

PIG

1.1 Objective:

To find out the top 10 stocks that gained most value in New York Stock Exchange between August 2009 and August 2010.

1.2 Dataset:

Source: <http://kumo.swcp.com/stocks/>

Data Format: csv, Number of records: 122,574, File size 5,202 KB

Sample data:

Date	StockSymbol	DayOpen	DayHigh	DayLow	DayClose	Volume
20090821	A	25.6	25.61	25.22	25.55	34758
20090824	A	25.64	25.74	25.33	25.5	22247
20090825	A	25.5	25.7	25.225	25.34	30891
20090826	A	25.32	25.6425	25.145	25.48	33334
20090827	A	25.5	25.57	25.23	25.54	70176
20090828	A	25.67	26.05	25.63	25.83	39694
20090831	A	25.45	25.74	25.31	25.68	51064
20090901	A	25.51	26.33	25.48	25.85	66422
20090902	A	25.97	25.97	24.96	25.22	64614
20090903	A	25.47	25.54	25	25.29	46369

1.3 Approach:

- MapReduce method of Pig is required as the dataset is large
- Import Daily Stocks file into HDFS
- Load the file into PIG
- Group records by stock symbol, calculate max & minimum values across the dataset for each stock symbol. Also calculate the difference between max and minimum values
- Sort the output on price difference in descending order. Limit the output to 10 records.
- Write the sort output to a file.

1.4 PIG Command:

PIG Command

```
DailyStocks = LOAD '/input/sp500hst.txt' using PigStorage(',') AS ( date:chararray, stocksymbol:chararray,
dayopen:double, dayhigh:double, daylow:double, dayclose:double, volume:long);
```

```

DailyPrices_grouped = GROUP DailyStocks by stocksymbol;
DailyPrices_highlows = FOREACH DailyPrices_grouped GENERATE group, MAX(DailyStocks.dayhigh),
MIN(DailyStocks.daylow), MAX(DailyStocks.dayhigh) - MIN(DailyStocks.daylow);
DailyPrices_highlows_sorted = ORDER DailyPrices_highlows BY $3 DESC;
DailyPrices_highlows_top10 = LIMIT DailyPrices_highlows_sorted 10;
STORE DailyPrices_highlows_top10 INTO '/output/pigStockTopTen';
STORE DailyPrices_highlows INTO '/output/pigStock';

```

PIG Execution

```

sushil@singlegen1:~$ pig /home/sushil/datasets/PIG_POC.pig
16/06/03 14:25:36 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
16/06/03 14:25:36 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
16/06/03 14:25:36 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2016-06-03 14:25:36,443 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01
2015, 11:43:55
2016-06-03 14:25:36,443 [main] INFO org.apache.pig.Main - Logging error messages to:
/home/sushil/pig_1464944136441.log
2016-06-03 14:25:36,747 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/sushil/.pigbootup
not found
2016-06-03 14:25:36,882 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine -
Connecting to hadoop file system at: hdfs://localhost/
2016-06-03 14:25:37,023 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine -
Connecting to map-reduce job tracker at: localhost:8021
2016-06-03 14:25:38,395 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script:
GROUP_BY,ORDER_BY,LIMIT
2016-06-03 14:25:38,434 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not
set... will not generate code.
2016-06-03 14:25:38,505 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer -
{RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter,
LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer,
PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2016-06-03 14:25:38,722 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100
optimistic? false
2016-06-03 14:25:38,769 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil
- Choosing to move algebraic foreach to combiner
2016-06-03 14:25:38,810 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary
Key Optimization for MapReduce node scope-62
2016-06-03 14:25:38,831 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before
optimization: 6
2016-06-03 14:25:38,834 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - Merged 2 map-only
splittees.
2016-06-03 14:25:38,834 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - Merged 2 out of total 3
MR operators.

