。  
  
99、来自《编程之美》的概率题：一个桶里面有白球、黑球各100个，现在按下述规则取球：的  
    i 、每次从通里面拿出来两个球；  
    ii、如果取出的是两个同色的求，就再放入一个黑球；  
    ii、如果取出的是两个异色的求，就再放入一个白球。  
问：最后桶里面只剩下一个黑球的概率是多少？  
  
100、给你一个自然数N，求[6,N]之内的所有素数中，两两之和为偶数的那些偶数。  
  
101、相似度计算用于衡量对象之间的相似程度，在数据挖掘，自然语言处理中是一个基础性计算，在广告检索服务中往往也会判断网民检索Query和Adword的主题相似度，假设Query或者Adword的主题属性定义为一个长度为10000的浮点数组Pr[10000]（称之为主题概率数组），其中Pr[i]表示Query或者Adword属于主题ID为I的概率，而Query和Adword的相似度简化定义为两者主题概率数组的内积：即sim(Query,Adword)=sum(QueryPr[i]\*AdwordPr[i](http://ask.julyedu.com/0%3C=i%3C10000" \t "_blank),在实际应用场景中，由于大多数主题的概率都为0，所以主题概率数组往往比较稀疏，在实现时会以一个紧凑型数组topic\_info\_t[]的方式保存，其中100<=数组大小<=1000,并按照topic\_id递增排列，0<=topic\_id<10000,0<topic\_pr<1.

Struct topic\_info\_t{

Int topic\_id;

Float topic\_pr;

};

现在给出Query的topic\_info\_t数组和N(N>=5000)个Adwords的topic\_info\_t数组，现要求出Query与Adwords的相似度最大值，即max(sim(Query,Adword[i]))(0<=i<N).  
Float max\_sim(comst vector<topic\_info\_t>&query\_topic\_info,  
Const vector<topic\_info\_t>adwords\_topic\_info[],  
Int adwords\_number);  
  
102、动态链接库和静态链接库分别有什么优缺点？  
  
103、轮询任务调度与抢占式任务调度的区别？  
  
104、待更新。  
105、待更新。  
  
106、长度为N（N很大）的字符串，求这个字符串里的最长回文子串。  
  
107、数轴上从左到右有n个点a[0],a[1]...a[n-1],给定一根长度为L的绳子，求绳子最多能覆盖其中的几个点。  
  
108、三色球排序的问题，相同的球放到一起，让你按顺序输出红白蓝三种颜色的球，可以用012来表示，要求只能扫描一次数组。  
点评：荷兰国旗问题，参见此文第8小节：[http://blog.csdn.net/v\_july\_v/ ... 11155](http://blog.csdn.net/v_july_v/article/details/6211155" \t "_blank)。  
  
109、实现C的strstr  
点评：手写字符串处理相关函数是面试中极为常见的一类题型。  
功能：从字符串str1中查找是否有字符串str2，  
-如果有，从str1中的str2位置起，返回str1中str2起始位置的指针，如果没有，返回null。  
C++参考代码：

//copyright@caopengcs 2013/10月

class Solution {

public:

    char \*strStr(char \*haystack, char \*needle) {

        // Start typing your C/C++ solution below

        // DO NOT write int main() function

        int i, j;

        for (i  = j  = 0; haystack[i] && needle[j];) {

            if (haystack[i] == needle[j]) {

                ++i;

                ++j;

            }

            else {

                i  = i  - j  + 1;

                j  = 0;

            }

        }

        return needle[j] ? 0 : (haystack  + i  - j);

    }

};

110、写一个memmove的函数  
点评：手写常见字符串处理函数是面试官很喜欢考的一类题型。  
  
111、JAVA里面的线程同步机制、异常处理机制、集合类、简单的设计模式、hashmap和hashtable的区别，及HashMap和ConcurrentHashMap的区别。  
  
112、给出数组A={a\_0,a\_1,a\_2,...,a\_n}（n是可变的），打印出所有元素的组合  
  
113、数组A中任意两个相邻元素大小相差1，现给定这样的数组A和目标整数t，找出t在数组A中的位置。  
  
114、求二叉树的面积（高乘宽），高为二叉树根到叶子节点的最大距离，宽慰二叉树最多的节点数。  
  
115、给了一个百度地图的截图，对于地图上的某一点，需要在地图上标注该点的信息，将信息抽象成一个矩形，可以在该点的左边标记，也可以在该点右边标记。但是任意两点标记后的矩形是不能有覆盖的，否则删除其中一个点  
    问题1，现给一固定区域，有n个点，设计一个算法，要求标记足够多的点  
    问题2，当点足够多时候，算法会遇到性能瓶颈，需要对算法重新优化。  
更多题目请参见：[http://blog.csdn.net/xyanghome ... 87771](http://blog.csdn.net/xyanghomepage/article/details/12687771" \t "_blank)。  
  
