Unit-III. Hash Tables

Hash Tables : Introduction, Hash Structure, Hash functions, Linear Open Addressing, Chaining and Applications.

Hash Table :-

The hash table contains key values with pointers to the corresponding records. Basic idea of hash table is that we have to place a key value into a location in the hash table. The location will be calculated from the key value it self.

This 1-to-1 correspondence b/w a key value and an index in the hash table is known as hashing (or) address calculation indexing.

Hashing Techniques :-

The main idea of hashing techniques is to find a 1-to-1 correspondence b/w a key value and an index in the hash table. where the key values can be placed.

It may be noted that the mapping is subjective. That is all key values are maped in to sum indexes and more than one key value may be maped into an index value. the function that defines this maping is called as hash function. ($H : K \to I$).

The hash function place a deminent role in hashing techniques. There are two principle criterias in

deciding a hash function.

(i) The function H should be very easy and quick to compute.

(ii) The function H should as far as possible give two different indices for two different key values.

Ex:- Hash Table

Key	Index
10	1
19	0
35	8
43	7
62	8
59	4
31	4
49	3
77	4
33	6

Index	Key
0	19
1	10
2	—
3	49
4	59,31,77
5	—
6	33
7	43
8	35,62
9	—

From the above hash table the hash functions does not distribute uniformly, ~~over~~ Sum entryies are empty and sum entries have more than one key value. Allotment of more than one key value in one location in the hash table is called collision. To remove these collision in the hash table we use various hash techniques.

Division Method:-

The division method is defined as follows..

A number h is larger than the number of keys in K. The hash function H is defined as

→ $H(k) = k \pmod h$: if index starts from 0.

→ $H(k) = k \pmod h + 1$: if index starts from 1.

where $k \in K$, a key value the operation MOD defines the modulo arithmetic operation which is equal to the remainder of driving k/h.

For example,

$k = 31$ and $h = 13$

$H(k) = k \pmod h$

$K(31) = 31 \pmod{13} = 5$

The number h is usually a prime number (or a number without small division and equal to the size of hash table.

Mid Square Method:-

The mid-square method is defined as follows. The hash function H is defined as $H(K) = x$, where x is obtained by selecting an appropriate number of bits (or digits) from the middle of the square in the key value k. This selection depends on the size of the hash table. It need to be emphasized that the same criteria should be used for selecting the bits (or) digits for all the keys.

For example,

$H(k) = x$	k : 1234	2345	3456
$H(12) = 14$	k^2 : 1522756	5499025	11943936
$= 4$	$H(k)$: 525	492	1496
$= 14$			

The mid square method has been criticized because of time consuming computation. But it gives a good result.

Folding Method:-

The method can be defined as follows:

The partition key into no. of parts $k_1, k_2, \ldots, k_n$ where each part except last has the same number of bits (or digits) as the required address of width. Then the parts one added together. that is

$H(k) = k_1 + k_2 + \cdots + k_n$. where if the keys are in the binary form the exclusive OR operation may be submitted for addition.

* There are many variations in this method, one is called the folding shifting method where the even parts: i.e, $k_2, k_4 \ldots$ are each reversed before the addition.
* Another variation is called fold boundary method where two boundary parts. i.e, k_1 & k_n are each reversed then added to all the parts.

Ex: K: 15222756	5499025	11943936
Chopping: 01 52 27 56	05 49 90 25	11 94 39 36
Fold shifting: 01+25+27+65	05+94+90+52	11+94+39+63
Fold boundary: 10+52+27+65	50+49+90+52	11+94+39+63
Pure Folding: 01+52+27+56	05+49+90+25	11+94+39+36

Folding method is also useful in converting multiword-keys into a single word. So that another hashing function can be used on that.

Digit Analysis Method:-

The basic idea of this hashing function is to form host addresses by extracting or shifting the extracted digits or bits of the original key.

Eg: K: 1522756

It can be transfer to hash address 625 by extracting the digits in even position & then reverse this combination.

For a given set of keys, the position in this keys and the same rearrangement pattern must be used for consistantly. The decision for extraction & then rearrangement is based on some analysis.

