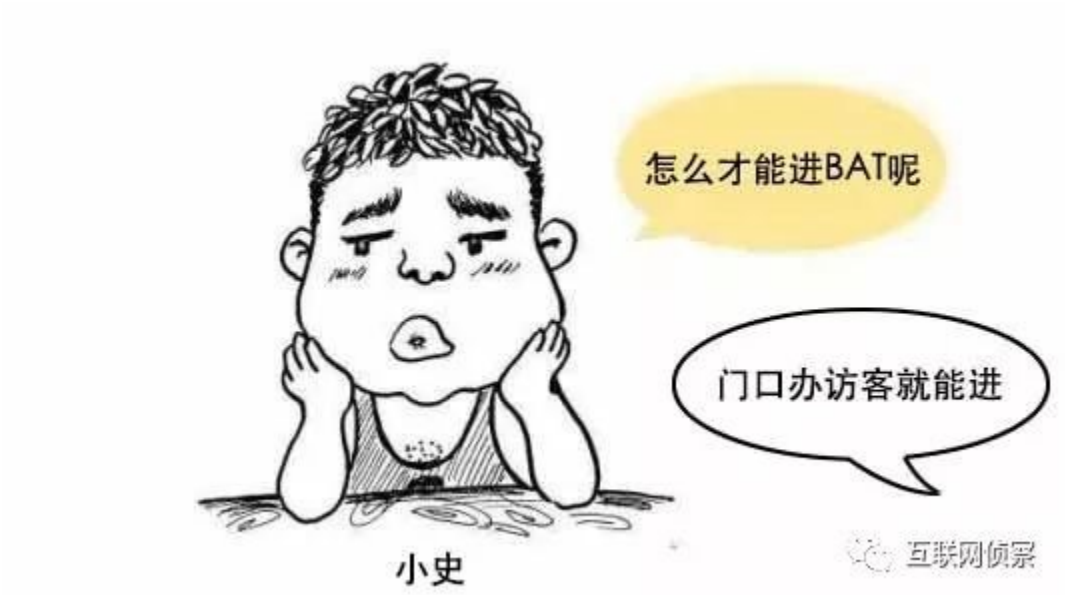
【面试现场】如何判断一个数是否在40亿个整数中？

channingbreeze 程序员乔戈里 1月31日

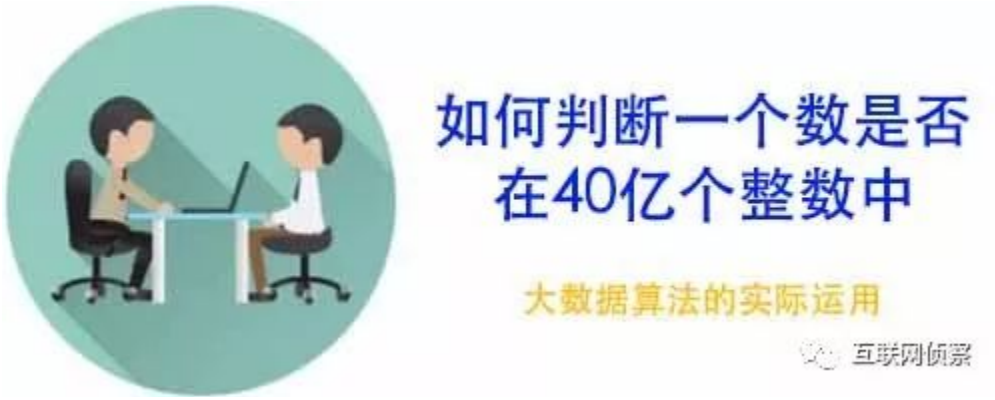
小史是一个应届生，虽然学的是电子专业，但是自己业余时间看了很多互联网与编程方面的书，一心想进BAT。

