**数据结构与算法题**

**\* 单向链表归并排序问题**

|  |  |
| --- | --- |
| 1 | public class Node { |
| 2 | public Node next; | |

|  |  |  |
| --- | --- | --- |
| 3 | public int value; | |
| 4 | } |

实现下面这个方法

node1是个一个链表的header，这个链表的元素是按照value值从小到大排序好的。

node2是个另一个链表的header，这个链表的元素是按照value值从小到大排序好的。

要求返回一个链表的header，其元素是input两个链表merge的集合，并要求按照value值从小到大排序

public Node merge(Node node1, Node node2);

出题者：黄宁海

考察点：代码能力，边界情况的考虑(两个链表相交，链表有环)

**\* 容器设计**

设计一个容器，主要支持3个操作：

1. 插入一个整数

2. 删除一个整数

3. 随机返回容器内的一个整数，每个整数被返回的概率均等

可以假设插入时整数都不重复，删除时容器内一定有你想删除的那个整数。同时系统提供一个完全随机的复杂度为O(1)的随机函数。需要上述3个操作都是O(1)，怎么设计？

出题者：黄宁海

考察点：

**\* 请反转一个已知头节点的单向链表**

出题者：候浩

考察点：

**\* 请写出一个有序二叉树的查找算法（提示：输入为根节点和节点值，输出为相应的节点）**

出题者：候浩

考察点：

**\* 在一个数组中找出第二大的数**

出题者：郭凛

考察点：边界情况，如：数组中所有数都一样大的情况，有两个最大数的情况，有两个第二大数的情况

**\* 计算Tribonacci队列**

计算Tribonacci队列，规则是T( n ) = T(n - 1) T(n - 2) T(n -3)，其中T(0) = T(1) = 1，T(2) = 2。

函数定义：int Tribonacci(int n) {}

出题者：黄宁海

考察点：

**\* 字符串解析函数设计**

在一个字符串中找到第一个只出现一次的字符。如abaccdeff，输出b

出题者：黄宁海

考察点：

**\* 计算N\*N的矩阵的n次幂**

写一个N\*N的矩阵的n次幂

函数定义：int[][] pow(int[][] mat, int n);

出题者：黄宁海

考察点：

**\* 设计表达式解析器**

写一个表达式解析器，要求不能用递归，表达式的内容包括0-9和+-\*/()

出题者：黄宁海

考察点：

**\* 股票最大收益**

百度是一家纳斯达克上市公司。现在假设你持有着百度的股票，但每天只有一个可买卖的时间点。给出一份百度股市价格表，描述了过去n（n>=2）天百度在可买卖时间点的股价，要求找出一个买点，一个卖点，使收益最大

出题者：黄宁海

考察点：

**\* 填字游戏**

假设有个N\*N的二维数组，表示一个N\*N的表格，每个格子里都有一个英文字母，现给出一个英文单词，需要实现一个函数检测出在表格中是否能通过连线（横竖斜均可）找到此单词。

出题者：黄宁海

考察点：

**\* 城管和钉子户问题**

和谐而方正的亚克西王国可以看成是一个N\*N的矩阵，王国中有存在的城管和钉子户这两种高度对立的群体，他们分布在N\*N的各个节点上。拆迁户为了自己的利益，前方百计要把城管干掉。王国中每一天里，拆迁户就会有一次行动，为了发扬游击战的作风，每次移动都是按照象棋中马的移动方式进行，在当拆迁户移动到城管所在的位置，这名城管就会光荣牺牲。

1. 目前，王国中有一个拆迁户，处于a,b的位置，王国中有N个城管(x1,y1)(x2,y2)…(xn,yn)，同一位置没有多个城管。请给出一个程序，证明这个拆迁户是否能消灭王国中所有的城管。

2. 随着斗争的扩大，一个拆迁户显然力不从心，而城管在斗争中，不断总结经验教训，懂得了使用路障来阻止拆迁户的进攻。所以每次拆迁户移动时，不能移动到路障所在的节点。现在王国中有N个拆迁户，分别处于(a1,b1) (a2,b2)… (an, bn)的位置，有M个城管，处于(x1,y1)(x2,y2)…(xm,ym)，有P个路障(i1,j1),(i2,j2)…(ip,jp)，任何一个节点上不能有多个城管，拆迁户和路障。请给出一个程序来证明，在给定的天数D天内，拆迁户是否能够灭掉王国中所有的城管。

