

MORE HASH TABLES

http://en.wikipedia.org/wiki/Hash_table

Primary Clustering

- Primary Clustering - this implies that all keys that collide at address b will extend the cluster that contains b

Quadratic Probing

- With quadratic probing we try to overcome the problem of clustering.
- So, when a collision occurs while trying to insert an item in the table, instead of looking at every cell until an empty one is found, a function is applied to find an empty cell.
- If $h(n)$ is occupied, try
 - $(h(n) + 1^2) \bmod \text{table-size}$,
 - $(h(n) + 2^2) \bmod \text{table-size}$,
 - and so on until an empty cell is found.
- Quadratic probing works well if the size of the table is a prime number and the table is less than half full. Quadratic probing may not get anywhere.

Quadratic Probing

- $f(i) = i^2$
- Example:
 $h(Kn) = n \% 11$
- Insert
M13
G7
Q17
Y25
R18
Z26
F6

Bucket	Data
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Secondary Clustering!

Secondary Clustering

- Secondary Clustering - is when adjacent clusters join to form a composite cluster

Re-hashing

- With rehashing, if a collision occurs while inserting a new data item into the hash table, then a second hash function is applied to the result of the first hash function to find an empty cell in the table.
- The re-hashing function can either be a new function or a re-application of the original one. As long as the functions are applied to a key in the same order, then a sought key can always be located.
- The second hash function has to be chosen with care:
 - The sequence should be able to visit all slots in the table.
 - The function must be different from the first to avoid clustering.
 - It must be very simple to compute.

Double Hash

- $f(i) = h_2(k) * i$
 - $h_2(k)$ is some second hash function
 - unique probe sequence for every key
- $bucket = (h(K) + h_2(K) * i) \% N$
- $h_2(k)$ should be relatively prime to N for all k
 - don't produce zero
- Example
 - $h(k) = k \% N$
 - $h_2(k) = 1 + (k \% (N - 1))$

Chaining

- Chaining is also called the bucket approach.
- It differs from the collision avoidance already discussed in that instead of each cell holding a single data item, several data items can be stored in the form of a linked list, and only the header of the linked list is placed in the table.
- The disadvantages of chaining include implementing a separate data structure, and dynamically allocating memory.

Chaining (Open Hashing)

- Each bucket is the head of a linked list
 - if you hash a key to a bucket, insert the data into the list
 - insert at front, back, or in sorted order.
 - why would this decision matter?

Problem

- Hash the keys M13, G7, Q17, Y25, R18, Z26, and F6 using the hash formula $h(Kn) = n \bmod 9$ with the following collision handling technique: (a) linear probing, (b) chaining
- Compute the average number of probes to find an arbitrary key K for both methods.
- $\text{avg} = (\text{summation of the \# of probes to locate each key in the table}) / \# \text{ of keys in the table}$