**Tìm hiểu về hash table**

# Khái niệm

1. Hàm Hash ?

a. Hàm băm là gì?

Hàm băm (hash function) là một hàm toán học chuyển đổi một thông điệp đầu vào có độ dài bất kỳ thành một giá trị băm có độ dài cố định (tuỳ thuộc vào thuật toán băm) đại diện cho dữ liệu ban đầu.

b. Đặc trưng của hàm băm:

Không thể dịch ngược từ giá trị băm quay trở lại chuỗi ký tự ban đầu.

Với mỗi một đầu vào ngẫu nhiên, hashing phải tạo ra được một giá trị băm tương ứng

Đầu vào khác nhau phải xuất ra được các giá trị băm khác nhau.

1. Hash table ?

- Bảng băm là một cấu trúc dữ liệu lưu trữ một tập hợp cho phép ta có thể nhanh chóng xác định xem một phần tử nào đó có nằm trong tập hợp hay không.

- Trong khoa học máy tính, bảng băm (hash tables) là một cấu trúc dữ liệu sử dụng hàm băm để ánh xạ từ giá trị xác định, được gọi là khóa (ví dụ như tên của một người), đến giá trị tương ứng (ví dụ như số điện thoại của họ)

Do đó, bảng băm là một mảng kết hợp. Hàm băm được sử dụng để chuyển đổi từ khóa thành chỉ số (giá trị băm) trong mảng lưu trữ các giá trị tìm kiếm.

(Hiểu) Đơn giản nó chỉ chuyển đổi từng chữ cái và ký hiệu thành một số và tổng hợp tất cả chúng.

1. Mục đích & vai trò

* Giúp cho việc tìm kiếm, tra cứu dữ liệu thông tin trở nên nhanh chóng dễ dàng, tiết kiệm thời gian cồn sức của người sử dụng.

1. Mô tả
2. Các bước thực hiện

Bước 1: Băm giá trị truyển vào bằng thuật toán băm.

Bước 2: Lưu giá trị vào bảng băm.

Bước 3: Thực hiện tìm kiểm giá trị cần tìm.

1. Mô tả chi tiết

Bước 1: Băm giá trị cần truyền vào

* Tùy vào mục đích của người thực hiện để dùng thuật toán băm phù hợp.
* Sau khi băm giá trị truyền vào, ta được một giá trị key tương ứng với 1 giá trị
* Rủi ro: sau khi băm sẽ có những giá trị key bị trùng nhau bởi sử dụng những thuật toán băm không phù hợp.

Bước 2:

* Lưu key là chỉ số (địa chỉ), giá trị là dữ liệu (gốc truyền vào).
* Rủi ro: khi key bị trùng sẽ dấn đến hiện tượng xung đột dữ liệu.

Bước 3:

* Thực hiện tìm kiếm theo chỉ số (key).

Biện pháp giải quyết rủi ro:

* Bước 1: tìm 1 thuật toán băm tốt
* Bước 2: Nếu bị xung đột thì sử dụng các kĩ thuật xử lý va chạm:

+ Separate chaining (open hashing): Để lưu giữ một phần tử trong bảng băm,chúng ta thêm nó vào một danh sách liên kết ứng với chỉ mục của nó. Nếu có sự va chạm xảy ra, các phần tử đó sẽ nằm cùng trong 1 danh sách liên kết.

+ Sử dụng các công thức để tìm ra một chỉ mục khác chưa có giá trị .