作者：NgShawn  
链接：[https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2](https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2" \t "_blank)  
来源：牛客网

## 智商题

1、36匹马，6条跑道，选出最快3匹，最少赛多少场？

**答案：**我的思路如下：首先这应该是多轮问题。

（1）初筛：36辆车分6组，每组6辆。这样，比下来，共需要6次。

（2）再筛：在初筛中，最坏情况下，有一组的前三名就是36辆车的前三名。因此，每个组都需要保留前三名（后三名就可以踢出去了。）

第一名的6辆车放在一起，跑一遍。取前三名。这样，第四五六组都可以被排除了。

这样共需要比赛1次。可以筛选出第一组的1-3名，第二组的1-2名，第三组的1名。

（3）总决赛：把再筛得到的6辆车放在一起，比一次，即可取出前三名。

因此，共需6+1+1=8次。

然后，问题就来了。总决赛的第一名已经在第二次赛完就已经得出来了。为什么第三次还需要它参赛呢？实际上，他是不需要参赛的。因此我们有了拓展问题，

6个跑道，8次比赛机会，最多可以从多少辆车中选出前3？

至少37辆车是肯定能选出来的。

两轮比赛，第一名直接晋级。最后剩下的那辆车代替第一名参加决赛。决赛取前两名晋级。

2、一个人要过一座80米的桥，每走一米需要吃一颗豆子，他最多可以装60颗豆子，问最少需要吃多少颗豆子才能走完桥？证明一下为什么你给的答案是最少的？桥长81米呢？当桥长n米，最多装m颗的时候结果用公式怎么表示？

**答案：**1.由条件1得一趟走不完，由条件3我们可以考虑折回尽量少的次数能否走完。

　　先考虑只折回一次，那么此时就要求得折回点的位置。

　　2.由条件3得折回的距离尽量短，那么可以考虑最短的折回距离，很明显应该是20米处；设折回点距离起始位置x米，则此时x=20;那么在x处，需要至少有60颗豆子，一次性拿60颗豆子走完最后一程。

　　那么问题就来了，怎么在x处囤积60颗豆子呢，假使他第一次信心满满的拿着60颗豆子，走到x处就只剩下40颗了，现在不够走完剩余的全程啊，只能放下一部分折回去再取了，那放多少呢，很明显由条件2得我们需要放下20颗，拿着20颗刚好可以折回到起点，继续拿着60颗豆子第二次来到x处，此时手上还有40颗，再捡起第一次放的20颗，共60颗豆子刚好可以走完全程。

在这种情况下，我们易知，他共走了（20）\*3+（60）= 120米，故吃掉120颗豆子。

**如果桥长81米呢？**

　　此时我们如果还是只折回一次的话，在21米处还是得有60颗豆子，也就是需要搬运60颗豆子到21米处；分析得不可能只折回一次就搬运60颗豆子到21米处，因此需要再设置一个折返点y，此时x=21，0<y<x；

　　刚才考虑了80米的情况，那么我们可以假设现在在80米的基础上加了1米，因此可以设y=1，即先折回两趟到1米处，这样y处就有60\*3-5\*(1)=175颗豆子，再从y到x折返1趟，x处就有60\*2-3\*(20)=60颗豆子，最后从x出发拿着60颗豆子就可以愉快的走过桥了。

　　因此他共走过5\*(1)+3\*(20)+(60) = 125米，故吃掉125颗豆子。

3、一个绳子烧完需要1个小时，假设所有绳子的材质都不一样，也不均匀，怎么取出1小时加 15分钟。

**答案：**一根绳子从两头烧，烧完就是半个小时。  一根要一头烧，一根从两头烧，两头烧完的时候（30分），将剩下的一根另一端点着，烧尽就是45分钟。再从两头点燃第三根，烧尽就是1时15分。

4、有100个黑球，100个白球。两个桶，桶的容量无限，每个球都可以任意放在任何一个桶中，没有限制，请设计一种分配方法，使得白黑球分配到两个桶之后, 某个人从某个桶中取出的球是白球的概率最大化。（这个人去第一个桶取球的概率是1/2,第二个桶也是1/2）

5、有1亿个货物，不能单个单个检测，只能通过两两对比来找出其中的次品，请设计一个算法来找出次品。

**答案：**由于每次天平能够提供的信息为左边重，右边重，一样重三个信息。  
 那么每次能够提供的信息即为3。合理的利用下，m次提供的信息为3^m个。  
而n个产品每个个体的信息为1，即总共需要n个信息。但就题主提的问题里还有一个信息就是次品是比其他轻还是重，由于这一个信息带来的实际是每一个产品的轻重属性，因此相当于把整个产品数量乘以2的。最后的信息量为2n。  
 那么要求就是提供的信息至少高于需求的信息量，即3^m≥2n。  
本题中m为6，那么n最大为不超过3^6/2的整数，为[729/2]即364个

6、现有8个螺丝零件,其中有一个略轻的次品,你能用天平(不用砝码)经过两次称量找出次品吗,写出你的算法 3^2≥8

**答案**：1、取6个螺丝,天平两边各放3个.  
2、a天平平衡：则测量剩下的2个,可判断重量轻的；  
b、天平不平衡：在较轻的一侧3个中,任取2个,分别放到天平两侧.入平衡,剩下的1个为次品；若不平衡,较轻的一侧为次品.

1. 有25匹马 ，5个跑道，一次只能比5匹马，得到跑得最快的前3，至少需要比几次？ **答案：**7次
2. 有3盏灯，房间外有3个开关，你只有1次机会进入房间，怎么判断哪个开关对应哪盏灯？

**答案：**先打开第一个开关，开几分钟，然后再开第二个开关，这个时候可以进屋了，现在所看见正在亮着的灯，很明显是第二开关控制的，然后用手摸一下另外两个灯，有热度的，就是第二个开关控制的，剩下的当然就第三个开关控制的。

9、农民分金条问题

题目：你让农民为你工作7天，给他的回报是一根金条。金条平分成相连的7段，你必须在每天结束时给他们一段金条，如果只许你两次把金条弄断，你如何给你的工人付费，保证该农民在七天中任意一天结束时都可以领到相应的报酬。（例：第一天结束时他可领到1/7，第三天结束时他可领到3/7）

