

11 词典 | 11A 散列 Hashing

#数据结构邓神

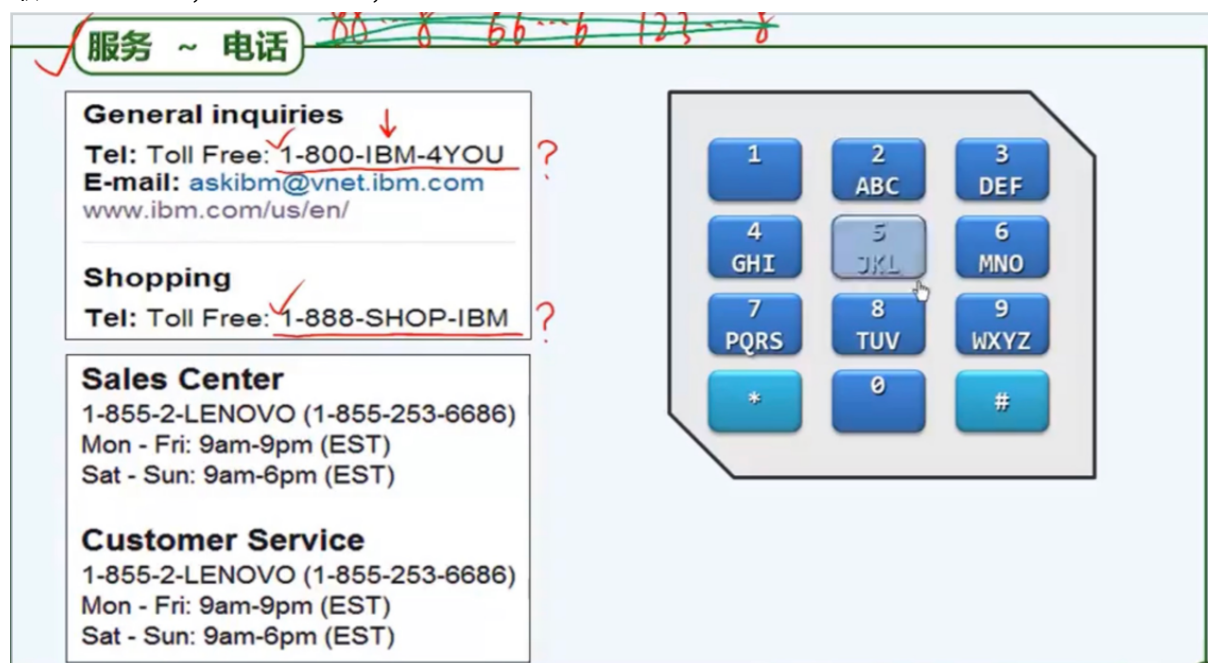
从服务到电话

当你需要获得每一个公司和电话，你应该查电话簿

没有电话簿怎么办？

为他们的服务取一个更加便于记忆的电话号码

取 88888888 , 66666666 , 12345678



来一看IBM电话

循值访问

Call By Rank : Vector

Call By Position : List

Call By Key : BST

Call By Value : Dictionary

电话簿

电话簿

❖ 需求：为一所学校制作电话簿

? 号码 \sim 个人（教员、学生、员工）或办公室

❖ 蛮力：使用数组，按电话号码索引

时间 = $O(1)$ ✓

❖ 以清华为例（2003）

✓ #可能的电话 = $R = 10^8 = 100M$ ✓

#实有的电话 = $N = 25,000 = 25K$

0000.0000	00000000
...	...
5153.xxxx	5153xxxx
...	...
6278.xxxx	6278xxxx
...	...
9999.9999	99999999

ZJ-xx#xxx

xx办公室

❖ 问题

空间 = $O(R + N) = O(100M + 25K)$

效率 = $25K / 100M = 0.025\%$

其它方面...

