



百度2015大数据云计算研发笔试卷

一. 问答题

1. 请简要描述一下Hadoop, Spark, MPI三种计算框架的特点以及分别适用于什么样的场景

正确答案: a) Hadoop

基于分布式文件系统HDFS的分布式批处理计算框架。适用于数据量大, SPMD(单程序多数据)的应用。

b) Spark

基于内存计算的并行计算框架。适用于需要迭代多轮计算的应用。

c) MPI

基于消息传递的并行计算框架。适用各种复杂应用的并行计算。支持MPMD(多程序多数据), 开发复杂度高

2. 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。

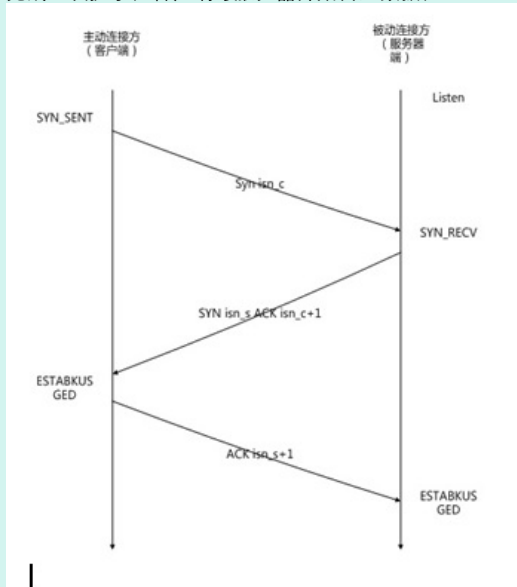
正确答案:

第一次握手: 建立连接时, 客户端调用发送syn包(syn=j)到服务器, 并进入SYN_SEND状态, 等待服务器确认;

第二次握手: 服务器端收到syn包, 必须确认客户的SYN (ack=j+1), 同时自己也发送一个SYN包 (syn=k), 即SYN+ACK包, 此时服务器进入SYN_RECV状态;

第三次握手: 客户端收到服务器的SYN + ACK包, 向服务器发送确认包ACK(ack=k+1), 此包发送完毕, 客户端和服务器进入ESTABLISHED状态, 完成三次握手。

完成三次握手, 客户端与服务器开始传送数据;



状态图如下:

相关系统调用: 客户端调用connect()开始建立连接, 连接建立好后退出

服务器端调用完listen()后就可以响应连接请求, 连接请求建立好后调用accept()把连接拿出开始通信

注意: accept()跟server建立连接没有关系, 它只是取出建立好连接的socket, 不参与连接建立的过程。

3. 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。

正确答案: int getMax(int a[],int len)

```
{
    int max1 = a[0]; //表示maxSum(n-2);
    int max2 = a[0]>a[1]? a[0]:a[1]; //表示maxSum(n-1);
    int max3 = 0; // n
    for(int i = 2; i<len; i++){
        max3 = Max(a[i],Max(max1+a[i],max2));
        // max3 = a[i]+max1>max2 ? a[i]+max1:max2; // 全部是负数也需要考虑的,这个没有
        max1 = max2;
        max2 = max3;
    }
    return max3;
}
```

```
int Max(int a,int b){
    if(a>b)
        return a;else
        return b;
}
```



4. 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 *));
```

正确答案：

5. 有编号1~100个灯泡，起初所有的灯都是灭的。有100个同学来按灯泡开关，如果灯是亮的，那么按过开关之后，灯会灭掉。如果灯是灭的，按过开关之后灯会亮。

现在开始按开关。

第1个同学，把所有的灯泡开关都按一次(按开关灯的编号： 1,2,3,.....100)。

第2个同学，隔一个灯按一次(按开关灯的编号： 2,4,6,.....,100)。

第3个同学，隔两个灯按一次(按开关灯的编号： 3,6,9,.....,99)。

.....

问题是，在第100个同学按过之后，有多少盏灯是亮着的？这些灯的编号是多少？要求给出解题思路或给出伪码。

正确答案：

10盏， 1, 4, 9, 16, 25, 36, 49, 64, 81, 100

按照同学来看，每个同学只会按是自己的倍数的灯。

那么我们转换成灯来看的话，每个灯只会被是自己的因子的同学按。

那么一个初始化为灭的灯，如何最后变成一盏亮的灯呢？

很明显，只有它有奇数个因子的时候，才有可能。

那么什么时候一个数可以有奇数个因子呢？

对于任意一个数N，都可以分解成 $N = a * b$ 的乘积，即任意一个数都可以分解成 M个($a * b$) 的乘积。

所以若想满足存在奇数个因子，a 必须等于 b。

即 $N = a^2$ ，所以只有平方数最后才满足要求，故可以在O(n)的时间复杂度解决该问题。

6. 打长沙麻将在一开始，只有庄家可得到十四张牌，其余的人十三张。现在庄家手里拿到十四张牌，他想请你写个程序帮忙判断一下，庄家是否已经胡牌。

如果你会打麻将，请忽略以下背景，如果不会，简单了解一下背景有助于理解本题：

长沙麻将打法简单、节奏快速，极易胡牌。长沙麻将共一百零八张牌：包括筒、索、万；不带东、南、西、北风、中、发、白。：

1、万子牌：从一万至九万，各4张，共36张。

2、筒子牌：从一筒至九筒，各4张，共36张。也有的地方称为饼，从一饼到九饼。

3、束子牌：从一束至九束，各4张，共36张。也有的地方称为条，从一条到九条。

组牌规则：

1，对子：两张一样花色，一样大小的牌，组成对子。

2，顺子：三张相同花色，连续的牌，组成顺子。

3，刻子：三张一样花色，一样大小的牌，组成刻子。

胡牌规则：每人有十四张牌，如果这十四张牌可以组成：一个对子，若干个顺子和刻子，则表示胡牌。比如以下牌型已经胡牌：

一万，一万，二万，三万，四万，二条，三条，四条，四条，四条，四条，五筒，六筒，七筒。

1：请描述你对这个问题的理解，并写出你的解题思路。

1.1，按花色细分处理，必须是一个花色的牌个数 3的倍数余2（留对子），其它花色的个数都是3的倍数。否则不能胡牌

1.2，从3的倍数余2的花色中选出一对，剩下的牌的处理和其它花色一样。如果没有对子，则不能胡牌。

1.3，对于某一个花色的牌，由于个数为3的倍数，判断其是否可以组成若干个顺子或刻子，否则不能胡牌。

1.4，对相同花色的牌进行排序和计数，判断第一张牌能否和其它牌组成顺子或刻子，若不能，则回溯。若能，由继续处理剩下的牌。

1.5，最后判断是否可以胡牌



2.请设计解决问题需要的数据结构。

需要设计一个花色的数据结构，包括type（花色）， id（牌的大小）， count（牌出现的次数）

正确答案：



技术QQ群：379386529



微博：<http://www.weibo.com/nowcoder>



微信

登录牛客网，参与以上题目讨论，查看更多笔试面试题