116、深度神经网络目前有哪些成功的应用？简述原因。  
   
117、列举不同进程共享数据的方式（至少三种）。  
   
118、对于N个样本，每个样本为D维向量，采用欧式距离使用KNN做类预测。  
1).给出预测时间复杂度。  
2).当N很大时，有哪些方法可以降低复杂度？  
3).k取值的大小对预测方差和偏差有何影响？  
  
119、给出一个数据A=[a\_0, a\_1, a-2, ... a\_n](http://ask.julyedu.com/%E5%85%B6%E4%B8%ADn%E5%8F%AF%E5%8F%98" \t "_blank)，打印出该数值元素的所有组合。  
   
120、有这样一个数组A,大小为n，相邻元素差的绝对值都是1，如A={4,5,6,5,6,7,8,9,10,9}。现在给定数组A和目标整数t，请找到t在数组中的位置。  
  
121、在平面上有一组间距为d的平行线，将一根长度为l(l<d)的针任意掷在这个平面上，求此针与平行线中任意一根相交的概率，用高等数学(微积分、概率的方法)求解，基于布丰投针的结论，任选一种编程语言(C/C++, matlab, python, java)，写出模拟投针实验(程序中允许把一个理想的Pi作为常量使用)，求解圆周率。  
  
122、关于K-means聚类算法，请回答以下问题：  
1).写出将N个样本X=(x1, ... xN)聚类成k类的k\_means聚类算法的优化目标；  
2).描述K-means终止的常用条件；  
 3).以Kmeans算法为例，描述Expectation-Maximization(EM)算法的基本原理与步骤。  
4).用伪代码给出基于MPI或者HADOOP的Kmeans并行算法。  
题目来源：[http://blog.csdn.net/luoweifu/ ... 85169](http://blog.csdn.net/luoweifu/article/details/12685169" \t "_blank)。  
  
123、简述计算机的存储系统分为哪几个层次，为什么这样的分层能够提高程序的执行效率。  
  
124、浮点数在计算中如何表示，如何对浮点数判等。  
  
125、简述TCP与UDP协议的差别，两者与HTTP的关系。并列举HTTP的方法，以及常见的返回状态码。  
  
126、设计一个反转字符串的函数 char \*reverse\_str(char \*str),不使用系统函数。  
  
127、给定一个字符串,（1，（2，3），（4，（5，6），7））,使它变为（1，2，3，4，5，6，7），设计一个算法消除其中嵌套的括号。(c/c++)  
  
128、使用C语言实现htonl（将long性转为网络字节码），不使用系统自带函数。  
  
129、面向对象是一种思想，使用C语言来实现下列问题。  
1). 如何定义一个类？  
 2). 如何创建以及销毁对象？  
3). 如何实现类的继承？  
题目来源：[http://blog.csdn.net/cocoarann ... 91025](http://blog.csdn.net/cocoarannie/article/details/12691025" \t "_blank)。  
  
130、数组A中任意两个相邻元素大小相差1，在其中查找某个数。  
数组A中任意两个相邻元素大小相差1，现给定这样的数组A和目标整数t，找出t在数组A中的位置。  
如数组：[1,2,3,4,3,4,5,6,5]，找到4在数组中的位置。  
这道题目最差时间复杂度也是O（N），所以重点在于能不能找到一种尽可能减少比较次数的方法。@jefflee 的方法就很不错，但感觉应该还可以继续优化？  
  
131、给定n个元素，打印出全排列  
比如输入1 2 3，打印出6种排列情况   
  
132、有两个不同的数在1-30之间（不包括1和30），甲知道两数之和，乙知道两数之积。乙问甲知道是那两个数吗？甲说不知道。甲同样反问乙，乙也说不知道。然后乙说我知道了，再然后甲说我知道了。请问是哪两个数？  
解析：前提是甲不能能通过两数之和确定具体哪两个数，乙也不能通过两数之积判断出具体哪两个数。  
然后当乙知道甲也不能确定的时候，乙却可以快速判断出来说明甲心里已经有了几个选项，然后根据甲不确定就可以排除掉不正确的。然后甲也是如此。  
来源：[http://ask.julyedu.com/question/261](http://ask.julyedu.com/question/261" \t "_blank)  
  
