

Student ID: 1121411

Student Name: 曾詠琪

Part A: Hash Table Definitions (Conceptual Understanding)

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

A1: 多個不同 key 的值算出相同的 hash value.

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

A2: 當不同的 value 在 hash table 上有同樣位置時, 這些 value 就在同一個 bucket 中.

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

A3: 空間使用率, 當 load factor (α) = 1 的時候表示 hash table 是滿的, 取數值越大表示要做例如 insert 的操作較費時

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

A4: ① 在一段連續放置的 data 後要找 slot 放, 即使算出來的 index 不同也被迫往下找其他 slot 放, 導致 data 都聚集在某一點。
② Linear probing

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

A5: ① 不同的值透過 hash function 計算出的 index 相同, 又透過一樣的找 slot 邏輯, 使得放入的 data 聚集在一起
② secondary clustering 是因為 hash function 所造成的

Q6. Briefly explain the difference between:

- Open addressing
- Separate chaining

A6: open addressing: 只用一個 array, 當發生 collision 時透過不同的 probing 找空的 slot.
separate chaining: 使用 array 搭配 linked list, 當發生 collision 時透過 linked list 將這些 data 串在一起儲存.

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.

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

Example:

Key = 8429 \rightarrow groups: 84 + 29 \rightarrow 113 \rightarrow 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: $h_1(27) = 7$ $h_1(57) = 7$
 $h_1(37) = 7$ $h_1(67) = 7$
 $h_1(47) = 7$

Q8. (Identify collision pattern)

From your results in Q1: 27.

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

A8: ① 每個 key 計算出的 hash value 都相同

② 算式本身不能使 key 有更多變化，且 mod 的數值本身數值小且非質數，使得結果只能產生出固定的 pattern。

Q9. (Compute using Hash Function 2)

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

A9: $h_2(1234) = 46 \bmod 11 = 2$
 $h_2(9217) = 109 \bmod 11 = 10$
 $h_2(4519) = 64 \bmod 11 = 9$
 $h_2(9902) = 101 \bmod 11 = 2$

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 h_1 ② 因為不算只是找 k 的餘數，還透過其他方式增加 k 的變化，且 mod 的數值改用質數，更能使 pattern 不可預測。
② h_2