出题者：肖潇

考察点：

**工程应用题**

**\* TinyURL服务设计**

最近上线的百度说吧（微博），由于微博每条博文只能写入140个字（一条手机短信长度），而URL往往太长以至于占掉了内容的大部分，因此需要实现一个TinyURL服务，当用户提交内容包含长URL时自动将其转换成TinyURL，例如：

用户输入：http://zhidao.baidu.com/question/189764699.html

实际输出：http://duurl.cn/a25Kf

其中duurl.cn是TinyURL服务的域名

现在假设你是TinyURL服务的设计者，你会如何来实现及规划？

可选的笔试或面试问题（按顺序，由易到难。如果后面的问题包含前面问题的答案，可以挑选一部分写在笔试题中，另一部分面试时问）：

1. tinyurl后面的编码如何生成*（Hash算法是否有思路？尽量短，6位以内，(大写+小写+数字)^6=560亿）*

2. 当用户给出一个已映射过的URL时如何找到其已存在的tinyurl*（为了节省空间）*

3. 如果URL数量达到十亿条，在一个服务器上存储不下，要分布到10个服务器上，怎么运作？

4. 如果产生了hash出来某个服务器数量特别多的情况，怎么运作？

5. 如果数量继续增加，10个服务器不够了，还需要增加服务器，但每次重算hash影响太大，怎么办？*（即要能随时增加服务器怎么设计？高可扩展性）*

6. 如果现在服务器已经增加到了50台，但是三天两头就有服务器会挂掉，怎么办？*（即要能随时减少服务器却保证服务尽量正常怎么设计？高可用性）*

出题者：黄宁海，郭凛

考察点：发散思维，在遇到没见过的问题时的思维开放性。对hash算法的理解和应用，对一致性hash算法的理解和应用，对工程上高可扩展性、高可用性的理解和应用

**\* 一亿条URL数据模糊匹配**

出题者：郭凛

考察点：发散思维

**\* 电骡(eMule)等P2P软件资源定位原理（未完待补充）**

出题者：郭凛

考察点：一致性hash的应用，与高可用性、高扩展性的分布式缓存、分布式存储原理相同

**\* 数据库索引设计**

如果让你来设计数据库索引，你会怎么做？（挑选怎样的数据结构）

可选问题与关注点：

1. 使用hash做索引有哪些优点和缺点？

2. 为什么会产生B-tree、B+tree这种数据结构，或是B-tree、B+tree这种数据结构主要是为了解决什么问题？

3. 为什么B+tree结点上的关键字要使用闭区间而不是开区间？这样做有什么好处？

4. 如果要做到文本字段like语句可以用到索引要怎么做？（MySQL Fulltext索引实现）

5. B+tree的分裂方式？

出题者：郭凛

考察点：Hash、B+Tree、二分查找、数据结构与实际工程场景的联系

**\* 百度百科词条自动绑定问题**

百度百科每个页面由词条和正文内容两部分组成。其中若正文内容中包含了已经存在的其他词条，会自动将那个词变为一个可点击链接指向词条，例如：

|  |
| --- |
| **百度**  百度（Nasdaq简称：BIDU）是全球最大的中文搜索引擎 |

其中第一行的“百度”是本页词条，“Nasdaq”和“中文搜索引擎”两个是已存在的词条被绑定上了链接，正文内容中的“百度”因为是当前页词条将不会被加上链接

问题：假设已有所有词条的列表的文本文件wordlist.txt，每行一个词，请写出替换算法。

（当多个词条匹配到相同内容时，优先匹配字数多的词条。例如：当正文内容包含“百度百科”时，优先匹配“百度百科”而不是“百度”）

可选问题与关注点（问题可以不写在试卷上，等到面试时再逐个询问）：

1. 是否注意到了英文匹配和中文匹配的差别？*（按whole word方式匹配，不能使用长词条优先匹配规则）*

2. 当发生可以同时匹配到的重叠词条时，如何匹配？*（例如：五环线，是匹配“五环”还是匹配“环线”）*

3. 当词条数量巨大时，如何保证效率？

4. 这个自动绑定在什么时机做更合适？各个时机做各有些什么优缺点？*（编辑时、提交时、显示时、定时）*

出题者：吴金亮、郭凛

考察点：对正则表达式原理的理解，对多字符字符串的跨界查找及替换等，边界情况考虑，工程上可能发生的问题

**\* 贴吧数据库结构与查询**

百度贴吧需要重新规划，而你被挑选为数据库方面的负责人，你注意到贴吧数据量非常大（上百亿个帖子），现在你需要完成以下工作：

1. 设计数据库结构，给出ER图

2. 给出取出主题内容及其回复内容的SQL

