

Hash Table Analysis

RUBRIC:

Extremely detailed statistical analysis using graphical annotation and a text explanation of what is happening

since need graphcal annotation, use excel make colour coded table + a graph to show the data

then explain the data

since we have data science students, use statistical analysis to explain ezpz lol

Hypothesis:

The greater the table size, the less number of collisions will occur.

Reason:

The greater the table size, the more space there is for the hash function to spread out the keys. This means that the keys will be more evenly distributed and there will be less collisions.

Hence, the greater the table size, the less number of collisions will occur.

Results:

Hash Base: 1, Table Size: 20021, Dataset: India Cities, Statistics: (2, 2, 1, 0)

Hash Base: 1, Table Size: 20021, Dataset: Australia Cities, Statistics: (32, 32, 1, 0)

Hash Base: 1, Table Size: 20021, Dataset: US Cities, Statistics: (5138, 10044, 21, 1)

Hash Base: 1, Table Size: 402221, Dataset: India Cities, Statistics: (0, 0, 0, 0)

Hash Base: 1, Table Size: 402221, Dataset: Australia Cities, Statistics: (2, 2, 1, 0)

Hash Base: 1, Table Size: 402221, Dataset: US Cities, Statistics: (240, 250, 2, 0)

Hash Base: 1, Table Size: 1000081, Dataset: India Cities, Statistics: (0, 0, 0, 0)

Hash Base: 1, Table Size: 1000081, Dataset: Australia Cities, Statistics: (0, 0, 0, 0)

Hash Base: 1, Table Size: 1000081, Dataset: US Cities, Statistics: (104, 106, 2, 0)

Hash Base: 9929, Table Size: 20021, Dataset: India Cities, Statistics: (3, 3, 1, 0)

Hash Base: 9929, Table Size: 20021, Dataset: Australia Cities, Statistics: (31, 35, 3, 0)

Hash Base: 9929, Table Size: 20021, Dataset: US Cities, Statistics: (5096, 10280, 27, 1)

Hash Base: 9929, Table Size: 402221, Dataset: India Cities, Statistics: (0, 0, 0, 0)

Hash Base: 9929, Table Size: 402221, Dataset: Australia Cities, Statistics: (1, 1, 1, 0)

Hash Base: 9929, Table Size: 402221, Dataset: US Cities, Statistics: (256, 268, 3, 0)

Hash Base: 9929, Table Size: 1000081, Dataset: India Cities, Statistics: (0, 0, 0, 0)

Hash Base: 9929, Table Size: 1000081, Dataset: Australia Cities, Statistics: (1, 1, 1, 0)

Hash Base: 9929, Table Size: 1000081, Dataset: US Cities, Statistics: (89, 90, 2, 0)

Hash Base: 250726, Table Size: 20021, Dataset: India Cities, Statistics: (3, 3, 1, 0)

Hash Base: 250726, Table Size: 20021, Dataset: Australia Cities, Statistics: (23, 24, 2, 0)

Hash Base: 250726, Table Size: 20021, Dataset: US Cities, Statistics: (5244, 10388, 24, 1)

Hash Base: 250726, Table Size: 402221, Dataset: India Cities, Statistics: (1, 1, 1, 0)

Hash Base: 250726, Table Size: 402221, Dataset: Australia Cities, Statistics: (2, 2, 1, 0)
Hash Base: 250726, Table Size: 402221, Dataset: US Cities, Statistics: (214, 218, 2, 0)
Hash Base: 250726, Table Size: 1000081, Dataset: India Cities, Statistics: (0, 0, 0, 0)
Hash Base: 250726, Table Size: 1000081, Dataset: Australia Cities, Statistics: (0, 0, 0, 0)
Hash Base: 250726, Table Size: 1000081, Dataset: US Cities, Statistics: (111, 112, 2, 0)

The first number in the statistics is the number of collisions. A collision is when two keys are mapped to the same index.

The second number is the total number of probes. A probe is

The third number is the probe max. This is the maximum number of probes that were required to find a key.

The last number is the rehash count which is the number of times the table was rehashed due to collisions.

the third number is the number of keys that were not inserted, and the fourth number is the number of keys that were inserted but not found.

Graph (made in Excel):

From testing, the factor that affects the statistics the most is how many items there are.

Every time we need to probe, it is counted as one conflict count.

Probe total is how many spots we need to probe when in total when inputting the items.

Probe max is the maximum distance of probe during all item insertion.

