

哈希表（散列表）原理详解

原创 2016年06月03日 15:23:19 标签：哈希表 / 散列表 62699 57

什么是散列表？

哈希表（Hash table，也叫散列表），是根据关键码值(Key value)而直接进行访问的数据结构。也就是说，它通过把关键码值映射到表中一个位置来访问记录，以加快查找的速度。这个映射函数叫做散列函数，存放记录的数组叫做散列表。

记录的存储位置=f(关键字)

这里的对应关系f称为散列函数，又称为哈希（Hash函数），采用散列技术将记录存储在一块连续的存储空间中，这块连续存储空间称为散列表或哈希表（Hash table）。

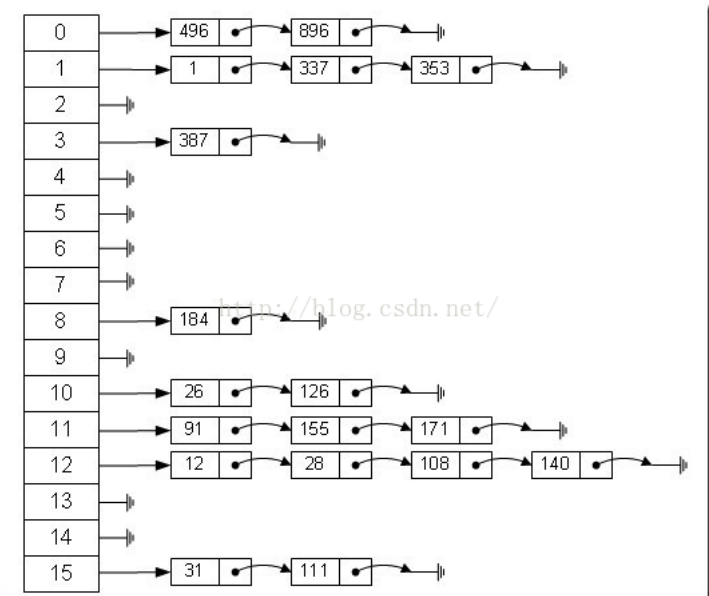
哈希表hashtable(key, value)就是把Key通过一个固定的算法函数既所谓的哈希函数转换成一个整型数字，然后就将该数字对数组长度进行取余，取余结果就当作数组的下标，将value存储在以该数字为下标的数组空间里。（或者：把任意长度的输入（又叫做预映射，pre-image），通过散列算法，变换成固定长度的输出，该输出就是散列值。这种转换是一种压缩映射，也就是，散列值的空间通常远小于输入的空间，不同的输入可能会散列成相同的输出，而不可能从散列值来唯一的确定输入值。简单的说就是一种将任意长度的消息压缩到某一固定长度的消息摘要的函数。）

而当使用哈希表进行查询的时候，就是再次使用哈希函数将key转换为对应的数组下标，并定位到该空间获取value，如此一来，就可以充分利用到数组的定位性能进行数据定位。

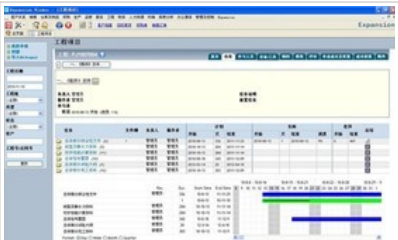
数组的特点是：寻址容易，插入和删除困难；

而链表的特点是：寻址困难，插入和删除容易。

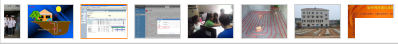
那么我们能不能综合两者的特性，做出一种寻址容易，插入删除也容易的数据结构？答案是肯定的，这就是我们要提起的哈希表，哈希表有多种不同的实现方法，我接下来解释的是最常用的一种方法——拉链法，我们可以理解为“链表的数组”，如图：



左边很明显是个数组，数组的每个成员包括一个指针，指向一个链表的头，当然这个链表可能为空，也可能元素很多。我们根据元素的一些特征把元素分配到不同的链表中去，也是根据这些特征，找到正确的链表，再从链表中找出这个元素。