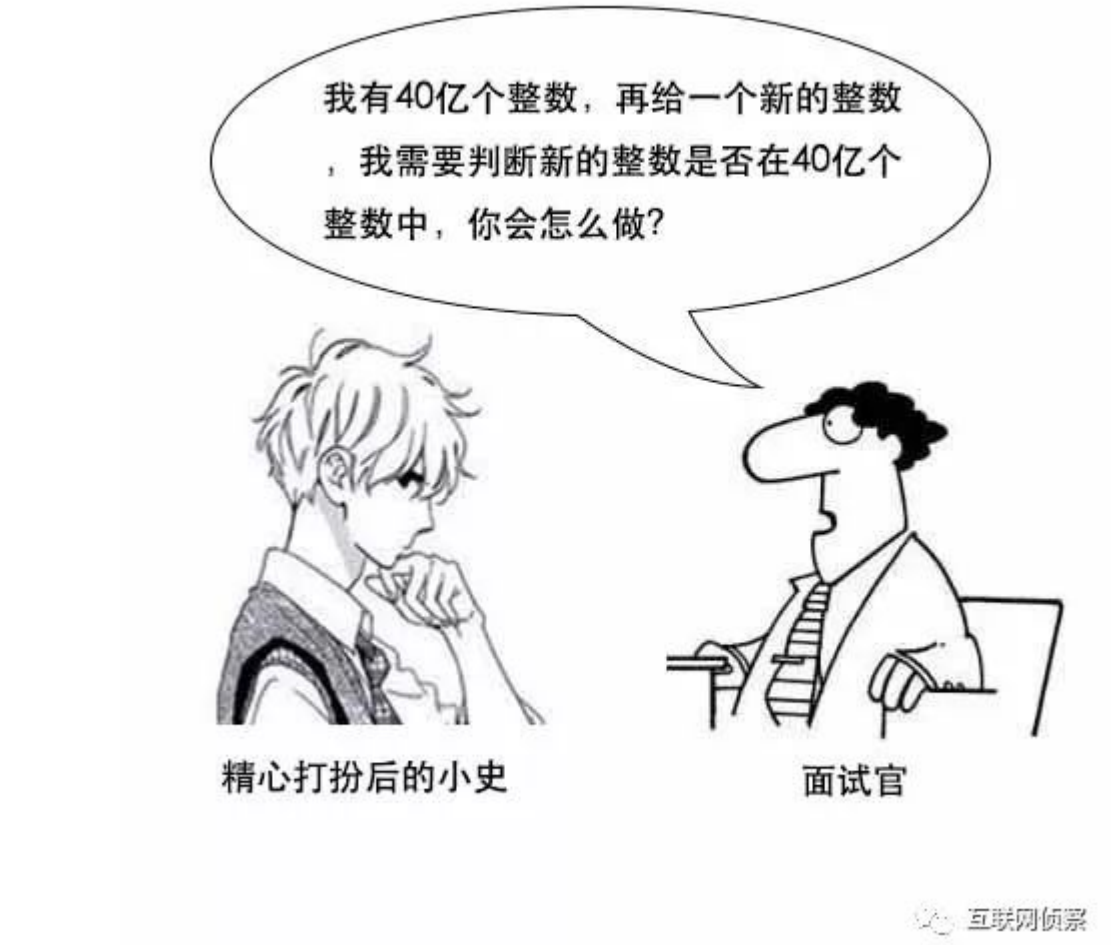


今天他就去BAT中的一家面试了。

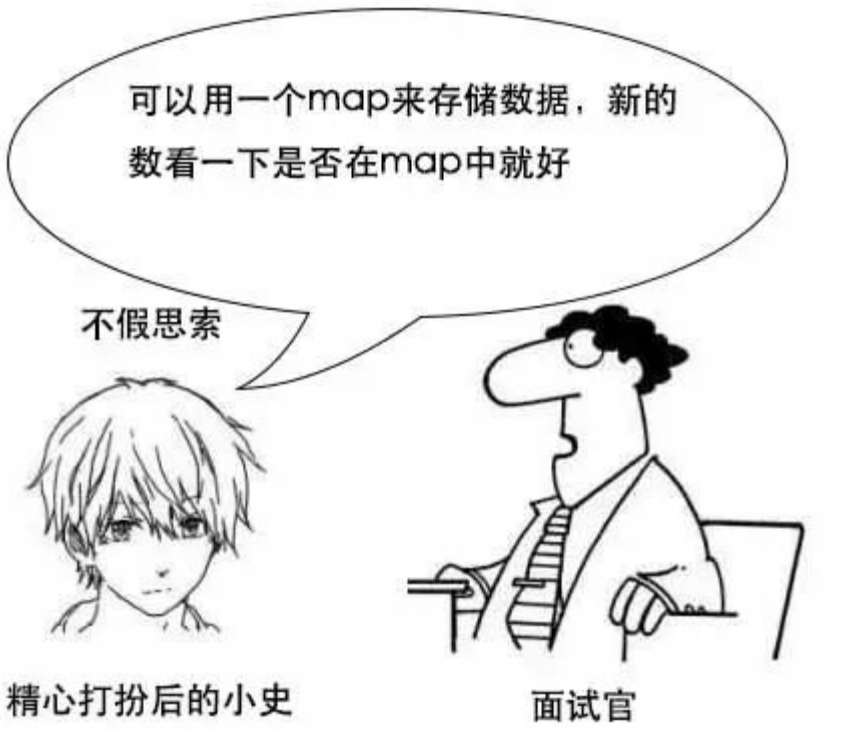
