第一部分 单选题（前10题，每题2分；后10题，每题3分，共50分，选对得满分，选错倒扣1分，不选得0分）  
  
1、假设把整数关键码K散列到有N个槽的散列表，以下哪些散列函数是好的散列函数（）  
A、h(K)=K mod N;  
B、h(K)=1;  
C、h(K)=K/N;  
D: h(K)=(K+rand(N)) mod N, rand(N)返回一个0到N-1的整数  
  
2. 下面排序算法中，初始数据集的排列顺序对算法的性能无影响的是（）  
A、堆排序      B、插入排序  
C、冒泡排序    D、快速排序  
  
3、下面说法错误的是：  
A、CISC计算机比RISC计算机指令多  
B、冯诺依曼机体系结构的主要特征是存储程序的工作方式  
C、增加流水线段数理论上可以提高CPU频率  
D、在指令格式中，采用扩展操作码设计方案的目的是为了保持指令字长不变而增加寻址空间  
  
4、不属于冯诺依曼机体系结构必要组成部分的是：  
A、CPU          B、Cache        C、RAM        D、ROM  
  
5、一个栈的入栈序列式ABCDE，则不可能的出栈序列是:  
A、DECBA          B、DCEBA          C、ECDBA       D、ABCDE  
  
6.你认为可以完成编写一个C语言编译器的设计语言是：  
A、汇编语言        B、C语言       C、VB语言       D、以上皆可  
  
7. 关于C++/JAVA类中的static成员和对象成员的说法正确的是：  
A、虚成员函数不可能是static成员函数  
B、static成员函数在对象成员函数中无法调用  
C、static成员变量在对象构造时生成  
D、static成员函数不能访问static成员变量  
8、



C、13个  
9、某进程在运行过程中需要等待从磁盘上读入数据，此时该进程的状态将：  
A、从就绪变为运行       B、从运行变为就绪  
C、从运行变为阻塞       D、从阻塞变为就绪  
  
10、下面算法的时间复杂度为：

**[cpp]** [view plaincopyprint?](http://blog.csdn.net/hackbuteer1/article/details/11931173)

1. **int** f(unsigned **int** n)
2. {
3. if(n == 0 || n == 1)
4. return 1;
5. else
6. return n\*f(n-1);
7. }

int f(unsigned int n)

{

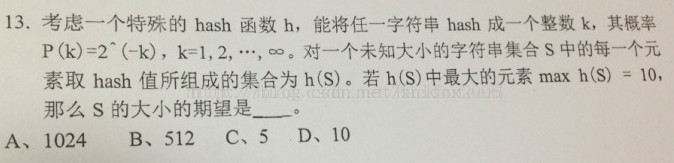
if(n == 0 || n == 1)

return 1;

else

return n\*f(n-1);

}

A、O(1)        B、O(n)        C、O(N\*N)       D、O(n!)  
  
11、n从1开始，每个操作可以选择对n加1或者对n加倍。若想获得整数2013，最少需要多少个操作。  
A、24        B、21        C、18       D、不可能  
  
12、对于一个具有n个顶点的无向图，若采用邻接表数据结构表示，则存放表头节点的数组大小为：  
A、n         B、n+1       C、n-1      D、n+边数  
  
  
  
14：如下函数，在32bit系统foo(2^31-3)的值是：

**[cpp]** [view plaincopyprint?](http://blog.csdn.net/hackbuteer1/article/details/11931173)

1. **int** foo(**int** x)
2. {
3. return x&-x;
4. }

int foo(int x)

{

return x&-x;

}

A、0         B、1          C、2         D、4  
  
15、对于顺序存储的线性数组，访问节点和增加、删除节点的时间复杂度为：  
A、O(n),O(n)         B、O(n),O(1)         C、O(1),O(n)          D、O(1),O(1)  
  
16、在32位系统环境中，编译选项为4字节对齐，那么sizeof(A)和sizeof(B)是：

**[cpp]** [view plaincopyprint?](http://blog.csdn.net/hackbuteer1/article/details/11931173)

1. struct A
2. {
3. **int** a;
4. **short** b;
5. **int** c;
6. **char** d;
7. };
8. struct B
9. {
10. **int** a;
11. **short** b;
12. **char** d;
13. **int** c;
14. };

struct A

{

int a;

short b;

int c;

char d;

