 **Bharatiya Vidya Bhavan’s**

**SARDAR PATEL INSTITUTE OF TECHNOLOGY**

(Autonomous Institute Affiliated to University of Mumbai)

Munshi Nagar, Andheri (W), Mumbai – 400 058.

COMPS Department

|  |  |
| --- | --- |
| **Experiment** | 9 |
| **Aim** | **Hashing Techniques: Write a program to implement Hash Table for the given input and solve the collision using using quadratic probing and linear probing and double hashing.** |
| **Name** | David Daniels |
| **UID** | 2023300038 |
| **Class** | Div -A |
| **Batch** | C |
| **Date of Submission** | 23-10-24 |

|  |  |
| --- | --- |
| **Theory** | **Hashing** is a technique used in data structures that efficiently stores and retrieves data in a way that allows for quick access. It involves mapping data to a specific index in a hash table using a **hash function** that enables fast retrieval of information based on its key. This method is commonly used in databases, **c**aching systems, and various programming applications to optimize search and retrieval operations. The great thing about hashing is, we can achieve all three operations (search, insert and delete) in O(1) time on average.  There are many applications of hashing,such as:   1. Database indexing: Hashing is used to index and retrieve data efficiently in databases and other data storage systems. 2. Dictionaries : To implement a dictionary so that we can quickly search a word 3. Password storage: Hashing is used to store passwords securely by applying a hash function to the password and storing the hashed result, rather than the plain text password. 4. Network Routing: Determining the best path for data packets 5. Cryptography: Hashing is used in cryptography to generate digital signatures, message authentication codes (MACs), and key derivation functions. 6. Load balancing: Hashing is used in load-balancing algorithms, such as consistent hashing, to distribute requests to servers in a network. 7. Blockchain: Hashing is used in blockchain technology, such as the proof-of-work algorithm, to secure the integrity and consensus of the blockchain. 8. Image processing: Hashing is used in image processing applications, such as perceptual hashing, to detect and prevent image duplicates and modifications. 9. File comparison: Hashing is used in file comparison algorithms, such as the MD5 and SHA-1 hash functions, to compare and verify the integrity of files. 10. Caching: Storing frequently accessed data for faster retrieval. For example browser caches, we can use URL as keys and find the local storage of the URL. 11. Symbol Tables: Mapping identifiers to their values in programming languages 12. Associative Arrays: Associative arrays are nothing but hash tables only. Commonly SQL library functions allow you retrieve data as associative arrays so that the retrieved data in RAM can be quickly searched for a key.   **Collisions and How to Handle Them**  Two or more keys can generate same hash values sometimes. This is called a collision. The new key should also be placed in the hash Somewhere. A collision can be handled using various techniques.  There are two major Types of Collision Resolution Techniques   1. **Separate Chaining Technique** 2. **Open Addressing technique**   **Open Addressing technique**  In this method, the values are all stored **in** the hash table itself. If collision occurs, we look for availability in the next spot generated by an algorithm. The table size at all times should be greater than the number of keys. *It is used when there is space restrictions, like in embedded processors.*  Point to note in delete operations, the deleted slot needs to be marked in some way so that during searching, we don’t stop probing at empty slots.  Types of Open Addressing:   1. Linear Probing: We linearly probe/look for next slot. If slot [hash(x)%size] is full, we try [hash(x)%size+1]. If that is full too, we try [hash(x)%size+2]…until an available space is found. Linear Probing has the best cache performance but downside includes primary and secondary clustering. 2. Quadratic Probing: We look for i²th iteration. If slot [hash(x)%size] is full, we try [(hash(x)+1\*1)%size]. If that is also full, we try [(hash(x)+2\*2)%size]…until an available space is found. Secondary clustering might arise here and there is no guarantee of finding a slot in this approach. 3. Double Hashing: We use a second hash function hash2(x) and look for i\*hash2(x) slot. If slot [hash(x)%size] is full, we try [(hash(x)+1\*hash2(x))%size]. If that is full too, we try [(hash(x)+2\*hash2(x))%size]…until an available space is found. No primary or secondary clustering but lot more computation here.   **Separate Chaining Technique**  The idea is to make each cell of the hash table point to a linked list of records that have the same hash function values. It is simple but requires additional memory outside the table. In this technique, the worst case occurs when all the values are in the same index or linked list, making the search complexity linear (n=length of the linked list). *This method should be used when we do not know how many keys will be there or how frequently the insert/delete operations will take place.*  To maintain the O(1) time of insertions, we make the new value as head of the linked list of the particular index.  Example of Separate Chaining Method | Download Scientific Diagram |
| **Algorithm** | 1. **Linear Probing** 2. **Calculate Initial Hash Index**:  * Use the hash\_fn function to compute the initial index rem for the given key based on the hash table's size.  1. **Check for Duplicate**:  * If the value at h->arr[rem] is equal to key, print a message indicating a duplicate exists, and exit the function.  1. **Check for Empty Slot**:  * If the status at h->status[rem] indicates it's empty ('E'):   + Insert key into h->arr[rem].   + Update the status at h->status[rem] to 'F' (indicating it's filled).  1. **Handle Collision**:  * If the slot is not empty, print a message indicating a collision has occurred. * Initialize a probe counter probe\_no starting at 1.  1. **Linear Probing**:  * Enter a loop that continues until an empty slot is found:   + Check again if h->arr[rem] is equal to key. If it is, print a duplicate message and exit.   + Increment rem to check the next slot (using modulo to wrap around the table).   + Print the current probe number and the new index being checked.   + Increment the probe\_no.  1. **Insert the Key**:  * Once an empty slot is found (when h->status[rem] is 'E'):   + Insert key into h->arr[rem].   + Update the status to 'F'.  1. **Quadratic Probing** 2. **Initialize Variables**:  * Create an ecount variable to keep track of the number of attempts made to insert the key. * Compute the initial index rem using the hash\_fn function.  1. **Check for Duplicate**:  * If the value at h->arr[rem] matches key, print a message indicating a duplicate exists, and exit the function.  1. **Print Initial Probe**:  * Print the details of the initial probing attempt using the calculated index rem.  1. **Check for Empty Slot**:  * If the status at h->status[rem] indicates it's empty ('E'):   + Insert key into h->arr[rem].   + Update the status at h->status[rem] to 'F'.  1. **Handle Collision with Quadratic Probing**:  * If the slot is occupied, print a message indicating a collision has occurred. * Initialize a temporary variable temp starting at 1.  1. **Quadratic Probing Loop**:  * Enter a loop that continues until an empty slot is found or the number of attempts (ecount) exceeds twice the size of the hash table:   + Check again if h->arr[rem] matches key. If it does, print a duplicate message and exit.   + Calculate the new index using the formula: rem = (rem + (temp \* temp)) % h->size.   + Print the current probing attempt details.   + Increment temp, and increase ecount by 1.  1. **Check for Infinite Loop**:  * If ecount exceeds h->size \* 2 - 2, print a warning message indicating potential infinite looping.  1. **Insert the Key**:  * Once an empty slot is found, insert key into h->arr[rem] and update the status to 'F'.  1. **Double Hashing** 2. **Calculate Initial Hash Index**:  * Compute the initial index rem for the given key using the hash\_fn function. * Store the original index in orem for later reference. * Calculate a secondary index rem2 using the hash\_fn1 function to assist in probing.  1. **Initialize Attempt Counter**:  * Set ecount to zero to keep track of the number of probing attempts.  1. **Check for Duplicate**:  * If the value at h->arr[rem] is equal to key, print a message indicating a duplicate exists and exit the function.  1. **Print Initial Probe Information**:  * Print details of the initial probing attempt using the calculated index rem.  1. **Check for Empty Slot**:  * If the status at h->status[rem] indicates it's empty ('E'):   + Insert key into h->arr[rem].   + Update the status at h->status[rem] to 'F' (indicating it's filled).  