项目管理 软件



联系我们



请扫描二维码联系客服  
webmaster@csdn.net  
400-660-0108  
QQ客服 客服论坛

关于 招聘 广告服务 百度  
©1999-2018 CSDN版权所有  
京ICP证09002463号

经营性网站备案信息  
网络110报警服务  
中国互联网举报中心  
北京互联网违法和不良信息举报中心



他的最新文章

更多文章

- 深度学习学习资料
- WPS文字设置奇偶页眉、下划线的方法步骤
- face94人脸图像样本库
- 用vector构造二维数组
- 求一个整数所有约数的和(除自身外)

文章分类

- 人脸识别 22篇
- 人脸对齐 6篇
- OpenCV 31篇
- 图像处理 15篇
- 基础知识 15篇
- OpenCV常见错误 19篇

展开

文章存档

- 2018年3月 1篇
- 2017年4月 2篇
- 2017年2月 5篇
- 2016年12月 8篇
- 2016年11月 4篇
- 2016年8月 1篇

展开



项目管理 软件



Go, STL, Docker, 容器, Java, C++, C, C#, JavaScript, Python, PHP, Ruby, Swift, Kotlin, Rust, Go, STL, Docker, 容器, Java, C++, C, C#, JavaScript, Python, PHP, Ruby, Swift, Kotlin, Rust

联系我们



请扫描二维码联系客服  
webmaster@csdn.net  
400-660-0108  
QQ客服 客服论坛

关于 招聘 广告服务 百度

©1999-2018 CSDN版权所有  
京ICP证09002463号

经营性网站备案信息  
网络110报警服务  
中国互联网举报中心  
北京互联网违法和不良信息举报中心

Hash的应用

1、Hash主要用于信息安全领域中加密算法，它把一些不同长度的信息转化成杂乱的128位的编码,这些编码值叫做Hash值。

也可以说，Hash就是找到一种数据内容和数据存放地址之间的映射关系。

2、查找：哈希表，又称为散列，是一种更加快捷的查找技术。我们之前的查找，都是这样一种方式：集合中拿出来一个元素，看看是否与我们要找的相等，如果不等，缩小范围，继续查找。而哈希表是完全另外一种思路：当我知道key值以后，我就可以直接计算出这个元素在集合中的位置，根本不需要一次又一次的查找！

举个例子，假如我的数组A中，第i个元素里面装的key就是i，那么数字3肯定是在第3个位置，数字10肯定是在第10个位置。哈希表就是利用这种基本的思想，建立一个从key到位置的函数，然后进行直接计算查找。

3、Hash表在海量数据处理中有着广泛应用。

Hash Table的查询速度非常的快，几乎是O(1)的时间复杂度。

hash就是找到一种数据内容和数据存放地址之间的映射关系。

散列法：元素特征转变为数组下标的方法。

我想大家都在想一个很严重的问题：“如果两个字符串在哈希表中对应的位置相同怎么办？”，毕竟一个数组容量是有限的，这种可能性很大。解决该问题的方法很多，我首先想到的就是用“链表”。我遇到的很多算法都可以转化成链表来解决，只要在哈希表的每个入口挂一个链表，保存所有对应的字符串就OK了。

散列表的查找步骤

当存储记录时，通过散列函数计算出记录的散列地址

当查找记录时，我们通过同样的是散列函数计算记录的散列地址，并按此散列地址访问该记录

关键字——散列函数（哈希函数）——散列地址

优点：一对一的查找效率很高；

缺点：一个关键字可能对应多个散列地址；需要查找一个范围时，效果不好。

散列冲突：不同的关键字经过散列函数的计算得到了相同的散列地址。

好的散列函数=计算简单+分布均匀（计算得到的散列地址分布均匀）

哈希表：一种数据结构，它可以提供快速的插入操作和查找操作。

优缺点

优点：不论哈希表中有多少数据，查找、插入、删除（有时包括删除）只需要接近常量的时间即O(1)的时间级。实际上，这只需要几条机器指令。

哈希表运算得非常快，在计算机程序中，如果需要在一秒钟内查找上千条记录通常使用哈希表（例如拼写检查器）哈希表的速度明显比树快，树的操作通常需要O(N)的时间级。哈希表不仅速度快，编程实现也相对容易。

