

Today's announcements:

MP6 available, due 4/12, 11:59p. EC due 4/5, 11:59p.

Exam 2: returned in section next week.

Today: Hashing - hash table consists of

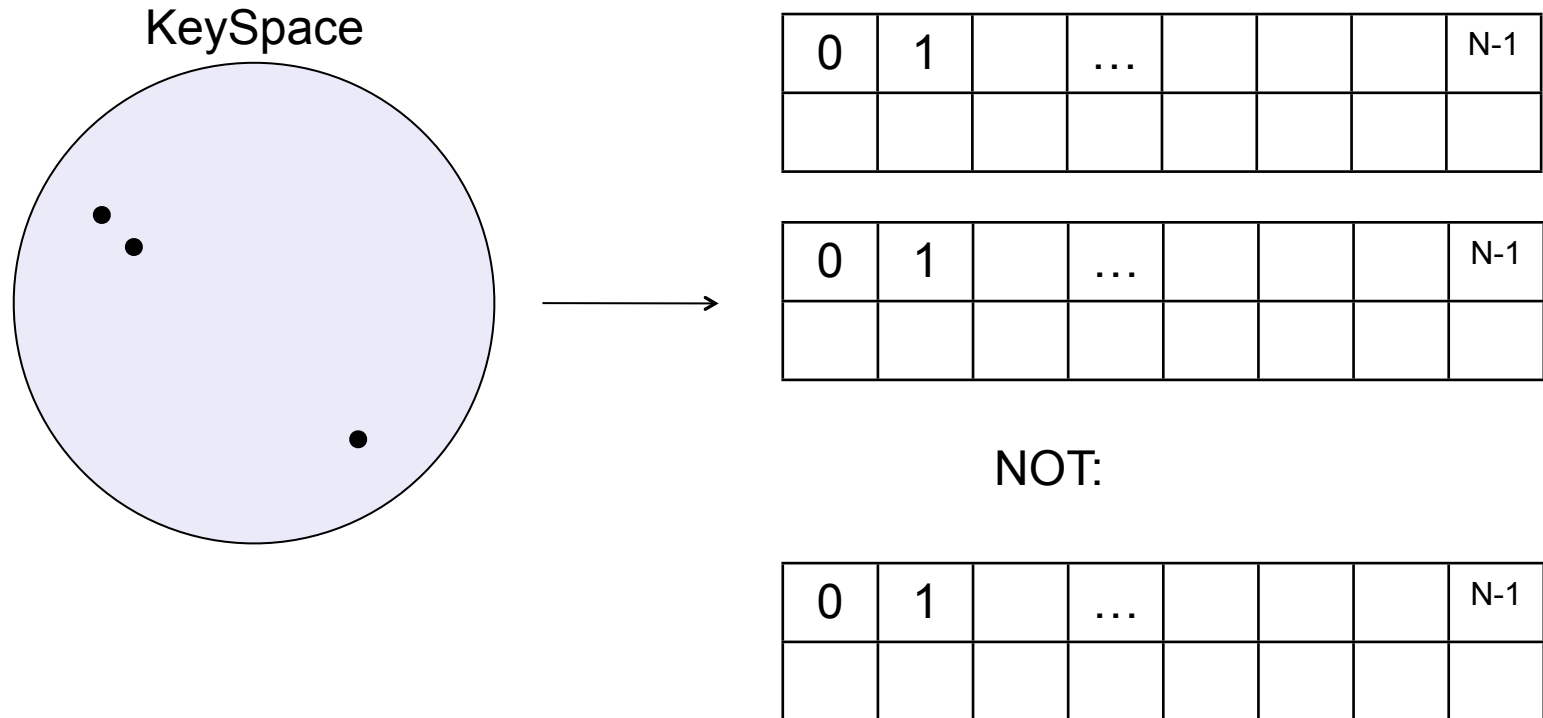
<http://groups.engin.umd.umich.edu/CIS/course.des/cis350/hashing/WEB/HashApplet.htm>

Hash Functions - why are general purpose hash functions so hard to create?

OBS1: If your keyspace is small, and known, then you can create a custom hash function that behaves well.

OBS2: If the distribution of your samples over the keyspace is known, then you can probably make a hash fn that behaves well.

BUT:



Typical approach: “randomize” the key via bit scrambling and then take a subset of the scrambled bits.

Hashing Strings (an example)

Given: 8 character strings are easy to hash

The idea: Select 8 random positions from long strings and hash that substring.