1. **Handle Collision with Double Hashing**:  * If the slot is occupied, print a message indicating a collision has occurred. * Initialize a temporary variable temp starting at 1.  1. **Probing Loop**:  * Enter a loop that continues until an empty slot is found or the number of attempts (ecount) exceeds twice the size of the hash table:   + Calculate the new index using the formula: rem = (orem + temp \* rem2) % h->size.   + Check if h->arr[rem] matches key. If it does, print a duplicate message and exit.   + Print the current probing attempt details, including the values used in the calculation.   + Increment temp and increase ecount by 1.  1. **Check for Infinite Loop**:  * If ecount exceeds h->size \* 2 - 2, print a warning message indicating potential infinite looping.  1. **Insert the Key**:  * Once an empty slot is found, insert key into h->arr[rem] and update the status to 'F'. |
| **Problem Solving** |  |
| **Program(Code)** | #include <stdio.h>  #include <stdlib.h>  #include <stdbool.h>  typedef struct hash  {      int \* arr;      int size;      char \* status;  }hash;  int hash\_fn (hash \* h ,int key)  {      return key%h->size;  }  void insert\_lin(hash \* h, int key)  {      int rem =hash\_fn(h,key);      printf("\n Applying Probe 0: (%d %% %d) %% %d = %d \n", key , h->size, h->size, rem);      if (h->arr[rem]==key)      {          printf("You Have a Duplicate: No need to insert: \n");          return;      }      if (h->status[rem]=='E')      {          h->arr[rem]=key;          h->status[rem]='F';      }      else      {          printf("A Collision has occured: \n");          int probe\_no=1;          //apply probing          while (h->status[rem]!='E')          {              if (h->arr[rem]==key)              {                  printf("You Have a Duplicate: No need to insert: \n");                  return;              }                rem= (rem+probe\_no)%h->size;              printf("Applying Probe %d: (%d+%d) %% %d = %d \n",probe\_no, key,probe\_no, h->size,  rem);              probe\_no++;          }          h->arr[rem]=key;          h->status[rem]='F';      }      printf("Element %d has been inserted at %d \n",key,rem);  }  void insert\_quad(hash \* h, int key)  {      int ecount=0;      int rem =hash\_fn(h,key);      if (h->arr[rem]==key)      {          printf("You Have a Duplicate: No need to insert: \n");          return;      }      int orem=rem;      printf("\n Applying Probe 0: (%d %% %d) %% %d = %d \n", key , h->size, h->size, rem);      if (h->status[rem]=='E')      {          h->arr[rem]=key;          h->status[rem]='F';      }      else      {          printf("A Collision has occured: \n");          int temp=1;          while (h->status[rem]!='E'&& ecount<h->size\*2)          {              if (h->arr[rem]==key)              {                  printf("You Have a Duplicate: No need to insert: \n");                  return;              }          rem=(rem + (temp\*temp))%h->size;          printf("Applying Probe %d: ( %d + %d ) %% %d = %d \n",temp,orem, temp\*temp , h->size,  rem);          temp++;          ecount++;          }          if (ecount>(h->size\*2)-2)          {          printf("Element %d is going in infinite loop: ",key);          return;          }          h->arr[rem]=key;          h->status[rem]='F';      }        printf("Element %d has been inserted at %d \n",key,rem);  }  int hash\_fn1 (hash \*h , int key)  {      return 7- ((key%5));  }  void insert\_doub(hash \* h, int key)  {      int rem =hash\_fn(h,key);      int orem=rem;      int rem2 = hash\_fn1(h,key);      int ecount=0;      if (h->arr[rem]==key)      {          printf("You Have a Duplicate: No need to insert: \n");          return;      }      printf("\n Applying Probe 0: (%d %% %d) %% %d = %d \n", key , h->size, h->size, rem);      if (h->status[rem]=='E')      {          h->arr[rem]=key;          h->status[rem]='F';      }      else      {          //apply probing          printf("A Collision has occured: \n");          int temp=1;          while (h->status[rem]!