Statistics:

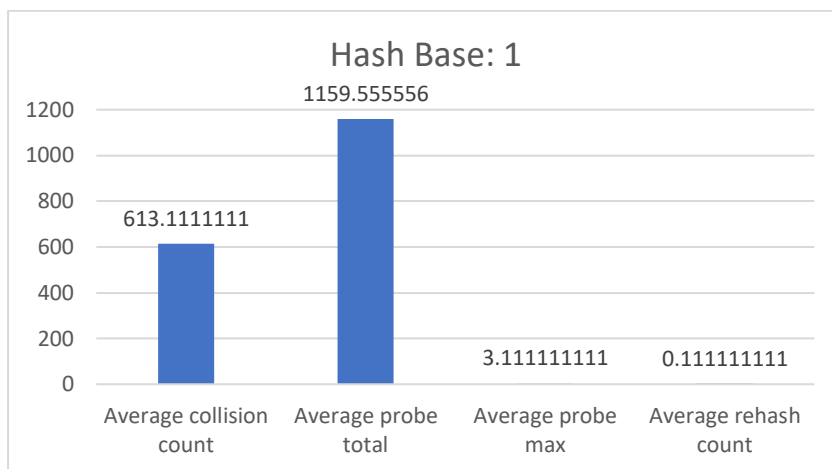
Hash Base: 1

Average Collision count : 613.11

Average Probe total: 1159.56

Average Probe max: 3.11

Average Rehash count: 0.11



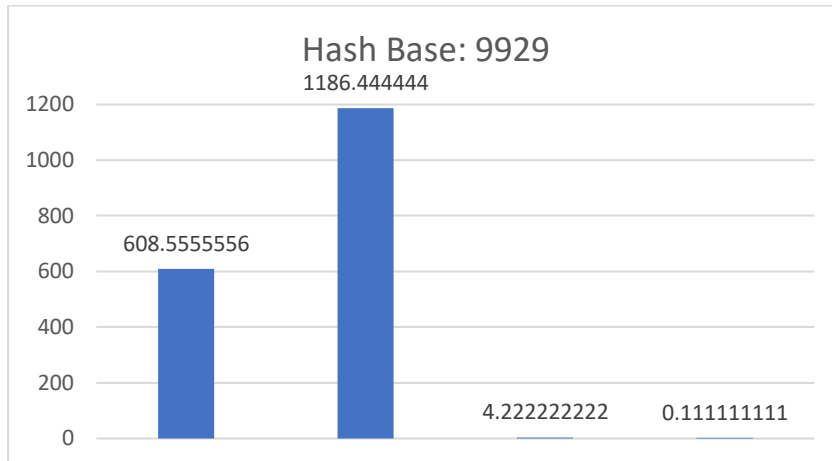
Hash Base: 9929

Average Collision count : 608.56

Average Probe total: 11586.44

Average Probe max: 4.22

Average Rehash count: 0.11



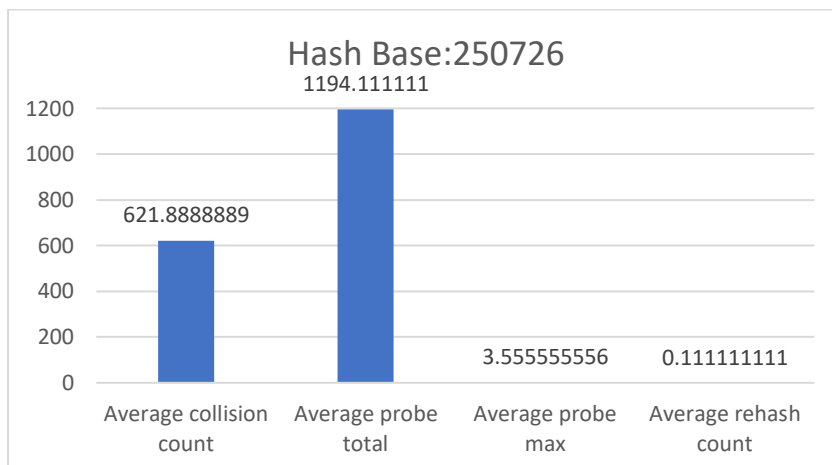
Hash Base: 250726

Average Collision count : 621.89

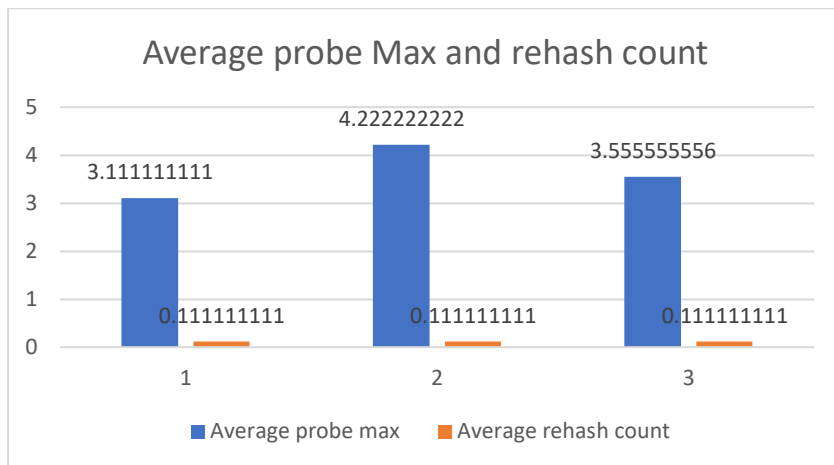
