PART – 3

PART 3.1 – Copy NYSE dataset (DailyPrices_A to DailyPrices_Z) to a folder in HDFS. No merging.

Write a MapReduce to find the Max price of stock_price_high for each stock. Capture the running time programmatically (or manually using a wristwatch or smartphone).

The time taken for the code to run 1min:08sec for NYSE_daily_stocks_AtoZ(26files).

PART 3.2 – Redo 3.1. Merge the NYSE files in a single file (that you may have done in your prev. HW) on HDFS. Now, repeat 3.1 on the single merged-file. Capture the running time. Did MapReduce on a single file run faster than running MapReduce on a bunch of files?

Interestingly the time taken for merged single file of NYSE_daily_stocks_AtoZ took 1min:22sec to run the code.