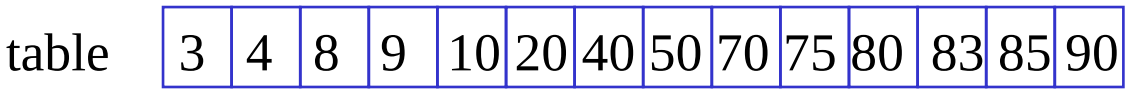
Méthodes de hachage

1. Introduction
2. Hachage ouvert
3. Hachage fermé
4. Implémentation des fonctions

Recherche dichotomique



d i f

4 ?

d i f

d i f

d i f

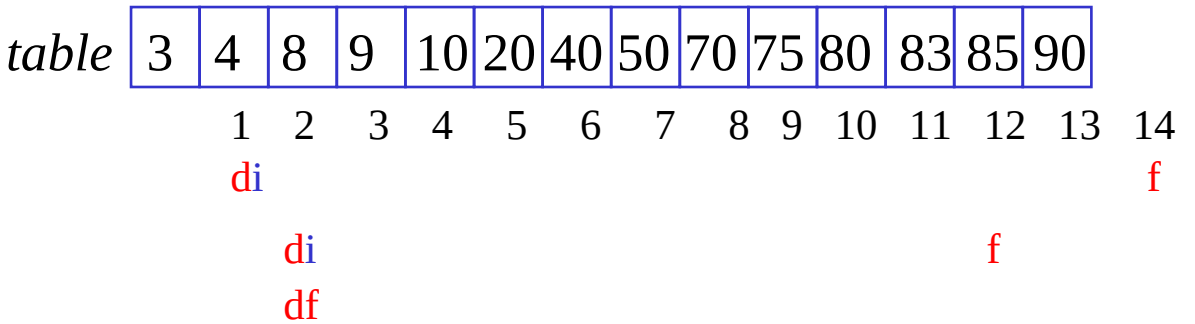
di f

df

```
fonction  ELEMENT (x, table, d, f ) ;
début
    si  (d = f)  alors
        si  (x = table [d ])  alors  retour  (vrai)
        sinon  retour (faux)
    sinon {
        i ← ⌊(d+f ) / 2⌋  ;
        si  (x > table [i])  alors
            retour  (ELEMENT (x, table, i+1, f ))
        sinon  retour  (ELEMENT (x, table, d, i))
    }
fin
```

Temps (ELEMENT sur *table* [1 … *n*]) = O (log *n*)

Recherche par interpolation



fonction ELEMENT (*x*, *table*, *d*, *f*) ;
début
  si (*d* = *f*) **alors**
    si (*x* = *table* [*d*] **alors retour** (vrai)
    sinon retour (faux)
  sinon {

$$i \leftarrow d + \left\lfloor \frac{x - table[d]}{table[f] - table[d]} \right\rfloor * (f - d);$$

    si (*x* > *table* [*i*] **alors**
      retour (ELEMENT (*x*, *table*, *i*+1, *f*))
    sinon retour (ELEMENT (*x*, *table*, *d*, *i*))
  }
fin

Idée : Établir une relation entre un élément et l'adresse à laquelle il est rangé en mémoire

Théorème : Si les éléments de *table* [1 … *n*] et *x* sont choisis uniformément dans un intervalle [*a*,*b*], le temps moyen d'exécution de la recherche par interpolation est O (log log *n*)

Une autre façon c’est d’utiliser
- hachage
- des tables de hachage
- des fonctionne de hachage

Une table de hachage est une structure de données qui implémente un tableau associatif (un dictionnaire)
Dans un tableau associatif, les données sont stockées sous la forme d’**une collection de paires clé-valeur**.
La position des données dans le tableau est déterminée par l’application d’un algorithme de hachage à la clé **-** un processus appelé **hachage**.
L’algorithme de hachage est appelé **fonction de hachage**.

- Les tables de hachage permettent une recherche très efficace.

- Dans le meilleur des cas, les données peuvent être extraites d’une table de hachage **en temps constant**.

- **La maintenance (ajout, mise à jour et suppression) de données** dans une table de hachage est également très efficace.

Comment implémenter une table de hachage ?

Pour implémenter une table de hachage :
1) Il faut **utiliser un tableau** car vous devez pouvoir accéder directement à chaque position du tableau.

2) Les positions au sein d'un tableau sont parfois appelées '**buckets** '; Chaque bucket est utilisé pour stocker des données.

3) La clé doit être stockée à côté des données

4) **La taille du tableau** doit être planifiée avec soin. Il doit être **suffisamment grand** pour stocker toutes les données, **mais pas trop grand** pour ne pas gaspiller de l'espace.

