

## CprE 419 Lab 7:

### Pig User Defined Functions for analyzing stock data

Shuo Wang

**Exp1:** Find the largest growth factor for each stock and report the top 5 stock with largest growth rate.

For finding the largest growth factor for each stock:

**Idea:** The largest growth factor is the highest the price divided by the lowest price in the analyzing period, subject to: the lowest price has to happen before the largest price

**Note:** The price of the last available day over the price of the first available day may not give the largest growth factor. For example, part of the records for ticker "HOT" is like this:

HOT	20000103	23.38
HOT	19971024	55.21
HOT	19920527	0.42
HOT	19900102	5.02

the largest growth factor is  $55.21/0.42 = 131.45$  rather than  $23.38/5.02 = 4.66$ . "Hot" with growth factor 131.45 is the top 1 ticker during period Jan 1st, 1990 – Jan 3rd, 2000.

**Algorithm:** sort the records by time for a certain ticker; scan from the first day to the last day while tracking the minimum price seen so far; calculate the growth rate by current price over minimum price; update the growth rate by the largest growth rate so far while scanning; report the growth rate after the scan.

For the detailed algorithm please refer to the Java source code in the submission file.

**Output:** the output for experiment 1 is shown below

Period Jan 1st, 1990 – Jan 3rd, 2000.

```
[shuowang@n0 ~]$ hcat /scr/shuowang/lab7/exp1/output1/part-r-00000
HOT      131.45238095238096
BIIB     65.58215962441315
QCOM     57.861635220125784
JDSU     47.03448275862069
EMC      39.75739644970414
```

Period Jan 2nd, 2005 – Jan 31st, 2014.

```
[shuowang@n0 ~]$ hcat /scr/shuowang/lab7/exp1/output2/part-r-00000
AIG      167.21621621621622
PCLN     65.05061927840603
C        53.950980392156865
THC      53.83516483516483
NFLX     44.73626373626374
```

**Exp2:** Find the moving average for six companies.

For the detailed algorithm please refer to the Java source code in the submission file.

Part of the output for experiment 2 is shown below

```
[shuowang@n0 ~]$ hcat /scr/shuowang/lab7/exp2/output/part-r-00000
GE      20131001      23.94      23.93
GE      20131002      24.14      23.9615
GE      20131003      24.22      24.019
GE      20131004      24.18      24.063
GE      20131007      23.84      24.1145
GE      20131008      23.92      24.144
GE      20131009      23.84      24.1565
GE      20131010      23.9      24.151
GE      20131011      24.32      24.138
```

For the entire output for experiment 2 please refer to the text file in the submission file.

## Appendix: files in submission .zip

1. exp1.pig and exp2.pig are pig file for running in n0
2. GROWTH.java is the Java source code of UDF for exp1.pig and GROWTH.jar is the compiled file
3. MOVAVG.java is the Java source code of UDF for exp2.pig and MOVAVG.jar is the compiled file
4. exp1\_output.txt and exp2\_output.txt are the output results for experiment 1 and 2, respectively