Dawson Murphy CS300 Assignment 4

Data size 50,000

BST

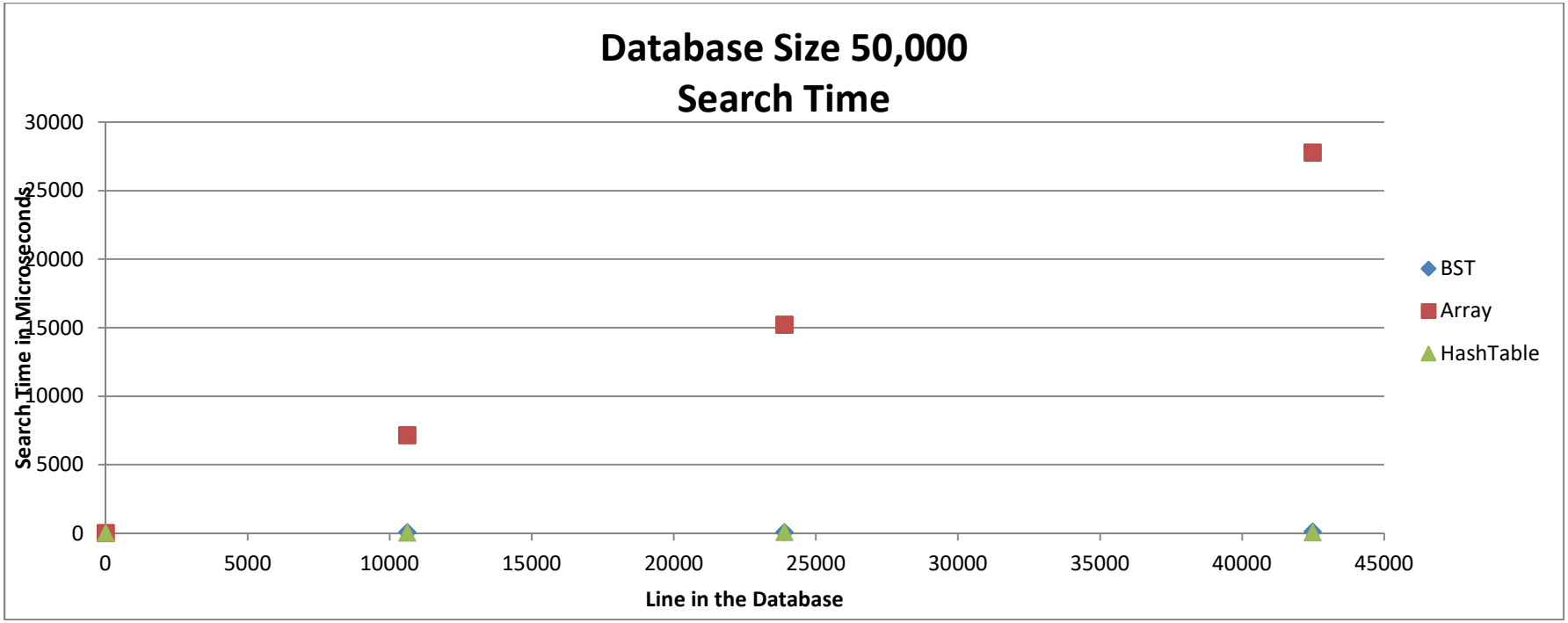
UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time
701197194311	5	8	8	8	8
799439078727	10,624	65	66	66	65.66666667
4605922001812	23,898	81	81	80	80.66666667
886970578424	42,478	104	103	103	103.3333333

HashTable

UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time
701197194311	5	7	7	7	7
799439078727	10,624	28	28	28	28
4605922001812	23,898	53	89	75	72.33333333
886970578424	42,478	56	55	55	55.33333333