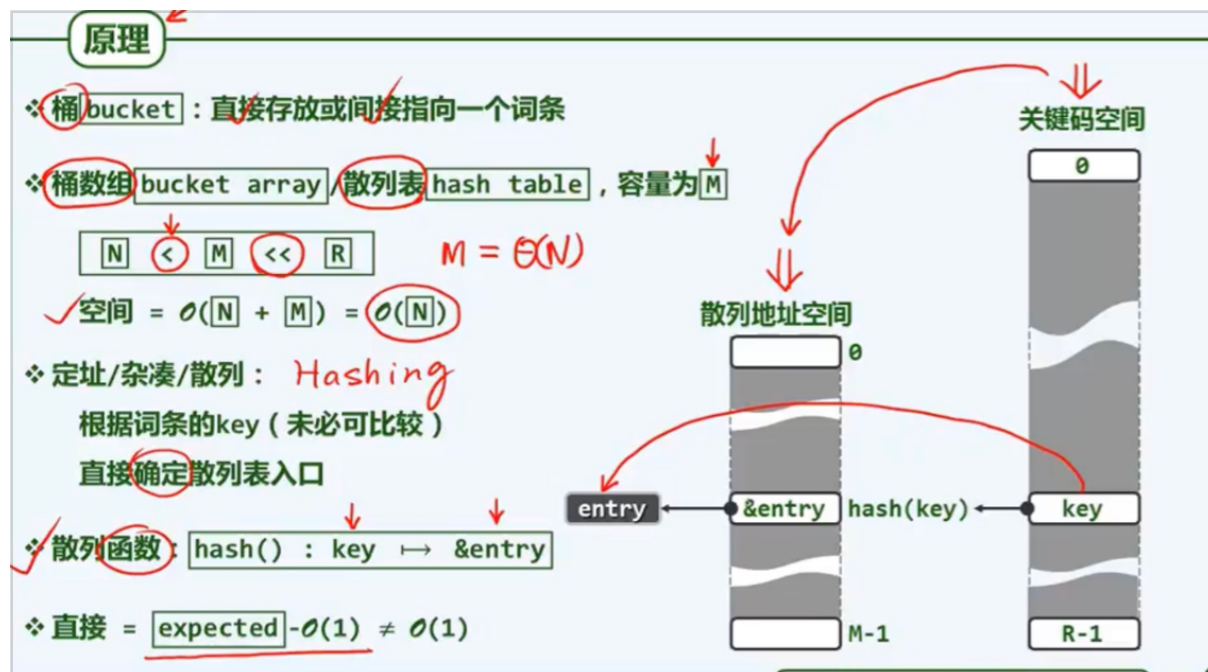
使用 8 维数组 直接使用 电话号码作为秩 时间 $O(1)$

空间复杂度：??? 你当我不存在是吗

$N \ll R$

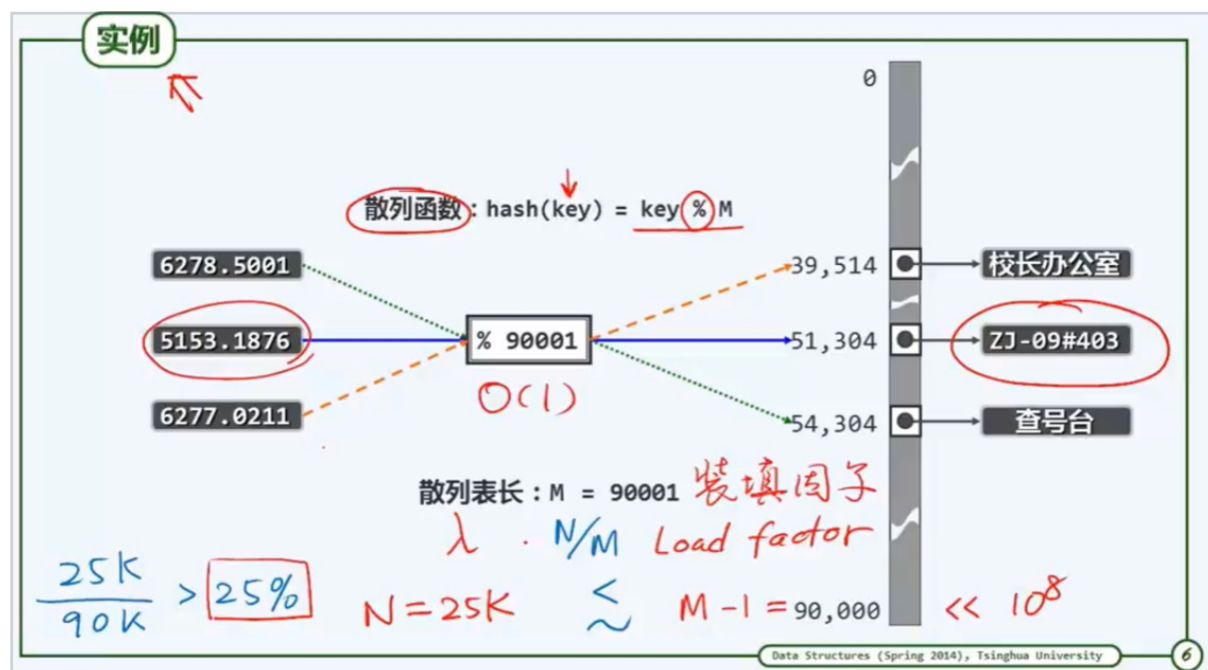
所需要处理的数据要远远小于全集

原理



