**Map\_Reduce Solution:**

Language of choice: Python

Below is the approach taken for designing Mapper and Reducer to solve the problem statement:

1. All the 4 CSV files are taken as input to Hadoop File system
2. The Input files (line by line) is massed to one set of Mapper/Reducer which does the following
   1. Mapper\_1:
      1. Prints all the rows which has EQ as second element in the input text
      2. The output of this is the filtered list with Series as “EQ”
   2. Reducer\_1:
      1. Does nothing, just prints the input as output and a new output Part file is generated
3. The output of first Map/Reduce function is passed as input to second round of Map/reduce process:
   1. Mapper\_2:
      1. From the input line by line it takes only the Symbol(name of stock), Close & Date element and prints out to the console (passes to reducer)
   2. Reducer\_2:
      1. For all the lines as input it takes the Name of the Stock (SYMBOL) and (Date) as the keys of a dictionary
      2. The Values will be the number in Close column
      3. The interim process output will be dictionary of SYMBOL & Date along with number of close values for each stock for all unique stocks and date in the input
      4. Then for each key in dictionary it calculates the Min, Max & Mean of all the corresponding values
      5. Output will be <Stock Name> <Date><Min><Max><Mean>
4. The output of this can be passed through another Map/Reduce which calculates the Standard Deviation and Pearson correlation.

**Hive Solution:**

Code with comments and the output CSV file which calculates Mean, Min, Max and STD\_Dev is attached in folder.