A bunch of strings:

Lookyhere, Huck, being rich ain't going
No! Oh, good-licks; are you in real dead
Just as dead earnest as I'm sitting here
nto the gang if you ain't respectable, y
Can't let me in, Tom? Didn't you let me
Yes, but that's different. A robber is m
irate is -- as a general thing. In most
Now, Tom, hain't you always ben friendly
ut, would you, Tom? You wouldn't do that
Huck, I wouldn't want to, and I DON'T wa
ay? Why, they'd say, 'Mph! Tom Sawyer's
t!' They'd mean you, Huck. You wouldn't
uck was silent for some time, engaged in
Well, I'll go back to the widder for a m
can come to stand it, if you'll let me
All right, Huck, it's a whiz! Come along
Will you, Tom -- now will you? That's go

Hashing Strings (an example)

Given: 8 character strings are easy to hash

The idea: Select 8 random positions from long strings and hash that substring.

A bunch of strings:

```
http://en.wikipedia.org/wiki/Le%C5%9Bna_Grobla
http://en.wikipedia.org/wiki/Blow_the_Man_Down
http://en.wikipedia.org/wiki/Swen_K%C3%B6nig
http://en.wikipedia.org/wiki/2/7th_Cavalry_Commando_Regiment_(Australia)
http://en.wikipedia.org/wiki/Salman_Ebrahim_Mohamed_Ali_Al_Khalifa
http://en.wikipedia.org/wiki/Alice_High_School
http://en.wikipedia.org/wiki/Beautiful,_Dirty,_Rich
http://en.wikipedia.org/wiki/RFA_Sir_Bedivere_(L3004)
http://en.wikipedia.org/wiki/Birthright_(band)
http://en.wikipedia.org/wiki/Jacky_Vimond
http://en.wikipedia.org/wiki/Vachon
http://en.wikipedia.org/wiki/McCarthy_%26_Stone
http://en.wikipedia.org/wiki/Salisbury,_New_Hampshire
http://en.wikipedia.org/wiki/A_Line_of_Deathless_Kings
http://en.wikipedia.org/wiki/Newfoundland_Irish
http://en.wikipedia.org/wiki/Beatrice_Politi
http://en.wikipedia.org/wiki/Bona_Sijabat
```

Collision handling - Separate Chaining: (an example of open hashing)

$S = \{16, 8, 4, 13, 29, 11, 22\}$

$|S| = n$

$h(k) = k \% 7$

0	
1	
2	
3	
4	
5	
6	

	Worst case	Under SUHA
Insert		
Remove/find: successful		
Remove/find: unsuccessful		

Collision Handling - Probe based hashing: (example of closed hashing)

$S = \{16, 8, 4, 13, 29, 11, 22\}$

$|S| = n$ $h(k) = k \% 7$

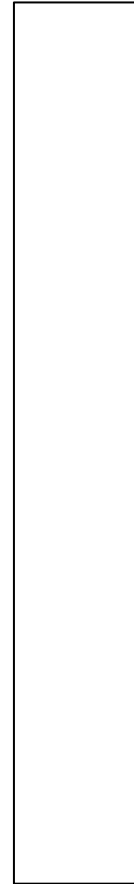
0	
1	
2	
3	
4	
5	
6	

Probe based hashing – 2 problems...

Removals:

0	22	
1	8	
2	16	
3	29	
4	4	
5	11	
6	13	

Clustering:



Probe based hashing: (double hashing)

$S = \{16, 8, 4, 13, 29, 11, 22\}$ $|S| = n$ $H(k,i) = h_1(k) + ih_2(k)$

0	
1	
2	
3	
4	
5	
6	

Hash table performance: expected # of probes for Find(key) under SUHA

Linear probing -

successful: $\frac{1}{2} (1 + 1/(1-\alpha))$

unsuccessful: $\frac{1}{2} (1 + 1/(1-\alpha))^2$

Double hashing -

successful: $-1/n (1-\alpha) / \alpha$

unsuccessful: $1/(1-\alpha)$

Separate chaining -

successful: $1 + \alpha / 2$

unsuccessful: $1 + \alpha$

Do not memorize these!

Observe:

- As α increases, running times increase...
- If α is held constant then running times are constant...

Which is better?

- Big records –
- Structure speed –

What structures do hash tables replace for us?

Miscellaneous Discussion –

Which hash table implementation is better?

- Big records –
- Structure speed –

What structures do hash tables replace for us?

Why do we talk about balanced BST if hashing is so great?

Applications of hashing?

Area of active research in mathematics to develop general purpose hash functions.

Interview questions!

Consider a stream of integers... you want to know immediately if the stream stops producing unique values.

0	1	2	3	4	5	6	7

Consider a stream of integers... you want to know immediately if you have seen a pair of numbers that sum to 225.

0	1	2	3	4	5	6	7

Interview questions!

Consider an array of n integers... determine if there are 3 integers whose sum is 0, and do it in $O(n^2)$ time.

.

0	1	2	3	4	5	6	7

<http://www.ardendertat.com/2012/01/09/programming-interview-questions/>