**答案：**将7份分成1/7，2/7，4/7.第一天结束时给1/7，第二天结束时给2/7，然后将农民手上的1/7要过来。第三天结束时将1/7给农民，第四天将4/7给农民，再将农民手里的3/7要过来，第五天再给1/7，第六天...

10、两柱香问题

题目：有两柱不均匀的香，每柱香燃烧完需要1个小时，问：怎样用两柱香切出一个15分钟的时间段？这个题的重点就是怎么切

**答案：**将甲香的一头点着，将乙香的两头点着，当乙香燃烧完时，说明已经过了半个小时，同时也说明甲香也正好燃烧了一半，此时，将甲香的另一头点着，从此时起到甲香完全烧完，正好15分钟。

11、瞎子翻牌

题目：给一个瞎子52张扑克牌，并告诉他里面恰好有10张牌是正面朝上的。要求这个瞎子把牌分成两堆，使得每堆牌里正面朝上的牌的张数一样多。瞎子应该怎么做？  
（瞎子摸不出牌是正面或者是反面，但是却可以随意翻动每一张牌）

**答案**：将52张牌分为2堆，一堆10张，另一堆42张，将10张的那一堆全部翻起来就行了。

12、100个苹果  
桌上有100个苹果，你和另一个人一起拿，一人一次，每次拿的数量大于等于1小于等于5，问：如何拿能保证最后一个苹果由你来拿？

**答案：**只需要你先拿，第一次拿4个，以后看对方拿的个数，根据对方拿的个数，保证每轮对方和你拿的加起来是6就行了，其实就是保证你拿到4，还要拿到10,16...直到94。

13：10斤酒两个桶

有三口酒缸，分别能装3斤；7斤；10斤。现在10斤的缸装满了酒，在没有称得情况下，怎么把这10斤酒平均分成两个5斤。

**答案**：

第一步，用10斤的先倒入3斤的，将3斤的装满，将3斤的倒入7斤的，再将10斤的缸子中的7斤倒入3斤的装满，将3斤的再倒入7斤的，最后再将10中剩下的4斤倒入3斤的缸子，此时，三个缸子的状态为，**10斤中有1斤，7斤的缸子中有6斤，3斤的缸子中有3斤。编程163**

第二步，用3斤的将7斤的装满，状态为：10斤中有1斤，7斤中有7斤，3斤中有2斤。

第三步，将7斤的缸子里的酒全部倒入10斤的缸子，状态：10斤的有8斤，7斤的有0斤，3斤的有2斤。

第四步，将3斤的倒入7斤的。状态为：10斤的有8斤，7斤的有2斤，3斤的有0斤。

第五步，用10斤的缸子将3斤的缸子装满，10斤的缸子中正好剩余5（8-3）斤，将3斤缸子里的倒入7斤缸子里，也正好5斤。正好实现

14：两个罐子+红球+蓝球问题

你有两个罐子，50个红色弹球，50个蓝色弹球，随机选出一个罐子，随机选取出一个弹球放入罐子，怎么给红色弹球最大的选中机会？在你的计划中，得到红球的准确几率是多少？

**答案：**一个罐子放一个红球,另一个罐子放49个红球和50个蓝球,概率接近75%.   
这是所能达到的最大概率了。

15：四个罐子问题

你有四个装药丸的罐子，每个药丸都有一定的重量，被污染的药丸是没被污染的重量＋1.只称量一次，如何判断哪个罐子的药被污染了？  
**答案**：给四个罐子编上号，1、2、3、4。然后1号拿一个，2号拿2个，3号拿3个，4号拿4个，称一下，若是都没被污染，应该重10个重量，若是11个重量就是1号罐，12就是2号罐，13就是3号罐，14就是4号罐；

16：果冻问题

你有一桶果冻，其中有黄色，绿色，红色三种，闭上眼睛，同时抓取两个果冻。抓取多少个就可以确定你肯定有两个同一颜色的果冻？  
**答案**：2次4个！

17：三人住旅馆

有三个人去住旅馆，住三间房，每一间房$10元，于是他们一共付给老板$30，第二天，老板觉得三间房只需要$25元就够了于是叫小弟退回$5给三位客人，谁知小弟贪心,只退回每人$1，自己偷偷拿了$2，这样一来便等于那三位客人每人各花了九元，于是三个人一共花了$27，再加上小弟独吞了不$2，总共是$29。可是当初他们三个人一共付出$30那么还有$1呢？  
**答案**：他们所消费的27元里已经包括小弟贪污的2元了，再加退还的3元=30元。这种题一定不要乱了阵脚，根据一条思路做：这30元现在的分布是：老板拿25元，伙计拿2元，三人各拿1元，正好！

18：蛋糕切8份问题

请把一盒蛋糕切成8份，分给8个人，但蛋糕盒里还必须留有一份。

**解答**：面对这样的怪题，有些应聘者绞尽脑汁也无法分成；而有些应聘者却感到此题实际很简单，把切成的8份蛋糕先拿出7份分给7人，剩下的1份连蛋糕盒一起分给第8个人。

19：拿最大钻石问题

一楼到十楼的每层电梯门口都放着一颗钻石，钻石大小不一。你乘坐电梯从一楼到十楼，每层楼电梯门都会打开一次，只能拿一次钻石，问怎样才能拿到最大的一颗？

**解答**：选择前五层楼都不拿，观察各层钻石的大小，做到心中有数。后五层楼再选择，选择大小接近前五层楼出现过最大钻石大小的钻石。

20：拿手电过桥问题

U2合唱团在17分钟内得赶到演唱会场，途中必需跨过一座桥，四个人从桥的同一端出发，你得帮助他们到达另一端，天色很暗，而他们只有一只手电筒。一次同时最多可以有两人一起过桥，而过桥的时候必须持有手电筒，所以就得有人把手电筒带来带去，来回桥两端。手电筒是不能用丢的方式来传递的。四个人的步行速度各不同，若两人同行则以较慢者的速度为准。Bono需花1分钟过桥，Edge需花2分钟过桥，Adam需花5分钟过桥，Larry需花10分钟过桥。他们要如何在17分钟内过桥呢？

