CS 225

Data Structures

March 16 – Hashing

Wade Fagen-Ulmschneider

A Hash Table based Dictionary

Client Code:

```
1  Dictionary<KeyType, ValueType> d;
2  d[k] = v;
```

A **Hash Table** consists of three things:

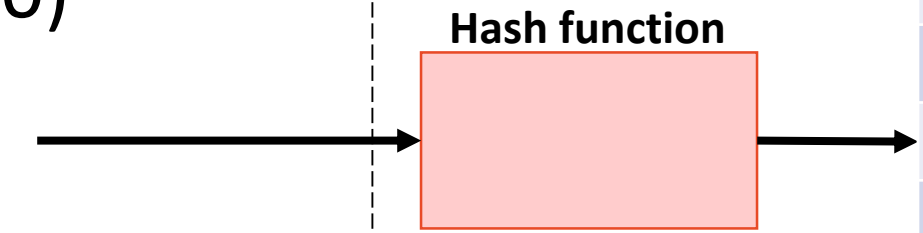
1. A hash function, f(k)

2. An array

3. A mystery element

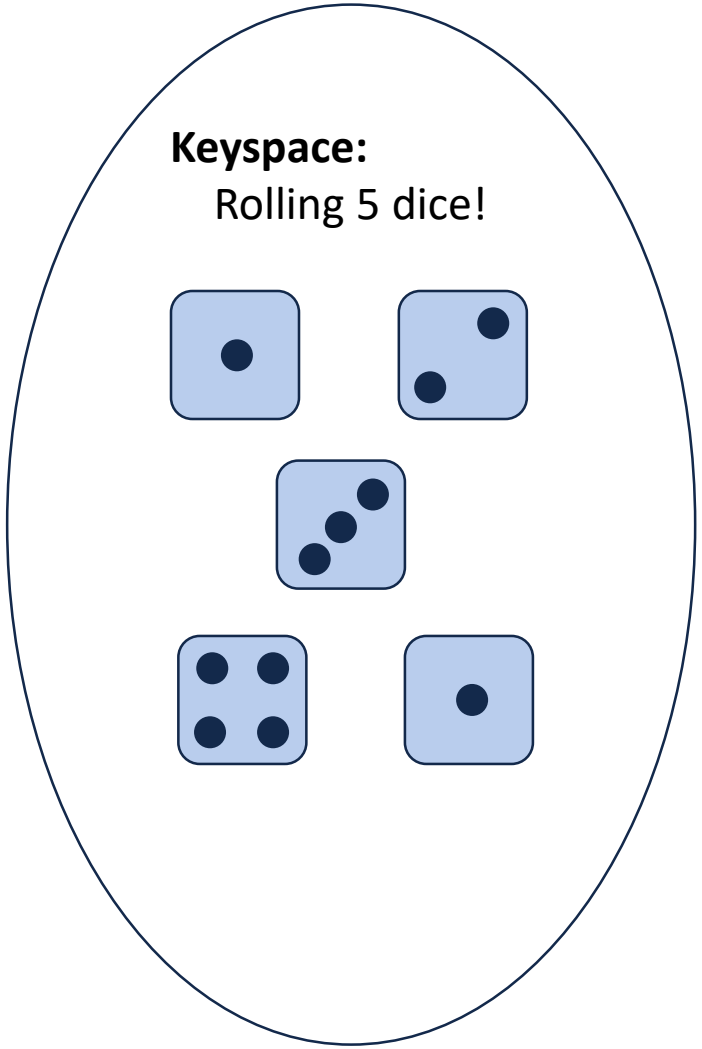
A Perfect Hash Function

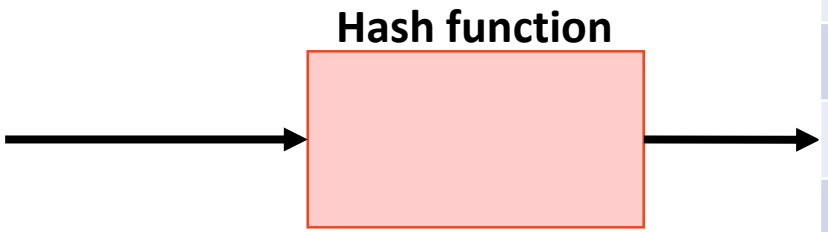
(Angrave, CS 241)
(Beckman, CS 421)
(Cunningham, CS 210)
(Davis, CS 101)
(Evans, CS 126)
(Fagen-Ulmschneider, CS 225)
(Gunter, CS 422)
(Herman, CS 233)



