Hashing

Storing & Retriving Data in O(1) time

Search Key (24, 52, 91, 67, 48, 83)

Hash Function (k mod 10, k mod n
mid square, folding)

Hash Table Size = n - 1

So if k mod 10

	0
91	1
52	2
83	3
24	4
	5
	6
67	7
48	8
	9

Hash table application
DB

For search/retrival, same method

O(1) time.

What if we add 62?

Collision

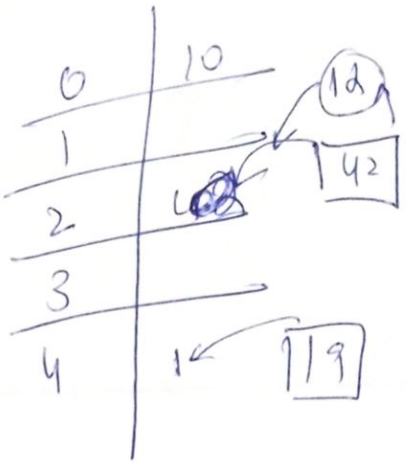
Collision Resolution Technique

- Chaining
  - Open hashing
    - ↓
    - open available space
- open addressing
  - closed hashing
    - ↳ Linear probing
    - ↳ Quadratic "
    - ↳ Double "

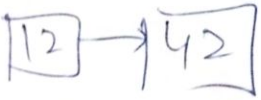
=> Chaining

key 42 19 10 12

k mod 5



node will be created { stored in node



Advantage: Deletion easy
O(1) Insertion

~~Search~~

Disadvantage
Searching is O(n)
Deletion could be O(n)

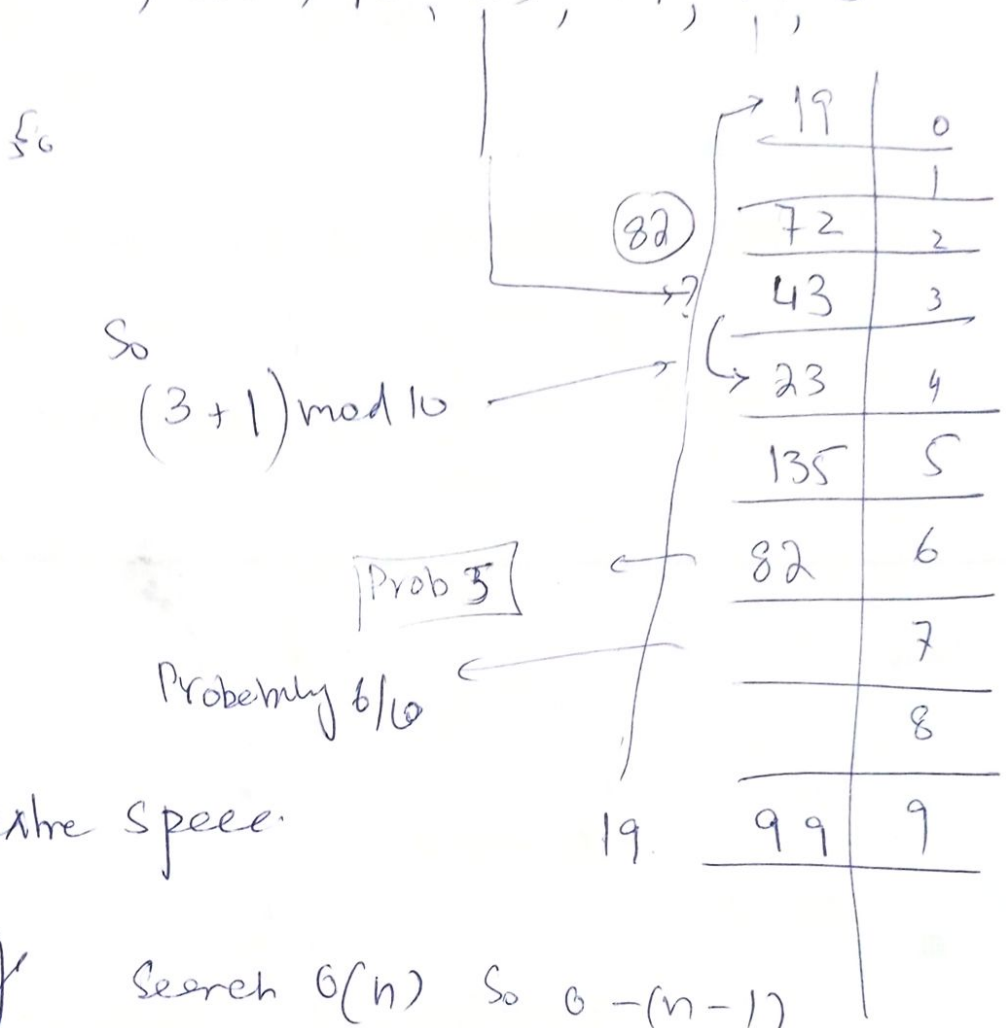
Linear Probing

$R(k) = k \bmod n$ $\quad (n = 10)$

$R(k, i) = (R(k) + i) \bmod n$

↳ collision number/probe num.

key = 43, 135, 72, 23, 99, 19, 82



So

$(3+1) \bmod 10$

Prob 3

Probability 6/10

Advantage

→ No extra space.

Disadvantage

Search $O(n)$ So $0 - (n-1)$

Deletion Difficult

↳ Assume we delete 43.

when traverse, 43 is empty

so 82 do not exist change

43 with # .So Searching continue

Primary clustering: it will make groups of elements - Probability of adding will increase

⇒ Primary cluster will increase search time

⇒ Secondary Clustering:

Visiting place by different elements which are already filled.

Quadratic Probing

$R(k) = k \bmod 10$

$R(k,i) = (R(k) + i^2) \bmod 10$

42 16 91 33 18 27 36 62

36	0
91	1
42	2
33	3
	4
	5
16	6
27	7
18	8
	9

Advantage
- No extra space
- Primary clustering resolved

$R(k) + i^2 \bmod 10$

$(6 + 1^2) \bmod 10$

7

$6 + 22 \bmod 10$

$(6 + 4) \bmod 10$

Issues
- Search $O(n)$
- Secondary clustering
- No guarantee of slot (after half filled)

Double Hashing

We use 2 hash functions

Key 20 34 45 70 56

$R_1(K) = K \bmod n$

use if collision $\begin{cases} R_2(K) = 8 - (K \bmod 8) \\ (R_1(K) + j R_2(K)) \bmod 11 \end{cases}$

for 45

$1 + 1 \left| 8 - (45 \bmod 8) \right. \bmod 11$

$4 \bmod 8$

make sure n = prime

	0
34	1
	2
	3
45	4
	5
70	6
	7
	8
20	9
	10

for 70

mod = (4)

So.

$8 - (70 \bmod 8)$

$8 - 6$

$(4 + 2 \times (2)) \bmod 11$

$6 \bmod 11$

Advantage = No space
= No primary clustering
= It secondary

Disadvantage = $O(n)$