LSDS Assignment 3

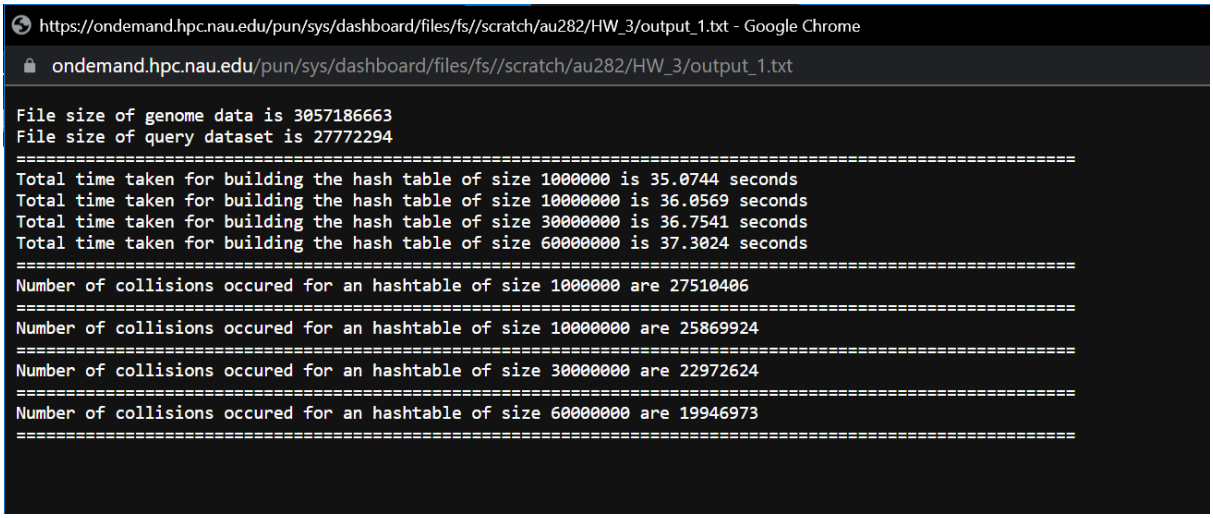
Command for running Problem A : srun homework_3 1 human.txt human_reads_trimmed.fa

Command for running Problem B : srun homework_3 2 human.txt human_reads_trimmed.fa

A) Assess the impact of the hash table size :

a. For each of your 4 hash table sizes, how many collisions did you observe while populating the hash?

Number of collisions observed while populating the hash table are mentioned in the screenshot below.



b. For each of your 4 hash table sizes, how long did it take you to populate the hash table? Do the timing results make sense? Explain.

There is no major difference here in time for populating the hash table of sizes 1 M till 60 M. This is because while populating the hash table I am inserting the nodes at the beginning of the linked list which optimizes things. As attached in the above screenshot, the time taken for populating the hash table are as follows:

For 1 Million : 35.0744 seconds
For 10 Million : 36.0569 seconds
For 30 Million : 36.7541 seconds
For 60 Million : 37.3024 seconds

B) Searching speed:

a. How long did it take to search for every possible 16-character long fragment of the subject dataset within the query dataset?

Total time taken for searching every fragment of the subject dataset in the query dataset is 4882.16 seconds.

b. How many such fragments did you find?

Total number of fragment hits are : 542456454

c. Print the first 10 fragments of the subject dataset that you found within the Query_HT.

The first 10 fragments of the subject dataset found in Query_HT is mentioned in the below screenshot.