

CSE 302 Homework Assignment 4

Due: April 18th, 2025, 11:59 PM

Instructions

Answer all questions carefully and completely. **All submissions are to be made via Gradescope. This assignment does not include programming. The submission will be a single file. Also see the last page for submission guidelines.**

Problems

Q1 (30 pts)

- (1) Explain the concept of **(Re)Sorting (On Demand)**. Make certain to include why (re)sorting on demand may still be useful or necessary despite us already having access to the Array-based Sorted List (ASList) and Binary Search Tree (BST) implementations.

- (2) Explain the principal differences between the **Insertion Sort** and **Heap Sort** algorithms. Be certain you: i. Provide a general idea of how the algorithms operate; ii. Explain which algorithm is generally the most efficient in the worst-case scenario; and iii. Explain which of the two algorithms is likely preferable if a collection is already sorted or almost entirely sorted. Your response should be 1-2 paragraphs in length.

Q2 (30 pts)

- (1) Explain what the **Set ADT** is, and what distinguishes it in practice from the two List ADTs (Unsorted List and Sorted List).

- (2) Provide an example when using the **bit-vector** implementation of a set can be advantageous, and another example when it is a poor choice.

Q3 (40 pts)

- (1) Explain what a **hash function** is, and provide an example of how one can be applied to: i. numeric data; and ii. non-numeric data.

- (2) Explain what a **hash table** is, and what performance advantage it carries relative to our prior data structure implementations.

- (3) Explain what issues can hamper this performance advantage, and what steps are taken to mitigate the slow-down.

- (4) Suppose employees at a company are assigned an 8-digit employee ID according to the following criteria:
- i. The first four digits correspond to a department code (ex: the accounting department may have code 0051, the IT department code 3505, etc.)
 - ii. The second four digits correspond the order in which the employee was hired within the company on the whole (ex: the second employee would have the last four digits “0002,” the one-hundredth “0100,” etc.)

Now, suppose you are given the task to store employee information in a small hash table by index and are given the following options for the hashing function based on the employee ID:

Option 1: Compute the hash index using the first four digits of the employee ID, mod 100.

Option 2: Compute the hash index using the last four digits of the employee ID, mod 100.

Assuming the company has several hundred employees and five departments, i. Explain which Option is better for this dataset, and why, and ii. Explain how the chosen hash function can be improved with a simple modification (there may be several valid answers to part ii, but there is one that is very straightforward).

Submission

Submit your solutions via Gradescope. Be sure to include:

- Follow this template.
- All written responses in a single document (.txt, .doc, or .pdf, preferably PDF) named “LN_FN_4” where LN is your last name, and FN is your first name.
- No code submission.