

11C 排解冲突 (2)

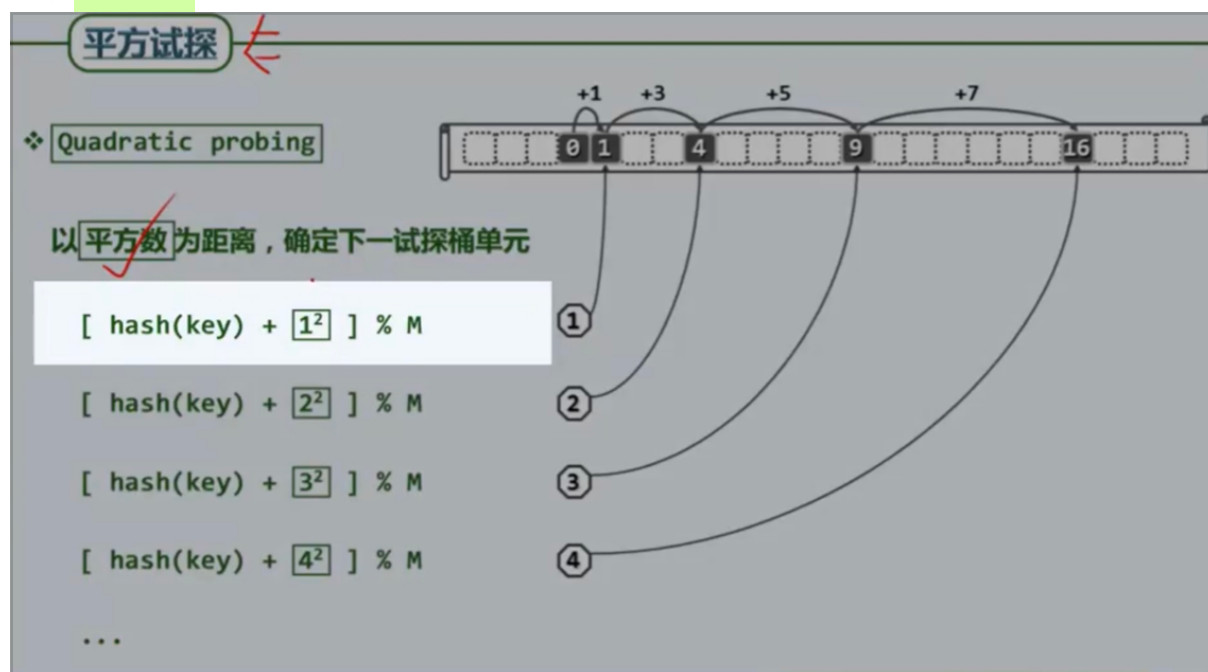
#数据结构邓神

平方试探 相对于线性试探

线性试探问题根源：线性试探位置间距太小

我们应该拉开距离

不妨 平方试探



一利一弊

优点：数据聚集现象有所缓解（很大程度上）

查找链条上，各个桶间距线性提升，一旦冲突可聪明的跳离是非之地

缺点：破坏了数据访问的局部性，如果涉及到外存，IO激增

缓存页面的规模为几个KB，不妨将他设置为1KB

如果只记录引用，每个桶4b

如果我们真的需要IO，需要连续发生16次冲突

但是如果真的发生，只能说散列函数太烂或者说运气太糟糕了

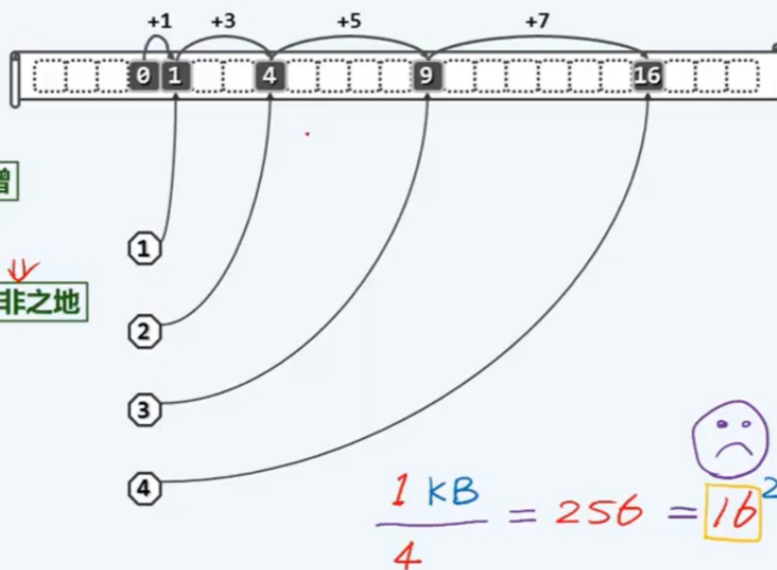
优点、缺点及疑惑

❖ 数据聚集现象有所缓解

查找链上，各桶间距线性递增

一旦冲突，可聪明地跳离是非之地

❖ 若涉及外存，I/O将激增



只要有空桶就一定能找出来吗

至多半载

装填因子必须要足够小

装填因子，须足够小！



M若为合数： $n^2 \% M$ 可能的取值必然少于 $\lceil M/2 \rceil$ 种——此时，只要对应的桶均非空...