简单的自我介绍后，面试官给了小史一个问题。

【面试现场】

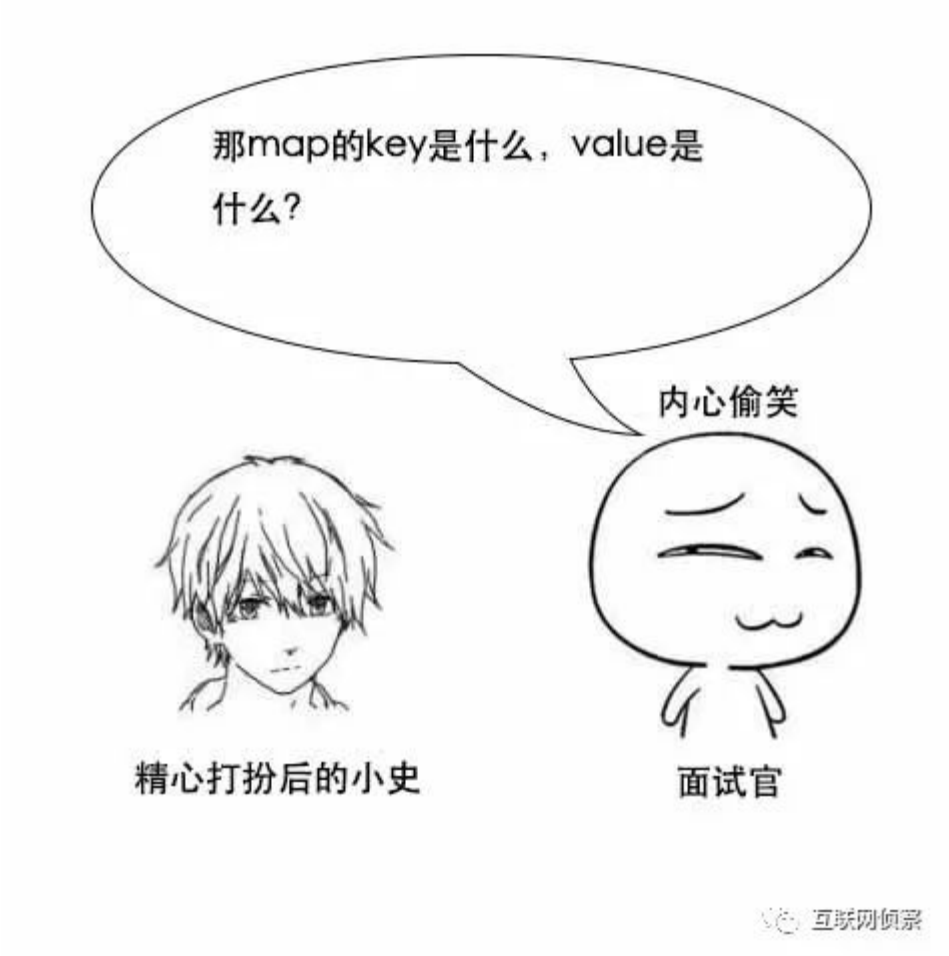




