CSCI 241

Scott Wehrwein

Hash Tables:
Collisions, Chaining, Load Factor

Goals

Understand how a Hash Table can be used to store a set of integers.

Know the definition of a collision and how to use the chaining strategy for collision resolution.

Know how to calculate the load factor of a hash table.

Direct Address Table

Why was this so easy?

The Set contents came from a small, fixed domain of possible values (e.g., 0..10).

Sets cannot have duplicates.

How can we make it useful?

Map any value onto the fixed domain (e.g., 0..10) using a hash function.

Reminder:
The Modulus Operator

`a % b` gives the remainder when dividing `a` by `b`:

```
12 % 8 => 4

24 % 10 => 4

4 % 10 => 4

28 % 14 => 0
```

Hash Functions

A hash function is a function that maps a value from some large (possibly infinite) domain to a non-negative integer that can be used as an array index.

Example: `h(x) = x % 10`

`h : int` → `0..10`

boolean[] A:

0	F
1	F
2	F
3	F
4	T
5	F
6	F
7	F
8	F
9	F

Hash Tables with Integers

A hash table stores a value at an index determined by their hash value (aka hash code).

insert(14)   (14 % 10) => 4

boolean[] A:

0	F
1	F
2	F
3	F
4	T
5	F
6	F
7	F
8	F
9	F

Hash Tables with Integers

A hash table stores a value at an index determined by their hash value (aka hash code).

boolean[] A:

0	F
1	F
2	F
3	F
4	T
5	F
6	F
7	F
8	F
9	F

insert(14) (14 % 10) => 4

contains(14)

uh oh...

Problem: which value was it?

Hash Tables with Integers

A hash table stores a value at an index determined by their hash value (aka hash code).

insert(14)   (14 % 10) => 4

contains(14)   (14 % 10) => 4   `true`

int[] A:

0	F
1	F
2	F
3	F
4	14
5	F
6	F
7	F
8	F
9	F

Hash Tables with Integers

A hash table stores a value at an index determined by their hash value (aka hash code).

int[] A:

0	F
1	F
2	F
3	F
4	14
5	F
6	F
7	F
8	F
9	F

insert(14) (14 % 10) => 4

contains(14) (14 % 10) => 4 `true`

insert(4)

uh oh...

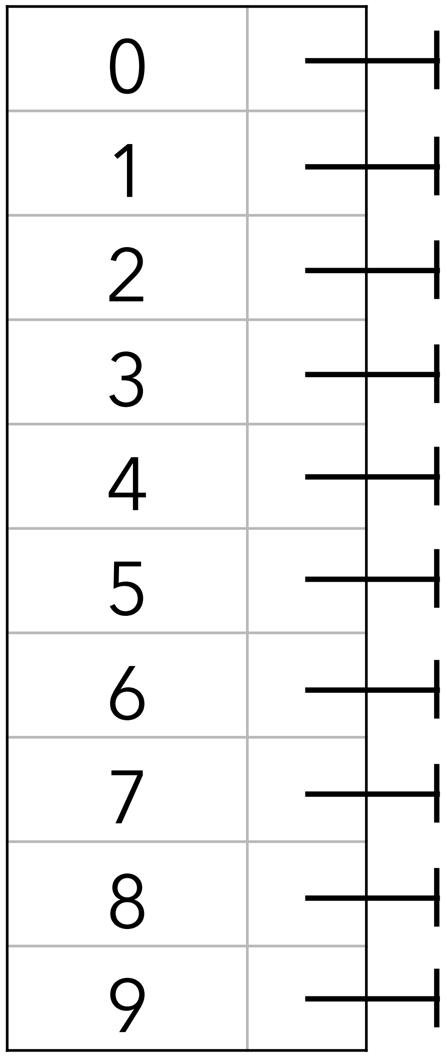
Problem: which values were they?