❖ $\{0, 1, 2, 3, 4, 5, \dots\}^2 \% 11 = \{0, 1, 4, 9, 5, 3\} \leftarrow$



M若为素数： $n^2 \% M$ 可能的取值恰好会有 $\lceil M/2 \rceil$ 种——此时，恰由查找链的前 $\lceil M/2 \rceil$ 项取遍

❖ **定理** 若M是素数，且 $\lambda \leq 0.5$ ，就一定能够找出；否则，不见得

M + Lambda

查找链前缀，必足够长！

❖ **反证**：假设存在 $0 \leq a < b < \lceil M/2 \rceil$ ，使得

M prime, $\lambda \leq 0.5$

沿着查找链，第a项和第b项彼此冲突

❖ 于是： a^2 和 b^2 自然属于 M 的某一同余类，亦即

$$\Rightarrow a^2 \equiv b^2 \pmod{M}$$

$$b^2 - a^2 = (b + a) \cdot (b - a) \equiv 0 \pmod{M}$$

❖ 然而： $0 < b - a < b + a < M$ —— 这与 M 为素数矛盾

$b+a$ 至少为2 这与 M 是素数的假设，相悖

双向平方试探

双向平方试探

❖ 自冲突位置起，依次向后试探

$$[\text{hash}(\text{key}) + 1^2] \% M$$

$$[\text{hash}(\text{key}) - 1^2] \% M$$

$$[\text{hash}(\text{key}) + 2^2] \% M$$

$$[\text{hash}(\text{key}) - 2^2] \% M$$

$$[\text{hash}(\text{key}) + 3^2] \% M$$

$$[\text{hash}(\text{key}) - 3^2] \% M$$

...



查找链，彼此独立？

| | $\pm i^2$ | -36 | -25 | -16 | -9 | -4 | -1 | 0 | 1 | 4 | 9 | 16 | 25 | 36 |
|---|-----------|-----|-----|-----|----|----|----|---|---|---|---|----|----|----|
| M | 5 | | | | | 1 | 4 | 0 | 1 | 4 | | | | |
| | 7 | | | | 5 | 3 | 6 | 0 | 1 | 4 | 2 | | | |
| | 11 | | 8 | 6 | 2 | 7 | 10 | 0 | 1 | 4 | 9 | 5 | 3 | |
| | 13 | 3 | 1 | 10 | 4 | 9 | 12 | 0 | 1 | 4 | 9 | 3 | 12 | 10 |

他可以取遍搜索的桶

只要装填因子不到100%，就一定可以找到一个空桶

❖ 正向和逆向的子查找链，各包含 $\lceil M/2 \rceil$ 个互异的桶

$-\lceil M/2 \rceil, \dots, -2, -1, \boxed{0}, 1, 2, \dots, \lceil M/2 \rceil$

除了公共的0，是否还有其他的公告的桶呢，如果没有总长就应该是m

不是

M = 13 的时候正向查找链表和逆向查找链是完全一样的，只有在某些素数表长可以行之有效，而有些却不行

两种素数有什么区别呢？

4K+3

4k + 3

❖ 两类素数：

3

5

7

11

13

17

19

23

29

31

...

| | | | | | | | | | | | | | | |
|-----------|----|-----|-----|-----|----|----|----|---|---|---|---|----|----|----|
| $\pm i^2$ | | -36 | -25 | -16 | -9 | -4 | -1 | 0 | 1 | 4 | 9 | 16 | 25 | 36 |
| M | 5 | | | | | 1 | 4 | 0 | 1 | 4 | | | | |
| | 7 | | | | 5 | 3 | 6 | 0 | 1 | 4 | 2 | | | |
| | 11 | | 8 | 6 | 2 | 7 | 10 | 0 | 1 | 4 | 9 | 5 | 3 | |
| | 13 | 3 | 1 | 10 | 4 | 9 | 12 | 0 | 1 | 4 | 9 | 3 | 12 | 10 |

有些素数关于4mod为3，和关于4的mod为1的两类素数

关于4mod为3都可以作为双向平方列表的长度

而为1的就不行

❖ 表长取作素数 $M = 4 \times k + 3$, 必然可以保证查找链的前 M 项均互异

泾渭分明