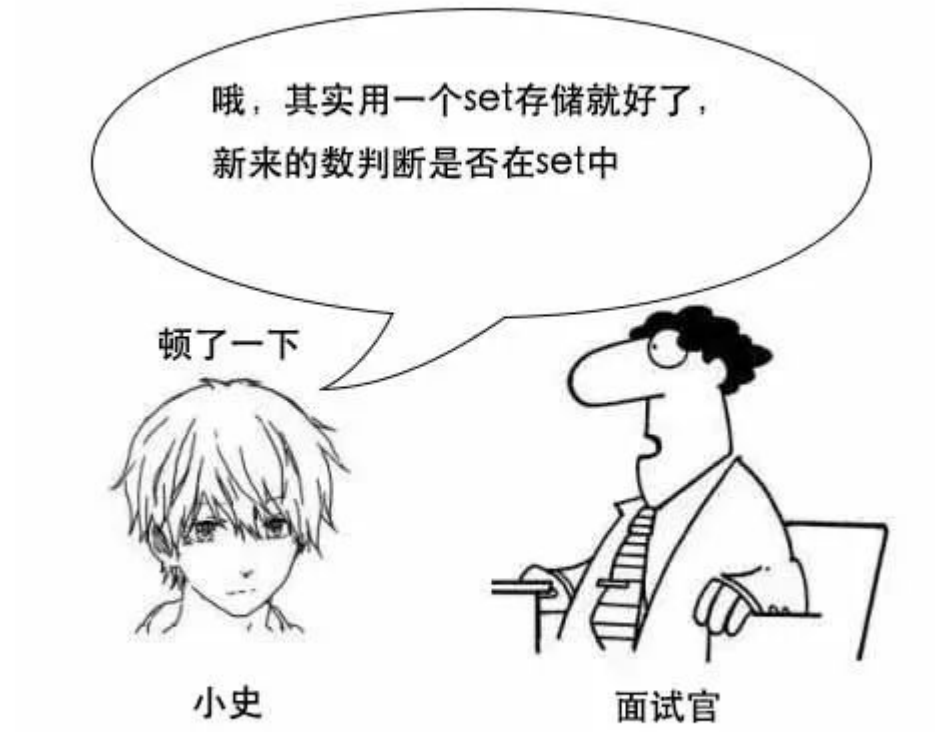
题目：我有40亿个整数，再给一个新的整数，我需要判断新的整数是否在40亿个整数中，你会怎么做?



