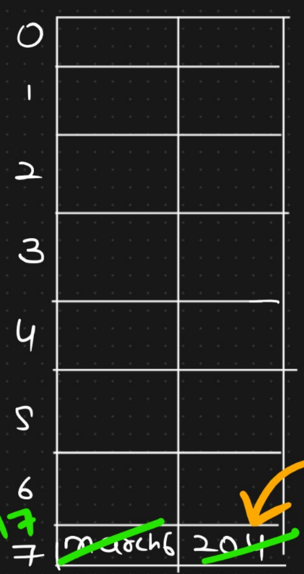
Collision Handling

- march 6 → 204
- march 17 → 206
- march 15 → 407

ASCII

march 6 — Hash function → 7

% m



Hashtable

arr(7)

Loose the data

march 17

206

collision

march 17 — Hash function → 7

1) chaining → Linked List / List ✓ → Live session ✓

extra space

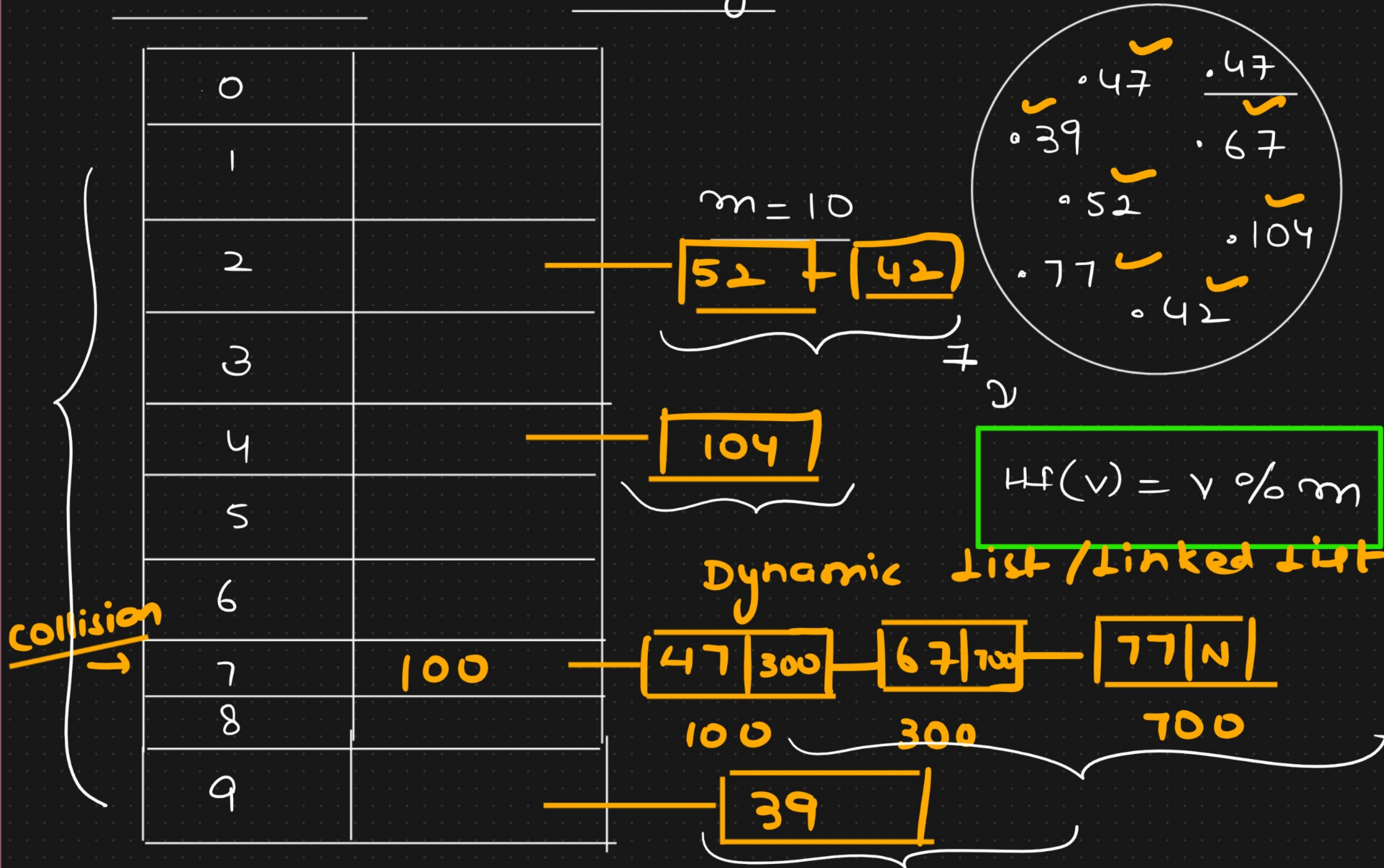
2) Open addressing → not using any extra space

