Hashing

- Hashing as a Dictionary implementation
- Need a hash function
- Need collision resolution
  - Separate Chaining
  - Open Addressing (linear, quadratic)
  - Other Approaches (second hash function)
Rehashing

Idea: When the table gets too full, create a bigger table (usually 2x as large) and hash all the items from the original table into the new table.

• When to rehash?
  › Separate chaining: full ($\lambda = 1$)
  › Open addressing: half full ($\lambda = 0.5$)
  › When an insertion fails
  › Some other threshold

• Cost of a single rehashing? $O(N)$
Rehashing Example

• Separate chaining example:
  \[ h_1(x) = x \mod 5 \] rehashes to \[ h_2(x) = x \mod 11. \]

\[
\begin{array}{c}
\lambda = 1 \\
\begin{array}{cccc}
0 & 1 & 2 & 3 \\
25 & 37 & 83 & \text{full} \\
\end{array} \\
\lambda = 5/11 \\
\begin{array}{cccccccc}
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
25 & 37 & 83 & \text{full} & \text{full} & \text{full} & \text{full} & \text{full} \\
\end{array}
\end{array}
\]
• Starting with table of size 2, double when load factor > 1.

Amortized analysis can show cost of inserting m keys < 3m
Hashing Summary

- Hash tables are one of the most important data structures.
- Hashing has many applications where operations are limited to find, insert, and delete.
- Can use either separate chaining or open hashing
  › Java uses separate chaining
- Rehashing has good amortized complexity.
- Also has a big data version to minimize disk accesses: extendible hashing (see textbook)
Java `hashCode()` Method

• Class `Object` defines a `hashCode()` method
  › Intent: returns a suitable hashcode for the object
  › Result is arbitrary int; must scale to fit a hash table (e.g. `obj.hashCode() % nBuckets`)
  › Used by collection classes like `HashMap`

• Classes should override with calculation appropriate for instances of the class
  › Calculation should involve semantically “significant” fields of objects
hashCode() and equals()

• To work correctly, particularly with collection classes (like HashMap), an Object’s hashCode() and equals() must obey this rule:

  if a.equals(b) then it must be true that a.hashCode() == b.hashCode()

• Why? Is the reverse also required?
A Generally Good hashCode()

- int result = 17;
- foreach field f
  - int fieldHashcode =
    - boolean: (f ? 1: 0)
    - byte, char, short, int: (int) f
    - long: (int) (f ^ (f >>> 32))
    - float: Float.floatToIntBits(f)
    - double: Double.doubleToLongBits(f), then above
    - Object: object.hashCode()
  - result = 31 * result + fieldHashcode
Really Great Practical Book

- Effective Java
- Joshua Bloch
1) 10 Points

Compute an appropriately tight $O$ (Big-Oh) bound on the running time of each code fragment, in terms of $n$. Assume integer arithmetic. Circle your answer for each fragment.

a)  
```plaintext
for(i = 0; i < n; i++) {
    for(j = 0; j < i; j++) {
        for(k = 0; k < j; k++) {
            sum++;
        }
    }
}
```
O($n^3$)

b)  
```plaintext
for(i = 0; i < n; i++) {
    for(j = 0; j < n; j++) {
        if(i == j) {
            for(k = 0; k < i; k++) {
                sum++;
            }
        } else {
            for(k = 0; k < i * i; k++) {
                sum++;
            }
        }
    }
}
```
O($n^4$)

c)  
```plaintext
for(i = n; i > 0; i = i / 2) {
    for(j = 0; j < n; j++) {
        sum++;
    }
}
```
O($n \log n$)

d)  
```plaintext
for(i = n; i > 0; i = i - (n / 2)) {
    for(j = 0; j < n; j++) {
        for(k = 0; k < n; k++) {
            sum++;
        }
    }
}
```
O($n^2$)
2) 10 Points

Prove by induction that:

\[ n^2 \geq 2n \text{ for every integer } n \geq 2 \]

**Base Case:**

\( n = 2 \quad 2^2 = 4 \geq 4 = 2 * 2 \)

**Inductive Hypothesis:**

\( n^2 \geq 2n \text{ for } 2 \leq k \leq n \)

**Show:**

\[ (k + 1)^2 \geq 2(k + 1) \]

\[ k^2 + 2k + 1 \geq 2k + 2 \]

\[ k^2 \geq 1 \]

\[ k^2 \geq 2k \geq 1 \text{ for } k \geq 2 \]
3) 10 Points

Consider the structure of tree a and tree b (node keys are intentionally omitted).

```
  tree a     tree b
    X    X
   / \   / \ 
  X   X  X   X
 / \ / \ / \ / \ 
/ x x x x x x
x x x x x 
```

For each question, answer "tree a", "tree b", "neither", or "both".