};

struct B

{

int a;

short b;

char d;

int c;

};

A、16,16               B、16,12            C、13,12       D、11,16  
  
17、袋中有红球，黄球，白球各一个，每次任意取一个又放回，如此连续抽取3次，则下列事件中概率是8/9的是：  
A、颜色不全相同            B、颜色全相同          C、颜色全不同         D、颜色无红色  
  
18、一个洗牌程序的功能是将n张牌的顺序打乱，以下关于洗牌程序的功能定义说法最恰当的是：  
A、任何连续位置上的两张牌的内容独立  
B、n张牌的任何两个不同排列出现的概率相等  
C、每张牌出现在n个位置上的概率相等  
D、每张牌出现在n个位置上的概率独立  
  
19、用两种颜色去染排成一个圈的6个棋子，如果通过旋转得到则只算一种，一共有多少种染色模式。  
A、10         B、14         C、15         D、16  
  
20、递归式的先序遍历一个n节点，深度为d的二叉树，则需要栈空间的大小为：  
A: O(logn)           B:O(nlogn)          C:O(n)             D:(d)  
  
第二部分 不定向选项（4题，每题5分，完全正确计5分，漏选计2分，不选计0分，多选、错选计-2分）  
21、两个线程运行在双核机器上，每个线程主线程如下，线程1：x=1;r1=y;线程2：y=1;r2=x;  
x和y是全局变量，初始为0。以下哪一个是r1和r2的可能值：  
A、r1=0,r2=0  
B、r1=1,r2=0  
C、r1=1,r2=1  
D、r1=0,r2=1  
  
22、关于Linux系统的负载（Load），以下表述正确的是：  
A: Load:2.5,1.3,1.1表示系统的负载压力在逐渐减小  
B: 通过就绪和运行的进程数来反映  
C: 通过TOP命令查看  
D: 通过uptime查看  
  
23、关于排序算法的以下说法，错误的是：  
A、归并排序的平均时间复杂度O(nlogn)，最坏时间复杂度O(n^2)  
B、堆排序平均时间复杂度O(nlogn)，最坏时间复杂度O(nlogn)  
C、冒泡排序平均时间复杂度O(n^2)，最坏时间复杂度O(n^2)  
D、快速排序的平均时间复杂度O(nlogn)，最坏时间复杂度O(N^2)  
  
24、假设函数rand\_k会随机返回一个【1，k】之间的随机数（k>=2）,并且每个证书出现的概率相等。目前有rand\_7,通过调用rand\_7（）和四则运算符，并适当增加逻辑判断和循环控制逻辑，下列函数可以实现的有：  
A、rand\_3         B、rand\_21          C、rand\_23         D、rand\_47  
  
第三部分 填空与问答  
  
  
25、某二叉树的前序遍历序列为-+a\*b-cd/ef，后序遍历序列为abcd-\*+ef/-，问其中序遍历序列是：  
  
26、某缓存系统采用LRU淘汰算法，假定缓存容量为4，并且初始为空，那么在顺序访问以下数据项的时候，1、5、1、3、5、2、4、1、2，出现缓存直接命中的次数是（），最后缓存中即将准备淘汰的数据项是（）  
  
27、有两个较长的单向链表a和b，为了找出节点node满足node in a 并且 node in b，请设计空间使用尽量小的算法。（用C/C++/JAVA或伪码表示都可以）  
  
28、当存储数据量超出单节点数据管理能力的时候，可以采取的办法有数据库sharding的解决方案，也就是按照一定的规律把数据分散存储在多个数据管理节点N中（节点编号0.1.2...N-1）。假设存储的数据是a，请完成为数据a计算存储节点的程序。（没学过C语言的同学也可以用伪码完成）

**[cpp]** [view plaincopyprint?](http://blog.csdn.net/hackbuteer1/article/details/11931173)

1. #define N 5
2. **int** hash(**int** element)
3. {
4. return element\*2654435761;
5. }
6. **int** shardingIndex(**int** a)
7. {
8. **int** p = hash(a);
9. //1
10. return p;
11. }

#define N 5

int hash(int element)

{

return element\*2654435761;

}

int shardingIndex(int a)