**解答：**

1和2先过去：2  
1回来：1   
5和10 过去：10  
2 回来：2  
1和2过去： 2  
共：17

21：两个桶称出准确的水

如果你有无穷多的水，一个3L的和一个5L的提桶，你如何准确称出4L的水？

**解答**：

Ａ、5L装满，倒入3L中，此时桶5有2L，并做个记号；B、将桶3倒掉，并将桶5 中的2L倒入桶3；Ｃ、向桶5倒入2L，并将桶3中的2L倒入到桶5，于是并构成了4L

22：诚实和说谎的连个人

一个岔路口分别通向诚实国和说谎国。来了两个人，已知一个是诚实国的，另一个是说谎国的。诚实国永远说实话，说谎国永远说谎话。现在你要去说谎国，但不知道应该走哪条路，需要问这两个人。请问应该怎么问?

**解答**:问其中一人：另外一个人会说哪一条路是通往诚实国的?回答者所指的那条路必然是通往说谎国的。

23：12个球一个天平问题

12个球一个天平，现知道只有一个和其它的重量不同，问怎样称才能用三次就找到那个球。13个呢?(注意此题并未说明那个球的重量是轻是重，所以需要仔细考虑)

**解答**：

12个球。 第一次：4，4

如果平了：那么剩下的球中取3放左边,取3个好球放右边，称：如果左边重，那么取两个球称一下，哪个重哪个是次品，平的话第三个重，是次品，轻的话同理如果平了，那么剩下一个次品，还可根据需要称出次品比正品轻或者重

如果不平： 那么不妨设左边重右边轻，为了便于说明，将左边4颗称为重球，右边4颗称为轻球，剩下4颗称为好球；取重球2颗，轻球2颗放在左侧，右侧放3颗好球和一颗轻球；**如果左**边重称那两颗重球，重的一个次品，平的话右边轻球次品；**如果右**边重称左边两颗轻球，轻的一个次品；如果平称剩下两颗重球，重的一个次品，**平的话**剩下那颗轻球次品

13个球。 第一次：4，4，如果平了剩5颗球用上面的方法仍旧能找出次品，只是不能知道次品是重是轻如果不平，同上

24：飞机加油问题

每个飞机只有一个油箱， 飞机之间可以相互加油（注意是相互，没有加油机）   一箱油可供一  架飞机绕地球飞半圈。为使至少一架飞机绕地球一圈回到起飞时的飞机场，至少需要出动几架飞机？（所有飞机从同一机场起飞，而且必须安全返回机场，不允许中途降落，中间没有飞机场）   
**解答**： 猜想 验证  
    猜想：  
    **至 少需要出动5 架飞机**。思路是这样的，一架飞机要想完成绕地球一周的飞行，**至少需要别的飞机给它提供1 箱油。**最划算的办法显然是，派飞机和它结伴飞行前四分之一周以及后四分之一周，（因为这两段路程距离基地近所花代价小。）由它独立飞行中间的半程。必须保 证两个加油点，前四分之一处，加满，后四分之一点，及时补充。那么必须有两架飞机与目标机结伴飞行四分之一周，这两架飞机需要做折返飞行，正好花费2 箱油。所以补充油的任务实际上该由另外两架飞机完成。这两架飞机飞八分之一周，做折返飞，正好富余1 箱油。因此，5 架飞机刚好完成任务。到了此时，问题只考虑了一半。能够提供多少油并不意味着就能够全部接受，受到结伴飞行的距离，即腾 出的油箱空间所限制。而以下做法正 好可以满足此条件。  
    验证:  
    3架飞机同时从机场出发，飞行八分之一周，各耗油四分之一。此时某架飞机给其余两架补满油，自己返回基地。 另一机和目标机结伴，飞至四分之一周，给目标机补满油，自己返回。目标机独自飞行半周，与从基地反向出发的一机相遇，2   机将油平分，飞至最后八分之一 处，与从基地反向出发的另一机相遇，各分四分之一油，返回。

25：汽车加油问题

一辆载油500升的汽车从A开往1000公里外的B，已知汽车每公里耗油量为1升，A处有无穷多的油，其他任何地点都没有油，但该车可以在任何地点存放油以备中转，问从A到B最少需要多少油

**解答**：严格证明该模型最优比较麻烦，但确实可证，大胆猜想是解题关键。题目可归结为求数列an=500/(2n 1)   n=0,1,2,3......的和Sn什么时候大于等于1000,解得n>6当n=6时，S6=977.57,所以第一个中转点离起始位置距离为1000-977.57=22.43公里.所以第一次中转之前共耗油22.43\*(2\*7 1)=336.50升此后每次中转耗油500升,所以总耗油量为7\*500 336.50=3836.50升。

26：猴子吃香蕉问题

一个小猴子边上有100 根香蕉，它要走过50 米才能到家，每次它最多搬50 根香蕉，每走1 米就要吃掉一根，请问它最多能把多少根香蕉搬到家里。   
**解答**：设 小猴从0 走到50, 到A 点时候他可以直接抱香蕉回家了, 可是到A 点时候他至少消耗了3A 的香蕉( 到A, 回0, 到A), 一个限制就是小猴只能抱50 只香蕉, **那么在A 点小猴最多49 只香蕉.100-3A=49**, 所以A=17.   这样折腾完到家的时候香蕉剩100-3A-(50-A)=50-2A=16.

