

CS 595 - Assignment 3

Name: Jagruti Vichare
Email ID: jvichare@hawk.iit.edu
CWID: A20378092

Question 6

The output of the program WordCount2.py is highlighted below in red

```
maria_dev@sandbox:~  
FILE: Number of large read operations=0  
FILE: Number of read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=1029  
HDFS: Number of bytes written=23  
HDFS: Number of large read operations=0  
HDFS: Number of read operations=9  
HDFS: Number of write operations=2  
Job Counters  
Data-local map tasks=2  
Launched map tasks=2  
Launched reduce tasks=1  
Total megabyte-milliseconds taken by all map tasks=6043750  
Total megabyte-milliseconds taken by all reduce tasks=1988250  
Total time spent by all map tasks (ms)=24175  
Total time spent by all maps in occupied slots (ms)=24175  
Total time spent by all reduce tasks (ms)=7953  
Total time spent by all reduces in occupied slots (ms)=7953  
Total vcore-milliseconds taken by all map tasks=24175  
Total vcore-milliseconds taken by all reduce tasks=7953  
Map-Reduce Framework  
CPU time spent (ms)=3670  
Combine input records=95  
Combine output records=4  
Failed Shuffles=0  
GC time elapsed (ms)=993  
Input split bytes=234  
Map input records=5  
Map output bytes=996  
Map output materialized bytes=65  
Map output records=95  
Merged Map outputs=2  
Physical memory (bytes) snapshot=528064512  
Reduce input groups=2  
Reduce input records=4  
Reduce output records=2  
Reduce shuffle bytes=65  
Shuffled Maps =2  
Spilled Records=8  
Total committed heap usage (bytes)=284164096  
Virtual memory (bytes) snapshot=5799043072  
Shuffle Errors  
BAD_ID=0  
CONNECTION=0  
IO_ERROR=0  
WRONG_LENGTH=0  
WRONG_MAP=0  
WRONG_REDUCE=0  
Streaming final output from hdfs:///user/maria_dev/tmp/mrjob/WordCount2.maria_dev.20180128.060813.499534/output.  
"a_to_n" 46  
"other" 49  
Removing HDFS temp directory hdfs:///user/maria_dev/tmp/mrjob/WordCount2.maria_dev.20180128.060813.499534...  
Removing temp directory /tmp/WordCount2.maria_dev.20180128.060813.499534...  
[maria_dev@sandbox ~]$
```

Question 10

The output of the program Salaries2.py is highlighted in red below

```
maria_dev@sandbox:~  
FILE: Number of read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=1669450  
HDFS: Number of bytes written=36  
HDFS: Number of large read operations=0  
HDFS: Number of read operations=9  
HDFS: Number of write operations=2  
Job Counters  
Data-local map tasks=2  
Launched map tasks=2  
Launched reduce tasks=1  
Total megabyte-milliseconds taken by all map tasks=7785250  
Total megabyte-milliseconds taken by all reduce tasks=2387250  
Total time spent by all map tasks (ms)=31141  
Total time spent by all maps in occupied slots (ms)=31141  
Total time spent by all reduce tasks (ms)=9549  
Total time spent by all reduces in occupied slots (ms)=9549  
Total vcore-milliseconds taken by all map tasks=31141  
Total vcore-milliseconds taken by all reduce tasks=9549  
Map-Reduce Framework  
CPU time spent (ms)=4770  
Combine input records=13818  
Combine output records=6  
Failed Shuffles=0  
GC time elapsed (ms)=1400  
Input split bytes=230  
Map input records=13818  
Map output bytes=129922  
Map output materialized bytes=96  
Map output records=13818  
Merged Map outputs=2  
Physical memory (bytes) snapshot=522235904  
Reduce input groups=3  
Reduce input records=6  
Reduce output records=3  
Reduce shuffle bytes=96  
Shuffled Maps =2  
Spilled Records=12  
Total committed heap usage (bytes)=275251200  
Virtual memory (bytes) snapshot=5798891520  
Shuffle Errors  
BAD_ID=0  
CONNECTION=0  
IO_ERROR=0  
WRONG_LENGTH=0  
WRONG_MAP=0  
WRONG_REDUCE=0  
Streaming final output from hdfs:///user/maria_dev/tmp/mrjob/Salaries2.maria_dev.20180128.064335.305800/output...  
"High" 442  
"Low" 7064  
"Medium" 6312  
Removing HDFS temp directory hdfs:///user/maria_dev/tmp/mrjob/Salaries2.maria_dev.20180128.064335.305800...  
Removing temp directory /tmp/Salaries2.maria_dev.20180128.064335.305800...  
[maria_dev@sandbox ~]$
```

Question 12

The first few lines of the output of program MovieCount.py is shown below

maria_dev@sandbox:~

Streaming final output from hdfs:///user/maria_dev/tmp/mrjob/MovieCount.maria_dev.20180128.065921.952784/output...

"1"	20
"10"	46
"100"	25
"101"	55
"102"	678
"103"	94
"104"	76
"105"	525
"106"	45
"107"	32
"108"	31
"109"	23
"11"	38
"110"	120
"111"	341
"112"	21
"113"	27
"114"	25
"115"	41
"116"	25
"117"	55
"118"	189
"119"	641
"12"	61
"120"	138
"121"	80
"122"	40
"123"	33
"124"	85
"125"	210
"126"	64
"127"	21
"128"	323
"129"	26
"13"	53
"130"	375
"131"	44
"132"	94
"133"	178
"134"	311
"135"	22
"136"	50
"137"	80
"138"	81
"139"	68
"14"	20
"140"	46
"141"	31
"142"	61
"143"	77
"144"	41
"145"	38
"146"	73