

CPSC 319

Assignment#5

Alex Yeap

[Alex.yeap@ucalgary.ca](mailto:Alex.yeap@ucalgary.ca)

Tutorial: T10

TA: Mobin Vahdati

## Questions

1. Briefly describe your hash function and how it works.
  2. Analyze the quality of your hash function based on the statistics reported. How could you improve it?
  3. Compare your hash function with an ideal hash function. How close is it to ideal performance?
  4. If the input data set contained numerals (for example, student ID numbers) instead of alphabetical characters, how would it influence the performance of your hash function? Justify your answer with a brief analysis.
1. My hash function converts a String into a unique Integer. A string is sent to the hash function which is converted to a char array name arr we then check if the array is odd and if it is, we will for loop until the array length - 1 and then grab the ascii value of each character concat every pair and add them all up. When the for loop is finished, we would add the last char not included in the for loop with and integer 32(ascii value of a space) if the char array is even the normal procedure of the for loop occurs. The added pairs of ascii values will be returned.
  2. We could improve the hash function by producing a more unique key modulus an ideal prime number that will provide us with less collisions
  3. It is fairly close to the ideal hash function as they are about 1 number off from one another

### Hash Function

```
PS C:\Users\AlexY\OneDrive - University of Calgary\CPSC 3
spaceStorage\5f130dec903c6e81b99075f3eb2df237\redhat.java
Records:11344.0
Total Reads: 38657
Average number of reads per record:3.4077045133991537
Load Factor: 71.42677244679511%
Hash Efficiency: 20.96037733493142
The size of Longest Search Chain:81
Successfully created and written in Outputfile
```

### My Hash Function

```
Records:11254.0
Total Reads: 37282
Average number of reads per record:3.3127767904745
Load Factor: 70.86009318725601%
Hash Efficiency: 21.389933177656218
The size of Longest Search Chain:68
Successfully created and written in Outputfile
```

4. It would improve the performance as they are already ints and are all unique for example each student would have their own unique student id number therefore we won't have to do any manipulations since there wont be any identical values