Student ID: 1131530   **Student Name:** 李昌㛎   80
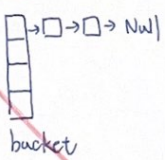
+40

Part A: Hash Table Definitions (Conceptual Understanding)

Q1. Define "collision" in the context of hash tables.

A1: 碰撞是指 2 個或多個 key 經過 hash function 被分到同個 index

Q2. What is a "bucket" in a hash table?

A2: 儲存位置

bucket

Q3. Define "load factor (α)" and explain why it affects performance.

A3: 商負載因子是 $\frac{n}{m}$, 代表 hash table 有多滿, 越滿會越多碰撞, 就會降低性能。
(n 數量, m table size)

Q4. What is "primary clustering," and which probing method suffers from it?

A4: 連續佔用的位置形成 clustering , 使新插入的需要更多時間找位置 , clustering 會越來越大。
受影響的方法 : linear Probing

Q5. What is "secondary clustering," and how is it different from primary clustering?

A5: 不同的 key . 如果經過 hash function 到同個 index . 會用相同的方式繼續排序
primary clustering 與 key 的初始雜湊值無關 . secondary clustering 有關 ,

Q6. Briefly explain the difference between:
- Open addressing
- Separate chaining

A6: Open addressing : 每個 bucket 儲存一個 linked list 處理 collision . (Array + linked list) → 結構
Separate chaining : 如果 index 發生 collision . 則 probe 其他位置 . (Array) → 結構 .

Part B: Hash Function Calculation (Collision & Pattern Observation)

Show your steps clearly.

Hash Function 1 — Division Method

$$h_1(k) = k \bmod 10$$

Hash Function 2 — Folding Method

Split key into two-digit chunks and sum the chunks.

$$h_2(k) = (sum\ of\ 2-digit\ groups) \bmod 11$$

Example:

Key = 8429 → groups: 84 + 29 → 113 → 113 mod 11 = 3

Q7. (Compute using Hash Function 1)

Given keys: 27, 37, 47, 57, 67

Compute their hash values using:

$$h_1(k) = k \bmod 10$$

A7: 27, 37, 47, 57, 67

→ 7, 7, 7, 7, 7 ✓

Q8. (Identify collision pattern)

From your results in Q1:

● What pattern do you observe?

● Explain why these keys collide.

A8: index 皆為7. 因為 m=10, 代表 index 每10次一循環. 所以相差10的 key 會碰撞

Q9. (Compute using Hash Function 2)

Compute $h_2(k)$ for: 1234, 9217, 4519, 9902

A9:
```
      92
      17        109
 109   99

1234 → 12+34 = 46 → 46 mod 11 = 2
9217 → 92+17 = 109 → 109 mod 11 = 10
4519 → 45+19 = 64 → 64 mod 11 = 9
9902 → 99+2 = 101 → 101 mod 11 = 2        A = 2, 10, 9, 2 ✓

      45
      19
      64   64
           55
```

Q10. (Compare distribution)

● Which hash function (h_1 or h_2) produced more collisions for the input set?

● Which seems to spread keys more <u>evenly</u>? 分散

● Provide 1–2 sentences of explanation.

A10: 1. h_1. 每10個連續數字一循環, 也就是會產生一樣的 index → 碰撞. 因為直接除.10

2. h_2. Folding Method 先進行一次前後先加的處理, 再除, 減少碰撞