Array

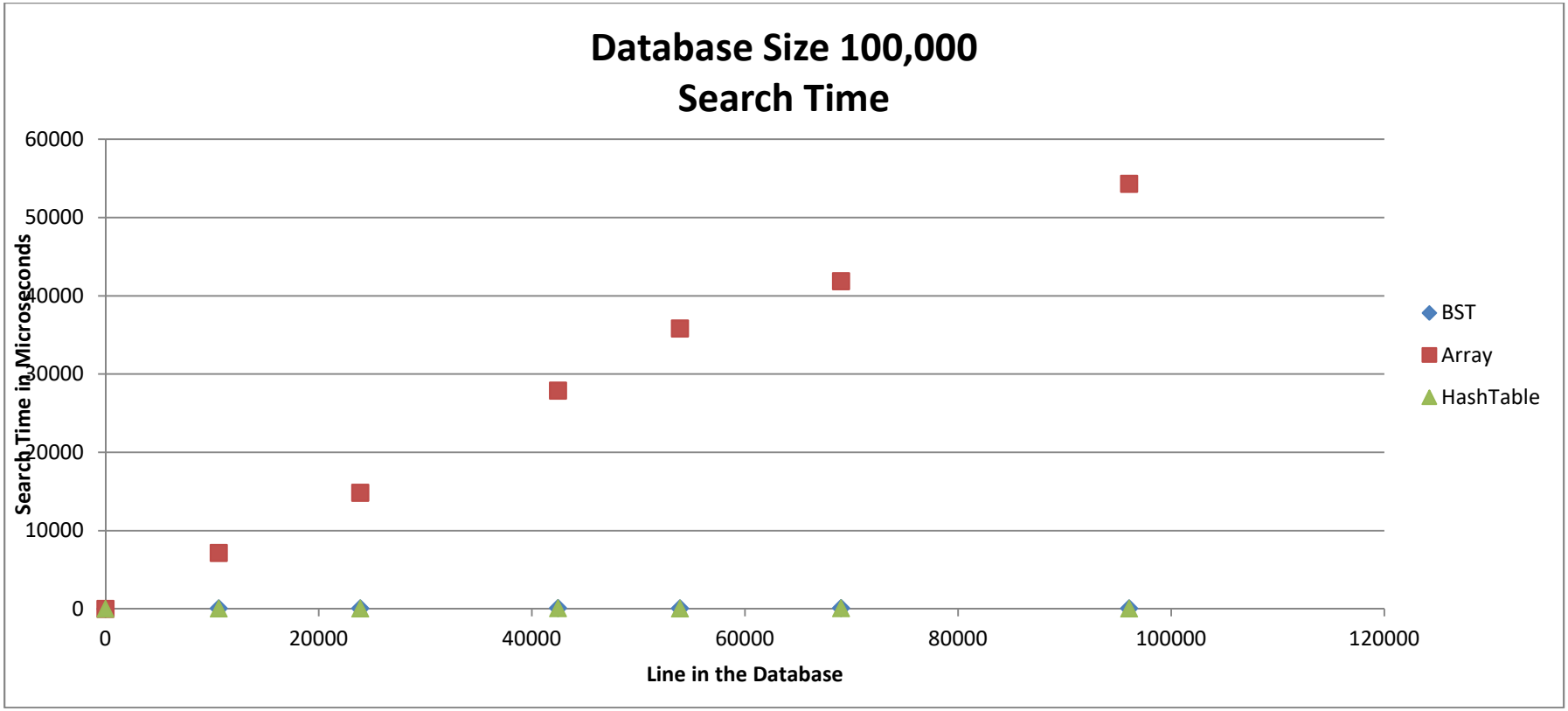
UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time
701197194311	5	7	6	7	6.666666667
799439078727	10,624	7134	7218	7089	7147
4605922001812	23,898	15281	15235	15130	15215.33333
886970578424	42,478	27349	28086	27890	27775


Database Size 50,000
Search Time
Search Time in Microseconds
30000
25000
20000
15000
10000
5000
0
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
Line in the Database
BST
Array
HashTable

Dawson Murphy CS300 Assignment 4

Data Size 100,00											
BST						HashTable					
UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time	UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time
701197194311	5	8	8	9	8.333333333	701197194311	5	7	7	7	7
799439078727	10,624	66	86	65	72.33333333	799439078727	10,624	28	29	28	28.33333333
4605922001812	23,898	81	81	81	81	4605922001812	23,898	51	51	51	51
886970578424	42,478	76	103	103	94	886970578424	42,478	55	54	54	54.33333333
662102111321	53,920	55	54	55	54.66666667	662102111321	53,920	36	35	38	36.33333333
35826072672	69,038	150	109	110	123	35826072672	69,038	60	54	53	55.66666667
352688190925	96,057	77	78	78	77.66666667	352688190925	96,057	50	50	49	49.66666667