Collision Resolution Techniques:-

There are several techniques are available to remove collision in hash table. There are 2 important techniques are available (1) Closed Hashing (Linear probing)
(2) Opened Hashing (Chaining)

Closed Hashing:-

The simplest method to resolve a collision is closed hashing. Suppose there is a hash table of size h & the key value is mapped to an address location i with hash function. The closed hashing can be started as follows

Start with the hash address the collision has occured let it be "i". Then follow the following sequence of locations in the hash table & do the sequential search,

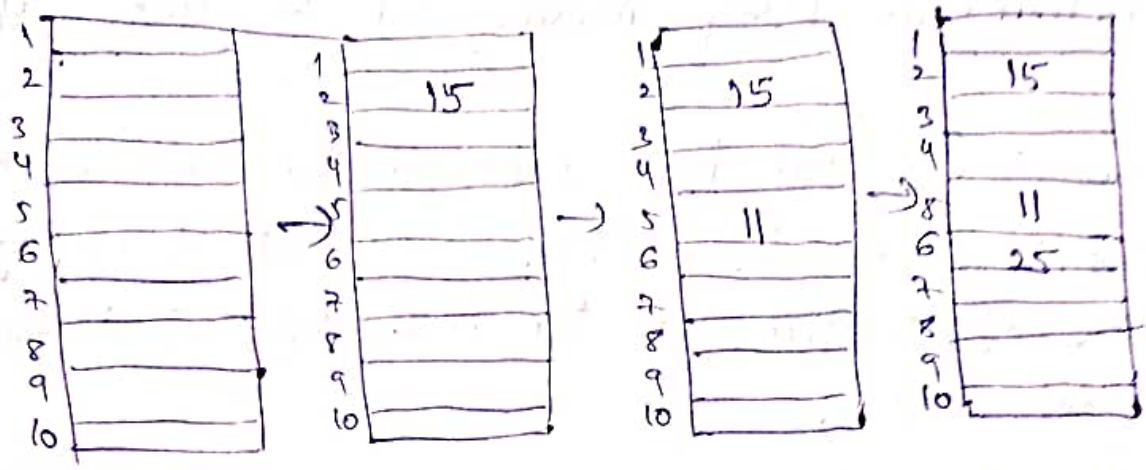
i, i+1, i+2, ---, h, 1, 2, ---, i-1

The search will continue until any of the following case occured.

(i) The key value is found
(ii) Any empty location is occured
(iii) The search reaches the location, where the search had started.

The first case defines to the successful search and the last two cases, defines to unsuccessful search. Here the hash table considered circular, so that when the location is reached the search proceeds to the first location. So this technique is called closed hash. Since, the technique search in a st.line it is also called linear probic. where probic means key comparisions. For example,

Assume that there is a hash table of size 10, the hash function uses the division method with remainder modulo 7. i.e, $H(k) = k \pmod 7 + 1$ and Consider the key value 15, 11, 25, 16, 9, 8, 12

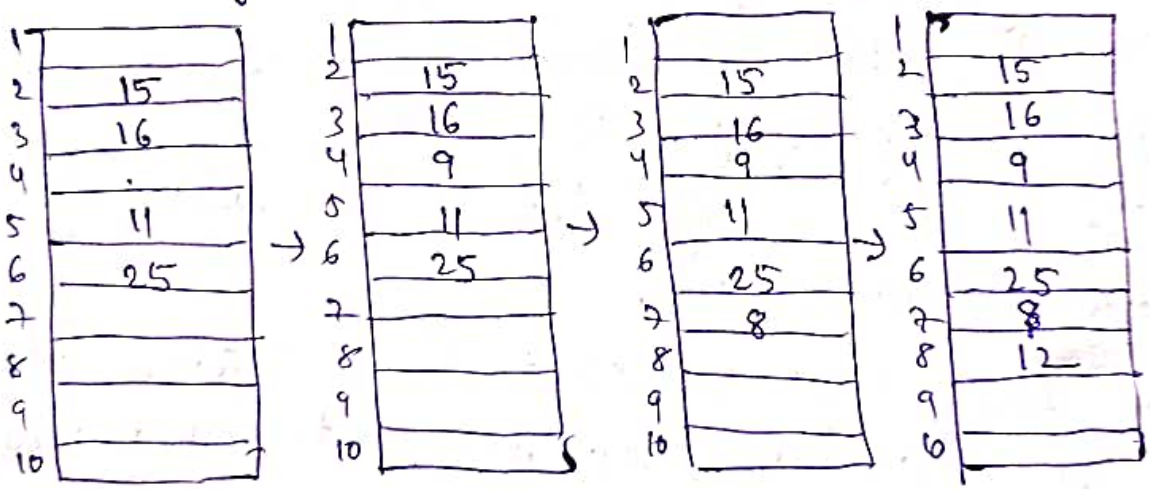


Initially the hash-table is empty

Insertion of 15

Insertion of 11

Insertion of 25



Insertion of 16

Insertion of 9

Insertion of 8

Insertion of 12

Drawbacks:-

The major drawback of closed hashing i.e, as half of the hash table is filled. There is tendency towards clustering. i.e, key values are clustered in

large group and as a result a sequential search becomes slower and slower. This kind of clustering is known as clustering.