- a) Linear Probing ✓
- b) Quadratic Probing ✓
- c) Double Hashing ✓

→ change Hash function

3) Load factor ✓

Hash Table   Chaining



Drawback   Worst case scenario

↳ 1) Extra time & extra space

search → $\Theta(n)$ ← TC

space complexity → $\Theta(n)$

Open Addressing

$hf(v) = v \% m$ ← size of hash table

1) Linear Probing

$\Rightarrow LP(v, i) = (hf(v) + i) \% m$

i → # collisions

Linear Probing → collision happened

⇓

move the value to next empty slot

$i = 0$

$LP(v, i) = hf(v) \% m$

50, 75, 99, 20, 35, 88, 45, 23, 55, 67

$50 \% 10 = 0$

$m = 10$

$LP(20, 0) = 0$

$LP(20, 1) = (0 + 1) \% 10 = 1$

$hf(20) = 0$

0	50
1	20
2	55
3	23
4	67
5	75
6	35
7	45
8	88
9	99

empty slot

65

65

size of Hash table

only value = m

$$hf(v) = (v \% m) \% m$$

$$= v \% m$$

Value = 45

$hf(45) = 45 \% 10 = 5$

$LP(45,0) = 5 \% 10 = 5 \rightarrow$ collision

$LP(45,1) = (5+1) \% 10 = 6 \rightarrow$ collision

$LP(45,2) = (5+2) \% 10 = 7$

Value = 55 (55, 65)

$hf(55) = 55 \% 10 = 5$

65, 0

$LP(55,0) = (5+0) \% 10 = 5$

65, 1

$LP(55,1) = (5+1) \% 10 = 6$

65, 2

$LP(55,2) = (5+2) \% 10 = 7$

Lot of time wastage

$LP(55,3) = (5+3) \% 10 = 8$

$LP(55,4) = (5+4) \% 10 = 9$

$LP(55,5) = (5+5) \% 10$

$10 \% 10 = 0$

$LP(55,6) = (5+6) \% 10$

$= 1$

65, 7

$LP(55,7) = (5+7) \% 10$

$= 2$

$LP(65,8) = (5+8) \% 10 = 3$

$\rightarrow$ traversed hash table $> m/2$

Time complexity

- Insertion (55) → $\Theta(n)$
- Searching → $\Theta(n)$
- Deletion → $\Theta(n)$

Worst case

Best case – $\Theta(1)$ (23)

Space complexity → $\Theta(1)$

→ open addressing techniques

Hashing DS

→ Drawback/Problem/Disadvantage

⇓

Linear Probing Technique

→ Primary clustering

(55, 65)

↓

Long Pattern/Path (Already followed by 55)

↳ Alot of time wastage to insert/search/delete

Quadratic Probing

$$\boxed{QP(v,i) = \left(hf(v) + c_1 i + c_2 i^2\right) \bmod m}$$

$hf(v) = v \bmod m$ $\quad$ $i \rightarrow$ collision num

$c_1, c_2 \rightarrow$ constant $= 1$

$m = 10$

50, 75, 99, 20, 35, 88, 45, 23, 55, 67

$\rightarrow$ not able to get an index $\rightarrow$ Data Loss

0	50
1	45
2	20
3	23
4	
5	75
6	
7	35
8	88
9	99

$hf(20) = 0$ $\quad$ $hf(20)$

$QP(20,0) = (0 + 0 + 0) \% 10 = 0 \rightarrow$ collision

$QP(20,1) = (0 + 1 + 1) \% 10 = 2 \% 10 = 2$

value = 35

$\boxed{c_1 = c_2 = 1}$

$hf(35) = 35 \% 10 = 5$

$$QP(v, i) = (hf(v) + c_1 \cdot i + c_2 \cdot i^2) \% m$$

$$QP(35, 0) = (5 + 0 + 0) \% 10$$

$$= 5$$

$$QP(35, 1) = (5 + 1 + 1) \% 10$$

$$7 \% 10 = 7$$

major Limitation of Quadratic Probing

$\Rightarrow$ value = 55

already Occupied

$\boxed{(5, 1, 7)}$

cycle

$QP(55, 0) = 5$ ✓

$QP(55, 1) = (5 + 1 + 1) \% 10 = 7$ ✓

$QP(55, 2) = (5 + 2 + 4) \% 10 = 1$

$QP(55, 3) = (5 + 3 + 9) \% 10 = 7$ ✓

$QP(55, 4) = (5 + 4 + 16) \% 10 = 5$ ✓

$QP(55, 5) = (5 + 5 + 25) \% 10 = 5$

Secondary clustering

Double Hashing

↳ $\boxed{\text{Two Hash functions}}$

$hf_1(v) = v \% m$

* $hf_2(v) = 1 + v \bmod (m-2)$

$$\boxed{DH(v,i) = \left(hf_1(v) + i\, hf_2(v)\right) \% m \quad (0 + 2*5) \% 10}$$

↓

collisions

50, 75, 99, 20, 35, 88, 45, 23, 55, 67

20 → 20, 5

↳ cycle

0	50
1	
2	
3	
4	
5	75
6	
7	
8	
9	99

$hf_1(20) = 20 \% 10$
$= 0$

$hf_2(20) = 1 + 20 \% 8$
$= 5$

$DH(20,1) = (0+5) \% 10$
$= 5$

$DH(20,2) = (0 + 2*5) \% 10$
$= 0$

$DH(20,3) =$
$(0 + 3*5) \% 10$
$= 5$

how much is the Load of the Hash table ← Load factor (α)

$n << m$ >0.5

$\boxed{\alpha = n/m}$ n = # data

m = size of hash table

0 to 1

$m = 10$

$n = 8$

$\boxed{\alpha = 0.8 \left(8/10\right)}$

$m = 100$ —— $\alpha = 8/100 = 0.08$