

# 12，阿里巴巴面试题

2018年4月27日 17:01

- 关于面试
- 面（笔）试特点
- 一些例题
  - 概率（例1，2）
  - 智力题（例3）
  - 数据结构与算法（例4，5，6）
  - 实际系统（例7）
- 总结
- 关于面试
- 各个公司有没有自己的题库
  - 题库里的题目来源（共享）
    - 员工
    - 网络
- 笔试和面试
  - 笔试
    - 无交流，写思路，题型丰富
  - 面试
    - 注重交流，写完整代码
- 面（笔）试特点
- 注重基础知识——题型多
  - 操作系统：缓存、进程、线程、死锁
  - 网络：TCP 协议、握手
  - 数据结构与算法：排序、二分、树、实际系统
  - 智力题（数学）

- 概率
- 组合数学

## ● 一些例题

### ○ 概率（例1）

/\*

问题：有8只球队，采用抽签的方式随机配对，组成4场比赛。假设其中有3只强队，那么出现强强对话（任意两只强队相遇）的概率为多少（概率）

分析：

基本事件总数： $7 \times 5 \times 3 \times 1$ ,

没有强强联合的概率数为从5个弱选三个配对，剩下的两弱一队，可能数为 $c(5,3) \times A(3,3)$   
= 60

那么强强相遇=  $105 - 60$

\*/

### 例2：

/\*

问题：有两个包，甲包有8个红球和2个蓝球，乙包有2个红球和8个蓝球，先抛硬币决定从甲包还是乙包取球。

确定后，一共取了11次球（都是从甲or乙），每次取一个球并且放回，11次的结果是7次红球，4次蓝球，问选中甲包取球的概率？(概率)

分析：明显，这是一个后验概率问题，

如果从甲取球，取7红4蓝的概率为 $c(11,4) \times (8/10)^7 \times (2/10)^4$

如果从乙，那么概率为 $c(11,4) \times (2/10)^7 \times (8/10)^4$

然后甲的处以（甲+乙）

\*/

### ○ 智力题（例3）

/\*

问题：在黑板上写下50个数字：1至50。在接下来的49轮操作中，每次做如下操作：选取两个黑板上的数字a和b，擦去，  
在黑板上写 $|b-a|$ 。请问最后一次动作之后剩下的数字可能是什么？为什么？（智力题）

分析：我们先不考虑剩下的数为什么数，先考虑剩下奇数or偶数

这个过程中，如果我们选中两个奇数，那么奇数数目-2, 偶数数目+1,

如果1奇数1偶数，那么奇数数目不变，偶数-1

如果两偶，那么奇数数目不变，偶数-1.

分析知，要么奇数数目不变，要么减少2, 那么奇数数目的奇偶性是不变的  
1-50中有25个奇数，那么最后剩下的一个肯定为奇数

因为剩下的数为 $|b-a|$ 的绝对值，所以范围定在1-50,  
如果想得到1, 那么将1, 2拿出来，先不动，把34配对，换成1, 56, 换1, 49, 50, 换1. 然后后面共有24个1, 两两消去  
如果想得到20, 那么将21, 1, 拿出来，先不动，然后23换1, 45。。。.  
最后1-49都能得到

\*/

来自 <<http://tool.oschina.net/highlight>>

## ○ 数据结构与算法（例4，5，6）

### ○ 例4

/\*

问题：有无序的实数列 $V[N]$ ，求里面大小相邻的实数的差的最大值，要求线性空间和线性时间。（数据结构与算法）

分析：注意，这里不是相邻元素最大的差，而是相邻大小实数最大的差，相当于在数轴上排好序，

假设数组中最大的为max，最小的为min，并且两者不相等

桶排，分 $n+1$ 个桶，每个桶左闭右开，最后一个桶双开，然后每个桶大小为 $(\max-\min)/(n+1)$ 。

然后定有一个空桶，所以桶内元素之间无意义，我们要记录的是桶之间的，然后记录每个桶的最大or最小值，

这道题我肯定做过

\*/

来自 <<http://tool.oschina.net/highlight>>

○

### ○ 例5

/\*

问题：有一个函数`int getNum()`，每运行一次可以从一个数组 $V[N]$ 里面取出一个数， $N$ 未知，当数取完的时候，函数返回NULL。

现在要求写一个函数`int get()`，这个函数运行一次可以从 $V[N]$ 里随机取出一个数，而这个数必须是符合 $1/N$ 平均分布的，

也就是说 $V[N]$ 里面任意一个数都有 $1/N$ 的机会被取出，要求空间复杂度为 $O(1)$ 。（数据结构与算法）

分析：蓄水池抽样算法

每个数，以 $1/i$ 的概率选择它。这样每个被选择且不被替换的概率都是 $1/n$

\*/

来自 <<http://tool.oschina.net/highlight>>

## ○ 例6

/\*

问题：A[i]是一个严格递增的整数数组，其中所有的数字都不相等，请设计一种算法，求出其中所有的A[i]=i的数字并分析时间复杂度，  
不分析复杂度不得分。（数据结构与算法）

分析：令B[i] = A[i] - i;

那么B[i+1] - B[i] = A[i+1]-A[i]-1, 因为A[i]是严格递增，因此A[i+1]-A[i]一定等于等于1, 因此B[i+1]-B[i]>=0

也就是B是递增的，不过不是严格递增，我们要找A[i] = i, 这时我们只需二分查找B[i] = 0的点即可

并且B数组我们无需事先求，我们只需在二分A时各减去i即可

\*/

来自 <<http://tool.oschina.net/highlight>>

## ○ 实际系统（例7）

## ○ 例7

/\*

问题：述有一大批数据，百万级别的。数据项内容是：用户ID、科目ABC各自的成绩。其中用户ID为0~1000万之间，且是连续的，  
可以唯一标识一条记录。科目ABC成绩均在0~100之间。有两块磁盘，空间大小均为512M，内存空间64M。

- 1). 为实现快速查询某用户ID对应的各科成绩，问磁盘文件及内存该如何组织；
- 2). 改变题目条件，ID为0~10亿之间，且不连续。问磁盘文件及内存该如何组织；
- 3). 在问题2的基础上，增加一个需求。在查询各科成绩的同时，获取该用户的排名，问磁盘文件及内存该如何组织。（实际系统）

分析： 开放问题

- 1) ID连续，可以用数组存

如果不从0开始，可以减一个偏移量，相当于存3000000个int（3M）（因为虽然id在0-1000万，但只有百万数据），

需要空间大概12M，内存够用

- 2) ID不连续，且id范围为0-10亿，需要map或者hash，并且需要对ID取余做映射放入hash表。空间复杂度大概12M（id本身所需空间） + hash的键的空间，

按内存要求，至少可以有12000000个(4倍)key（48M），这样一共耗费60M。相当于将0-10亿id映射到0-12000000个int之间。

- 3) 需要对100万的数据作排序

取决于hash表的大小，如果内存不够，需要作外部归并排序

内存排小块，写入磁盘文件

文件不断作归并排序

得到rank值，还可以存入内存

优化思路：

各科成绩0-100，可以考虑压缩，不一定要存一个int（32bit，其实可以存 $2^{32}$  约10亿的数据），用char就可以。

排序策略不一定用快排，可以计数排序（总分是有范围的：0-300）

\*/

来自 <<http://tool.oschina.net/highlight>>

## ○ 总结

- 笔试注重基础
- 面试注重交流
  - 不要把面试当成笔试
  - 给面试官积极的情绪
  - 没有标准答案——开放问题
  - 多提假设，简化问题
- 函数头部要自己写出
- 无固定套路
- 多总结、思考、归纳