**数据结构与算法处理面试题**

1、时针走一圈，时针分针重合几次

2、N\*N的方格纸,里面有多少个正方形

3、2000万个整数，找出第五十大的数字？

4、百度POI中如何试下查找最近的商家功能(提示：坐标镜像+R树)。

5、一个文件中有100万个整数，由空格分开，在程序中判断用户输入的整数是否在此文件中。说出最优的方法

6、两个不重复的数组集合中，这两个集合都是海量数据，内存中放不下，怎么求共同的元素？

7、烧一根不均匀的绳，从头烧到尾总共需要1个小时。现在有若干条材质相同的绳子，问如何用烧绳的方法来计时一个小时十五分钟呢？

8、万亿级别的两个URL文件A和B，如何求出A和B的差集C(提示：Bit映射->hash分组->多文件读写效率->磁盘寻址以及应用层面对寻址的优化)

9、怎么在海量数据中找出重复次数最多的一个？

10、海量日志数据，提取出某日访问百度次数最多的那个IP。

11、在2.5亿个整数中找出不重复的整数，注，内存不足以容纳这2.5亿个整数。

12、搜索引擎会通过日志文件把用户每次检索使用的所有检索串都记录下来，每个查询串的长度为1-255字节。

13、有一个1G大小的一个文件，里面每一行是一个词，词的大小不超过16字节，内存限制大小是1M。返回频数最高的100个词。

14、有10个文件，每个文件1G，每个文件的每一行存放的都是用户的query，每个文件的query都可能重复。要求你按照query的频度排序。

15、给定a、b两个文件，各存放50亿个url，每个url各占64字节，内存限制是4G，让你找出a、b文件共同的url？

16、一个文本文件，大约有一万行，每行一个词，要求统计出其中最频繁出现的前10个词，请给出思想，给出时间复杂度分析。

17、腾讯面试题：给40亿个不重复的unsigned int的整数，没排过序的，然后再给一个数，如何快速判断这个数是否在那40亿个数当中？

附、100w个数中找出最大的100个数。