Average Probe total: 1194.11

Average Probe max: 3.55

Average Rehash count: 0.11

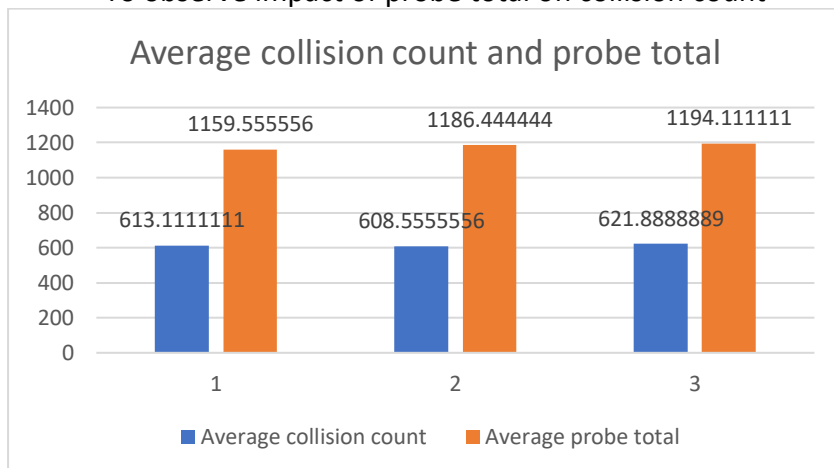


Average Probe Max and Rehash Count

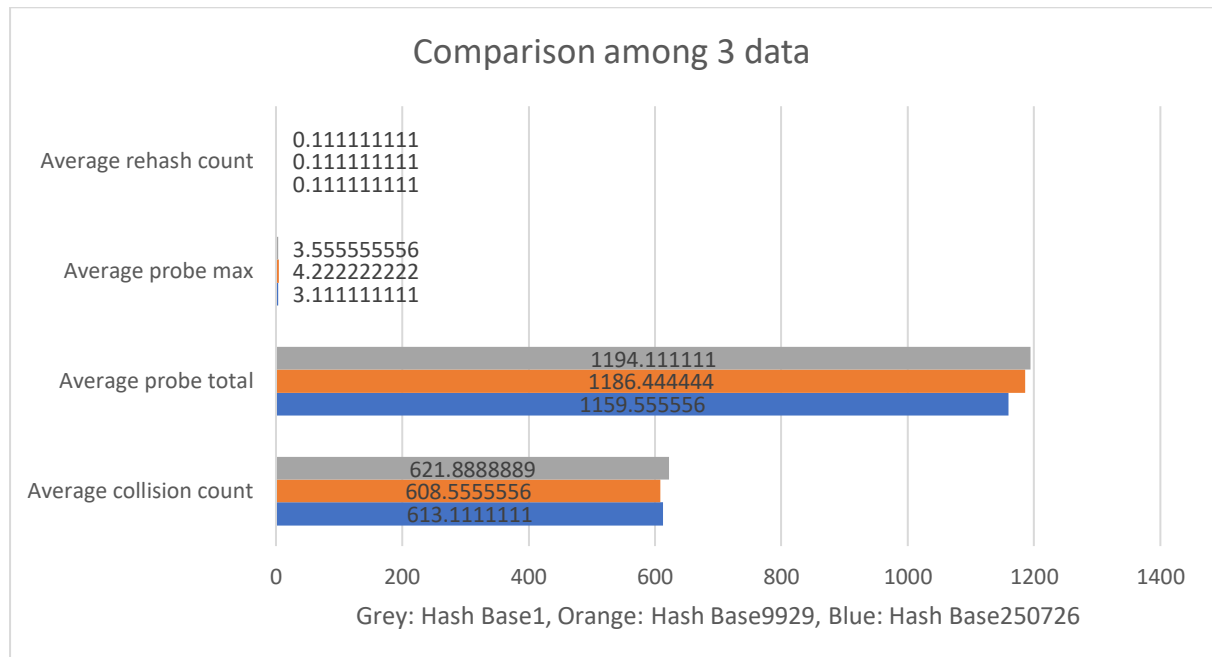


Average Collision Count and Probe Total

- To observe impact of probe total on collision count



Comparison of Hash Base 1, 9929, and 250726.



Analysis and Findings:

The greater the average probe total, the higher the value for average of collision count.

Conclusion:

In conclusion,