

Week 3

I ran the MaxTemperature script on 3 data sets. Following are the outputs and execution time:

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: ~/hadoop-book/ch02-mr-intro/src/main/java — s
Map input records=5000000
Map output records=4642067
Map output bytes=41778603
Map output materialized bytes=51062785
Input split bytes=904
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=51062785
Reduce input records=4642067
Reduce output records=1
Spilled Records=9284134
Shuffled Maps =8
Failed Shuffles=0
Merged Map outputs=8
GC time elapsed (ms)=3167
CPU time spent (ms)=21880
Physical memory (bytes) snapshot=1924378624
Virtual memory (bytes) snapshot=16911880192
Total committed heap usage (bytes)=1391427584
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=1030902727
File Output Format Counters
  Bytes Written=9

real    1m4.189s
user    0m3.444s
sys     0m0.120s
vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$ Hadoop fs -ls /user/$USER/output1
Hadoop: command not found
vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -ls /user/$USER/output1
-ls: Unknown command
[vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -ls /user/$USER/output1
Found 2 items
-rw-r--r--  1 vagrant supergroup          0 2017-02-10 20:14 /user/vagrant/output1/_SUCCESS
-rw-r--r--  1 vagrant supergroup          9 2017-02-10 20:14 /user/vagrant/output1/part-r-00000
[vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output1/part-r-00000
1990    607
vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$
```

```

    Total time spent by all map tasks (ms)=1816038
    Total time spent by all reduce tasks (ms)=318858
    Total vcore-seconds taken by all map tasks=1816038
    Total vcore-seconds taken by all reduce tasks=318858
    Total megabyte-seconds taken by all map tasks=1859622912
    Total megabyte-seconds taken by all reduce tasks=326510592
Map-Reduce Framework
    Map input records=35000000
    Map output records=30592765
    Map output bytes=275334885
    Map output materialized bytes=336520775
    Input split bytes=6840
    Combine input records=0
    Combine output records=0
    Reduce input groups=2
    Reduce shuffle bytes=336520775
    Reduce input records=30592765
    Reduce output records=2
    Spilled Records=61185530
    Shuffled Maps =60
    Failed Shuffles=0
    Merged Map outputs=60
    GC time elapsed (ms)=28745
    CPU time spent (ms)=176790
    Physical memory (bytes) snapshot=13301149696
    Virtual memory (bytes) snapshot=114581987328
    Total committed heap usage (bytes)=9769713664
Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
File Input Format Counters
    Bytes Read=7993006187
File Output Format Counters
    Bytes Written=18

real    6m40.488s
user    0m4.712s
sys     0m0.288s
[vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output3/par
1990    607
1992    605
vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$ █

```

```

liushuxiao — sliu102@euca-192-168-166-33: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh sliu102@64.131.111.40 -i ./sliu...
Map input records=163661993
Map output records=148281528
Map output bytes=1334533752
Map output materialized bytes=1631098422
Input split bytes=39005
Combine input records=0
Combine output records=0
Reduce input groups=4
Reduce shuffle bytes=1631098422
Reduce input records=148281528
Reduce output records=4
Spilled Records=296563056
Shuffled Maps =269
Failed Shuffles=0
Merged Map outputs=269
GC time elapsed (ms)=80451
CPU time spent (ms)=1468400
Physical memory (bytes) snapshot=122682884096
Virtual memory (bytes) snapshot=1054556901376
Total committed heap usage (bytes)=141138329600

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=36100200483
File Output Format Counters
Bytes Written=36

real    9m39.384s
user    0m8.640s
sys     0m0.804s
sliu102@euca-192-168-166-33:~/hadoop-book/ch02-mr-intro/src/main/java$

```

with combiner:

```

liushuxiao — sliu102@euca-192-168-166-33: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh sliu102@64.131.111.40 -i ./sliu...
Map input records=39300822
Map output records=36931296
Map output bytes=332381664
Map output materialized bytes=1071
Input split bytes=8820
Combine input records=36931296
Combine output records=63
Reduce input groups=1
Reduce shuffle bytes=1071
Reduce input records=63
Reduce output records=1
Spilled Records=126
Shuffled Maps =63
Failed Shuffles=0
Merged Map outputs=63
GC time elapsed (ms)=17636
CPU time spent (ms)=330200
Physical memory (bytes) snapshot=26740273152
Virtual memory (bytes) snapshot=249981194240
Total committed heap usage (bytes)=29354885120

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=8427702060
File Output Format Counters
Bytes Written=9

real    4m56.648s
user    0m8.012s
sys     0m0.688s
sliu102@euca-192-168-166-33:~/hadoop-book/ch02-mr-intro/src/main/java$

```

```
liushuxiao — sliu102@euca-192-168-166-33: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh sliu102@64.131.111.40 -i ./sliu...

Map-Reduce Framework
  Map input records=80991273
  Map output records=74324416
  Map output bytes=668919744
  Map output materialized bytes=2255
  Input split bytes=19800
  Combine input records=74324416
  Combine output records=133
  Reduce input groups=2
  Reduce shuffle bytes=2255
  Reduce input records=133
  Reduce output records=2
  Spilled Records=266
  Shuffled Maps =132
  Failed Shuffles=0
  Merged Map outputs=132
  GC time elapsed (ms)=36987
  CPU time spent (ms)=658380
  Physical memory (bytes) snapshot=54992732160
  Virtual memory (bytes) snapshot=519407804416
  Total committed heap usage (bytes)=63140528128

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=17699950687
File Output Format Counters
  Bytes Written=18

real    6m49.294s
user    0m8.356s
sys     0m0.828s
```

```
liushuxiao — sliu102@euca-192-168-166-33: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh sliu102@64.131.111.40 -i ./sliu...

Map input records=163661993
Map output records=148281528
Map output bytes=1334533752
Map output materialized bytes=4606
Input split bytes=39005
Combine input records=148281528
Combine output records=272
Reduce input groups=4
Reduce shuffle bytes=4606
Reduce input records=272
Reduce output records=4
Spilled Records=544
Shuffled Maps =269
Failed Shuffles=0
Merged Map outputs=269
GC time elapsed (ms)=76506
CPU time spent (ms)=1149420
Physical memory (bytes) snapshot=124306468864
Virtual memory (bytes) snapshot=1054557708288
Total committed heap usage (bytes)=145273389056

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=36100200483
File Output Format Counters
  Bytes Written=36

real    5m24.838s
user    0m8.096s
sys     0m0.588s
sliu102@euca-192-168-166-33:~/hadoop-book/ch02-mr-intro/src/main/java$
```