133、子query统计和重要子query识别  
问题定义：  
当query A切词后的term集是query B切词后的term集的真子集时，称query A为query B的子query，例如：  
“刘德华”的切词结果为“刘德华”；  
“刘德华电影”的切词结果为“刘德华 电影”；  
“刘德华最新电影”的切词结果为“刘德华 最新 电影”；  
“刘德华电影下载”的切词结果为“刘德华 电影 下载”；  
根据以上切词结果，刘德华”是“刘德华电影”，“刘德华最新电影”, “刘德华电影下载”的子query；  
“刘德华电影”是“刘德华最新电影”, “刘德华电影下载”的子query；  
但是，“刘德华电影下载”和“刘德华最新电影”互相不是对方的子query。  
现有亿级的用户query，并且知道每个query的查询次数，要求：  
（1）列出一个query的全部子query，写出C语言实现。  
（2） query中的不同term对这个query的重要性不同的，例如“刘德华 电影 下载”中“刘德华”和“电影”的重要性比“下载”重要，因为：“刘德华 电影“所表达的查询需求，与”刘德华 下载“或者”电影 下载“相比，更接近原query的需求。根据(1)中的统计的子query数据，请给出一种思路，来计算一个query中的所有子query的重要性排序。  
如果认为子query数据的信息不够充分，请给出还需要哪些信息，以及获得这些信息的途径，给出算法思路描述，必要的符号和推理公式即可。  
  
134、给定多个集合，求他们的笛卡尔积。  
比如给定{a, b}, {1,2,3}，结果为{a, 1}, {a, 2}, {a, 3}, {b,1}, {b, 2}, {b, 3}；  
要求时间和空间复杂度尽可能低，不要使用递归，不要使用类似树的非递归实现。  
  
135、一个单词单词字母交换，可得另一个单词，如army->mary，成为兄弟单词。提供一个单词，在字典中找到它的兄弟。描述数据结构和查询过程。  
  
136、假设张三的mp3里有1000首歌，现在希望设计一种随机算法来随机播放。与普通随机模式不同的是，张三希望每首歌被随机到的改了吧是与一首歌的豆瓣评分（0~10分）成正比的，如朴树的《平凡之路》评分为8.9分，逃跑计划的《夜空中最亮的星》评分为9.5分，则希望听《平凡之路》的概率与《夜空中最亮的星》的概率比为89:95,。现在我们已知这1000首歌的豆瓣评分：  
1).请设计一种随机算法来满足张三的需求。  
2).请写代码实现自己的算法。  
**社区讨论地址**：[http://ask.julyedu.com/question/127](http://ask.julyedu.com/question/127" \t "_blank)。  
  
137、给定任意一个正整数，求比这个数大且最小的“不重复数”，“不重复数”的含义是相邻两位不相同，例如1101是重复数，而1201是不重复数。  
  
138[6,N]之内的所有素数中，两两之和为偶数的那些偶数  
RT  
其中N是个自然数，请给出算法描述，代码与时间复杂度分析。  
  
139在由N个正整数的集合S中，找出最大元素C，满足C=A + B  
RT  
其中A,B都是集合S中元素，请给出算法描述，代码与时间复杂度分析。  
  
140、请列举出你熟悉的知名论坛/社区的名称、URL、优势以及原因。  
  
141、如何提高老年人和儿童使用手机百度的频率？  
  
142、百度卫士新推出看片保护（观看视频时防止病毒侵害）功能， 请针对该功能设计一个具体的运营规划。  
  
143、Cookie、sessionStorage、localStorage的区别  
  
144、javascript中call()方法和apply()方法的区别。  
  
145、什么是 “use strict”? 使用它的好处和坏处是什么？  
  
146、写一段简单的正则表达式，匹配并取出字符串”[https://www.baidu.com/s?cl=3](https://www.baidu.com/s?cl=3" \t "_blank)”中的域名部分（注：域名部分非固定）  
  
147、用原生javascript编写程序：创建一个ul无序列表元素添加到body中，ul下包含5个li元素，每个li元素包含一个text类型元素，text元素内容可自定义。  
  
