Task 2

What is collision?

A collision occurs when two different keys hash to the same index in a hash table.

why collision happens?

1- Pigeonhole Principle

This principle states that if n items are put into m containers, with n>m , at least one container must contain more than one item. Applied to hash tables, it means that if you have more keys than slots, some slots must contain more than one key.

2-Hash Function Characteristics

Even with a good hash function, collisions can still happen because it converts an input of arbitrary size to a fixed-size output (the hash value). This reduction in size means different inputs can produce the same output.

The possible techniques to solve the collision problem:

1-Separate Chaining

Separate chaining involves maintaining a list of all elements that hash to the same slot. For example: Linked Lists Each slot in the hash table contains a linked list of all elements that hash to the same slot.

2-Open Addressing

Open addressing handles collisions by finding another open slot within the hash table. The main techniques include:

Linear Probing: If a collision occurs, the algorithm checks the next slot in a linear sequence until an empty slot is found.

Quadratic Probing: Instead of checking the next slot linearly, the algorithm uses a quadratic function to determine the next slot, which reduces clustering compared to linear probing.

Double Hashing: It uses two hash functions to calculate the index and the step size for probing. This helps distribute keys more uniformly and reduces clustering, especially primary clustering.

How It Works

1. **Primary Hash Function:** Computes the initial index.
2. **Secondary Hash Function:** Computes the step size for probing.

When a collision occurs, the next index to check is determined by adding the step size computed by the secondary hash function to the current index. This process repeats until an empty slot is found or the desired key is located.

3-Robin Hood Hashing: It aims to balance the probe lengths of keys. The key idea is that during insertion, if a new key collides with an existing key, the new key will "steal" the spot if it is farther away from its original hashed position than the existing key. This helps to minimize the variance in probe lengths, leading to more uniformly distributed search times.

Key Concepts

1. **Probe Length:** The number of slots a key has to skip over before finding an empty slot or its intended position.

2. **Robin Hood Principle:** If a key being inserted has a greater probe length than the key at the current position, they swap places.

4-Cuckoo Hashing: This technique uses two hash functions and two hash tables. If a collision occurs in the first table, the element is moved to the second table, potentially displacing another element, which then gets rehashed into the first table, and so on.