a) Which are perfect?
   neither

b) Which are complete?
   both

c) Which are full?
   both

d) Which could be the structure of an array-based binary heap?
   both

e) Which could be the structure of a leftist heap?
   both

f) Which could be the structure of a skew heap?
   both

g) Which could appear within a binomial queue?
   tree a

h) Which could be the structure of a binary search tree?
   both

i) Which could be the structure of an AVL tree?
   both

j) Which could be the structure of a splay tree?
   both
4) 10 Points

Consider the following binary min-heap:

```
    4
   / \
  7   10
 / \   \
4   12
```

Perform the following operations in order, drawing the resulting heap after each operation and using it as the starting point for the next operation. You need only show the result of the operation, but showing your work will allow us to award partial credit. If the space here is insufficient, use the back of this sheet (clearly labeling your work). Circle the final state of the queue so we can distinguish it from intermediate work.

a) DeleteMin

![DeleteMin](image1)

b) Insert 9

![Insert 9](image2)

c) Insert 6

![Insert 6](image3)

d) DeleteMin

![DeleteMin](image4)

e) DeleteMin

![DeleteMin](image5)

f) Draw an array-based representation of your heap from step e.

![Array Representation](image6)

g) In your array-based representation, what is the index of:

- the parent of the node at index $i$: $\frac{i}{2}$
- the left child of the node at index $i$: $2i$
- the right child of the node at index $i$: $2i + 1$
5) 10 Points

Consider the following leftist

Perform the following operations in order, drawing the resulting heap after each operation and using it as the starting point for the next operation. You need only show the result of the operation, but showing your work will allow us to award partial credit. If the space here is insufficient, use the back of this sheet (clearly labeling your work). Circle the final state of the queue so we can distinguish it from intermediate work.

a) Annotate the above tree with the null-path length of each node.

b) Insert 7

c) DeleteMin

d) DeleteMin
6) 10 Points

Consider the following skew heap:

```
  2
 / \  
 a   5
 / \  
12   a
 /  
10
```

Perform the following operations in order, drawing the resulting heap after each operation and using it as the starting point for the next operation. You need only show the result of the operation, but showing your work will allow us to award partial credit. If the space here is insufficient, use the back of this sheet (clearly labeling your work). Circle the final state of the queue so we can distinguish it from intermediate work.

a) Insert 7

```
  2
 / \  
 a   5
 / \  
12   a
 /  
10
```

b) DeleteMin

```
  2
 / \  
 a   5
 / \  
7   12 8
  
```

```
  4
 / \  
5   8
 / \  
7   12 8
  
```

```
  4
 / \  
5   8
 / \  
7   12 8
  
```

```
  4
 / \  
5   8
 / \  
7   12 8
  
```

Unique ID: «Unique_ID»
7) 10 Points

Consider the following binomial queue min-heap, which currently contains 3 trees:

```
  4
 / \ 8
/   \  \
16  14
```

a) Perform a deleteMin operation on this binomial queue. You need only show the result of the operation, but showing your work will allow us to award partial credit. If the space here is insufficient, use the back of this sheet (clearly labeling your work). Circle the final state of the queue so we can distinguish it from intermediate work.

b) If your answer for step a is correct, is it the only possible correct answer? Why?

No, made an arbitrary choice when choosing which two trees of size 4 to combine

c) If you merged four identical copies of your result from step a, the resulting binomial queue would contain how many trees of what size?

```
13 x 4 items = 52 items = 32 + 16 + 4
```
8) 10 Points

Consider the following AVL tree:

```
  10
 / \  
  5  15
   /  
  2
```

Perform the following operations in order, drawing the resulting tree after each operation and using it as the starting point for the next operation. You need only show the result of the operation, but showing your work will allow us to award partial credit. If the space here is insufficient, use the back of this sheet (clearly labelling your work). Circle the final state of the queue so we can distinguish it from intermediate work.

a) Insert 1

```
  10
 /     \  
  5       15
   /  \   
  2     7
```

b) Insert 7

```
  10
 /     \  
  5       15
   /  \   
  2     7
```

c) Insert 13

```
  10
 /     \  
  5       15
   /  \   
  2     7
```

d) Insert 11

```
  10
 /     \  
  5       15
   /  \   
  2     7
```
9) 10 Points

Consider the following splay tree:

```
10
 / \
9  15
 / \
2
```

Perform the following operations in order, drawing the resulting tree after each operation and using it as the starting point for the next operation. You need only show the result of the operation, but showing your work will allow us to award partial credit. If the space here is insufficient, use the back of this sheet (clearly labeling your work). Circle the final state of the queue so we can distinguish it from intermediate work.

a) Insert 1

![Diagram of tree after inserting 1]

b) Insert 7

![Diagram of tree after inserting 7]

c) Delete 5

![Diagram of tree after deleting 5]