Student ID: 1133317   Student Name: 吳杰恩   70

Part A: Hash Table Definitions (Conceptual Understanding)

+30

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

A1: 兩個 value 要 assign 到陣列中時 index 相同

(hash function 產生之 hash value)

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

A2: 代表一個分割的區塊

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

A3: 能代表資料的離散程度，一般而說 ≤ 0.7 為佳

$0 \leq \alpha \leq 1$

+5 太高資料擁擠易發生 collision，太低則浪費空間。

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

A4: linear probing 易造成 primary clustering，也就是連續幾個 index 都被佔用的情悅下，必須不斷往下找才找得到空位。

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

A5: 碰撞後又再碰撞一次，

+5 使用 probing 重新計算新的 hash value
secondary clustering

Q6. Briefly explain the difference between:

- Open addressing *Array
- Seperate chaining Array + Linked list

A6: 在記憶體空間分配上有很大不同

Open Adress 用大 Array 包住所有 hash values，可能會有資料過於分散導致效率低

b→a→b→a seperate chaining 多了 index 來在不連續記憶體中排序

hash value
[0|...] | | |99991

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 mod 10 = 7   37 mod 10 = 7   47 mod 10 = 7   57 mod 10 = 7   67 mod 10 = 7

Q8. (Identify collision pattern)

From your results in Q1 7

- What pattern do you observe?
- Explain why these keys collide.

A8: 1. 所有 key 的 hash value 都是 7 發生碰撞
2. hash function 設計過於簡單, 無法達到將資料分散的效果。(只有 0~9 size 小)
mod 10 只要尾數相同必撞, 應改成 mod 一質數   index

Q9. (Compute using Hash Function 2)

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

A9: $h_2(1234)$ = 12 + 34 = 46 $h_2(9217)$ = 92 + 17 = 109, $h_2(4519)$ = 45 + 19 = 64

$h_2(9902)$ = 99 + 2 = 101

Q10. (Compare distribution)

- Which hash function (h_1 or h_2) produced more collisions for the input set?
- Which seems to spread keys more evenly?
- Provide 1–2 sentences of explanation.

A10: 1. Hash Function 1
2. Hash Function 2

2-2 這個 function 具有將大數化小的作用, 藉由四則運算(相加)相較 mod 更具隨機性