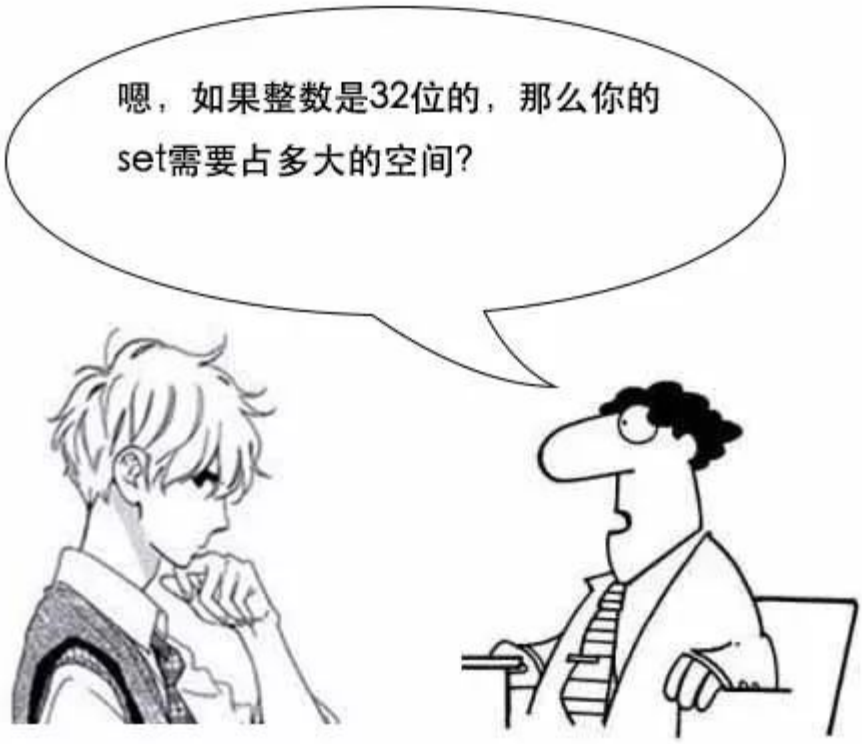
互联网侦察


那map的key是什么，value是什么?
内心偷笑
精心打扮后的小史
面试官

互联网侦察

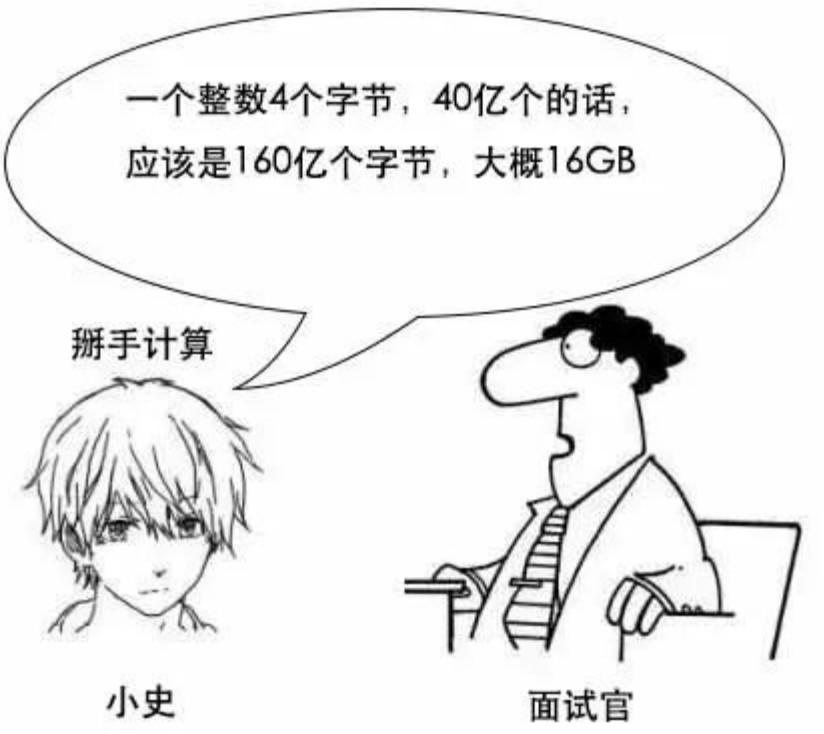

哦，其实用一个set存储就好了，新来的数判断是否在set中
