

Week 3

Part 1: I ran the MaxTemperature script on 3 data sets. Following are the outputs and execution time:

1990:

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh * vagrant ssh — 164x46
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/$USER/output1
Found 2 items
-rw-r--r-- 1 vagrant supergroup      0 2017-02-10 20:14 /user/vagrant/output1/_SUCCESS
-rw-r--r-- 1 vagrant supergroup      9 2017-02-10 20:14 /user/vagrant/output1/part-r-000000
vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output1/part-r-000000
1990    607
vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$
```

1990&1992:

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh • vagrant ssh — 164x46
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/$USER/output3/part-r-00000
1990    607
1992    605
vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$
```

1990-1993:

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: ~/hadoop-book/ch02-mr-intro/src/main/java — ssh • vagrant ssh — 164x46
Total time spent by all reduce tasks (ms)=541533
Total vcore-seconds taken by all map tasks=3114199
Total vcore-seconds taken by all reduce tasks=541533
Total megabyte-seconds taken by all map tasks=3188939776
Total megabyte-seconds taken by all reduce tasks=554529792
Map-Reduce Framework
  Map input records=50000000
  Map output records=42671468
  Map output bytes=384043212
  Map output materialized bytes=469386676
  Input split bytes=10032
  Combine input records=0
  Combine output records=0
  Reduce input groups=3
  Reduce shuffle bytes=469386676
  Reduce input records=42671468
  Reduce output records=3
  Spilled Records=85342936
  Shuffled Maps =88
  Failed Shuffles=0
  Merged Map outputs=88
  GC time elapsed (ms)=51784
  CPU time spent (ms)=285500
  Physical memory (bytes) snapshot=19448356864
  Virtual memory (bytes) snapshot=167180423168
  Total committed heap usage (bytes)=14234288128
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=11753148425
File Output Format Counters
  Bytes Written=27

real    11m6.822s
user    0m5.820s
sys     0m0.440s
[vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$ hadoop fs -cat /user/$USER/output4/part-r-00000
1990    607
1991    607
1992    605
vagrant@pleasechangeme:~/hadoop-book/ch02-mr-intro/src/main/java$
```

Part 2: I ran the MaxTemperatureWithCombiner script on 3 data sets. Following are the outputs and execution time:

1990:

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: ~/hadoop-book1/ch02-i
Failed Shuffles=0
Merged Map outputs=8
GC time elapsed (ms)=3884
CPU time spent (ms)=21680
Physical memory (bytes) snapshot=1848012800
Virtual memory (bytes) snapshot=16928657408
Total committed heap usage (bytes)=1338572800
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    1m13.994s
user    0m3.672s
sys     0m0.144s
[vagrant@pleasechangeme:~/hadoop-book1/ch02-mr-intro/src/main/java$ time hadoop
ar mt.jar MaxTemperatureWithCombiner /user/$USER/tempdata/9092 /user/$USER/tem
```

1990&1992:

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: ~/hadoop-book1/ch02-mr...

Shuffled Maps =60
Failed Shuffles=0
Merged Map outputs=60
GC time elapsed (ms)=32622
CPU time spent (ms)=177460
Physical memory (bytes) snapshot=13206110208
Virtual memory (bytes) snapshot=114708758528
Total committed heap usage (bytes)=9629925376

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    7m24.710s
user    0m5.292s
sys     0m0.468s
vagrant@pleasechangeme:~/hadoop-book1/ch02-mr-intro/src/main/java$
```

1990-1993:

```
ubuntu-xenial-large-521 — vagrant@pleasechangeme: ~/hadoop-book1/ch02-mr...

Shuffled Maps =63
Failed Shuffles=0
Merged Map outputs=63
GC time elapsed (ms)=25033
CPU time spent (ms)=153170
Physical memory (bytes) snapshot=13826244608
Virtual memory (bytes) snapshot=120335798272
Total committed heap usage (bytes)=10108272640

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=8295719911
File Output Format Counters
  Bytes Written=27

real    6m25.279s
user    0m4.652s
sys     0m0.220s
vagrant@pleasechangeme:~/hadoop-book1/ch02-mr-intro/src/main/java$
```