148、假设有一个基础对象叫“动物”，拥有以下属性：腿的数量、是否有尾巴，有另外一个对象叫“猫”，拥有“动物”对象的属性，并增加一个属性为：动物名称，再增加一个方法，返回动物名称+腿的数量+是否有尾巴的描述，请使用javascript原型链继承来创建以上2个对象。  
  
149、请解释tcp连接建立过程，如果可能，请结合相应系统调用函数解释交互过程。  
  
150、给定一个整数的数组，相邻的数不能同时选，求从该数组选取若干整数，使得他们的和最大，要求只能使用o(1)的空间复杂度。要求给出伪码。  
  
151、二分查找是常用的编程方法，请用完整代码实现该函数（不许调用库函数）  
void \*bsearch(const void \*key, const void \*base, size\_t nel, size\_t width, int (\*compar) (const void \*, const void \*));  
  
152、对于Edit控件，你如何抓防止密码框内容被抓取？  
  
153、DNS欺骗的方式有哪些？  
  
154、列举两种应用层中简单的跨进程<span>DLL</span>注入的方法。  
  
155、假设有如下所示的一个数字金字塔，现在，要求写一个程序来查找从顶点到底部任意处结束的路径，使路径经过的数字的和最大，并输出该路径的最大和。比如以下金字塔的和最大路径的和为7+3+8+7+5=30。  
  
7  
3 2  
8 1 0  
2 7 4 4  
4 5 2 6 5  
  
156、假设有如下字符串： (234453)[234]{2324} 现在，要求编程分析其括号配对是否正确。请自行选择下列两种方案之一实现该程序：  
方案一：不考虑括号优先级，只考虑配对正确性；方案二：考虑括号优先级，比如{1[2（3）4]5} 是正确的。但是[1{2}3]是不正确的。   
  
157、百度是一个大型网站，内部含有多个产品线，比如广为人知的贴吧、知道、空间等应用。然而设计这些应用的统一登录平台却是一件非常艰巨的挑战。需要考虑到通用性和安全性。  
1). 对于一个Web应用程序，主要的身份验证和凭证保持的方法主要有cookie和session两种。他们又是如何起作用的？各有哪些优缺点？  
2). 影响到cookie值作用范围的因素有哪些？请一一说明。  
3) .从安全角度来考虑，一个大型网站的单点登录可能会引入哪些安全问题？如何设计安全的在线单点登录系统？   
  
158、HTML的Doctype作用? 严格模式与混杂模式如何区分？它们有何意义?  
  
159、请用CSS实现如下图的样式，相关尺寸如图示，其中dom结构为：  
<div id=”demo”></div>  
  
160、简述document.write和 innerHTML的区别。  
  
161、你知道的，javascript语言的执行环境是"单线程模式"，这种模式的好处是实现起来比较简单，执行环境相对单纯；坏处是只要有一个任务耗时很长，后面的任务都必须排队等着，会拖延整个程序的执行，因此很多时候需要进行“异步模式”，请列举js异步编程的方法。  
  
162、用户从手机的浏览器访问[www.baidu.com](http://www.baidu.com/" \t "_blank)，看到的可能跟桌面PC电脑，是不太一样的网页效果，会更适合移动设备使用。请简要分析一下，实现这种网页区分显示的原因及技术原理。  
  
163、Flappy Bird是风靡一时的手机游戏，玩家要操作一只小鸟穿过无穷无尽的由钢管组成的障碍。如果要你在HTML前端开发这个游戏，为了保证游戏的流畅运行，并长时间运行也不会崩溃，请列举开发要注意的性能问题和解决的方法。  
  
164、如下图，请实现表格信息的排序功能，当点击表头的属性区域，将表格信息进行排序切换功能，即第一次点击为降序排序，再一次点击进行升序排序。  
  
165、C++有哪些数据类型?为什么long和int都是4字节?  
  
166、JAVA和C++的区别是什么?分别用在什么情景比较好?  
  
167、给定一个文件每一行是字符串，找出所有的逆序对，比如abc和cba是逆序的对。  
  
168、给定一个奇数n，比如n=3，生成1到n平方的数，如1到9，填入九宫格，使得横竖斜的和都相等。  
  
169、C和C++有什么区别，能用C实现C++所有功能吗?C能实现多态吗?  
  
170、25匹马，5条赛道，一匹马一个赛道，比赛只能得到5匹马之间的快慢程度，而不是速度，求决胜1，2，3名至少多少场。  
  
171、请用c++ 实现stl中的string类，实现构造，拷贝构造，析构，赋值，比较，字符串相加，获取长度及子串等功能。