Performance Evaluation

1) Single Fault Handling

For all file, lets analyze the total time elapsed to complete a job and no of faults occured.

Following are 30 observations in tabular format.
For varying fail probability for a task: 10, 20, 30, 40, 50
And for varying chunk Size: 50,000 and 100000
Every entry indicate the time elapsed in millisecond and number of faults occurred

File 1: 200,000 and mergeNumber =10

	10	20	30	40	50
50,000	43632, 37	45446, 45	46808, 67	47299, 76	49773, 89
100,000	13660, 23	15732, 30	16209, 45	17652, 71	19248, 98

File 1: 2,000,000 and mergeNumber =10

	10	20	30	40	50
5,00,000	227432, 18	249845, 33	257834, 53	273245,79	301548, 271
1,00,000	91578, 32	105904, 55	124583, 69	139856, 88	158739, 134

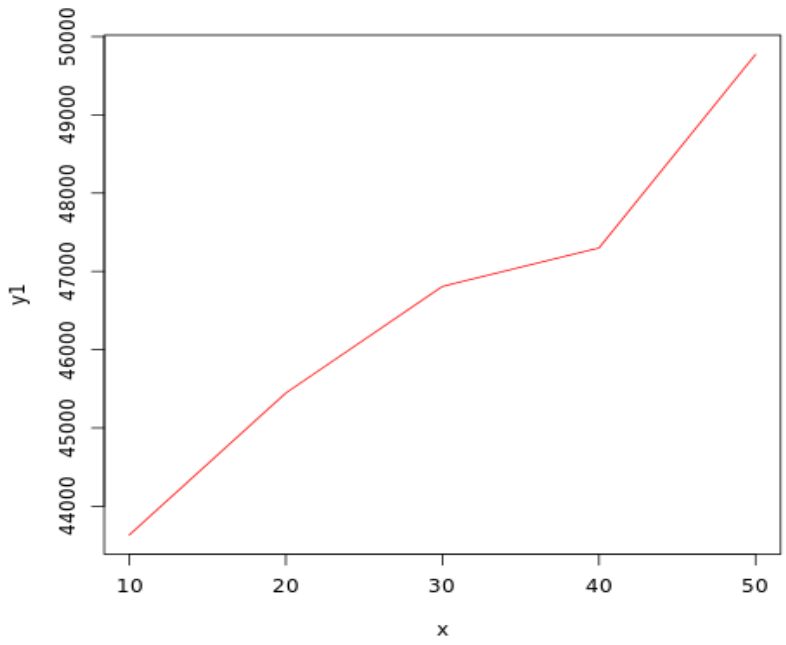
File 1: 20,000,000 and mergeNumber =10

	10	20	30	40	50
5,000,000	4067296, 94	4169034, 43	4251957, 78	4367983,110	4539476, 209
10,000,000	2294569, 59	2400891, 39	2498342, 21	2600034, 82	2812129, 114

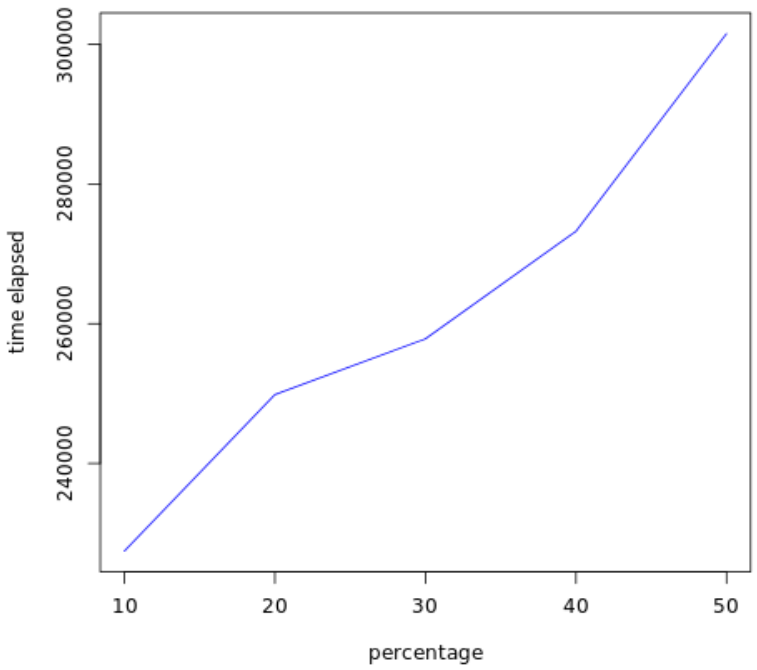
We observed that the general trend is as the chunk size increases, the time required to execute the task decreases. This can be very well supported by the fact that for small chunks, the number of sort and merge jobs will be greater. Hence, the time required will be also be greater.

