

PART – 3

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

```

C:\Program Files\Git\cmd> Command Prompt
D:\hadoop\sbin>hdfs dfs -mkdir /stocks
D:\hadoop\sbin>hdfs dfs -copyFromLocal D:/stocks/*.csv /stocks
D:\hadoop\sbin>hdfs dfs -ls /stocks/
Found 26 items
-rw-r--r-- 1 91833 supergroup 40990992 2023-06-23 14:50 /stocks/NYSE_daily_prices_A.csv
-rw-r--r-- 1 91833 supergroup 32034760 2023-06-23 14:50 /stocks/NYSE_daily_prices_B.csv
-rw-r--r-- 1 91833 supergroup 45790256 2023-06-23 14:50 /stocks/NYSE_daily_prices_C.csv
-rw-r--r-- 1 91833 supergroup 19234471 2023-06-23 14:50 /stocks/NYSE_daily_prices_D.csv
-rw-r--r-- 1 91833 supergroup 22104043 2023-06-23 14:50 /stocks/NYSE_daily_prices_E.csv
-rw-r--r-- 1 91833 supergroup 17387253 2023-06-23 14:50 /stocks/NYSE_daily_prices_F.csv
-rw-r--r-- 1 91833 supergroup 22608522 2023-06-23 14:50 /stocks/NYSE_daily_prices_G.csv
-rw-r--r-- 1 91833 supergroup 23127143 2023-06-23 14:50 /stocks/NYSE_daily_prices_H.csv
-rw-r--r-- 1 91833 supergroup 20680033 2023-06-23 14:50 /stocks/NYSE_daily_prices_I.csv
-rw-r--r-- 1 91833 supergroup 9537527 2023-06-23 14:50 /stocks/NYSE_daily_prices_J.csv
-rw-r--r-- 1 91833 supergroup 14782892 2023-06-23 14:50 /stocks/NYSE_daily_prices_K.csv
-rw-r--r-- 1 91833 supergroup 12958785 2023-06-23 14:50 /stocks/NYSE_daily_prices_L.csv
-rw-r--r-- 1 91833 supergroup 38124545 2023-06-23 14:50 /stocks/NYSE_daily_prices_M.csv
-rw-r--r-- 1 91833 supergroup 31488945 2023-06-23 14:50 /stocks/NYSE_daily_prices_N.csv
-rw-r--r-- 1 91833 supergroup 8865718 2023-06-23 14:50 /stocks/NYSE_daily_prices_O.csv
-rw-r--r-- 1 91833 supergroup 31943478 2023-06-23 14:50 /stocks/NYSE_daily_prices_P.csv
-rw-r--r-- 1 91833 supergroup 190989 2023-06-23 14:50 /stocks/NYSE_daily_prices_Q.csv
-rw-r--r-- 1 91833 supergroup 16808595 2023-06-23 14:50 /stocks/NYSE_daily_prices_R.csv
-rw-r--r-- 1 91833 supergroup 31852353 2023-06-23 14:50 /stocks/NYSE_daily_prices_S.csv
-rw-r--r-- 1 91833 supergroup 28754690 2023-06-23 14:50 /stocks/NYSE_daily_prices_T.csv
-rw-r--r-- 1 91833 supergroup 9951590 2023-06-23 14:50 /stocks/NYSE_daily_prices_U.csv
-rw-r--r-- 1 91833 supergroup 9503196 2023-06-23 14:50 /stocks/NYSE_daily_prices_V.csv
-rw-r--r-- 1 91833 supergroup 15972013 2023-06-23 14:50 /stocks/NYSE_daily_prices_W.csv
-rw-r--r-- 1 91833 supergroup 3613198 2023-06-23 14:50 /stocks/NYSE_daily_prices_X.csv
-rw-r--r-- 1 91833 supergroup 686216 2023-06-23 14:50 /stocks/NYSE_daily_prices_Y.csv
-rw-r--r-- 1 91833 supergroup 2093424 2023-06-23 14:50 /stocks/NYSE_daily_prices_Z.csv
D:\hadoop\sbin>

```

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).

[illegible]

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?

[illegible]

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