1. (20%) Runtime analysis. Given the following piece of code:

```java
public static void mystery(int n)
{
    if (n==1) {
        System.out.println("Something");
    }
    else {
        System.out.println("Something");
        mystery(n/2);
    }
}
```

Let $T(n)$ be the running time of mystery in terms of its argument $n$.

a. (5%) write the $T(n)$ formula as shown in class and explain briefly.

$$T(n) = T(n/2) + C$$
where C is a constant. The program makes a recursive call of size $n/2$ and a constant number of operations (if statement + boundary condition).

b. (5%) What is the runtime of this function based on the $T(n)$ formula you derived? The runtime is $O(\log n)$, similar to binary search we saw in class - every call cuts the input size by half and, so $O(\log n)$ calls at most are needed, each one does a constant operation.

c. (5%) Exactly how many times will “Something” be printed for $n = 7$?

Three times. One for $n = 7$, then for $n = 3$ and then for $n = 1$ (remember that integer division rounds down).

d. (5%) This part is unrelated to (a-c) above. A given algorithm runs as $O(2^n)$ where $n$ is the input size. If the algorithm takes 10 seconds to run on an input of size 5, what is the input size that makes the algorithm run for approximately 40 seconds? Explain. The answer is 7.

If $O(2^5) \approx 10$ seconds, then multiplying both sides by 4 gives us $4 \times O(2^5) = O(2^7) \approx 40$. Another way to look at it: If the runtime is $O(2^n)$, then adding 1 to the input size doubles the runtime, and adding 2 increases the runtime 4-fold, hence 7.
2. (25%) Java implementation and collections: For each of the following statements say whether it is possible or not using only the API collections we learned in class. Explain each answer briefly.

(a) (5%) A program has set up a Map from String to Integer, i.e. `Map<String, Integer>`. Now it wants to make two different keys, “x” and “y” map to the same Integer 6.

Yes, two different keys can map to the same values. The keys have to be unique but not the values.

(b) (5%) A program has set up a Map from String to Integer. Now it wants to set up a “reverse” map from Integer to String. All it has to do is to allocate new `Map<Integer, String>`, get all the HashEntries from the original map, reverse the role of key and value and insert it to the new map, and the result will be exactly the original map with reverse roles.

No, because values are not always unique and they won’t necessarily be able to make a unique set of keys.

(c) (5%) The program has set up a List of integers, `List<Integer>`, and added several elements to it. Now it wants to determine the smallest number in less than $O(n)$ time.

No, lists are not sorted.

(d) (5%) Same as (c) above only using a `HashSet<Integer>`.

No, HashSets are not sorted.

(e) (5%) Same as (c) above only using a `TreeSet<Integer>`.

Yes, we can find the minimum in a TreeSet in $O(1)$ or $O(\log n)$ (depending on the implementation). In any case, it’s less than $O(n)$
3. Hash tables (15%): In each one of the following questions provide a brief explanation (no more than a sentence or two).

(a) (5%) Let $H$ be a hash table where collisions are handled by separate chaining. Re-hashing is used whenever the load factor (ratio of items in the table and the size of the table) exceeds $\frac{1}{2}$. Assume that the initial size of $H$ is 2 and re-hashing doubles the size of the table. After inserting 10 items with different keys, what is the size of $H$?

The size is 32. Let’s trace the inserts: After the first two inserts we have to rehash to 4 (two and not one because the load factor has to strictly exceed $\frac{1}{2}$). After one more insert we have three items and rehash again to 8. After two more inserts we have five items and rehash to 16. After four more inserts we have 9 items and rehash to 32. Insert item 10 and we’re done.

(b) (5%) Consider an initially empty hash table $H$. In the worst case scenario, what is the time complexity (big-Oh) to insert $n$ keys into the table if separate chaining is used to resolve collisions? Suppose that each entry stores a linked list, and when adding a new element to an unordered linked list, such an element is inserted in the beginning of the list.

Every insert takes exactly $O(1)$ if we insert in the beginning of the list, because we need constant time to find the entry and constant time to insert in the beginning of a list. A total of $n$ insert gives us $O(n)$ worst case run time.

(c) (5%) In the hash table in (b) above, what would be the worst case run time if we maintained every entry as a sorted data structure, say a TreeSet (the sorting is by key)?

In the worst case scenario all the items go to the same slot, so we have to create and maintain a sorted set of $n$ keys, overall $O(n \log n)$ inserts.
4. Induction (25%). Prove the following formulas. Don’t forget to explicitly state the three stages:

- Base case
- Inductive hypothesis for $1 < k < n$
- The inductive step from $n - 1 \rightarrow n$ (or $n \rightarrow n + 1$ if it works better)

(a) (13%) For all $n \geq 1$, show that $n^3 + 2n$ is divisible by 3.

For your convenience, $(n + 1)^3 = n^3 + 3n^2 + 3n + 1$.

- Base case: When $n = 1$, $n^3 + 2n = 1 + 2 = 3$.
- Assume this is true for any $1 \leq k \leq n$.
- Prove for $n + 1$:
  \[(n + 1)^3 + 2(n + 1) = n^3 + 3n^2 + 3n + 1 + 2n + 2 = n^3 + 2n + 3n^2 + 3n + 3.\]

$n^3 + 2n$ is divisible by 3 by inductive hypothesis. The rest is multiples of 3, obviously divisible by 3.

(b) (12%) Show that every tree with $n$ nodes has exactly $n - 1$ edges. You may define a tree as an undirected graph with no cycles.

**Hint:** In the inductive step, instead of adding a node to a tree of size $n - 1$, start from a non-empty tree of size $n$ and remove an edge from it.

- Base case: A tree with one node has zero edges.
- Assume this is true for any tree of size $1 \leq k < n$.
- Prove for $n$:
  We can take any non-empty tree $T$ of size $n$ and remove one arbitrary edge from it (assume it has at least one edge. If it has no edges we’re back to the base case). Trees don’t have cycles, so removing an edge disconnects the tree and we get two non-empty subtrees – $T_1$ and $T_2$, both of size $< n$, so by inductive hypothesis $T_1$ has $m$ nodes and $m - 1$ edges. $T_2$ has $n - m$ nodes and $n - m - 1$ edges. In total, both subtrees then have $n$ nodes and $m - 1 + n - m - 1 = n - 2$ edges. Put back the edge you removed and you get the tree $T$ which has the same $n$ nodes, the edges of $T_1$ and $T_2$ and the edge you just added, so $n - 1$ in total.
(a) (13%) There are two types of professional wrestlers: "babyfaces" ("good guys") and "heels" ("bad guys"). Between any pair of professional wrestlers, there may or may not be a rivalry. Suppose we have \( n \) professional wrestlers and we have a list of \( r \) pairs of wrestlers for which there are rivalries. Give an \( O(n + r) \) time algorithm that determines whether it is possible to designate some of the wrestlers as babyfaces and the remainder as heels such that each rivalry is between a babyface and a heel. If it is possible to perform such a designation, your algorithm should produce it. You may re-use any algorithm we learned in class as-is (without re-writing all the stages).

Create a graph such that every wrestler is a vertex and there's an edge representing each rivalry. To test whether we can designate some wrestlers as babyfaces and some as heels, we should check whether the graph is bipartite (any algorithm mentioned in class or the text will do).

b. (12%) Trace the run of Depth-First search (DFS) algorithm starting from \( v_1 \) in the graph below. For tracing, use the same notation as in the class notes and HW3 solutions. In case of multiple options follow numerical order. Mark an * near the edges that participate in the final tree.