```

```
2016-06-03 14:25:38,834 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after
optimization: 4
2016-06-03 14:25:38,922 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings
are added to the job
2016-06-03 14:25:38,931 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2016-06-03 14:25:38,933 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected,
estimating # of required reducers.
2016-06-03 14:25:38,934 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator:
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2016-06-03 14:25:38,939 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator -
BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=5203962
2016-06-03 14:25:38,939 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2016-06-03 14:25:38,939 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be
converted run in-process
2016-06-03 14:25:39,100 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/pig-0.15.0-core-h1.jar to DistributedCache through /tmp/temp-
1115249976/tmp1384142684/pig-0.15.0-core-h1.jar
2016-06-03 14:25:39,130 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-
1115249976/tmp1424613870/automaton-1.11-8.jar
2016-06-03 14:25:39,155 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-
1115249976/tmp201002403/antlr-runtime-3.4.jar
2016-06-03 14:25:39,204 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/guava-11.0.jar to DistributedCache through /tmp/temp-1115249976/tmp-
1135655096/guava-11.0.jar
2016-06-03 14:25:39,224 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/joda-time-2.5.jar to DistributedCache through /tmp/temp-
1115249976/tmp1269448269/joda-time-2.5.jar
2016-06-03 14:25:39,261 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up multi store job
2016-06-03 14:25:39,273 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false,
will not generate code.
2016-06-03 14:25:39,273 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move
generated code to distributed cacche
2016-06-03 14:25:39,273 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []
```

2016-06-03 14:25:39,389 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

2016-06-03 14:25:39,430 [JobControl] WARN org.apache.hadoop.mapred.JobClient - No job jar file set. User classes may not be found. See JobConf(Class) or JobConf#setJar(String).

2016-06-03 14:25:39,502 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2016-06-03 14:25:39,502 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2016-06-03 14:25:39,511 [JobControl] INFO org.apache.hadoop.util.NativeCodeLoader - Loaded the native-hadoop library

2016-06-03 14:25:39,511 [JobControl] WARN org.apache.hadoop.io.compress.snappy.LoadSnappy - Snappy native library not loaded

2016-06-03 14:25:39,514 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2016-06-03 14:25:39,892 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_201606031312_0019

2016-06-03 14:25:39,893 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases DailyPrices_grouped,DailyPrices_highlows,DailyStocks

2016-06-03 14:25:39,893 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: DailyStocks[1,14],DailyStocks[-1,-1],DailyPrices_highlows[7,23],DailyPrices_grouped[5,22] C: DailyPrices_highlows[7,23],DailyPrices_grouped[5,22] R: DailyPrices_highlows[7,23]

2016-06-03 14:25:39,893 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_201606031312_0019

2016-06-03 14:25:39,911 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete

2016-06-03 14:25:39,912 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_201606031312_0019]

2016-06-03 14:25:48,571 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12% complete

2016-06-03 14:25:48,574 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_201606031312_0019]

2016-06-03 14:25:56,646 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 16% complete

2016-06-03 14:25:56,646 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_201606031312_0019]

2016-06-03 14:25:59,166 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 25% complete

2016-06-03 14:25:59,167 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_201606031312_0019]

2016-06-03 14:26:05,378 [main] INFO org.apache.pig.tools.pigstats.JobStats - using output size reader: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.FileBasedOutputSizeReader

2016-06-03 14:26:05,392 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

```
2016-06-03 14:26:05,393 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2016-06-03 14:26:05,395 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected,
estimating # of required reducers.
2016-06-03 14:26:05,395 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator:
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2016-06-03 14:26:05,405 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator -
BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=19404
2016-06-03 14:26:05,407 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2016-06-03 14:26:05,407 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be
converted run in-process
2016-06-03 14:26:05,618 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/pig-0.15.0-core-h1.jar to DistributedCache through /tmp/temp-
1115249976/tmp1079108953/pig-0.15.0-core-h1.jar
2016-06-03 14:26:05,731 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-1115249976/tmp-
1038541678/automaton-1.11-8.jar
2016-06-03 14:26:05,842 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-
1115249976/tmp1303711394/antlr-runtime-3.4.jar
2016-06-03 14:26:05,985 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/guava-11.0.jar to DistributedCache through /tmp/temp-
1115249976/tmp2025267628/guava-11.0.jar
2016-06-03 14:26:06,101 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/joda-time-2.5.jar to DistributedCache through /tmp/temp-
1115249976/tmp1631214300/joda-time-2.5.jar
2016-06-03 14:26:06,141 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2016-06-03 14:26:06,145 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false,
will not generate code.
2016-06-03 14:26:06,150 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move
generated code to distributed cacche
2016-06-03 14:26:06,151 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []
2016-06-03 14:26:06,356 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s)
waiting for submission.
2016-06-03 14:26:06,564 [JobControl] WARN org.apache.hadoop.mapred.JobClient - No job jar file set. User classes
may not be found. See JobConf(Class) or JobConf#setJar(String).
```