The following are some solutions known to avoid this situation.

(i) Random Probing
(ii) Double Hashing (or) Rehashing
(iii) Quadratic Probing

(i) Random Probing:-

This method uses a pseudo random number generator to generate a random sequence of locations rather than an ordinary sequence as in the case of linear probing method.

The random sequence generated by the pseudo random number generator contains all the positions b/w 1 & h the highest locations of the hash table i.e, $i = (i+m) (\text{Mod } h) + 1$ where 'i' is a number in the sequence, m & h are integers and they are relatively prime to each other.

For example, $m=5, h=7, i=2$

The above mentioned pseudo random number generator generates a sequence as 8, 3, 9, 4, 10, 5, 11, 6, 1, 7, 2.

Here it stops producing the number when the 1st location is duplicated.

Double Hashing:-

If the same sequence of location is generated for two different keys by the random probing method then clustering may takes place. This kind of cluster is known as secondary clustering. An alternative approach to avoid the secondary clustering problem is to use a second hash function in addition to the first one.

This second hash function result in the value of m for the pseudo random number generator as employed in the random probing method.

This second function should be selected in such a way that the hash addresses generated by the two m & h are relatively prime. For example,

$H_1(k)$ is the initially used hash function &
$H_2(k)$ is the second one then the functions are

$H_1(k) = k \pmod h + 1$

$H_2(k) = k \pmod{(h-4)} + 1$

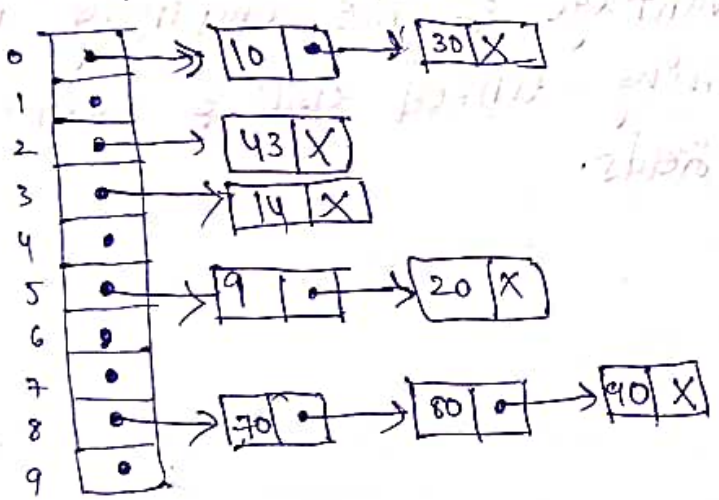
Quadratic Hashing:-

Quadratic probing is a collision resolution method that eliminates the primary clustering problem of linear probing. For linear probing if there is a collision at location 'i' then the next locations $i+1, i+2, i+3, \ldots, i+n$ are probed but in

quadratic probing the next locations to be probed are $i+1^2$, $i+2^2$, $i+3^2$, ..., $i+n^2$. If h is the size of the hash table and H(k) is the hash function then the quadratic probing searches the location is $H(k) = (i^2 \bmod h)$ (for $i = 1, 2, 3, \ldots, n$)

Open Hashing :-

To resolve collision problem another hashing method is used i.e, Open hashing also called chaining. The chaining method is defined as follows. The chaining method uses a hash table as an array of pointers.
i.e, each pointer points to a linked list.



The hash addresses for any key is desided by its last digits. For a given key value the hash address is calculated. It then searches the linked list pointed by the pointers at that location. If the element is found it returns the pointer to the node containing that key value else insert the element at the end of that list.

Advantages and Disadvantages of chaining:

(i) An overflow situation never arises.

(ii) Collision resolution can be acheived very effectively. If the list maintain an order of keys. So that keys can be searched quickly.

(iii) Insertion & deletion becomes a quick and an easy task in open hashing. Deletion proceeds in the same way as deletion of a node in a single linked list finally open hashing is best suitable in application where the no. of key values varies open hash ~~xxxx~~ uses dynamic storage management policy.

The only disadvantage of the chaining method is that of maintaining linked list & extra storage space for link fields.