Key	Value

A Perfect Hash Function





Key	Value
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

Hash Function

Our **hash function** consists of two parts:

- A **hash**:

- A **compression:**

Choosing a good hash function is tricky…

- Don’t create your own (yet*)
- Very smart people have created very bad hash functions

Hash Function

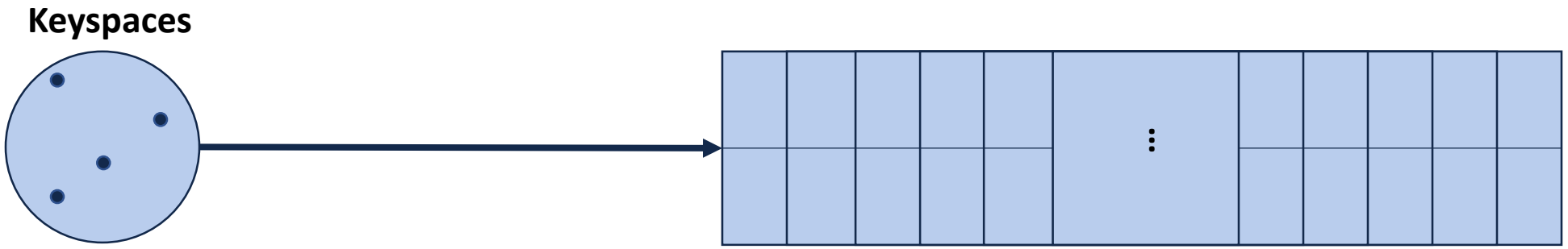
Characteristics of a good hash function:

1. Computation Time:

2. Deterministic:

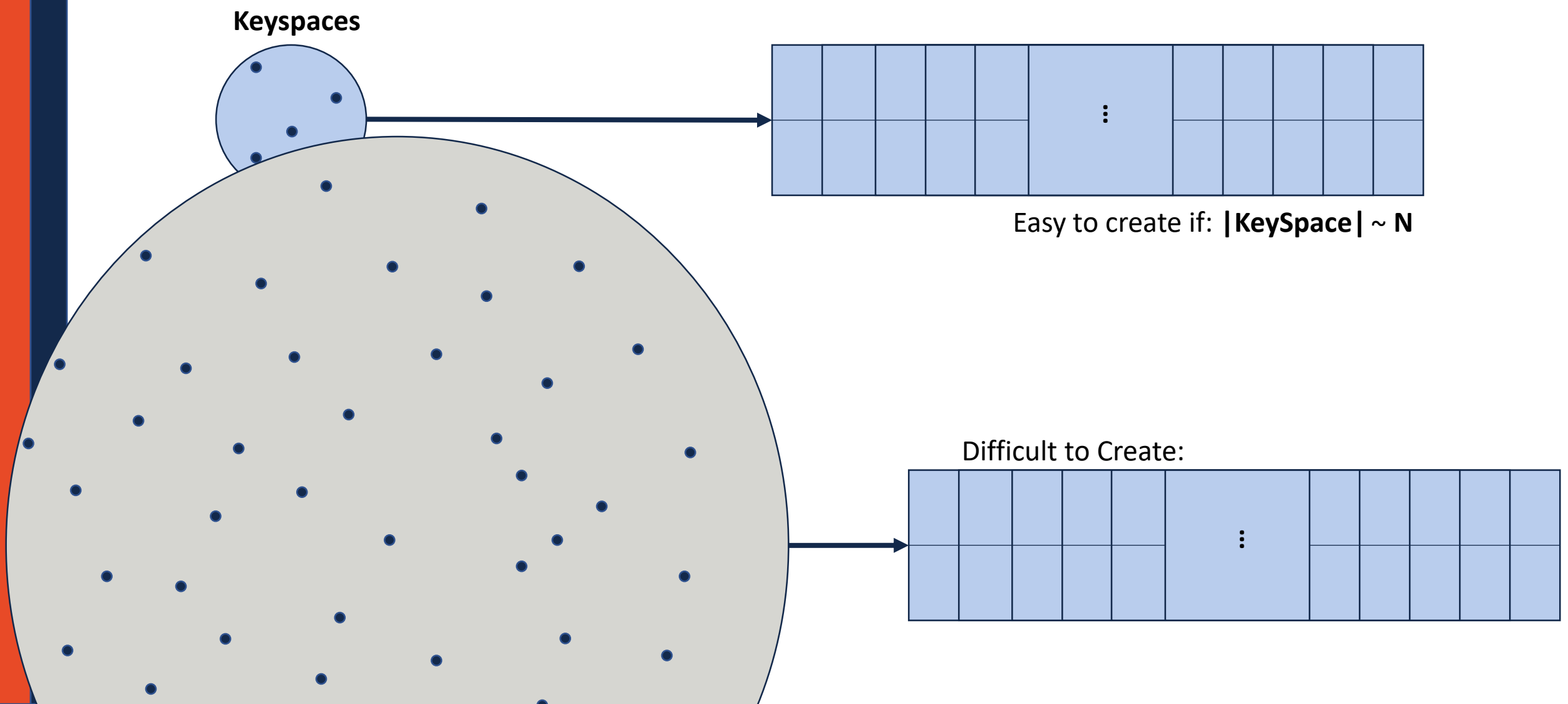
3. Satisfy the SUHA:

General Purpose Hash Function

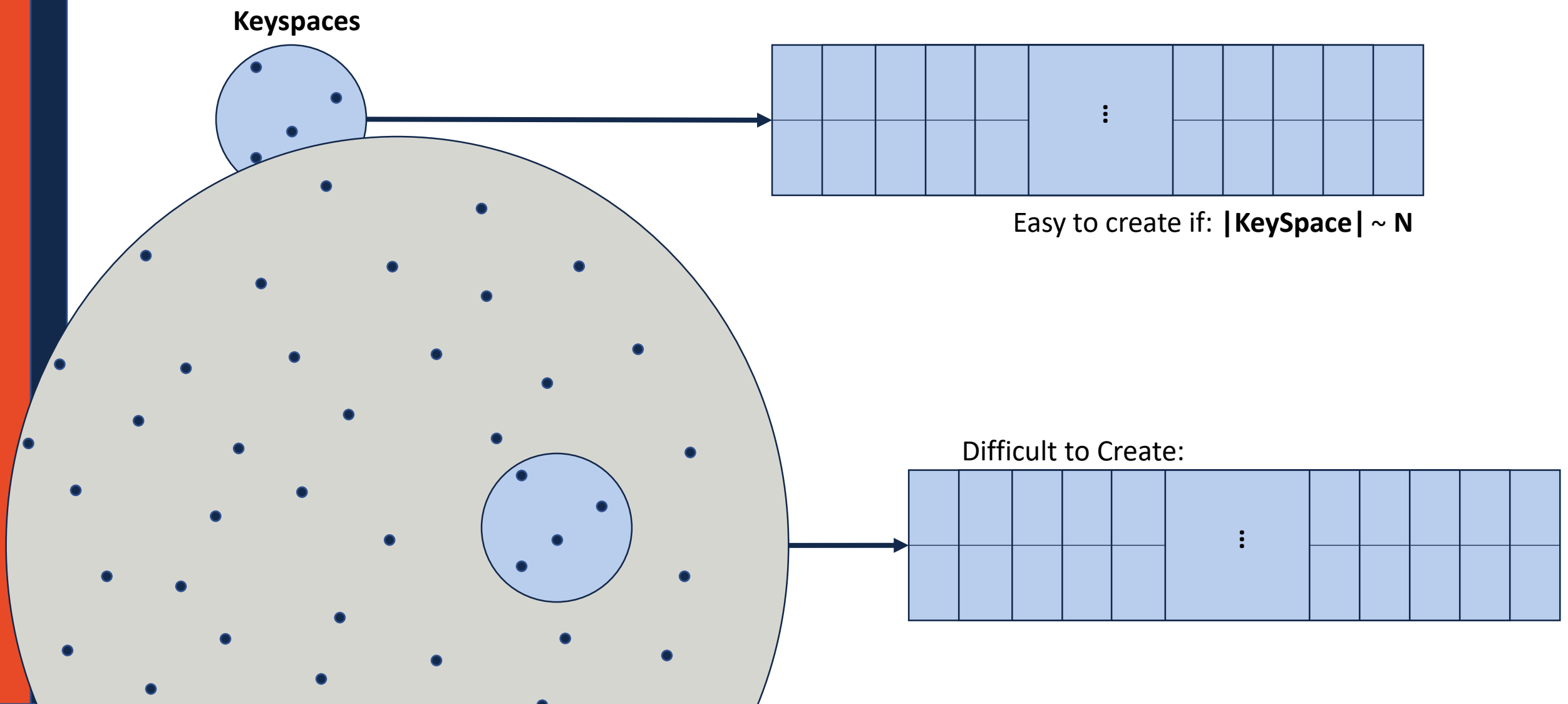


Easy to create if: **|KeySpace| ~ N**

General Purpose Hash Function



General Purpose Hash Function



Hash Function

Given: Easy to create a hash function of strings of length 8.

Idea: Map 40 character things to length 8:

```
Alice was beginning to get very tired of
 sitting by her sister on the bank, and
of having nothing to do: once or twice s
he had peeped into the book her sister w
as reading, but it had no pictures or co
nversations in it, ‘and what is the use
of a book,’ thought Alice ‘without pictu
res or conversations?’  So she was consi
dering in her own mind (as well as she c
ould, for the hot day made her feel very
 sleepy and stupid), whether the pleasur
e of making a daisy-chain would be worth
 the trouble of getting up and picking t
he daisies, when suddenly a White Rabbit
 with pink eyes ran close by her.  There
 was nothing so very remarkable in that;
 nor did Alice think it so very much out
 of the way to hear the Rabbit say to it
self, ‘Oh dear! Oh dear! I shall be late