2016-06-03 14:26:06,622 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2016-06-03 14:26:06,622 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

2016-06-03 14:26:06,630 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2016-06-03 14:26:07,850 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_201606031312_0020

2016-06-03 14:26:07,850 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases DailyPrices_highlows_sorted

2016-06-03 14:26:07,850 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: DailyPrices_highlows_sorted[9,30] C: R:

2016-06-03 14:26:07,851 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_201606031312_0020

2016-06-03 14:26:14,700 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 37% complete

2016-06-03 14:26:14,700 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_201606031312_0020]

2016-06-03 14:26:22,272 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 41% complete

2016-06-03 14:26:22,274 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_201606031312_0020]

2016-06-03 14:26:24,288 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete

2016-06-03 14:26:24,288 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_201606031312_0020]

2016-06-03 14:26:27,848 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

2016-06-03 14:26:27,850 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2016-06-03 14:26:27,851 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.

2016-06-03 14:26:27,851 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1

2016-06-03 14:26:27,851 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process

2016-06-03 14:26:27,962 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/sushil/pig-0.15.0/pig-0.15.0-core-h1.jar to DistributedCache through /tmp/temp-1115249976/tmp-698568264/pig-0.15.0-core-h1.jar


```
2016-06-03 14:26:28,053 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-
1115249976/tmp1000629308/automaton-1.11-8.jar

2016-06-03 14:26:28,123 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-1115249976/tmp-
1808410393/antlr-runtime-3.4.jar

2016-06-03 14:26:28,194 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/guava-11.0.jar to DistributedCache through /tmp/temp-1115249976/tmp-
1953651948/guava-11.0.jar

2016-06-03 14:26:28,233 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/joda-time-2.5.jar to DistributedCache through /tmp/temp-
1115249976/tmp34141648/joda-time-2.5.jar

2016-06-03 14:26:28,247 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2016-06-03 14:26:28,248 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false,
will not generate code.

2016-06-03 14:26:28,249 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move
generated code to distributed cacche

2016-06-03 14:26:28,249 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []

2016-06-03 14:26:28,307 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s)
waiting for submission.

2016-06-03 14:26:28,385 [JobControl] WARN org.apache.hadoop.mapred.JobClient - No job jar file set. User classes
may not be found. See JobConf(Class) or JobConf#setJar(String).

2016-06-03 14:26:28,422 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input
paths to process : 1

2016-06-03 14:26:28,422 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil -
Total input paths to process : 1

2016-06-03 14:26:28,425 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil -
Total input paths (combined) to process : 1

2016-06-03 14:26:28,809 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId:
job_201606031312_0021

2016-06-03 14:26:28,810 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases
DailyPrices_highlows_sorted

2016-06-03 14:26:28,810 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M:
DailyPrices_highlows_sorted[9,30] C: R:

2016-06-03 14:26:28,811 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at:
http://localhost:50030/jobdetails.jsp?jobid=job_201606031312_0021

2016-06-03 14:26:34,354 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 62% complete
```

```
2016-06-03 14:26:34,354 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are
[job_201606031312_0021]
2016-06-03 14:26:42,593 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 66% complete
2016-06-03 14:26:42,593 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are
[job_201606031312_0021]
2016-06-03 14:26:44,105 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 75% complete
2016-06-03 14:26:44,105 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are
[job_201606031312_0021]
2016-06-03 14:26:48,682 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings
are added to the job
2016-06-03 14:26:48,684 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2016-06-03 14:26:48,686 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected,
