Samuel Wolfe

March 12, 2024

MSBA 214-70

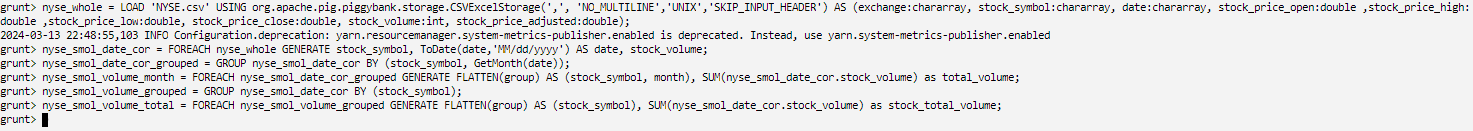
Midterm Part 2 Question 2



nyse\_whole = LOAD 'NYSE.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER') AS (exchange:chararray, stock\_symbol:chararray, date:chararray, stock\_price\_open:double ,stock\_price\_high:double ,stock\_price\_low:double, stock\_price\_close:double, stock\_volume:int, stock\_price\_adjusted:double);

nyse\_smol\_date\_cor = FOREACH nyse\_whole GENERATE stock\_symbol, ToDate(date,'MM/dd/yyyy') AS date, stock\_volume;

1. The total volume traded for each stock for each month
   1. nyse\_smol\_date\_cor\_grouped = GROUP nyse\_smol\_date\_cor BY (stock\_symbol, GetMonth(date));
   2. nyse\_smol\_volume\_month = FOREACH nyse\_smol\_date\_cor\_grouped GENERATE FLATTEN(group) AS (stock\_symbol, month), SUM(nyse\_smol\_date\_cor.stock\_volume) as total\_volume;
2. The total volume traded for each stock.
   1. nyse\_smol\_volume\_grouped = GROUP nyse\_smol\_date\_cor BY (stock\_symbol);
   2. nyse\_smol\_volume\_total = FOREACH nyse\_smol\_volume\_grouped GENERATE FLATTEN(group) AS (stock\_symbol), SUM(nyse\_smol\_date\_cor.stock\_volume) as stock\_total\_volume;



For question 1, I took it literally as just group them by stock and then by month and output the stocks total sum for that month. It’s to long to get even multiple screen shots as the data runs off the CLI. Here is a screenshot of the last few stocks.

A screenshot of a computer code

Description automatically generated

For the second question I also took it literally and grouped by stock and then grabbed the total. The pictures below do show GJE twice, but that is due to the scroll amount for my system, it only displays once.

A screenshot of a computer

Description automatically generated A screenshot of a computer code

Description automatically generated A number and numbers on a white background

Description automatically generated