27：拿硬币问题（与问题4类似）

16 个硬币，A 和B 轮流拿走一些，每次拿走的个数只能是1 ，2 ，4 中的一个数。谁最后拿硬币谁输。问：A 或B 有无策略保证自己赢？   
   博弈类问题，分清两概念   
   必胜态：有一种方法导致下一状态为必败态   
   必败态：每一种方法导致下一状态为必胜态   
  **解决**办法：递推   
  1: 必败   
  2: 必胜：取1 ，导致变为1 状态( 必败)  
  3: 必胜：取2-> 必败态   
  4: 必败：取1 或2 或4 均导致必败态或直接失败   
   以些类推知16 为必败态，即后手必胜   
  剩2 个时, 取1 个必胜;  
  剩3 个时, 取2 个必胜;  
  剩4 个时, 如果对手足够聪明则必败;  
  剩5 个时, 去1 个必胜...  
  记作 2(1) 3(2) 4(x) 5(1) 6(2) 7(x) 8(1) ...  
  从中找出规律:  
  当剩余个数K=3N-2,N 为自然数时, 只要对手足够聪明则必败.  
  当K=3N-1 时, 有必胜策略:   取1 个;  
  当K=3N 时, 有必胜策略: 取2 个;  
  所以, 当16 个时, 后取者有必胜策略.

28：平均分问题（与问题5类似）

有三个酒杯，其中两个大酒杯每个可以装8两酒，一个可以装3两酒。现在两个大酒杯都装满了酒，只用这三个杯子怎么把酒平均的分给4个人喝？   
**解答**：用 一个三位数表示三个杯，880 ，前两个为8 升的杯最后一个3 升。开始：880\_853A 喝掉3 升变为：850\_823\_B 喝掉2 升为：803\_830\_533\_560\_263\_281A 喝掉1 升(A 已经喝4 升完毕）为：280\_253\_550\_523\_820\_802\_703\_730\_433\_460\_163\_181CD 各喝一升为：080\_053\_350\_323CD 各喝3 升B 喝2 升，分水结束，ABCD 四人各喝4 升。

作者：NgShawn  
链接：[https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2](https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2" \t "_blank)  
来源：牛客网

## 大数据

* 100亿数字，怎么统计前100大的？
* 10亿个url，每个url大小小于56B，要求去重，内存4G。
* 1KW句子算相似度（还是那套分块+hash/建索引，但是因为本人不是做这个的，文本处理根本说一片空白，所以就不误导大家了），之后就是一直围绕大数据的题目不断深化。
* Q1：给定一个1T的单词文件，文件中每一行为一个单词，单词无序且有重复，当前有5台计算机。请问如何统计词频？
* Q2：每台计算机需要计算200G左右的文件，内存无法存放200G内容，那么如何统计这些文件的词频？
* Q3：如何将1T的文件均匀地分配给5台机器，且每台机器统计完词频生成的文件只需要拼接起来即可（即每台机器统计的单词不出现在其他机器中）
* 一个大文件A和一个小文件B，里面存的是单词，要求出在文件B中但不在文件A中的单词。然后大文件A是无法直接存到内存中的。
* 一道题目是如果有一个人注册一个qq，如何保证这个qq号码和之前已存在的qq号码不重复呢？
* 扔硬币，连续出现两次正面即结束，问扔的次数期望
* 有100W个集合，每个集合中的word是同义词，同义词具有传递性， 比如集合1中有word a, 集合2中也有word a, 则集合1，2中所有词都是同义词，对这100W个集合进行归并，同义词都在一个集合当中。
* 有几个 G 的文本，每行记录了访问 ip 的 log ，如何快速统计 ip 出现次数最高的 10 个 ip，如果只用 linux 指令又该怎么解决；
* 海量数据的topk问题

作者：NgShawn  
链接：[https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2](https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2" \t "_blank)  
来源：牛客网

## 计算机基础

* Linux下的一些指令，$$（进程id），$?（上一条命令退出时状态），怎么查看进程，按照内存大小，CPU占用排序等等。
* Linux的命令：pwd、ln、which
* Linux线程通信
* hash表是怎么实现的？有冲突的时候怎么处理？
* linux 文件词频统计
* 介绍一下hash，怎么解决冲突。
* 你说一下hashmap的原理
* 内存泄露出现原因。
* 悲观锁乐观锁
* 把两个表按id合并怎么搞？
* 数据库transaction
* 浅拷贝深拷贝
* 第二题是两题 sql ，涉及 join,group by,max,min,sum,count 等操作的结合，以及同个题目多种写法。
* 线程安全是什么意思？新线程什么情况下会影响原有线程？
* 网络基础TCP三次握手
* 计算机网络：描述他发一句hello world到我这边显示，中间经历了哪些过程，我从应用层开始一层层往下分析答的，主要说http和tcp，网络层和链路层有些忘，但主要的几个协议和子网划分什么的也答了，面试官比较满意
* 词向量的推导，混合高斯，linux硬链接，三次握手，linux inode
* 进程线程的区别

