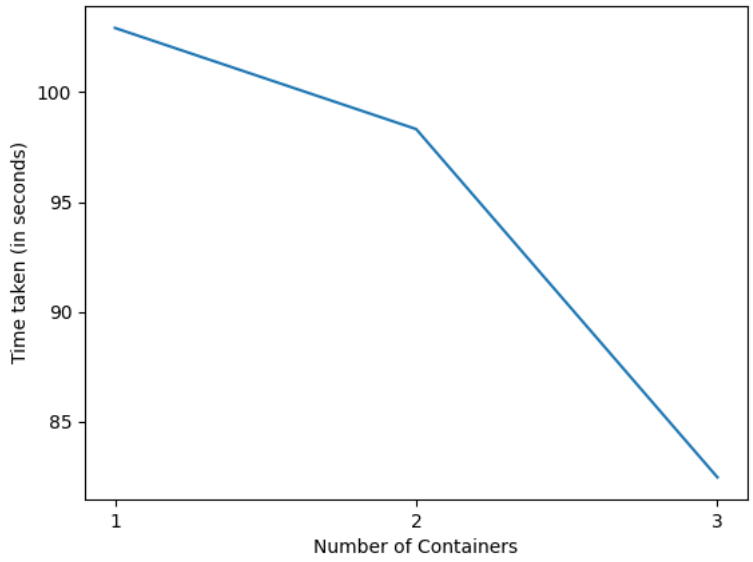
MPI Analysis

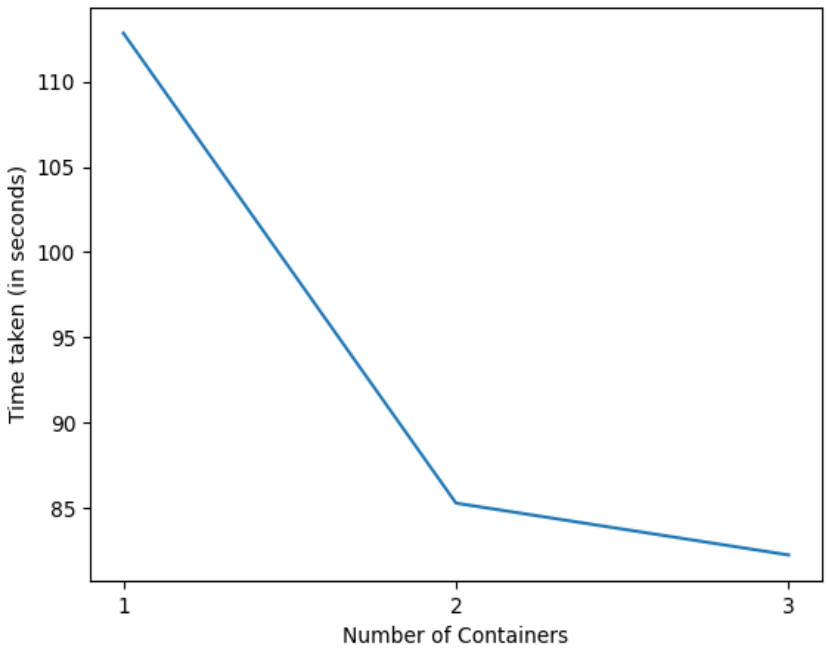
X-axis = number of worker containers

Y-axis = Time taken(in seconds)

Q1.