散列表应该足以容纳所有的待存放的元素

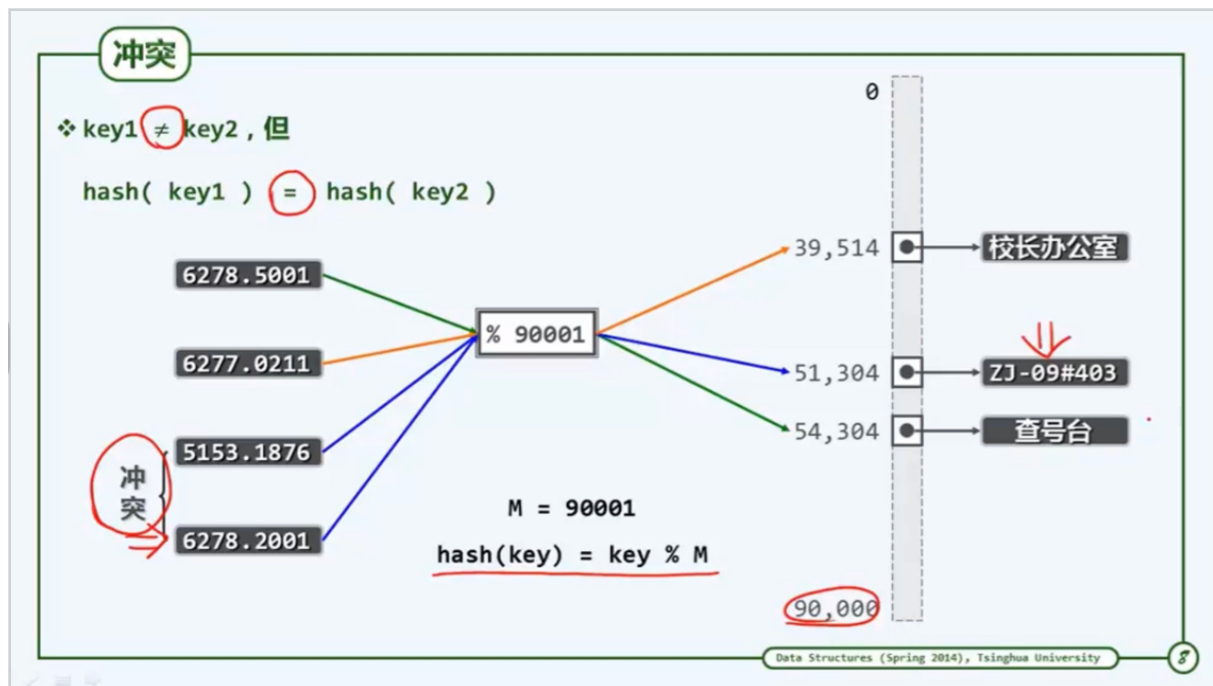
实例



空间利用率上升至至少为 1/4

时间利用率还是为 $O(1)$

冲突



冲突是无法杜绝的!

因为散列函数是将 R 映射到 N 而 $R \gg N$

所有冲突是一定会发生的!

只要策略和方法得到, 我们可以将冲突控制的非常好