Module	BIG DATA and HADOOP
Instructor	Dr. Vinay Kulkarni
Assignment Start Date	14-05-2019
Assignment Submission Date	31-05-2019
Total Marks	100

Problems to be solved using the Hadoop Ecosystem VM:

1. Create **mapper.py** and **reducer.py** programs and execute them using Hadoop Streaming to accomplish the following with the file ‘all.csv’:
   a. Extract all records related to the GEOMETRIC for any given month (eg. APRIL-2011) and find out the ‘count’ of the records (ie. the number of days in the month on which GEOMETRIC was traded)
   b. Extract all records related to WIPRO and calculate the descriptive statistics: mean, median, variance and standard deviation of the ‘closing price’
   c. Print out the dates on which the equity closing prices of Larsen and Toubro (LT) are within Rs 10 of each other.
   - In each of the above cases, submit the following:
     - The python programs
     - Outputs generated
     - Marks: [20 x 3 = 60]

2. Create the programs - **mapper.py** and **reducer.py** to create the counts of all the ‘words’ present in the three files that you have already copied into the HDFS ‘input’ directory. Ensure that words get properly split before word-count. This can be done by specifying the following additional “delimiters” : comma, full stop, quotes, inverted commas, dash, all mathematical symbols, other symbols like | % $ #, etc. Submit the programs as well as the output generated. **[15 Marks]**

3. Go through the HDFS Shell documentation available at hadoop.apache.org and try out 10 HDFS commands (other than those used in class – ls, mkdir, rm, copyFromLocal, copyToLocal). Explain the commands you have tried out and provide screenshots as proof of having executed the commands. Prepare a document (eg. MS Word) outlining the steps you have taken and the screenshots. **[10 Marks]**

4. From the internet find out **three** cases / articles which highlight case where Hadoop has been successfully implemented. Summarize each case in not more than 15 lines and provide a link (URL) to the case. **[5 x 3 = 15 Marks]**