='E' && ecount<h->size\*2)          {              rem= (orem + temp\*rem2) %h->size;              if (h->arr[rem]==key)              {                  printf("You Have a Duplicate: No need to insert: \n");                  return;              }              printf("Applying Probe %d: ( %d + %d\*%d ) %% %d = %d \n",temp ,orem,temp,rem2,h->size,rem);              temp++;              ecount++;          }          if (ecount>(h->size\*2)-2)          {          printf("Element %d is going in infinite loop: \n",key);          return;          }          h->arr[rem]=key;          h->status[rem]='F';      }          printf("Element %d has been inserted at %d \n",key,rem);  }  void print\_hash(hash\* h)  {      printf("\n Index  Element Stored  Status \n");      for (int i = 0; i < h->size; i++)      {          printf("%d  \t    %d\t \t %c \n",i , h->arr[i], h->status[i]);      }  }  void choice(hash \* h)  {      printf("Enter 1 to Do linear probing \nEnter 2 to do Quadratic probing: \nEnter 3 to do Double Hashing \n");      int num=0;      scanf("%d",&num);      switch (num)      {      case 1:      {          printf("Welcome to linear probing \n Enter 1 to insert \n Enter 2 to print the entire Table \n Enter 0 to Exit \n");          while (true)          {              int choice=0;              printf("Enter choice \n");              scanf("%d",&choice);              if (choice==0)              {                  printf("\nGoodbye\n");                  break;              }              switch (choice)              {              case 1:              {                  printf("Enter number to insert: ");                  scanf("%d", &num);                  insert\_lin(h,num);                  break;              }              case 2:              {                  print\_hash(h);                  break;              }                default:              {                  printf("Error");                  break;              }              }          }            break;      }        case 2:      {          printf("Welcome to Quadratic probing \n Enter 1 to insert \n Enter 2 to print the entire Table \n Enter 0 to Exit \n");          while (true)          {              int choice=0;              printf("Enter choice \n");              scanf("%d",&choice);              if (choice==0)              {                  printf("\nGoodbye\n");                  break;              }              switch (choice)              {              case 1:              {                  printf("Enter number to insert: ");                  scanf("%d", &num);                  insert\_quad(h,num);                  printf("\n");                  break;              }              case 2:              {                  print\_hash(h);                  break;              }                default:              {                  printf("Error");                  break;              }              }          }          break;      }          case 3:      {          printf("Welcome to Double Hashing \n Enter 1 to insert \n Enter 2 to print the entire Table \n Enter 0 to Exit \n");          while (true)          {              int choice=0;              printf("Enter choice \n");              scanf("%d",&choice);              if (choice==0)              {                  printf("\nGoodbye\n");                  break;              }              switch (choice)              {              case 1:              {                  printf("Enter number to insert: ");                  scanf("%d", &num);                  insert\_doub(h,num);                  printf("\n");                  break;              }              case 2:              {                  print\_hash(h);                  break;              }                default:              {                  printf("Error");                  break;              }              }          }          break;        }      default:      {          printf("Error");          break;      }        }  }  int main(int argc, char const \*argv[])  {      hash \* h = (hash \* ) malloc(sizeof(hash));      printf("Enter size of the hash table: ");      int hash\_size=10;      scanf("%d",&hash\_size);        h->size=hash\_size;      h->arr=(int \*)malloc(h->size \*sizeof(int));      h->status=(char \*)malloc(h->size \*sizeof(char));      for (int i = 0; i < h->size; i++)      {          h->status[i]='E';          h->arr[i]=-1;      }      choice(h);      free(h->arr);      free(h->status);      free(h);      return 0;  } |
| **Output** | **LINEAR**        **Quadratic**                    **Double** |
| **Conclusion** | Thus, we implemented a hash table that employs three collision resolution techniques: linear probing, quadratic probing, and double hashing. Each method effectively manages collisions, but with different performance implications. Linear probing is simpler but may lead to clustering, while quadratic probing mitigates this issue through better distribution. Double hashing offers an efficient alternative by utilizing a secondary hash function. |