## 概率题

* 100人坐飞机，第一个乘客在座位中随便选一个坐下，第100人正确坐到自己坐位的概率是？
* X是一个以p的概率产生1,1-p的概率产生0的随机变量，利用X产生1/2概率是0,1/2概率是1的随机变量。
* X，Y均服存于 [0,1] 的均匀分布，求X+Y。
* 一个国家重男轻女，只要生了女孩就继续生，直到生出男孩为止，问这个国家的男女比例？
* 一个有7个格子的环，三种颜色染色，相邻不能颜色重复，问多少种方案
* 一个袋子里有很多种颜色的球，其中抽红球的概率为1/4，现在有放回地抽10个球，其中7个球为红球的概率是多少？
* 一枚硬币，扔了一亿次都是正面朝上，再扔一次反面朝上的概率是多少？
* 一道概率题，54张牌，平均分成三堆，大小王在同一堆的概率？
* 一道概率题，一个六位的密码，由0~9组成，问你正过来看和倒过来看密码是一样的概率。
* 一道组合数学题。10盏灯，灭三盏，两头的必须亮着，不能灭掉相邻的两盏灯，问组合数？
* 三个硬币，分别是正正，反反，正反。随机抛一个硬币，结果是正面，问选的是那个硬币
* 个人玩游戏，100个球，每次挑5个，如何保证必胜。52张牌，四个人抽，黑桃A和红桃A同时在一个人手里的概率。
* 好像是问有70%的人喜欢玩游戏，30%的人不喜欢玩游戏，现在推送的资源必须是50%游戏，50%非游戏。问怎么分配比较合理。
* 有n个elements和1个Compare(A, B)函数，用Compare函数作为排序算法中的比较算子给elements排序。Compare函数有p的可能比较错。排序完取Top m个元素，本来就在Top m并被正确分在Top m的元素个数是x。问x的数学期望。
* 有两个随机数产生器，R1以0.7的概率产生1，以0.3的概率产生0，而R2以0.3的概率产生1，0.7的概率产生0.问如何组合这两种产生器，使新得到的随机数产生器以0.5的概率产生1，0.5的概率产生0。随机数产生器可复用。
* 有两枚硬币A和B，A正面的概率为0.6，B正面的概率为0.5.现在扔了一枚硬币显示为正面，问：该枚硬币是A的概率是多少？
* 概率题：有种癌症，早期的治愈率为0.8，中期的治愈率为0.5，晚期的治愈率为0.2.若早期没治好就会转为中期，中期没治好就会变成晚期。现在有一个人被诊断为癌症早期，然后被治愈了，问他被误诊为癌症的概率是多少？
* 给一个函数，返回0和1，概率为p和1-p，请你实现一个函数，使得返回01概率一样。
* 给定一个分类器p，它有0.5的概率输出1，0.5的概率输出0。Q1：如何生成一个分类器使该分类器输出1的概率为0.25，输出0的概率为0.75？Q2：如何生成一个分类器使该分类器输出1的概率为0.3，输出0的概率为0.7？
* 问了一个概率题 54张牌，分成6份，每份9张牌，大小王在一起的概率

## HR常问问题

* 为什么不读博、对读博报以什么态度。
* 为什么选择百度，谷歌百度都给你offer你选哪个。
* 为什么选择跨专业学计算机？
* 为什么选择阿里
* 以后可能要学习很多新技术，你怎么看。
* 你平时喜欢做什么？看过哪些书？最近在看什么书？
* 你觉得最有挑战的项目是什么。
* 你觉得最难忘的事情是什么？
* 你认为你的优（缺）点是什么。
* 你还有什么想问的？
* 加班怎么看。
* 印象最深刻的事？
* 压力最大的情况是什么时候。
* 在面试过程中觉得自己那些当面有进步
* 场景分析题，有一个任务给你，要求一个月完成，但是以目前的能力一个月完成不了，现在你知道有一个同事擅长这部分工作，但是他有自己的活，帮助你就可能耽误他的进度，问你咋办。
* 大学令你觉得最不爽的事情是什么
* 如何学习的？
* 如何看待加班。
* 实习期间项目，在组内担任的角色，是否熟悉其他组员的工作。
* 家庭教育观念？
* 家里什么情况？独生子女？
* 将来的职业规划？
* 工作地点
* 工作地点的问题
* 平时有什么兴趣爱好。
* 我觉得我会先去专心钻研技术，到达一定的
* 最后问了一下我兴趣爱好
* 有什么问题问我。
* 有没其他offer
* 有没有想过去创业公司
* 现在在哪里实习？实习主要做些什么？
* 简单介绍一下自己
* 聊聊offer情况，有什么考虑之类的。
* 聊聊实验室生活。
* 能不能来北京
* 自己有什么优点缺点？
* 自己本科生和研究生相比有哪些进步
* 要求用两个字评价大学生涯。
* 讲一下你觉得你突出的地方，有亮点的地方。
* 评价一下你自己的优点缺点？
* 详细介绍项目。
* 说下你的优缺点
* 说说你的经历。
* 说说你自己的性格。
* 说说研究生阶段最有成就的事，遇到问题具体怎么解决的。
* 请你说一下你对应聘该岗位的优势。
* 遇到的最大挫折是什么。
* 问你的职业规划，遇到挑战怎么处理，有没有之前和同事发生过较大分歧。

作者：NgShawn  
链接：[https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2](https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2" \t "_blank)  
来源：牛客网