We also observed that the general trend is as fail probability increases, the time required to execute the task increases. This can be very well supported by the fact that forlarge fail probabilities, the number of tasks failed will be greater. Hence, the time required will be also be greater as we go from left to right.

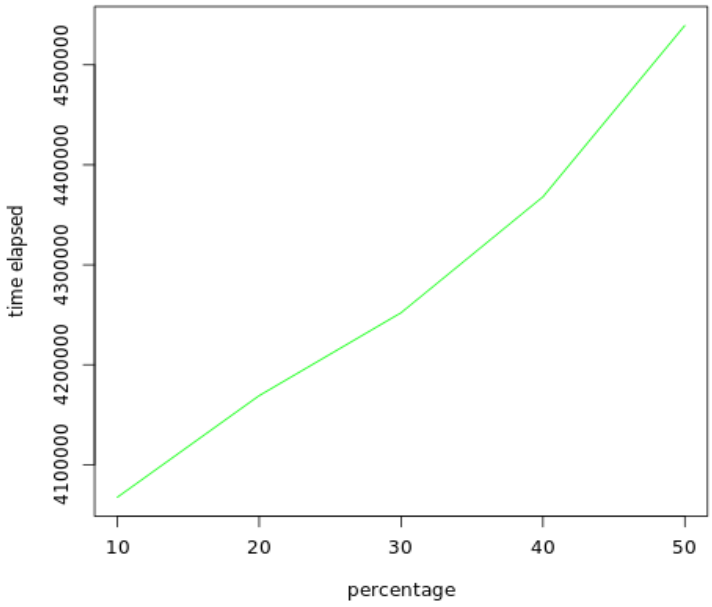
Following are the graphs for the first rows of all tables:
File 1:



File 2:



File 3:



2) Proactive Fault

For all file, lets analyze the total time elapsed to complete a job and no of faults occurred.
Following are 30 observations in tabular format.
For varying fail probability for a task: 10, 20, 30, 40, 50
And for varying chunk Size: 50,000 and 100000
And for x=3 where x is # of Randomly chosen node
Every entry indicate the time elapsed in millisecond and number of faults occurred

File 1: 200,000 and mergeNumber =10

	10	20	30	40	50
50,000	131765, 43	136563, 63	140459, 87	142343, 110	150967, 89
100,000	41211, 29	47331, 56	48989, 75	53129,	57983, 98

File 1: 2,000,000 and mergeNumber =10

	10	20	30	40	50
5,00,000	684329, 37	749872, 67	774371, 53	820568,162	912328, 400
1,00,000	274983, 62	318562, 55	374398, 69	420451, 181	478003, 273

File 1: 20,000,000 and mergeNumber =10

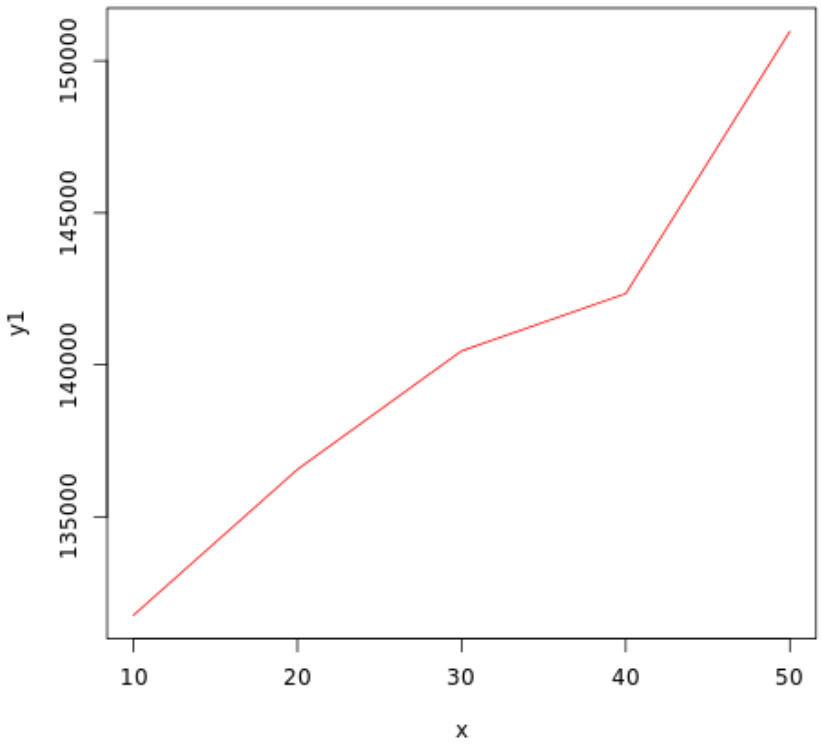
	10	20	30	40	50
5,000,000	12469690,193	12545823,87	12760926,165	13145298,236	13690321,392
10,000,000	6889743, 133	7212783, 86	7549627, 49	7816739, 178	84893239,250

We observed that the general trend is as the chunk size increases, the time required to execute the task decreases. This can be very well supported by the fact that for small chunks, the number of sort and merge jobs will be greater. Hence, the time required will be also be greater.