5) Une table de hachage efficace **doit toujours avoir de l'espace libre**.

des fonctionne de hachage

Une fonction de hachage est un algorithme qui convertit une clé de hachage en une valeur de hachage.

Les principales exigences d'une fonction de hachage:

1) produisent toujours la même valeur de hachage pour la même clé

2) fournir une distribution uniforme des valeurs de hachage. (chaque valeur a une probabilité égale d'être générée)

3) minimiser le clustering (la collision); Cela se produit lorsque de nombreuses clés différentes produisent la même valeur de hachage. Lorsque deux ou plusieurs clés différentes produisent la même valeur de hachage, on parle de « **collision** ».

4) être très rapide à calculer.

Exemple 1 :

supposons que les **clés sont des nombres à 5 chiffres** et que nous avons **une table de hachage avec 97 'bucket'**.

```
FUNCTION hash_integer(hash_key, number_of_slots)

// Générer la valeur de hachage à l'aide de l'opératrice modulo

hash_value = hash_key MOD number_of_slots

// Return the hash value

RETURN hash_value

ENDFUNCTION
```

Le resultat de hachage

Key	Hashing function	Hash value
12345	12345 MOD 97	26
67564	67564 MOD 97	52
34237	34237 MOD 97	93
23423	23423 MOD 97	46
00332	00332 MOD 97	41

如果要存n个数，一般mod比n小的最大素数

Exemple 2 (hashage des chaîne de caractère) :

supposons que les **clés sont des chaîne de caractères comme TH5L** et que nous avons **une table de hachage avec 97 'bucket'**.

```
FUNCTION hash_string(hash_key, number_of_slots)
    // Initialise total
    total = 0

    // Repeat for every character in the hash key
    FOR i = 0 TO LEN(hash_key) - 1
       ascii_code = ASC(hash_key[i])
       total = total + ascii_code
   NEXT i

    // Generate the hash value using the modulo operator
    hash_value = sum MOD number_of_slots

    // Return the hash value
    RETURN hash_value
ENDFUNCTION
```

Le resultat de hachage

Par exemple si on considère A5RD:

ASC(A)+ASC(5)+...= 65 + 53 + 82+ 68 = 268

Et finalement: 268 MOD 97 = 74

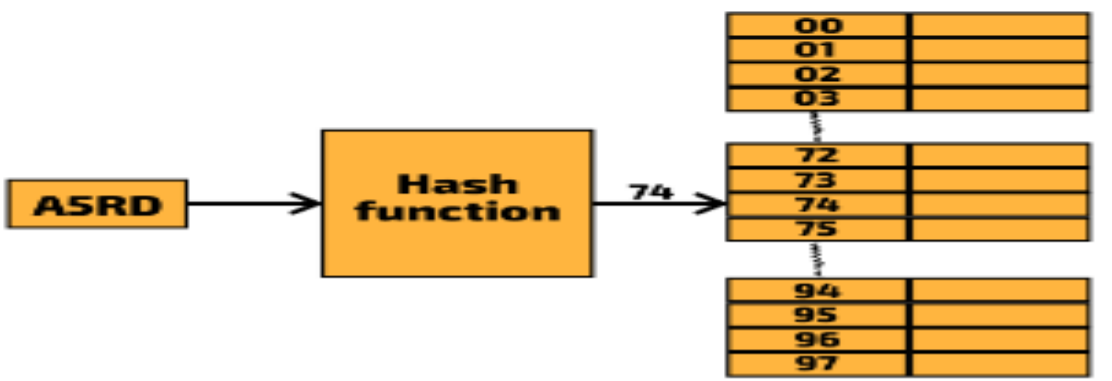
Les données associées au code A5RD seront stockées à la position 74 dans la table de hachage.

insérer des données dans une table de hachage

Il faut utiliser **la fonction de hachage pour générer l'index** de la position dans le tableau qui sera utilisé pour stocker les données.

Il faut stocker la clé à côté de données .

Hash key	Hashing function	Hash value	
A5RD	Apply hashing function	Hash value = 74	