## 开放题

* 2016年每个项目有个上线和下线时间段，统计每天在线的项目数量
* 一堆问题和答案的pair，算它们的相关性
* 一面现场面，自我介绍加挑一个项目细讲，还有场景题，第一题是QQ添加好友按名称搜索时，怎么区别广告号，诈骗号；
* 为什么之前没有深度网络出现（数据量不够+机器性能）
* 为今日头条设计一个热门评论系统，支持实时更新。
* 从项目中在哪一方面体会最深。
* 假设一个文档，连续的K个词，认为是一个时间窗口，一个时间窗口的词有关系，如何得到所有的时间窗口。
* 假设你拥有一切搜索数据，问怎么在不同场景下进行推荐，具体场景忘了（核心点：共线性、语义相似度、主题聚类等等）
* 假设有100W个单词，如何存储（我答的是trie树，面试官问每个节点会有很多子节点，每个子节点是一个指针，占用8个字节，如何节省空间，我说不知道，面试官提示双数组trie树）
* 假设要对一场nba球赛进行自动解说，会遇到哪些困难，又该怎么解决呢？
* 做过哪些项目？项目中遇到哪些难点，你是怎样解决的？
* 关于集群调度的一些经验 trick 掌握多少；
* 分词时，为了提高效率，怎么存储词典？（键树）如何压缩存储？
* 在微信的场景下，如何判断用户的职业？开放问题
* 场景题如何鉴别淘宝上卖假货的商家，价格维度可以用什么策略等
* 如何做一个新闻推荐
* 如何在语料中寻找频繁出现的字串，分析复杂度。
* 如何用尽可能少的样本训练模型同时又保证模型的性能；
* 如何预测双十一支付宝的负载峰值。
* 对推荐算法的未来看法。
* 平面上有n个点，让你设计一个数据结构，能够返回这个这n个点中距离某特定点最近的一个点。一开始讲了下kd树，然而太复杂面试官不满意，就讲了一个类似GeoHash的方案。
* 建立一个数据结构，基于此写一段程序用于存储sparse vector，同时编写一个函数实现两个sparse vector的相加运算
* 很多单词，如何计算单词之间的相似度（或者对单词进行分类）
* 怎么预测降雨量。
* 我只有一大批实体词， 如何对他们进行聚类（无监督聚类）， 如何找出这些词中， 哪些词之间有关系， 是强关系还是弱关系， 具体是什么关系，（如刘德华和朱丽倩 属于娱乐分类， 是强关系， 关系为夫妻）
* 拼车软件是如何定价的以及如何优化。
* 推荐算法（基于用户的协同过滤，基于内容的协同过滤）
* 推荐系统的冷启动问题如何解决
* 文本挖掘中，分词算法？如何选取特征？如何进行相似度计算，文本聚类结果如何评估？
* 无给定条件，预测蔬菜价格。
* 有100W个集合，每个集合中有一些词，对于每个集合，找出他是哪些集合的真子集。
* 有一堆已经分好的词，如何去发现新的词？
* 比赛相关问题提特征特征选择等
* 海量的 item 算文本相似度的优化方法；
* 特征工程经验。
* 用两分钟介绍自己的项目，创新点在哪里。
* 用户给三个item（query），如何给出查询网页。
* 第三题是如何鉴别实施诈骗的QQ用户；
* 第二题是微信朋友圈内容的安全鉴别；
* 第四题是如何做反作弊，例如公众号的刷阅读量。
* 系统设计题，给一个query，如何快速从10亿个query中找出和它最相似的 （面试官说可以对每个query找1000个最相似的，存起来，每天离线更新）
* 线性代数：特征线性依赖，出现冗余，会导致什么问题？
* 给一堆数据找找到最佳拟合的直线，数据有较多噪声
* 给你一个系统（面试官好像是无人车部门的），后台的逻辑已经实现了，但是前端加载很慢，怎么检测。
* 给你两个文件a和b，大小大概100M，两个文件每行一个整数，要求找到两个文件中相同的整数，存到文件c里，问我怎样尽快的完成这项工作？
* 给出一个算法实现如何确定快递邮件上的地址，要求从国家到省市到县到乡镇的一个识别，要求效率高（有陷阱，比如有的人把县写到市的前面，有人喜欢写地域名称的省略词比如安徽省写成安徽或者皖）。
* 给定淘宝上同类目同价格范围的两个商品A和B，如何利用淘宝已有的用户、商品数据、搜索数据、评论数据、用户行为数据等所有能拿到的数据进行建模，判断A和B统计平均性价比高低。统计平均性价比的衡量标准是大量曝光，购买者多则高。
* 给很多单词，统计某个子串出现次数，我给的方法还是用Trie，只不过一个单词要分成多个插入到Trie数中就行了。
* 给很多单词，要求统计出现某个前缀出现次数。
* 统计全球会弹钢琴的人数，我用机器学习的思路答的，面试官还比较满意
* 自己项目中有哪些可以迁移到其他领域的东西。
* 讲了讲自己在深度学习的认识，问的问题是几个具体场景的设计，包括怎么从海量数据中提取热点问题。
* 设计 LRU 系统
* 设计一个合理的电梯调度策略，调度两个电梯 ，考虑满足基本的接送需求，满足能耗最小，满足用户等待时间最短
* 设计一个系统可以实时统计任意ip在过去一个小时的访问量；
* 设计一个结构存取稀疏矩阵（面试官最后告诉我了一个极度压缩的存法，相同行或列存偏差，我当时没听懂，还不懂装懂，最后还是没记住）
* 设计实现一个git diff
* 说一下最能代表你技术水平的项目吧？
* 项目：具体问了特征怎么做的。
* （难到我了，我想的方法不好，面试告诉我了他的想法，类似于一个进程调度问题，每一时刻只可能有一个用户按按钮，把这条指令接收，判断当前电梯能否满足，能满足就执行，不能满足则放入一个队列里，实际情况还要细化）

作者：NgShawn  
链接：[https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2](https://www.nowcoder.com/discuss/33737?type=2&order=0&pos=20&page=2" \t "_blank)  
来源：牛客网