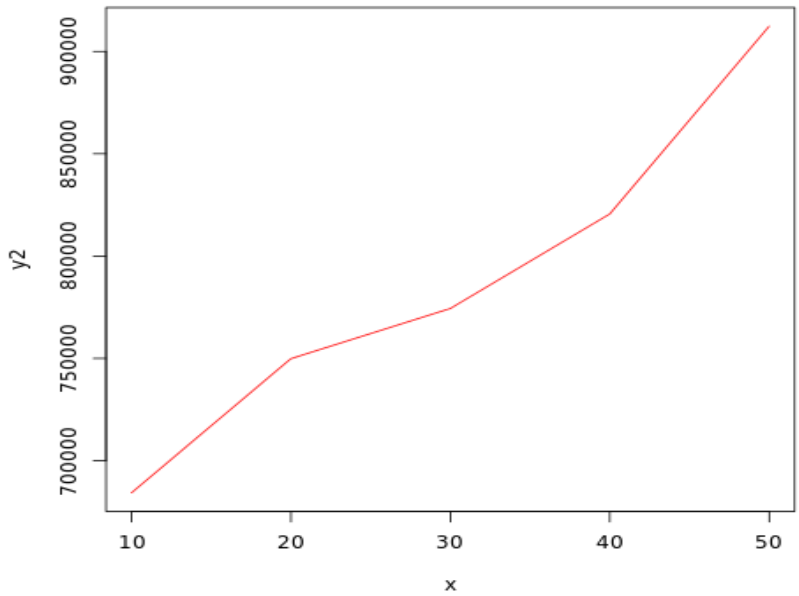
We also observed that the general trend is as fail probability increases, the time required to execute the task increases. This can be very well supported by the fact that forlarge fail probabilities, the number of tasks failed will be greater. Hence, the time required will be also be greater as we go from left to right.

We also observed that the time required has become almost 3 fold to the previous case

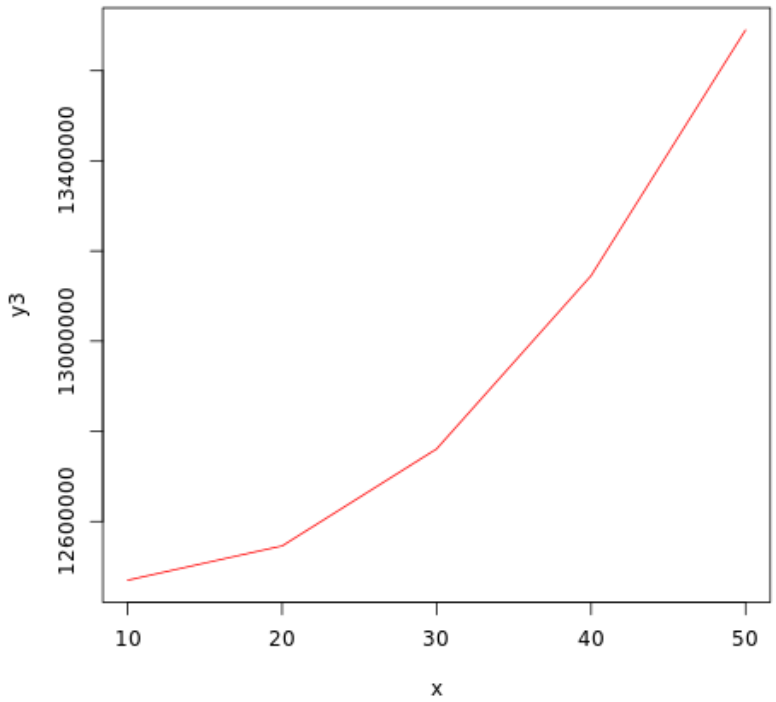
Following are the graphs for the first rows of all tables:
File 1



File 2


y2
700000
750000
800000
850000
900000
10
20
30
40
50
x

File 3:


y3
12600000
13000000
13400000
10
20
30
40
50
x