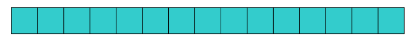
ExtendableArrays (aka, ArrayLists)

David Kauchak
cs62
Spring 2021

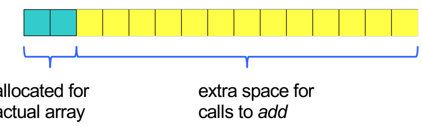
1

Extendable arrays



Allocate extra, unused memory and save room to add elements

For example: new ArrayList(2)

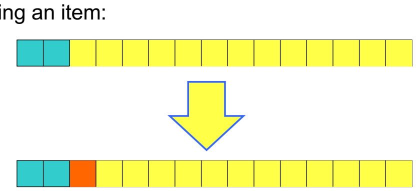


2

Extendable arrays

Allocate extra, unused memory and save room to add elements

Adding an item:



Running time: O(1) Problems?

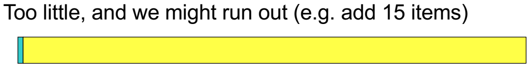
3

Extendable arrays

Allocate extra, unused memory and save room to add elements

How much extra space do we allocate?

Too little, and we might run out (e.g. add 15 items)



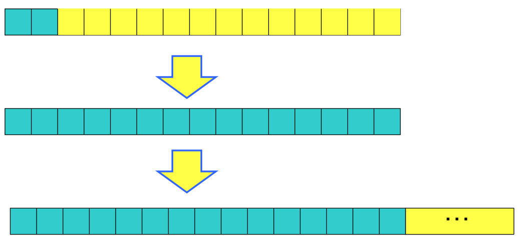
Too much, and we waste lots of memory Ideas?

4

Extendable arrays

Allocate some extra memory and when it fills up, allocate some more and ***copy the current data***

For example: new ArrayList(2)

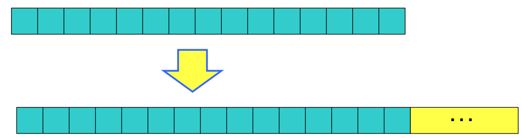


5

Extendable arrays

Allocate some extra memory and when it fills up, allocate some more and ***copy the current data***

For example: new ArrayList(2)



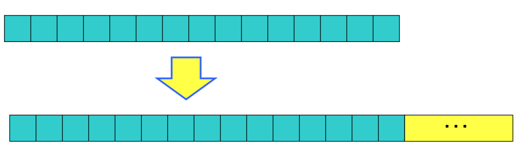
Running time: O(n)

6

Extendable arrays

Allocate some extra memory and when it fills up, allocate some more and ***copy the current data***

For example: new ArrayList(2)



How much extra memory should we allocate?

7

Extendable arrays

Allocate some extra memory and when it fills up, allocate some more and ***copy the current data***

For example: new ArrayList(2)

What is the best case running time of add? O(1)

What is the worst case running time of add? O(n)

Can we bound this tighter?

8

Extendable arrays

Challenge: most of the calls to *add* will be O(1)

How else might we talk about runtime?

What is the ***average*** running time of *add* in the ***worst case*** for **multiple calls to add**?

Note this is different than the *average-case* running time

9

Amortized analysis

What does "amortize" mean?

am·or·tized | **am·or·tiz·ing**

Definition of AMORTIZE

1 **:** to pay off (as a mortgage) gradually usually by periodic payments of principal and interest or by payments to a sinking fund

2 **:** to gradually reduce or write off the cost or value of (as an asset) <*amortize* goodwill> <*amortize* machinery>

— **am·or·tiz·able** *adjective*

10

Amortized analysis

There are many situations where the worst case running time is bad

However, if we average the operations over *n* operations, the average time is more reasonable

This is called *amortized* analysis

- This is different than average-case running time, which requires reasoning about the input/situations that the method will be called
- The worse case running time doesn't change

11

Amortized analysis

Many approaches for doing amortized analysis

Aggregate method

- figure out the big-O runtime for a sequence of *n* calls
- divide by *n* to get the average run-time per call

12

Amortized analysis

Assume we start with an empty array with 1 location and we **double the size** of the arraylist each time we fill it up

What is the cost to insert n items?

total_cost(n) = basic_cost(n) + double_cost(n)

```
public void add(E item) {
    if (size == data.length){
        resize(2 * data.length);
    }
    data[size] = item;
    size++;
}
```

double_cost

basic_cost

13

basic cost(n)

What is the cost for one add? O(1)

```
public void add(E item) {
    if (size == data.length){
        resize(2 * data.length);
    }
    data[size] = item;
    size++;
}
```

14

basic cost(n)

What is the cost for n adds? O(n)

```
public void add(E item) {
    if (size == data.length){
        resize(2 * data.length);
    }
    data[size] = item;
    size++;
}
```

15

double_cost(n)

Start with array of size one

16

double_cost(n)

Start with array of size one

double_cost

number added: 1

17

double_cost(n)

Start with array of size one

double_cost

What needs to happen now?

number added: 2

18

double_cost(n)

Start with array of size one

double_cost

1

number added: 2

19

double_cost(n)

Start with array of size one

double_cost

1

number added: 2

20

double_cost(n)

Start with array of size one

double_cost

1

What needs to happen now?

number added: 3

21

double_cost(n)

Start with array of size one

double_cost

1

2

number added: 3

22

double_cost(n)

Start with array of size one

double_cost

1

2

number added: 3

23

double_cost(n)

Start with array of size one

double_cost

1

2

number added: 4

24

double_cost(n)

Start with array of size one

double_cost
1
2

What needs to happen now?

number added: 5

25

double_cost(n)

Start with array of size one

double_cost
1
2
4

What needs to happen now?

number added: 5

26

Amortized analysis

Assume we start with an empty array with 1 location and we **double the size** of the arraylist each time we fill it up

What is the cost to insert n items?

$$\text{total\_cost}(n) = \text{basic\_cost}(n) + \text{double\_cost}(n)$$

$$\text{basic\_cost}(n) = O(n) \qquad \text{double\_cost}(n) \leq 1+2+4+8+16+\ldots+n = 2n$$

$$\text{total\_cost}(n) = O(n)$$

amortized *O(1)*

27

Amortized analysis vs. worse case

What is the worst case for *add*?

- Still O(n)
- If you have an application that needs it to be O(1), this implementation **will not work!**

amortized analysis give you the cost of *n* operations (i.e. average cost) **not** the cost of any individual operation

28

Extendable arrays

What if instead of doubling the array, we increase the array by a fixed amount (call it *k*) each time

Is the amortized run-time still O(1)?

- No!
- Why?

29

Amortized analysis

Consider the cost of *n* insertions for some constant *k*

$$\text{total\_cost(n)} = \text{basic\_cost(n)} + \text{double\_cost(n)}$$

$$\text{basic\_cost(n)} = O(n)$$

$$\text{double\_cost(n)} = k+2k+3k+4k+5k+\ldots+n$$

$$= \sum_{i=1}^{n/k} ki$$

$$= k\sum_{i=1}^{n/k} i$$

$$= k\frac{\frac{n}{k}\left(\frac{n}{k}+1\right)}{2} = O(n^2)$$

30

Amortized analysis

Consider the cost of *n* insertions for some constant *k*

$$\text{total\_cost(n)} = O(n) + O(n^2)$$

$$= O(n^2)$$

amortized *O(n)!*

31