

**CSE488: Big Data Analytics**

**[SPRING 2023]**

**Assignment**

**Offline Assignment(Computing average and performance comparison)**

**Submitted by:**

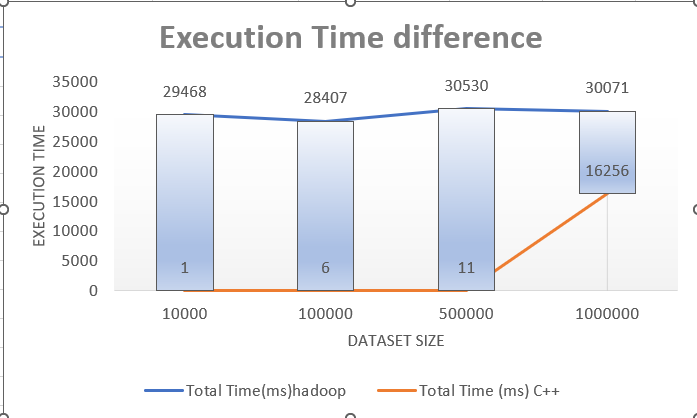
Student ID: Mujahidul Islam

Student Name: 2019-2-60-072

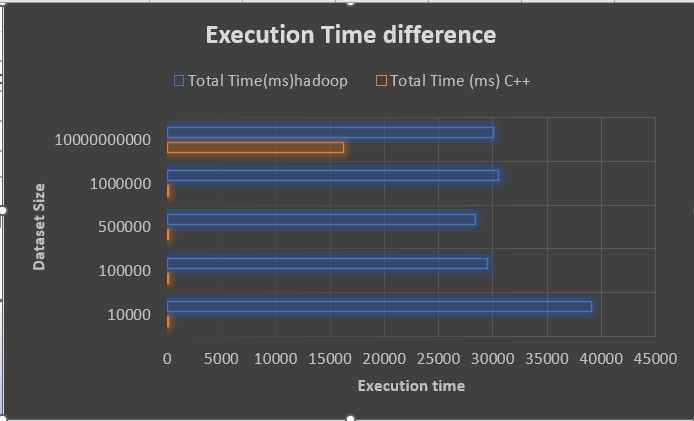
Comparison Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset Size** | **Hadoop MapReduce** | | | **C++** |
|  | **Mapper time(ms)** | **Reducer time(ms)** | **Total Time(ms)hadoop** | **Total Time (ms) C++** |
| **10000** | **26096** | **12955** | **39051** | **1** |
| **100000** | **12989** | **16479** | **29468** | **1** |
| **500000** | **12057** | **16350** | **28407** | **6** |
| **1000000** | **12834** | **17696** | **30530** | **11** |
| **10000000000** | **13196** | **16875** | **30071** | **16256** |

Line Plot:



Bar Plot:



Interpretation:

MapReduce is distributed programming platform which is built to handle larger amount of dataset. Whereas C++ is static and do not calculate in different node. So in smaller amount of data C++ is quite efficient to use but where the larger data comes Mapreduce takes lesser time.

If we look at the graphs above, We can see that for data size (10,000-10,00,000) C++ doesn’t take that much time. But in case of 10e10 data it takes a much larger amount time.

Whereas in MapReduce platform the data is 10000 or 10e10 , it takes more or less same amount of time.

So it be said that, for smaller datasets, C/C++ program would perform better than a MapReduce program but for a larger dataset MapReduce program would perform better