Data: NYSE.CSV

- Write a series of Pig Latin commands to identify:

1. The total volume traded for each stock for each month.

```
grunt> data = LOAD 'Sudi/NYSE.csv' USING PigStorage(',') AS (
>>     exchange: chararray,
>>     stock_symbol: chararray,
>>     date: chararray,
>>     stock_price_open: float,
>>     stock_price_high: float,
>>     stock_price_low: float,
>>     stock_price_close: float,
>>     stock_volume: int,
>>     stock_price_adj_close: float
>> );
2024-03-14 00:35:07,187 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.ena
bled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
grunt>
```

```
grunt> data_with_month = FOREACH data GENERATE
grunt>     stock_symbol,
>>     GetMonth(ToDate(date, 'MM/dd/yyyy')) AS month,
>>     stock_volume;
```

```
grunt> NYSE_Data = LOAD 'Sudi/NYSE.csv' USING PigStorage(',') AS (
>>     exchange: chararray,
>>     stock_symbol: chararray,
>>     date: chararray,
>>     stock_price_open: float,
>>     stock_price_high: float,
>>     stock_price_low: float,
>>     stock_price_close: float,
>>     stock_volume: int,
>>     stock_price_adj_close: float
>> );
2024-03-13 23:56:42,383 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.ena
bled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
grunt>
grunt> NYSE_with_date = FOREACH NYSE_Data GENERATE
grunt>     exchange,
>>     stock_symbol,
>>     ToDate(date, 'MM/dd/yyyy') AS date,
>>     stock_price_open,
>>     stock_price_high,
>>     stock_price_low,
>>     stock_price_close,
>>     stock_volume,
>>     stock_price_adj_close;
grunt>
grunt>
```

Extract the month from the datetime value.

```
grunt> data_with_month = FOREACH NYSE_with_date GENERATE
grunt>     exchange,
>>     stock_symbol,
>>     GetMonth(date) AS month,
>>     stock_volume;
grunt>
```

```
grunt> volume_per_month = FOREACH (GROUP data_with_month BY (stock_symbol, month)) GENERATE
grunt>     FLATTEN(group) AS (stock_symbol, month),
>>     SUM(data_with_month.stock_volume) AS total_volume;
grunt>
```

2. The total volume traded for each stock.

```
grunt> volume_per_stock = FOREACH (GROUP NYSE_with_date BY stock_symbol) GENERATE
grunt>     group AS stock_symbol,
>>     SUM(NYSE_with_date.stock_volume) AS total_volume;
```

```
grunt> -- Calculate total volume traded for each stock
grunt> volume_per_stock = FOREACH (GROUP data BY stock_symbol) GENERATE
grunt>     group AS stock_symbol,
>>     SUM(data.stock_volume) AS total_volume;
grunt>
```

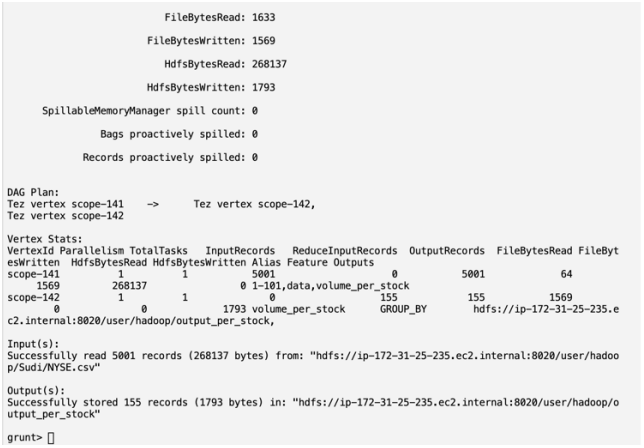
3. Store the results in HDFS and display them.

```
STORE volume_per_month INTO 'output_per_month' USING PigStorage(',');

DUMP volume_per_month;
```

```
STORE volume_per_stock INTO 'output_per_stock' USING PigStorage(',');

DUMP volume_per_stock;
```



```
                        FileBytesRead: 1633

                     FileBytesWritten: 1569

                        HdfsBytesRead: 268137

                     HdfsBytesWritten: 1793

      SpillableMemoryManager spill count: 0

                Bags proactively spilled: 0

             Records proactively spilled: 0


DAG Plan:
Tez vertex scope-141    ->      Tez vertex scope-142,
Tez vertex scope-142

Vertex Stats:
VertexId Parallelism TotalTasks   InputRecords   ReduceInputRecords  OutputRecords  FileBytesRead FileByt
esWritten  HdfsBytesRead HdfsBytesWritten Alias Feature Outputs
scope-141          1          1           5001                   0           5001             64
      1569         268137               0 1-101,data,volume_per_stock
scope-142          1          1              0                 155            155           1569
         0              0          1793 volume_per_stock      GROUP_BY        hdfs://ip-172-31-25-235.e
c2.internal:8020/user/hadoop/output_per_stock,

Input(s):
Successfully read 5001 records (268137 bytes) from: "hdfs://ip-172-31-25-235.ec2.internal:8020/user/hadoo
p/Sudi/NYSE.csv"

Output(s):
Successfully stored 155 records (1793 bytes) in: "hdfs://ip-172-31-25-235.ec2.internal:8020/user/hadoop/o
utput_per_stock"

grunt>
```

```
2024-03-14 01:12:21,438 INFO util.MapRedUtil: Total input paths to process : 1
(GA,9531100)
(GB,7681400)
(GD,146828800)
(GE,2141242700)
(GF,6057900)
(GG,198803800)
(GR,52768300)
(GS,276912700)
(GT,123810700)
(GU,5211500)
(GY,14105700)
(GAB,15470700)
(GAH,425600)
(GAI,675700)
(GAJ,275000)
(GAM,2037400)
(GAP,14445200)
(GAR,7100)
(GAS,16713800)
(GAT,116700)
(GBB,13700)
(GBE,2833700)
(GBF,75400)
(GBL,1701400)
(GBX,4668200)
(GCA,3401700)
(GCF,191300)
(GCH,5151000)
(GCI,62263600)
(GCO,20883600)
(GCS,1587000)
(GCV,748100)
(GDF,1747500)
(GDI,14871600)
(GDL,602100)
(GDO,8000)
(GDP,8523900)
```

need to use ToDate operator to change the time format from String to DateTime