顿了一下
小史
面试官

互联网侦察

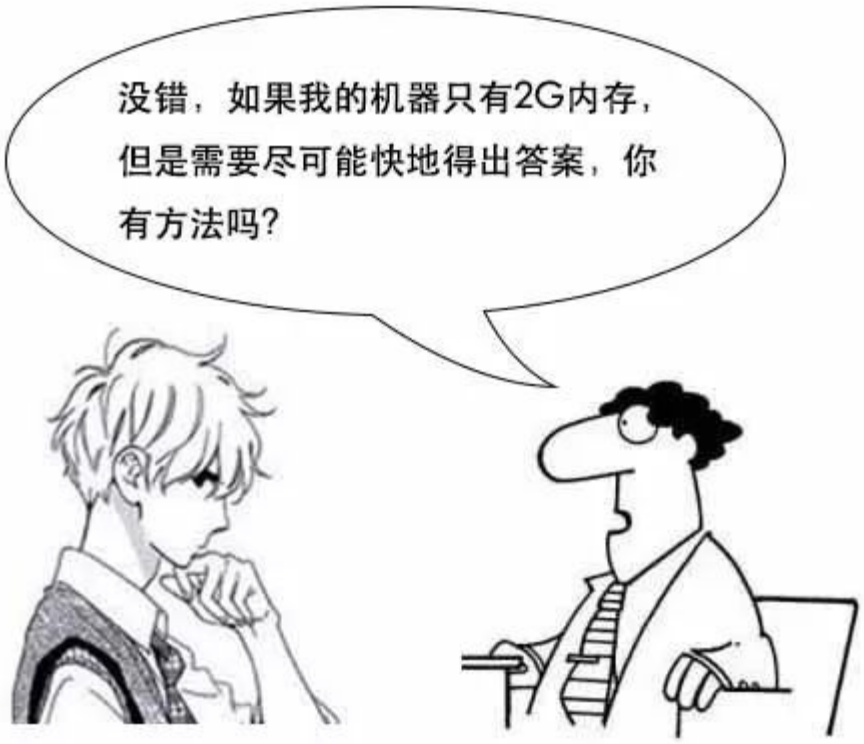


小史　　面试官

互联网侦察

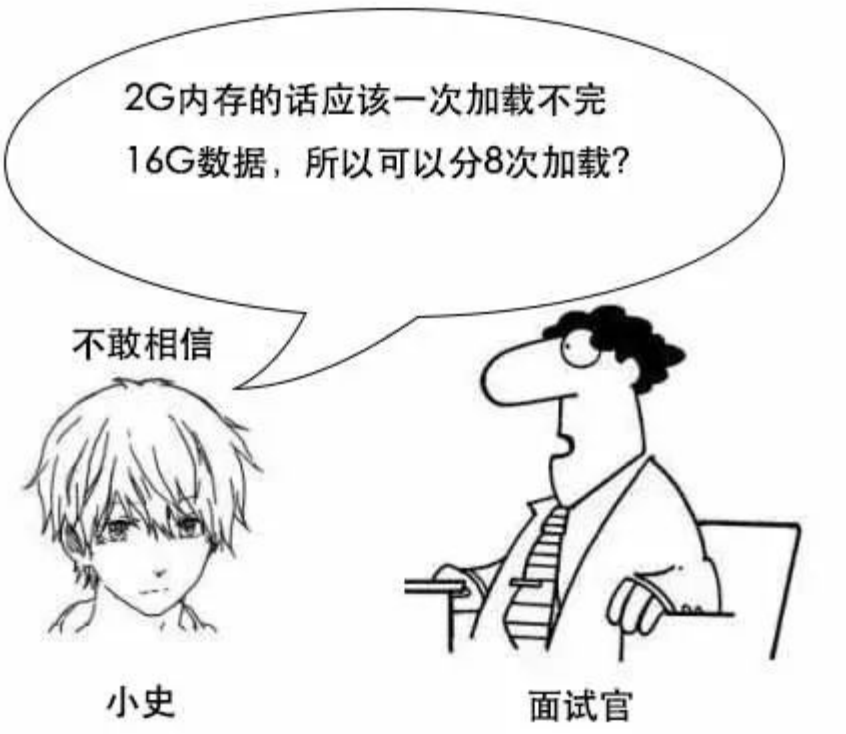


互联网侦察

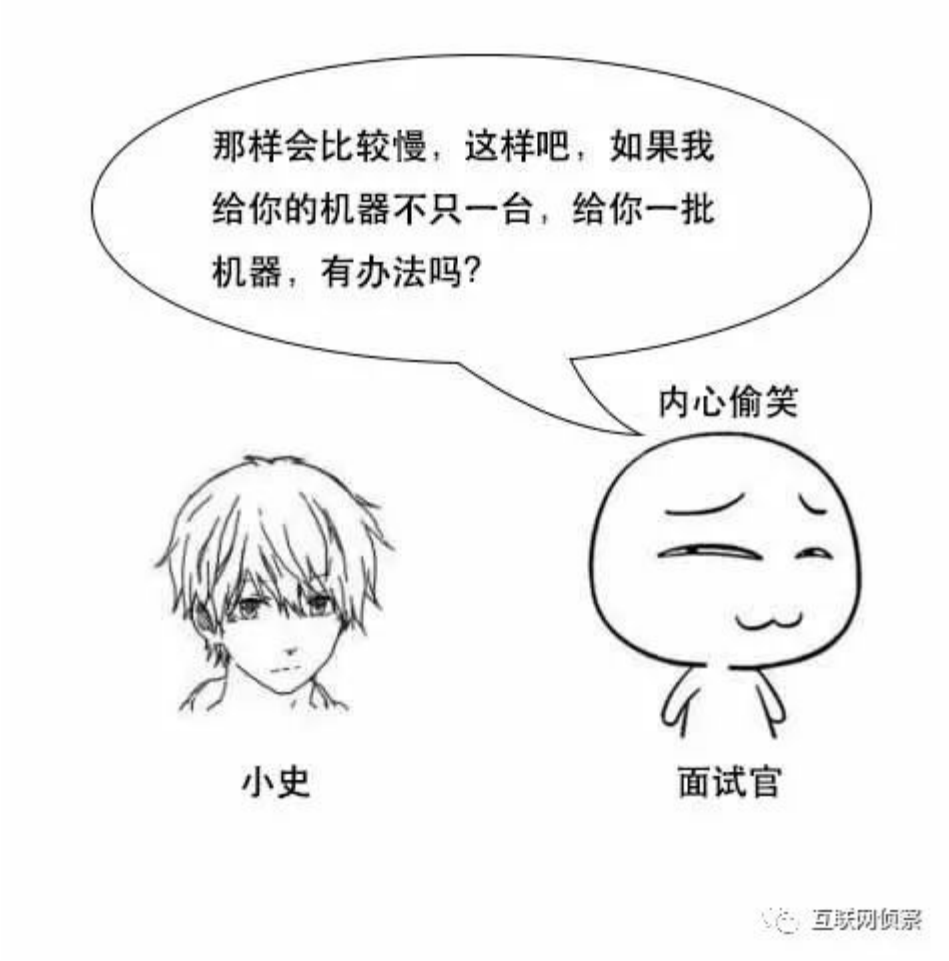