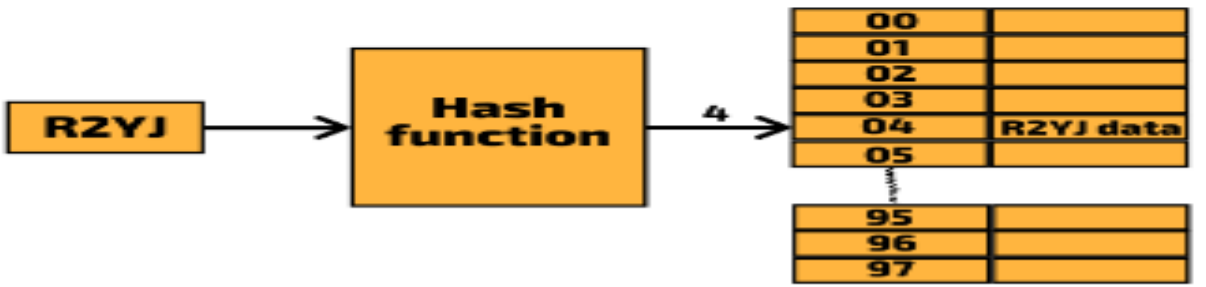


récupérer une valeur à partir d’une table de hachage

la fonction de hachage est appliquée à la clé afin de générer l’index de la position dans le tableau

c’est là que se trouveront les données associées à cette clé

Key	Hash function	Hash value	
R2YJ	Apply hash function	Hash value = 4	

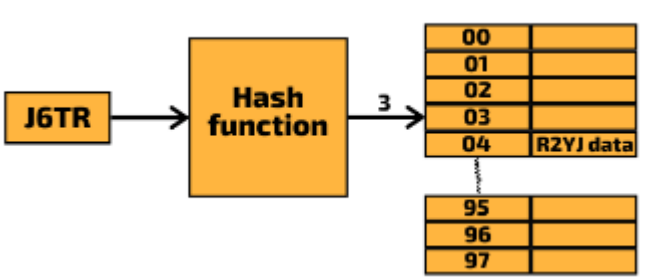


Recherche d’un élément qui n’existe pas

la fonction de hachage est appliquée à la clé afin de générer l’index de la position dans le tableau

Hash key	Hash function	Hash value	
J6TR	Apply hashing function	Hash value = 3	

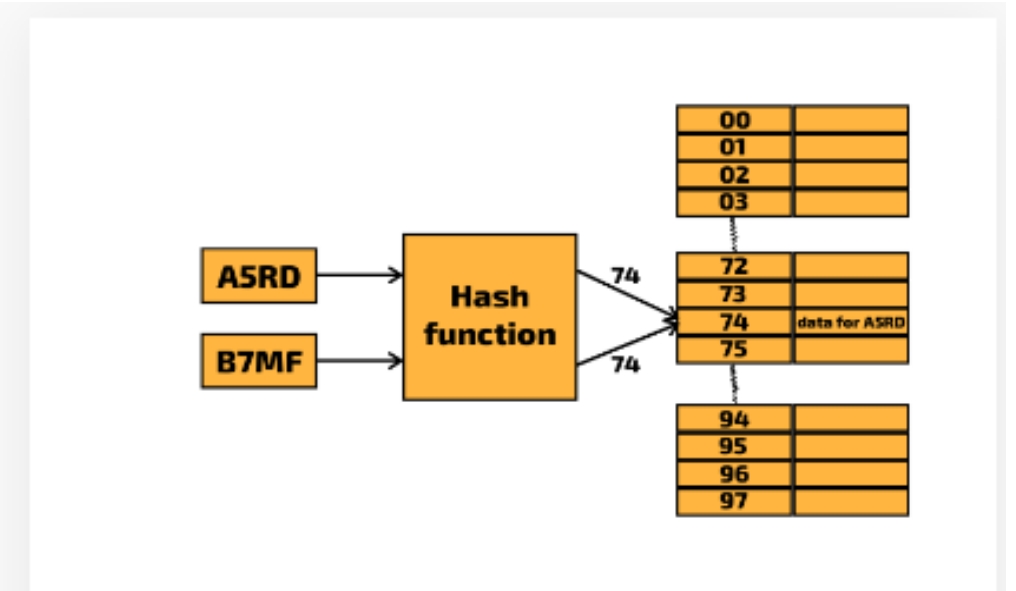
S’il n’y a pas de données à la position 3 de la table de hachage, vous savez qu’elles n’existent pas ;
Vous n’avez pas besoin de chercher dans l’ensemble du tableau pour le prouver.



collision

Une collision se produit lorsque la fonction de hachage produit la même valeur de hachage pour deux clés ou plus

Hash key	Hash function	Hash value	
J6TR	Apply hashing function	Hash value = 3	



Les problème de collision

1) vous ne pouvez pas stocker deux ensembles de données au même endroit. Vous ne pouvez pas recupérer des information

2) Quelle que soit la qualité de votre fonction de hachage, vous devrez probablement faire face à des collisions.

Résolution des collisions

Hachage ouvert : avec listes, triées ou non.

Hachage fermé : linéaire, quadratique, aléatoire, uniforme, double, ...

填充因子：需要放n个元素/有m个空间

按经验，一般取0.7比较好

开放地址法：找不到向后找别的空位

拉链法：以链表形式存在一个位子里