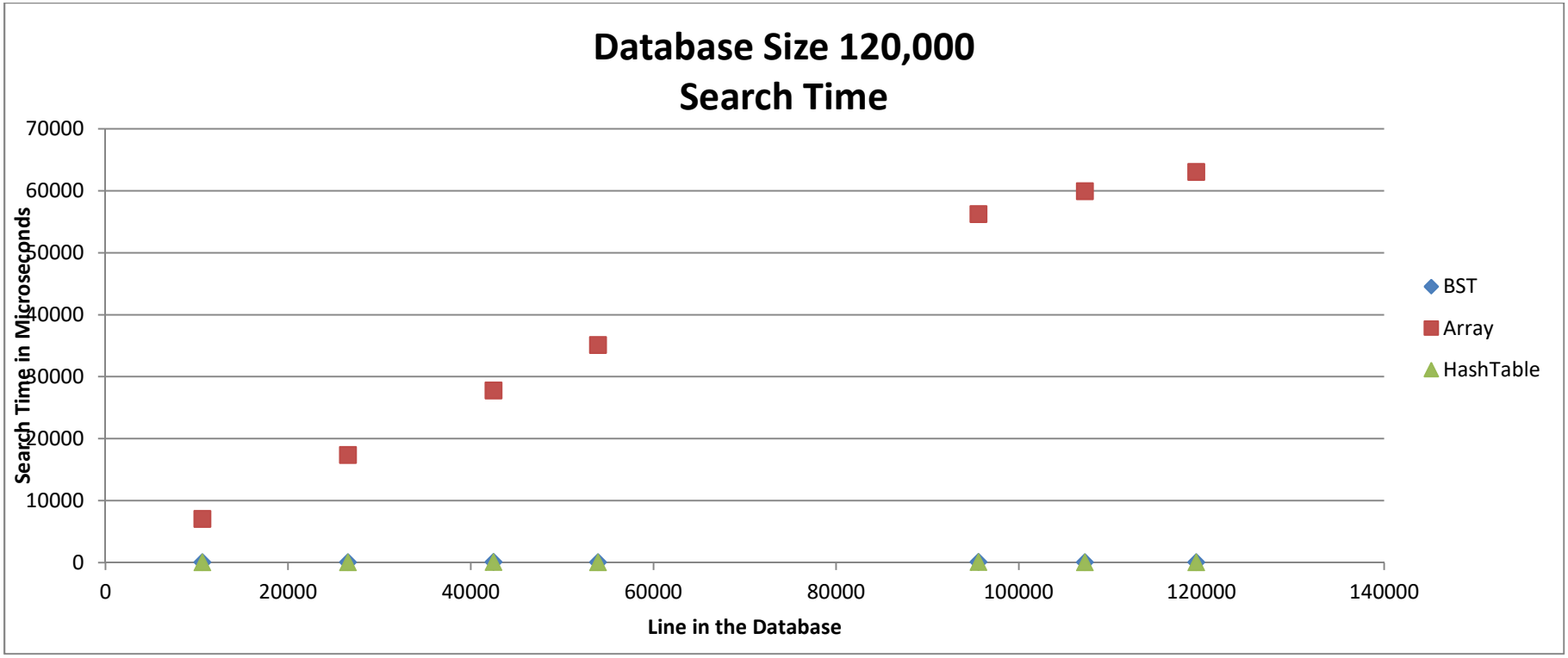
Array					
UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time
701197194311	5	7	7	8	7.333333333
799439078727	10,624	7153	7145	7103	7133.666667
4605922001812	23,898	15104	14314	15092	14836.66667
886970578424	42,478	28006	27979	27618	27867.66667
662102111321	53,920	36225	36516	34755	35832
35826072672	69,038	42317	40377	42887	41860.33333
352688190925	96,057	55448	53229	54230	54302.33333


Database Size 100,000
Search Time
Search Time in Microseconds
60000
50000
40000
30000
20000
10000
0
0
20000
40000
60000
80000
100000
120000
Line in the Database
BST
Array
HashTable