Hash Tables with Integers

A hash table stores a value at an index determined by their hash value (aka hash code).

insert(14)

LinkedList<Integer>[] A:



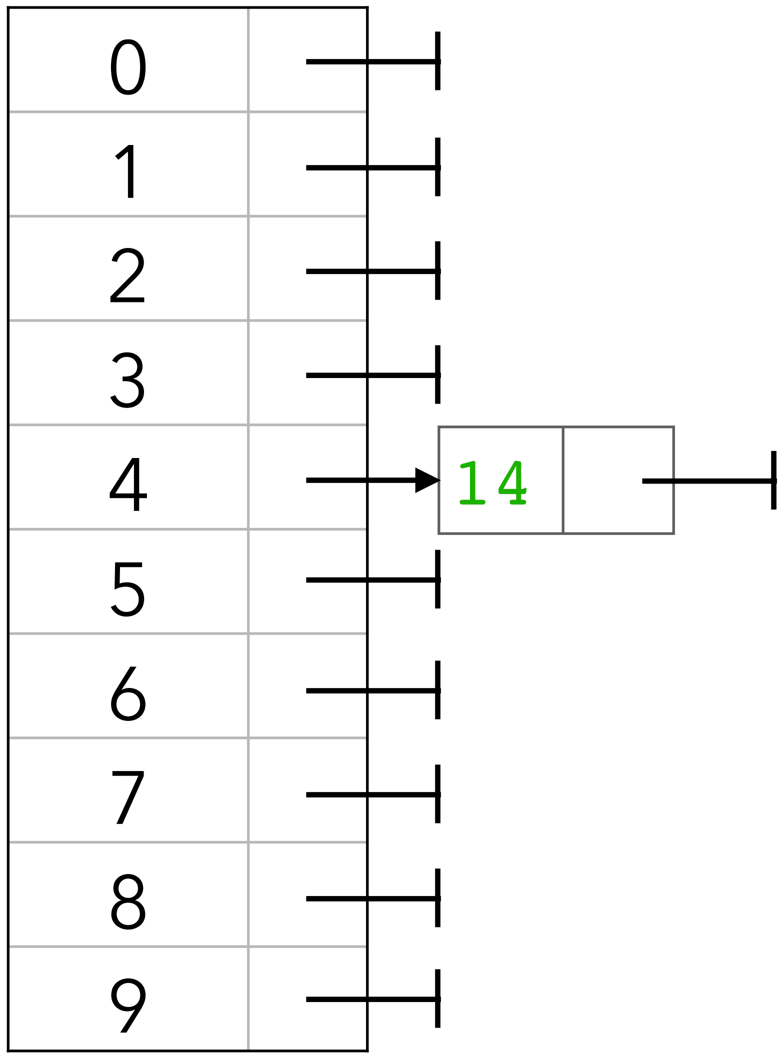
Hash Tables with Integers

A hash table stores a value at an index determined by their hash value (aka hash code).

LinkedList<Integer>[] A:

insert(14)

contains(14) `true`



Hash Tables with Integers

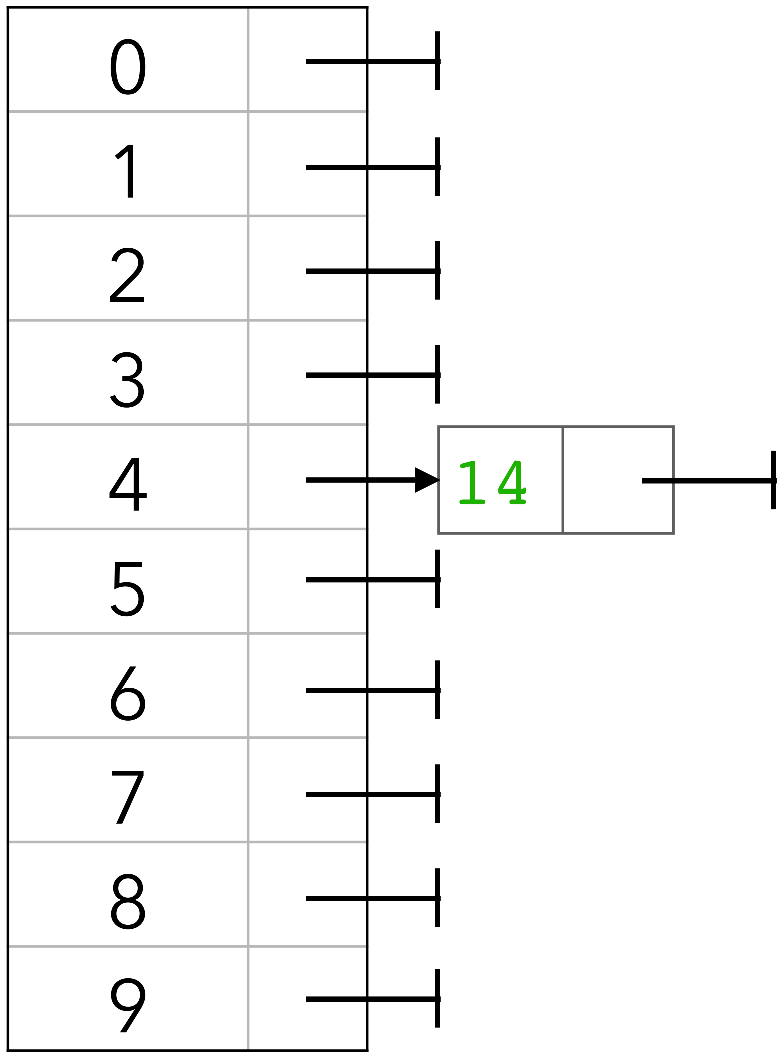
A hash table stores a value at an index determined by their hash value (aka hash code).