3. 你会采取一些什么措施来保证访问效率和高可用性？请描述方案，需要时给出相应的存取数据的代码

出题者：吴金亮

考查点：数据库范式及反范式的理解、SQL语言、数据库的垂直分区和水平分区以及shard的相关算法理解

**程序设计**

**\* Number类面向对象设计**

有Number、Integer、Double三个类，其中Integer和Double均继承于Number，现希望实现三个类能满足如下代码：

|  |  |
| --- | --- |
| 1 | Number a = new Integer(3); |
| 2 | Number b = new Double(2.1); | |

|  |  |
| --- | --- |
| 3 | Number c = a.add(b);  // 3 + 2.1 = 5.1 |

要求，不允许使用instanceof指令

出题者：黄宁海

考察点：

**编程语言相关题**

**\* [C] 变量内存地址分配问题**

有下面的一个程序

|  |  |
| --- | --- |
| 01 | #include<iostream> |
| 02 | #define ARRAY\_SIZE 10 | |

|  |  |
| --- | --- |
| 03 |  |
| 04 | using namespace std; | |

|  |  |
| --- | --- |
| 05 |  |
| 06 | void natural\_numbers() { | |

|  |  |
| --- | --- |
| 07 | int i; |
| 08 | int array[ARRAY\_SIZE]; | |

|  |  |
| --- | --- |
| 09 |  |
| 10 | while( i <= ARRAY\_SIZE) { | |

|  |  |  |
| --- | --- | --- |
| 11 | array[i] = i – 1; | |
| 12 | i = i + 1; |

|  |  |  |
| --- | --- | --- |
| 13 | } | |
| 14 |  |

|  |  |  |
| --- | --- | --- |
| 15 | cout << array[10] << endl; | |
| 16 | } |

请写出程序输出结果

答案：由于i的内存地址在array之后分配，因此array[10]将指向i所在地址，因此程序最后会进入死循环

出题者：肖潇

考察点：C语言变量内存地址分配

**\* [C/C++] 小端法和大端法表示问题**

编写函数is\_little\_endian，在小端法表示的机器上编译和运行时返回1,在大端法表示的机器上编译和运行时返回0，这个程序应该能够运行在任何机器上，无论机器的字长是多少。

出题者：肖潇

考察点：小端法和大端法的理解以及编程能力

**\* [Java] 多线程问题**

实现一个方法，要求其中的逻辑只能被最多两个线程执行(如果超过两个线程则第三个线程阻塞直到调用方法逻辑的线程低于两个)

出题者：郭凛

考察点：多线程的理解以及

**\* [Java] 统计Web类执行时间**

如果要记录Servlet每个请求消耗的时间，该如何做？如果要记录每个Service的执行时间怎么做？如果要单独记录Servlet的执行时间（不包含其调用Service的时间）怎么做？

出题者：郭凛

考察点：Servlet容器的理解，AOP的应用，ThreadLocal的了解

**\* [C++] 在C++程序中，整型（int）文字常量存放在：**

A. 常量区

B. 代码区

C. 堆区

D. 栈区

出题者：肖潇

考察点：对C++内存模型的理解

**\* [C++] 下面关于main()函数的描述中，正确的是：**

A. main()函数一定是C++程序中第一个被执行的函数

B. main()函数可以重载

C. main()函数的返回类型应该是int而不是void

D. 在main()函数中一定要显式使用return语句返回一个整数值

出题者：肖潇

考察点：对C++ main函数的熟悉程度

**\* [C] 请自定义函数实现int strlen(const char\* str)，你能想到几种算法？**

出题者：候浩

考察点：

**\* [C] 请自定义实现void\* memmove(void \*dest, const void \*src, size\_t n)**

出题者：候浩

考察点：