Today's announcements:

MP6 available, due 11/17, 11:59p. EC due 11/10, 11:59p

Hashing - using "hash tables" to implement _____

Suppose we have the following info…

Locker Number	Name
103	Jay Hathaway
92	Linda Stencel
330	Bonnie Cook
46	Rick Brown
124	Kim Petersen
…	…

…and we want to be able to retrieve a name, given a locker number.

Now suppose our keys are not so nicely described…

Course Number -> Schedule info

Color -> BMP

Vertex -> Set of incident edges

Flight number -> arrival information

URL -> html page

dice roll -> payoff amt

Some general vocabulary

A *dictionary* is a structure supporting the following:

```
void insert(kType & k, dType & d)
void remove(kType & k)
dType find(kType & k)
```

An *associative array* is a dictionary w a particular interface–

Overloads the [] operator for insert and find:

```
myDictionary["Miguel"]  = 22;
dType d = myDictionary["Miguel"];
```

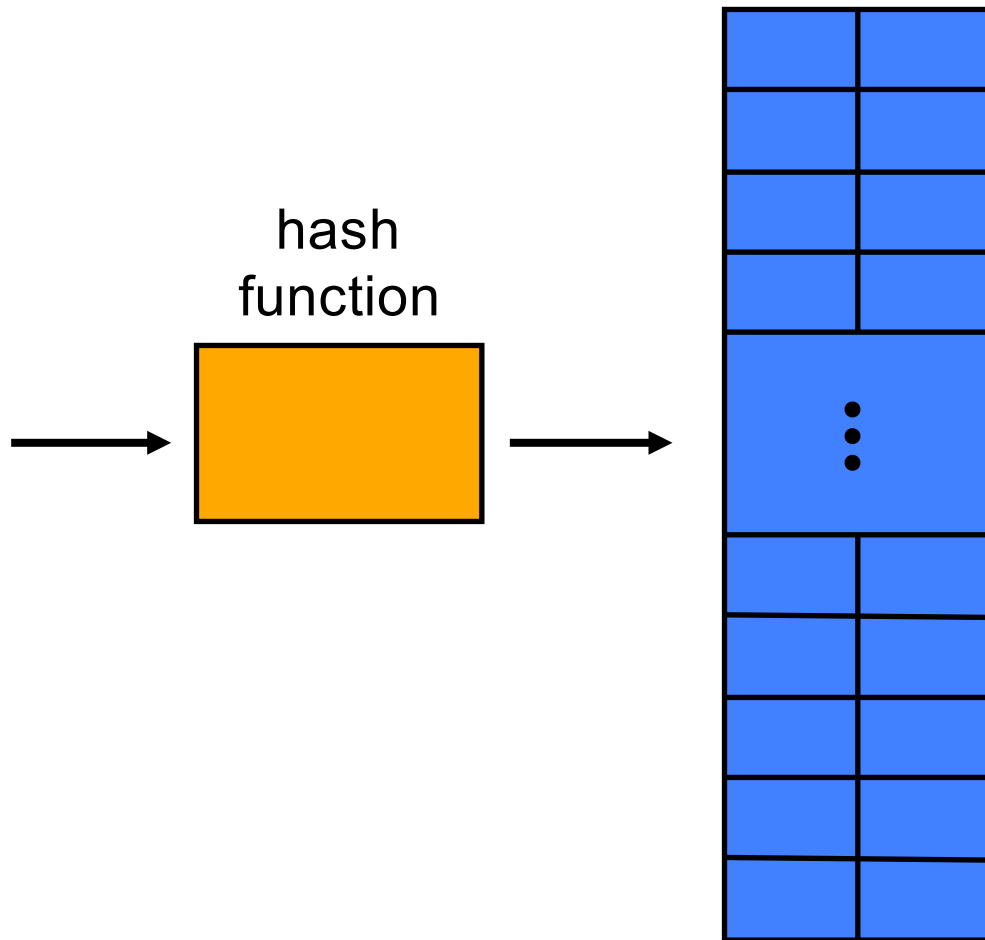|							

Hashing:

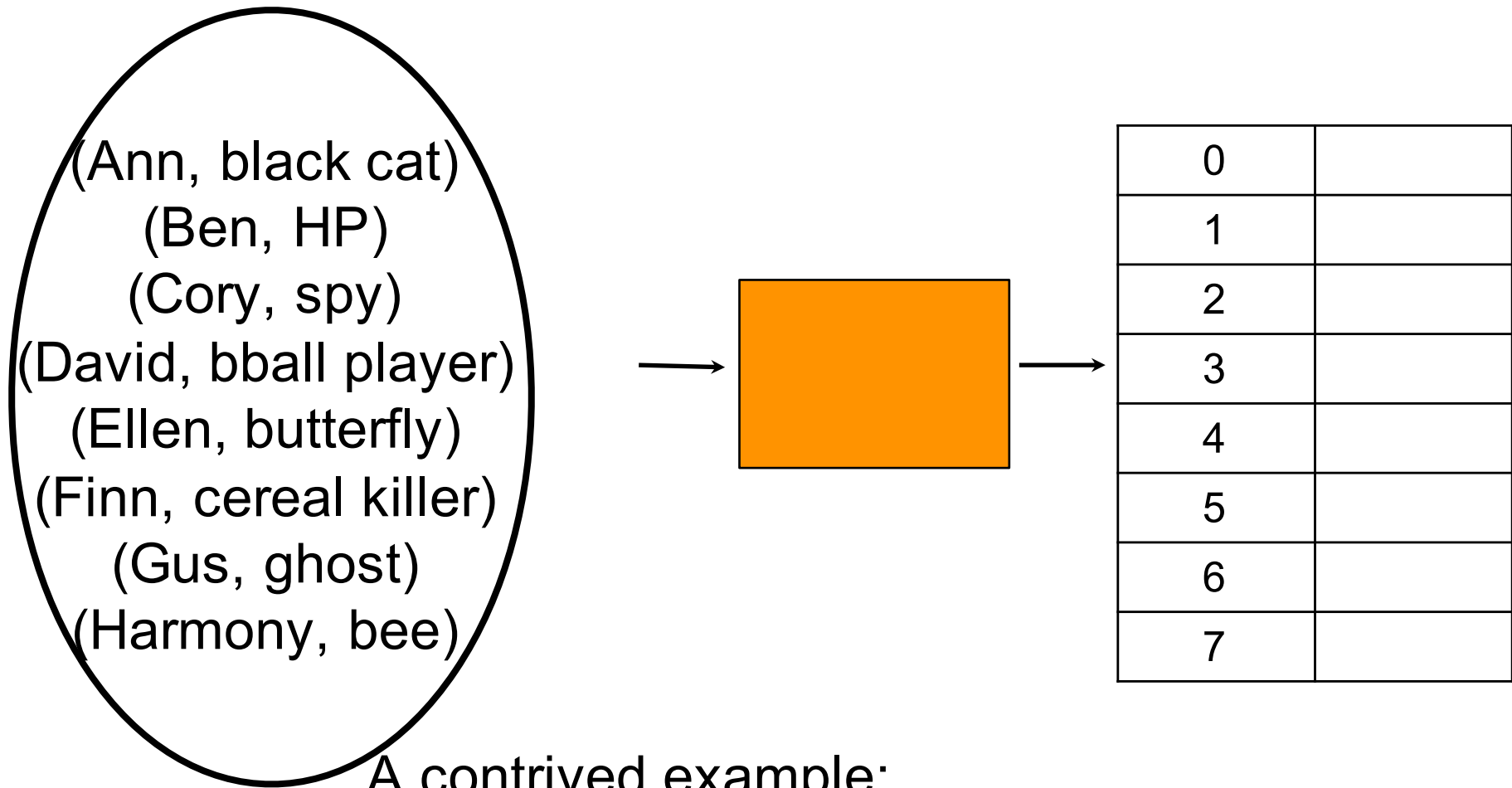(defn) Keyspace – a (mathematical) description of the keys for a set of data.

Goal: use a function to map the keyspace into a small set of integers.

What's fuzzy about this goal?

Basic Idea: we seek a mapping, `h(k)`

hash
function

A perfect hash function:

(Ann, black cat)
(Ben, HP)
(Cory, spy)
(David, bball player)
(Ellen, butterfly)
(Finn, cereal killer)
(Gus, ghost)
(Harmony, bee)

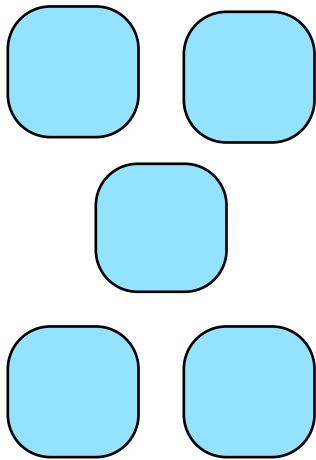|	
0	
1	
2	
3	
4	
5	
6	
7	

A contrived example:
these keys have a fabulous hash fn.
   a. each key hashes to a different int
   b. collection of keys hash to a seq of ints

A perfect hash function?

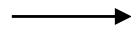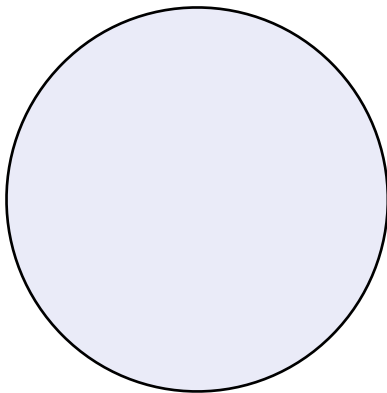Roll 5 dice:

Is this a good hash function?

0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	

A Hash Table consists of:

•An array

•A hash function

•A collision resolution strategy

KeySpace

0	1		...				N-1

Easy, if
|KeySpace| ~ N

keys used:
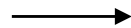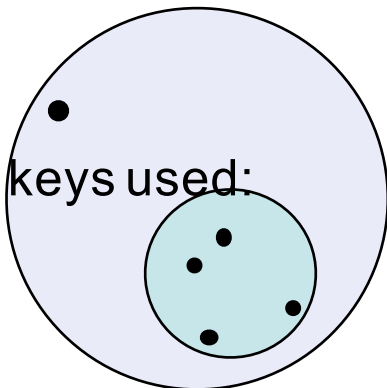
0	1		...				N-1

Hash Functions:

- Consist of 2 parts:
  - A Hash: Function mapping a key to an integer `i`
  - A compression: function mapping `i` into the array cells 0 to N-1.

- Choosing a hash function is tricky...
  - Don't create your own (yet)
  - Smart people can produce poor hash functions (what's a bad hash function?)
    - Knuth's multiplicative hash in "the Art of Computer Programming"

- Characteristics:
  - Computed in _____ time.

  - Deterministic.

  - Satisfy the SUHA.