!’ (when she thought it over afterwards,
 it occurred to her that she ought to ha
```

Idea: Map 40 character things to length 8:

```
https://en.wikipedia.org/wiki/Main_Page
https://en.wikipedia.org/wiki/Battle_of_
https://en.wikipedia.org/wiki/Vector_Gen
https://en.wikipedia.org/wiki/2017_Austr
https://en.wikipedia.org/wiki/19th_Natio
https://en.wikipedia.org/wiki/Japanese_g
```

Hash Function

In CS 225, we focus on **general purpose** hash functions.

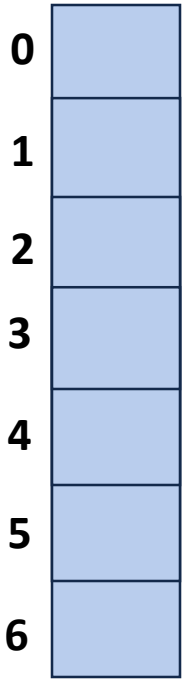
Other hash functions exists with different properties
(eg: cryptographic hash functions)

(Example of open hashing)

Collision Handling: Separate Chaining

S = { 16, 8, 4, 13, 29, 11, 22 } **|S| = n**

h(k) = k % 7 **|Array| = N**



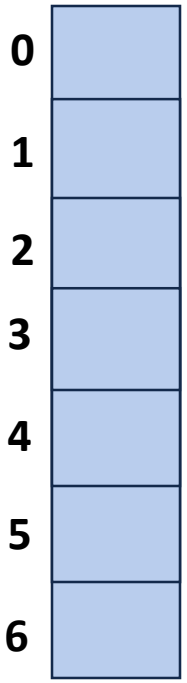
	Worst Case	SUHA
Insert		
Remove/Find		

(Example of closed hashing)

Collision Handling: Probe-based Hashing

S = { 16, 8, 4, 13, 29, 11, 22 } **|S| = n**

h(k) = k % 7 **|Array| = N**



Try h(k) = (k + 0) % 7, if full…

Try h(k) = (k + 1) % 7, if full…

Try h(k) = (k + 2) % 7, if full…

Try …