Part 1.4 - Report

For a given B and H, we randomly generate 10^15 pairs of 32-bit integer key and payload, and insert them into the splash table using the build method. The followings are the results:

B	R	S	H	Load Factor
2	1000	15	2	0.8967
2	1000	15	2	0.8902
2	1000	15	2	0.8937
2	1000	15	2	0.8918
2	1000	15	2	0.8927
2	1000	15	2	0.8912
2	1000	15	2	0.896
2	1000	15	2	0.8941
2	1000	15	2	0.8898
2	1000	15	2	0.8972
			Average	0.89334

B	R	S	H	Load Factor
4	1000	15	2	0.9778
4	1000	15	2	0.98
4	1000	15	2	0.977
4	1000	15	2	0.9779
4	1000	15	2	0.978
4	1000	15	2	0.976
4	1000	15	2	0.9784
4	1000	15	2	0.9767
4	1000	15	2	0.9766
4	1000	15	2	0.9777
			Average	0.9776

B	R	S	H	Load Factor
4	1000	15	3	0.9984
4	1000	15	3	0.9987
4	1000	15	3	0.9986
4	1000	15	3	0.9987
4	1000	15	3	0.9984
4	1000	15	3	0.9988
4	1000	15	3	0.9986
4	1000	15	3	0.9983
4	1000	15	3	0.9986
4	1000	15	3	0.9987
			Average	0.9986

B	R	S	H	Load Factor
4	1000	15	4	0.9998
4	1000	15	4	0.9997
4	1000	15	4	0.9997
4	1000	15	4	0.9997
4	1000	15	4	0.9999
4	1000	15	4	0.9998
4	1000	15	4	0.9997
4	1000	15	4	0.9996
4	1000	15	4	0.9995
4	1000	15	4	0.9996
			Average	0.9997

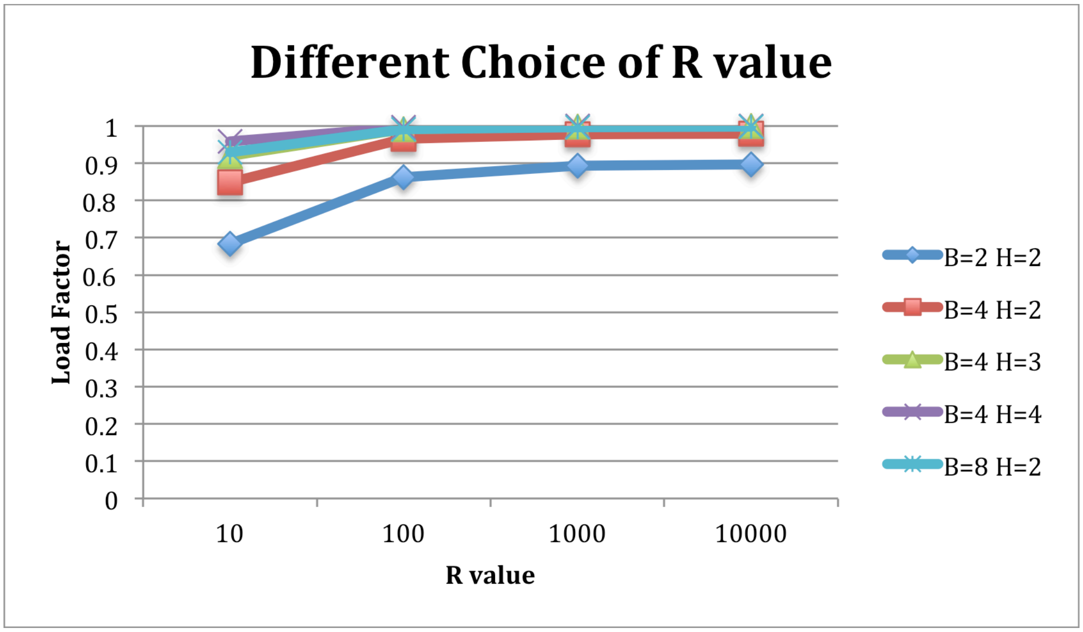
B	R	S	H	Load Factor
8	1000	15	2	0.9965
8	1000	15	2	0.9973
8	1000	15	2	0.9977
8	1000	15	2	0.9975
8	1000	15	2	0.998
8	1000	15	2	0.9973
8	1000	15	2	0.9971
8	1000	15	2	0.9978
8	1000	15	2	0.9968
8	1000	15	2	0.9966
			Average	0.9973

B	H	Load Factor
2	2	0.89334
4	2	0.97761
4	3	0.99858
4	4	0.9997
8	2	0.99726

Part 1.4 - Report

From above we can see that our results agree with the reference document. Note that we have used R = 1000 and S = 15 to give the most stable load factor.

We also observe that different R value can have impacts on load factor. For a given B, H and S, the larger the R value, the larger the load factor. However, the curve is not linear; when R reaches over 1000, it seems not to have significant impact on load factor. The following chart shows the observation.



When B = 1 and H = 2 (given R = 1000 and S = 15), we found that the load factor is about 0.5135 (ranging from 0.47 to 0.53) on average for about 50 test cases, but we did not find any earlier failure that describes in the assignment section 1.4. The load factor is lower, because the bucket size is too small (only 1), and the number of hash function is not large enough to spread the keys evenly into different buckets, thus insertion will fail earlier when comparing with previous tests.

In conclusion, our implementation shows a very similar result with the reference document. Note that we use S = 15 for all the tests, because this number is large enough yet not too small, thus giving us the best deviation.