Dawson Murphy CS300 Assignment 4

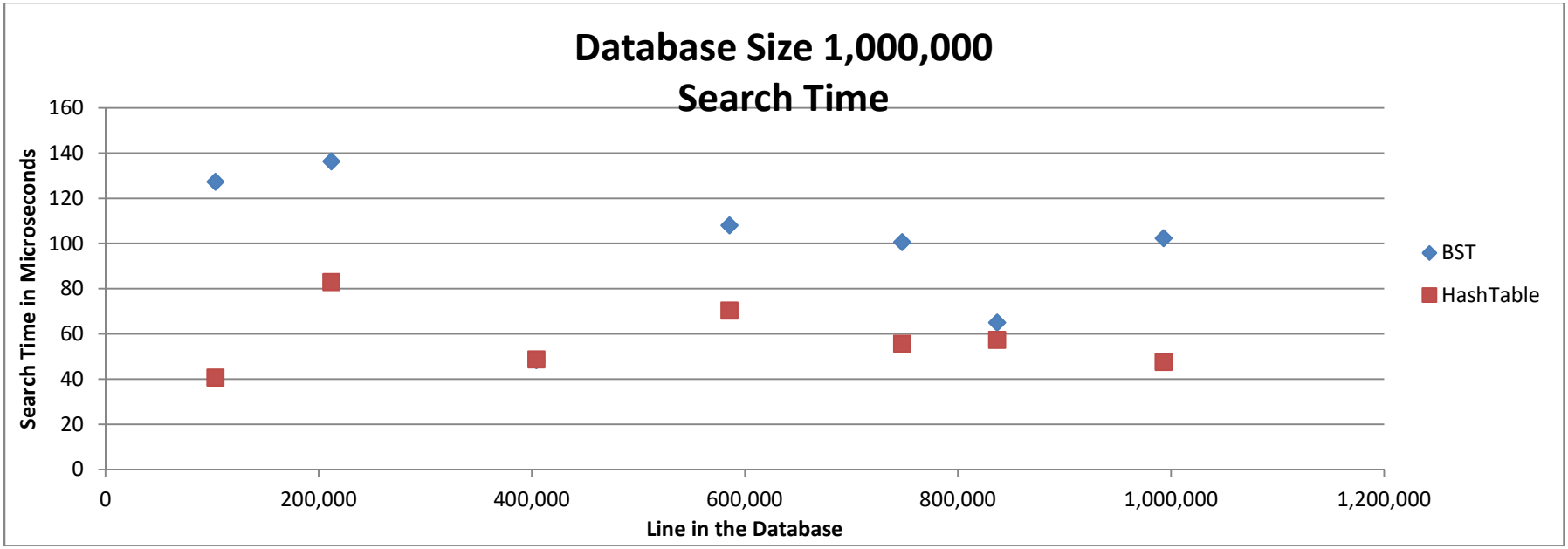
Data Size 120,000											
BST						HashTable					
UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time	UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time
799439078727	10,624	65	65	78	69.33333333	799439078727	10,624	28	29	29	28.66666667
751492439181	26,551	49	50	49	49.33333333	751492439181	26,551	33	34	33	33.33333333
886970578424	42,478	104	104	105	104.3333333	886970578424	42,478	56	54	55	55
662102111321	53,920	57	55	55	55.66666667	662102111321	53,920	36	35	36	35.66666667
603899051650	95,572	92	91	116	99.66666667	603899051650	95,572	57	58	57	57.33333333
7051057658877	107,224	95	95	65	85	7051057658877	107,224	47	47	46	46.66666667
79118004132	119,423	72	71	70	71	79118004132	119,423	29	30	29	29.33333333

Array					
UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time
799439078727	10,624	7028	7165	7020	7071
751492439181	26,551	17309	17500	17342	17383.66667
886970578424	42,478	27731	27635	27917	27761
662102111321	53,920	35015	36070	34254	35113
603899051650	95,572	57937	51934	58877	56249.33333
7051057658877	107,224	62740	55446	61530	59905.33333
79118004132	119,423	62440	64693	61982	63038.33333



Note: This is as large an array as could be made in the Virtual Machine on my computer.

Data Size 1,000,000					
BST					
UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time
7501254400781	103,204	143	133	106	127.3333333
32692189206	211,977	146	146	117	136.3333333
697675600385	404,421	48	49	48	48.33333333
42799703694	585,710	107	108	109	108
738329022525	747,476	99	102	101	100.6666667
634479066344	836,727	65	65	65	65
72512070512	992,908	96	114	97	102.3333333
HashTable					
UPC Code	Line # in Database	Test 1	Test 2	Test 3	Average Time
7501254400781	103,204	40	42	40	40.66666667
32692189206	211,977	84	82	83	83
697675600385	404,421	49	48	49	48.66666667
42799703694	585,710	62	88	61	70.33333333
738329022525	747,476	55	56	56	55.66666667
634479066344	836,727	58	56	58	57.33333333
72512070512	992,918	47	48	48	47.66666667



As shown during the tests, the data structures have different performances when searching for an item in a dataset. The Array is quick near the beginning of the data set, but much longer when looking for an object deeper in the dataset. The Binary Search Tree takes longer the further into the dataset it gets, but eventually gets to a point where the time it takes to find one item vs the time to find another much further in the dataset is negligible. As for HashTable, it seems to take about the same amount of time to find an item at the front as at the back. The main difference in search times is how many items are stored at a particular hash. Changing the size of the dataset does not seem to have a consequential effect on the search time of objects that existed in smaller data sets.