如果不需要有序遍历数据，并且可以提前预测数据量的大小。那么哈希表在速度和易用性方面是无与伦比的。

**缺点：**它是**基于数组**的，数组创建后难于扩展，某些哈希表被基本填满时，性能下降得非常严重，所以程序员**必须要清楚表中将要存储多少数据**（或者准备好定期地把数据转移到更大的哈希表中，这是个费时的过程）。



元素\*\*转变为数组下标的方法就是散列法。散列法当然不止一种，下面列出三种比较常用：

1，除 列法

最直观的一种，上图使用的就是这种散列法，公式：

$$\text{in} = \text{value} \% 16$$

学过汇编的都知道，求模数其实是通过一个除法运算得到的，所以叫“除法散列法”。

2，平方散列法

求index是非常频繁的操作，而乘法的运算要比除法来得省时（对现在的CPU来说，估计我们感觉不出来），所以我们考虑把除法换成乘法和一个位移操作。公式：

$$\text{index} = (\text{value} * \text{value}) >> 28$$
（右移，除以 $2^{28}$ 。记法：左移变大，是乘。右移变小，是除。）

如果数值分配比较均匀的话这种方法能得到不错的结果，但我上面画的那个图的各个元素的值算出来的index都是0——非常失败。也许你还有个问题，value如果很大，value \* value不会溢出吗？答案是会的，但我们这个乘法不关心溢出，因为我们根本不是为了获取相乘结果，而是为了获取index。

3，斐波那契（Fibonacci）散列法

平方散列法的缺点是显而易见的，所以我们能不能找出一个理想的乘数，而不是拿value本身当作乘数呢？答案是肯定的。

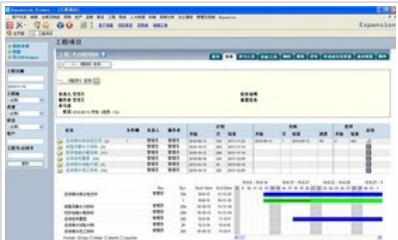
- 1，对于16位整数而言，这个乘数是40503
- 2，对于32位整数而言，这个乘数是2654435769
- 3，对于64位整数而言，这个乘数是11400714819323198485

这几个“理想乘数”是如何得出来的呢？这跟一个法则有关，叫黄金分割法则，而描述黄金分割法则的最经典表达式无疑就是著名的斐波那契数列，即如此形式的序列：0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 2584, 4181, 6765, 10946, ...。另外，斐波那契数列的值和太阳系八大行星的轨道半径的比例出奇吻合。

