

PART - 6

Redo Part5 of this assignment, but cram multiple values (max stock_volume, min stock_volume, max stock_price_adj_close) into a Text object with some delimiter. Use a Combiner. Compare the running time of Part 2 to Part 3. You could measure the running time programmatically, or use your smartphone's timer.

Implemented the combiner class in addition to the part-5 code. Java file has been provided in the folder.

Part -2 (Access.log file)

The time taken to run code with combiner: **18.61sec**

The time taken to run code without combiner: **20.17sec**

Part -3 (NYSE_AtoZ dataset)

The time taken for the code to run for individual files: **1min:08sec**

The time taken for the code to run for merged file: **1min:22sec**

Part – 6 (NYSE_AtoZ dataset)

The time taken by the code to run for individual files with combiner and multiple fields of mapreduce :**1min:23sec**

```
Administrator: Command Prompt
D:\hadoop>bin\hadoop jar D:\NYSEDataAnalysis\1.0-SNAPSHOT.jar com.mycompany.nysedataanalysis.NYSEDataAnalysis -input /stocks/*.csv -output /output
2023-06-27 14:45:35,108 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-06-27 14:45:35,383 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy thi
s.
2023-06-27 14:45:35,395 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/91833/.staging/job_1687890659913_0001
2023-06-27 14:45:35,572 INFO InputFileInputFormat: Total input files to process : 26
2023-06-27 14:45:35,675 INFO mapreduce.JobSubmitter: number of splits:26
2023-06-27 14:45:35,840 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1687890659913_0001
2023-06-27 14:45:35,840 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-06-27 14:45:35,984 INFO conf.Configuration: resource-types.xml not found
2023-06-27 14:45:35,985 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-06-27 14:45:36,389 INFO impl.VarClientImpl: Submitted application application_1687890659913_0001
2023-06-27 14:45:36,430 INFO mapreduce.Job: The url to track the job: http://Manohar-Laptop:8088/proxy/application_1687890659913_0001/
2023-06-27 14:45:36,433 INFO mapreduce.Job: Running Job: job_1687890659913_0001
2023-06-27 14:45:43,574 INFO mapreduce.Job: Job job_1687890659913_0001 running in uber mode : false
2023-06-27 14:45:43,575 INFO mapreduce.Job: map 0% reduce 0%
2023-06-27 14:46:00,348 INFO mapreduce.Job: map 12% reduce 0%
2023-06-27 14:46:01,361 INFO mapreduce.Job: map 23% reduce 0%
2023-06-27 14:46:15,714 INFO mapreduce.Job: map 46% reduce 0%
2023-06-27 14:46:29,639 INFO mapreduce.Job: map 58% reduce 0%
2023-06-27 14:46:30,658 INFO mapreduce.Job: map 62% reduce 0%
2023-06-27 14:46:34,879 INFO mapreduce.Job: map 65% reduce 0%
2023-06-27 14:46:38,980 INFO mapreduce.Job: map 65% reduce 22%
2023-06-27 14:46:42,100 INFO mapreduce.Job: map 72% reduce 22%
2023-06-27 14:46:45,262 INFO mapreduce.Job: map 77% reduce 26%
2023-06-27 14:46:46,302 INFO mapreduce.Job: map 81% reduce 26%
2023-06-27 14:46:49,418 INFO mapreduce.Job: map 83% reduce 26%
2023-06-27 14:46:52,549 INFO mapreduce.Job: map 96% reduce 28%
2023-06-27 14:46:54,591 INFO mapreduce.Job: map 100% reduce 28%
2023-06-27 14:46:55,602 INFO mapreduce.Job: map 100% reduce 100%
2023-06-27 14:46:55,613 INFO mapreduce.Job: Job job_1687890659913_0001 completed successfully
2023-06-27 14:46:55,684 INFO mapreduce.Job: Counters: 55
File System Counters
  FILE: Number of bytes read=108234
  FILE: Number of bytes written=6674142
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
HDFS: Number of bytes read=511088609
HDFS: Number of bytes written=98767
HDFS: Number of read operations=83
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
  Killed map tasks=1
  Launched map tasks=26
  Launched reduce tasks=1
  Data-local map tasks=26
  Total time spent by all maps in occupied slots (ms)=345143
  Total time spent by all reduce tasks in occupied slots (ms)=38834
  Total time spent by all map tasks (ms)=345143
  Total time spent by all reduce tasks (ms)=38834
  Total vcore-milliseconds taken by all map tasks=345143
  Total vcore-milliseconds taken by all reduce tasks=38834
  Total megabyte-milliseconds taken by all map tasks=353426432
  Total megabyte-milliseconds taken by all reduce tasks=39766016
Map-Reduce Framework
  Map input records=9211037
  Map output records=9211031
  Map output bytes=330816029
  Map output materialized bytes=108384
  Input split bytes=3042
  Combine input records=9211031
  Combine output records=2853
  Reduce input groups=2853
  Reduce shuffle bytes=108384
  Reduce input records=2853
  Reduce output records=2853
  Spilled Records=5706
  Shuffled Maps =26
  Failed Shuffles=0
  Merged Map outputs=26
  GC time elapsed (ms)=849
  CPU time spent (ms)=109447
  Physical memory (bytes) snapshot=9666576384
  Virtual memory (bytes) snapshot=11808860800
  Total committed heap usage (bytes)=7191142408
  Peak Map Physical memory (bytes)=38222336
  Peak Map Virtual memory (bytes)=448204800
  Peak Reduce Physical memory (bytes)=289616896
  Peak Reduce Virtual memory (bytes)=438481408
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDT=0
File Input Format Counters
  Bytes Read=511085627
```