insert(14)

contains(14) `true`

insert(4)

LinkedList<Integer>[] A:



0
1
2
3
4 → 14
5
6
7
8
9

Hash Tables with Integers

A hash table stores a value at an index determined by their hash value (aka hash code).

insert(14)

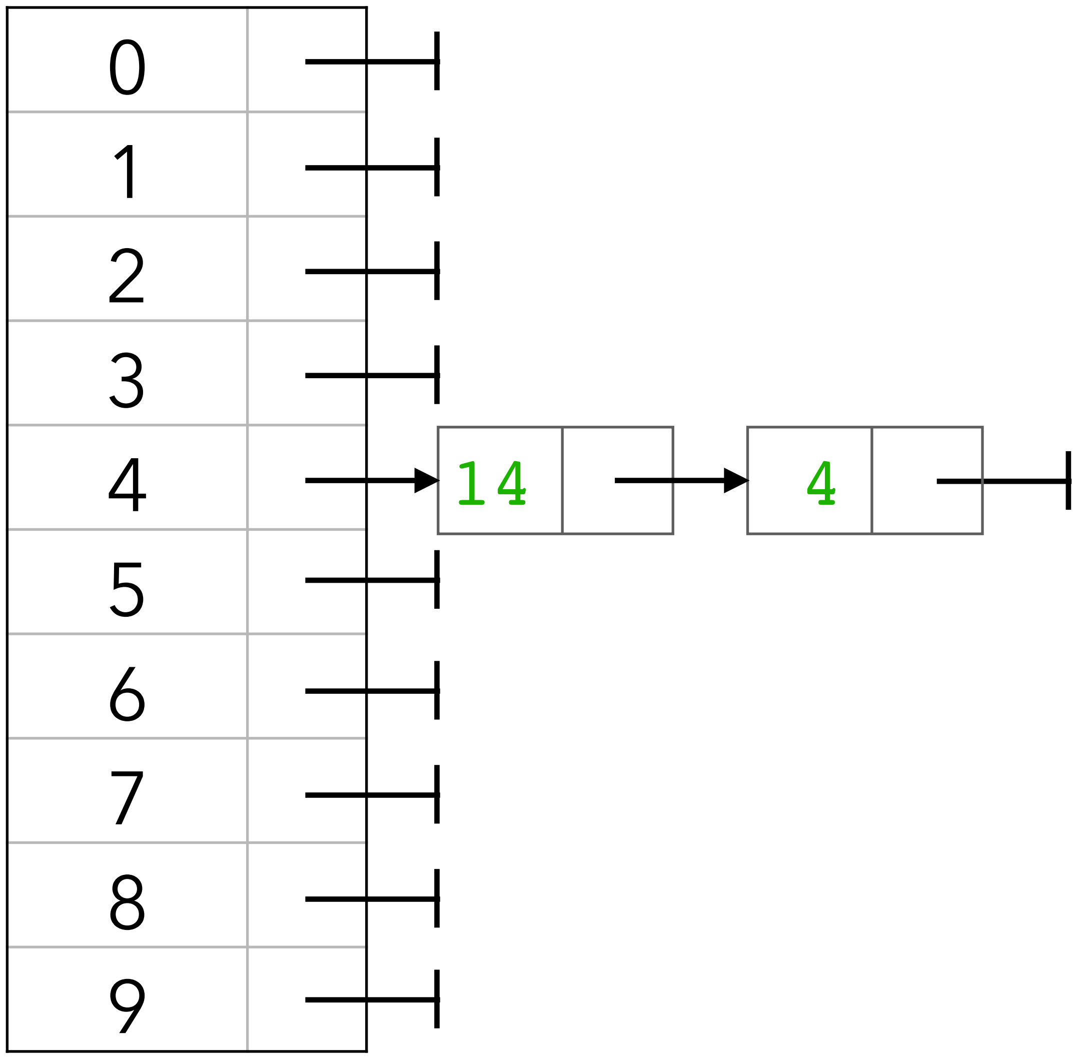
contains(14) `true`

insert(4)

This is a collision: when two values map to the same bucket.