小史　　面试官

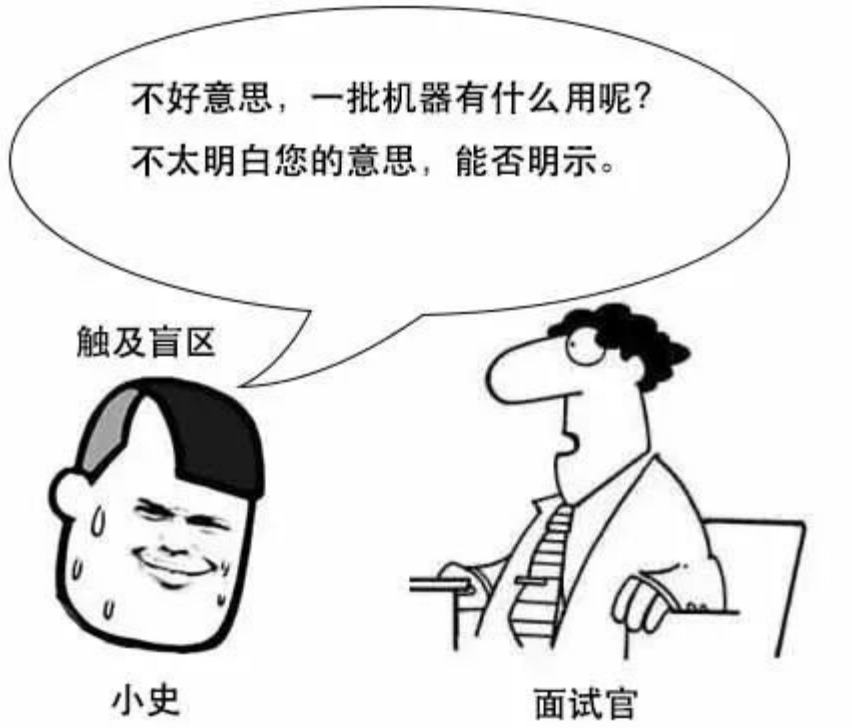
互联网侦察



互联网侦察



互联网侦察



互联网侦察

【请教大神】

小史回到学校，把面试的情况和计算机学院的吕老师说了一下。

小史忙拉着吕老师问，为什么我说分8次加载数据，面试官会说太慢了呢?