## 编程题

* 1~n这n个数现在去掉两个，如何找到去掉的两个数。 假设去掉的两个数是a和b，那么通过求和，平方和可以知道a+b和a^2+b^2，然后解方程就行了。
* char a[4] = {1, 2, 3, 4}; char \*b = a; b[0] = 100; 请问输出a的结果是什么？
* 一个 N\*M 的矩阵，从左上走到右下最小需要（N+M）步走完，问一共有多少种走法。
* 一个严格递增的数组，将前缀取一部分放在后面，在修改后的数组上找到最小的数。(剑指Offer原题)
* 一个大写字符串如ABABB（len<1000），代表游客进游乐场的顺序及从哪个入口进入，要求每个入口(不多于26个入口)从第一个游客直到该入口的最后一个游客，检票员都不能离开，问最少检票人数K。
* 一个字符数组中，每个字符都出现了3次，只有一个出现了2次，如果快速找出这个出现2次的？
* 一个字符矩阵，只可能是R,G,B三种字符。判断是否满足某个条件。这个条件是每种符号连成一个长方体，三个长方体长宽一致,且横着平行
* 一个广告，它有一个id，一个上线时间，一个下线时间，现在我有很多这样的广告，如果现在给你一个时间，告诉我有多少个广告在这个时间在线的
* 一个数据流中，如何采样得到100个数，保证采样得到的100个数是随机的？
* 一个数组中某个数出现次数大于一半，最快找出该数。
* 一个数组只有一个数字是单独出现，其他出现了三次。
* 一个数组存着1-1000连续的整数，假如我取出其中一个数，怎么能快速找到（用类二分查找）
* 一个数组存着负数与正数，将正数放在前面，负数放在后面
* 一个运算序列只有+、\*、数字，计算运算序列的结果。（Code）
* 一堆ip地址区间，不会重叠，来一个新的ip地址，看它在不在，在哪个区间。
* 一维数组，swap 其中的几对数字（每个数字只属于一次 swap 操作），实现查找（与二分有关）；
* 一维有序数组，经过循环位移后，最小的数出现在数列中间，如果原数组严格递增或递减，如何找这个最小数；
* 一维有序数组，经过循环位移后，最小的数出现在数列中间，如果原数组严格递增，如何找这个最小数。
* 一维有序数组，经过循环位移后，最小的数出现在数列中间，如果原数组非严格递增或递减，如何找这个最小数；
* 一维有序数组，经过循环位移后，最小的数出现在数列中间，数组可能是递增、递减、递减后递增、递增后递减四种情况，递增递减都是非严格的，如果有转折点，返回转折点的值，否则返回-1；
* 一道题：给定一个整数数组，里面有两个数相同，其他数都是不同的，如何尽快找到这两个数（答，用hash表，O（N），有更好的方法么？）
* 一题是多位数用链表存储（ e.g. 123 用 1->2->3 存储），实现相加功能函数
* 不创建临时产量换两个数
* 两个同样大小有序数组求中位数，写代码
* 两个大整数相乘。（Code）
* 两棵树相加——对应位置两棵树都有值则相加，对应位置只有一棵树有值则取该值；
* 中序遍历二叉树，利用O(1)空间统计遍历的每个节点的层次。（Bug Free Code）
* 中缀表达式转逆波兰表达式，逆波兰表达式求值；
* 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。
* 二分查找
* 二分查找，查找target，在区间[start，end]之间，如果有重复元素，返回最后一个下标，其他情况返回-1
* 二叉树前序递归遍历算法（手写代码）
* 二叉树的前中后遍历
* 二叉树的文件存储，也就是序列化。
* 二叉树遍历，描述下层序遍历。
* 二维数组，每行递增，每列递增，任意交换其中的两数，发现并恢复。
* 二维数组，每行递增，每列递增，实现查找。
* 二维数组，每行递增，每列递增，求第k大的数。
* 什么样的数据结构可以满足多次插入删除，取最小数，给出时间复杂度。
* 介绍二叉树前序遍历非递归遍历算法（手写代码）
* 介绍大顶堆和小顶堆
* 从一组数中找出和为sum的三个数（leetcode原题，先sort再找，并且剪枝），写代码，四个数呢？说思路。
* 假设有个M\*N的方格，从最左下方开始往最右上方走，每次只能往右或者往上，问有多少种走法，假设中间有若干个格子不能走，又有多少种走法。
* 允许两个元素交换一次的最大连续子序列和。
* 全排列
* 全排列。
* 冒泡排序(手写代码)
* 写 find 函数，在目标串中匹配模式串（要考虑中文字符的情况）
* 写一个二叉树的非递归的后续遍历
* 写一个简单的正则匹配表达式(将文本中的123.4匹配出来)
* 写个动态规划，最长公共子序列
* 判断一个字符串是否为另外一个字符串旋转之后的字符串
* 前k大的数
* 单链表的翻转
* 去掉连续的重复数字，输出新数组，例如：1，2，2，2，1，3，5——> 3，5。
* 去除字符串S1中的字符使得最终的字符串S2不包含’ab’和’c’。（Code）
* 合法括号匹配
* 在一个字符串中，找出最长的无重复字符的字串
* 在二叉树结点结构中加一个指针域，使其指向层次遍历的下一个结点，特别地，每一层的最后一个结点为空。（Code）
* 堆排序(手写代码)
* 堆是怎么调整的。
* 复杂链表的复制。
* 如果给出一个二叉搜索树的后续能不能建立（可以，因为只要将遍历结果排序就可以得到中序结果）。
* 字符串反转（手写代码）
* 字符串移位，给出字符串abc##dfg##gh，实现将所有#移至字符串串头。输出####abcdfggh。
* 字符串转整数
* 字符串，给一个url，求中间的site
* 字符串，给一个url，求中间的site。
* 定义满足$n=x^a+y^b$（$x，y，a，b$是非负整数）的n是神奇数。如$4 = 2^0 + 3^1,17 = 2^3 + 3^2$。输入l和r，请求出闭区间$[l,r]$里，最长的一段不含有神奇数的连续区间长度。$x,y,l,r<=10^{18},x>=2,y>=2$，如$3\ 5\ 10\ 22$，在$[10,22]$区间内，$x=3,y=5$的条件下，区间内[14]是神奇数，所以最长的区间是$[15,22]$长度为$8$，如$2，3，1，10$，在$[1,10]$区间内，$x=2，y=3$的条件下，$2，3，4，5，7，9$都是神奇数，所以最长的区间只有长度$1$。
* 实现栈，使得 添加、删除、max 操作的复杂度为 O(1)。
* 对于一个字符串，请设计一个算法，只在字符串的单词间做逆序调整，也就是说，字符串由一些由空格分隔的部分组成，你需要将这些部分逆序。给定一个原字符串A和它的长度，请返回逆序后的字符串。
* 对于一个字符串，请设计一个算法，将字符串的长度为len的前缀平移到字符串的最后。
* 寻找字符串中第一个只出现一次的字符；
* 将字符串连续重复出现的字符删到只剩一个，这个可以用双指针，时间复杂度n，空间复杂度1。
* 常用排序算法的时间和空间复杂度
* 平衡二叉树是什么
* 归并排序(手写代码)
* 快速排序(手写代码)
* 快速排序+二分查找
* 手写快排(easy)
* 打印数组的组合数。
* 打印螺旋数组；
* 把一个bst转化成一个双向链表。
* 把一个字符串的大写字母放到字符串的后面，各个字符的相对位置不变，不能申请额外的空间。例如AbcDeFGhi ->bceiADFG
* 排序二叉树转双向链表。（Code）
* 描述Dijkstra最短路径算法
* 插入排序(手写代码)
* 数列中找第 k 大的数字（与快排或堆排序有关）
* 数据解压缩，3(a4(ab)) -> aababababaababababaabababab；
* 数组有只有一个数出现一次，其他数都出现三次，找出那个数。
* 旋转数组
* 最少时间复杂度求数组中第k大的数。（Code）
* 最短路径代码。
* 最长公共子串（动态规划有关）；
* 最长公共子序列
* 有一堆无向好友列表 1-2， 3-4, 2-3 之类的，问能不能把这些用户划分两组，组内都不互为好友。
* 有序数组寻找和为某数的一对数字；
* 正数数组，找三个数使积最小，问有多少种选择。
* 母鸡、公鸡和小鸡问题：公鸡五块一只，母鸡三块一只，小鸡一块三只，用100元买100只鸡的所有方法。
* 求double类型的二进制1的个数。
* 求二叉树最近公共祖先(leetcode原题)
* 求连续子数组最大乘积，还让考虑边界问题（最后问了：连乘有可能导致溢出，存不下了）
* 用一个队列，将每个二叉树的root先放入队列。
* 用数组实现队列，各操作的复杂度分析。
* 用速度不同的指针可以判断链表中是否有环，问两速度满足怎样的关系可以保证发现环。
* 直接插入排序写代码
* 看段代码，问输出是啥。(就是段求二进制中1的个数)
* 矩阵求最长连续递增的路径长度
* 矩阵求最长连续递增的路径长度。
* 第一题是链表倒数第 k 节点；第二题是二叉树打印路径，第三题是矩阵中将 0 元素所在行列全置 0 的最优空间解法
* 第二轮是写出一个算法输出二叉树的 s 序列，何为 s 序列，比如现在有个二叉树 1-2,3-4,5 6,7 这是一颗完全二叉树， S 序列输出就是按照 1237654 这个顺序输出，用两个栈就能实现比较简单。
* 算法题，也只记得一个了：存在一个数组，大小98，里面的元素均为在[1,100]，且无重复， 不申请额外空间的情况下，在时间复杂度为O(N)情况下，找出缺失的两个元素值。
* 给一个n\*n的矩阵，矩阵中满足每行每列都是递增的，要查找矩阵是否存在某个数.(leetcode原题)
* 给一个数组，只有一个元素出现了一次，其他都出现了两次，找出出现一次的数。
* 给一个数组，数组种存在一种数，它的左边都比它小，右边都比它大，找出所有这些数的位置。
* 给一个股票，n天的价格，只能两次买入卖出，而且只能只能先卖再买，问最多赚多少钱？
* 给一个股票，n天的价格，只能进行一次买入和卖出，问最多赚多少钱？
* 给一个股票，n天的价格，可以买入卖出k次，而且只能只能先卖再买，问最多赚多少钱？
* 给一个股票，n天的价格，可以无限次买入卖出，问最多赚多少钱？
* 给了一个链表，第1个结点标号为1，把链表中标号在M到N区间的部分反转。
* 给你一个无重复的数组输出全排列。
* 给你一颗二叉树按层输出每一层输出后都换行
* 给出一个二维矩阵，从（0,0）出发走到右下角，只能向右或向下走，找到一条路径，是这条路径上的总和最大。
* 给出一段代码问代码作用（二进制数据1的个数）
* 给出一颗二叉树，两个叶节点，找到这两个叶节点互连通的一条最短路径。
* 给定一个数组，只有一个元素出现了一次，其他都出现了3次，找出出现一次的数。
* 给定一个数组，有两个元素出现了一次，其他都出现了两次，找出两个出现一次的数。
* 给定一个正整数向量，判断这个向量是否存在一个片段，使得反转这个片段后能够使该向量升序排列。如：[1, 2, 4, 3]，就可以通过反转[4, 3]使得向量变为[1, 2, 3, 4]。
* 给定二叉树的先序跟后序遍历，能不能将二叉树重建（不能，因为先序：父节点-左节点-右节点，后序：左节点-右节点-父节点，两者的拓扑序列是一样的，所以无法建立）
* 给定循环递增数组 $a=[7,8,9,1,2,3]$和一个值$k=2$,返回该值得再数组中的下标。
* 给定数组A[]={1,4,7,...}和一个数T。求和为T的A中的数最少要几个。A中的数可复用。 我写了个递归，面试官不建议使用，因为效率不高。但没有反对。
* 给定数组，寻找 next big（堆排序有关）；
* 给我一个数组［1，2，5，10，20，50，100］，可以从里面取若干个数，要求得出和为100的不同取法有多少？（说出思路）
* 统计数列中的逆序对（归并排序有关）；
* 编程题：实现求正整数平方根整数部分的函数（使用梯度下降）
* 翻转二叉树（Code）
* 若干个二叉树，如何按照层序遍历
* 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。
* 输入一个大长方形，长宽ab，和一堆小长方形。选择两个小长方形，它能放进大长方形，而这个小长方形面积和最大