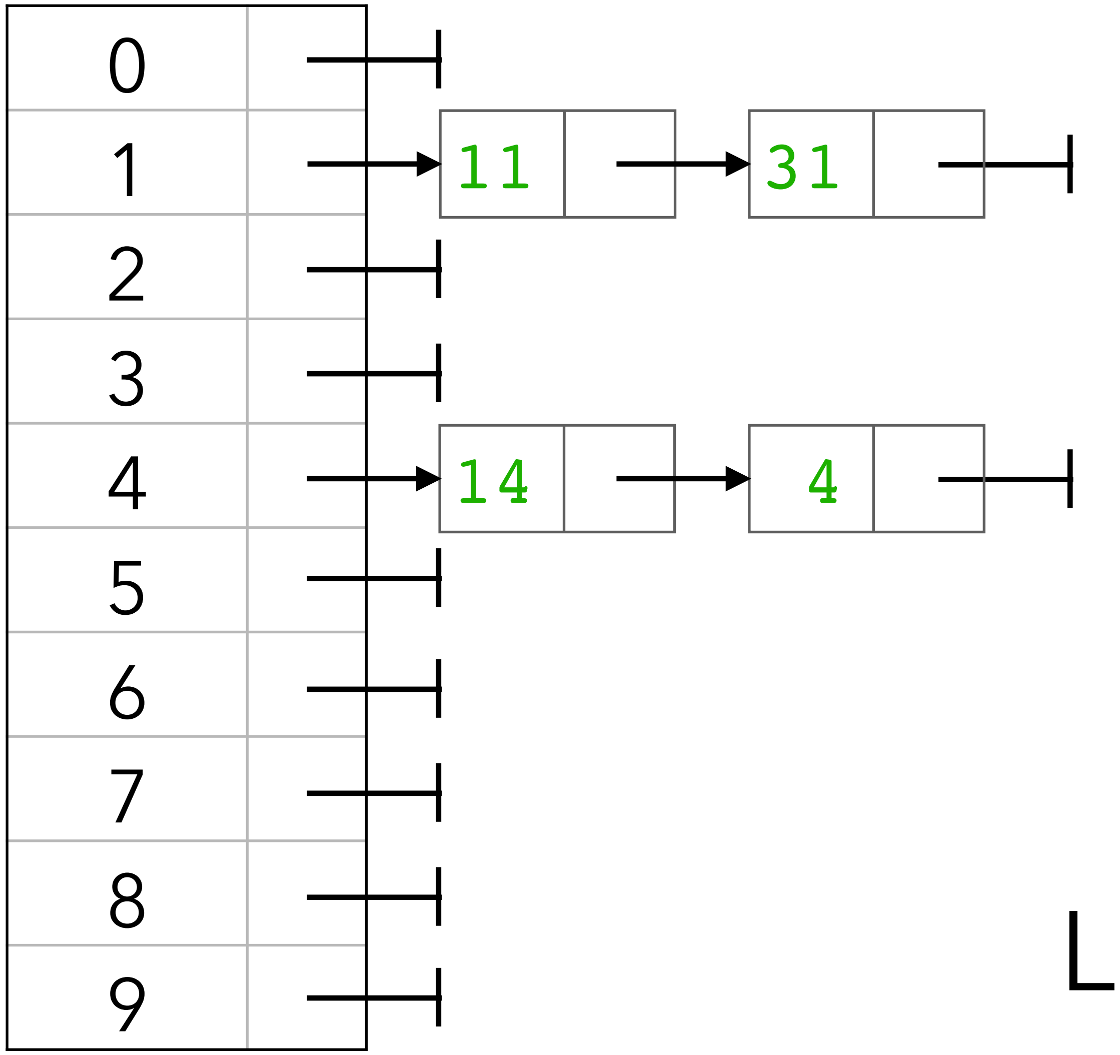
This hash table uses chaining for collision resolution.

LinkedList<Integer>[] A:



Hash Tables: Load Factor

$$\text{Load factor } \lambda = \frac{\text{\# entries in table}}{\text{size of the array}}$$



4 entries

10 buckets

Load factor: 0.4