Hadoop Stock Exchange Case study

Mapreduce Hands-on

1] Write a map reduce program to split the NYSE_daily.txt into 4 files based on Ticker

hadoop jar nyse.jar casestudy1.SplitData /user/training/stock/input/NYSE_daily.txt /case1

hadoop dfs –get /case1 output/

```
bitnami@ip-172-31-1-112:~$ hadoop jar nyse.jar casestudy1.SplitData /user/training/stock/input/NYSE_daily.txt /case1
16/07/16 13:17:41 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/07/16 13:17:42 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with T
oolRunner to remedy this.
16/07/16 13:17:42 INFO input.FileInputFormat: Total input paths to process : 1
16/07/16 13:17:42 INFO mapreduce.JobSubmitter: number of splits:1
16/07/16 13:17:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1468674155193_0001
16/07/16 13:17:43 INFO impl.YarnClientImpl: Submitted application application_1468674155193_0001
16/07/16 13:17:43 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1468674155193_0001/
16/07/16 13:17:43 INFO mapreduce.Job: Running job: job_1468674155193_0001
16/07/16 13:17:49 INFO mapreduce.Job: Job job_1468674155193_0001 running in uber mode : false
16/07/16 13:17:49 INFO mapreduce.Job:  map 0% reduce 0%
16/07/16 13:17:54 INFO mapreduce.Job:  map 100% reduce 0%
16/07/16 13:18:01 INFO mapreduce.Job:  map 100% reduce 25%
16/07/16 13:18:04 INFO mapreduce.Job:  map 100% reduce 50%
16/07/16 13:18:05 INFO mapreduce.Job:  map 100% reduce 100%
16/07/16 13:18:06 INFO mapreduce.Job: Job job_1468674155193_0001 completed successfully
16/07/16 13:18:06 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=3021950
                FILE: Number of bytes written=6634947
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=3194226
                HDFS: Number of bytes written=2907144
                HDFS: Number of read operations=15
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=8
        Job Counters
                Killed reduce tasks=1
                Launched map tasks=1
                Launched reduce tasks=4
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=2445
                Total time spent by all reduces in occupied slots (ms)=24253
                Total time spent by all map tasks (ms)=2445
                Total time spent by all reduce tasks (ms)=24253
                Total vcore-milliseconds taken by all map tasks=2445
                Total vcore-milliseconds taken by all reduce tasks=24253
                Total megabyte-milliseconds taken by all map tasks=2503680
                Total megabyte-milliseconds taken by all reduce tasks=24835072
```

2] Write a Map Reduce Program to find the maximum Variation for each stock ticker

hadoop jar nyse.jar casestudy2.MaxVariation /case1/p /case2*

hadoop dfs –get /case2 output/

```
bitnami@ip-172-31-1-112:~$
bitnami@ip-172-31-1-112:~$ hadoop jar nyse.jar casestudy2.MaxVariation /case1/p* /case2
16/07/16 13:32:58 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/07/16 13:32:58 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with T
oolRunner to remedy this.
16/07/16 13:32:59 INFO input.FileInputFormat: Total input paths to process : 4
16/07/16 13:32:59 INFO mapreduce.JobSubmitter: number of splits:4
16/07/16 13:32:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1468674155193_0002
16/07/16 13:32:59 INFO impl.YarnClientImpl: Submitted application application_1468674155193_0002
16/07/16 13:32:59 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1468674155193_0002/
16/07/16 13:32:59 INFO mapreduce.Job: Running job: job_1468674155193_0002
16/07/16 13:33:05 INFO mapreduce.Job: Job job_1468674155193_0002 running in uber mode : false
16/07/16 13:33:05 INFO mapreduce.Job:  map 0% reduce 0%
16/07/16 13:33:16 INFO mapreduce.Job:  map 100% reduce 0%
16/07/16 13:33:21 INFO mapreduce.Job:  map 100% reduce 100%
16/07/16 13:33:22 INFO mapreduce.Job: Job job_1468674155193_0002 completed successfully
16/07/16 13:33:22 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=1630270
                FILE: Number of bytes written=3851624
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2907564
                HDFS: Number of bytes written=5879
                HDFS: Number of read operations=15
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=4
                Launched reduce tasks=1
                Data-local map tasks=4
                Total time spent by all maps in occupied slots (ms)=36174
                Total time spent by all reduces in occupied slots (ms)=2359
                Total time spent by all map tasks (ms)=36174
                Total time spent by all reduce tasks (ms)=2359
                Total vcore-milliseconds taken by all map tasks=36174
                Total vcore-milliseconds taken by all reduce tasks=2359
                Total megabyte-milliseconds taken by all map tasks=37042176
                Total megabyte-milliseconds taken by all reduce tasks=2415616
        Map-Reduce Framework
                Map input records=57391
                Map output records=57391
```