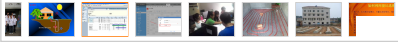
对我们常见的32位整数而言，公式：

$$\text{index} = (\text{value} * 2654435769) >> 28$$

如果用这种斐波那契散列法的话，那上面的图就变成这样了：



项目管理 软件



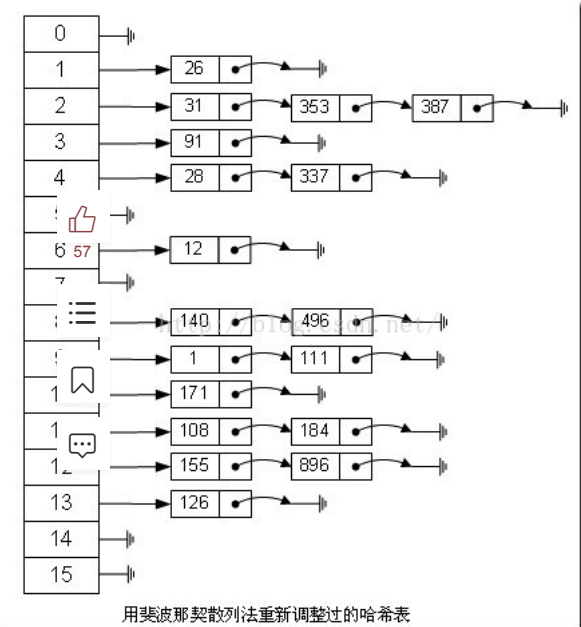
联系我们



请扫描二维码联系客服  
webmaster@csdn.net  
400-660-0108  
QQ客服 客服论坛

关于 招聘 广告服务 百度  
©1999-2018 CSDN版权所有  
京ICP证09002463号

经营性网站备案信息  
网络110报警服务  
中国互联网举报中心  
北京互联网违法和不良信息举报中心



用斐波那契散列法重新调整过的哈希表

注：用斐波那契散列法调整之后会比原来的取模散列法好很多。

适用范围

快速查找，删除的基本数据结构，通常需要总数据量可以放入内存。

基本原理及要点

hash函数选择，针对字符串，整数，排列，具体相应的hash方法。

碰撞处理，一种是open hashing，也称为拉链法；另一种就是closed hashing，也称开地址法，opened addressing。

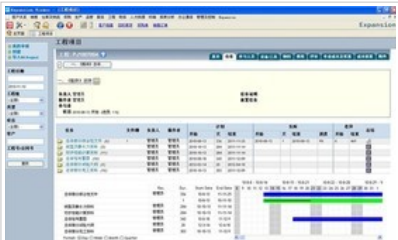
散列冲突的解决方案：

- 1.建立一个缓冲区，把凡是拼音重复的人放到缓冲区中。当我通过名字查找人时，发现找的不，就在缓冲区里找。
- 2.进行再探测。就是在其他地方查找。探测的方法也可以有很多种。
  - (1)在找到查找位置的index的index-1，index+1位置查找，index-2，index+2查找，依次类推。这种方法称为线性再探测。
  - (2)在查找位置index周围随机的查找。称为随机在探测。
  - (3)再哈希。就是当冲突时，采用另外一种映射方式来查找。

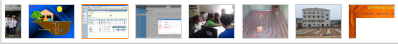
这个程序中是通过取模来模拟查找找到重复元素的过程。对待重复元素的方法就是再哈希：对当前key的位置+7。最后，可以通过全局变量来判断需要查找多少次。我这里通过依次查找26个英文字母的小写计算的出了总的查找次数。显然，当总的查找次数/查找的总元素数越接近1时，哈希表更接近于一一映射的函数，查找的效率更高。

扩展

d-left hashing中的d是多个的意思，我们先简化这个问题，看一看2-left hashing。2-left hashing指的是将一个哈希表分成长度相等的两半，分别叫做T1和T2，给T1和T2分别配备一个哈希函数，h1和h2。在存储一个新的key时，同时用两个哈希函数进行计算，得出两个地址h1[key]和h2[key]。这时需要检查T1中的h1[key]位置和T2中的h2[key]位置，哪一个位置已经存储的（有碰撞的）key比较多，然后将新key存储在负载少的位置。如果两边一样多，比如两个位置都为空或者都存储了一个key，就把新key存储在左边的T1子表中，2-left由此而来。在查找一个key时，必须进行两次hash，同时查找两个位置。



项目管理 软件



联系我们



请扫描二维码联系客服  
webmaster@csdn.net  
400-660-0108  
QQ客服 客服论坛

关于 招聘 广告服务 百度  
©1999-2018 CSDN版权所有  
京ICP证09002463号

经营性网站备案信息  
网络110报警服务  
中国互联网举报中心  
北京互联网违法和不良信息举报中心



问题实例（海量数据处理）

我们知道hash 表在海量数据处理中有着广泛的应用，下面，请看另一道百度面试题：  
题目：海量日志数据，提取出某日访问百度次数最多的那个IP。  
方案：IP的数目还是有限的，最多2^32个，所以可以考虑使用hash将ip直接存入内存，然后进行统计。



版权声明：本文为@那年聪聪 原创文章，未经博主允许不得转载。  
http://blog.csdn.net/duan19920101/article/details/51579136



严禁讨论涉及中国之军/政相关话题，违者会被禁言、封号！



yanwenwennihao2018-03-19 10:59

回复11楼

这个总结太详细了，谢谢博主分享



Bruce\_\_0002018-02-04 11:48

2条回复 回复10楼

“哈希表hashtable(key, value) 就是把Key通过一个固定的算法函数既所谓的哈希函数转换成一个整型数字，然后就将该数字对数组长度进行取余，取余结果就当作数组的下标，将value存储在以该数字为下标的数组空间里。” 如果key值一个是4一个是7 而数组长度是3 取余的话不就一样了吗？



qq\_407785362018-01-20 10:11

1条回复 回复9楼

“好的散列表=计算简单+分布均匀”，可是怎么能保证“分布均匀”呢？

查看 17 条热评

哈希表的使用



wqf3632007年03月12日 17:30

19189

数据结构：hash\_map原理 这是一节让你深入理解hash\_map的介绍，如果你只是想囫圇吞枣，不想理解其原理，你倒是可以略过这一节，但我还是建议你看看，多了解一些没有坏处。hash\_map基于ha...