输入一个宿舍楼亮灯描述图，计算把所有灯关掉的最短时间，管理员起点在左下角，只能在最左或最右的楼梯往上一层，不可往下一层。每次往上一层花费1分钟，每次往左或往右一间宿舍花费1分钟，关灯不花时间。输入的高<=15，宽<=100。  
如图:

0010

0100

从左下角开始，最短花费时间是先往右（关灯），再往左，再上一层，再往右两次（关灯）完成：5

再如：

001000

000010

000010

最短时间是先往右四次（关灯），往右一次，上一层，往左一次（关灯），往右一次，上一层，往左三次（关灯），完成，12

* 输入两个正数数组，在两个数组分别选一个数，要求第一个数组选的数的下标小于第二个数组选的数的下标。使得两个数的乘积最大。
* 输出字符串中的所有重复子串，例如：abcab，输出: a, b, ab
* 连续子数组最大和
* 迷宫的深度搜索、广度搜索；
* 选取任意数据结构实现添加、删除、随机返回三个功能，分析复杂度。
* 选择排序(手写代码)
* 链表上的快速排序。
* 长度为N的序列Sequence=abc......Z，问有多少不同的二叉树形态中序遍历是这个。（Code）
* 问了一两个算法题，记不清了，只记得其中一个是：找数组中2个出现两次的数字，还有3个两次的数字
* 问了一个1的平方加到100的平方结果
* 非常经典的0-1背包问题