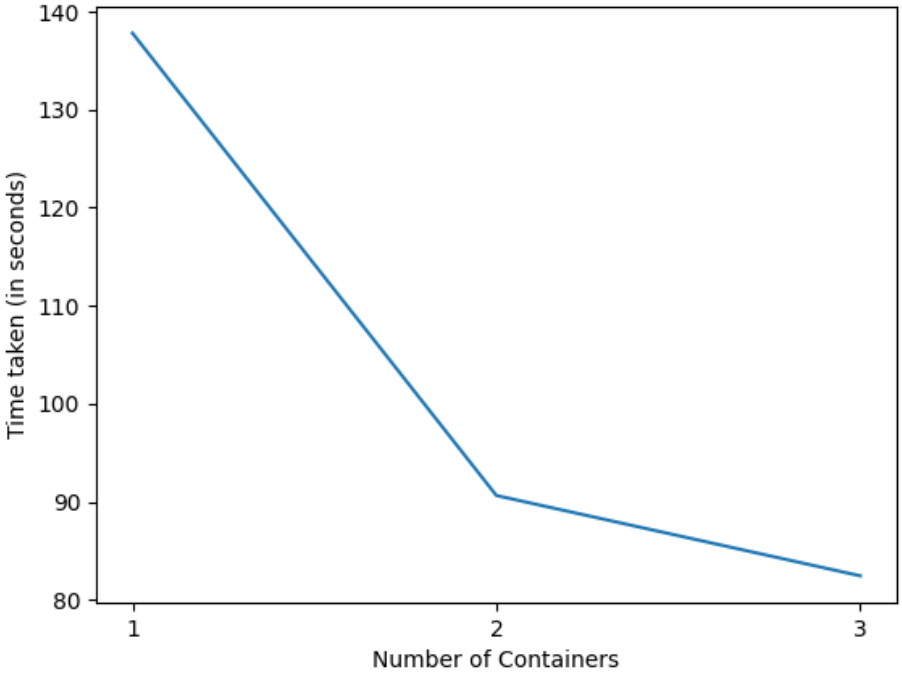


Q2.

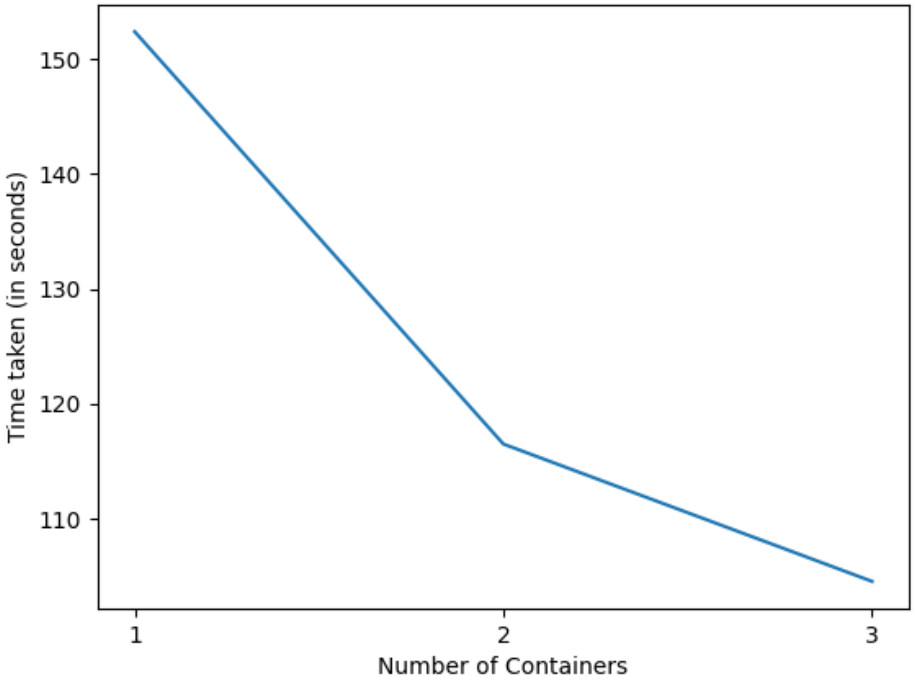


MPI Analysis

Q3.



Q4.



MPI Analysis

All the four graphs show a similar pattern. As we **increase the number of worker containers** from 1 to 3, there is a **significant decrease in time taken for completion of tasks**. This is because with increase in number of worker containers, smaller chunks of our dataset can be distributed by the master across workers, which eventually would take comparatively less time for computing, processing, and delivering the results. These independent chunks of data are aggregated back by master to deliver the final result of the task.

When there is only **one worker container**, all the 6.4 million records in the datasets are computed and processed by a single worker. Average **execution time** is around **120 seconds.**

When number of **worker containers are increased to 2**, the load is distributed, and each worker now manages 3.2 million records only. This results in an **average speed up (decrease in execution time) of about 22 seconds in comparison to a single worker container** which is a considerable increase in performance.

On further increasing the **number of worker containers to 3**, the computation load is even distributed further (~2.1 million) and again we see speed up in performance. Although not as large as the last one, **there is an average drop of 6 seconds in computing the output as compared to working with 2 worker containers.**