吕老师：哈哈，从磁盘加载数据是磁盘io操作，是非常慢的，你每次都要加载这么大的数据，还要8次，我估计你找一个数的时间可以达到分钟甚至小时级了。

小史：那如果是你，你会怎么办呢？

吕老师：其实面试官已经提示得比较明显了，他说给你一批机器，就是暗示你可以用分布式算法。你把数据分散在8台机器上，然后来一个新的数据，8台机器一起找，最后再汇总结果就行了。

小史：这样的话能快多少？

吕老师：这样应该能达到秒级。小史，你可以自己分析分析。

小史：我想想……哦，这样做的话，因为每台机器都可以一次性把数据读入内存，在比较的时候不用来回加载数据了，所以可以节省加载数据的开销！这真是个好办法。

【更好方案】

吕老师：其实这并不是最好方法，我这还有一种毫秒级的方法，想不想知道啊？

小史：当然想啊，快教教我。

小史：哦，对哦，这样我就申请40亿个位就好了，新的数转换成一个位，然后判断一下这个位是0还是1就行了。

吕老师：小史啊，考虑问题要考虑清楚啊，如果是40亿个位，那么这40亿个位哪些是0，哪些是1呢？来了一个新的数，怎么判断是否在40亿个位之中？

小史：我想想，对啊，40亿个位，40亿个数，那么每个位都是1，这。。。

吕老师：其实你可以想想，32位int的范围，总共就是2的32次方，大概42亿多点。所以你可以申请2的32次方个位。

小史：意思是我把整个整数范围都覆盖了，哦，对哦。这样一来，就可以做了，1代表第一个位，2代表第二个位，2的32次方代表最后一个位。40亿个数中，存在的数就在相应的位置1，其他位就是0。

吕老师：没错，那来了一个新的数呢？

小史：新的数就去找相应的位，比如来了一个1234，就找一下第1234位，如果是1就存在，是0就不存在啦。

吕老师：没错，那么这样的话，需要多大内存呢？

小史：我想想啊，2的32次方个位，相当于2的29次方个字节，哇，才500MB，真是节省了不少内存呢。

小史：这么厉害的算法，你是怎么想到的？

吕老师：其实这是一种非常有名的大数据算法，叫位图法，英文名叫bitmap。顾名思义，就是用位来表示状态，从而节省空间。明天正好我有一节课，就讲位图法，你可以来听一听。

【吕老师的课】

第二天，吕老师开始上课，他一开始就抛出了小史遇到的面试题。

吕老师：同学们，这道题是BAT公司的一道面试题，大家有什么思路吗？

话音刚落，蛋哥就站起来回答。蛋哥是吕老师最得意的门生，以思维活跃著称。

蛋哥：我觉得可以这样。首先，32位int的范围是42亿，40亿整数中肯定有一些是连续的，我们可以先对数据进行一个外部排序，然后用一个初始的数和一个长度构成一个数据结构，来表示一段连续的数，举个例子。

如果数据是1 2 3 4 6 7……这种的，那么可以用(1,4)和(6,2)来表示，这样一来，连续的数都变成了2个数表示。

来了一个新数之后，就用二分法进行查找了。

这样一来，最差情况就是2亿多的断点，也就是2亿多的结构体，每个结构体8个字节，大概16亿字节，1.6GB，在内存中可以放下。

吕老师：嗯，非常好，不仅给出了方案，还能主动分析空间和可行性。

小史听完后深感佩服，问题的解决方法绝对不止一种，只要肯动脑筋，即使没有学过bitmap算法，也能有别的方法来解决问题。

END

觉得文章不错的，欢迎点好看和转发，长按下图关注程序员乔戈里，收看更多精彩。

今日问题：面试中你还遇到类似的问题吗？欢迎留言！

文章转载自公众号 互联网侦察