重温数据结构：哈希 哈希函数 哈希表



u0112408772016年10月27日 00:49

6406

点击查看 Java 集合框架深入理解 系列， - ( ° - ° )つ口 乾杯~ 在学习 HashMap 前，我们先来温习下 Hash(哈希)的概念。什么是 HashHash（哈希），又称“散列” ...

有眼睛的扫地机器人，看得见才能不漏扫

首发预定，减200元

哈希表



ligenyingsr2013年09月18日 15:40

2445

\* 若结构中存在关键字和K相等的记录，则必定存储在f(K)的位置上。由此，不需比较便可直接取得所查记录。这个对应关系称为散列函数(Hash function)，按这个思想建立的表为散列表。 ...

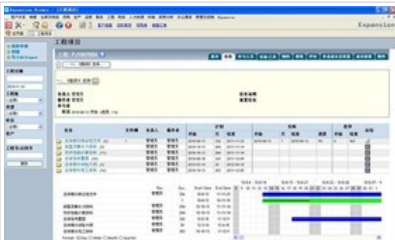
哈希表总结



chenhuajie1232013年07月01日 00:32

32477

哈希表的概念 哈希表(Hash Table)也叫散列表，是根据关键码值（Key Value）而直接进行访问的数据结构。它通过把关键码值映射到哈希表中的一个位置来访问记录，以加快查找的...



项目管理 软件



联系我们



请扫描二维码联系客服  
webmaster@csdn.net  
400-660-0108  
QQ客服 客服论坛

关于 招聘 广告服务 百度  
©1999-2018 CSDN版权所有  
京ICP证09002463号

经营性网站备案信息  
网络110报警服务  
中国互联网举报中心  
北京互联网违法和不良信息举报中心

一步一步写算法（之hash表）

feixiaoxing 2011年10月18日 18:43 130641

【声明：版权所有，欢迎转载，请勿用于商业用途。 联系信箱：feixiaoxing @163.com】 hash表，有时候也被称为散列表。个人认为，hash表是介于链表和二叉树之间...

IT人一秒懂升职加薪的奥义

动动鼠标就能实现的新购满额减再返最高6000



海量数据管理——分治和hash映射

yangquanhui1991 2016年08月10日 16:06 5940

什么是Hash Hash，一般翻译做“散列”，也有直接音译为“哈希”的，就是把任意长度的输入（又叫做预映射，pre-image），通过散列算法，变换成固定长度的输出，该输出就是散列值。这...

哈希表心得

haoel 2003年04月02日 09:02 31392

最近我在做一个项目，其中要用到一个数据结构——Hash Table（哈希表），以前只有理论知识，现在却发现很不简单，所以写下来和大家共分享。我们知道，哈希表是一个固定大小的数组，数组的每个元素是一个...

分布式缓存--系列1 -- Hash环/一致性Hash原理

当前，Memcached、Redis这类分布式kv缓存已经非常普遍。从本篇开始，本系列将分析分布式缓存相关的原理、使用策略和最佳实践。我们知道Memcached的分布式其实是一种“伪分布式”，也就是它...

chunlongyu 2016年11月25日 01:26 2099

排序算法系列：基数排序

u013761665 2016年06月16日 23:14 20953

今天要说的这个排序算法很特殊，它不需要直接对元素进行相互比较，也不需要将元素相互交换，你需要做的就是对元素进行“分类”。这也是基数排序的魅力所在，基数排序可以理解成是建立在“计数排序”的基础之上的一种...

哈希表

lcj110011 2014年05月22日 16:12 499

以前对哈希表都是半懂不懂的，

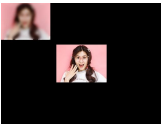
哈希表的实现 除留余数法

hmh86 2017年01月02日 11:31 2971

查找有两种方式，比较式查找和计算式查找，而计算式查找则通过哈希表来实现。给定表M，存在函数f(key)，对任意给定的关键字值key，代入函数后若能得到包含该关键字的记录在表中的地址，则称表M为哈希(H...

码农不会英语怎么行？英语文档都看不懂！

软件工程出身的英语老师，教你用数学公式读懂天下英文→



哈希表的创建

sanqima 2015年10月06日 22:28 4657

在记录的存储位置和它的关键字之间建立一个确定的对应关系f，使每个关键字和表中唯一的存储位置相对应，称这个对应关系f为哈希(散列)函数，根据这个思想建立的表为哈希表。 若key1≠key2, 而f...