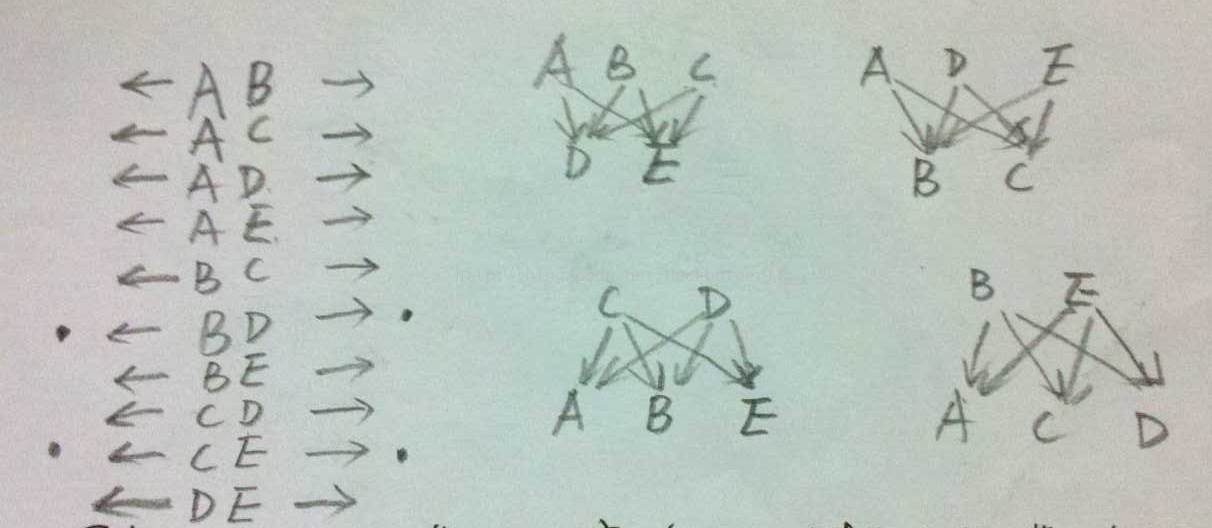
{

int p = hash(a);

//1

return p;

}

空格1处：     p %= N;  
29、宿舍内5个同学一起玩对战游戏，每场比赛有一些人作为红方，另一些人作为蓝方，请问至少需要多少场比赛，才能使任意两个人之间有一场红方对蓝方和一场蓝方对红方的比赛？  
答案为4场。  
  
第四部分 JAVA选做题  
1、以下每个线程输出的结果是什么？（不用关注输出的顺序，只需写出输出的结果集即可）  
  
  
2、一个有10亿条记录的文本文件，已按照关键字排好序存储，请设计算法，可以快速的从文件中查找关键字的记录。  
  
C++选做题  
Part I  
假设你有一台计算机，配置如下：  
48GB内存  
16核CPU，3.0GHz  
12块2TB SATA硬盘  
有两个数据文件A和B，A的大小是40GB，B的大小是2TB，A和B的文件格式一样，都包含等长的100字节的记录，记录的前20个字节表示key，后80个字节表示value，所有的key和value都由数字和大小写字母组成（0-9 A-Z a-z），同一个文件中的key没有排序，也没有重复。  
文件A和B都切成了1GB（1\*10^9字节）的数据块（名为A000001、A000002......A000010、B000001、B000002......B002000），均匀分布在6块硬盘上。  
请问如何用最快的方法找到A和B之间共同的key，以及他们对应的value值（建议输出格式如下所示：<key><空格><A中对应value><空格><B中对应value>）  
  
  
请描述你的方法里面用到的关键的数据结构和算法，估算这个方法需要的内存空间和运算时间，并说明你的推导过程。  
  
  
Part II  
如果你有100台服务器，每台配置如上描述，它们通过千兆网络组成一个集群，任意两台之间的带宽可以达到1000Mbps，同时假设文件A和B的大小也放大100倍（各位4TB和200TB），并且被切分成1GB的碎片，均匀分布在100台服务器上。  
请问如何用最快的方法找到A和B之间共同的key，以及他们对应的value值（建议输出格式如下所示：<key><空格><A中对应value><空格><B中对应value>）  
  
  
请描述你的方法里面用到的关键的数据结构和算法，估算这个方法需要的内存空间、网络流和运算时间，并说明你的推导过程。