estimating # of required reducers.
2016-06-03 14:26:48,687 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2016-06-03 14:26:48,687 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be
converted run in-process
2016-06-03 14:26:48,844 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/pig-0.15.0-core-h1.jar to DistributedCache through /tmp/temp-1115249976/tmp-
1963761679/pig-0.15.0-core-h1.jar
2016-06-03 14:26:48,869 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-
1115249976/tmp1185263039/automaton-1.11-8.jar
2016-06-03 14:26:48,901 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-1115249976/tmp-
413921285/antlr-runtime-3.4.jar
2016-06-03 14:26:48,938 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/guava-11.0.jar to DistributedCache through /tmp/temp-1115249976/tmp-
1607590308/guava-11.0.jar
2016-06-03 14:26:48,964 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar
file:/home/sushil/pig-0.15.0/lib/joda-time-2.5.jar to DistributedCache through /tmp/temp-
1115249976/tmp1615526304/joda-time-2.5.jar
2016-06-03 14:26:48,974 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2016-06-03 14:26:48,976 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false,
will not generate code.
```



```

2016-06-03 14:26:48,976 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move
generated code to distributed cacche
2016-06-03 14:26:48,977 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []
2016-06-03 14:26:49,025 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s)
waiting for submission.
2016-06-03 14:26:49,064 [JobControl] WARN org.apache.hadoop.mapred.JobClient - No job jar file set. User classes
may not be found. See JobConf(Class) or JobConf#setJar(String).
2016-06-03 14:26:49,081 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input
paths to process : 1
2016-06-03 14:26:49,081 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil -
Total input paths to process : 1
2016-06-03 14:26:49,083 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil -
Total input paths (combined) to process : 1
2016-06-03 14:26:49,526 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId:
job_201606031312_0022
2016-06-03 14:26:49,527 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases
DailyPrices_highlows_sorted
2016-06-03 14:26:49,527 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M:
DailyPrices_highlows_sorted[9,30] C: R:
2016-06-03 14:26:49,528 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at:
http://localhost:50030/jobdetails.jsp?jobid=job_201606031312_0022
2016-06-03 14:26:54,574 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 87% complete
2016-06-03 14:26:54,574 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are
[job_201606031312_0022]
2016-06-03 14:27:02,233 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 91% complete
2016-06-03 14:27:02,233 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are
[job_201606031312_0022]
2016-06-03 14:27:04,248 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are
[job_201606031312_0022]
2016-06-03 14:27:09,318 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2016-06-03 14:27:09,367 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion PigVersion  UserId StartedAt   FinishedAt   Features
1.2.1 0.15.0 sushil 2016-06-03 14:25:38 2016-06-03 14:27:09 GROUP_BY,ORDER_BY,LIMIT

```

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature	Outputs
job_201606031312_0019	1	1	5	5	5	5	10	10	10	10			
DailyPrices_grouped,DailyPrices_highlows,DailyStocks GROUP_BY,MULTI_QUERY,COMBINER /output/pigStock,													
job_201606031312_0020	1	1	2	2	2	2	10	10	10	10			DailyPrices_highlows_sorted
SAMPLER													
job_201606031312_0021	1	1	2	2	2	2	9	9	9	9			DailyPrices_highlows_sorted
ORDER_BY,COMBINER													
job_201606031312_0022	1	1	2	2	2	2	9	9	9	9			DailyPrices_highlows_sorted
/output/pigStockTopTen,													

Input(s):

Successfully read 122574 records (5204324 bytes) from: "/input/sp500hst.txt"

Output(s):

Successfully stored 524 records (15404 bytes) in: "/output/pigStock"

Successfully stored 10 records (307 bytes) in: "/output/pigStockTopTen"

Counters:

Total records written : 534

Total bytes written : 15711

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

```

job_201606031312_0019 -> job_201606031312_0020,
job_201606031312_0020 -> job_201606031312_0021,
job_201606031312_0021 -> job_201606031312_0022,
job_201606031312_0022

```

2016-06-03 14:27:09,432 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

2016-06-03 14:27:09,473 [main] INFO org.apache.pig.Main - Pig script completed in 1 minute, 33 seconds and 120 milliseconds (93120 ms)