HashTable和HashMap的区别详解

fujiakai 2016年06月04日 19:35 53136

一、HashMap简介 HashMap是基于哈希表实现的，每一个元素是一个key-value对，其内部通过单链表解决冲突问题，容量不足（超过了阈值）时，同样会自动增长。 Has...

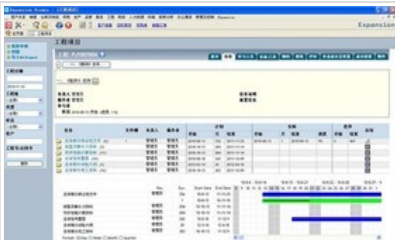
哈希表的详解

yzl\_rex 2012年08月16日 10:03 5944

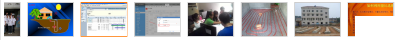
哈希表（Hash table，也叫散列表），是根据关键码值(Key value)而直接进行访问的数据结构。也就是说，它通过把关键码值映射到表中一个位置来访问记录，以加快查找的速度。这个映射函数叫做散列...

哈希表

luobo140716 2017年01月14日 20:28 230



项目管理 软件



联系我们



请扫描二维码联系客服  
webmaster@csdn.net  
400-660-0108  
QQ客服 客服论坛

关于 招聘 广告服务 百度  
©1999-2018 CSDN版权所有  
京ICP证09002463号

经营性网站备案信息  
网络110报警服务  
中国互联网举报中心  
北京互联网违法和不良信息举报中心

哈希表无论是在面试中，还是在日常编程中，都有着举足轻重的地位，我们虽然不用完完全全自己去构建一个哈希表的数据结构，但是也应该知道哈希表是什么，它的原理是什么，它有什么好处等等这些内容 哈希表是什...

散列表 哈希表 原理 python实现

waltonhuang 2016年08月26日 09:44 3145

参考算法导论引文 散列表(hash table)是实现字典操作的一种有效的数据结构。尽管最坏的情况下，散列表中查找一个元素的链表中查找的时间相同，达到了O(n)。然而实际应用中，散列的查找的...



57

简单介绍了哈希表作用及程序举例

mark20170902 2016年07月15日 10:07 540

转自：http://blog.csdn.net/songzi1111/article/details/10985429 转自：http://blog.csdn.net/yleek/article/d...



达内可告

有在达内...的吗

百度广告

处理哈希冲突的线性探测法

leex\_brave 2016年05月24日 15:28 8588

哈希表，是根据关键字（Key value）而直接访问在内存存储位置的数据结构。也就是说，它通过计算一个关于键值的函数，将所需查询的数据映射到表中一个位置来访问记录，这加快了查找速度。这个映射函数称做散...

哈希表入门题目总结(HDU 1280、1425、2027、3833、1496、2648 POJ 1...

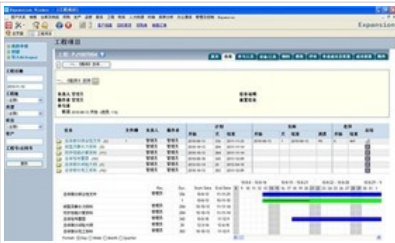
今天做了一天的哈希表题目，虽然不是哈希算法，但是感觉先把这哈希表搞定了，再学哈希算法的时候会快一些，所以今天搞得现在都有点头痛了，看了一天的电脑..... 快要睡觉了，整理一下今天的成果。整理一下题目与思想...

u011466175 2013年12月22日 01:07 1492

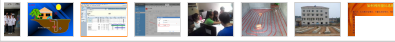
散列表（哈希表）查找算法

xiaoxiaoxuewen 2012年05月15日 22:46 6901

散列方法不同于顺序查找、二分查找、二叉排序树及B-树上的查找。它不以关键字的比较为基本操作，采用直接寻址技术。在理想情况下，无须任何比较就可以找到待查关键字，查找的期望时间为O(1)。散列表的...



项目管理 软件



联系我们



请扫描二维码联系客服  
webmaster@csdn.net  
400-660-0108  
QQ客服 客服论坛

关于 招聘 广告服务 百度  
©1999-2018 CSDN版权所有  
京ICP证09002463号

经营性网站备案信息  
网络110报警服务  
中国互联网举报中心  
北京互联网违法和不良信息举报中心